A Technology Shift towards Triangular Electrodes from Square Electrodes in Design of Digital Microfluidic Biochip

Piyali Datta[1], Amartya Dutta[1], Riya Majumder[1], Arpan Chakraborty[1], Debasis Dhal[2], and Rajat Kumar Pal[1]

[1]*Department of Computer Science and Engineering, University of Calcutta*
92, A. P. C. Road, Kolkata – 700 009, West Bengal, India
piyalidatta150888, dutta.kultu, riyamanai, arpanc250506, and *pal.rajatk@gmail.com*
[2]*Department of Information Technology, Triguna Sen School of Technology, Assam University, Silchar, Cachar – 788 011, Assam, India*
todevcse@rediffmail.com

***Abstract*—Digital Microfluidic Biochip (DMFB) is a ground-breaking invention in many areas of Microelectronics, Biochemistry, and Biomedical sciences. It is also known as 'Lab-on-a-Chip' for its performance as an alternative for laboratory experiments. Various types of diagnosis procedures are performed on it through a sequence of modular operations like sample preparation, mixing, and detection. Mixing is the most important operation of DMFB as the outcome of the experiment is almost dominated by mixing. A good mixing of corresponding sample and reagent gives proper result while an improper mixing leads to erroneous result, which may be the reason to discard the assay. So, our objective is the betterment of mixing operation. In this paper, we have proposed a design using equilateral triangular electrodes instead of square electrodes, maintaining all the constraints required to ensure safe droplet movement and other modular operations, while improvement of the mixing operation is the key design issue.**

Keywords: ***Digital microfluidics; Electrowetting on dielectric; Square electrodes; Triangular electrodes; Mixing; Detection; Assay; Principal diagonal.***

I. Introduction

Microfluidics based biochips combine Electronics with Biology, which open new application areas. These are used for different clinical diagnosis, massive parallel DNA sequencing, automated drug discovery, and real time bio-molecular recognition. Digital microfluidic biochip, which is a popular biomedical analytic instrument now-a-days, is based on Electro-Wetting-on-Dielectric (EWOD) theory [3]. Here discrete droplet of nanoliter volume can be moved on over the patterned array of electrodes. Due to dynamic reconfigurability, different types of operations like mixing, storing, etc. can be done in the same set of microfluidic cells in different time span. Discrete droplets can move through the electrodes under clock pulse in *digital manner*; hence the term *digital microfluidics* comes into picture. This approach exhibits dynamic reconfigurability, by which without affecting the functionality, the droplets can be relocated on other places of the array, consists of square electrodes [1,2].

In this paper we have proposed a design using equilateral triangular electrodes instead of square electrodes, maintaining all the constraints required to ensure safe droplet movement and other modular operations. This may be regarded as a technological shift where mixing operation is better performed than a set of square electrodes due to more zigzag path that enhances the scope and quality of mixing. As mixing is the most prior operation in case of diagnostic use of biochip, this is undoubtedly an improvised design procedure.

II. Preliminaries and Inherent Constraints

In this section, we briefly discuss on the working principle of DMFB, microfluidic architectural choice, droplet type, and the modular operations in the context of square electrodes. Electrowetting on dielectric is the phenomenon where electric field modifies the wetting characteristic of a polarizable and/or conductive liquid droplet in contact with a hydrophobic, insulated electrode [1,2,3]. According to the Lippmann-Young equation, the electric field generated due to the voltage gradient reduces the interfacial tension between the liquid and the surface, hence can manipulate the droplets [3]. Again droplets are sandwiched between two parallel plates, where the bottom is the chip surface holding the electrode array and the top is the continuous ground plate. The upper plate contains the continuous ground electrode of transparent, conductive indium tin oxide (ITO). Intermediate gap is filled with silicon fluid preventing droplet evaporation as well as reducing the contamination on the surface due to the use of various types of biological samples. The top and bottom surfaces are covered by a thin hydrophobic film, made by Teflon AF. This reduces the wet ability of a surface and improves the capacitance between the droplets and the solid electrodes [1,2,3].

Some basic definitions have been stated in the existing articles that comprise droplet formation, routing, mixing, and detection in the context of DMFB.

Droplet formation is a process of making droplet of a desired size from a source (of larger volume) of the material (sample or reagent or wash droplet) we like to dispose into the array perpendicularly by activation and deactivation of adjacent electrodes [1,2].

A *routing path* is the passageway that a droplet uses for its movement following adjacent cells of an array through a synchronized activation and deactivation of the electrodes [1,2,5].

A *mixer* is a module in an array where the most important task of mixing happens. Here different

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

sample(s) and reagent(s) come from their respective sources and mix for detection. This operation takes a maximum amount of time needed for an assay. A mixer can also be used for splitting of a droplet or dilute a sample droplet. A typical mixer takes 1000-2000 clock cycles [1,3,4].

Detection site is a small module usually formed by a single cell in the array that helps to detect the parameters present in a sample from its mixed droplet. Usually, the number of detection sites is not many (as it is a costly module) and their locations are also tentatively fixed [1,3].

An *assay* is a whole operation that includes creation of droplets, their routing, mixing, and detection (of a sample's state). We usually deploy an array of electrodes that are activated and deactivated in a preferred synchronized fashion, and all subsequent steps of an assay are tracked to meet the objective(s) affirmed by the assay [1].

Though we are supposed to satisfy all the constraints in isolation, maintaining all the constraints for some bioassay may introduce the problem of *cross contamination* [4]. This problem tells that an overlapping region may occur for two routing paths of two different droplets that are not supposed to mix.

Constraints in Performing Bioassay Operations:

A bioassay operation involves routing, mixing, detecting some parameters present in a sample, and many others. Naturally, some problem related constraints are there; some of which are fluidic constraint, electrode constraint, time constraint, and area constraint, as follows.

Fluidic constraint: During droplet routing, in static condition, at least one cell is supposed to be kept in between two electrodes containing two droplets to prevent unintended mixing. During movement of droplets following a particular direction, we may observe that at least a gap to two electrodes is must to avoid unwanted mixing. Hence static and dynamic fluidic constraints are introduced, as these are necessary for a pair of droplets for their minimum separation on a bioassay [4,6].

Electrode constraint: In case of pin constrained design, more than one electrode is controlled by a single pin. This may introduce unwanted effect of voltage on some electrodes, this imposes several constraints during routing. If we can make proper voltage assignment over the pins, correct movement of droplets can be guaranteed [1,2,5].

Timing constraint: Timing constraint in droplet routing is given by an upper bound on droplet transportation time. It is defined to have the proper synchronization among all the bioassay operations held in different modules [1,5].

Area constraint: In order to perform all the bioassays in a minimum chip area in view of all the abovementioned constraints, area constraints are introduced. A mixing region is supposed to be located in a proper position for utilization of total array area. So a design must support how capably a chip of some fixed area can be utilized [1].

III. Problem Formulation and Proposed Design

In this section we propose the design of equilateral triangular electrodes as well as we demonstrate droplet position, movement, and modification of fluidic constraint due to this replacement.

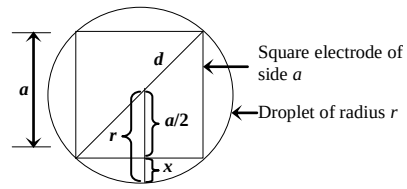


Fig. 1: Top view of a droplet placed on a charged square electrode.

A droplet moves from one electrode to the next due to the gradient of electrostatic force when the electrodes are deactivated and activated serially. To maintain the voltage level constant while shifting from square to triangular electrode, we consider the overlapped region of a droplet over its immediate neighbours where the droplet may move in one clock pulse, is fixed. Hence we achieve a relation between the sides of a triangular electrode with that of a square electrode, while keeping the side of the square electrode as the reference. Fig. 1 depicts the top view of a droplet of radius r placed on a charged square electrode. The portion (here we consider the part of the radius which is out of the electrode) of droplet outside the square is denoted by x, as shown in Fig. 1.

A. Calculation of the value of x:

Let the diagonal of the square = d and the side of the square = a. Hence, $d = \sqrt{2}a$, and the radius of the droplet (r) = $d/2 = a/\sqrt{2}$. Now, $a/2 + x = r = a/\sqrt{2}$; so, $x = 0.207a$. Here we consider the following two assumptions:

i) The volume of the droplet which we keep on a square electrode is same for a triangular electrode.
ii) The portion of the droplet beyond the side of a square electrode is same for the equilateral triangular electrode as well.

Satisfying the above assumptions, we next calculate the side of an equilateral triangular electrode which is most fitted for holding a droplet of radius r while the electrode is charged.

B. The side of the equilateral triangular electrode:

Let, the radius of the droplet which is kept on the equilateral triangle be r_1. From Fig. 2, $r_1 = (1/3)h + x$, where h is the height of the triangle. Now, $h = (\sqrt{3}/2)y$, where y is the side of the equilateral triangle. Therefore, $r_1 = (1/3)\times(\sqrt{3}/2)\times y + 0.207a = y/(2\sqrt{3}) + 0.207a$.

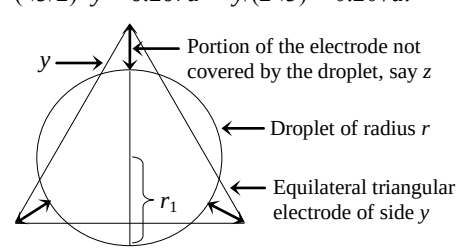


Fig. 2: Top view of a droplet placed on a charged equilateral triangular electrode.

As the volume of the droplet is same, keeping on both square and equilateral triangular electrode, the radii must be same for the above two cases. So, radius of the droplet on the equilateral triangular electrode (r_1) = radius of the

droplet on the square electrode (r). So, $a/\sqrt{2} = y/(2\sqrt{3}) + 0.207a$, and $y = 1.73a = \sqrt{3}a$. Hence, $y > a$, i.e., the length of one side of an equilateral triangular electrode is greater than the side of a square electrode. From Fig. 2, we can say that, $(2/3)h$ = radius of the droplet + length beyond the droplet up to a vertex of the triangular electrode (say z).

Hence, $(2/3)\times(\sqrt{3}/2)\times(\sqrt{3}a) = (a/\sqrt{2}) + z$ (putting the value of h), and $z = 0.3a$, where a is the length of a side of the square electrode.

The path of movement of the droplet through the triangular electrodes is *zigzag*, because the center of masses of two horizontally consecutive equilateral triangles are not at the same horizontal line, whereas this path is straight for the square electrodes, for the location of their center of masses at the same horizontal and vertical lines. Again the gap between two square electrodes is kept $a/2$ [1]. If the gap is greater than $a/2$, then two droplets can stay on two adjacent square electrodes (obviously in absence of voltage). This violates the fluidic constraints of droplet transportation on microfluidic biochip. If the gap is less than $a/2$, then there is a chance of unintentional mixing of two droplets on two adjacent square electrodes. We are to keep this value constant for triangular electrodes as well, as our objective is to provide the same voltage source as for the square electrodes, i.e., we consider the least gap between two triangular electrodes same as $a/2$.

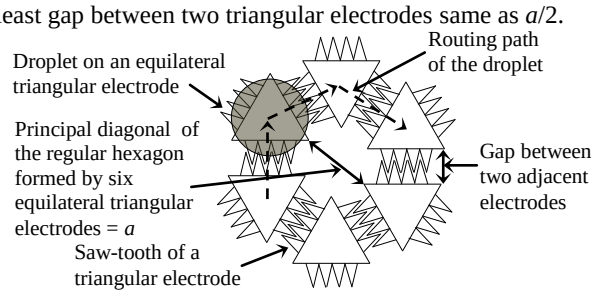


Fig. 3: Top view of a droplet on an equilateral triangular electrode. Here saw-teeth of adjacent electrodes are not short circuited.

The junction of four square electrodes forms a square of side $a/2$. Hence, the length of principal diagonal of the square formed by four adjacent square electrodes is therefore $a/\sqrt{2}$, whereas the junction of six equilateral triangular electrodes forms a regular hexagon of side $a/2$. So, the length of the principal diagonal = 2×(length of a side of the regular hexagon) = $2\times(a/2) = a$.

C. Fluidic constraints for triangular electrodes:

Throughout the droplet routing, a minimum spacing between droplets must be maintained to prevent the unintentional mixing, except the requirement of merging of droplets. Fig. 3 sows the top view of equilateral triangular electrodes with saw-teeth, holding one droplet, and the movement of droplet from one electrode to another electrode. For multiple droplet routes, droplet may intersect or overlap with each other. Hence, fluidic constraint rules are to be introduced to avoid the unintentional behaviour. At first we consider two droplets D_i and D_j. To avoid unintentional mixing, it is assumed that their initial locations at time t should not directly or diagonally adjacent to each other.

Let us represent the microfluidic array by 2D co-ordinate (X, Y). At time t, location of D_i is represented by $X_i(t)$ and $Y_i(t)$. We discuss the fluidic constraints for triangular electrodes in two phases, one for X-coordinate, and other for Y-coordinate.

Fluidic Constraints along *X*-Coordinate: This can be categorized into two parts: (a) Static fluidic constraints and (b) Dynamic fluidic constraints.

(a) Static Fluidic Constraints:

Rule 1: $|X_i(t) - X_j(t)| \geq 2$. So, the minimum spacing between two droplets at time t, on same horizontal row is at least two (i.e., two electrodes). Here in Fig. 4(a), $D_i(t)$ and $D_j(t)$ are on electrodes 1 and 3, respectively. To avoid unintentional mixing, the minimum spacing between them must be of two electrodes.

Rule 2: $|X_i(t+1) - X_j(t+1)| \geq 2$. At time t+1, the minimum difference between the X-coordinate of D_i and D_j must be two, such that D_i and D_j are not adjacent to each other. Here in Fig. 4(b), at time t+1, D_i goes from electrode 1 to electrode 2, and at the same time D_j moves from electrode 5 to electrode 4. So, the gap between the new locations of the two droplets at time t+1 is of two electrodes.

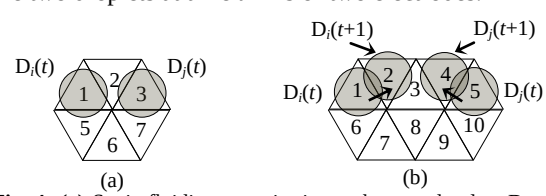


Fig. 4: (a) Static fluidic constraint is met by two droplets D_i and D_j along X-axis at time t. **(b)** Static fluidic constraint is met by two droplets D_i and D_j along X-axis at time t+1.

(b) Dynamic Fluidic Constraints:

Rule 3: $|X_i(t+1) - X_j(t)| \geq 2$. The least distance along X-coordinate for D_i at time t+1 and D_j at time t must be two, to avoid unwanted mixing. Similarly, $|X_i(t) - X_j(t+1)| \geq 2$ is also true. So, finally we can conclude that the minimum separation between two droplets situated in the same horizontal row is of two electrodes in both static and dynamic cases.

Fluidic Constraints along *Y*-Coordinate: Now we are concerned about the Y-coordinate of the droplets D_i and D_j. In the array of the equilateral triangular electrodes we notice that there are two types of triangular electrode as shown in Fig. 5.

If a droplet is placed on a *Type 1* electrode, then for this type of electrode the movement of droplet to the upward direction is restricted as movement of a droplet is only possible to the electrodes that are adjacent through an edge of the triangle. Similarly, if a droplet is placed on a *Type 2* electrode, then the movement of the droplet is restricted to the downward direction for the same reason. Again, according to the calculation of principal diagonal of the inter-electrode gap, which is hexagonal in shape, we know that there is a safe distance between a *Type 1* (*2*) electrode and the electrode just above (below) it. Hence, we may conclude that these two droplets may safely be kept in these two electrodes as they are not adjacent to each other with respect to fluidic constraints.

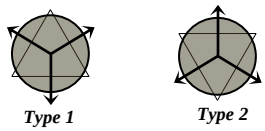


Fig. 5: Two types of electrode with droplet movement directions.

(a) Static Fluidic Constraints:

Rule 1: $Y_i(t)-Y_j(t) \geq 2$, if D_i is on a *Type 1* electrode, else $Y_i(t)-Y_j(t) \geq 1$, if D_i is on a *Type 2* electrode, where D_i is the droplet that has been considered to have higher *Y*-coordinate than the other. So, the least distance between the *Y*-coordinates of D_i and D_j depends on the position of the droplets. Here, in Fig. 6(a), D_i and D_j are on electrodes 7 and 11, respectively, keeping their safe separation.

Rule 2: At time t+1, Rule 1 must be maintained again depending on the relative positions of the droplets. In Fig. 6(b), at time t+1, D_i may go from electrode 7 to 6, and at the same time D_j may move from electrode 11 to 10. So the difference between the new locations of these two droplets at time t+1 is one electrode, disobeying the fluidic constraint; this often may lead to unintentional mixing.

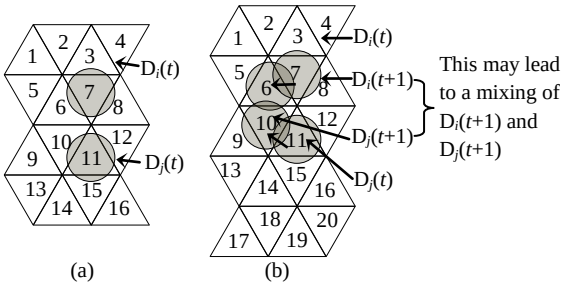


Fig. 6: (a) Static fluidic constraint is met by two droplets D_i and D_j along *Y*-axis at time *t*. **(b)** Static fluidic constraint is met by two droplets D_i and D_j along *Y*-axis at time t+1.

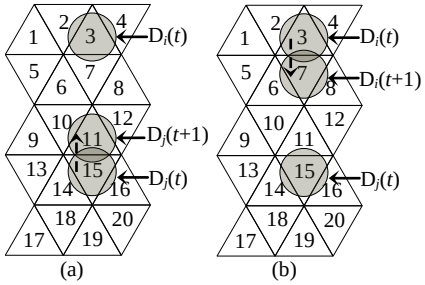


Fig. 7: (a) and **(b)** Dynamic fluidic constraints along *Y*-axis is met by two droplets D_i and D_j.

Table I: Table of comparison between square and triangular electrodes.

Type of electrode	Square electrode	Triangular electrode
Length of one side	a	$1.732a$
Overlapping length	$0.207a$	$0.207a$
Length of the principal diagonal	$a/\sqrt{2}$	a
Gap between two consecutive electrodes	$a/2$	$a/2$
Number of direct neighbours	4	3
Path measured between two electrodes	Manhattan distance	Number of electrodes passed by a droplet
Area of electrode	a^2	$1/2\times(3/2)a\times(\sqrt{3}a)=3\sqrt{3}a^2/4$

(b) Dynamic Fluidic Constraints:

Rule 3: Throughout the movement of one or more droplets, the static fluidic constraints at time t and t+1 are to be considered simultaneously that also satisfy the dynamic fluidic constraints as well, as shown in Fig. 7.

Table I shows a comparative study between a square and a triangular electrode citing some variables that are the most important issues to be considered in designing a chip.

VI. Applications and Importance

Digital microfluidic biochip is a revolutionary invention in modern science due to its versatile utilizations as well as reduced cost. It can perform enzymatic analysis (e.g., glucose and lactate assays), DNA analysis (e.g., PCR and nucleic acid sequence analysis), toxicity monitoring, and several biomedical and biochemical testing efficiently, which are of ultimate importance now-a-days. Droplet speed is high (up to 25 cm/sec for square electrodes) that makes the time of routing negligible. DMFB is of utmost importance as most of the electrolyte solutions work as biological samples or reagents that are utilized without any loss. DMFB is extremely energy efficient as it needs nanowatts or microwatts amount of power per movement of droplets. Among all the modular operations, the importance of mixing procedure is dominating, as it is the key of proper result of an assay performed on a biochip. A good mixing of sample and reagent offers desired result while an improper mixing leads to discard the whole assay. Triangular electrodes offer betterment in mixing as the accumulated droplets of samples and reagents are moved through a zigzag path and the scope of diffusion between the droplets is higher than moving through a straight path as happens in the case of square electrodes.

VI. Conclusion

In this paper, we have introduced a new design of digital microfluidic biochip which is a better alternative over square electrodes. In case of triangular electrodes, as the centers of the electrodes follow a zigzag path, hence the mixing operation becomes better, and as result we may consider it as a successful shift to a new design concept. The pin constrained design may be planned in a better way such that the number of pins is reduced than that of a chip containing square electrodes, which may decrease the required voltage making the chip more power efficient.

Reference

[1] Chakrabarty K. and F. Su, "*Digital Microfluidic Biochips: Synthesis, Testing, and Reconfiguration Techniques*," CRC Press, 2007.

[2] Chakrabarty K. and T. Xu, "*Digital Microfluidic Biochips Design Automation and Optimization*," CRC Press, 2010.

[3] Fair R. B., "Digital microfluidics: Is a true lab-on-a-chip possible?," *Microfluid Nanofluid*, Springer, vol. 3, pp. 245-281, 2007.

[4] Paik P., V. K. Pamula, M. G. Pollack, and R. B. Fair, "Electro-wetting based droplet mixers for microfluidic systems," *Lab on a Chip*, vol. 3, pp. 28-33, 2003.

[5] Su F., W. Hwang, and K. Chakrabarty, "Droplet routing in the synthesis of digital microfluidic biochips," *DATE*, pp. 323-328, 2006.

[6] Xu T. and K. Chakrabarty, "A droplet-manipulation method for archiving high throughput in cross-referencing based digital microfluidic biochips," *TCAD*, vol. 27, pp. 1905-1917, 2008.

Diffusion Tensor Based Global Tractography of Human Brain Fiber Bundles

Fahmida K. Chowdhury*, Eddie Jacobs, Jakir Hossen, Teddy Salan

The Department of Electrical and Computing Engineering, University of Memphis, Memphis, TN. USA 38152

*Email: fkchwdhr@memphis.edu

Abstract— **Tractography is a non-invasive process for reconstruction, modelling and visualization of neural fibers in the white matter (WM) of the human brain. It has emerged as a major breakthrough for neuroscience research due to its usefulness in clinical applications. In this research, we have investigated deterministic and probabilistic tractography approaches. We have evaluated the performance of different approaches on fiber bundle tracking using diffusion tensor imaging (DTI) data having multiple gradient directions. Experimental results show that global tractography is best for reconstruction of kissing and crossing fibers compared to deterministic tractography. We have also shown that DTI acquisition with a higher number of gradient directions provides better tracking results given limitations of acquisition and computation time.**

Index Terms— **Tractography, diffusion tensor imaging, white matter, gradient directions.**

I. INTRODUCTION

Diffusion tensor imaging (DTI) based tractography has been given much attention in recently due to its applications in several medical fields such as patients with acute stroke or brain tumors, neurodegenerative disorders including multiple sclerosis (hardening or thickening of tissue), epilepsy, Alzheimer disease, autism, and movement disorders [3]. DTI enables visualization and characterization of white matter fascicles in two and three dimensions. In general, it measures the displacement of water molecules on the micron scale and yields information about white matter (WM) fibers that pass within a voxel. Water molecules contain hydrogen nuclei (protons), which become aligned in a magnetic field. Motion or diffusion of water molecules is found to be much faster along the WM fibers than perpendicular to them [2]. The difference between the two motions (parallel and perpendicular to fibers, also termed diffusion anisotropy) is the basis of DTI.

DTI gradient directions are the list of vectors that describe the diffusion weighting directions for DTI acquisition on a MRI scanner. To observe the diffusion in all direction, many diffusion weighted images with diffusion weighting gradients are acquired in different directions. From these diffusion-weighted images, a 3D description of the direction for the diffusion of water within a voxel is inferred. Diffusion is anisotropic in nature and an ellipsoid is used to represent diffusion directions. These directions of water diffusion are tracked in DTI based tractography. Two different approaches to tractography [1] are considered in this research: deterministic and probabilistic. In deterministic tractography, the reconstruction of long neuronal pathways occurs in small successive steps by following the local direction of the fiber in a given voxel. This method is computationally efficient but minor errors in the determination of a local step may significantly affect the final path, making the method unstable. More recently, probabilistic approaches are predominant among tractography methods. Global tractography is a probabilistic approach. The global methods [8] try to reconstruct the fiber paths simultaneously while finding the configuration that minimizes the difference between the measured data and the reconstruction. This tracking is more stable in the presence of noise and imaging artifacts in the data.

In this research, we have compared the performances of deterministic and global tractography based on the simulation using various gradient directions data. We investigate the reconstruction performance considering crossing and kissing fibers and the effects of multi-gradient direction data. The background is discussed in section 2. Section 3 presents the description of the data sets. Methodology, experimental results, and the conclusion are discussed in sections 4, 5 and 6 respectively.

II. BACKGROUND

The development of magnetic resonance imaging (MRI) has led to the design of numerous imaging techniques. Among them DTI assesses white matter changes in the patients that are not normally seen on conventional MRI.

The relationship between the signal intensity of the diffusion weighted images S, diffusion sensitizing field gradient based on Stejeskal Tanner spin echo scheme and the signal value S_0 without the gradient is given in (1):

$$S = S_0 e^{-\gamma^2 G^2 \delta^2 (\Delta - \frac{\delta}{3}) D_{app}} \tag{1}$$

where γ is the gyromagnetic ratio of proton, δ and G represent the duration and the magnitude of the motion probing (or diffusion sensitizing field) gradient, Δ is the time between the centers of the pair of gradient pulses, and D_{app} is a scalar value called the apparent diffusion coefficient (ADC) which reflects molecular diffusivity under motion restriction such as fluid viscosity.

In DTI, the diffusion coefficient is a 3 ×3 symmetric positive semi-definite matrix as shown in (2).

$$\mathbf{D} = \begin{matrix} D_{xx} & D_{xy} & D_{xz} \\ D_{xy} & D_{yy} & D_{yz} \\ D_{xz} & D_{yz} & D_{zz} \end{matrix} \tag{2}$$

The largest eigenvector of the diffusion tensor as shown in Fig. 1 is assumed to be oriented parallel to the local fiber

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

tracts, which can be reconstructed in a brain dataset by following paths in the estimated eigenvector field.

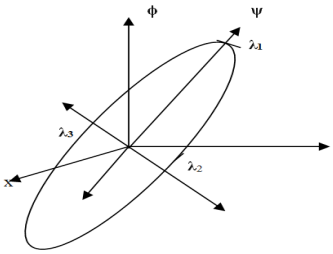


Fig. 1 Diffusion tensor represented by an ellipsoid ($\lambda_1 \geq \lambda_2 \geq \lambda_3$)

One of the most important factors in DTI acquisition is gradient direction. At least 6 gradient directions are required to calculate the diffusion tensor [4]. If gradient directions are increased, more diffusion weighted images are used to calculate the diffusion tensor, resulting in more accurate tensor estimation but much longer imaging time [7].

A long thin ellipsoid means very good diffusion for water along the long axis of that ellipsoid. The mean diffusivity (MD) and fractional anisotropy (FA) as shown in (3) and (4) respectively are the most widely used indices of DTI for representing the motional anisotropy of water molecules, being sensitive to the presence and integrity of WM fibers [11][12]. Color coded FA maps are the way to show the directional information embedded in DTI.

$$\mathrm{MD} = \frac{\lambda_1+\lambda_2+\lambda_3}{3} \tag{3}$$

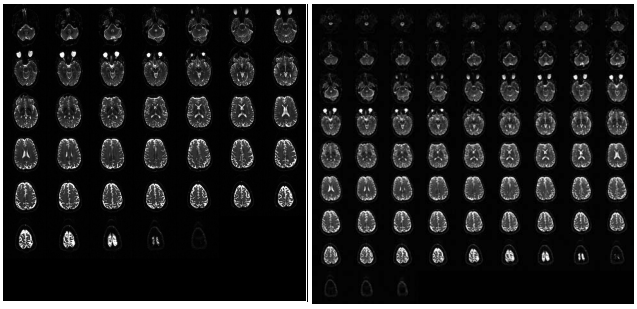
$$\mathrm{FA} = \sqrt{\frac{3}{2}} * \sqrt{\frac{(\lambda_1-D)^2+(\lambda_2-D)^2+(\lambda_3-D)^2}{\lambda_1^2+\lambda_2^2+\lambda_3^2}} \tag{4}$$

These indices help to detect brain diseases. For example, in case of pediatric multiple sclerosis, the value D increases and FA decreases compare to their normal value [9] [10].

Deterministic tractography methods are primarily based upon streamline algorithms where the local tract direction is defined by the major eigenvector of the diffusion tensor [1]. Crossing and kissing white matter tracts create a significant challenge for reconstruction. These limitations can be solved using global tractography.

III. DATA SETS

Data provided by St Jude children's research hospital was used for this research. They have diffusion MRI scanners of 1.5 and 3 Teslas. Samples of the DTI data sets provided are shown in Fig. 2.



(a) (b)
Fig. 2 Example data (a) 12 directions (b) 20 directions

We have conducted research on four data sets: (1) 12 gradient directions, 4 averages (40 slices and each 3 mm thick) (2) 20 gradient directions, 3 averages (75 slices and each 2 mm thick) (3) 30 gradient directions, 2 averages (70 slices and each 2 mm thick), and (4) 64 gradient directions, 2 averages (96 slices and each 1.7 mm thick).

IV. METHODOLOGY OF TRACTOGRAPHY

We have done the simulation for both deterministic and probabilistic tractography. There are various deterministic and probabilistic tracking techniques such as connect the voxel [13], Fiber Assignment by Continuous Tracking (FACT) [11], path- integral method [14], Gibbs tracking [5], spin glass model [15]. FACT and Gibbs tracking techniques have been considered for our simulation. Camino [16] and DTI & fiber tools [16] are used for simulation. The simulation procedure is discussed step by step in the following sections.

A. Pre-processing

For both of the tractography techniques, the first step in processing the data is ordering and sorting. In this step artifacts and the quality of data [6] are controlled. At first, we have to change the raw image format because medical imaging supports some specific image format. After that the data was pre-processed using several tools. Pre-processing involved artifact removal, skull stripping and reformatting.

B. Diffusion Tensor Calculation

Voxel based single and multi diffusion tensor models for the orientation distribution function (ODF) were fit to the data by Camino. After calculating FA and MD, we obtained the orientation distribution function (ODF) as shown in Fig. 3. It is used to describe the directionality of multimodal diffusion in regions with complex fiber architecture present in the brain. The maximum diffusivity is normalized to voxel size. The color indicates the direction of maximum diffusivity i.e. red means left to right, green indicates anterior-posterior and blue for superior-inferior. This step is also performed in Gibbs tracking tensor calculation but is done only for those pixels having signal intensity above the threshold value 40 in the average image.

C. Deterministic Tractography

The tractography process was done in three steps: region of interest (ROI) selection, track, and visualization [6]. ROI was selected using the ITK snap tool [16]. Corpus callosum (CC) is the largest white matter structure in brain that connects the left and right cerebral hemispheres. So selecting CC as a ROI will give the maximum number of fiber connections. Using the ROI, Camino tracked the fiber connections and the visualization is done by the Paraview tool.

D. Global Tractography

Gibbs tracking is a powerful global tractography method. This method includes three main steps [5]: (1) Creation of a trial fiber configuration, (2) Calculation of the corresponding DW signal (simulated signal), and (3) Adjustment of the trial configuration with the experimentally measured signal to minimize the difference.

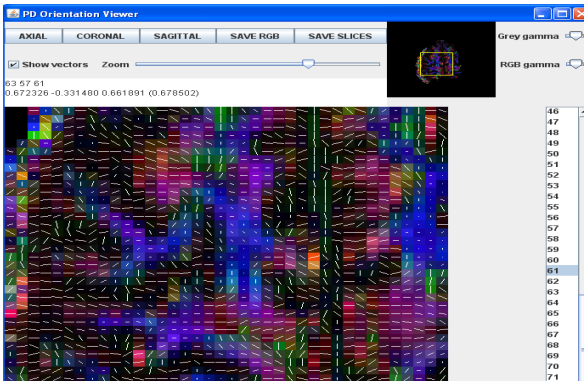


Fig. 3 Orientation distribution function of deterministic tractography.

The building elements of reconstructed fibers are small straight cylinders whose length, position and orientation can vary continuously. These cylinders are re-oriented through a process similar the distribution of states in the cooling of an ideal gas. The reconstruction begins at very high "temperature" where the cylinders are randomly distributed in the space occupied by the white matter as shown in Fig 4. The interaction between them results in building long fiber chains with decreasing temperatures. The neuronal fibers start and end on predefined surfaces, for example at the boundary with the grey matter. Each cylinder contributes a signal typical for parallel fibers to all voxels it crosses [5].

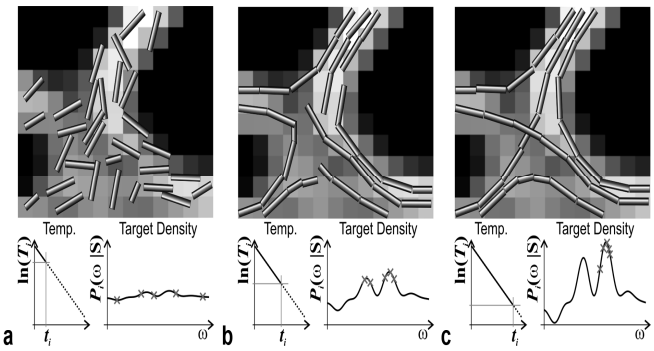


Fig. 4 Principle of Gibbs tracking method; (a) at high temperature (b) at lower temperature (c) temperatures close to zero. [5]

The sum of such contributions is the anisotropic part of the signal. This described method is known as the Gibbs Tracking [5], keeping in mind its close analogy to statistical physics. On the other hand, the described simulation can be considered as a Bayesian approach based on a spatial point processes. The interactions between the cylinders represent the a priori probability and the likelihood function represents the similarity of the measured signal and simulated signal.

Before the visualization step, the diffusion tensor images are provided with the tracking parameters. The tracking can be started by dense reconstruction that counts many fibers and takes in the range of 12-24 hours of computation or sparse reconstruction that shows fewer fibers and has reconstruction times around 1 hour.

After tracking completion, fibers are sent to a fiber viewer to show the fiber distribution of the whole brain, the spatial diffusion distribution, and the fiber distribution in a single slice.

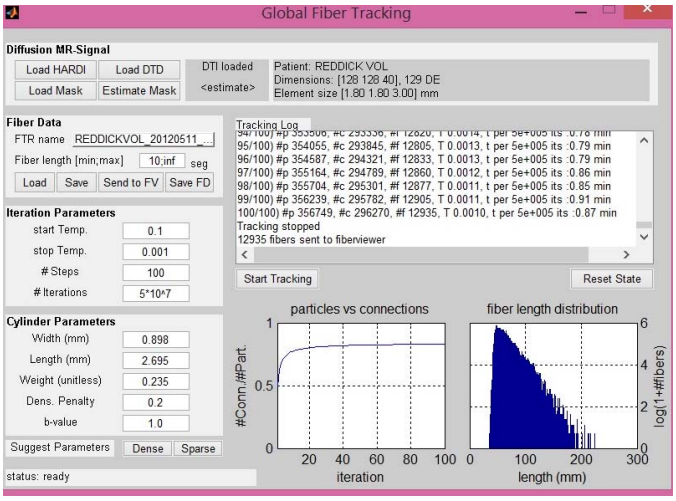


Fig. 5 Gibbs tracking of 12 D gradient directions using DTI & Fiber tools

V. EXPERIMENTAL RESULTS

We obtained the fiber trajectories after completing the deterministic and global tractography simulations. All the prepocessing and tensor calculation steps of simulation are run on an intel PC core i5, 1.8 GHz and 8 GB RAM with Windows 8™ as an operating system. Complete simulation of 12 directions data is possible on a laptop with the configuration as mentioned for prepocessing steps. Simulation of 20, 30 and 64 directions data sets are run on a high performance computing platform with 512 GB RAM.

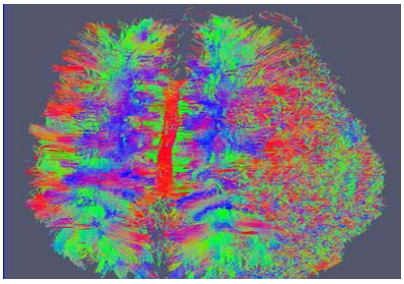
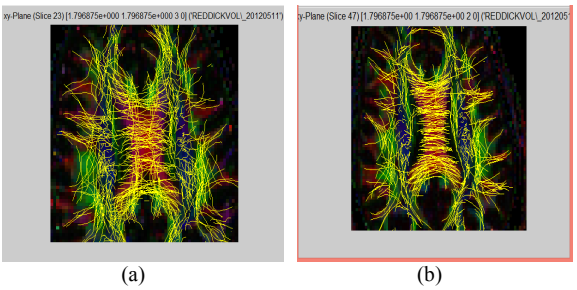


Fig. 6 Deterministic tractography

Visualization of fibers by deterministic tractography is shown in Fig. 6. A microscopic observation by an expert points out the limitation of this technique. The reconstruction does not work well where fiber crossing and kissing occur, though it takes only couple of hours for the simulation to execute. The tractography results for the Gibbs tracking method are shown in Fig. 7 and 8.

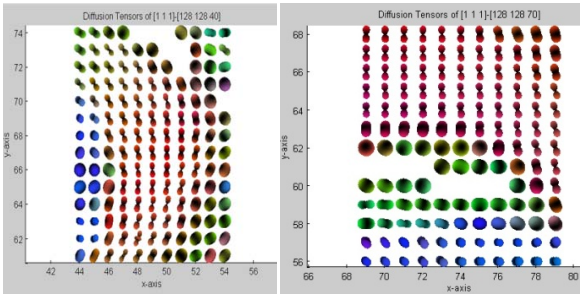
Fig. 7 Fiber tract for 20 D



(a) (b)

Fig. 8 Fiber cut (a) 12 D (b) 20 D

Fiber tract shows a projection of the complete actual fiber subset and fiber cut indicates the curve segments inside the actual slice of the current fiber structure subset. Spatial diffusion distribution represents diffusivity of the tensor in a particular position or region of interest as an ellipsoid as shown in Fig. 9. The maximal diffusivity is normalized to the voxel size. The color indicates the direction of diffusivity of the spatial distribution inside the current voxel.



(a) (b)

Fig. 9 Spatial distribution Red: left to right; Green: Anterior- posterior; Blue: Superior to inferior (a) 12 D (b) 30 D

For different positions or slices, we get different FA and MD values. The value of FA is normalized from 0 to 1. FA close to 0 indicates isotropic diffusion and anisotropic nature increases as FA values tend to 1. Here we have taken positions near to corpus callosum for different gradient directions and their corresponding FA and MD values are shown in Table 1.

TABLE I
COMPARISON OF FA AND MD VALUES FOR DIFFERENT GRADIENT DIRECTIONS

Gradient directions	Position	FA	MD
12 D	[67 68 23]	8.376×10^{-1}	6.730×10^{-4}
20 D	[66 70 47]	8.269×10^{-1}	6.686×10^{-4}
30 D	[67 69 43]	8.184×10^{-1}	6.604×10^{-4}
64 D	[67 64 47]	8.456×10^{-1}	6.292×10^{-4}

TABLE II
NUMBER OF FIBERS VS GRADIENT DIRECTIONS.

Gradient directions	Reconstruction process	Number of fibers	Time (Hours)
12 D	Sparse	12935	0:26
	Dense	80165	10:25
20 D	Sparse	13527	0.30
	Dense	83454	11:35
30 D	Sparse	13457	0:35
	Dense	88835	12:45
64 D	Sparse	14597	0:42
	Dense	99491	14:18

Tracking is started by selecting the required parameters like cylinder parameters, starting and ending temperature etc. An example of parameter sets for sparse reconstruction is provided in Fig. 5. In dense reconstruction, the weight of the cylinder was set to a low value e.g. 0.059. It can produce more fibers compared to the sparse process though it takes long computation time. Table 2 shows how the number fibers found are increased with an increase in gradient directions.

VI. CONCLUSION

Diffusion tensor imaging is a powerful tool for the visualization of white matter structures. Although deterministic approaches are computationally efficient, they have some limitations which can be overcome by global tractography. The characteristic features of the traced neural fiber networks, such as structure, topology, and connectivity, can serve as biomarkers for various dementias and brain diseases, and can be used in various clinical applications. It is shown that gradient directions have positive effect on tracking though the time requirement introduces artifacts in the data due to patient motion which limits the clinical application. There should be a trade off between the number of gradient directions and the specific clinical application. Developing a tractography algorithm to overcome difficulties associated with the current techniques is a topic of our future research.

ACKNOWLEDGMENT

We are grateful to Dr. Gene Reddick from St. Jude Children's reasearch hospital for the dataset, continuous support and comments on results.

References

[1] Derek K. Jones, *Diffusion MRI: Theory, Methods and Applications*, Oxford University Press, UK, Sep-2012.
[2] Basser PJ, Pierpaoli C, *A simplified method to measure the diffusion tensor from seven MR Images,* Magn Res Med 39:928–934, 1998.
[3] Mark A. Horsfield and Derek K. Jones, *Application of diffusion-weighted and diffusion tensor MRI to white matter diseases-a review*, NMR biomed, 15(7-8):570-7, Nov-Dec. 2002
[4] H. Nia, V. Kavcicb, T. Zhuc, S. Ekholma and J. Zhonga,c, *Effects of Number of Diffusion Gradient Directions on Derived Diffusion Tensor Imaging Indices in Human Brain,* AJNR 27: 1776-1781, Sept. 2006 .
[5] Kreher BW., Mader I. et al.," *Gibbs tracking: a novel approach for the reconstruction of neuronalpathway,"* Wiley Inter Science, 2008.
[6] Soares, J. M., Marques, P., Alves, V., and Sousa, N, *A hitchhiker's guide to diffusion tensor imaging*. Front. Neurosci. 7:31. doi: 10.3389, 2013.
[7] Na Zhang, Zhen-Sheng Deng, Fang Wang and Xiao-Yi Wang, *The effect of different number of diffusion gradients on SNR of diffusion tensor-derived measurement maps*, J. Biomed. Sci and Engg, 96-101, Feb 2009.
[8] Marco Reisert, Irina Mader, Valerij Kiselev et. al, *Global fiber reconstruction becomes practical,* NeuroImage 54, 955–962; (2011).
[9] Mara Cercignani, Matilde Inglese et. Al.,*Mean diffusivity and fractional anisotropy histograms of patients with multiple sclerosis,* AJNR 2001, 22:952-958
[10] M.S. Vishwas, B. C. Healy et al, "*Diffusion tensor analysis of pediatric multiple sclerosis and clinically isolated syndromes,"* AJNR 2013, 34:417-423
[11] Mori S, Crain B, Chacko VP et. Al,, *"Three dimensional tracking of axonal projections in the brain"* Ann Neurol 45:265–269, 1999.
[12] S. Mori, and P.C.van Zijl. "*Fiber tracking: principles and strategies–a technical review*". NMR Biomed, 15(7-8): 468-480, 2002.
[13] Thomas E. Conturo et al., *"Tracking neuronal fiber pathways in the living human brain"* Neurobiology, Applied Physical Sciences, Vol. 96, pp. 10422–10427, August 1999
[14] Peter J. Basser, et al., *"In Vivo Fiber Tractography Using DT-MRI Data,"* Magnetic Resonance in Medicine 44:625–632 (2000).
[15] Fillard P, Poupon C, Mangin JF., "A novel global tractography algorithm based on an adaptive spin glass model," Med Image Comput Comput Assist Interv. 2009; 12 (Pt 1):927-34.
[16] https://tbi.ci.uchicago.edu/tools

3D QUANTITATIVE BRAIN TUMOR GROWTH MODEL BASED on CELL PROLIFERATION and DIFFUSION

Sohana Tanzeem,[1,*] Wilburn E. Reddick,[2] and Khan M. Iftekharuddin[3]

[1]Electrical and Electronic Engineering Department, East West University
Plot No-A/2, Jahurul Islam City, Aftabnagar, Dhaka-1219, Bangladesh.

[2] Department of Radiological Sciences, St. Jude Children Research Hospital, Memphis, Tennessee, USA.

[3] Electrical and Computer Engineering Department, Old Dominion University, USA.

[*]stanzeem@ewubd.edu

Abstract— **The focus of this work was to develop a 3D mapping of brain tumor (glioma) growth based on cell proliferation and diffusion. In this mathematical model, we incorporated high resolution brain tissue maps (white and gray matter) from an anonymized pediatric patient and initialized the model with a single voxel seed point of tumor with a Gaussian distribution. We used this model to investigate the ratio of growth rate to the diffusion coefficient (ρ/D) which determines the proportion of tumor that is detectable. After expansion of the tumor growth model to three dimensions and solving the differential equations for our specific starting conditions, we performed several simulations to assess tumor growth patterns. After observing the performance of the model at varying time points across a one year time frame with different values for ρ/D, we ascertained that the tumor diffused more rapidly than the cell proliferated for a short period of time followed by an exponential growth in detectable tumor size.**

Index Terms—**Proliferation, Diffusion, Tumor.**

I. Introduction

Glioma are major intracranial tumors that constitute more than 50% of all primary brain tumors and are one of the major causes of morbidity and mortality in the world [1]. They are thought to arise from the supporting glial cells of brain that have lost or ceased to respond to normal growth regulatory mechanism, presumably through mutations and/or altered gene expression. One of the fundamental problems in treating glioma is their highly diffusive and infiltrative nature that makes it difficult to differentiate between healthy brain tissue and the tumor boundary as seen in various conventional imaging modalities [2]. In the tumor mass model [3], it is shown that tumor growth follows the 'universal' growth law curve which is exponential in nature. The basic rate of tumor growth can be calculated by deducting the tumor cell death from the rate of tumor cell birth [4]. A tumor growth model is proposed based on diffusion and cell proliferation [5].

Our hypothesis is that 3D diffusion mapping will provide a more sensitive method to better delineate tumor growth patterns by focusing on the extent of tumor invasiveness. The focus of our work is to develop a 3D mapping of brain tumor (glioma) growth based on cell proliferation and diffusion. The universal growth curve [3] gives us the idea for using exponential growth in our 3D model. The 3D growth model of PK Burgess et al. [6] incorporates tumor radius. A radius is defined there at which the glioma is usually fatal and then ρ and D are investigated as indicators of how long it may take the tumor to reach this critical radius. But actual tissue segmentation has not been used here for the image sets. Uniform diffusion properties have been assumed. The situation after the tumor has destroyed the underlying neuroanatomical structure has been observed but it has not been defined if it is in gray or white matter. However, in our 3D model, we have defined human brain as inhomogeneous considering two types of diffusion coefficients in gray and white matter. Also, we have considered a specific tumor cell density as a threshold of detection for MR images.

II. Method

We propose a 3D brain tumor growth model based on cell proliferation and diffusion that incorporates actual tissue mapping of white and gray matter from a pediatric subject. The tumor seed is assumed to be a Gaussian distribution with a mean of zero and a standard deviation of one (measured in millimetres). We begin by checking the performance of the model in 1D.

A. The Mathematical Model

Original works from Burgess et al. [6] Cruywagen et al. [7], Tracqui et al. [8] and Woodward et al. [9] develop a mathematical model of glioma growth based solely on cell proliferation and diffusion.

$$\frac{dc}{dt} = \nabla.(D\nabla c) + \rho c \qquad (1)$$

where $c(x,t)$ is the number of cells, ρ (1/day) represents the net rate of growth of cells including proliferation and death (or loss), D (mm^2/day) is the diffusion coefficient of cells in brain tissue and ∇ represents the spatial gradient.

The diffusion term describes the active migration of the glioma cells using a simple Fickian diffusion where cells move from regions of higher to lower densities.

We propose to solve equation (1) through equations (2) and (3).

$$\frac{dc}{dt} = D\nabla^2 c + \rho c \qquad (2)$$

$$c = T(t)R(x,y,z) \qquad (3)$$

First, we assume that diffusion D is homogeneous in each region and c is the number of cells which is a function of time $T(t)$ and position $R(x,y,z)$.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

The 3D model is developed as follows.

$$c(x,y,z,t) = c_0 \frac{1}{\Pi^{3/2} G^3} . e^{-\frac{x^2+y^2+z^2}{G^2}} . e^{\rho t} \quad (4)$$

where c_0 is the initial number of cells and

$$G^2 = a^2 \left(1 + \frac{4Dt}{a^2} \right) \quad (5)$$

where a is the standard deviation for the seed point.

We have assumed here that tumor growth occurs in two phases ([7]-[9]). In the first phase, the tumor cells only proliferate to form a small dense lesion and only at some later time the tumor cells begin to diffuse. The growth model is applied only in the second phase. Without this assumption, the model simulates the case of gliomatosis cerebri that is tumor cells can be found throughout the brain but no single bulk tumor is identifiable [5]. One special characteristic of the analysis is the use of a zero diffusion coefficient in all voxels not segmented as gray or white matter. This constrains the tumor from growing outside the head or into the ventricles.

B. Parameter Estimation

Parameter estimates in any realistic mathematical model are crucial. Doubling times for glioma have been estimated at 1 week to 12 months covering the range of high to low grade glioma [10]. To relate linear velocity of the detectable tumor margin, v, with the proliferation rate , ρ, and a random walk diffusion coefficient D, we used Fisher's approximation $D \approx \frac{v^2}{4\rho}$ ([6],[11]). This approximation comes from the observation that a population governed by growth and diffusion alone expands at a rate of $2\sqrt{\rho D}$ for a large period of time. The diffusion coefficients $D_w = \frac{v_w^2}{4\rho}$ and $D_g = \frac{v_g^2}{4\rho}$ are assigned the experimentally observed linear velocities v_g and v_w in gray and white matter, respectively ([7]-[9]). From the CT scans, the speed of advance of the tumor margin across the corpus callosum (white matter) is two to three times as fast as that in (predominately) gray matter. We therefore, estimate $v_w > 2.2 v_g$ at 0.18 mm/day and $D_w > 5 D_g$ at 0.22 mm^2/day. Parameter values are shown in Table I.

In our model, we have used empirically estimated parameters. We get the idea as to what the parameter values usually can be in case of high and low grade glioma from Swanson et al.'s model [5]. We also get the idea of the ratio between the diffusion coefficients of gray and white matter, but we considered two more empirical ideas for our estimation. First, we have used a pediatric brain which is structurally different than the adult one. So the diffusion coefficient will be smaller and the growth rate for the pediatric glioma should be faster than the adult glioma. Second, Swanson et al. [5] derived their parameters using contrast enhanced CT images, but our model is based on MR images. Therefore, detection threshold should be different in our case.

TABLE I
ESTIMATED PARAMETER VALUES

	Swanson et al. [5]	**High grade glioma**	**Low grade glioma**
Doubling Time (days)	56	28	56
Growth rate, ρ(/day), corresponding to doubling time	0.012	0.036	0.018
Linear velocity of the detectable tumor margin in gray matter, v_g (mm/day)	0.08	0.046	0.018
Diffusion coefficient in gray matter, D_g (mm^2/day)	0.13	0.015	0.005
Linear velocity in white matter, v_w (mm/day)	0.176	0.08	0.03
Diffusion coefficient in white matter, D_w (mm^2/day)	0.65	0.045	0.015

III. RESULTS

We investigated the performance of this model under several experimental conditions and reported those observed results. These investigations began with a simple 1D implementation to gain more experience with the affect and interaction of the variable parameters. Next, we implemented the full 3D version of the model in Matlab using an actual tissue map obtained from a pediatric patient treated for acute lymphoblastic leukaemia at St. Jude Children's Research Hospital, Memphis, Tennessee, USA. Diffusion and proliferation rates were chosen to model both high grade and low grade glioma and tested for seed points arising from both white matter and gray matter.

The first study conducted began with a 1D investigation of tumor cell density changes with time. In an early phase of growth, diffusion is the dominant factor for a period of time. After that time, the situation is reversed, with cell proliferation becoming the more dominant factor. Figures 1 and 2 demonstrate the results for 1D model experiment.

We initialized a 1 mm^3 Gaussian seed voxel in white matter of a segmented image and observed the growth of tumor (both high and low grade) in serial time steps from three orthogonal views. Figures 3 and 4 demonstrate the results.

Later we initialized a 1 mm^3 Gaussian seed voxel in gray matter of a segmented image and observed the growth of tumor (both high and low grade) in serial time steps from

three orthogonal views. Figures 5 and 6 demonstrate the results.

IV. Discussions

The general observations from figures 3, 4, 5 and 6 are as follows. First, in the high-grade case, ρ / D results in a larger portion of the tumor remaining detectable relative to the low grade case. Second, high-grade glioma take less time to diffuse than low-grade. Third, detectable tumor was found to spread faster and to become larger when arising in white matter than gray matter

V. Conclusions

In the initial preliminary simulations conducted in this study using the tumor growth model, the three key observations as discussed in section IV were consistent with clinical experiences.

We have implemented a tumor growth model in 3D which used real patient 3D tissue maps from an MR imaging set. This use of real imaging from the patient will be important for future testing in actual brain tumor patients since their distribution of white and gray matter (D_w, D_g) will be different for each patient and needs to be based on actual imaging. Since our model is truly 3D, the results can be overlaid on patient imaging and viewed in orthogonal planes. The model we have implemented is a necessary and essential first step in any investigation of more elaborate tumor models.

Acknowledgement

I would like to convey my gratitude to Dr. Anisul Haque, Professor, Department of Electrical and Electronic Engineering of East West University for helpful discussion.

References

[1] T. Roose, S.J. Chapman, and P.K. Maini, "Mathematical models of avascular tumor growth", *SIAM Review,* Vol. 49, No.2, pp. 179-208, 2007.

[2] C. Hogea, C. Davatzikos, and G. Biros, "An image-driven parameter estimation problem for a reaction-diffusion glioma growth model with mass effects", *J Math Biology*, Vol. 56, no. 6, pp.793-825,2008.

[3] C.Guiot, P.G. Degiorgis, P.P. Delsanto, P. Gabriele, and T.S. Deisboeck, "Does Tumor Growth follow a 'universal law'?", *J Theoretical Biology*, Vol.225, No. 2, pp. 147-151, 2003.

[4] Alexander Herman, "Towards a general model for solid tumor growth", SFI REU Project, Wesleyan University, 2002.

[5] K.R. Swanson, E.C. Alvord Jr, and J.D. Murray, "A quantitative model for differential motility of gliomas in gray and white matter", *Cell Proliferation*, Vol. 33, No. 5, pp. 317-329, 2000.

[6] P.K. Burgess, P.M. Kulesa, J.D. Murray, and E.C. Alvord Jr., "The interaction of growth rates and diffusion coefficients in a three dimensional mathematical model of gliomas", *J Neuropathology and Experimental Neurology*, Vol. 56, No. 6, pp. 704-713, 1997.

[7] G.C. Cruywagen, D.E. Woodward, P. Tracqui, G.T. Bartoo, J.D. Murray, and E.C. Alvord Jr., "The modeling of diffusive tumors", *J Biological Systems*, Vol. 3, pp. 937-945, 1995.

[8] D.E. Woodward, J. Cook, P. Tracqui, G.C. Cruywagen, J.D. Murray, and E.C. Alvord Jr., "A mathematical model of glioma growth: the effect of extent of surgical resection.", *Cell proliferation*, Vol. 29, No. 6, pp. 269-288, 1996.

[9] P. Tracqui, G.C. Cruywagen, D.E. Woodward, G.T. Bartoo, J.D. Murray, and E. C. Alvord Jr., "A mathematical model of glioma growth : the effect of chemotherapy on spatio-temporal growth", *Cell Proliferation*, Vol. 28, No. 1, pp. 17-31, 1995.

[10] E.C. Alvord Jr. and C.M. Shaw, "Neoplasms affecting the nervous system in the elderly", In: *The Pathology of the Aging Human Nervous System*, S. Duckett ed., pp.210-281, Philadelphia: Lea& Febiger, 1991.

[11] JD Murray (1993), *Mathematical Biology*, Heidelberg : Spinger-Verlag.

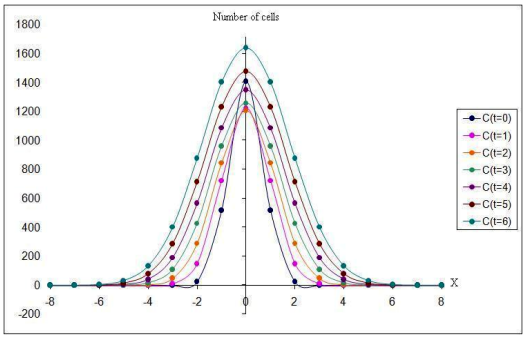


Fig. 1 Change of intensity and spread of cells with time (0-6 days). High grade glioma (ρ=0.036/day, Dw=0.045 mm2/day, Dg=0.015 mm2/day)

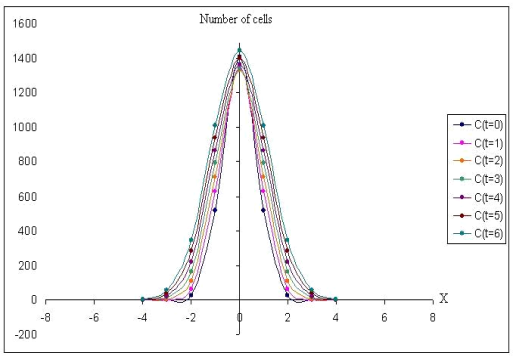


Fig. 2 Change of intensity and spread of cells with time (0-6 days). Low grade glioma (ρ=0.018/day, Dw=0.015 mm^2/day, Dg=0.005 mm^2/day)

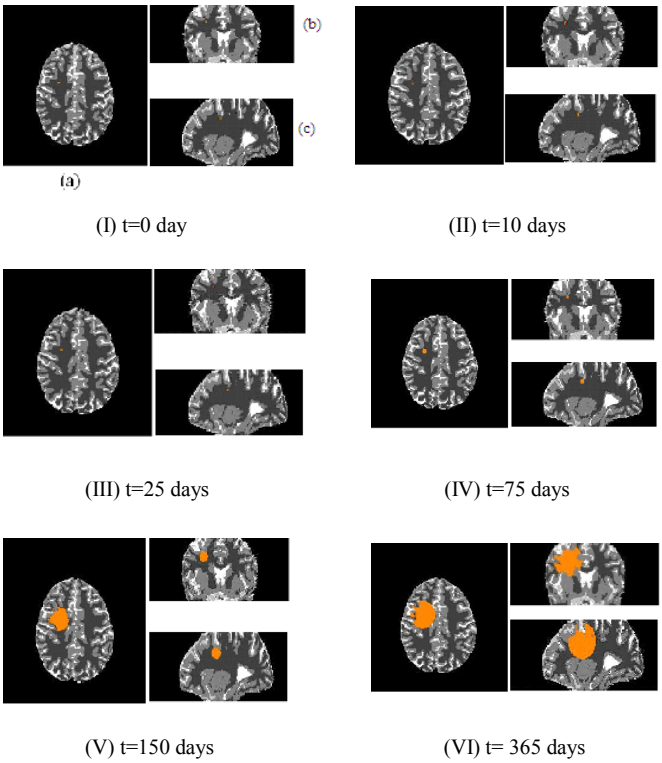
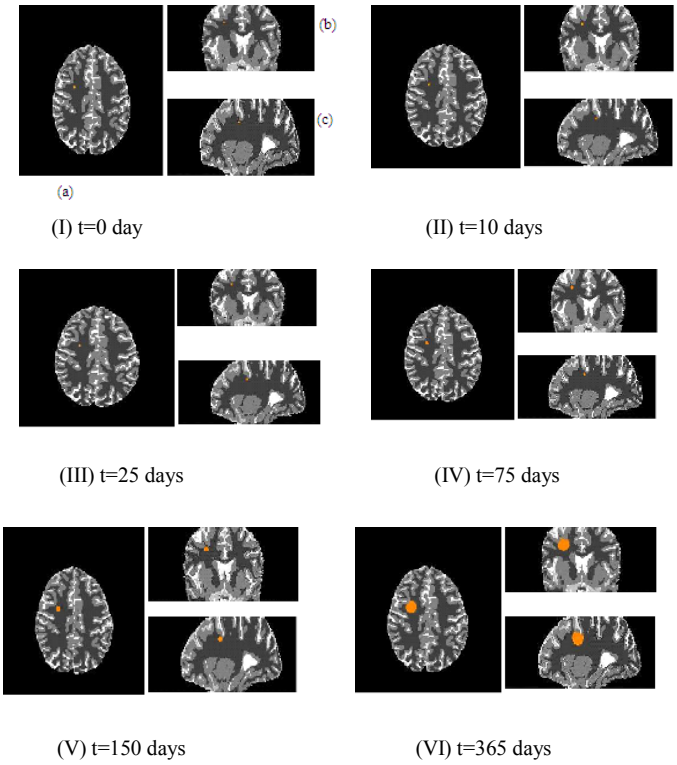


Fig.3 (I)-(VI): High-grade glioma (ρ=0.036/day, D_w=0.045mm^2/day, D_g=0.015 mm^2/day) from three orthogonal views while seed voxel was initiated in white matter. a) transverse, b) coronal , c) sagittal plane.

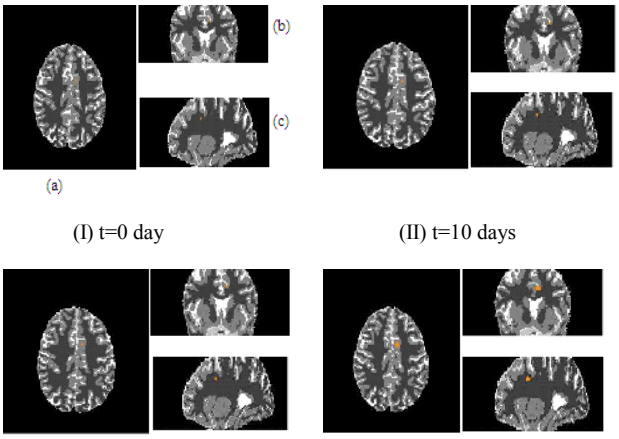


(I) t=0 day (II) t=10 days

(III) t=25 days (IV) t=75 days

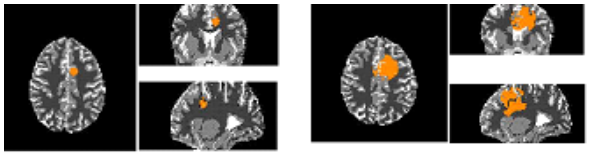
(V) t=150 days (VI) t=365 days

Fig.4 (I)-(VI): Low-grade glioma (ρ=0.018/day, D_w =0.015mm^2/day, D_g =0.005 mm^2/day) from three orthogonal views while seed voxel was initiated in white matter. a) transverse, b) coronal , c) sagittal plane.



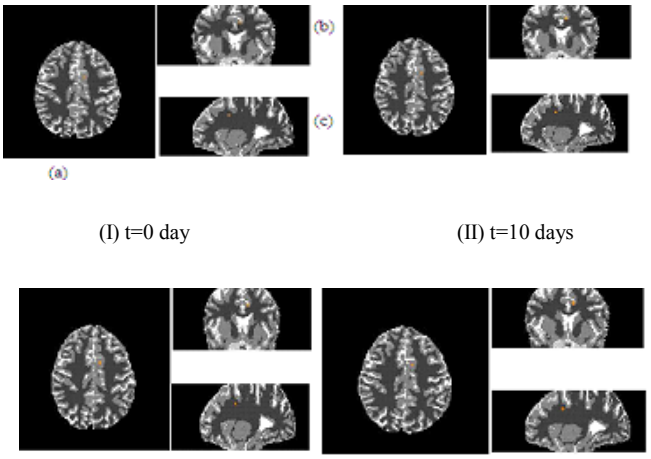
(I) t=0 day (II) t=10 days

(III) t=25 days (IV) t=75 days



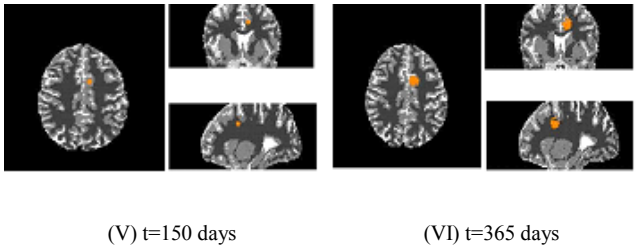
(V) t=150 days (VI) t=365 days

Fig.5 (I)-(VI): High-grade glioma (ρ=0.036/day, D_w=0.045mm^2/day, D_g =0.015 mm^2/day) from three orthogonal views while seed voxel was initiated in gray matter. a) transverse, b) coronal , c) sagittal plane.



(I) t=0 day (II) t=10 days

(III) t=25 days (IV) t=75 days



(V) t=150 days (VI) t=365 days

Fig.6 (I)-(VI): Low-grade glioma (ρ=0.018/day, D_w=0.015mm^2/day, D_g =0.005 mm^2/day) from three orthogonal views while seed voxel was initiated in gray matter. a) transverse, b) coronal , c) sagittal plane.

LED's Optimum Duty Cycle in Heart Rate and Oxygen Saturation Level Combinedly Measurement Device for Low Power Consumption

Umme Salma[1], Md. Jannatul Ferdous[2], Mohammad Rafiqul Alam[3], Tilottoma Barua[4] and Esrat Jahan[5]

[1,3,4,5] Department of Electrical and Electronic Engineering, Chittagong University of Engineering and Technology (CUET)
[2]Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology (BUET)
[1]Sumi.7445@gmail.com, [2]j.ferdous63@yahoo.com, [3]mra_cueteee@yahoo.com

Abstract— **Heart rate and oxygen saturation level measurement device needs to flash two LEDs alternatively at working mode. These flashing LEDs dominate the power requirement of the whole system. To save power dissipation in the LEDs, the LED drive signals are chopped with a small duty cycle along with high switching frequency. But this small duty cycle lowers the signal SNR and leads to an inaccurate result. So we need to find out an optimum duty cycle which can satisfy both criterions while the device works. This paper made an investigation to find out an optimum duty cycle of LED drive signals which can fulfil not only the low power consumption criterion but also ensures the data accuracy of the device. At first, theoretical design has been given. Then a prototype has been constructed and implemented by which the data accuracy variation at different duty cycle has been checked out. Finally we got the promising result at 5% duty cycle.**

Index Terms— **Duty cycle, low power consumption, data accuracy.**

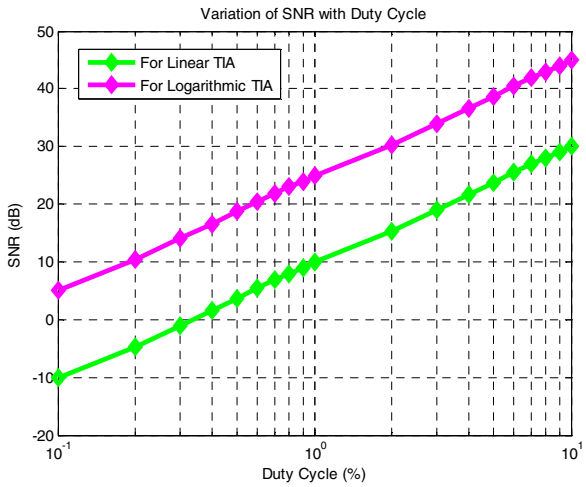


Fig. 1 Variation of SNR with Duty Cycle

I. Introduction

Low power consumable device which can measure the heart rate and oxygen saturation level (SpO2) combinedly is still a major challenge to design in health care system[1]-[2]. To measure the heart rate and oxygen saturation level, circuit needs to flash two LEDs (Red and Infrared) alternatively [2]-[5]. These LEDs are one of the dominant sources of power consumption. To save power dissipation in LEDs, it is necessary to chop the LED drive signals with small duty cycle. Different pulse oximeter devices use this technique. For example, in [2] and [6] a duty cycle of 3% and 0.8% with a LED switching frequency of 75Hz was proposed in order to dissipate low power. But the main problem arises with low duty cycle as it lowers the Signal to Noise ratio (SNR) of the device logarithmically which is depicted in Fig.1. This lower SNR at low duty cycle exacerbates the output signal remarkably and leads to an inaccurate result [7]-[9]. Our primary objective is to design a device which can measure heart rate and oxygen saturation level simultaneously at low power without sacrificing its accuracy. In order to do that, we have to maintain sufficient SNR at the detector or to increase the number of amplifier-filter block. Increasing the number of amplifier-filter block causes high installation cost as well as more power consumption at working mode.

So in this paper, we prefer to increase the duty cycle of LED drive signal at an optimum value for data accuracy as well as low power consumption. In this regard, we have carried out an investigation to set the perfect duty cycle for both LEDs which will provide sufficient SNR.

At 100 Hz frequency of LED drive signal, we found that 5% duty cycle can provide almost 39dB SNR which ensures low power consumption as well as data accuracy of the device. Variation of collected data and its accuracy (both for heart rate and oxygen saturation level) at 5% duty cycle, less than 5% duty cycle and more than 5% duty cycle have been shown in experimental result section. We have observed the output up to 10% duty cycle because we did not get any significant change in average data error between 10% duty cycle and 5% duty cycle. Moreover, we have got acceptable average error (2.29% for heart rate and 1.90% for SpO_2) at 5% duty cycle. If we would consider more than 10% duty cycle, then there will be a little data accuracy (negligibly small) but the power consumption by LEDs will be very high. This is totally contrary to the objective of our work. At last, a logarithmic transimpedance amplifier (TIA) has been used in the circuit which ensures faithful amplification at low duty cycle of LEDs. The use of logarithmic TIA also leads the energy efficient characteristics of the device. Power consumption at different block of the proposed circuit has also been measured and compared with other existing devices.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

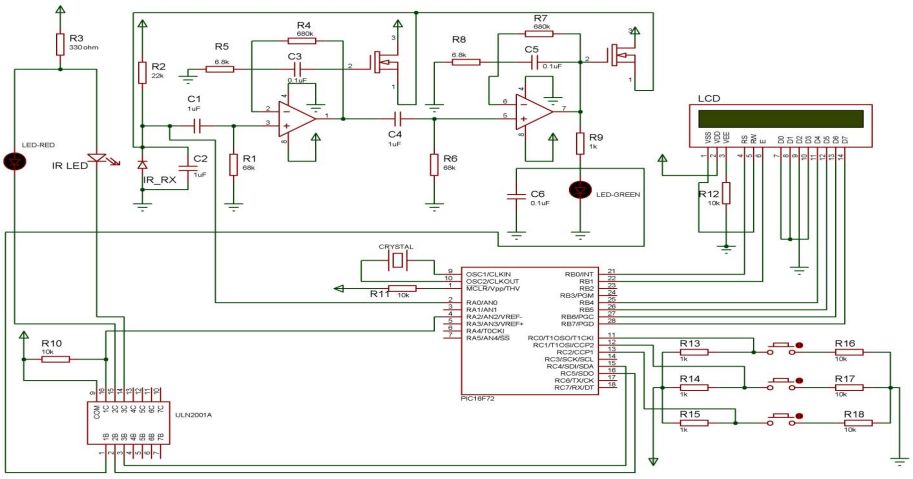


Fig. 2 Proposed circuit architecture

II. Proposed Circuit Architecture and Operating Principle

Fig.2 shows the architecture of the proposed circuit. Its overview has been given below.

A. Signal conditioning Circuit

The circuit consists of a red LED, an infrared (IR) LED and a photo detector. To observe the variation of data accuracy and power dissipation, the LED drive signals are chopped with differentl duty cycle with 100Hz switching frequency. The switching frequency must be kept much higher than the heart rate frequency so that the aliasing effects are minimal. To drive the LEDs, we have used a driver circuit ULN 2003A and to control the driving pulse frequency, duty cycle etc., a low cost microcontroller (PIC16F72) has been used. The microcontroller also assists to count the heart rate and oxygen saturation level. An LCD display is interfaced with microcontroller to display the heart rate and oxygen saturation level simultaneously.

B. Transimpedance Amplifier

In our circuit, a logarithmic transimpedance amplifier (TIA) is used which employs a logarithmic characteristics to convert current to voltage.

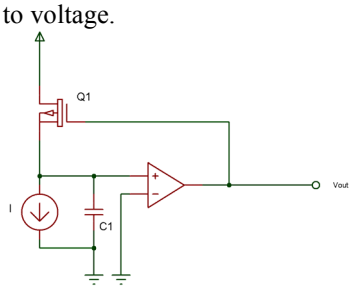


Fig.3 Logarithmic TIA

In Fig.3, the output voltage of logarithmic TIA is-

$$Vout = U_T/K \ln(I_1/I_0) \quad (1)$$

Where K_1 is the sub threshold exponential coefficient of Q_1 and U_T is the thermal voltage. The key incentive behind the employment of the logarithmic TIA is that it is inherently sensitive to the contrast (ac/dc) of the input photo current signal [2].

C. Filter circuit

Both LED consists of a large DC component. To block this DC component, a high pass filter (HPF) has been used in the circuit. C2 and R1 are the HPF components. In addition to that, to extract the blood pulsation signal, a low pass filter (LPF) is used. The cut-off frequency of this LPF needs to be low enough to remove the switching frequency (100Hz) components. The cut-off frequency of the LPF is-

$$f_c = 1/2\pi R_4 C_3 \quad (2)$$

$$= 1/2\times 3.1416\times 680\times 10^3\times 0.1\times 10^{-6}$$

$$= 2.34 \text{ Hz}$$

Where R4 and C3 are the feedback resistance, capacitance of the LPF in Fig.2. The maximum measurable heart rate is about-

$$2.34\times 60 = 140 \text{ bpm}$$

The red LED, IR LED and photo detector are mounted on a finger clip. At first, light is emitted from the LEDs alternatively, passing through the finger and reaches at the photo detector which is placed on the other side of the finger. But a portion of light is absorbed by the finger while passing through it. The signal received by photo detector posses both DC and AC components which have been removed by LPF and HPF. To get an amplified output, energy efficient logarithmic TIA is used. This amplification has been performed by two stage amplifier block. The output of amplifier and filter circuit then feed to microcontroller.

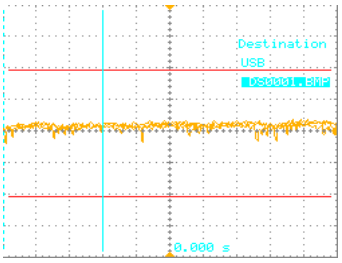
Microcontroller counts the number of pulses over a fixed time interval and thus obtains the heart rate of the subject.

Moreover, by analysing the absorption of two wavelength of two LED, it is possible to measure the actual difference in the absorption spectra of oxihemoglobin (HbO2) and hemoglobin (Hb) which helps to calculate the oxygen saturation level. This value is displayed in LCD with the help of microcontroller.

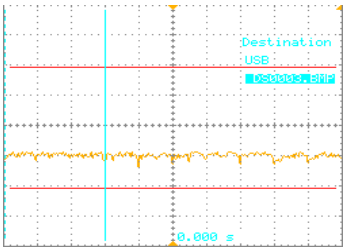
III. Experimental Results and Analysis Details

In this section, experimental analysis has been carried out with the constructed prototype and the collected data has been analysed in details. Real constructed circuit is shown in Fig.5 by which the experiment has been conducted.

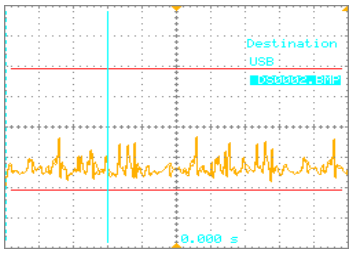
The output of the receiver is connected to an Oscilloscope to obtain the signal wave shape which is depicted in Fig.4. At first, the output of the receiver has been taken at below 5% duty cycle (at 2% duty cycle) of LED drive signal. This raw output signal is demonstrated by Fig.4(a). The signal illustrates that it is corrupted with heavy noise. So the data collected from this signal shows much error which has been summarised in TABLE I and TABLE II.



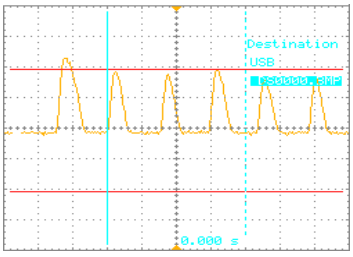
(a) 2% Duty Cycle,y axis 1V/Div.



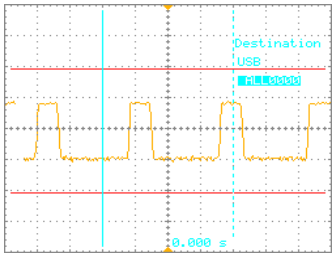
(b) 4% Duty Cycle, y axis 1V/Div.



(c) 5% Duty Cycle, y axis 1V/Div.



(d) Filtered Signal, y axis 1V/Div.



(e) Amplified output, y axis 2.5V/Div.

Fig. 4 Oscilloscope output , X axis 0.427s/Div.

To minimize the error rate, we increase the value of the duty cycle of LED drive signal at 4%. The output signal is now less corrupted compare to previous one. At 5% duty cycle, the device has given a promising result with small error rate. Fig.4(c) shows the raw signal taken at 5% duty cycle where as Fig.4(d) and Fig.4(e) are the filtered and amplified output signal.

The prototype was used to measure the heart rate and SpO2 of 15 healthy subjects at 2%, 4%,5% and 10% duty cycle. At the same time their heart rates were measured by OMRON and SpO2 was measured by Avant®4000. We have selected OMRON and Avant®4000 as a reference device for heart rate and oxygen saturation measurement. Both values were compared in TABLE I and TABLE II. TABLE I shows that at 2% and 4% duty cycle, the device average error is 7.12% and 5.63% respectively where at 5% duty cycle the error comes down at 2.29%. Furthermore, 10% duty cycle gives an error of 2.12%. So statistically it is evident that the error minimizes comprehensively when the duty cycle increases from 4% to 5%. But the error minimization is negligibly small from 5% to 10% duty cycle. Moreover, 2.29% is an acceptable error rate (supported by [9]) which we have got at 5% duty cycle. We can run our device at 10% duty cycle or more instead of 5% to get better data accuracy but at the same time we have to sacrifice high power loss at larger duty cycle by two LEDs. This phenomenon has been depicted in Fig. 6(c).Even there will be no major or substantial improvements in error percentile while the device working at more than 5% duty cycle. So 5% duty cycle is the optimum value for the device. Same analysis is true for oxygen saturation data which is sum up in TABLE II.

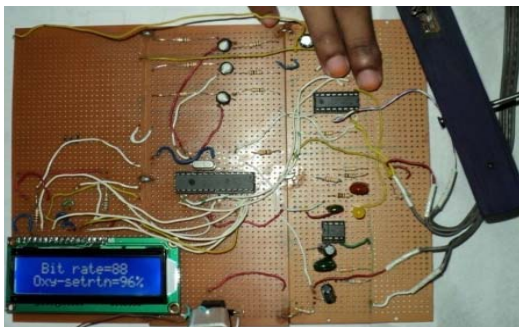


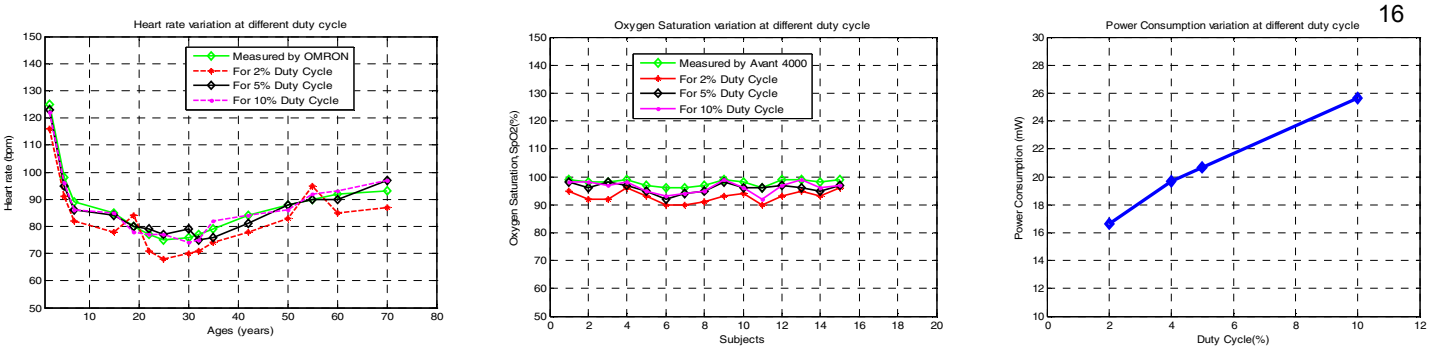
Fig.5 Constructed prototype

TABLE I
HEART RATE VARIATION AND ERROR MEASUREMENT AT DIFFERENT DUTY CYCLE (D)

Ages	OMRON bpm	bpm for 2% D	bpm for 4% D	bpm for 5% D	bpmfor 10% D	Average Error(%)
2	125	116	117	123	122	**7.12%**(for 2% D)
5	98	91	93	95	96	
7	89	82	84	86	86	
15	85	78	79	84	85	
19	80	84	76	80	78	**5.63%** (for 4% D)
22	77	71	73	79	77	
25	75	68	69	77	77	
30	76	70	72	79	74	
32	77	71	73	75	75	**2.29%** (for 5% D) ***Acceptable***
35	79	74	75	76	82	
42	84	78	80	81	84	
50	88	83	83	88	86	
55	90	95	86	90	92	**2.12%** (for 10% D)
60	92	85	87	90	93	
70	93	87	98	97	97	

TABLE II
OXYGEN SATURATION (SPO2) VARIATION AND ERROR MEASUREMENT AT DIFFERENT DUTY CYCLE (D)

Subjects #	Avant 4000	SpO2 for 2% D	SpO2 for 4% D	SpO2 for 5% D	SpO2 for 10% D	Average Error(%)
1	99	95	95	98	98	**5.10%**(for 2% D)
2	98	92	94	96	98	
3	98	92	92	98	97	
4	99	96	96	97	98	
5	97	93	93	95	95	**4.16%**(for 4% D)
6	96	90	91	92	93	
7	96	90	93	94	94	
8	97	91	92	95	95	
9	99	93	96	98	99	**1.90%**(for 5% D) ***Acceptable***
10	98	94	94	96	96	
11	96	90	92	96	92	
12	99	93	95	97	97	
13	99	95	95	96	99	**1.64%**(for 10% D)
14	98	93	93	95	96	
15	99	96	96	97	97	



(a) (b) (c)

Fig. 6 MATLAB plot curves showing variation of heart rate, oxygen saturation and power consumption with duty cycle

To illustrate the variation and error at different duty cycle, we have plotted the data in MATLAB. Fig. 6 displays all the curves with the contrast more vividly. State of power consumption (which is measured) with increasing duty cycle is also depicted in Fig. 6(c).

We carefully measure the power requirement of various stages of our device running at 5% duty cycle which is listed in TABLE III. The table shows just LED and processing power consumption, not LCD display or other equipment's consumption. We have used 50mW two LEDs.

TABLE III
POWER CONSUMPTION CALCULATION

Power Consumption Block	Value
LEDs	5mW
Microcontroller+Driver	15mW
Transimpedance amplifiers	100 uW
Filters	550 uW
Total	20.65mW

Power consumption of our device adds up to 20.65mW which is relatively low. Power mostly spent by the LEDs, driver and microcontroller circuit. Finally, TABLE IV compares the power dissipation of our device with several commercial devices. Power dissipation data for several commercial devices has been collected from [2]. This table of comparison shows that, our device which operates at 5% duty cycle consumes less power relative to other devices.

TABLE IV
POWER CONSUMPTION COMPARISON

Type	**Consumed Power**
Our Devive	20.65 mW
Wrist ox 3100	55 mW
Avant®4000	71 mW
Xpod	60 mW
Device proposed in [2] (works at 3%Duty cycle)	4.8mW

IV. CONCLUSIONS

The main aim of our work was to investigate in to optimum duty cycle of LED drive signals of heart rate and oxygen saturation level measurement device. In order to minimize the power loss and retain the data accuracy, this optimum duty cycle of LED drive signal is a pivotal parameter. After constructing a device prototype, we have observed the heart rate and oxygen saturation level data at different duty cycle and tried to pick up the best possible optimum one. At 5% duty cycle, the device provides a very promising result with an acceptable average error. At last, we have measured the power consumption at 5% duty cycle which exhibits low power consumption property of the device compare to other existing one.

APPENDIX

We have calculated error percentile using the following equation-

$$Error = \frac{ActualData - MeasuredData}{ActualData} \times 100\%$$

REFERENCES

[1] Matthew James Hayes and Peter R. Smith, "A New Method for Pulse Oximetry Possessing Inherent Insensitivity to Artifact", IEEE Trans. on Biomed. Engg, Vol. 48, No. 4, pp. 452–461,April 2001

[2] M. Tavakoli, L. Turicchia, and R. Sarpeshkar, "An ultra-low-power pulse oximeter implemented with an energy efficient transimpedance amplifier." IEEE Trans. on Biomed. cir. And sys. Vol. 4, No. 1, pp. 27–38,February 2010.

[3] Mohamed A. Zaltum, M. Shukri Ahmad, Ariffuddin Joret, and M. Mahadi Abdul Jamil, "Design and Development of a portable Pulse Oximetry system" , pp 37-40, December 2008.

[4] H. Shim, J.H. Lee, S.O. Hwang, H.R. Yoon, and Y.R. Yoon,"Development of heart rate monitoring for mobile telemedicine using smartphone", 13thInternational Conference on Biomedical Engineering (ICBME 2008), Singapore, 2008, pp. 1116-1119.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[5] M.M.A. Hashem, Rushdi Shams, Md. Abdul Kader, and Md. Abu Sayed, "Design and Development of a Heart Rate Measuring Device using Fingertip" International Conference on Computer and Communication Engineering (ICCCE), Kuala Lumpur ,2010.

[6] P.O. Isaacson, D.W. Gadtke, V.D. Heidner, and N.F. Nordling. Single channel pulse oximeter, 1988. US Patent 4,773,422. 36, 38, 39, 41, 78, 177..

[7] http://www.alldatasheet.com/datasheetpdf/pdf/80247/ATMEL/ATMEGA8.html.

[8] R.S. Khandpur,"Handbook of Biomedical Instrumentation", Tata Macgraw-Hill Education, April 1, 2003, second edition

[9] Konstantinos N. Glaros, "Low power pulse oximetry and transimpedance amplifier," Ph.D thesis, Imperial College London,October 2011.

Methylsorb: A Simple Method for Quantifying DNA Methylation Using DNA-gold Affinity Interactions

Muhammad J. A. Shiddiky[1,*] Abu Ali Ibn Sina[1], Laura G. Carrascosa[1], Ramkumar Palanisamy[1], Sakandar Rauf[1], and Matt Trau[1, 2]

[1]Australian Institute for Bioengineering and Nanotechnology (AIBN), Corner College and Cooper Roads (Bldg 75), The University of Queensland, Brisbane QLD 4072, Australia

[2]School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, QLD 4072, Australia.

*m.shiddiky@uq.edu.au

The analysis of DNA methylation is becoming increasingly important both in the clinic and also as a research tool to unravel key epigenetic molecular mechanisms in biology. Current methodologies for the quantification of regional DNA methylation are largely affected by comprehensive DNA sequencing methodologies which tend to be expensive, tedious, and time-consuming for many applications. Herein, we report an entirely new DNA methylation detection method referred to as 'Methylsorb', which is based on the inherent affinity of DNA bases to the gold surface (i.e., the trend of the affinity interactions is adenine > cytosine ≥ guanine > thymine) [1]. Since the degree of gold-DNA affinity interaction is highly sequence-dependent, it provides a new capability to detect DNA methylation by simply monitoring the relative adsorption of bisulfite treated DNA sequences onto a gold chip. Because the selective physical adsorption of DNA fragments to gold enable a direct read-out of regional DNA methylation, the current requirement for DNA sequencing is obviated. To demonstrate the utility of this method we present data on the regional methylation status of two CpG clusters located in the EN1 and MIR200B genes in MCF7 and MDA-MB-231 cells. The methylation status of these regions was obtained from the change in relative mass on gold surface with respect to relative adsorption of an unmethylated DNA source and this was detected using surface Plasmon resonance (SPR) in a label-free and real-time manner. We anticipate that the simplicity of this method, combined with the high level of accuracy for identifying the methylation status of cytosines in DNA, could find broad application in biology and diagnostics.

***Index Terms*— DNA methylation • Gold-DNA affinity • DNA adsorption • Surface Plasmon resonance**

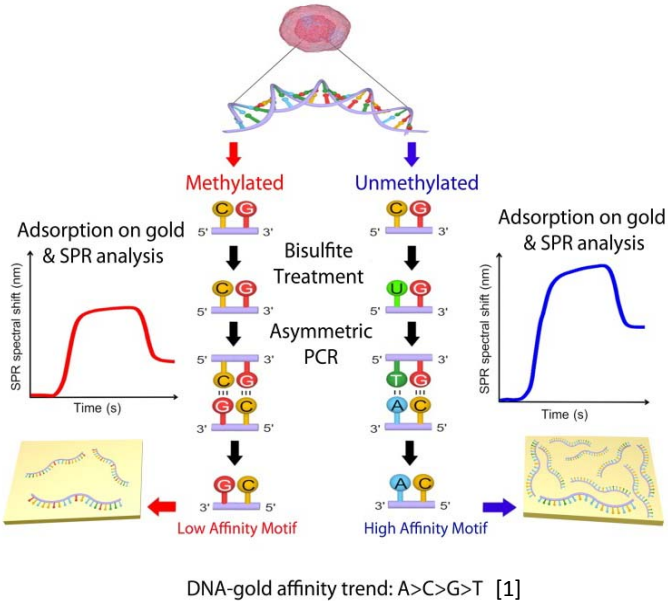
I. Introduction

DNA methylation is a prevalent epigenetic alteration in cancer [2], and hence considered one of the most promising biomarkers for cancer diagnosis, prognosis and therapy monitoring [3]. Early detection of aberrant DNA methylation offers a great potential to improve survival rates and decrease the cost of treatment by ameliorating prescription of ineffective therapies. DNA methylation strategies typically involve bisulfite [4] modification of DNA to convert methylation events to base changes followed by comprehensive sequencing to read each individual base pair in the genomic region of interest [5]. Over the past few years, much attention has also been focused on detecting DNA methylation from bisulfite treated DNA using sequence-sensitive methods such as methylation specific PCR (MSP) [6], combined bisulfite restriction analysis (COBRA) [7], sequencing [5] or bio-sensing [8] coupled with mass spectroscopy, electrochemical or optical readouts. While most of these methodologies have significantly improved the analysis performance, their practical application is restricted due to the complicated surface and conjugation chemistries, multistep analysis procedures or the use of labels. Therefore, we believe that the development of a simple, low cost and robust detection method for DNA methylation is essential to enable the uptake and full utilization of such biomarkers in the clinic.

In this report, we have envisioned an extremely simple method referred to as 'Methylsorb', which is an effective and potential alternative to traditional methods and offers a label-free and real-time detection of DNA methylation. The underlying principle of Methylsorb involves the use of the sequence-dependent affinity interaction of DNA bases towards gold substrates. Since (i) bisulfite treatment converts methylation events into base changes, and (ii) DNA base-gold affinity interaction (*i.e.,* DNA adsorption) is highly sequence-dependent [1], [9]-[10], we hypothesized that the direct detection/quantification of adsorbed (*via* DNA-gold affinity interaction) bisulfite treated-DNA sequences could be an extremely simple way to analyse DNA methylation events. The schematic of the Methylsorb approach is illustrated in Scheme 1. Initially, genomic DNA from cells was treated with sodium bisulfite to convert unmethylated cytosines into uracil, leaving methylated cytosines unchanged. Bisulfite converted samples were then amplified via a subsequent asymmetric PCR amplification step, which converts all uracils into thymines in the sense strand (*i.e.,* the DNA strand running from 5' to 3' end) or into adenines in the antisense strand (*i.e.,* the DNA strand running from 3' to 5' end). Asymmetric amplification ensures that only ss-DNA sequences corresponding to the antisense strand are amplified. This step therefore introduces an accurate and unique way to create specific base changes on DNA, which ultimately can alter DNA-gold affinity interaction, and thus give quantitative information on the original methylation status of the sample. This is because (i) it only generates ss-DNA, which is more prone than dsDNA to uncoil sufficiently to expose its bases and rapidly interact with the gold surface [11], and (ii) it correlates methylation profile with gold-DNA affinity behaviour, as only the antisense strand, which becomes guanine-enriched for methylated or adenine-enriched for unmethylated samples, is amplified. Since adenine bases are

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

reported to have stronger affinity towards gold than guanine [1], [9]-[10], the adenine-enriched ssDNA sequences (*i.e.,* unmethylated sequence) are expected to exhibit higher adsorption on gold than the methylated guanine-enriched sequences. The differential affinity interaction (*i.e.,* differential adsorption) of adenine or guanine-enriched sequences, corresponding to the DNA base-changes associated with DNA methylation, can be determined in a single step via direct quantification of adsorbed DNA sequences in-real time.



Scheme 1. Methylsorb approach. Genomic DNA is derived from breast cancer cell lines. Bisulfite treatment and asymmetric PCR processing of these samples generate adenine-enriched (blue) or guanine-enriched (red) antisense ss-amplicons for unmethylated and methylated templates respectively. The ss-DNA amplicons are directly adsorbed on gold and their adsorption is monitored in real-time using surface Plasmon resonance readout.

We used this approach to quantify the DNA methylation status of two CpG clusters located within the Engrailed homeobox1 (*EN1*) and MIR200B gene promoter regions on DNA derived from MCF7 and MDA-MB-231 cells respectively. As a standard reference for unmethylated DNA sample, we used whole genome amplified (WGA) DNA. To quantify adsorbed DNA sequences on gold chip, we used a surface Plasmon resonance (SPR) biosensing approach that detects changes in refractive index over time (*i.e.,* changes in the SPR spectral shift) at the sensing surface, which are directly proportional to the relative mass increase associated with target adsorption, thus enabling the real-time and label-free monitoring of targets. This method has previously been used to detect regional DNA methylation using proteins with affinity to CpG rich regions [12] or molecular inversion probes [13]. Detecting DNA methylation through this gold-driven adsorption process in combination with SPR offers several critical improvements to conventional methylation technologies that include: (i) minimal sample and operational requirements, (ii) no use of surface-bound receptors and thus avoids modification procedures for the recognition layer of the SPR surface and (iii) label-free and real-time monitoring that avoids any artifacts generally associated with the use of labels.

II. Materials and Methods

A. DNA Sample Preparation

Genomic DNA from MCF7 and MDA-MB-231 cells was extracted from 105 cells plates using the DNeasy blood and tissue kit (QIAGEN Pty. Ltd., Venlo, Netherlands) according to manufacturer's instructions. The purified DNA was bisulfite treated using a MethylEasyXceed kit (Human Genetic Signatures Pty. Ltd., New South Wales, Australia) according to the manufacturer's instructions. To obtain ss-stranded amplified amplicons containing the target sequence, samples were then amplified in a 60 µL reaction containing 1.5 unit Taq DNA polymerase (AmpliTaq DNA Polymerase, Applied Biosystems, Australia), 0.7X PCR buffer (AmpliTaq 10X PCR buffer) 0.2 mM each dNTP, 0.1% tween, 125 nM of forward primer and 375 nM reverse primer (Supplementary table 1). Cycling was carried out in a Bio-Rad thermo cycler (MJ Mini Personal Thermal Cycler) using the following conditions: denaturation at 94 ºC for 10 min followed by 50 cycles of 94 ºC for 30 s, 58 ºC for 45 s and 72 ºC for 30 s.

B. Methylation detection with SPR biosensor

All experiments were performed in a custom-made SPR platform. Initially, the SPR sensor chips (5nm Ti and 50 nm Au) were cleaned by rinsing with hot acetone, ethanol and deionised water and dried under flow of nitrogen gas. Subsequently, the chips were dipped in piranha solution (70% H_2SO_4–30% H_2O_2) for few seconds, rinsed with water and dried under flow of nitrogen gas. Following this step, the chip was placed in the SPR platform and sodium saline citrate (SSC) 5X buffer (0.75 M in NaCl, 0.075 M in sodium citrate, pH 7) was flowed in continuous at a flow rate of 0.6 mL/h. Then, 250 µL of DNA samples (10 µL of asymmetric PCR output spiked in 250 µL of SSC5X buffer) were directly injected in the SPR flow system at the same flow rate. The SPR signals were monitored using custom-made Labview software.

III. Results and Discussion

To investigate the strategy of gold–DNA affinity interactions for detecting DNA methylation levels, we designed synthetic DNA (Fig.1A) similar to the bisulfite treated and asymmetric amplified sequences derived from the *EN1* gene. The *EN1* gene comprises of eight potentially methylated CpG sites within a span of 53 bases downstream to the transcription start site and identified as a potential biomarker in several cancer types [14].

We initially performed our experiment at pH 3 because DNA adsorption extent and kinetics are reported to be facilitated at pH 3 [15]. However, we have observed that adsorption at pH 3 limits the ability of the Methylsorb approach to differentiate methylated and unmethylated targets. At low pH (*i.e.* pH = 3) adenines and cytosines become protonated, leading to high DNA adsorption compared to that under neutral pH conditions [15]. Since both methylated and unmethylated sequences contain large numbers of cytosines and adenines bases, they could be strongly adsorbed on gold substrate within a short period of time at pH <7. Thus, we chose to carry out our experiments at the solution pH 7 that does adsorb the desired unmethylated targets, but do not strongly adsorbed the methylated one.

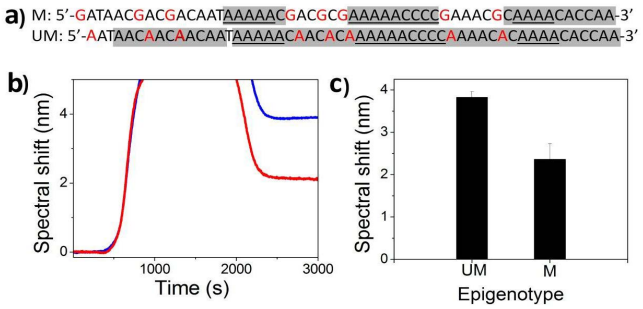


Fig. 1 Detection of DNA methylation in synthetic samples. (a) DNA Sequence. High and low affinity sequence motifs are noted (*i.e.,* DNA sequence motifs of minimum 4 identical bases are underlined; sequence motifs corresponding to poly A/C mixed bases are highlighted in grey colour; the A/G differences between the two strands are highlighted in red font). (b) SPR sensorgrams showing the spectral shift generated during the adsorption of methylated (red) and unmethylated (blue) synthetic *EN1* region onto the gold chip. (c) Corresponding mean values of SPR spectral shift for methylated (M) and unmethylated (UM) regions. Each bar represents the average of three separate trials (n=3), and error bars represent the standard deviation of measurements within experiments (relative standard deviation (%RSD) was found to be <8% for n= 3).

Fig. 1b illustrates the SPR signals obtained for methylated and unmethylated synthetic DNA samples at pH 7. The adenine enriched unmethylated DNA demonstrated 40% enhanced adsorption towards gold surfaces in comparison to the guanine enriched methylated DNA (see Fig. 1c, 3.82 ± 0.13 versus 2.29 ± 0.18 nm). This adsorption behaviour is possibly due to the higher affinity interaction of adenine bases towards gold surface in comparison to guanine. This behaviour is in accordance with the previously observed base-dependent adsorption (*i.e.,* A >C ≥ G >T) of DNA [1] onto gold surfaces. Notably, this outcome indicates that the Methylsorb approach can effectively distinguish methylated and unmethylated epigenotypes in synthetic samples even though they involved only 8 CpG sites. It is also important to note that for a given number of DNA bases, a DNA containing homo-polymer (*e.g.*, a sequence with clustered poly-A motifs) and a DNA containing more distributed DNA bases (*e.g.*, adenine bases spread across the sequence) might essentially not exhibit the same adsorption behaviour. We analysed the sequence composition of the methylated and unmethylated epigenotypes and investigated the presence of any high gold-affinity motif containing (i) homo-polymers (*e.g.*, poly A) with a minimum of four bases and, (ii) polymers enriched with a mixture of adenine and cytosine bases. The identified homo-polymers (underlined) and polymers enriched with A, C bases (highlighted in grey) are represented in Fig. 1a. It is evident that the methylated and unmethylated DNA contains similar homo-polymers both in numbers (*e.g.*, 3 poly A and 1 poly-C motif) and in length. However, the polymers enriched with A,C bases were significantly larger in length for unmethylated sequence. This is because bisulfite treatment and PCR processing of DNA not only lead to A/G-enrichment, but also result in GC or AC enriched motifs in the methylated or unmethylated sequences (Scheme 1 and Fig. 1A). Additionally, bisulfite conversion (followed by asymmetric PCR) of DNA also increases the adenine contents in the antisense strand for both epigenotypes. This enhances the possibility of forming larger AC motifs in the unmethylated sequences. We believe that this differential length of poly-AC motifs in methylated and unmethylated sequences is the key factor to offer a distinguishable SPR signal in our method.

To assess the applicability of this methodology in genomic DNA samples, we chose to investigate (i) *EN1* region in DNA derived from MCF7 cells, and (ii) a region containing 8 CpG sites in a span of 65 bases within the promoter region of the *MIR200B* gene derived from the MDA-MB-231 cells. Both the genes were identified previously as a potential biomarker in several cancer types [14], [16]. We simultaneously analysed the same *EN1* and *MIR200B* regions on whole genomic amplified (WGA) human DNA. This DNA is fully unmethylated (adenine enriched), and hence DNA amplicons derived from WGA are expected to give maximum achievable SPR signal for this region. Since methylated amplicons are guanine-enriched and they have lower adsorption affinity compared to the WGA DNA, there will be a clear correlation between the presence of methylation and lowering of the SPR signal from that of the WGA DNA.

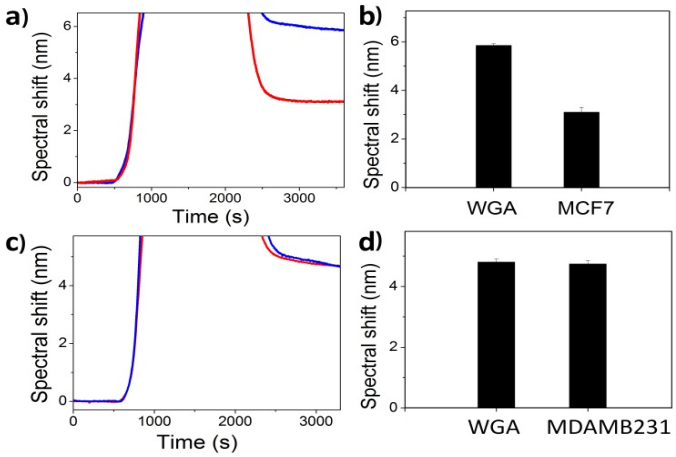


Fig. 2 Quantification of DNA methylation in samples derived from cancer cell lines and whole genome amplified DNA. Representative SPR sensorgrams showing the spectral shift during real-time adsorption of antisense amplicons derived from the (a) *EN1* region in MCF7 cells (red) and WGA (blue) DNA and the (c) *MIR200B* gene region in MDA-MB-231 (red) and WGA (blue) DNA. Corresponding mean SPR spectral shift are shown in the right panels (b, d). Each bar represents the average of three separate trials (n=3), and error bars represent the standard deviation of measurements within experiments (%RSD was found to be <7%).

Fig. 2a illustrates the SPR signals obtained for the direct adsorption of MCF7 and WGA DNA amplicons onto a gold surface. Similar to the responses obtained for *EN1* synthetic DNA samples, we observed an enhanced adsorption behaviour in the case of unmethylated WGA DNA (A>G; dark blue line) that is evident from a larger spectral shift (Fig. 2b, 5.85 ± 0.07 (MCF7) versus 3.1 ± 0.2 nm (WGA)) in comparison to methylated MCF7 DNA (G>A; red line) indicating the methylated status of this region in MCF7 cells. Responses were in agreement with our previous sequencing analysis data of this same region on MCF7 cells (see Supporting Information of reference 21). Subsequently, the Methylsorb analysis of *MIR200B* region resulted in almost identical SPR signals for MDA-MB-231 (Fig. 2c; red) and WGA DNA (Fig. 2c; blue). The negligible difference in SPR spectral shift (Fig. 2d; 4.74 ± 0.09 (MDA-MB-231) versus 4.8 ± 0.1 nm (WGA)) values suggested that the *MIR200B* region in MDA-MB-231 could possibly be unmethylated. Sequencing analysis data confirmed the unmethylated status of *MIR200B* sequence in MDA-MB-231 DNA cell line. This data contrasts with previous observations, where a strong CpG methylation (*i.e.*, hypermethylation) across the promoter of the *MIR200B* in

MDA-MB-231 cell line has been identified [28]. It has been reported that replicas of the same cell line in vitro could exhibit differential DNA methylation at certain regions [17], we believe that this could explain the discrepancy between our results and those from the literature.

Clinical samples often contain a mixture of tumour and normal cells [18] where the presence of unmethylated sequences can potentially influence the sensitivity of the diagnostic assay. Therefore, we investigated the ability of our approach to detect heterogeneous DNA methylation (*i.e.*, methylated targets in the presence of designated proportions of unmethylated DNA) by measuring the adsorption of DNA samples containing mixture of methylated and unmethylated sequences. Samples were prepared by quantifying the copy numbers of MCF7 and WGA DNA followed by mixing designated proportions of methylated and unmethylated targets (*i.e.*, MCF7/WGA DNA at 0%:100%, 25%:75%, 50%:50%, 75%:25%, and 100%:0%).

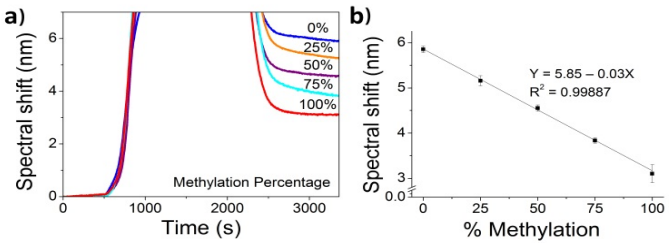


Fig. 3 Detection of DNA methylation in heterogeneous samples. (a) Representative sensorgrams showing the SPR spectral shift for the real-time adsorption of heterogeneous sample mixtures containing designated percentage of DNA derived from MCF7 and WGA (*i.e.,* MCF7:WGA DNA at 0%:100%, in blue; 25%:75%, in orange; 50%:50%, in purple; 75%:25%, in light blue and 100%:0%, in red). (b) Corresponding calibration plot. Each data point represents the average of three separate trials, and error bars represent the standard deviation of measurements within experiments.

Fig. 3a represents the SPR sensograms of the heterogeneously methylated DNA samples. It was evident that the SPR signal was a function of methylation percentage in the heterogeneous mixture. It was also noted that there was a linear correlation (R^2 = 0.9988) in signal enhancement with a decrease in methylation percentage (Fig. 3b). This is presumably due to the increase in the adenine nucleotide content with increase in the percentage of unmethylated DNA in the mixture resulting in an enhanced signal. Additionally, a significant difference in signal for samples containing 25% and 0% methylated targets (Fig. 3b; 5.85 ± 0.07 versus 5.16 ± 0.11 nm) indicate that our approach was sensitive enough to detect at least 25% methylation changes. However, we believe that further optimization to the operating parameters (*e.g.*, flow rate, buffer composition and multiplexed SPR platform) can improve the detection sensitivity of our method and potentially be integrated into clinical settings.

IV. Conclusions

We have developed a simple, cost-effective and real-time approach with the capacity to sensitively detect DNA methylation levels in genomic DNA samples. Our experimental findings demonstrate the accurate interrogation of CpG regions on multiple genes using SPR biosensor and simultaneous validation using conventional bisulfite sequencing. This unique label-free and rapid (*e.g.,* ~40 minutes) approach based on the preferential nucleotide affinity towards gold can effectively detect DNA methylation levels thereby eliminating the need for sequencing approaches. We have demonstrated the feasibility of using this approach to sensitively (*e.g.,* 25% methylated target) and specifically detect methylated targets in the presence of large excess of unmethylated DNA. This method has also the potential to detect global hypomethylation, one of the most important biomarkers of cancer, since the large cohort of CpG regions involved in this type analysis might promote a large difference in the adsorption profiles of normal and hypomethylated samples. We envisage that this simple and rapid approach can potentially find its relevance in diagnostic settings.

Acknowledgment

This work was supported by the UQ fellowship (2012001456) and ARC DECRA and DP (DE120102503, DP140104006). We also acknowledge funding received by our laboratory from the National Breast Cancer Foundation of Australia (CG-12-07). This NBCF grant has significantly contributed to the environment that stimulates the research described here

References

[1] H. Kimura-Suda, D.Y. Petrovykh, M.J. Tarlov, and L.J. Whitman, Base-dependent competitive adsorption of single-stranded DNA on gold, *J. Am. Chem. Soc.* 125, 9014-9015, 2003.

[2] M. Suzuki and A. Bird, DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.* 9, 465-476, 2008.

[3] H. Heyn and M. Esteller, DNA methylation profiling in the clinic: applications and challenges. *Nat. Rev. Genet.* 13, 679-92, 2012.

[4] S.J. Clark, A. Statham, C. Stirzaker, P.L. Molloy and M. Frommer, DNA methylation: Bisulphite modification and analysis. *Nature Protocols.* 1, 2364, 2006.

[5] E. Lee, J. Luo, J. Wilson and H. Shi, Analyzing the cancer methylome through targeted bisulfite sequencing. *Cancer Lett.* 340, 171-8, 2013.

[6] J.G. Herman, J.R. Graff, S. Myohanen, B.D. Nelkin and S.B. Baylin, *Methylation-specific PCR: A novel PCR assay for methylation status of CpG islands. Proc. Natl. Acad. Sci. USA.* 93, 9821-9826,1996.

[7] Z. Xiong and P.W. Laird, COBRA: a sensitive and quantitative DNA methylation assay. *Nucleic Acids Res.* 25, 2532-2534, 1997.

[8] K.M. Koo, E.J.H Wee, S. Rauf, M.J.A. Shiddiky and M. Trau, Microdevices for Detecting Locus-Specific DNA Methylation at CpG resolution. *Biosens. Bioelectron.* 15, 278-85,2014.

[9] L. Demers *et al.* Thermal desorption behaviour and binding properties of DNA bases and nucleosides on gold. *J. Am. Chem. Soc.* 124, 11248-9, 2002.

[10] A. Opdahl, D.Y. Petrovykh, H. Kimura-Suda, M.J. Tarlov and L. Whitman, *J. Proc. Natl. Acad. Sci. USA.* 2007, *104*, 9-14.

[11] H. Li and L. Rothberg, Colorimetric detection of DNA sequences based on electrostatic interactions with unmodified gold nanoparticles. *Proc. Natl. Acad. Sci. USA.* 101, 14036–14039, 2004.

[12] Y. Yu *et al.* Direct DNA methylation profiling using methyl binding domain proteins. *Anal. Chem.* 82, 5012-9, 2010.

[13] L.G. Carrascosa *et al.* Molecular inversion probe-based SPR biosensing for specific, label-free and real-time detection of regional DNA methylation. *Chem. Commun.* (Camb) 50, 3585 – 3588, 2014.

[14] P. Karpinski, D. Ramsey, Z. Grzebieniak, M.M. Sasiadek and N. Blin, The CpG island methylator phenotype correlates with long-range epigenetic silencing in colorectal cancer. *Mol. Cancer Res.* 6, 585, 2008.

[15] X. Zhang, B. Liu, N. Dave, M. Servos and J. Liu, Instantaneous attachment of an ultrahigh density of nonthiolated DNA to gold nanoparticles and its applications. *Langmuir* 28, 17053-17060, 2012.

[16] C. Köhler *et al.* Analyses in human urothelial cells identify methylation of miR-152, miR-200b and miR-10a genes as candidate bladder cancer biomarkers. *Biochem. Biophys. Res. Commun.* 438, 48-53, 2013.

[17] S. Cocozza, G. Scala, G. Miele, I. Castaldo and A. Monticelli, A distinct group of CpG islands shows differential DNA methylation between replicas of the same cell line in vitro. *BMC Genomics* 14, 692, 2013.

[18] S. Wang *et al.* Relationship between tumor DNA methylation status and patient characteristics in African-American and European-American women with breast cancer. *PLoS ONE.* 7, e37928, 2012.

Finite Element Modeling and Analysis of Thrombosis in Middle Cerebral Artery

Md. Sarwar Jahan[1], Md. Faruque Hossain[1*], Md. Maodudul Hasan[1], and Abbas Z. Kouzani[2]

[1]Department of Electronics and Communication Engineering, Khulna University of Engineering and Technology, Khulna 9203, Bangladesh

[2]School of Engineering, Deakin University, Warun Ponds, Victoria 3216, Australia

[*]fhossain@mail.kuet.ac.bd

Abstract—**Thrombotic stroke, which is caused by blood clot in the cerebral artery, is a major source of increased mortality and morbidity. Considering as efficient and fastest methods, mathematical approaches have gained significant importance for analyzing and understanding the biological events like thrombosis. This paper presents a computational model to analyze the effects of thrombosis using the theory of coupled fluid dynamics-structure interaction. The finite element method is used for the modeling of thrombosis (blood clot) of different stages in the middle cerebral artery with physiological compliance. The developed model is used to investigate the consequences that occur due to the various sizes of clots in the artery in the form of blood flow velocity, blood pressure, and artery wall stress. Such numerical assessment will facilitate better understanding of the biophysical process in case of thrombosis and thus would support medical practitioners to take faster curing steps.**

Index Terms—**Thrombosis, Middle cerebral artery, Blood clot, Finite element modeling, Thrombotic stroke.**

I. Introduction

Thrombosis, which is caused by the formation of blood clot (thrombus) inside a blood vessel, is one of the major health problems around the world. Thrombus formation or blood clot is an inherent characteristic of blood to protect the person from continuous haemorrhage. When a blood vessel is injured, the body uses platelets (thrombocytes) and fibrin to form a blood clot to prevent blood loss. Even when a blood vessel is not injured, blood clots may still form inside the blood vessel due to different associative factors such as repeated formation of calcium, cholesterol and fat carried in the blood, and disorders of blood clotting [1]. Blood clot caught on the plaque of the vessel walls may mostly or fully stop the blood flow. Stopped blood flow for a few minutes may cause neuronal injury (i.e., tissue death) due to the deprivation of oxygen and nutrient from blood, and leads to thrombotic stroke [2].

Diagnosis of stroke can be made by clinical assessment. Different imaging methods such as computed tomography (CT), magnetic resonance imaging (MRI), and contrast venography are widely used for the detection of thrombosis and its effect as stroke [3], [4]. Early prevention of a thrombus formation can be achieved by medical treatment. However, the early detection of thrombosis and prediction of its possible effects is often difficult by medical examination, since it often produces no specific symptoms before neuronal injury or stroke. In most cases, diagnosis processes are done only after the stroke happened. Thus, there arises a necessity to develop a mathematical approach to understand and analyze the effects of various thrombosis states.

A number of works have been devoted to the computational modeling of blood flow and thrombosis [4]–[8]. However, most of these works have considered 2D approaches [4], [6] and concentrated on the analysis of venous thrombosis [5]–[7]. Some of them considered only blood flow without considering the artery wall, which could not give many expected results such as pressure on the artery wall [4], [8]. Thus, the existing works are not sufficient to represent the realistic scenario of thrombosis. Since the works on the computational modeling of thrombosis are in their early stage, more efforts are required for the better understanding of thrombosis.

This paper presents an effective numerical modeling of thrombosis in the middle cerebral artery (MCA) and thus, demonstrates its various effects leading to a thrombotic stroke. The finite element (FE) method is used for the three dimensional (3D) modeling of various clot conditions inside the artery whose result is then employed to investigate a detailed scenario of blood flow including velocity magnitude, pressure distribution along blood and stress to artery wall. The critical conditions for the brain stroke are discussed with the obtained results. Our computational approach may have greater impact in diagnosing the effects of thrombosis in a shorter time period.

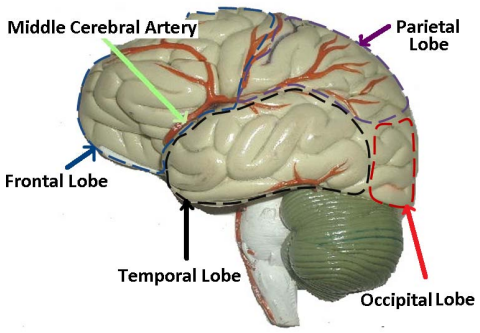


Fig. 1 Pictorial view of human brain.
(Courtesy: www.drcink.net/circulation/middlecerebralartery.jpg)

II. MCA Anatomy and Clinical Significance

The middle cerebral artery (MCA) is one of the three major paired arteries that supply blood to the cerebrum. Fig. 1 shows a pictorial view of the human brain indicating the MCA. The artery supplies a portion of the frontal lobe and the lateral surface of the temporal and parietal lobes, including the primary motor and sensory areas of the face, throat, hand and arm, and the areas for speech [9].

The MCA is the most common site of stroke [9], [10]. Occlusion of this artery results in MCA syndrome, potentially showing the defects as paralysis, weakness, or sensory loss of the contralateral face and arm, damage to the dominant

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

hemisphere (usually the left hemisphere) resulting in aphasia, damage to the non-dominant hemisphere (usually the right hemisphere) resulting in contralateral neglect syndrome, and large MCA infarcts especially during the acute period.

III. Mathematical Model

The dynamics of the fluid flow plays an important role to identify the degree of obstruction due to blood clot in cerebral artery. Computational fluid dynamics (CFD) is used here as a tool to study the flow dynamics. Fluid flow through veins or artery has the similarity to that of fluid flow through pipes or tubes. Such flow can be of two types as laminar flow and turbulent flow. Laminar flow occurs when a fluid flows in parallel layers, with no disruption between the layers. In laminar flow, the streamlines are straight lines. In turbulent flow, the fluid moves in a random path. Usually, this type of flow can be found at the junction of blood vessels. For the sake of simplicity, we chose the MCA without branch, and therefore considered the laminar flow model of blood. A widely used commercial software COMSOL MULTIPHYSICS is used for the implementation of the model. The model consists of two distinct but coupled procedures: first, a fluid-dynamics analysis including a calculation of velocity field and pressure distribution in the blood and second, a mechanical analysis of the deformation of the artery.

A. Laminar Flow Model

The blood behaves as a non-Newtonian fluid in which the relation between the stress and strain is non-linear or complex. As observed practically, the flow is considered in this model to be non-Newtonian, incompressible, viscous and laminar. Here, we used the 3D Navier-Stokes equation which defines a differential change in momentum within the blood as:

$$\rho\frac{\partial u}{\partial t}+\rho(u\cdot\nabla)u=\nabla\cdot\left[-\rho I+\mu(\nabla u)+(\nabla u)^T\right]+F \quad (1)$$

$$\rho\nabla\cdot u=0 \quad (2)$$

Here ρ is the density of blood, μ is the dynamic viscosity, F is the force, and u is the velocity vector. This relation coupled with the field continuity equation Δu=0, completely defines the dynamics of blood flow through the artery. They are valid for the blood being incompressible laminar flow.

The Navier-Stokes equations are complemented with the following boundary conditions. On the solid wall of artery, the no-slip condition is imposed.

$$u=0 \quad (3)$$

At the inlet of artery, a fully developed velocity profile is assumed. Here U_0 is the normal inflow velocity.

$$u=U_0 n \quad (4)$$

No viscous stress boundary condition specifies vanishing viscous stress along with a Dirichlet condition on the pressure:

$$p=p_0;\quad \left[(\nabla u)+(\nabla u)^T\right]n=0 \quad (5)$$

To solve the Navier-Stokes equation, the boundary conditions being the known values of velocity and pressure are introduced at inlet and outlet respectively [11]-[13].

B. Structural mechanics model

The structural mechanics of the artery is defined by the global equilibrium equations based on the stress components. In a general form, this is given by:

$$-\nabla\sigma=F_v \quad (6)$$

where σ is the stress tensor. The artery itself was taken to consist of a Neo-Hookean hyperelastic material.

The total strain tensor is written in terms of the displacement gradient.

$$\varepsilon=\frac{1}{2}\left[(\nabla u)^T+\nabla u\right] \quad (7)$$

This extension of Hooke's law is thought to better approximate the stress-strain relationship that exists within the artery. The above equations are complemented with following boundary conditions:

Roller boundary condition:

$$n\cdot u=0 \quad (8)$$

Fixed boundary load condition:

$$\sigma\cdot n=F_A \quad (9)$$

The boundary conditions introduced within the structural mechanics module take into account the coupling of the two physics. This is done by taking the computed pressure obtained from the Nervier-Stokes equations and utilizing this as a load onto the artery walls [13].

IV. Finite Element Model of Thrombosis

Figure 2 shows the geometry of the MCA [Fig. 2(a)] considered for the analysis and its finite element model [Fig. 2(b)]. The average length of the MCA is 1.6 cm with a range of 0.5 cm to 3 cm. The diameter ranges from 0.3 cm to 0.5 cm [14]. The average thickness of the MCA tunica media at its origin was 0.012 cm [15]. Here we used artery length 3 cm keeping diameter of inner and outer part at 0.488 cm and 0.5 cm, respectively.

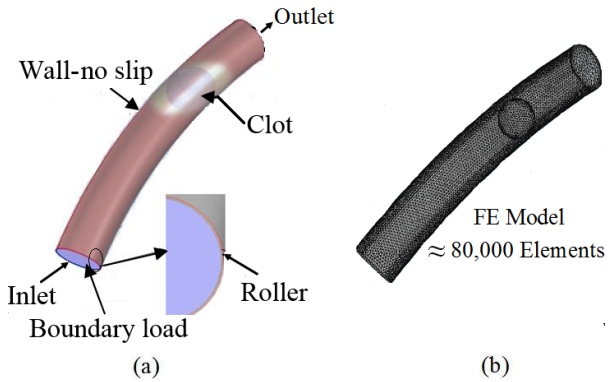


Fig. 2 (a) MCA geometry consisting blood clot inside, and indicating the boundary conditions, and (b) FE model of the proposed MCA geometry.

The velocity of blood flow through MCA has been reported as 0.583 m/s [16]. A time-dependent flow problem was first solved by setting this velocity condition at the inlet and zero pressure condition at the outlet of the selected artery. Flow through the vessel is driven by a pressure difference between the inlet and the outlet of the system. For structural mechanical analysis we used stationary solver. The material properties used in the simulation are summarized in Table I.

TABLE I
Material Properties Used in the FE Analysis [17]

Material	Property	Value
Blood	Density	1060 kg/m^3
	Dynamic Viscosity	0.005 Pa.s
Artery wall	Density	960 kg/m^3
	Poisson Ratio	0.45
	Young's Modulus	100 MPa

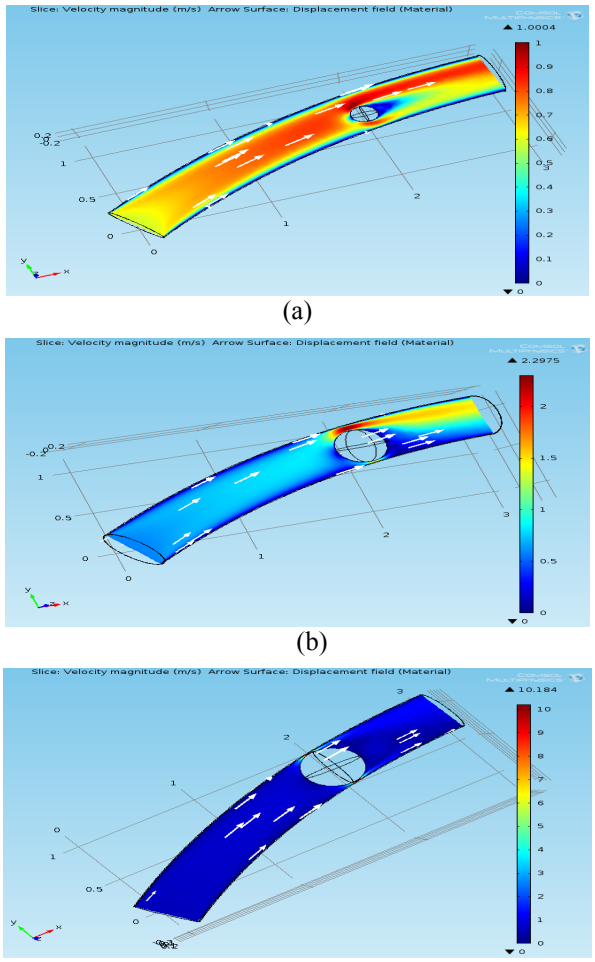


Fig. 3 Velocity magnitude of blood through the MCA with clot radius of (a) 0.1 cm (b) 0.2 cm and (c) 0.23 cm.

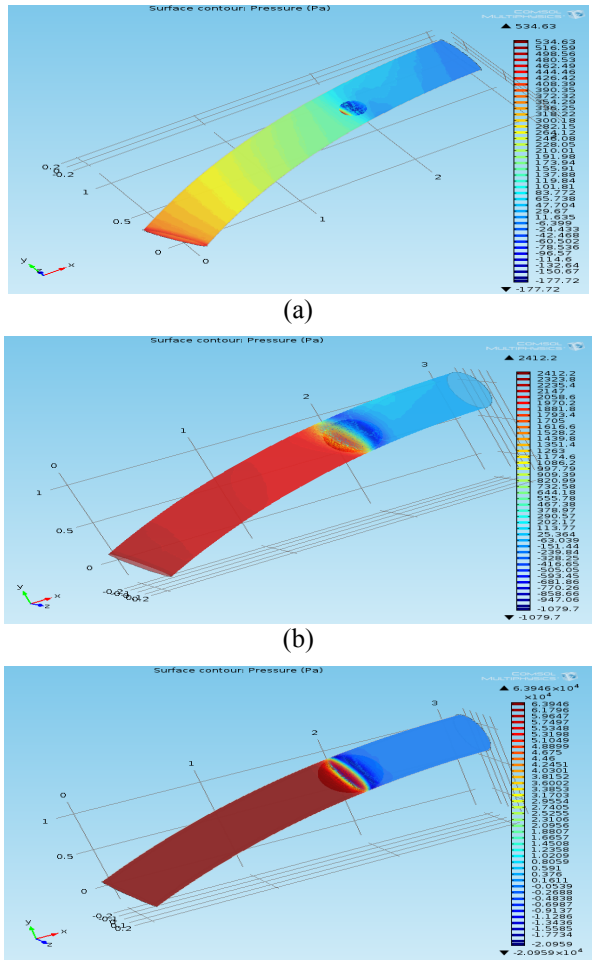


Fig. 4 Influence of blood clot on the pressure distribution along blood for clot radius of (a) 0.1 cm (b) 0.2 cm and (c) 0.23 cm.

The model consists of about 80,000 tetrahedral elements and the mesh for the blood and artery regions enables us to solve for 168960 and 218763 degrees of freedom, respectively.

V. Simulation Results and Discussions

We have designed 3D computational models mimicking an artery, and incorporated different occlusion bodies inside this artificial arterial structure. These occlusion bodies are representative of blood clots or thrombus as seen in a large vessel thrombosis. Figs. 3-5 show the MCA thrombosis mechanism with occlusion of blood into a spherical clot with radius 0.1, 0.2 and 0.23 cm detailing velocity, pressure distribution among blood, and stress to artery wall structure using a colour table. These clot sizes represent a thrombus at its initial formation stage (20 % of artery diameter) to almost blocking stage (92 % of artery diameter). With the rise of clot size, velocity of blood, pressure among blood and stress on the wall increases abruptly. Considering Fig. 3(a)-(c), we can see that the velocity rises by 9 times (1 m/s to 9 m/s) near the clot region when the clot size increases form 20 % to 92 % of the artery diameter (0.5 cm).

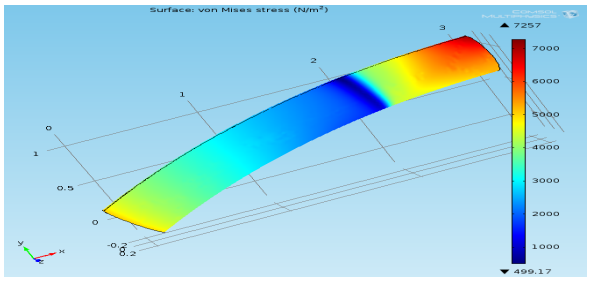
At the same time, developed blood pressures are inversely correlated with the velocity. For the occlusion, blood pressure rises before the clot and decreases after the clot. From Fig. 4(a)-(c), we can see that pressure rises abruptly from 2412.2 Pa to 6.39×10^4 Pa which may cause a thrombotic stroke. Stroke condition varies for different people. When diastolic pressure rises above approximately 12000 Pa (90 mm Hg), it increases the risk of stroke [18].

Fig. 5 shows the variation of stress on artery wall due to thrombus of various dimensions. With the rise of clot size, stress also rises. But this stress rising condition is highly non-linear. Maximum stress on artery wall raises from 3000 to 6.65×10^4 N/m^2 in various clot conditions. It is enough to cause haemorrhage if it finds a weak point on artery wall [19]. Brain haemorrhage is a type of stroke wherein damage induced in the artery results in localized bleeding in the brain and destroys the tissue and cell in the vicinity. It is a severe condition, wherein the person needs immediate medication/treatment. If the person is not subjected to proper treatment in time, it can result in further complications like loss of brain function and cause coma or death. However, the maximum tolerable stress of artery wall is not specified and depends on individuals conditions.

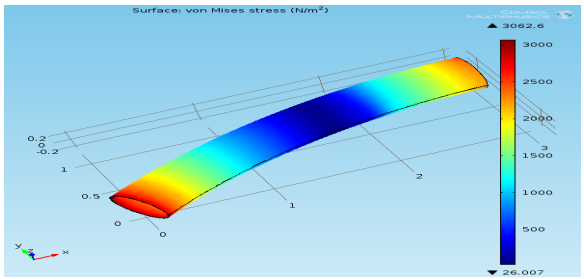
Fig. 6 shows inverse correlation between velocity and pressure with varying coordinates, and displays relative condition of velocity and pressure near clot. We can divide the pressure as positive and negative. Positive pressure works against the flow of blood. On the other hand, negative pressure increases the blood flow velocity. In various clot conditions, near clot we see the rise of negative pressure. That's why velocity there gets increased. With the rise of positive pressure, velocity of blood decreases.

VI. Conclusions

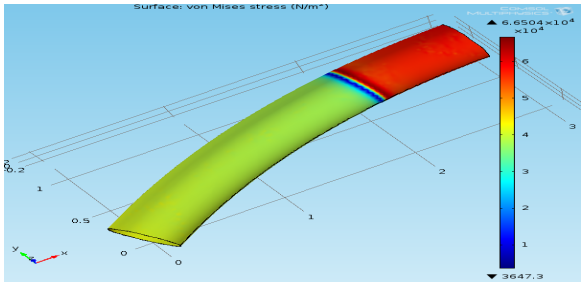
This works presented a three dimensional FE model and its numerical simulation results for the study of the blood flow along the middle cerebral artery (MCA) in the presence of clot



(a)



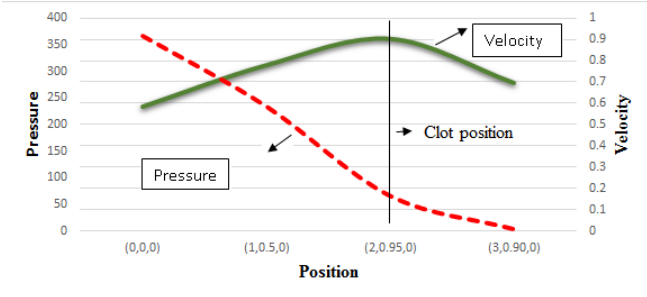
(b)



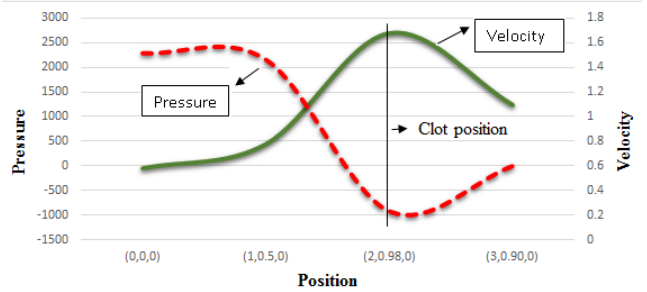
(c)

Fig. 5 Variation of stress on artery wall due to blood clot of different sizes of (a) 0.1 cm (b) 0.2 cm and (c) 0.23 cm.

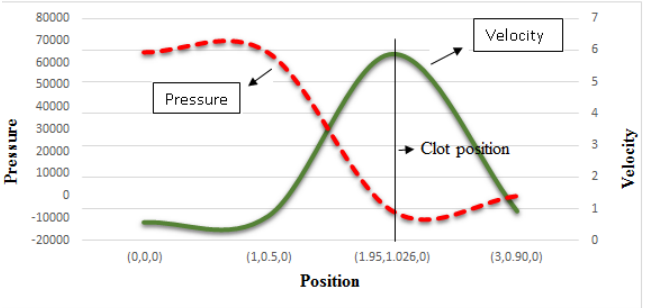
(thrombus) considering fluid-structure interaction. The thrombus formation in the middle cerebral artery causes disturbance in blood flow. The model is used to explore the effects of minimal to maximal clot size formation in the blood flow. With the increase of clot size, the blood pressure before the clot rises significantly which then may result in a stroke. Before the clot, the stress on the artery wall also rises with the increase in clot size. Such stress on the artery wall may cause haemorrhage in a weak point. We believe that our proposed model will be helpful to predict the effects of thrombosis in the shorter period of time, and thus would enable the physicians and the scientists to take a better preventive measure.



(a)



(b)



(c)

Fig. 6 Velocity and pressure curves around the thrombosis in the MCA for various clot sizes of (a) 0.1 cm (b) 0.2 cm and (c) 0.23 cm.

REFERENCES

[1] "Types of stroke," National Stroke Association [Online], Available: http://www.stroke.org/site/PageServer?pagename=type.

[2] "Rapid Recovery Hyperbarics," [Online]. Available: http://www.rapidrecoveryhyperbarics.com/strokeresearch29.html

[3] "Thrombotic Stroke," Drugs.com [Online], Available: http://www.drugs.com/ health-guide/thrombotic-stroke.html

[4] N. Kumar G. Ramunigari, and D. Roy, "Numerical simulations of thrombosis," *Chron. of Young Scient.*, vol. 4, no. 2, pp. 130-137, 2013.

[5] V. Sheth, and A. Ritter, "Using computational fluid dynamics model to predict changes in velocity properties in stented carotid artery," in *Proc. of the COMSOL conference*, 2010, paper 8800.

[6] Z. Xu, N. Chen, S. C. Shadden, J. E. Marsden, M. M. Kamocka, E. D. Rosen, and M. Alber, "Study of blood flow impact on growth of thrombi using a multiscale model," *Soft Matter*, vol. 5, pp. 769-779, 2008.

[7] U. Usman, and M. Wasay, "Mechanism of neuronal injury in Cerebral Venous Thrombosis," *J. Pak. Med. Assoc.*, vol. 56, pp. 509-512, Nov. 2006.

[8] B. Liu, "Computer simulations of flows in curved tubes with stenosis," in *Proc. of the COMSOL conference*, 2005, paper 1012.

[9] "Blood Vessels of the Brain," The Internet Stroke Center [Online], Available: http://www.strokecenter.org/professionals/brain-anatomy/ blood-vessels-of-the-brain/.

[10] "Anatomy of the Brain," University of British Columbia neuroanatomy website [Online], Available: http://www.neuroanatomy.ca/ stroke_model/mca_info.html.

[11] R. K. Bansal, *A Textbook of Fluid Mechanics and Hydraulic Machines,* Revised 9th ed., Laxmi Publications (P) Ltd., New Delhi, 2010.

[12] P. N. Modi, and S. M. Seth, *Hydraulics and Fluid Mechanics including Hydraulic Machines*, 14th ed., Standard Book House, Delhi, 2005-2006.

[13] *COMSOL Multiphysics Reference Manual*, COMSOL, 1998-2013.

[14] A digital library of anatomy information homepage on anatomy atlases.[Online].Available:http://www.anatomyatlases.org/AnatomicVariants/Cardiovascular/Text/Arteries/CerebralMiddle.shtml

[15] Idowu, Sakunbi, Malamo, and Akang, "Anomalies, aneurysm and histology of the middle cerebral artery in Nigerian Africans," *Int. J. Morphol.*, vol. 26, pp. 1023-1027, 2008.

[16] W. A. Kofke, P. Brauer, and R. Policare, "Middle cerebral artery blood flow velocity and stable xenon-enhanced computed tomographic blood flow during balloon test occlusion of the internal carotid artery," *Stroke*, vol. 26, pp. 1603-1606, 1995.

[17] A. A. Dobre, and A. M. Morega, "Blood flow-vessel interaction in a subclavian aneurysm", *INCAS Bulletin*, vol. 3, pp. 59-65, 2011.

[18] "High Blood Pressure (Hypertension)," National Stroke Association [Online], Available: http://www.stroke.org/site/PageServer?pagename = highbloodpressure.

[19] H. J. Steiger , R. Aaslid , S. Keller , and H. J. Reulen , "Strength, elasticity and viscoelastic properties of cerebral aneurysms," *Heart and Vessels*, vol. 5, pp. 41-46, 1989.

Performance Analysis of BPPM and M-ary PPM in Free-Space Optical Communication Systems under Gamma-gamma Distribution Using APD and PIN

Md. Athaharul Islam * and Ashraful Bari Chowdhury

B.Sc. Student, Department of EEE, Ahsanullah University of Science and Technology, Dhaka, Bangladesh.
* E-mail: a1b2i3r45@gmail.com

Abstract—**Free-space optical (FSO) communication offers highly directional, high bandwidth communication channels, which is why it is often used in digital communication systems. However, atmospheric turbulence has a voluminous degradation in the performance of FSO communication system. Therefore, different modulation techniques are introduced to improve the performance of the FSO communication systems. In this paper, the modulation schemes considered for the proposed FSO system, are binary pulse position modulation (BPPM) and M-ary pulse position modulation (M-ary PPM). The channel fading is characterized using gamma-gamma distribution. APD and PIN photodiode are used as photodetectors. Performance of the FSO systems has been compared in terms of their bit error rates (BERs). The results obtained in the paper show the suitability of BPPM over M-ary PPM for both photodetectors and also indicate the suitability of APD over PIN photodiode for the considered modulation techniques.**

Keywords—Avalanche photodiode (APD), Bit error rate (BER), Free-space optics (FSO), Gamma-gamma distribution model, M-ary pulse position modulation (M-ary PPM).

I. INTRODUCTION

A Free Space Optical communication system is generally thought of as a system in which, there are no apertures or materials between the transmitter and receiver, i.e. the transmission medium is the air of the respective environment and the link is line-of-sight (LOS). Though, we assume an LOS path exists between the transmitter and receiver array, the transmitted field from laser will propagate through an atmosphere and may experience several effects [1]. So, the performance of FSO link is inherently affected by atmospheric conditions.

Turbulence is caused by the inhomogeneity of temperature and pressure and is responsible for refractive index variation of the air [2]. Turbulence causes amplitude and phase fluctuations in the optical beam traveling through the atmosphere. Such fluctuations deteriorate signal intensity at the receiver, increase bit error rate (BER). The effect of atmospheric turbulence on light waves has been extensively studied and various models for atmospheric channel have been proposed from time to time. Among them, the most widely used model in FSO communication systems for the probability of density function (PDF) is log-normal distribution due to its simplicity. However, it is only applicable to the weaker turbulence regime [3]. As the strength of turbulence increases, multiple scattering effects become important and significant deviations from log-normal statistics are exhibited in experimental data. Hence, we consider gamma-gamma distribution in this paper. Gamma–gamma distribution is a two parameter distribution which is based on a doubly stochastic theory of scintillation. Moreover, it has been shown that gamma–gamma distribution is valid for both weaker and stronger turbulence regimes [4].

For a high bit rate optical communications, APD is frequently the photodetector of choice due to its internal gain, which provides better sensitivity and fast response times than PIN photodiode. APDs are generally recommended for high bandwidth applications or where internal gain is needed to overcome high pre-amp noise [5]. However, APD has many drawbacks too, some of them are the high cost, the complex circuit, the temperature sensitivity and the need of high voltage power. This makes PIN more reliable than APD especially in lower data rates. Here we use silicon (Si) APD in the system because of the ionization ratio (ς).

Pulse position modulation (PPM) using intensity modulation with direct detection (IM/DD) is among the most widely considered modulation techniques for FSO system. PPM achieves high power efficiency and improves the system performance at the expense of reduced bandwidth efficiency compared to other modulation schemes such as on-off keying (OOK) [6]. Adopting low and high order M-ary PPM for FSO is shown to enhance the power and bandwidth efficiencies as compared to conventional modulation techniques [6], [7].

In this paper, we propose an analytical approach to evaluate the performance of a FSO system, which adopts low and high order M-ary PPM using two different type of receivers (APD and PIN); in order to make a comparison of performance between them. The performance results are evaluated in terms of bit error rate (BER) assuming that, the channel is modelled using gamma-gamma distribution.

II. SYSTEM MODEL

In the proposed FSO system the source bits are modulated using low and high order M-ary PPM. The transmitted light propagates over gamma-gamma distribution channel and additive white Gaussian noise (AWGN) due to shot noise and thermal noise are added at the receiver input. As receiver, we use PIN and APD to convert the light intensity into electrical current. It is assumed that full channel state information is available at the receiver. For M-ary PPM, one data symbol consists of M number of time slots, thereby transmitting

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

$L=log_2^M$ bits per symbol. In the transmitter, the signals are described by the waveforms,

$$X = S_M(t)=\begin{cases} 1, & for\ t \in (\frac{(M-1)T_s}{L}, \frac{MT_s}{L}) \\ 0, & elsewhere \end{cases} \quad (1)$$

At the receiver the received signal P in terms of intensity after optical/electrical conversion, is given by [8];

$$P = \eta h I_o + n \quad (2)$$

Here, $I_0 =$ the average transmitted light intensity, $I = hI_0 =$ the corresponding received intensity in an ON PPM slot, $h =$ the channel fading coefficient, $n =$ receiver noise.

III. CHANNEL MODEL

Gamma-Gamma Distribution:

The gamma-gamma model first proposed by Al-Habash et al. [9] is valid for all turbulence regimes from weak to strong regimes. It is based on the assumption that the fluctuation of the laser beam propagating through the turbulence medium consists of the refraction (I_1) and scattering (I_2) effects. Subsequently, the normalized irradiance is expressed as the normalized product of two independently statistical random processes I_1 and I_2 ; they are given in (3).

$$I = I_1 I_2 \quad (3)$$

Here, I_1 and I_2 represent the refraction and scattering effects, respectively and both are governed by the gamma distribution [9]. Thus, the Probability density function (PDF) of the gamma-gamma model for the received irradiance fluctuation is derived as (4).

$$f(I)=\frac{2(\alpha\beta)^{(\alpha+\beta)/2}}{\Gamma(\alpha)\Gamma(\beta)} I^{\frac{(\alpha-\beta)}{2}-1} K_{(\alpha-\beta)}\left(2\sqrt{\alpha\beta I}\right),\ I>0 \quad (4)$$

Here, I is the signal intensity, $\Gamma(.)$ is the gamma function, and $K_{(\alpha-\beta)}$ is the modified Bessel function of the second kind and order $\alpha\beta$. α and β are PDF parameters describing the scintillation experienced by plane waves [9].

IV. THEORETICAL ANALYSIS

The generalized BER of M-ary PPM in the gamma-gamma channel, P_b^M is given by [10].

$$P_b^M \leq \frac{M}{2}\sum_{i=-N, i\neq 0}^{N} w_i |x_i| Q\left\{\frac{E\{K_s\}x_i^2}{\sqrt{F\,E\{K_s\}x_i^2 + K_n}}\right\} \quad (5)$$

Here, M is the modulation level. In this paper, we have considered the value of M to be 2, 8 and 256.

w_i and x_i $(i = -N, -N+1, \ldots\ldots -1, 1, 2, \ldots\ldots N)$ are the weight factors and the zeros of the Hermite polynomial, respectively. These are well tabulated and are easily available in the literature/internet. The value of N can be chosen depending on the desired value of accuracy.

$E\{K_s\}$ is the average photons per PPM slot and it is a function of the mean (m_R) and the variance of the gamma-gamma distribution and has the form

$$E\{K_s\} = e^{\left(\frac{\sigma_R^2}{2} + m_R\right)} \quad (6)$$

Here, σ_R^2 is the Rytov variance. In (5), K_n is the total noise per PPM slot which results from background noise and thermal noise, is given by [11].

$$K_n = \frac{2\sigma_n^2}{(E\{g\}q)^2} + 2FK_b \quad (7)$$

Here, K_b is the average background noise photons per slot, $E\{g\}$ is the average gain of the APD, q is the electron charge and the variance, σ_n^2 of the thermal noise in a PPM slot is defined by [11]

$$\sigma_n^2 = \left(\frac{2KTT_{slot}}{R_L}\right) \quad (8)$$

Here, T is the effective absolute temperature of the receiver, K is the Boltzmann constant, R_L is the APD load resistance, T_{slot} is the PPM slot duration which is related to the data rate, R_b, by [11]

$$T_{slot} = \frac{\log_2(M)}{MR_b} \quad (9)$$

Considering the statistical nature of avalanche multiplication by using McIntyre formula, the excess noise factor, F in (5) and (7), is given by [12],

$$F = \varsigma E\{g\} + (1-\varsigma)\left(2 - \frac{1}{E\{g\}}\right) \quad (10)$$

Here, ς is the ionization factor. The McIntyre formula can be approximated for $\varsigma < 0.1$ and $E\{g\} > 20$ without significant loss of accuracy and so the noise factor, F, of APD is defined by [11],

$$F \approx 2 + \varsigma E\{g\} \quad (11)$$

Using the derived value of F from (11), first in (7) and then in (5), the BER expression for M-ary PPM using APD photodector is obtained.

Similarly, the BER expression for M-ary PPM using PIN photodiode is attained by setting both noise factor *(F)* and gain ($E\{g\}$) to unity in (10) and then using the obtained values in (7) and (5) [10].

V. RESULTS AND DISCUSSIONS

Based on the described model, the BER of 10^{-9}, which is considered as a practical performance target for FSO link [13], is calculated for gamma-gamma distribution and the obtained results are displayed in Figs. 1-8. The simulation is performed using matlab. System parameters used for computation are shown in Table I [6], [10].

TABLE I. SYSTEM PARAMETERS

Parameter Name	Value
Average gain, $E\{g\}$ (for PIN)	1
Background noise, K_b	10 photons per PPM slot
Data rate	2.4 Gbps
Ionization factor, ς	0.028
Length, L	100 m
Load resistance, R_L	50Ω
Modulation level, M	2, 8, 256
Rytov variance, σ_R^2	0.7
Temperature, T	300 K
Wavelength, λ	850×10^{-9} m

Variations of BER with the average number of photons received per PPM bit (logarithmic), which are equal to $log_{10}E\{K_s\}$/(number of bits in PPM symbol), are shown in the following figures.

Fig. 1, 2 and 3 depict the plots of bit error rate vs. average number of photons per PPM bit for different values of APD gain; respectively for M = 2, 8 and 256. From these figures it is clear that, for all modulation levels (M=2, 8, 256); increase in APD gain causes decrease in average photon number per PPM bit, indicating improvement in power efficiency.

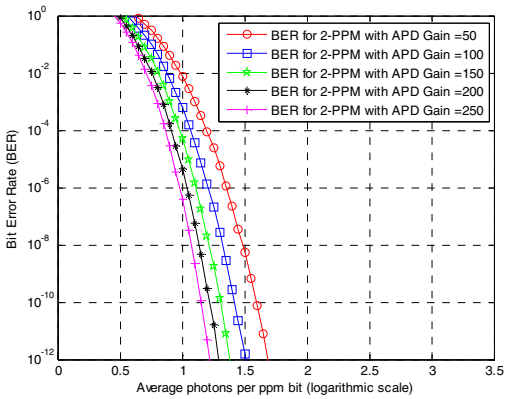


Fig. 1 Plots of BER vs. Average photons per PPM bit for 2-PPM.

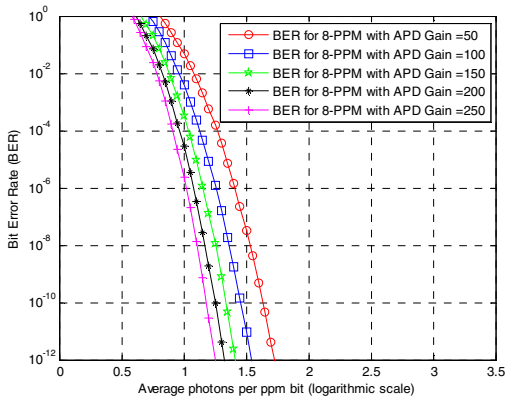


Fig. 2 Plots of BER vs. Average photons per PPM bit for 8-PPM.

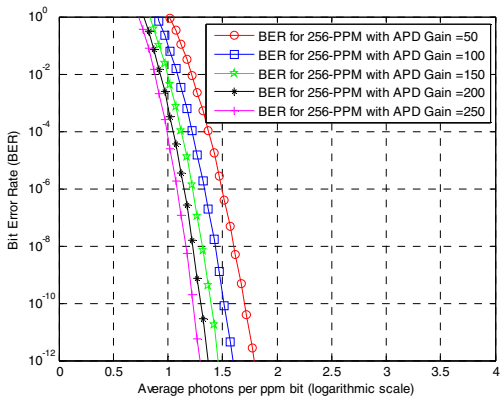


Fig. 3 Plots of BER vs. Average photons per PPM bit for 256-PPM.

Fig. 4 depicts the plots of bit error rate vs. APD gain for 256-PPM with different values of average photons per PPM bit ($E\{K_s\}$). It clearly shows that, up until APD gain ($E\{g\}$) reaches 150; increase in average photons per PPM bit ($E\{K_s\}$) causes decrease in BER. However, from then on, despite the increase in $E\{K_s\}$ along with $E\{g\}$, BER does not decrease anymore. This is because, a large $E\{g\}$ leads to an increase in excess noise factor *(F)*; which impedes the reduction of BER.

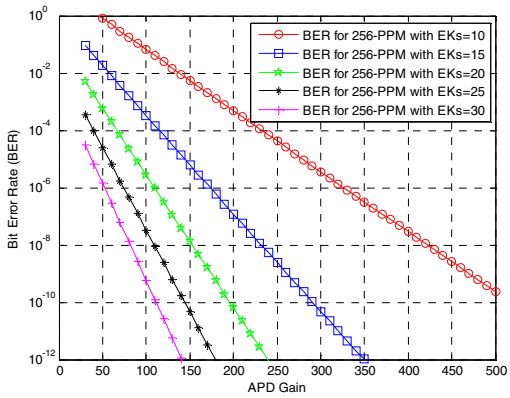


Fig. 4 Plots of BER vs. APD gain for 256-PPM.

Fig. 5, 6 and 7 illustrate the plots of bit error rate vs. average photons per PPM bit for low and high order (M=2, 8, 256) M-ary PPM with the value of gain, $E\{g\} = 150$ for APD under gamma-gamma distribution. From these figures it is clear that, for all cases (M=2, 8, 256), the average number of photons per PPM bit ($E\{K_s\}$) is almost 300 photons fewer, with APD than with PIN photodiode. This significant improvement in performance with APD, is due to the high internal gain ($E\{g\}$) of the APD.

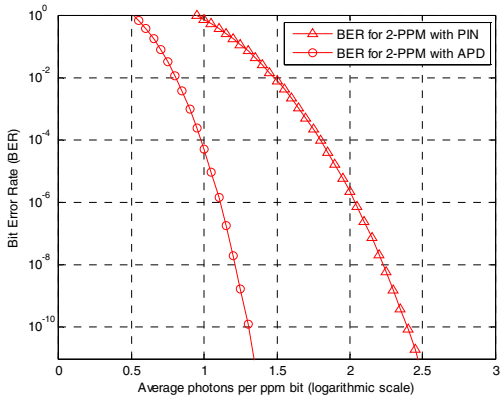


Fig. 5 Plots of BER vs. Average photons per PPM bit for 2-PPM with both PIN and APD photodetectors.

Fig. 8 represents a comparative illustration of Fig. 5, 6 and 7. It portrays the plots of bit error rate vs. average number of photons per PPM bit for low and high order (M=2, 8, 256) M-ary PPM using both PIN and APD photodetectors with the value of gain, $E\{g\} = 150$ for APD; under gamma-gamma distribution. Fig. 8 clearly shows that, as the modulation level (M) increases from 2 to 8 and then finally to 256, average number of photons per PPM bit ($E\{K_s\}$) slightly increases with higher values of M for both PIN and APD photodetectors; which implies that the performance of the system, slightly worsens with higher order PPM.

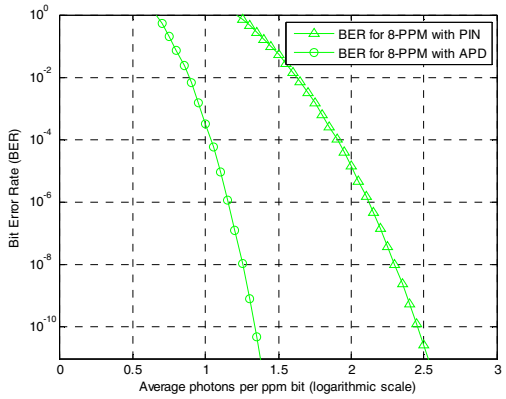


Fig. 6 Plots of BER vs. Average photons per PPM bit for 8-PPM with both PIN and APD photodetectors.

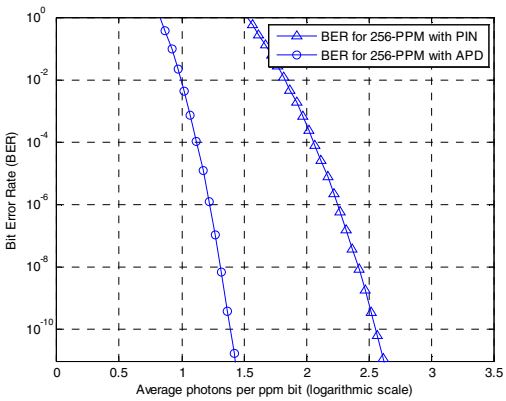


Fig. 7 Plots of BER vs. Average photons per PPM bit for 256-PPM with both PIN and APD photodetectors.

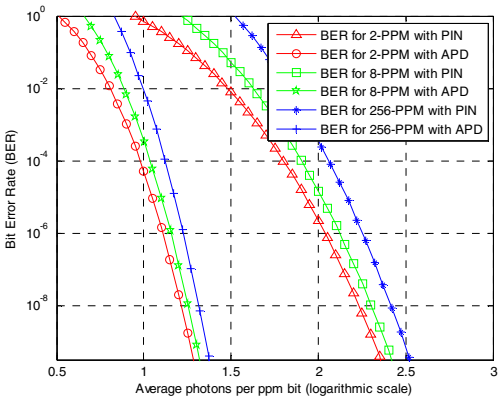


Fig. 8 Variations of BER with Average photons per PPM bit for different values of M with both PIN and APD photodetectors.

VI. CONCLUSIONS

In this paper, we have analyzed a FSO system employing low and high order M-ary PPM under gamma-gamma distribution using PIN and APD photodetectors. From the results we have concluded that, APD is better than PIN photodiode as photodetector with M-ary PPM in the cases of M=2, 8 and 256; due to the significant reduction of average number of photons per PPM bit with APD compared to PIN. It implies that, with M-ary pulse position modulated systems, APD is a more power-efficient solution compared to PIN photodiode. Moreover, we have found the optimal values of the APD gain up to which the excess noise factor does not deteriorate the system performance.

From the results, we have also concluded that, due to the slight reduction in the value of the average number of photons per PPM bit with BPPM compared to the higher order M-ary PPM, BPPM system is preferable over higher order M-ary PPM under gamma-gamma distribution for both PIN and APD photodetectors. Finally, it can be said that the optimal performance of the proposed system is achieved for BPPM with APD as receiver.

REFERENCES

[1] Majumdar A.K., "*Free-space laser communication performance in the atmospheric channel*", Journal of Optical and Fiber Communications Reports 2(4), 2005, pp. 345–396.

[2] S. Karp, R. M. Gagliardi, S. E. Moran, and L. B. Stotts, "*Optical channels*," New York, Plenum press, 1988.

[3] Bobby Barua, Mirza Moazzem Hossain, Md. Rezwan Islam, Md. Khairul Bashar, "Performance Evaluation of Different Type of Channel Models in FSO Communication," *International Journal of Science and Advanced Technology* (ISSN 2221-8386), Volume 1 No 5 July 2011.

[4] M. Uysal, J. T. Li, and M. Yu, "Error rate performance analysis of coded free-space optical links over gamma-gamma atmospheric turbulence channels," *IEEE Transactions on Wireless Communications,* vol. 5, pp. 1229-1233, June 2006.

[5] Joe C. Campbell, "Recent advances in telecommunications avalanche photodiodes," *Journal of Lightwave Technology*, vol. 25, no. 1, pp. 109-121, Jan. 2007.

[6] M. R. Abaza, N. A. Mohammed, and M. H. Aly, "BER performance of M-ary PPM free-space optical communications with channel fading," in *Proceeding of the 8th International Conference on High Capacity Optical Networks & Enabling Technologies (HONET 2011)*, Riyadh, Saudi Arabia, Dec. 19–21, 2011, pp. 111–115.

[7] J. Yan, Z. Zheng, W. Hu, and A. Xu, "Improved performance of M-ary PPM free-space optical communication systems in atmospheric turbulence due to forward error correction," in *Proceeding of the 10th International Conference on Communication Technology (ICCT)*, Guilin, China, Nov. 27–30, 2006, pp. 1–4.

[8] Bobby Barua, Dalia Barua "Evaluate the Performance of FSO Communication link with Different Modulation Technique under Turbulent Condition", *14th International Conference on Computer & Information Technology*, Dec 22-24, 2011, ICCIT 2011.

[9] M.A. Al-Habash, L.C. Andrews and R.L. Phillips, *Optical Engineering 40*, pp. 1554-1562 (2001).

[10] Jagtar Singh, V. K. Jain, "Performance analysis of BPPM and *M*-ary PPM Optical Communication Systems in Atmospheric Turbulence", *IETE Technical Review*, vol. 25, no. 4, pp. 146-153, Jul. Aug. 2008.

[11] K. Kiasaleh, "Performance of APD-based, PPM free-space optical communication systems in atmospheric turbulence," *IEEE Transactions on Communications,* vol. 53, no. 9, pp. 1455–1461, Sep. 2005.

[12] McIntyre, R.J., *Multiplication Noise in Uniform Avalanche Diodes, IEEE Transactions on Electron Devices*, ED-13, pp. 164-168 (1966).

[13] Murat Uysal and Jing Li, "BER performance of Coded Free-Space Optical Links over Strong Turbulence Channels," *Proceeding of the 59th Vehicular Technology Conference (VTC)*, Milan, pp. 352-356, 2004.

Long-Reach Gbit/s Visible Light Communication exploiting White Hot-Spot

G. Cossu[1], P. K. Choudhury[2,*], A. M. Khalid[1], R. Corsini[1] and E. Ciaramella[1]
[1]Scuola Superiore Sant'Anna, TeCIP, Pisa, Italy
[2]Dept. of ECE, Khulna University of Engineering & Technology, Bangladesh
[*] E-mail: pallab@ece.kuet.ac.bd

***Abstract*— We demonstrated a visible light communication system based on the 'White-Hot Spot' concept, using a commercially available white-phosphorus LED, modulated by an adaptive DMT signal. For a long wireless distance spanning from 1.5 m to 7 m, a record transmission rate was achieved, ranging from 1.2 Gbit/s to 693 Mbit/s, within the Hot-Spot area. The presented results can be useful to design high-speed optical wireless transmission links.**

Keywords*— *Visible Light Communication, Light Emitting Diode, Discerete Multitone Modulation

I. INTRODUCTION

Visible light communication (VLC) using light emitting diodes (LEDs) are considered a potential solution for indoor wireless communications, because it provides a synergy between illumination and communication. Moreover, spectrally efficient modulation formats, mostly Discrete Multi-Tone (DMT) [1], allow VLC to be used for high speed communications [2], [3].

Compared to radio frequency systems, VLC has the advantage of good spatial confinement, with the ability to maintain sufficient power density over quite small areas. Recently, a localized VLC link type, referred as 'White Hot-Spots' (W-HS), was suggested to ensure intense and focused light for data communication [4]. This approach enables very high data rate with reduced channel distortion and interference compared to conventional uniform lighting modes [5]. This W-HS scheme has a great potential for applications that require high-speed wireless links in small beam footprint. Examples can be personal passenger light in the aircraft or illumination of desktops in cubicle offices. Clearly, in order to meet the end-user requirements the W-HS must be shaped, adjusting the distance, the divergence angle and the corresponding illuminated area.

Here we experimentally demonstrate high-speed VLC systems using different W-HS areas by means of a phosphorescent white LED and adaptive DMT modulation. In the hot-spot area the maximum data rate was between 1.2 Gbit/s and 693 Mbit/s for distances spanning from 1.5 m to 7 m. In particular, considering the typical ceiling-desk distance of 2.5 m, the achieved capacity was 1.1 Gbit/s which is the highest capacityobtained at this distance to the best of our knowledge.

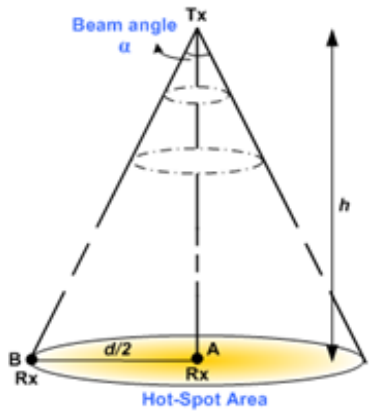


Fig. 1. *White Hot-spot scenario for VLC link. Tx: transmitter source; Rx: receiver.*

II. WHITE HOT-SPOT

Figure 1 shows a W-HS scenario. The power distribution of common LEDs generates a W-HS circular area, whose boundary circle is defined by the radius where the optical intensity is halved (point B) from the value at the centre (point A). This white area will be referred as 'Hot-spot Area' (HS-A). The HS-A diameter d is given by $d = 2h \tan(\alpha/2)$ where α is the beam divergence of the TX, h is the distance between the transmitter and (horizontal) plane where the receiver can be placed. Both parameters may be varied depending on the application (e.g., coverage area diameters of 1.5 m for cubical offices and 0.5 m for aircraft cabin seat [6]). From simple trigonometric calculation these coverage areas will be achieved by the divergence range from 16° to 20°. A spotlight with low emission angle can be used; in our experiment we used a commercial wide Field of View (FOV) LED chips and small FOV lens in order to reduce the divergence of the LEDs. View (FOV) LED chips and small FOV lens in order to reduce the divergence of the LEDs.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

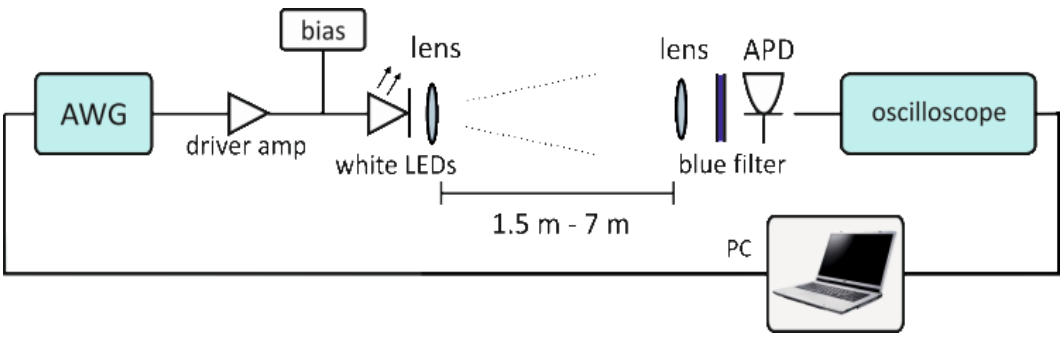


Fig.2: *Experimental setup for W-HS demonstration. AWG: Arbitrary Waveform Generator; APD: Avalanche PhotoDiode*

III. Experimental Set-up

In order to emulate the scenario of a VLC system exploiting the W-HS solution, we realized the experimental set-up shown in Fig. 2. The optical transmitter (Tx) was a board hosting three low cost phosphorus-white LEDs, closely mounted and connected in series. This source generates a total luminous flux of 400 lm at nominal driving currents of 350 mA, with 120° Lambertian emission, reduced down to 18° by using a proper low-cost lens. The receiver module (Rx) was placed in front of the LED and aligned to ensure the direct line-of-sight (LOS) between the Tx and Rx. The analogue receiver was an APD module consisting of an avalanche photodiode (APD, 0.42 A/W responsivity at 620 nm; 3.14 mm^2 active area) and an integrated TIA (Trans-Impedance Amplifier, 280 MHz 3-dB bandwidth). A dichroic optical band pass filter (Semrock, 97% transitivity at 452 nm, with 45 nm FWHM bandwidth) was placed in front of the APD to reject the slow phosphorescent components and a bi-convex glass lens (60 mm focal length) was used to focus the light onto the APD active area

The DMT waveform, consisting of 512 subcarriers within 205 MHz total bandwidth (optimized according to the channel response), was generated offline and uploaded into the Arbitrary Waveform Generator (AWG). The first carrier, corresponding to DC, was not modulated. The received signal was processed offline by a common PC. Before transmission, we first sent a signal probe having BPSK modulation on all the subcarriers; then we measured the Error Vector Magnitude (EVM) values of the received constellations and from these we evaluated the signal-to-noise ratio (SNR) values on the subcarriers [7] . In order to maximize the system capacity we adapted the modulation order and the power level depending on the channel SNR for each subcarrier [8]. The optimization of the system was performed by using the bit-power loading process to maintain the BER below the FEC limit [9]. We also applied a linear power pre-emphasis of 10 dB in the transmitted signal to take advantage of available electrical power. The DMT signal was first amplified by a power amplifier (25 dB gain, and 29 dBm minimum output power at 1-dB compression), then superimposed to a dc bias current (300 mA) via bias-T and sent to the LED. The output signal from the APD module was then sampled and stored by means of a real-time oscilloscope (Lecroy, 2 GSa/s) for offline demodulation. In DMT process the training sequence with zero-forcing equalization was used for estimation and correction of the channel. After the demodulation and signal correction, the BER value was computed comparing the original sequence with the recovered bits.

IV. Experimental Results and Discussions

We experimentally evaluated the system capacity for different HS-A values, keeping the beam divergence (18°) constant and varying the distance between transmitter and receiver *h*. At any HS-A condition, we measured the maximum data rate in the best and worst case; in the first case, we placed the receiver at the centre (point A in Fig. 1) and measured the maximum capacity. Then the receiver was moved to the boundary (point B) and we measured the minimum achievable data rate (worst case). Since the VLC link should not affect the general illumination level, we controlled the intensity level by a lux-meter in front of the Rx to be sure that the above level did not exceed the standard value for the working environments. Indeed we varied the distance *h* from 1.5 to 7 m, with an illuminance level from 460 to 17 lux (at point A) and from 380 to 14 lux (at point B). These values are well below the typical levels of the brightness in working environments (European Standard recommends 500 lx [10]).

Finally we report the results as a function of the height *h*, in order to evaluate the performance at usual distances between ceiling and working plane. We observed a good signal quality at every receiver position (point A and B). For

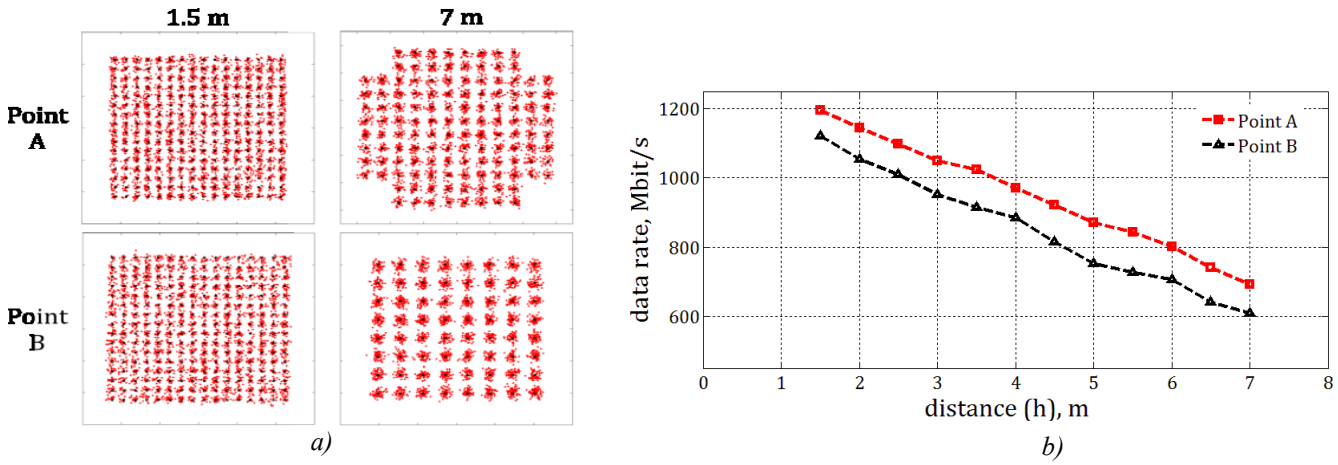


Fig.3: *Achieved capacity.*
a) received constellation diagrams: at point A, h=1.5 m and h=7 m (first row); at point B, h=1.5 m and h=7 m (second row).
b) capacity vs. distance h for both point A and B for each W-HS;

instance, Fig. 3*a* shows the received constellations of 2nd subcarrier at the distance of 1.5 m and 7 m for both the points A and B. In Fig. 3b we show the measured achievable data rate for the two positions i.e., at positions A and B varying h from 1.5 m to 7 m. When increasing the distance, the system capacity was mostly affected by the reduction in the received optical power. Note that the difference of data rate between the positions A and B of HS-A, which is mainly due to difference in the illuminance level, is approximately 100 Mbit/s. The data rate measured in the HS-A, considering *h*=1.5 m, is in the range of 1.1 Gbit/s (point B in Fig. 1) and 1.2 Gbit/s (point A). On the other hand, at *h*=7 m, a record system capacity was achieved within the range of 609 Mbit/s (point B) and 693Mbit/s (point A). The above mentioned data rates and system parameters allow to design a Gbit/s VLC system for LED based lighting. As an example, in an aircraft cabin scenario (h = 1.5 m, d = 50 cm) exploiting LED based lighting system [6], a guaranteed data capacity of 1.1 Gbit/s could be achieved by our VLC system. Moreover, in a typical Indoor application, where the ceiling-desk distance is around 2.5 m, the VLC transmission link can provide more than 1 Gbit/s within a circular area having of 80 cm diameter.

V. CONCLUSIONS

We experimentally demonstrated high capacity LED-based W-HSs for long distance indoor wireless communication. The proposed solution provides a record speed of 1.1 Gbit/s within coverage circle having 50 cm diameter at 1.5 m distance from the source. In a wider circular are (h =7 m) having 2.2 m diameter , 609 Mbit/s was also demonstrated.

REFERENCES

1. Armstrong, J., "OFDM for Optical Communications," Lightwave Technology, Journal of , vol.27, no.3, pp.189,204, Feb.1, 2009
2. A. M. Khalid, R. Corsini, P. Choudhury, E. Ciaramella, "1-Gb/s Transmission Over a Phosphorescent White LED by Using Rate-Adaptive Discrete Multitone Modulation" Photonics Journal, IEEE, Volume: 4 , Issue: 5Page(s): 1465 - , 1473Oct. 2012
3. G. Cossu, A. M. Khalid, P. Choudhury, R. Corsini, and E. Ciaramella, "3.4 Gbit/s visible optical wireless transmission based on RGB LED", Optics Express, Vol. 20 Issue 26, pp.B501-B506 (2012)
4. D.C. O'Brien et al.: 'Short-Range Optical Wireless Comm-unications', Wireless world research forum, Vol. **2**, wiley, 2006.
5. T.Borogovac et al.: Proc Globecom 2010.
6. C. Quintana, et.al.: 'Ethernet to Visible-Light Communications Adapter for In-Flight Passenger Data Networking', IEEE CWCN2011, Shanghai (China), 28-30 Oct. 2011
7. R. A. Shafik, et.al.:, 'On the extended relationships among EVM, BER and SNR as performance metrics', in Proc. ICECE, Dec. 19–21, 2006, pp. 408–411.
8. D.Hughes-Hartogs.: 'Ensemble modem structure for imperfect transmission media', U.S. Patent 4 833 796, May 1989.
9. ITU-T Recommendation, G.709/Y.1331.
10. European standard EN 12464-1: Lighting of indoor work places, 2003.

Analysis of DCO-OFDM and Flip-OFDM for IM/DD Optical-Wireless System

Rashed Islam, Pallab Choudhury* and Muhammad Ashiqul Islam

Department of Electronics and Communication Engineering
Khulna University of Engineering & Technology (KUET)
Khulna-9203, Bangladesh
E-mail: pallab_eee@yahoo.com*

Abstract— **The paper analyzes the performance of two forms of orthogonal frequency division multiplexing (OFDM) techniques called as DC biased Optical OFDM (DCO-OFDM) and Flip- OFDM for intensity modulated direct detection (IM/DD) system. The aforementioned OFDM schemes are compared in terms of peak-to-average power ratio (PAPR) and bit error rate (BER) as a function of electrical energy per bit to noise power ratio ($E_{b(elec)}/N_0$) when the signal is affected by additive white Gaussian noise (AWGN). In particular, 7-dB and 13-dB DC bias are used for DCO-OFDM in BER analysis and finally compared with flip-OFDM for the different data formats such as 4-,16-,64-,256- QAMs. The results show that flip-OFDM performs better in BER analysis with reduced value of ($E_{b(elec)}/N_0$) and thus, it can be considered as a power efficient unipolar modulation format. However, Flip-OFDM signal provides a high PAPR value that can deteriorate the overall system performance. Finally, OFDM signal performance tradeoff is measured as a function of transmitted optical power, QAM size, and spectral efficiency.**

Index Terms— **DCO-OFDM, Flip-OFDM, PAPR, BER, Intensity Modulation, Direct Detection.**

I. Introduction

Orthogonal frequency division multiplexing (OFDM) is a fast-progressing multi-carrier modulation scheme that is widely used in most of major wired and wireless communication systems now-a-days [1][2]. Especially in radio frequency (RF) domain, OFDM has been employed as a core part of the physical interface in the communication systems such as worldwide interoperability for microwave access (WiMAX), long term evolution (LTE) mobile network. Moreover, OFDM also has a great potential to use in optical system including optical wireless, multimode and single mode fibre optic communication system [3].

For the cost effective design of OFDM based optical system, the earlier investigations are mainly focused on intensity modulated (IM) optical transmitter with direct detection (DD) optical receiver [4]. For IM/DD system, OFDM signal needs to be unipolar i.e., real and positive, which is quite different from conventional OFDM scheme with bipolar symbols used in RF communication. The real valued OFDM signal is generated by Hermitian symmetric input data with Inverse Fast Fourier Transform (IFFT). However, to generate OFDM signal with positive values, several forms of OFDM schemes have been proposed for IM/DD optical system such as DC-biased optical OFDM (DCO-OFDM), Asymmetrically clipped optical OFDM (ACO-OFDM), flip-OFDM [5][6][7].

In DCO-OFDM, a DC-bias is added to make the OFDM signal positive, where the required bias value depends on the peak-to-average power ratio (PAPR) of OFDM symbols. As the OFDM signal has high PAPR value, the transmit optical power need to be increased, which eventually makes the DCO-OFDM power inefficient. For a power efficient modulation format, ACO-OFDM is proposed to avoid the use of any DC-bias. In ACO-OFDM, only the odd-subcarriers are used to carry the information and clipped the negative values at the transmitter. It is shown that ACO-OFDM is more power efficient than DCO-OFDM, however, in terms of bandwidth utilization, ACO-OFDM is inefficient due to the use of half number of subcarriers for data. Recently, another modulation technique is proposed for unipolar signal called as flip-OFDM, where the positive real time domain signal is transmitted in two consecutive OFDM symbols. It is shown that flip-OFDM provides 50% less implementation complexities compared to ACO-OFDM [5].

This paper provides a comparative analysis between the conventional DCO-OFDM and recently proposed flip-OFDM for IM/DD optical system in terms of bit error rate (BER) performance, PAPR label and optical power efficiency. The BER curves are plotted against electrical energy per bit to noise power ratio assuming that the received signal is corrupted with Additive White Gaussian Noise (AWGN) [8]. The PAPR of DCO- and Flip- OFDM are computed against a range of predetermined threshold values in dB unit. Moreover, the effect of average transmitted optical power on the BER performance is also analysed for both the aforementioned OFDM schemes with respect to normalised bandwidth/bit rate. These results have practical implications in both the wireless and optical fibre IM/DD communication systems.

The rest of the paper is organised as follows. Section II gives a description of system blocks of the two OFDM schemes. Section III provides comparison among those OFDM schemes in terms of BER vs. ($\boldsymbol{E_{b(elec.)}/N_o}$), PAPR and $\mathbf{E_{b(opt.)}/N_0}$ vs. normalised bandwidth / bit rate. Finally, we provide the conclusion in section IV.

II. Study of Direct Detection OFDM Schemes

In this section, we provide brief description about the DCO-OFDM and flip-OFDM systems consecutively.

A. DCO-OFDM

DCO-OFDM transmitter and receiver are shown in Fig.1. At the first step of modulation procedure, a high rate binary

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

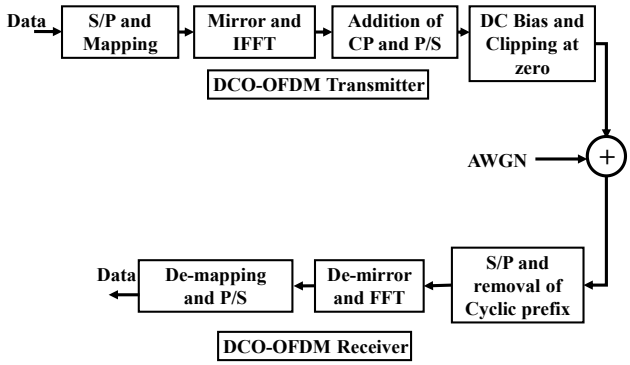


Figure 1: Implementation blocks for DCO-OFDM system

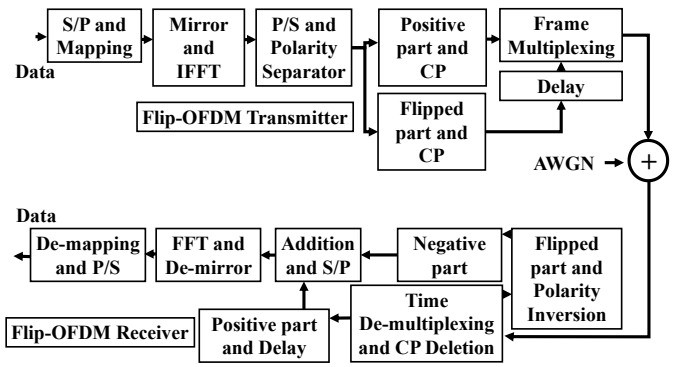


Figure 2: Implementation blocks for Flip-OFDM system

serial data sequence is divided into a number of parallel lower rate data streams that are transmitted simultaneously over a number of subcarriers. Each data stream goes through constellation mapping where every log_2(M) bits are mapped onto complex values according to an M-ary quadrature amplitude modulation technique. For $X_n = A_n + jB_n$, where, $n = 0, 1\ldots, N$-1 denotes the subcarrier number, Hermitian symmetry is taken by complex conjugate mirroring as defined below,

$$X_{N-n} = X_n^* \tag{1}$$

where $n = 0,1,\ldots,N/2$ and $X_0 = X_n = 0$. By taking N-point IFFT, real value k^{th} sample is denoted by x_k as,

$$x_k = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} X_n \exp\left(\frac{j2\pi nk}{N}\right) \tag{2}$$

Thus a DCO-OFDM symbol is generated with N point real value bipolar data sequence. Afterwards, cyclic prefix (CP) is added in each OFDM symbol and sends to parallel to serial converter for serial data transmission. Finally, DC bias is added and clipped the signal at zero label to ensure non-negative data transmission.

Additive white Gaussian noise (AWGN) is considered in the propagation channel to include noise effect on received data. At the receiver, DCO-OFDM frame is divided into (N+N_{CP}) parallel data streams. The number of DCO-OFDM symbols obtained from this serial to parallel conversion process is exactly equal to that of symbols transmitted. Cyclic prefix is then removed from each symbols and demodulation is performed successively using N point FFT. The complex valued data streams can be generated from the FFT operation as,

$$X_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} x_k \exp\left(\frac{-j2\pi nk}{N}\right) \tag{3}$$

The output of this FFT process is de-mirrored and applied to M-ary QAM demodulator to recover the original transmitted binary data streams. At the end, parallel data streams are again converted into serial data streams in binary form and compare with transmit data for BER measurement.

B. Flip-OFDM

Fig.2 shows the block diagram of the generation and reception of Flip-OFDM signal. In the transmitter, the real bipolar OFDM signal is generated following the same procedure like DCO-OFDM signal. However, the incoming bipolar signal is then decomposed into

$$x(k) = x^+(k) + x^-(k) \tag{4}$$

where the positive and negative parts of the Flip-OFDM are defined as,

$$x^+(k) = \begin{cases} x(k) & if\ x(k) \geq 0 \\ 0 & otherwise \end{cases}$$
$$x^-(k) = \begin{cases} x(k) & if\ x(k) \leq 0 \\ 0 & otherwise \end{cases} \tag{5}$$

where $k = 0,1,2,\ldots,N$-1. Two successive OFDM symbols are used to transmit the aforementioned components. The positive part $x^+(k)$ is transmitted in the first sub-frame, whereas the flipped signal $-x^-(k)$ is transmitted in the second sub-frame. Cyclic prefix is added to both the OFDM frames resulting the length of (N+N_{CP}), where N_{CP} is the length of cyclic prefix. Finally, the frames are multiplexed with all the positive real values and send to the desired propagation channel.

At the Flip-OFDM receiver, bipolar OFDM symbols are reconstructed using two data frames namely positive and flipped parts as shown in Fig.2. Firstly, the cyclic prefixes are removed from both of data frames and after this subtraction operation is performed which provides original bipolar signal as,

$$y(k) = y^+(k) - y^-(k) \tag{6}$$

where $y^+(k)$ and $y^-(k)$ represent positive and flipped data frames respectively. FFT operation is performed on this bipolar signal to generate the original complex conjugate data sequences. De-mirroring and QAM demodulation processes are used to reconstruct the binary stream from the complex valued signal.

III. COMPARISON OF DCO-OFDM AND FLIP-OFDM

A. $(E_{b(elec.)}/N_o)$ *vs. BER :*

The ($E_{b\,(elec.)}$ /N_0) vs. BER curves were computed for both type of OFDM signals to compare their performance in AWGN channel. The BER curves of 7 dB biased DCO-OFDM for 4, 16, 64, and 256-QAM are shown in fig.3. To reach BER at 10^{-4}, 600 OFDM symbols are used with 1024

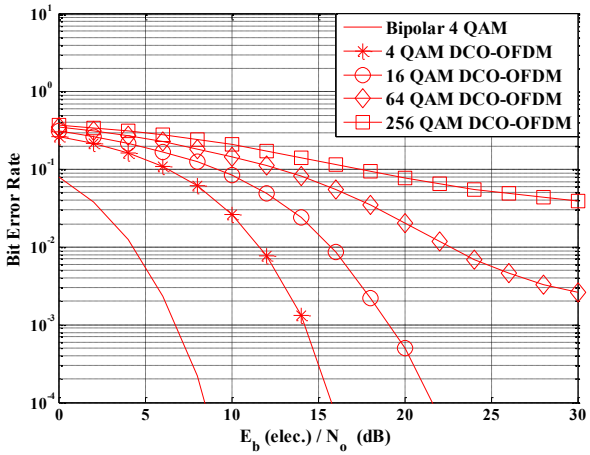


Figure 3: BER analysis for 7 dB biased DCO-OFDM

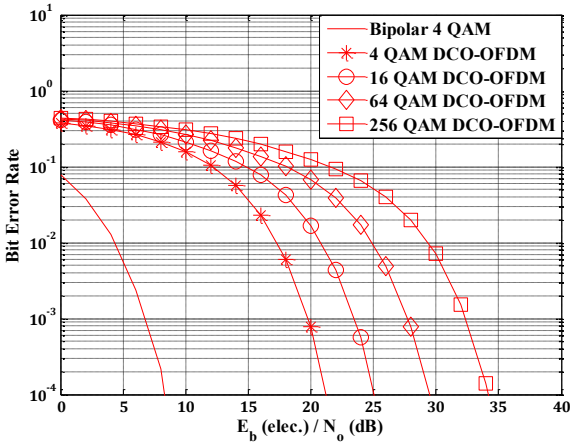


Figure 4: BER analysis for 13 dB biased DCO-OFDM

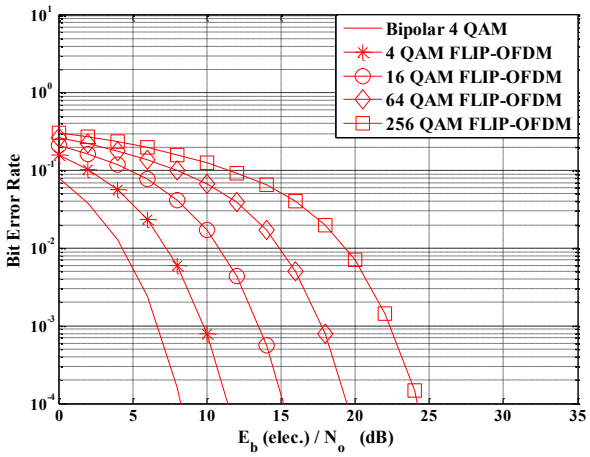


Figure 5: BER analysis for Flip-OFDM

subcarriers in each symbol and the cyclic prefix length was 16. Fig.3 shows that clipping noise dominants for higher constellation size especially in 64- and 256-QAM, which eventually provides the BER floor. However, 4-and 16-QAM modulation formats are still give a better BER performance with the required electrical power higher than the conventional bipolar OFDM. With the increase of DC bias to 13 dB, the clipping noise can be reduced that improves the BER performance for high constellation size modulation formats as shown in Fig. 4. For DCO-OFDM as justified from both the figures 3 and 4, the required $E_b(elec.)/N_0$ for a given BER is approximately equal to the $E_b(elec.)/N_0$ for a bipolar signal plus the bias level in dB. The BER vs $E_b(elec)/N_0$

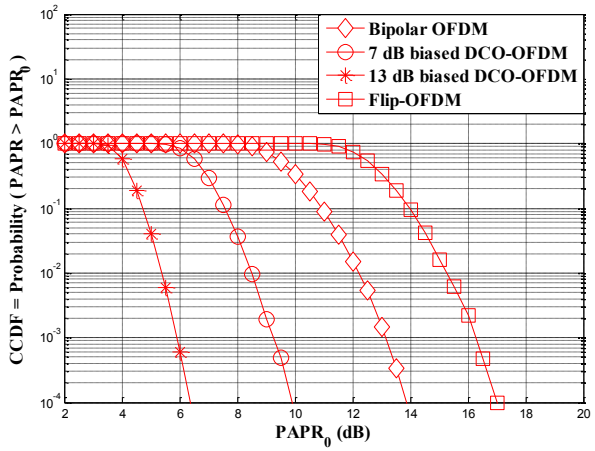


Figure 6: CCDF curves for DCO-OFDM and Flip-OFDM

performance for Flip-OFDM, as shown in fig.5, is more power efficient compare to DCO-OFDM as it requires much less power to achieve the target BER and also to maintain the same data rate. This is mainly due to the average symbol energy which is quite low compare to the DCO-OFDM.

B. *PAPR :*

PAPR is the ratio of peak signal power to average signal power. OFDM signal inherently possess high PAPR as it has large number of subcarriers overlapped to each other. OFDM signal with high PAPR requires a linear power amplifier in the front end transmitter which eventually increases the system implementation cost. Therefore, it is desirable to minimize the PAPR value for the generation of OFDM signal. Fig. 6 shows the PAPR performance for DCO- and Flip OFDM as defined by [9]:

$$PAPR_{dB} = 10 \times \log_{10}\left(\frac{\max[(x_{DCO\,or\,Flip})^2]}{E[(x_{DCO\,or\,Flip})^2]}\right) \qquad (7)$$

The PAPR performance is normally expressed in complementary cumulative distributive function (CCDF). CCDF is defined as the probability that the PAPR exceeds a certain value $PAPR_0$,

$$CCDF = 1 - P(PAPR_{DCO\,or\,Flip}(dB) \le PAPR_0) \qquad (8)$$

For this PAPR measurement, we used an OFDM signal with the IFFT size of 256 and 4-QAM data modulation for each carrier. Fig. 6 shows that Flip-OFDM undergoes PAPR penalty compare to DC biased OFDM signal. In particular, PAPR ratio is around 3.7 dB high than the bipolar OFDM. Moreover, the PAPR value is increased to around 7.2 dB and 10.2 dB for flip-OFDM in compare with 7 dB and 13 dB biased DCO-OFDM. The increase of PAPR for flip-OFDM is mainly due to low value of total average power that arises from many zero values in both positive and flipped frame. However, for DCO-OFDM the large bias voltage improves the average power as well as reduced the negative peaks and thus, 13-dB bias OFDM performs better compare to 7-dB bias for PAPR measurement as shown in fig. 6.

C. *Bandwidth vs.* ($E_{b(opt.)}/N_o$)

We now compare the DCO-OFDM and flip-OFDM in

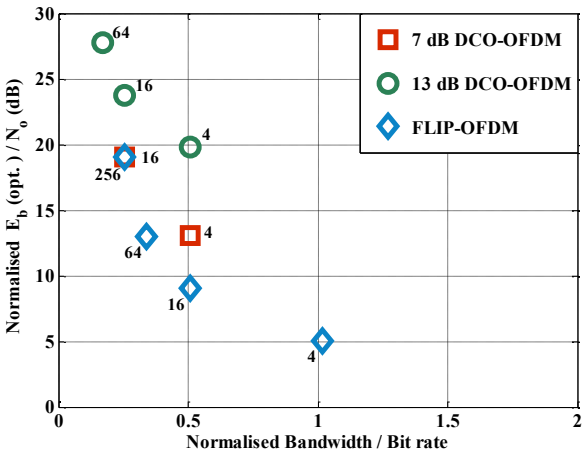


Figure 7: $E_{b(opt.)}/N_0$ for BER 10^{-3} vs Normalized bandwidth per bit rate for DCO- and Flip- OFDM

terms of transmitted optical power to noise power ($E_{b(opt.)}/N_0$) and spectral efficiency as these parameters have the importance in practical implementation. This simulation is performed by adopting unity optical power ($P_{opt.}^{DCO} = P_{opt.}^{FLIP} = 1$) for both Flip and DCO-OFDM. As a result, we can derive the relationship between electrical and optical power for this two OFDM schemes as defined by [7],

$$\frac{E_{b,opt.}^{DCO}}{N_0} = \frac{k^2}{1+k^2}\frac{E_{b,elec.}^{DCO}}{N_0} \tag{9}$$

$$\frac{E_{b,opt.}^{FLIP}}{N_0} = \frac{1}{\pi}\frac{E_{b,elec.}^{FLIP}}{N_0} \tag{10}$$

Here, constant k is obtained by equating the numbers 7 and 13 with the statement $10log_{10}(k^2+1)$ for DCO-OFDM with 7 and 13 dB bias respectively. The simulation results shown in fig.7 clearly reveals that normalised bandwidth per bit rate of 7-dB and 13-dB biased DCO-OFDM with 4-QAM data format are identical and reach to around 0.5. However, Flip-OFDM shows the value of 1.0, which indicates the necessity of double bandwidth for the same modulation format. The reason behind this large bandwidth for Flip-OFDM is the bipolar samples that are divided into positive and flipped parts with equal length each. However, in terms of optical power efficiency, flip-OFDM outperforms the DCO-OFDM as also predicted before. In particular, flip-OFDM shows the power value of 5 dB for 4-QAM data signal, whereas the power increases to 13 dB and 20 dB for the DCO-OFDM with 7-dB and 13-dB bias respectively. With the high constellation size, the performance of flip-OFDM is close to the DCO-OFDM especially with 7 dB bias condition.

IV. Conclusion

In this paper, the performance of Flip-OFDM is compared with DCO-OFDM on the basis of three parameters, BER as a function of ($E_{b(elec.)}/N_o$), PAPR and ($E_{b(opt.)}/N_o$) as a function of normalized bandwidth/bit rate. Flip-OFDM has better BER performance compare to DCO-OFDM, however, it shows high PAPR value mainly due to lower average power. Moreover, flip-OFDM exhibits considerable power gain for the same normalised bandwidth/bit rate with respect to DCO-OFDM. On the other hand, the performance of DCO-OFDM depends on the bias value, where the large bias is useful to reduce the clipping noise.

References

[1] J. Armstrong, "OFDM for Optical Communications," *J. Lightw. Tech.*, vol. 27, no. 3, pp. 189–204, Feb. 2009.

[2] W. Sheih, I. Djordjevic, "OFDM for Optical Communications", ELSEVIER, 2009.

[3] D. Barros, S. Wilson, and J. Kahn, "Comparison of orthogonal frequency-division multiplexing and pulse-amplitude modulation in in-door optical wireless links," *IEEE Trans. Commun.*,vol.60, no.1, pp. 153–163, Jan. 2012.

[4] J. Armstrong, B. J. C. Schmidt, D. Kalra, H. A. Suraweera, and A. J. Lowery, "Performance of asymmetrically clipped optical OFDM in AWGN for an intensity modulated direct detection system," *Proc. Of IEEE* Globecomm, 2006.

[5] S.D. Dissanayake, J. Armstrong, "Comparison of ACO-OFDM, DCO-OFDM and ADO-OFDM in IM/DD Systems," *J. of Lightwave Technology*, vol.31, no.7, pp.1063, 1072, April1, 2013.

[6] N. Fernando, Y. Hong, and E. Viterbo, "Flip-OFDM for unipolar communication systems," *IEEE Trans. Comm.*, vol. 60, no. 12, pp. 3726 3733, Dec. 2012.

[7] P. Saengudomlert, "On the Benefits of Pre-Equalization for ACO-OFDM and Flip-OFDM Indoor Wireless Optical Transmissions Over Dispersive Channels," *J. of Lightwave Technology*, vol.32, no.1, pp.70,80, Jan.1, 2014.

[8] Ranjha, Bilal; Kavehrad, Mohsen, "Hybrid asymmetrically clipped OFDM-based IM/DD optical wireless system," IEEE/OSA J. *Optical Communications and Networking,*, vol.6, no.4, pp.387,396, April 2014.

[9] H. Seung and J. Lee, "An overview of peak-to-average power ratio reduction techniques for multicarrier transmission," *IEEE Wirel. Commun.,* vol. 12, no. 2, pp. 56–65.

PMD Tolerance of 107 Gb/s PM-QPSK System at Optimum Transmitting Power

Kazi Abu Taher, Satya P. Majumder and B. M. Azizur Rahman
Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh (1st and 2nd author) and School of Engineering and Mathematical Sciences, City University London, UK
Email: kataher@yahoo.com, spmajumder2002@yahoo.com and B.M.A.Rahman@city.ac.uk

***Abstract*—To provide higher capacity optical fiber networks, a 107 Gb/s transmission system is simulated by using the combined features of polarization multiplexing (PM), quadrature phase shift keying (QPSK) and forward error correction (FEC). The lower tolerance to linear and nonlinear fiber impairments remains a major concern at a higher bit rate. To address the limitations set by the linear impairments, coherent detection of multilevel modulation formats with polarization division multiplexing appears as a promising solution by reducing the symbol rate to one-forth. Electronic dispersion compensation (EDC) and constant modulus algorithm (CMA) are used at the receiver for the linear dispersion compensation and forward error correction (FEC) respectively. Fiber spans each of 90 km are used in this simulation making the transmission length 1980 km. The bit rate of the modulating signals is kept at 26.75 Gb/s and the aggregate transmission rate of 107 Gb/s by utilizing the combined features of PM and QPSK. The power sharing of both the polarization components in the presence of polarization mode dispersion (PMD), fiber nonlinearities and random birefringence is a very important for the system design. In this paper, the PMD tolerance of the system at different transmitting powers with and without the presence of nonlinearities and birefringence are simulated and evaluated in terms of the bit error rate (BER). Transmitting power of 2.518 mW is found as the optimum value for single channel PM-QPSK systems at PMD 0.05 ps/sqrt(km).**

***Keywords*—Polarization multiplexing, quadrature phase shift keying, coherent detection, electronic dispersion compensation.**

I. Introduction

Transmission in optical fiber communication systems is impaired and ultimately limited by chromatic dispersion, amplified spontaneous emission noise from amplifiers, polarization effects and fiber nonlinearities [1]. The transmitting power level has a direct implication in setting the operating parameters for high bit transmission. At higher power level, the nonlinear parameters play dominant role [2]. For conventional coherent direct-detection single-carrier systems, the impairment induced by a constant differential group delay (DGD) scales with the square of the bit rate, resulting in drastic PMD induced degradation for high speed transmission systems [3,4]. Complementary features of PM and QPSK provide a way out of achieving the bit rate. At this high bit rate, the PMD effects are difficult to analyze because they are stochastic in nature. Hence modeling the effects of PMD is not straight forward. Random change of the fiber birefringence affects the correlation length and the beat length of the fiber [5,6]. The power sharing between the orthogonal polarization components of the optical field along the slow and fast axis becomes important. On the other hand Karr nonlinearity is the main source of nonlinearity is deterministic and can be modeled to analyze its effects on the data transmission. Karr nonlinearity leads to phase rotation that is proportional to the intensity at every point in time [7,8]. When the nonlinearity couples with other transmission impairments such as PMD, it can lead to complex dynamics. Thus PMD which is stochastic in nature and nonlinearity which is deterministic in occurrence should be addressed together for modelling high speed data transmission systems [9,10]. Nonlinear birefringence is another source of impairment in fiber communication [11].

If $A_x(z,t)$ and $A_y(z,t)$ as the complex envelops of the x and y components of the electromagnetic field of the propagating mode, then [12],

$$E_x(z,t) = A_x(z,t)\exp\left[j(k_0 z - \omega_0 t)\right] \quad \text{..... (1a)}$$

$$E_y(z,t) = A_y(z,t)\exp\left[j(k_0 z - \omega_0 t)\right] \quad \text{.....(1b)}$$

In the presence of chromatic dispersion and Kerr third-order nonlinearities, the propagation of light in a single linearly birefringent section is given by the coupled nonlinear Schodinger equation [12]:

$$-j\frac{\partial A_x}{\partial z} = (\beta_x(\omega_0) - \beta_0)A_x + j\left.\frac{d\beta_x}{d\omega}\right|_{\omega_0}\cdot\frac{\partial A_x}{\partial t} - \frac{1}{2}\left.\frac{d^2\beta_x}{d\omega^2}\right|_{\omega_0}\cdot\frac{\partial^2 A_x}{\partial t^2} + j\frac{\alpha}{2}A_x + \gamma\left[|A_x|^2 + \frac{2}{3}\left|A_y\right|^2\right]A_x + \frac{\gamma A_y^2 A_x^*}{3} \quad \text{..... (2a)}$$

$$-j\frac{\partial A_y}{\partial z} = (\beta_y(\omega_0) - \beta_0)A_y + j\left.\frac{d\beta_y}{d\omega}\right|_{\omega_0}\cdot\frac{\partial A_y}{\partial t} - \frac{1}{2}\left.\frac{d^2\beta_y}{d\omega^2}\right|_{\omega_0}\cdot\frac{\partial^2 A_y}{\partial t^2} + j\frac{\alpha}{2}A_y + \gamma\left[\left|A_y\right|^2 + \frac{2}{3}|A_x|^2\right]A_y + \frac{\gamma A_x^2 A_y^*}{3} \quad \text{..... (2b)}$$

where the first terms on the right hand side expresses the phase velocity as well as the effect of the birefringence, the second term the group velocity, third term the dispersion, fourth term the loss, fifth and sixth terms the nonlinearity. Each of these terms has its own length scale. The birefringence, represented by the first term will rotate the state of polarization of the propagating light, completing a full circle around the s_1 axis of the Poincare sphere every beat-length. When power and intensity dependent nonlinear effects in optical fibers become dominant, the optical field can induce nonlinear birefringence.

Digital signal processing along with phase and polarization diverse coherent receiver proved to be a promising technology for high data transmission networks. EDC could compensate 107,424 ps/nm of chromatic dispersion over 6400 km of standard single mode fiber [13]. Enhanced FEC (EFEC) system has excellent error correction capability and can correct an error up to approximately an error rate of 10^{-3} by performing Concatenated Coding [14]. Though PM-QPSK system offers a solution for achieving higher transmission rate, the design of the system needs careful selection of the operating parameters. Increase in transmitting power causes an increase in fiber nonlinearity putting a limiting concern. Hence this paper reports the optimum transmitting power of the system and its PMD tolerance.as a noble techniwu

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

II. SYSTEM SETUP

Designing WDM channels at 107 Gbit/s rate is more challenging than 26.75 Gbit/s. But the features of PM and QPSK are combined to achieve the same high transmission rate through a network which is designed to handle only one fourth of the transmission rate. In QPSK system, two data streams are taken as input to modulate the phase of the carrier. The phase of each of the bisected part of coherent source light is similar and taken as the reference phase. One part of the bisected light is used as the carrier for one data stream and the data stream amplitude modulates the source light. Another data stream is also used to amplitude modulate (AM) the other part of the bisected source light. The phase of the amplitude modulated optical signal that corresponds to second data stream is further phase modulated by a constant bias voltage. Then the AM optical signal that carries first data stream and the amplitude as well as phase modulated optical signal that carries second data stream are combined by a beam combiner. Figure 1(a) depicts the QPSK modulator module. This modulator takes the source light from a coherent source which is bisected using a polarization maintaining splitter. Upper amplitude modulator implements AM to the coherent optical signal using Data 1. Lower amplitude modulator implements AM to the optical signal using Data 2. Output of the lower modulator is then sent to the phase modulator. The phase shift bias voltage generates 180^0 phase shift on the optical signal. The output of the Phase Modulator and the output of the upper Amplitude Modulator are combined using the Beam Combiner.

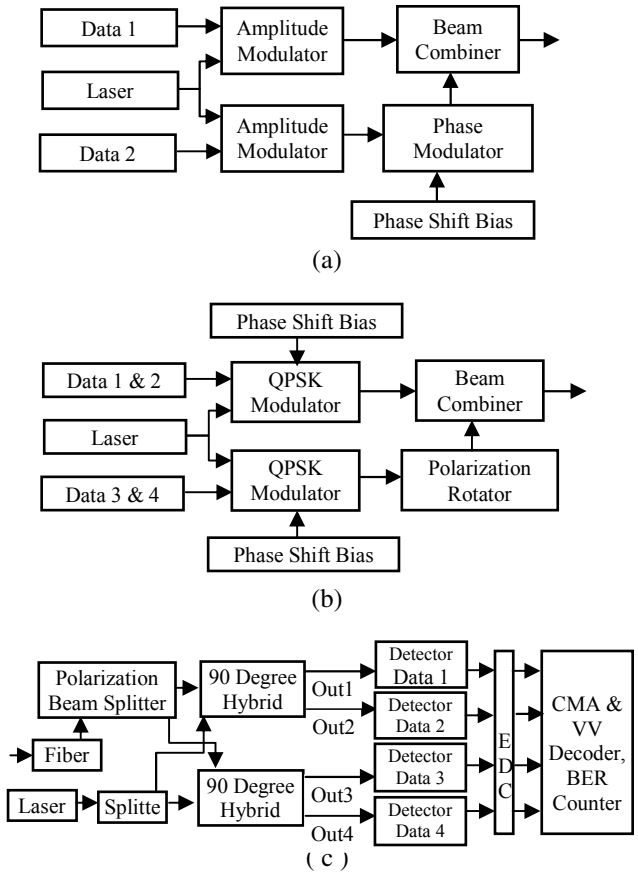


Figure 1. (a) the QPSK modulator section, (b) the polarization multiplexing section and (c) the receiver section.

Fig. 1(b) shows the polarization multiplexing that is consist of two QPSK modulator modules, one polarization rotator and one beam combiner. Both the QPSK Modulators take the coherent light from the same source. Two QPSK modulator modules take four data streams. Output of the lower QPSK Modulator that corresponds to Data 3 and 4 are passed through the polarization rotator. Polarization rotator rotates the angle of the optical field by 180^0 that relates to the S_3 of Poincare Sphere. The beam combiner is used to combine the output of the first QPSK modulator and the output of the polarization rotator. Thus by using the features of PM, two QPSK data are accommodated within the band limit. Finally four data streams are sent within the bandwidth of one data stream. But the maintenance of optimum power level in two polarization modes is essential to keep the error of the transmission with in tolerance.

The receiver section as shown in Figure 1(c), consists of the 90 degree hybrid, detector, EDC) and error correction module. A laser source is used at the receiver section to supply the reference light to two 90 degree hybrids through a splitter. The outputs 1 and 2 from the upper hybrid correspond to the modulated signal of data 1 and 2. Similarly the output 3 and 4 of the lower hybrid corresponds to the modulated signals of data 3 and 4. Four detectors are used to detect the data 1-4. Then the detected data are passed to EDC. Finally constant modulus algorithm is used for implementing the forward error correction. Viterbi and Viterbi algorithm is used for decoding the data. For encoding and decoding data, binary polynomials of degree 11 are used. Polynomials 1-4 are used for input data 1-4 respectively.

To simulate the PM-QPSK system, necessary noise loading is implemented and channel losses are considered. Standard loss of 0.22 dB/km is included in the in the simulation. Uncompensated fiber of dispersion 3.8 ps/nm/km at reference frequency of 193 THz is considered. Fiber spans each of 90 km followed by an amplifier with amplification enough to cover up the loss suffered in the fiber are used. Total 22 spans are implemented resulting the 1980 km of fiber.

III. SIMULATION AND DISCUSSION

The simulation is carried out with the help of OptSim and MATLAB software where total 65536 bits are used. Since the bit error rate (BER) is based on the actual counting of the errors, BER with the probability less than 10^{-6} is shown as zero. But with forward error correction tools as used in this simulation, any error with BER less than 10^{-3} can be automatically corrected [13,14]. 65536 bits are found reasonably sufficient to give reliable results. Parameters like transmitting power, PMD, nonlinearities, birefringence and fiber length are taken into consideration in this simulation. Output of the booster at transmitter end just before the fiber is taken as the transmitting power. The booster amplification is varied from -4 to 5 dBm which corresponds to 0.632556 to 5.02457 mW. The PMD coefficient is varied from 0.02 to 0.2 ps/sqrt(km) to study the change of PMD. To simulate the effect of the variations of nonlinearities, the fiber nonlinear coefficient is varied from 0.05 to 2.5 1/Watt/km. The fiber birefringence is varied statistically following the Maxwell probability distribution. The fiber length is kept 1980 km to study the effect of power, nonlinearity and birefringence. The analytical simulation of Eq 1 is carried out in six steps. Firstly, impact of PMD alone on BER is simulated. The impact of changing PMD in the absence of nonlinearities and birefringence on the BER of the system is depicted in Figure 2. The level is varied within 0.5-5.5 mW and the simulated PMD values are 0.02, 0.1 and 0.2 ps/sqrt(km). The results show that the BER is well below 10^{-3}. Secondly, impact of

PMD and birefringence are simulated as shown in Figure 3. The birefringence are varied following the Maxwell distribution. This addition of birefringence caused the BER to increase. But the BER is well below 10^{-3} for almost all transmitter powers.

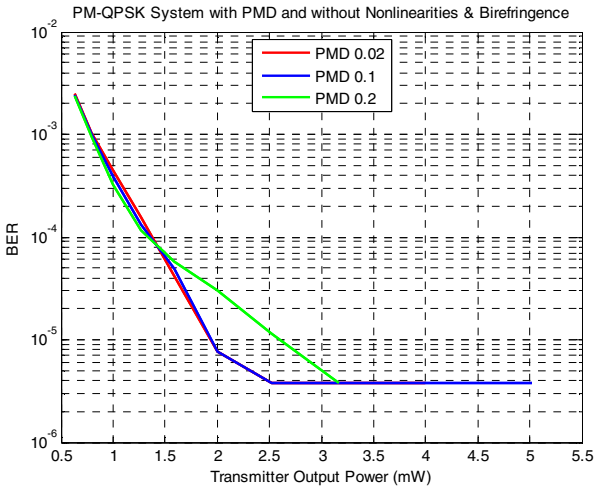


Fig. 2. BER-Power curve of PM-QPSK system with PMD and without nonlinearities & birefringence.

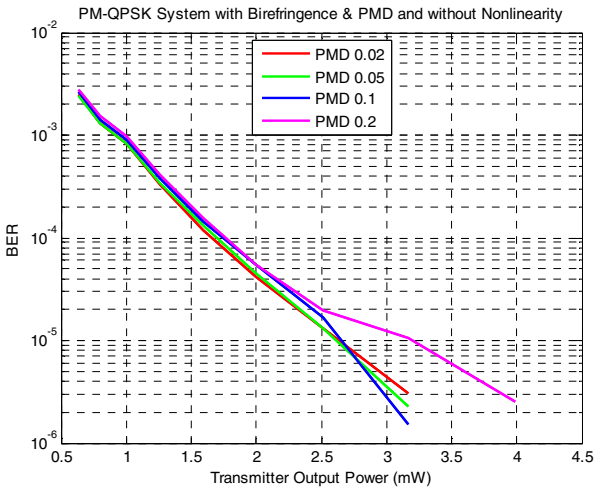


Figure 3. BER-Power curve of PM-QPSK system with PMD & birefringence and without nonlinearities.

In the third step, the simulation is carried out considering then impact of nonlinearities only, in the absence of PMD and birefringence as shown in Figure 4. The BER level goes down at power level around 1 mW. But soon with the increase of the transmitter power, the BER increases almost monotonically. The power level is varied from 0.5-5 mW for nonlinearities 0.5-2.5 1/W/km with interval of 0.5 1/W/km. When the power level is increased with low nonlinearity, the BER remains within the acceptable limit. For nonlinearity around 0.5 1/W/km, the BER remains low. But when the nonlinearity is increased more than 1 1/W/km, the BER rises sharply with the increased power and the BER crosses the acceptable limit. It is also noticeable that for the standard value of nonlinearity of 1.5 1/W/km, the BER goes beyond acceptable level. In the fourth step, PMD and nonlinearities are taken into account. Here the nonlinearities are kept fixed at the standard value of 1.5 1/W/km and the PMD values are changed Figure 5 shows the BER-power curve with nonlinearities and PMD in to consideration but the effect of the random birefringence is omitted. As seen from the Figure 5, with the increase in the transmitting power, the BER increases having the similar pattern caused by the nonlinear effects alone. But the combined effect of nonlinearity and PMD has changed the curve pattern significantly. The effect of the higher

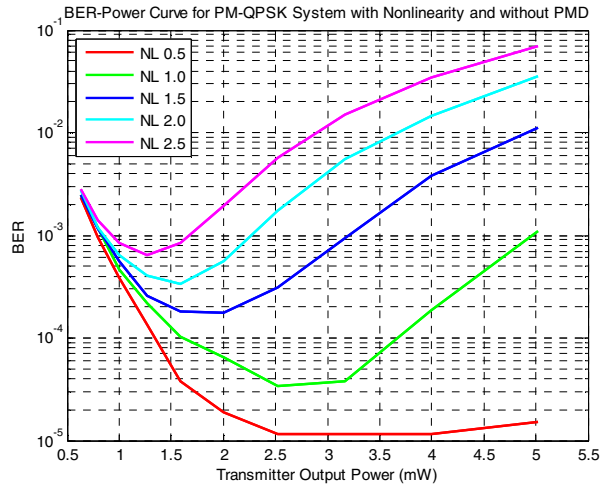


Figure 4. BER-Power curve of PM-QPSK system with nonlinearities and without PMD & birefringence.

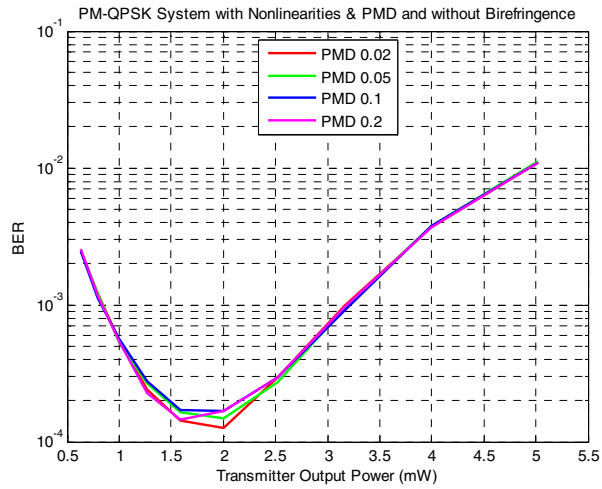


Figure 5. BER-Power curve of PM-QPSK system with nonlinearities & PMD and without birefringence.

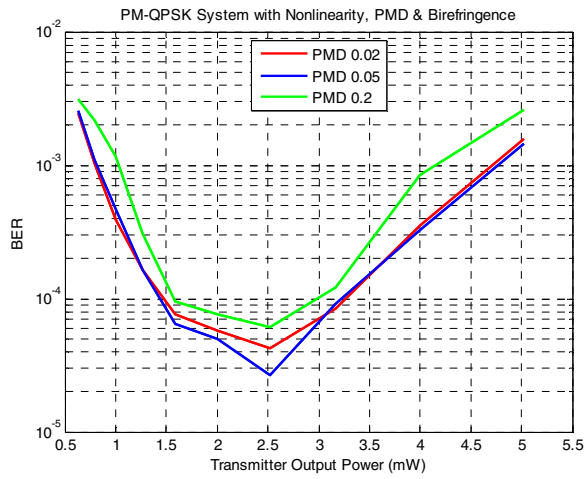


Figure 6. BER-Power curve of PM-QPSK system with nonlinearities, PMD and birefringence.

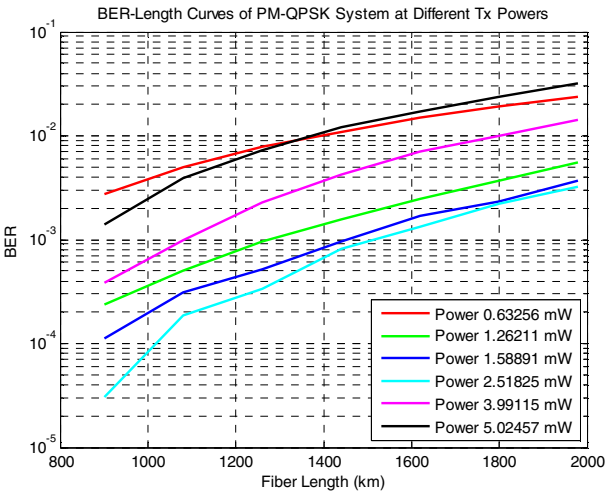


Figure 7. BER-Length curve of PM-QPSK system with nonlinearities, PMD and birefringence for different transmitter power levels.

nonlinearity is reduced in terms of BER and the effect of low nonlinearity is further impaired by increasing the BER. In general the curve pattern follows the effect of the nonlinearities at power level more than 2.5 mW. The reason of such change is due to the fact that in the absence of random birefringence, the impairment caused by the PMD is reduced considerably. In the absence of random birefringence, the nonlinear cross polarization between the polarization components is also reduced.

In the fifth step, the combined effect of PMD, nonlinearity and birefringence are simulated as shown in Figure 6. It shows the BER-power curves describing their combined effect. The nonlinearity coefficient is kept fixed at 1.5 1/W/km. Comparing the Figures 2-6, it is clear that the impact of PMD and birefringence is not detrimental with the increase of power. But nonlinearity alone pushes the BER towards the unacceptable level as the power level increases beyond 1.5 mW. At lower PMD value, the nonlinearities will have dominant effect and the BER increases. But the PMD together with birefringence has a higher tolerance to higher power level and reduces the detrimental effect of nonlinearities and pushes the optimum power level to the higher value. The characteristic curves show that the system gives the best performance for PMD = 0.05 ps/sqrt(km) and nonlinearity = 1.5 1/W/km at transmitting power level = 2.518 mW.

Finally in the sixth step, the simulation is carried out varying the length of the fiber and keeping the PMD set at 0.5 ps/sqrt(km) and the nonlinearity value at 1.5 1/W/km and the birefringence at Maxwell distribution. The simulated transmitted powers are varied 0.63256 - 5.02457 mW as shown in Figure 7. The system has the best performance at transmitter power level of 2.518 mW having the BER within 10^{-3} up to a distance of 1550 km. The results show that if the power level is either increased or decreased from the optimum value of 2.518 mW, the BER increases. Transmitter power levels 0.63256 and 5.02457 show the worst performance and the BER level is unacceptable. Power levels 3.99115, 1.26211 and 1.58891 have the acceptable BER limits of 10^{-3} for fiber lengths 1090, 1275 and 1450 km respectively.

IV. Conclusions

The PM-QPSK system gives an important and easily implementable technique for achieving higher bit rates at 107 Gb/s including 7% overhead for EFEC. The combined features of polarization multiplexing and QPSK in conjunction with EDC and FEC algorithm made it an alternative for future high speed communication. The combined features made it possible to quadruple the bit rate. The overall penalty of PMD-nonlinearity impaired system is lower than the sum of the individual penalties. Impact of increasing transmitter powers on BER in the presence of random birefringence, PMD and nonlinearity is a proven tool. The results show that the PM-QPSK system has the best performance at transmitting power level of 2.518 mW at PMD of 0.05 ps/sqrt(km) and nonlinearity of 1.5 1/W/km.

Acknowledgment

This paper is prepared as part of the PhD research work under the Strongties project of Erasmus Mundus Scholarship program in collaboration between City University London, United Kingdom and Bangladesh University Engineering and Technology, Dhaka, Bangladesh.

References

[1] C. R. Menyuk and B. S. Marks, “Interaction of Polarization Mode Dispersion and Nonlinearity in Optical Fiber Transmission Systems”, Journal of Lightwave Technology, vol. 24, pp. 2806-2826 Jul 2006.

[2] G. P. Agrawal, “Nonlinear Fiber Optics” 5th edn, London, Academic Press, 2013.

[3] I. P. Kaminow, T. Li and A. E. Willner, “Optical Fiber Telecommunications”, ELSEVIER Press, Feb 2008.

[4] L. Huiyang, Z. Xiaoguang and X. Wei, “Research on Polarization Mode Dispersion Compensation Performance of Optical DQPSK Modulation Format”, Proc. of The International Society for Optical Engineering Vol. 7136 713635 pp.1-8, 2008.

[5] H. H. Yaffe, and D. L. Peterson, “Experimental Determination of System Outage Probability Due to First-Order and Second-Order PMD”, Journal of Lightwave Techno., vol. 24, pp. 4155-61, Nov 2006.

[6] H. Kogelnik, L. E. Nelson, and J. P. Gordon,“ Emulation and Inversion of Polarization-Mode Dispersion”, Journal of Lightwave Technology, vol. 21, pp. 482-495, Feb 2003.

[7] M. Karlsson and H. Sunnerud, “Effects of Nonlinearities on PMD-Induced System Impairments”, Journal of Lightwave Technology, vol. 24, pp. 4127-4137, Nov 2006.

[8] N. J. Frigo, F. Bucholtz, and C. V. McLaughlin, “Polarization in Phase Modulated Optical Links: Jones- and Generalized Stokes-Space Analysis”, Journal of Lightwave Technology, vol. 31, pp. 1503-1511, May 2013.

[9] C. Laperle, B. Villeneuve, Z. Zhang and M. O’Sullivan, “WDM Performance and PMD Tolerance of a Coherent 40-Gbit/s Dual-Polarization QPSK Transceiver”, Journal of Lightwave Technology, vol. 26, pp. 168-175, Jan 2008.

[10] J. N. Damask, “Methods to Construct Programmable PMD Sources—Part I: Technology and Theory”, Journal of Lightwave Technology, vol. 22, pp. 997-1005, Apr 2004.

[11] J. Renaudier, G. Charlet and S. Bigo, “Linear Fiber Impairments Mitigation of 40-Gbit/s Polarization-Multiplexed QPSK by Digital Processing in a Coherent Receiver”, Journal of Lightwave Technology, vol. 26, pp. 36-42 , Jan 2008.

[12] M. Brodsky, N. J. Frigo and M. Tur, “Optical Fiber Telecommunications V A: Components and Subsystems,” 5th edn, Elsevier Inc, 2008.

[13] S. J. Savory, G. Gavioli and K. P. Bayvel, “Electronic Compensation of Chromatic Dispersion using a Digital Coherent Receiver”, Optical Society of America, vol. 15, pp. 2120-2126, March 2007.

[14] M. Kamakura, K. Takatori and H. Narita, “Optical Communication Card and Optical Transmission Device”, Patent No. US8,626,001 B2, 7 Jan 2014.

Unconstraint Adaptive Frequency-Domain Equalizer for Coherent Optical Receivers

Md. Saifuddin Faruk

Department of Electrical & Electronic Engineering, Dhaka University of Engineering & Technology, Gazipur, Bangladesh

Abstract—**In this paper, unconstraint adaptive frequency-domain equalizer (FDE) based on the constant-modulus algorithm (CMA) is investigated for equalization of linear transmission impairments in coherent optical receivers. With unrepeated 200-km transmission of 10-Gbaud quadrature phase-shift keying (QPSK) signal, it is verified that, compared with conventional constraint FDE, the proposed unconstraint FDE causes a very small sensitivity penalty; however, can be applied with much less computational complexity.**

***Index Terms*—optical fiber communication, digital signal processing, equalization, adaptive filter.**

I. Introduction

Recent advancement of digital coherent receivers enable compensation of fiber transmission impairments such as chromatic dispersion (CD) and polarisation-mode dispersion (PMD) in digital domain [1]. Generally we use adaptive finite-impulse-response (FIR) filters which tap coefficients are updated symbol-by-symbol manner using any gradient-decent based algorithm such as constant-modulus algorithm (CMA) or decision-directed least-mean-square (LMS) algorithm. However, the computational complexity of such approach increases rapidly with increase of number of filter taps [2]. Frequency-domain approach can reduce this computation cost significantly by using block-by-block processing and fast implementation of discrete Fourier transform (DFT).

The conventional adaptive frequency-domain equalizer (FDE) such as fast LMS [3] can operate with input signal sampled at symbol rate. However, the receiver with two-fold oversampled input sequences is preferred since such an oversampling rate reduces the aliasing effect significantly. To enable adaptive FDE based on two-fold oversampled input signal, previously we propose the concept of splitting the equalizer in even/odd sub-equalizer [4,5].

In this paper, we further modified the equalizer with even/odd sub-equalizer concept by eliminating the gradient constraint used in the configuration. Such a modification yields a more computational efficient structure. The principle of operation of the proposed scheme is verified with 10-Gbaud quadrature phase-shift keying (QPSK) 200-km unrepeated transmission experiments. It is found that, the bit-error-rate (BER) performance of the proposed unconstraint FDE is very close to that obtained by the constraint FDE; however, the proposed FDE can be implemented with reduced computational complexity.

Throughout the remainder, time- and frequency-domain variables are denoted by lower- and upper-case characters, respectively, while boldface characters denote vectors. Furthermore, the symbol $\otimes$ is element-by-element multiplication; $\mathbf{O}_L$ is a column vector with L zeros; $\mathbf{I}_L$ is a column vector with L ones; superscripts $(\bullet)^e$ and $(\bullet)^o$ correspond to even and odd sub-equalizer parameters, respectively; and conj(•) is the conjugate operator.

II. Adaptive FDE with Even/Odd Sub-Equalizer Concept

The adaptive FDE with even/odd sub-equalizer concept that enable processing of two-fold oversampled input sequence is shown in Fig. 1 [4]. For simplicity of explanation, we consider the configuration for single-polarization transmission; however, the configuration can be easily modified for polarization-multiplexed system by using two-by-two butterfly configuration of the equalizer [5]. The equalizer is composed of two sub-equalizers. Each sub-equalizer processes input sequences at the symbol rate, which are taken from even or odd sequences of data sampled at the rate of two samples/symbol. The overlap-save method is employed to perform linear convolution effectively [6].

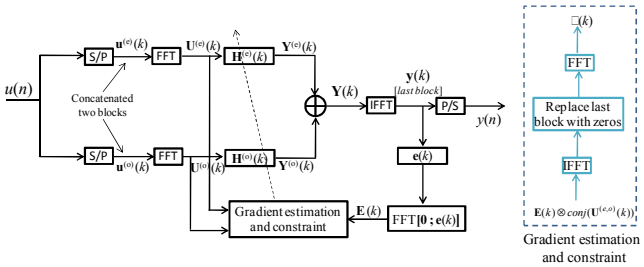


Fig.1: Block diagram of adaptive FDE with even/odd sub-equalizer concept. Inset is the detail of gradient estimation and constraint operation.

Considering that a block length equals the sub-equalizer filter length L, in the case of 50% overlap (hence, FFT length N=2L), the k-th block of the frequncy-domain filter output can be written as

$$\mathbf{Y}(k) = \mathbf{U}^{(e)}(k) \otimes \mathbf{H}^{(e)}(k) + \mathbf{U}^{(o)}(k) \otimes \mathbf{H}^{(o)}(k), \tag{1}$$

where $\mathbf{U}^{(e,o)}(k)$ and $\mathbf{H}^{(e,o)}(k)$ are two column vectors of the length N having elements of the frequency-domain input sequences and filter-tap coefficients, respectively.

Then the time-domain output is computed using overlap-save method as

$$\mathbf{y}(k) = \text{last } L \text{ data of } IFFT\{\mathbf{Y}(k)\}. \tag{2}$$

The error vector in CMA is calculated in the time domain as

$$\mathbf{e}(k) = [\mathbf{I}_L - \mathbf{y}(k) \otimes \text{conj}\{\mathbf{y}(k)\}] \otimes \mathbf{y}(k), \tag{3}$$

which is transformed into the frequency domain as

$$\mathbf{E}(k) = FFT[\mathbf{O}_L ; \mathbf{e}(k)]. \tag{4}$$

The time-domain gradient vectors are then computed as

$$\boldsymbol{\varphi}^{(e,o)}(k) = \text{first } L \text{ data of } IFFT\{\mathbf{E}(k) \otimes conj(\mathbf{U}^{(e,o)}(k))\}. \tag{5}$$

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

Finally, the updating process of frequency-domain filter-tap coefficients based on gradient decent algorithm is expressed by

$$\mathbf{H}^{(e,o)}(k+1) = \mathbf{H}^{(e,o)}(k) + \mu FFT[\boldsymbol{\varphi}^{(e,o)}(k); \mathbf{O}_L], \qquad (6)$$

where μ is a step-size parameter.

III. Proposed Unconstraint Adaptive FDE

For the equalizer configuration shown in the Fig.1, a time-domain constraint is imposed for gradient vector estimation by discarding last L elements of IFFT of $\mathbf{E}(k) \otimes conj(\mathbf{U}^{(e,o)}(k))$ as shown in Eq. (5) and replacing the elements so discarded by a block of L zeros before reapplying the FFT as shown in Eq. (6). However, according to [7], under fairly mild conditions, the FDE algorithm can still work well even this gradient constraint is dropped. Based on it, we propose a unconstraint adaptive FDE having even/odd sub-equalizer concept as shown in Fig.2 which can be consider a modified version of FDE shown in Fig.1. The filtering and filter adaption process for the proposed FDE can be done using the following set of equations:

$$\mathbf{Y}(k) = \mathbf{U}^{(e)}(k) \otimes \mathbf{H}^{(e)}(k) + \mathbf{U}^{(o)}(k) \otimes \mathbf{H}^{(o)}(k)$$

$$\mathbf{y}(k) = \text{last } L \text{ data of } IFFT\{\mathbf{Y}(k)\}$$

$$\mathbf{e}(k) = [\mathbf{I}_L - \mathbf{y}(k) \otimes \text{conj}\{\mathbf{y}(k)\}] \otimes \mathbf{y}(k),$$

$$\mathbf{E}(k) = FFT[\mathbf{O}_L; \mathbf{e}(k)]$$

$$\boldsymbol{\varphi}^{(e,o)}(k) = \mathbf{E}(k) \otimes conj(\mathbf{U}^{(e,o)}(k))$$

$$\mathbf{H}^{(e,o)}(k+1) = \mathbf{H}^{(e,o)}(k) + \mu FFT[\boldsymbol{\varphi}^{(e,o)}(k); \mathbf{O}_L]$$

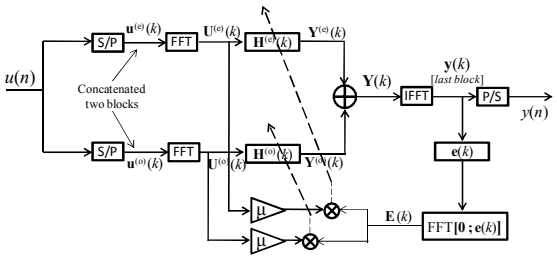


Fig.2: Proposed unconstraint adaptive FDE with even/odd sub-equalizer concept.

It is important to note that, elimination of gradient constraint ensure reduction of 2 FFTs and 2 IFFTs operation. Considering the classic radix-2 algorithm, which requires $N\log_2(N)/2$ multiplications to execute FFT of N complex numbers [2], we can achieve a reduction of computational complexity of $4L\log_2(2L)$ complex multiplications in the proposed unconstraint FDE compare to that in constraint FDE as shown in Fig.1.

IV. Experimental Results and Discussion

In order to verify the effectiveness of the proposed algorithm, a 10-Gbaud QPSK unrepeated transmission experiment is conducted as show in Fig. 3.

In the transmitter side, a continuous wave was generated from a distributed-feedback laser diode (DFB-LD). The wavelength of the DFB-LD was 1552nm and its 3-dB linewidth was about 150 kHz. An NRZ 10-Gbaud QPSK signal was generated by a $LiNbO_3$ optical IQ modulator (IQM). The IQM was driven by two streams of electrical signals generated from two independent ports of an arbitrary waveform generator (AWG) with a pseudo-random binary sequences (PRBS) length of 2^9-1. The optical signal was then transmitted through a 200-km standard single-mode fiber (SMF). The accumulated CD of such system was about 3200 ps/nm.

In front of the receiver, the state of polarization of the incoming signal was controlled manually with a polarization controller (PC) so that it align with that of local oscillator (LO). Then, the received average power P_{in} was controlled by a variable optical attenuator (VOA). After that, the signal was pre-amplified by an Erbium-doped fiber amplifier (EDFA), and detected by a phase-diversity homodyne receiver, where another DFB-LD having the same characteristics was used as a LO. The frequency offset between transmitting laser and the LO was set below 10 MHz. The output from the receiver was asynchronously sampled at 20 Gsample/s by a 2-channel analog-to-digital converter (ADC) with 8-bit resolution. The digitized signals were stored in a personal computer (PC) for offline digital signal-processing (DSP).

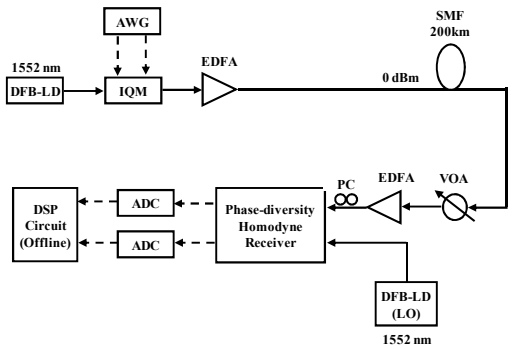


Fig.3 : Experimental setup to verify the effectiveness of the proposed adaptive FDE

In the DSP, clock recovery [8] and equalization is done by either constraint adaptive FDE or proposed unconstraint adaptive FDE. The value of step-size parameter is optimized manually so that the BER is minimum. Then carrier-phase is recovered using Viterbi-Viterbi algorithm. Finally, the symbols are decoded and BER is measured using direct counting method.

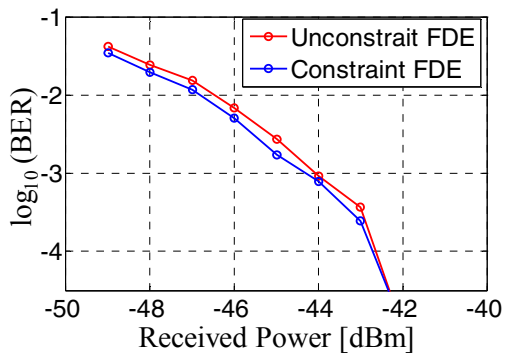


Fig. 4: The BER performance of constraint and unconstraint adaptive FDE

The BER characteristic is shown in Fig. 4. It is found that the constraint FDE has slightly better performance compare to unconstraint FDE. For a BER of 10^{-3} the sensitivity penalty is only 0.3 dB for unconstraint FDE. Such small sensitivity penalty is acceptable since we can implement it with a lower computational complexity.

V. Conclusions

Unconstraint adaptive FDE is successfully introduced in digital coherent optical receivers. Such an equalization approach reduces the computational complexity significantly while keeping the system performance almost same as conventional constraint FDE. The effectiveness of the proposed FDE is confirmed with 10-Gbaud QPSK transmission experiments.

Acknowledgment

The author would like to thank Prof. K. Kikuchi of the University of Tokyo for his valuable discussion.

References

[1] S. J. Savory, "Digital filters for coherent optical receivers," *Opt. Express*, vol. 16, pp. 804-817, 2008.

[2] B. Spinnler, "Equalizer design and complexity for digital coherent receivers," *IEEE J. Sel. Topics Quantum Electron.*, vol. 16, pp. 1180 - 1192, 2010.

[3] J. J. Shynk, "Frequency-domain and multirate adaptive filtering," *IEEE Signal Process. Mag.*, vol. 9, pp. 14-37, 1992.

[4] M. S. Faruk and K. Kikuchi, "Frequency-Domain Adaptive Equalization in Digital Coherent Receivers," in *Proc. CLEO'11*, 2011, paper CThH7.

[5] M. S. Faruk and K. Kikuchi, "Adaptive frequency-domain equalization in digital coherent optical receivers," *opt. express*, vol. 19, no. 13, pp. 12789-12798.

[6] S. Haykin, Adaptive Filter Theory,3rd ed., Prentice Hall, 2001.

[7] D. Mansour and A. H. Garay, "Unconstraint frequency-domain adaptive filter," *IEEE Trans. Acoust. , Speech and Signal Process.*, vol. ASSp-39, no. 5, pp. 583-594, 1982.

[8] K. Kikuchi, "Clock recovering characteristics of adaptive finite-impulse-response filters in digital coherent optical receivers," *Opt. express*, vol. 19, pp. 5611-5619, 2011.

An Enhanced Design of PAPR Reduction Technique

Md. Munjure Mowla[1*], Zaki Muhammad Shahriar[2], and S.M. Mahmud Hasan[3]

[1,2,3] Department of Electronics & Telecommunication Engineering
Rajshahi University of Engineering & Technology
Kazla-6204, Rajshahi, Bangladesh.
[*]rimonece@gmail.com

Abstract - **Next generation wireless communication networks (NGN) uses one of the emerging multicarrier multiplexing transmission techniques which is known as orthogonal frequency division multiplexing (OFDM). It provides several advantages such as offering greater immunity to multipath fading & impulse noise, eliminating inter-symbol interference (ISI), inter-carrier interference (ICI) & the need for equalizers. However, OFDM have a generic problem of high peak to average power ratio (PAPR) which is defined as the ratio of the peak power to the average power of the OFDM signal. In this paper, an improved design of amplitude clipping & filtering method is proposed and implemented which shows the significant improvement in case of PAPR reduction compare to an existing method.**

Index Terms - **Bit Error Rate (BER), Complementary Cumulative Distribution Function (CCDF), Peak to Average Power Ratio (PAPR).**

I. Introduction

In this modern era, the explosive growth of new multimedia wireless applications have resulted in an massive demand for technologies that support very high speed transmission rates, mobility and efficiently utilize the available spectrum and network resources. OFDM is one of the best solutions to achieve this goal and it offers a promising choice for future high speed data rate systems. OFDM has been standardized as part of the IEEE802.11a and IEEE 802.11g for high bit rate data transmission over wireless LANs. One of the major drawbacks of OFDM signal is the high peak-to-average power ratio (PAPR) of the transmit signal. As the data rates and mobility supported by the OFDM system raise, the number of subcarriers also raise, which in turn leads to high PAPR. As future OFDM-based systems may push the number of subcarriers up to meet the higher data rates and mobility demands, there is a need to mitigate the high PAPR [1]. If the peak transmit power is limited by either regulatory or application constraints, the effect is to reduce the average power allowed under multicarrier transmission relative to that under constant power modulation techniques. This in turn reduces the range of multicarrier transmission. Moreover, to prevent spectral growth of the multicarrier signal in the form of intermodulation among subcarriers and out-of-band radiation, the transmit power amplifier must be operated in its linear region (i.e., with a large input back-off), where the power conversion is inefficient. This may have a deleterious effect on battery lifetime in mobile applications. As portable devices have a finite battery life it is important to find ways of reducing the PAPR allowing for a smaller more efficient HPA, which in turn will mean a longer lasting battery life. In many low-cost applications, the drawback of high PAPR may outweigh all the potential benefits of multicarrier transmission systems [2]. Several attractive research works have been carried out to reduce PAPR with the expense of increase transmit signal power, bit error rate (BER), computational complexity and data rate loss etc. So, a system trade-off is required. The clipping & filtering method was first introduced by X. Li and L.J. Cimini, showed the effects of clipping and filtering on the performance of OFDM [3]. A repeated operation of clipping and filtering (RCF) technique was then implemented by J.Armstrong resulted that the PAPR of an OFDM signal could be reduced without any increase in the out-of-band power by clipping the oversampled time domain signal followed by filtering using an FFT-based, frequency domain filter designed to reject out-of-band discrete frequency components [4]. A novel iterative clipping & filtering (ICF) at receiver was proposed by H.Chen and A.Haimovich to estimate and cancel the distortion caused by clipping noise [5]. A This SCF shows that actually in one iteration, obtained the same PAPR reduction as that of the previous ICF technique with several iterations. The computational complexity was therefore significantly reduced. Simulations show that, in relatively large clipping cases, SCF technique had better PAPR reduction and out-of-band radiation than and very close BER performance to ICF. Recently, a FIR based bandpass filter achieves better performance used by Yong Soo Cho et. al.[6] . It introduces less noise into the signal, causes less peak regrowth and needs less computation.

In this simulation, a special type of IIR based peak resonator filter is proposed which reduces the PAPR significantly compare to existing method [6] but at the slight increase of BER.

II. OFDM System Model

OFDM is a special form of multicarrier modulation (MCM) with densely spaced subcarriers with overlapping spectra, thus allowing multiple-access. MCM works on the criteria of transmitting data by dividing the stream into several bit streams, each of which has a much lower bit rate and by using these sub-streams to modulate several carriers. In multicarrier transmission, bandwidth divided in many non-overlapping subcarriers but not necessary that all subcarriers are orthogonal to each other shown in figure 1, where Δf shows the frequency spacing between two sub-carriers. In OFDM the sub-channels overlap each other to a certain extent. The information sequence is mapped into symbols, which are distributed and sent over the N sub-channels, one symbol per channel. To permit dense packing and still ensure that a minimum interference between the sub-channels is encountered, the carrier frequencies must be chosen carefully according to their orthogonal properties. By using orthogonal carriers, frequency domain can be viewed so as the frequency

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

space between two sub-carriers is given by the distance to their first spectral null [2].

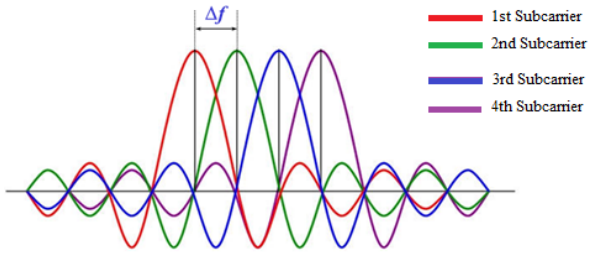


Fig. 1 An OFDM Signal with 4 sub-carriers

III. PROPOSED METHOD

As the main focus of this research work is to reduce PAPR, so, in this simulation, we have done trade-off between PAPR reduction with BER increment. *Yong Soo et.al* [6] show that if clipped signal passes through a composed filter (FIR based BPF) before passing a low pass filter (LPF) to reduce out-of-band radiation, then it causes less BER degradation with medium amount of PAPR reduction. Considering this concept, we have designed a scheme for clipping & filtering method where clipped signal passes through a composed filter (IIR based resonator filter) before passing a LPF, and then it causes a little bit more BER degradation but significant amount of PAPR reduction than existing method [6].

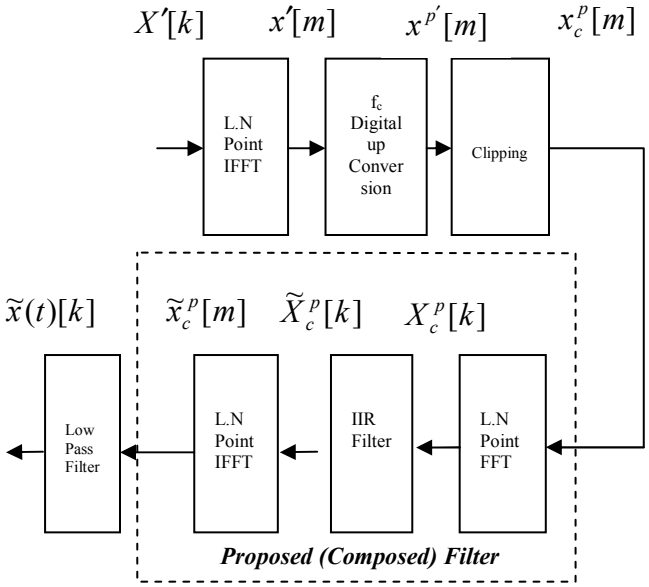


Fig.2 Block Diagram of Proposed Clipping & Filtering Method

This proposed scheme is shown in the figure 2. It shows a block diagram of a PAPR reduction scheme using clipping and filtering, where L is the oversampling factor and N is the number of subcarriers. The input of the IFFT block is the interpolated signal introducing $N(L-1)$ zeros (also, known as zero padding) in the middle of the original signal. Let, $x_c^p[m]$ denote the clipped version of $x^p[m]$ which is expressed as,

$$x_c^p[m] = \begin{cases} -A & x^p[m] \le -A \\ x^p[m] & |x^p[m]| < A \\ A & x^p[m] \ge A \end{cases} \quad (1)$$

Where, A is the pre-specified clipping level. After clipping, the signals are passed through the proposed filter (Composed Filter). The filter itself consists on a set of FFT-IFFT operations where filtering takes place in frequency domain after the FFT function. The FFT function transforms the clipped signal $x_c^p[m]$ to frequency domain yielding $X_c^p[k]$. The information components of $X_c^p[k]$ are passed to a IIR based peak resonator filter producing $\tilde{X}_c^p[k]$.This filtered signal is passed to the unchanged condition of IFFT block and the out-of-band radiation that fell in the zeros is set back to zero. The IFFT block of the filter transforms the signal to time domain and thus obtain $\tilde{x}_c^p[m]$. Table I shows the values of parameters used in the QPSK & QAM system for analyzing the performance of clipping and filtering technique. We have simulated the both methods (Existing and Proposed) with the same parameters.

TABLE I
PARAMETERS USED FOR SIMULATION

Parameters	Value
Bandwidth (BW)	1 MHz
Over sampling factor (L)	8
Sampling frequency, f_s = BW*L	8 MHz
Carrier frequency, f_c	2 MHz
No. of Subscribers (N)	128
CP / GI size	32
Modulation Format	QPSK and QAM
Clipping Ratio (CR)	0.8, 1.0, 1.2, 1.4, 1.6

IV. SIMULATED RESULTS

At first, we simulate the PAPR distribution for CR values =0.8, 1.0, 1.2, 1.4, 1.6 with QPSK modulation & N=128 and then, with QAM modulation and N=128. PAPR distributions for different CR value are shown in figure 3 & figure 4 in case of existing & proposed method respectively.

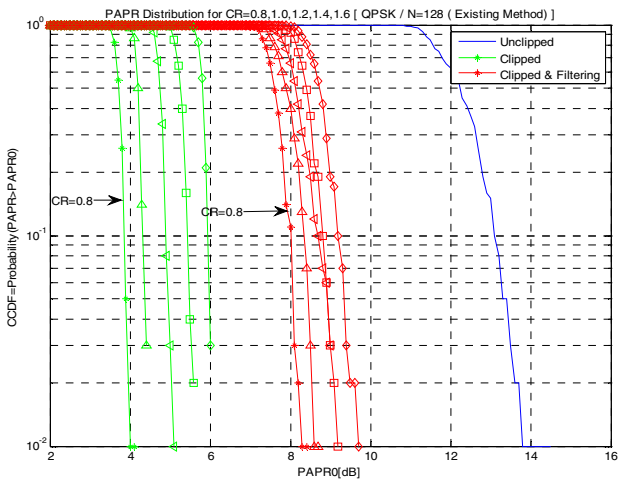


Fig. 3 PAPR Distribution [QPSK /N=128 (Existing)]

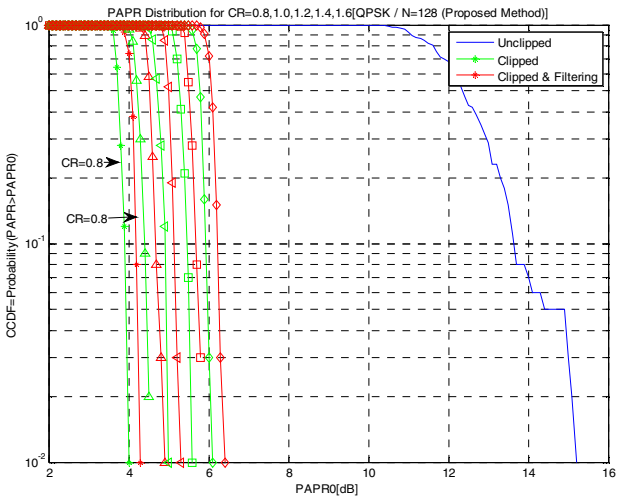


Fig.4 PAPR Distribution [QPSK /N=128 (Proposed)]

From table II, it is clearly observed that the proposed method reduces PAPR significantly with respect to existing method analysis [6].

TABLE II
COMPARISON OF EXISTING WITH PROPOSED METHOD FOR PAPR VALUE [QPSK /N=128]

CR value	PAPR value (dB)(Existing)	PAPR value (dB)(Proposed)	Improvement in PAPR value (dB)
0.8	8.101	4.217	3.884
1.0	8.353	4.679	3.674
1.2	8.721	5.134	3.587
1.4	8.817	5.681	3.136
1.6	9.221	6.225	2.996

TABLE III
COMPARISON OF EXISTING WITH PROPOSED METHOD FOR PAPR VALUE [QAM /N=128]

CR value	PAPR value (dB)(Existing)	PAPR value (dB)(Proposed)	Improvement in PAPR value (dB)
0.8	8.231	4.184	4.047
1.0	8.323	4.652	3.671
1.2	8.652	5.111	3.541
1.4	8.929	5.710	3.219
1.6	9.151	6.272	2.879

Figure 5 & figure 6 shows the simulation results for QAM modulation & N=128 for existing & proposed method. From table III, it is obvious that the proposed method reduces PAPR significantly. Now, if we analyse the values for different CR values in case of QPSK & QAM to show the effect of modulation on proposed filter design, it is observed that QAM provides less PAPR than QAM up to CR value=1.2. But, with the increasing value of CR, QPSK results less PAPR than QAM. So, for low CR value (More amount of clipping), QAM is more suitable than QPSK in case of proposed filter.

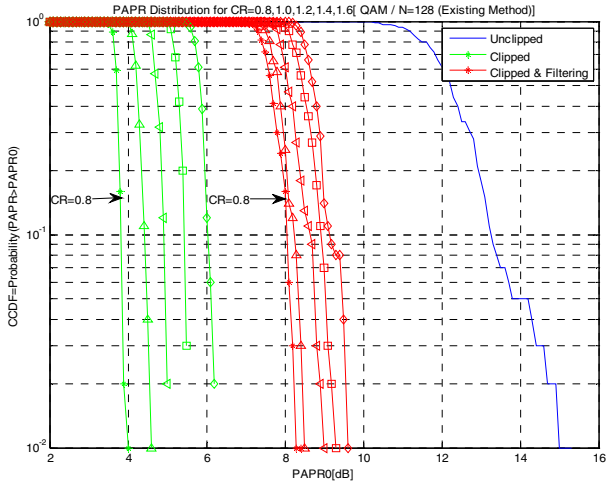


Fig.5 PAPR Distribution [QAM /N=128 (Existing)]

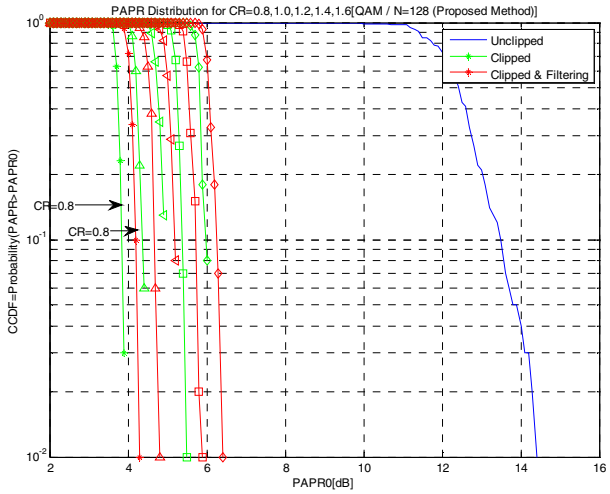


Fig.6 PAPR Distribution [QAM/N=128 (Proposed)]

The clipped & filtered signal is then passed through the AWGN channel and BER are measured for both existing & proposed methods. Figure 7 & 8 shows the BER curve for QPSK in case of existing & proposed method correspondingly. Figure 9 & 10 shows the BER curve for QAM in case of existing & proposed method respectively.

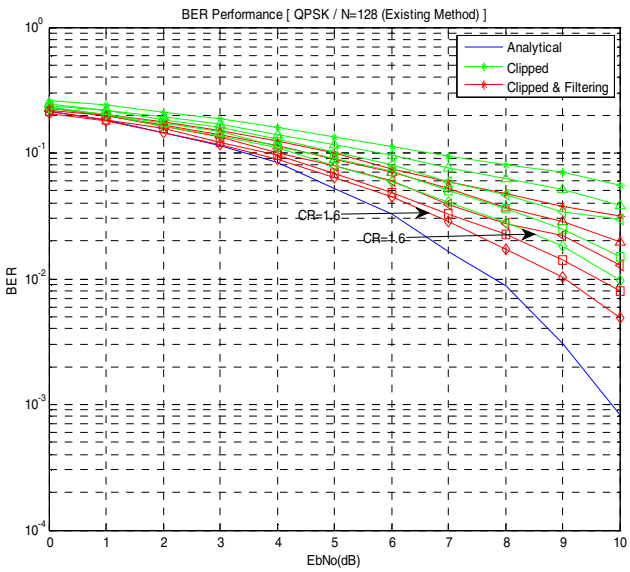


Fig. 7. BER Performance [QPSK /N=128 (Existing)]

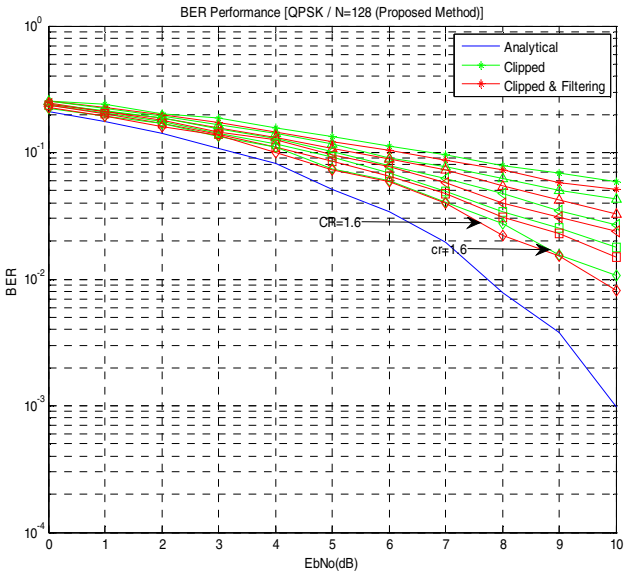


Fig. 8 BER Performance [QPSK /N=128 (Proposed)]

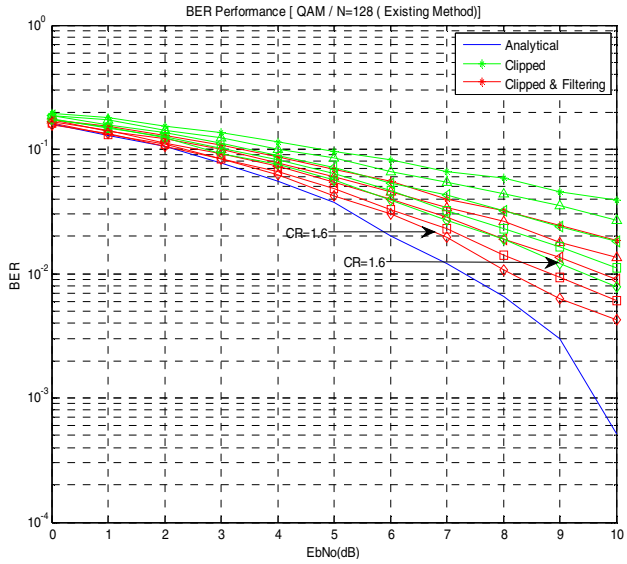


Fig. 9 BER Performance [QAM /N=128 (Existing)]

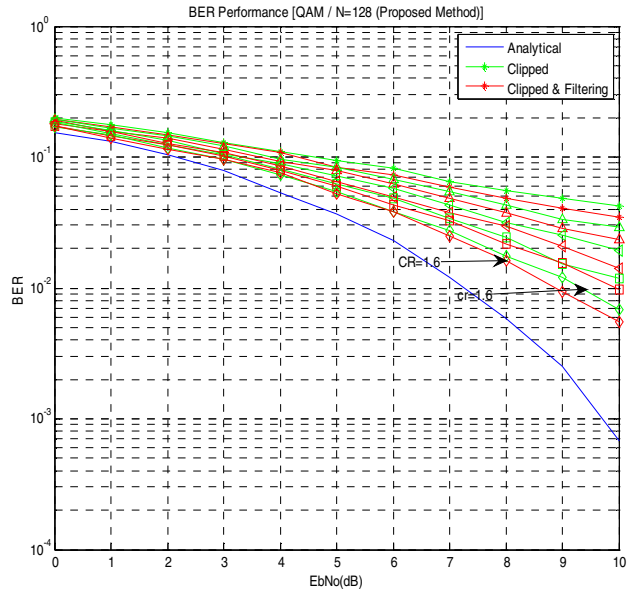


Fig. 10. BER Performance for [QAM /N=128 (Proposed)]

TABLE IV
COMPARISON OF EXISTING WITH PROPOSED METHOD FOR BER VALUE [QPSK /N=128]

CR value	BER value (Existing)	BER value (Proposed)	Difference in BER value
0.8	0.07413	0.10481	-0.03068
1.0	0.06984	0.08886	-0.01902
1.2	0.05982	0.07685	-0.01703
1.4	0.04891	0.06535	-0.01644
1.6	0.04449	0.05937	-0.01488

TABLE V
COMPARISON OF EXISTING WITH PROPOSED METHOD FOR BER VALUE [QAM /N=128]

CR value	BER value (Existing)	BER value (Proposed)	Difference in BER value
0.8	0.05563	0.07354	-0.01791
1.0	0.04614	0.06244	-0.0163
1.2	0.04016	0.04961	-0.00945
1.4	0.03268	0.04346	-0.01078
1.6	0.03055	0.03799	-0.00744

From table IV and table V, it is noted that BER are increased in both the cases. Now, if we compare the values for different CR values in case of QPSK & QAM to show the effect of modulation on proposed filter design, it is observed that for the same simulation parameters, QAM provides less BER degradation than QPSK in all cases. So, QAM is more suitable for proposed method in all aspects.

V. CONCLUSIONS

In this paper, an improved scheme of amplitude clipping & filtering based PAPR reduction technique has been analysed where PAPR reduces significantly compare to an existing method with slightly increase of BER in case of both QPSK & QAM modulations. It is also observed from this simulation, that for low CR (more amount of clipping), QAM is better candidate in case of PAPR reduction as well as BER performance for all amount of clipping. In order to estimate the OFDM system performance in real world, multipath Rayleigh fading can be a major consideration in next time.

REFERENCES

[1] Y. Rahmatallah and S.Mohan, "Peak-To-Average Power Ratio Reduction in OFDM Systems: A Survey and Taxonomy", *IEEE Commun. Surveys & Tutorials,* vol. 15, no. 4, pp. 1567-1592, First Quarter 2013.

[2] S. H. Han and J. H. Lee, "An overview of peak-to-average power ratio reduction techniques for multicarrier transmission", *IEEE Wireless Commun,* vol. 12, no.2, pp.56-65, Apr. 2005

[3] Xiaodong Li and Leonard J. Cimini, Jr.,"Effects of Clipping and Filtering on the Performance of OFDM", *IEEE Communs Letter*, vol. 2, no. 5, pp. 131–133, 1998.

[4] J. Armstrong,"Peak-to-Average Power Reduction for OFDM by Repeated Clipping and Frequency Domain Filtering", *Electronics Letters,* vol. 38, no. 5, pp. 246-247, 2002.

[5] H.Chen and A. Haimovich, "An Iterative Method to Restore the Performance of Clipped and Filtered OFDM Signals", *IEEE International Conference of Communications, (ICC'03).* vol.5, pp.3438-3442, 2003.

[6] Yong Soo Cho, Jaekwon Kim, Won Young Yang and Chung-Gu Kang, *MIMO OFDM Wireless Communications with MATLAB*, Singapore: John Wiley & Sons (Asia) Pte Ltd, 2010.

A Simple Wideband Microstrip Bandstop Filter for WLAN and WiMAX Band Rejection Purpose

Lakhindar Murmu[1*] and Sushrut Das [2]
[1, 2] Department of Electronics Engineering
Indian School of Mines, Dhanbad, India.
[*]lakhndar.kgec25@gmail.com

Abstract— **In this paper, a simple hexagonal wideband microstrip bandstop filter with an open-end stepped impedance stub has been proposed for the rejection of WLAN and WiMAX bands. Controlling the dimensions of the shunt open-end stepped impedance stub, the rejection bandwidth and the level of rejection can be controlled. The structure is very simple and is also easy to fabricate. The final filter structure provides a stop band that extends from 3.19 GHz to 5.36 GHz within and more than -20dB rejection and hence it can reject the 3.5 GHz WiMAX bands and 5.2 GHz WLAN frequency bands. The simulated response has been validated after comparing it with the measured data.**

Index Terms—**Open loop hexagonal filter, bandstop filter (BSF), stepped impedance resonator (SIR).**

I. Introduction

With rapid growth of communication technologies, WLAN and WiMAX communications have been developed a lot. These technologies have become so popular that in present days it has become almost a part of our life and almost every corner of the world is flooded with the signals from WLAN and WiMAX. These high speed WLAN and WiMAX signals propagate rapidly through the infrastructure of office and home environments and couple with other RF systems. These coupled signals propagate along the transmission line and circuits of the RF system and may result in an undesired behaviour of the system. Therefore to protect the RF systems from these signals bandstop filters can be used.

Recently, numerous methods and structures have been proposed for realizing the bandstop filters [1-2]. Based on the formulae of SIR, bandstop filters with high skirt selectivity as well as wide rejection bandwidth and miniature circuit configuration were presented in [1,2]. Divyabramham et al. [2] presented a design technique of sharp-rejection wideband bandstop filters by using a lossless transmission line model. In the year 2008, a new compact bandstop microstrip filter using two cascaded ring resonators was presented and the equivalent circuit model was developed to characterize the proposed bandstop structure [3]. One year later, in [4] a technique to design a dual-band bandstop filter (DBBSF) was presented with a dual-mode loop resonator.

The hexagonal geometry promises better confinement in the microstrip circuits due to their large interior angles (as compared to rectangle and square counterparts).Open loop hexagonal multiplexer for communication system were proposed by R. Kumar and G.A. Edae [5]. In that paper they have presented the tri-band band multiplexer topology based on the hexagonal close loop resonators of different size which were capacitive coupled from a single input. In [6] a novel microstrip bandstop filter was proposed, where signal interference technology was applied for achieving the bandstop behaviour. Two open loop resonators were introduced to the conventional transmission line with larger electrical length. Then, two transmission zeros were obtained in the stopband. Meanwhile, several transmission poles were used to improve the flatness of the passband. In comparison with the traditional bandstop filters using signal interference technique, the proposed filter has wider stopband and flat passband. In the same year, a new and miniature circuit configuration for wideband bandstop filter (BSF) operating at X-band range was presented [7]. The presented planar microstrip filter consists of a three-section stepped impedance resonator (SIR) paralleled with three open-circuited stepped impedance stubs (SISs).

In this paper the authors have proposed a compact open loop hexagonal resonator filter with open-end stepped impedance stub which has been designed for the rejection of 3.5 GHz WiMAX band and 5.2 GHz WLAN frequency band. Commercial IE3D simulator has been used to design the structure.

II. Open Loop Hexagonal Filter Design

An open loop hexagonal loop filter with direct coupled tapered ports is shown in Fig.1. The length L (total loop length, L= [(6×a) –g]) determines the centre frequency of filter. A Pair of 50-Ω collinear tapered lines feed with the resonator. RT/Duroid 6010 with a thickness of 1.27 mm and a relative dielectric constant of 10.2 has been used as the substrate. The simulated response i.e. insertion loss and return loss versus frequency is shown in Fig.2. The figure reveals that it has a band pass response from DC to 2.2 GHz (up to 2.2 GHz) and a band stop response from 2.2GHz to 5.4 GHz.

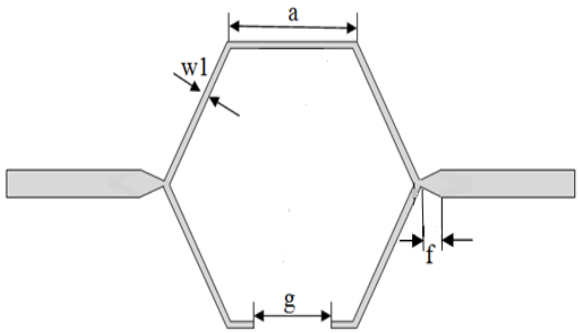


Fig.1. Hexagonal open loop bandstop filter (Filter A).

Inspired by the band stop behaviour of a simple hexagonal open loop, a second hexagonal bandstop filter, but this time loaded with open ended stepped impedance stub, has been

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

developed as shown in Fig.3. The filter structure has been studied by varying the dimensions of the stubs of the hexagonal resonator. The frequency response of the proposed filter is shown in Fig.4 for different parameter values. The respective filter parameters are tabulated in table 1.

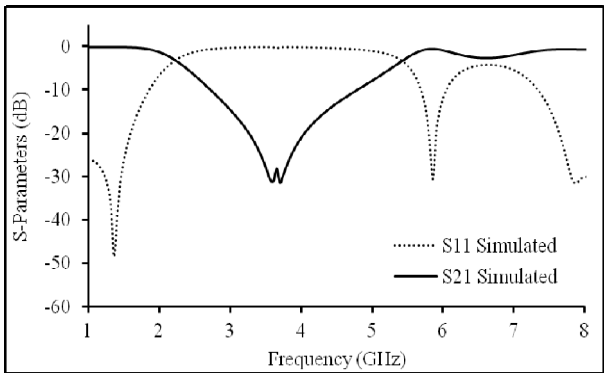


Fig.2. Simulated response of hexagonal open loop bandstop filter (Filter A).

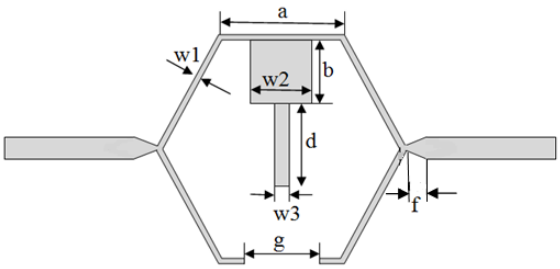


Fig.3. Hexagonal resonator bandstop filter with open-end stubs (Filter B).

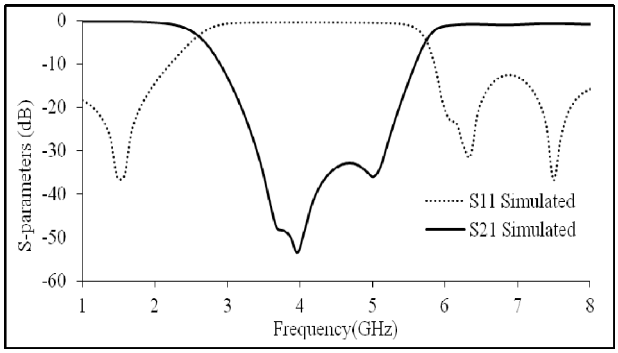


Fig.4. Simulated frequency response of the proposed (Filter B).

Table 1: Geometrical dimensions of the filters A-B

Filter	A	B
a	6mm	6mm
b	--	3mm
d	--	4mm
f	2.05mm	2.05mm
g	1mm	2mm
w1	0.85mm	0.2mm
w2	--	3mm
w3	--	1mm

III. Fabrication and Measurement Results

All Based on the simulated results obtained from IE3D, the filter has been fabricated with the proposed dimensions: a=6mm, b=3mm, d=4mm, w1=0.2mm, w2=3mm, w3=1mm, g=2mm (Filter B) as shown in Fig. 5 and then measured for frequency response. The measurement was carried out using an Agilent E5071C Vector Network Analyzer and the measured response is shown in Fig.6. Simulated frequency response of the proposed filter is also shown in same figure. The measure frequency response shows the 3dB cut-off frequencies of 2.62 GHz and 5.63 GHz. The highest insertion loss in the low and in the high frequency pass-band is better than 0.3 dB and 1.5 dB, respectively. The highest return loss in the stop-band is better than 0.4 dB. The measured result provides a stop band that extends from 3.03 GHz to 5.28 GHz within and more than -20dB rejection and hence it can reject the 3.5 GHz WiMAX bands and 5.2 GHz WLAN frequency bands. The measured result shows a good agreement with the simulated data. The result shows a passband return loss of about 20dB & stopband return loss about 0.4 dB.

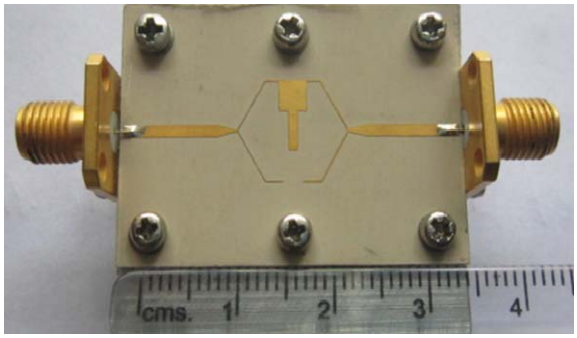


Fig.5. Photograph of bandstop filter (Filter B).

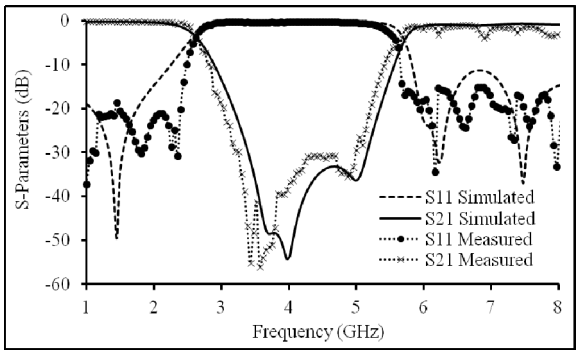


Fig.6. Measured and simulated frequency response of the proposed filter (Filter B).

The stop band insertion loss is better than 30dB. Fig. 6 also shows that at the rejection band plenty of ripples are present whereas the simulation is smooth. The ripples are due to the connector's flanges and screws that are missing in the simulation model.

IV. Conclusions

A Compact open loop hexagonal resonator bandstop filters with and without a loaded open stub have been presented in this paper. The bandwidth of the filter with a loaded open stub has been observed to have a better performance. The optimization of the various parameters of the filter has been accomplished to achieve a good response. The proposed filter is compact, easy to fabricate and easy to integrate with MIC/MMIC microwave circuits. The final filter structure provides a stop band that extends from about 3.19 GHz to 5.36 GHz and hence it can reject the 3.5 GHz WiMAX and

5.2 GHz WLAN frequency bands. The simulated response has been validated after comparing it with the measured data. In the measured results, the centre frequency of the fabricated filter has been shifted towards a lower frequency. This is due to the fabrication tolerance.

REFERENCES

[1] Wang Y.-Z. and M.-L. Her, "Compact microstrip bandstop filters using stepped-impedance resonator (SIR) and spur-line sections," IEE Proc.-Microw. Antennas Propag. Vol. 153, No. 5, pp. 435-440, 2006.

[2] Divyabramham, K., M.K. Mandal and S. Sanyal, "Sharp-Rejection Wideband Bandstop Filters," IEEE Microwave and Wireless Components Letters, Vol. 18, No. 10, pp. 662-664, 2008.

[3] A. Boutejdar, A. Batmanov, A. Omar and E. Burte, "A new method to improve the reject band of a 5.6 GHz bandstop filter using λ/2 open-loop ring microstrip resonators," Asia-Pacific Microwave Conference, APMC,pp.1-4, 2008.

[4] H.-K. Chiou and C.-F. Tai, "Dual-band microstrip bandstop filter using dual-mode loop resonator," Electronics Letters, Vol.45, No.10, pp.507-509, 2009.

[5] Raj Kumar and Girma Alemu Edae, "On the Design of Close and Open Loop Hexagonal Multiplexer for Communication System," 2nd International Conference and workshop on Emerging Trends in Technology (ICWET), Proceedings published by International Journal of Computer Applications® (IJCA), pp.7-14, 2011.

[6] Hao-Ran Zhu , Wei Shen and Jun-Fa Mao, "A novel bandstop filter with wide stopband using signal interference technology", International Workshop on Microwave and Millimeter Wave Circuits and System Technology (MMWCST), pp.1-4, 2012.

[7] Hui Chen, Hong-sheng Zhong, Zhi-qin Zhao and Yu-xing Zhang, "A compact X-band microstrip bandstop filter using stepped impedance line with open-end stepped impedance stubs," International Conference on Microwave and Millimeter Wave Technology (ICMMT), Vol.2, pp.1-3, 2012.

A Novel DCSK Communication Scheme for PLCC based DAS

Subhra J. Sarkar[1], Palash K. Kundu[2]

[1]Dept. of Electrical Engg.
Batanagar Institute of Engg., Management & Sc. Kolkata, India, subhro89@gmail.com
[2]Dept. of Electrical Engg.
Jadavpur University, Kolkata, India, palashm.kushi@gmail.com

Abstract- **With the increasing popularity of power line communication technologies, modifications over Spread Spectrum Technologies were made so as to develop a suitable alternative for power line communication. Differential Code Shift Keying (DCSK) is one such technology developed by Yitran Technologies & patented on May 16th, 2000. In this paper, it is proposed to develop this DCSK Communication Scheme which can be used for PLCC system for DAS application.**

Keywords- **Power Line Carrier Communication (PLCC), DCSK, Chirp, Superchirp, Spread Spectrum Technology**

I. INTRODUCTION

In this modern world, digital communication is an interesting field of research as it plays a crucial role in our daily life. New technologies are developing day to day to reduce the effect of noise present in the transmission medium on the modulated signal. The effect of noise becomes more prominent if the transmission medium is a power line cable as it not designed for it. Noise levels in the networks are often excessive; cable attenuation in the systems at the frequencies of interest is often very large; important channel parameters such as impedance and attenuation are hard to predict due to their unpredictable nature [1].

But modern day automation prefers the power line carrier communication i.e. PLCC for low power remote monitoring & control. But due to noisier environment, the quality & speed of data transfer reduces significantly with the conventional modulation techniques. So in order to overcome these problems, new modulation techniques are developed. Differential Code Shift Keying (D.C.S.K.) is one such technique which is suitable especially for PLCC.

In this paper attempt has been made to develop a system for implementing DCSK modulation scheme employing MATLAB software. Separate programs are designed at transmitting & receiving end for single character & string. Then the principle is applied to design a multi-drop system where device address & the corresponding data are merged together in a single string at transmitting end, which is then separated at the receiving end to display the device address & the data value from the particular device.

II. POWER LINE CARRIER COMMUNICATION

In the recent times, there are various communication technologies for transmitting information from one point to another through suitable communication medium which may be wired or wireless. As mentioned previously, PLCC is a method of communication via power line. In today's world, every house and building has properly installed electricity connections. If the existing AC power lines can be used as a medium to transfer the information, it becomes easier to connect the houses with a high speed network access point & hence eliminating the need of Ethernet cable. So, it can also be used in building/home automation [7, 11]. The data collected from different sensors is transmitted on these power lines thereby reducing the maintenance cost of the additional wiring arrangements. This technology is also used to provide Internet connection in some countries [2]. PLCC can be employed successfully in automatic meter reading, home automation & control, voice or data communication etc. While dealing with PLC communication, commonly used digital communication techniques like FSK & PSK suffers with serious problems due to power line characteristics. At high frequencies (i.e. at communication frequencies), power line characteristics varies greatly if compared with that at power frequency. The difference in characteristic impedance & load impedance at such high frequencies often results high attenuation at such communication frequencies. Apart from it, significant noise is produced by various devices connected to the system thereby degrading the performance of conventional digital communication techniques. It was found that 'Spread Spectrum Modulation Technique' is gives better results under power line conditions [1, 10].

III. DIFFERENTIAL CODE SHIFT KEYING (D.C.S.K.) MODULATION TECHNIQUE

Spread-spectrum techniques spreads any signal (electrical, electromagnetic, or acoustic signal) generated with a particular bandwidth deliberately in the frequency domain,

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

resulting the formation of a signal with wider bandwidth [8]. Spread spectrum uses wideband, noise-like signals which are hard to detect, intercept, or demodulate. It was its anti-jam feature which made it popular, especially in military environments [3, 12].

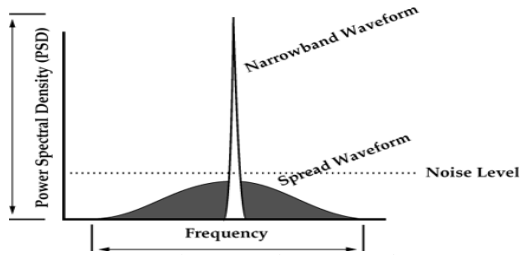


Fig.1. Spread-spectrum scheme

The high bandwidth of spread spectrum results a much lower spectral power density (in Watts per Hertz), than narrow band transmitters. Due to this low spectral power density, spread-spectrum and narrowband signals can occupy the same band, with little or no interference [5]. The block diagram of spread-spectrum scheme is as shown below.

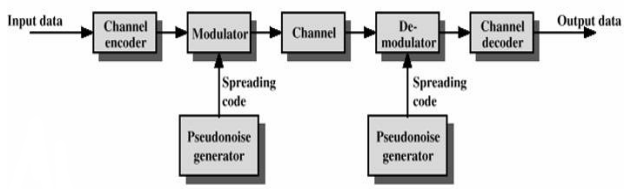


Fig.2. Block diagram of Spread-Spectrum Communication scheme

In the block diagram it is clearly visible that spreading codes are required at both modulating & demodulating ends which must be supplied from same source or different sources producing same spreading codes & thus need some synchronization mechanism. This synchronization can be implemented by using an acquisition method in combination with a tracking loop or other tracking mechanism. In noisy, unpredictable environment like that in AC power line, the tracking loop often fails frequently causing loss of information. The system compensating these problems is generally large, complex & expensive. Apart from it, these systems can transmit only one or two bits per symbol [3]. These drawbacks led to the development of a modified spread spectrum technology, Differential Code Shift Keying (DCSK) that is especially meant for power line communication scheme [6]. It modulates the symbol by cyclic shifting of the basic symbol in the transmitter, and detecting that shift in the receiver [4].

As mentioned previously, in this modulation technique data transmission is by the cyclic shifting of the spreading waveforms which may be any waveform having suitable auto correlation. Chirp is the commonly used waveform for this purpose, which is a frequency modulated signal segment whose frequency increases or decreases with time [9]. These chirp waveforms of constant length are then rotated suitably depending on the data value & are termed as 'Unit Symbol Time' (UST) [3].

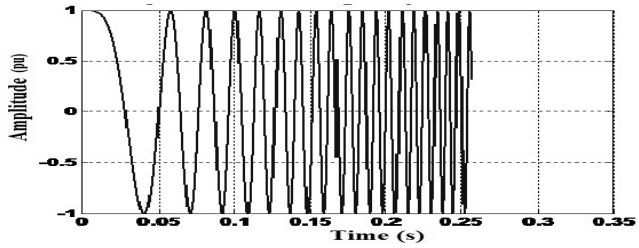


Fig.3. A linear chirp waveform

Let us illustrate this chirp rotation of the linear chirp as shown above. For performing the rotation, let us divide the chirp waveform in 255 equal segments each of time interval of 1ms. This rotation of waveform may in clockwise or counter-clockwise direction. In counter-clockwise rotation, segments are circularly rotated by 1 place in counter-clockwise direction. The anti-circularly rotated chirp rotated by 100 units counter-clockwise will be is as shown below.

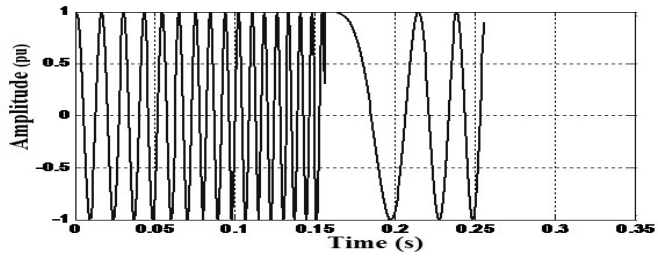


Fig.4. Counter-clockwise rotated chirp by hundred units

In clockwise rotation, segments are circularly rotated by 1 place in clockwise. The circularly rotated chirp will be rotated by 100 units clockwise is as shown below.

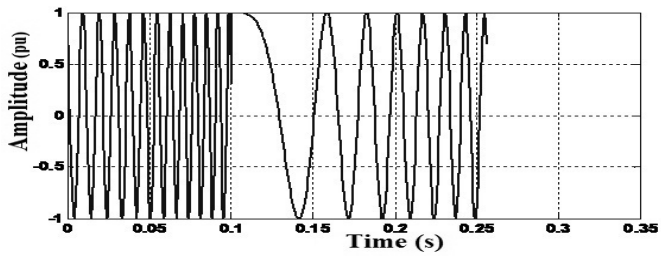


Fig.5. Clockwise rotated chirp by hundred unit

There may be two alternatives for designing DCSK communication scheme. In the first arrangement, the data is conveyed in the amount of rotation of the chirp. Alternatively, data may be conveyed in the differential shift between two consecutive chirps [3]. In this paper, the first case is only considered. Generally data is in the form of string which can be transmitted is in the form of a collection of a no. of shifted chirp waveforms (USTs) forming a superchirp or symbol stream [3]. A superchirp of character string "$jS" is shown below. As the given string possesses three characters, so three

shifted chirp waveforms placed in sequences will form the superchirp waveform. The arrows drawn in the bottom of superchirp indicates the amount of rotation of individual chirps corresponding to the ASCII value of the characters present in the string.

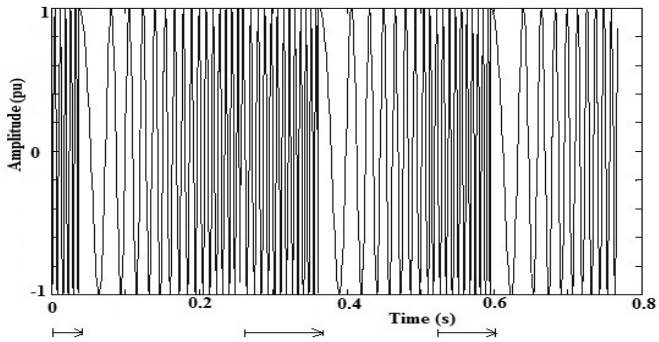


Fig.6. A superchirp or symbol stream of character string "$jS"

At the receiving end, demodulation is done by comparing the individual USTs with the non-rotated chirp waveform to determine the amount of rotation. This is done by the series of rotations & simultaneous comparison gives the transmitted data [3]. This modulation scheme has the following advantages:

a. High transmission reliability
b. Simple & fast synchronization
c. Immediate recovery from severe fading
d. Higher data throughput rate using the same symbol time duration as in direct sequence spread spectrum communication system.
e. Robust against channels characterized by frequency varying signal to noise ratios.
f. Low cost implementation in a single VLSI ICs.

IV. RESULT AND ANALYSIS

At the transmission end, the character is encoded in waveform as shifted chirp & is transmitted. The algorithm of transmitting a character employing DCSK communication scheme is as below.

ch: Input character
fi: Initial frequency of chirp waveform
ff: Final frequency of chirp waveform

STEP 1: Enter ch, fi & ff.
STEP 2: Determine the ASCII value of ch.
STEP 3: If the no. of significant bits of ff is n, then chirp length is given by 255.99 $X10^{-n}$ s.
STEP 4: Produce chirp of desired length with given fi & ff.
STEP 5: Divide chirp in 256 equal sections.
STEP 6: cnt:=0
STEP 7: do cnt!=ASCII
STEP 8: : Rotate the chirp 1 unit counter clockwise.
STEP 9: : Increment cnt by 1.
STEP 10: End of the loop.
STEP 11: Transmit rotated chirp.
STEP 12: End.

At the receiving end, the transmitted waveform is obtained & is continuously compared with the non-shifted chirp signal as used at the transmitting end to determine the amount of rotation. The algorithm of decoding the modulated signal at the receiving end employing DCSK communication scheme is as below.

r(t): Non-shifted reference chirp
u(t): Shifted received chirp

STEP 1: Obtain r(t) & u(t).
STEP 2: Split u(t) in 256 equal sections.
STEP 3: cnt:=0.
STEP 4: do r(t) & u(t) are same.
STEP 5: : Rotate u(t) 1 unit clockwise.
STEP 6: : Increment cnt by 1.
STEP 7: End of the loop.
STEP 8: ASCII value of transmitted character=cnt.
STEP 9: Obtain the character.
STEP 10: End.

The very same logic can be implemented for string transmission using DCSK communication scheme. Then output will be a superchirp, comprising a no. of chirps rotated corresponding to the characters in the string.

str: Input string
fi: Initial frequency of chirp waveform
ff: Final frequency of chirp waveform

STEP 1: Enter str, fi & ff.
STEP 2: Obtain chirps for each character in str of desired length with given fi & ff.
STEP 3: Merge individual chirps for each character to form a superchirp.
STEP 4: Transmit superchirp.
STEP 5: End.

At the receiving end this received superchirp is sub-divided in the corresponding chirps. These chirps are then continuously compared with the non-shifted chirp signal to determine the amount of rotation & hence the ASCII value of character. Thus all the characters of the data string can be obtained. The flowchart of the demodulating end is as below.

s(t): Superchirp
r(t): Non-shifted reference chirp

STEP 1: Split s(t) in individual chirps.

STEP 2: Determine the characters transmitted via individual chirps by comparing with r(t) & store the characters separately.
STEP 3: Merge the characters to form the string.
STEP 4: End.

Utilizing the above communication method, a PLCC based data acquisition system can be implemented consisting of PLCC modem, microcontroller, sensor arrays and indicators as shown in fig. 7. Here a single phase 230V, 50 Hz A.C. supply has been utilized for communication channel run over the moderate distance within the site area.

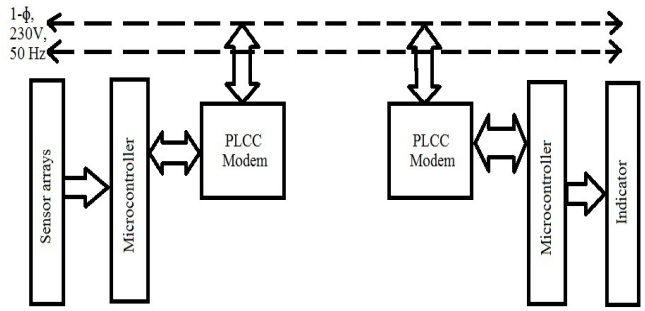


Fig.7. Proposed system: PLCC base DAS

V. DISCUSSION AND CONCLUSION

The algorithm is employed for developing MATLAB programs for character/string transmission & reception using PLCC scheme. For single character transmission & reception programs, the execution time is very small. But while executing the string transmission & reception program, it was found that the execution time required varies as some function of no. of characters in the string. The variation of execution time with the increase of characters in the string for both transmission & reception program is shown below.

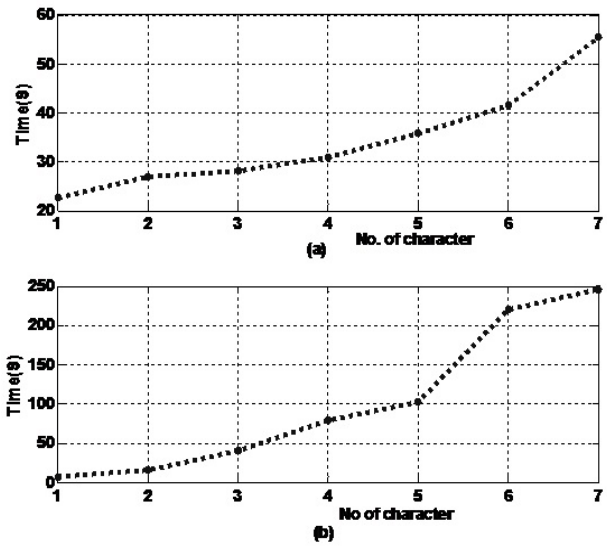


Fig.8. Variation of execution time with the variation of no. of characters in (a) Transmission program; (b) Reception program

From fig. 8, it is clearly visible that though the system is running successfully, it suffers serious problems regarding rate of data transfer for string transmission & reception. With increased no. of characters in the string, the execution time also increases significantly for both transmission & reception programs thereby limiting its rate of data transfer. Thus modifications are needed to overcome its limitations.

VI. REFERENCES

[1] Phil Sutterlin and Walter Downey, Echelon Corporation. *A Power Line Communication Tutorial- Challenges & Technologies*.

[2] Khyrram Hussain Zaberi, "Power Line Carrier (PLC) Communication Systems". M.S. thesis, Royal Institute of Technology, Stockholm, Sweden, September 2003.

[3] Dan Raphaeli, "Spread Spectrum Communication System utilizing Differential Code Shift Keying", US Patent 6064695, May 16th, 2000.

[4] DCSK Modulation. Yitran Technologies website. [Online]. Available: http://www.yitran.com/index.aspx?id=3387

[5] Hui Xiao, Xuebin Wu, Xiaojiao Tong, Rong Lian, Zeng Xiangjun and Sheng Su, "Application of Spread Spectrum Communication in Medium Voltage Power Line Carrier", *Asia and Pacific Transmission and Distribution Conference and Exhibition: 2005, IEEE/PES,* 2005, pp 1-4.

[6] Shivaram K., Rajendra N., Mahesh K., Murthy K.N.B. and Jawahar P., "Power line communication based automation system using a handheld Wi-Fi device", *IEEE 16th International Symposium on Consumer Electronics (ISCE), 2012* ,2012, pp 1-6.

[7] Miura N., Sato H., Narita H. and Takaki M., "Automatic meter-reading system by power line carrier communications", *IEE Proceedings C Generation, Transmission and Distribution*, Vol. 137, Issue: 1, pp. 25-31, Jan. 1990.

[8] Wikipedia: Spread Spectrum. [Online]. Available: http://en.wikipedia.org/wiki/Spread_spectrum

[9] Wikipedia:Chirp.[Online].Available:http://en.wikipedia.org/wiki/Chirp

[10] B.A.Mork, D.Ishchenko, X.Wang, A.D.Yerrabelli, R.P.Quest & C.P.Kinne, "Power Line Carrier Communication System Modeling", *International Conference on Power Systems Transients (IPST-05)*, 2005.

[11] Tosaphol Ratniyomchai, Uthai Jaithong, & Thanatchai Kulworawanichpong, "Power Line Carrier for Power Telemetering", International Science Index, Vol. 5, No. 8, pp. 544-547, Aug. 2011.

[12] Hui Xiao, Xuebin Wu, "Design of Medium Voltage Power Line Spread Spectrum Carrier Communication System Based on Embedded System", *5th International Conference on Wireless Communications, Networking and Mobile Computing, 2009. WiCom '09*, 2009, pp 1-4.

Probabilistic Trust Management in Wireless Communication System

Swarnali Hazra
Computer Science and Engineering
University of Calcutta
Kolkata, India
swarnali.cse@gmail.com

S.K.Setua
Computer Science and Engineering
University of Calcutta
Kolkata, India
sksetua@gmail.com

***Abstract*—Wireless communication system consists of wireless communicating entities that transmit data through bandwidth flexibly and conveniently. Wireless entities have the freedom to change their position which implies wireless communications support mobility in the system. With such benefits, wireless communication systems are challenged by many security issues. In this paper, we have prevented malicious behaviour of compromised communicating entity or any other malevolent entity to disrupt the communication system with our probabilistic approach of trust based security management. We have proposed Probabilistic Trust Management (PTM) in wireless communication system with consideration of current and previous interactions along with indirect evaluations depending on timing constraint. Our simulation results show the efficiency of proposed PTM in wireless communication system.**

Keywords—probability; direct interaction; indirect interaction; posetive interaction; negative interaction; trustor; trustee.

I. INTRODUCTION

In wireless communication system, wireless entities communicate flexibly and conveniently with each other. Wireless communication system gives flexibility to change position of entities on their need and provides convenient communication between them. Wireless communication is already influence our social interaction but it has many security issues. When wireless entities communicate with each other, any misbehaving entity can falsify them and degrade the expected performance of the communication system. The communicating entities may also be compromised. These necessitate a level of belief or trust in between communication entities.

To prevent such malicious behaviour in wireless communication system, we have proposed PTM (Probabilistic Trust Management) in wireless communication system. Trust [1] is a security approach in which a binary relationship is maintained between communicating devices based on computed trust level with respect to their interactions. Trust level evaluating entity is considered as trustor (TR) and interacting entity with TR for trust evaluation is considered as Trustee (TE). The major disadvantage of traditional trust management is the consideration of trust and untrust notion. The long chain of trustor and trustee necessitates the security analysis with a probabilistic approach. There exists uncertainty to be trusted of trustee to trustor which implies that there is always the probability of being untrusted of trustee to trustor. In our approach, evaluated probabilistic value decides the trust level in qualitative manner rather than quantitative. Our probabilistic trust evaluation is not bound to current direct interactions depending on expectations, but extended to previous direct interactions and notifications with respect to timing constraint.

Related works are discussed in section II . Our proposed probabilistic Trust Management (PTM) is discussed in section III. Section IV shows the simulation results and section V concludes the presented work.

II. RELATED WORK

In [3], authors proposed a probabilistic framework that models agent inter actions as a Hidden Markov Model (HMM). They maintain trust with fully probabilistic representation and reputation while modeling agents with dynamic behaviour. In [5], authors present probability theory and statistics based a novel Peer to Peer trust model to warrant the density of transaction data and the accuracy of computing results. The history records of transactions are considered to form out the trust value of peer with maximum likelihood estimation and hypothesis testing. In [6], authors proposed a probabilistic trust prediction model depending on trust-message passing which takes advantage of the two types of information considered as explicit information and implicit information. In [4], authors proposed probabilistic model to handle trust in P2P system. This proposal supports local computation and simple propagation form of the trust of peers into classes of other peers. In [2], authors quantified the uncertainty of probabilistic trust models using confidence intervals. Here, authors proposed an uncertainty driven method which reduces the risk of the trust-based utility maximization.

III. PTM: PROBABILISTIC TRUST MANAGEMENT

In our probabilistic approach, trust is measured at TR with its computed probability on TE's legitimacy. ${}_{TR}^{c}FP_{TE}^{t_{cur}} \in [0,1]$, where ${}_{TR}^{c}FP_{TE}^{t_{cur}}$ is final probability of legitimacy about TE evaluated by TR at current time t_{cur} on context c. ${}_{TR}^{c}T_{TE}^{t_{cur}} \propto {}_{TR}^{c}FP_{TE}^{t_{cur}}$, T denotes trust value which directly proportional to ${}_{TR}^{c}FP_{TE}^{t_{cur}}$. Our approach defines highest negative Trust: ${}_{TR}^{c}HNT_{TE}^{t_{cur}}$, highest positive trust: ${}_{TR}^{c}HPT_{TE}^{t_{cur}}$,

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

Positive trust: ${}_{TR}^{c}PT_{TE}^{t_{cur}}$, Negative trust: ${}_{TR}^{c}NT_{TE}^{t_{cur}}$ and uncertain trust ${}_{TR}^{c}UT_{TE}^{t_{cur}}$.

$$ {}_{TR}^{c}FP_{TE}^{t_{cur}} = \begin{cases} {}_{TR}^{c}HNT_{TE}^{t_{cur}}, & For\ {}_{TR}^{c}FP_{TE}^{t_{cur}} = 0 \\ {}_{TR}^{c}NT_{TE}^{t_{cur}}, & For\ 0 < {}_{TR}^{c}FP_{TE}^{t_{cur}} < 0.5 \\ {}_{TR}^{c}UT_{TE}^{t_{cur}}, & For\ {}_{TR}^{c}FP_{TE}^{t_{cur}} = 0.5 \\ {}_{TR}^{c}PT_{TE}^{t_{cur}}, & For\ 0.5 < {}_{TR}^{c}FP_{TE}^{t_{cur}} < 1 \\ {}_{TR}^{c}HPT_{TE}^{t_{cur}}, & For\ {}_{TR}^{c}FP_{TE}^{t_{cur}} = 1 \end{cases} \quad (1) $$

$$ {}_{TR}^{c}FP_{TE}^{t_{cur}} = f({}_{TR}^{c}FDP_{TE}^{t_{cur}}, {}_{TR}^{c}DIP_{TE}^{t_{cur}}) \quad (2) $$

Describing "(2)", final probability is the function of probability of TE's legitimacy in direct monitoring (${}_{TR}^{c}FPD_{TE}^{t_{cur}}$) and probability of TE's legitimacy in indirect monitoring (${}_{TR}^{c}FPI_{TE}^{t_{cur}}$). Following Bernoulli's probability combination we can formulate final probability as per "(3)".

$$ {}_{TR}^{c}FP_{TE}^{t_{cur}} = \sum_{i=1}^{n} Wi \times P_{Si} \quad (3) $$

where, $P_{Si} \in ({}_{TR}^{c}FPD_{TE}^{t_{cur}}, {}_{TR}^{c}FPI_{TE}^{t_{cur}})$ and i = 1...n and n=total numbers of probabilities for TE. For our work, n=2 and we have referred $Ps_1 = {}_{TR}^{c}FPD_{TE}^{t_{cur}}$ and $Ps_2 = {}_{TR}^{c}FPI_{TE}^{t_{cur}}$.

A. Probability in Direct Monitoring

We assume total numbers of past interactions is n[Int] and current interaction is $(n[Int]+1)^{th}$ interaction. TR has the knowledge about total n[Int] numbers of direct interactions with TE at different past time instances. $n[Int_P] \in n[Int]$, where $n[Int_P]$ is the total numbers of positive direct interactions. Total numbers of negative direct interactions $n[Int_N]=n[Int]- n[Int_P]$. Bernoulli distribution governs the probability of positive direct interaction as per "(4)".

$$ P(Int_i = true) = p $$
$$ P(Int_i = false) = 1 - p $$

$$ where\ \ i = 1,2, \ldots (n[Int] + 1) \quad (4) $$

P is an unknown measure of interaction being positive and p denotes probability. Int_i denotes ith direct interaction. True value denotes 1 and false value denotes 0.

- Int_i = true , it denotes the ith interaction is positive by legitimate TE.
- Int_i = false , it implies that TE's misbehaviour is identified and it is negative interaction by malevolent TE.
- $n[Int]_{lgtm}$ = number of interactions that returns true value with respect to legitimacy of TE.

Every Int_i is independent to each other for all i's. Since binomial distribution provides discrete probability distribution, we can represent the observation of success that $n[Int]_{lgtm} = n[Int_P]$ among n[Int] numbers of interactions as $P(n[Int]_{lgtm}=$ true | p) following binomial distribution shown in "(5)".

$$ P\left(\left(n[Int]_{legitimate} = n[Int_P] \middle| p\right) \right) = $$

$$ \binom{n[Int]}{n[Int_P]} p^{n[Int_P]} (1-p)^{n[Int]-n[Int_P]} \quad (5) $$

The probability of $(n[Int]+1)^{th}$ interaction that returns true value after performing n[Int] numbers of interactions by TE towards TR, is represented as $P(n(Int)+1_{lgtm}=$ true). Using Bayesian approach TR can evaluate $P(Int_{(n[Int]+1)}=$ true | $n[Int]_{lgtm} = n[Int_P]$) for TE. From Bayes' theorem TR calculate the probability distribution function for $P(Int_{(n[Int]+1)}=$ true) as follows:

$$ P\left(Int_{(n[Int]+1)} = true \,\middle|\, n[Int]_{lgtm} = n[Int_P] \right) $$

$$ = \frac{P(Int_{(n[Int]+1)}=true\,, n[Int]_{lgtm}=n[Int_P])}{P(n(Int)_{lgtm}=n[Int_P])} \quad (6) $$

From marginal distribution we get "(7)".

$$ P\left(n(Int)_{lgtm} = n[Int_P] \right) $$

$$ = \int_0^1 P\left(n(Int)_{lgtm} = n[Int_P] \middle| p \right) f(p) dp \quad (7) $$

On considering other hand, the evaluation of $P\left(Int_{(n[Int]+1)} = true\,,\ n[Int]_{lgtm} = n[Int_P]\right)$ as follows:

$$ P\left(Int_{(n[Int]+1)} = true\,,\ n[Int]_{lgtm} = n[Int_P]\right) = $$

$$ \int_0^1 P\left(Int_{(n[Int]+1)} = true \middle| p\right) P\left(n[Int]_{lgtm} = n[Int_P] \middle| p\right) f(p) dp \quad (8) $$

Since, $Int_{(n[Int]+1)}$ = true = p, we get "(9)" from "(8)" as follows:

$$ P\left(Int_{(n[Int]+1)} = true\,,\ n[Int]_{lgtm} = n[Int_P]\right) $$

$$ = \int_0^1 p.P\left(n[Int]_{lgtm} = n[Int_P] \middle| p\right) f(p)\, dp \quad (9) $$

Replacing the respective values in "(6)" from "(7)" and "(9)" we get "(10)" as belolow:.

$$ P\left(Int_{(n[Int]+1)} = true \,\middle|\, n(Int)_{lgtm} = n(Int_P) \right) $$

$$ = \frac{\int_0^1 p.P(n[Int]_{lgtm}=n[Int_P]|p) f(p)\, dp}{\int_0^1 P(n(Int)_{lgtm}=n[Int_P]|p) f(p)\, dp} \quad (10) $$

Here no prior information exists about p's distribution for TE. As a consequence, with an uniform prior distribution, f(p)=1, similar to [7] is shown in "(11)".

$$ P\left(Int_{(n[Int]+1)} = true \,\middle|\, n(Int)_{lgtm} = n(Int_P) \right) $$

$$ = \frac{\int_0^1 p.P(n[Int]_{lgtm}=n[Int_P]|p)\, dp}{\int_0^1 P(n(Int)_{lgtm}=n[Int_P]|p)\, dp} \quad (11) $$

Replacing the value of $P(n[Int]_{lgtm} = n[Int_p] \mid p)$ in "(11)" by the value of "(5)" we get "(12)".

$$P\left(Int_{(n[Int]+1)} = true \mid n(Int)_{lgtm} = n(Int_P) \right) = \frac{\int_0^1 p \binom{n[Int]}{n[Int_P]} p^{n[Int_P]} (1-p)^{n[Int]-n[Int_P]}\, dp}{\int_0^1 \binom{n[Int]}{n[Int_P]} p^{n[Int_P]} (1-p)^{n[Int]-n[Int_P]}\, dp} = \frac{n[Int_P]+1}{n[Int]+2} = \frac{n[Int_P]+1}{(n[Int_P]+n[Int_N])+2} \quad (12)$$

From "(12)", final probability in direct interaction ${}_{Ri}^{c}FPD_{TE}^{t_{mr}}$ is evaluated as per "(13)".

$${}_{Ri}^{c}FPD_{TE}^{t_{mr}} = f\left(P\left(Int_{(n[Int]+1)} = true \mid n[In]_{lgtm} = n[Int_P] \right), W\right) = \frac{\sum_{i=1}^{n}\left(\frac{n[Int_P]_{total.i}}{n[Int]_{total.i}} \times n_i[Int_P]\right) + 1}{\sum_{i=1}^{n}\left(\frac{n[Int_P]_{total.i}}{n[Int]_{total.i}} \times n_i[Int_P]\right) + \sum_{i=1}^{n}\left(\frac{n[Int_N]_{total.i}}{n[Int]_{total.i}} \times n_i[Int_n]\right)) + 2} \quad (13)$$

where, $n[Int_P]_{total.i}$ denotes the total number of positive interactions upto ith interaction. $n[Int_N]_{total.i}$ denotes the total number of negative interactions upto ith interaction. $n[Int]_{total.i}$ denotes the total number of interactions upto ith interaction.

B. Probability in Indirect Interactions

TR's 1-hop distant nodes (except TE) having the evaluated final probability knowledge about TE considered as recommenders ($R=\{R_1,R_2,....R_n\}$). Ri (ith recommender) sends the evaluated final probability about TE to TR as recommendation probability (${}_{Ri}^{c}P_{TE}^{t_0}$, t_0 is the computing time instant). Allowance of giving recommendation probability by every recommender follows exponential distribution as in "(14)".

$$f(t,\alpha) = \alpha e^{-\alpha t} \quad , for \;\; t \geq 0 \;\; and \; \alpha > 0 \quad (14)$$

Time $t = t_0, t_1, t_2, ... , t_{cur}$. when $t = t_{cur}$, t is considered as current time instant and when $t = t_0$, t is considered as computing time instant. α is the rate parameter of the distribution.

Using "(14)", the allowed recommendation probability distribution curve over time follows f{t} as shown in "(15)".

$$f\{t\} = f_R\{ \left({}_{Ri}^{c}P_{TE}^{t_0} \times \alpha e^{-\alpha t_0}, t_0\right), \left({}_{Ri}^{c}P_{TE}^{t_0} \times \alpha e^{-\alpha t_1}, t_1\right) , \dots\dots\dots , \left({}_{Ri}^{c}P_{TE}^{t_0} \times \alpha e^{-\alpha t_{cur}}, t_{cur}\right) \} \quad (15)$$

Here, Ri is the ith recommender and ${}_{Ri}^{c}P_{TE}^{t_0}$ is the recommendation probability from Ri. Since nature of the curve is decaying over time, we are considering the lastly evaluated final probability of Ri as recommendation probability.

Allowed recommendation probability by Ri at current time instant is shown in "(16)"

$${}_{Ri}^{c}AlRP_{TE}^{t_{cur}} = {}_{Ri}^{c}P_{TE}^{t_0} \times \alpha e^{-\alpha t_{cur}}, for \; t \geq 0 \; and \; \alpha > 0 \quad (16)$$

Every ith recommender sends their ${}_{Ri}^{c}AlRP_{TE}^{t_{cur}}$ to TR as recommendation. After getting allowed recommendation from different recommenders, TR considers every Ri's ${}_{Ri}^{c}AlRP_{TE}^{t_{cur}}$ as accepted probability (${}_{Ri}^{c}AcPi_{TE}^{t}$) with respect to its acceptance rate according to "(17)".

$${}_{Ri}^{c}AcRP_{TE}^{t} = {}_{Ri}^{c}AlRP_{TE}^{t_{cur}} \times acceptance\ rate \quad (17)$$

where, acceptance rate on i[th] allowed recommendation probability is evaluated as per "(18)".

$$Acceptance\ rate = acceptance\ level \times \alpha e^{-\alpha t} \quad (18)$$

Acceptance level of TR on Ri is the lastly evaluated final probability by TR for Ri (${}_{TR}^{c}Pi_{Ri}^{t_0}$). Acceptance level of any recommender follows similar exponential distribution as in "(14)". Cumulative accepted recommendation probability (${}_{R}^{c}CAcRP_{TE}^{t}$) for TE is evaluated by TR as per "(19)".

$${}_{R}^{c}CAcRP_{TE}^{t} = \frac{1}{n}\sum_{i=1}^{n} {}_{Ri}^{c}AcRP_{TE}^{t} \quad (19)$$

Final probability on indirect monitoring (${}_{R}^{c}FPI_{TE}^{t}$) is evaluated by TR for TE as per "(20)".

$${}_{R}^{c}FPI_{TE}^{t} = \sum_{i=1}^{n} W_i \times {}_{R}^{c}IP_{class:i}\,{}_{TE}^{t} \quad (20)$$

where class:i denotes the ith class. i=1,2.... n. Here n=2. class:1= positive and class:2=negative. $W_2 = (1-W_1)$. W_1 and W_2 are the corresponding weight.

IV. Simulation

Our simulation result shows the probability of legitimacy from direct monitoring with consideration of positive and negative type of interaction. In Fig.1, we have shown the probabilities of legitimacy where TE performs continuous positive and continuous negative direct interactions with TR. Fig.2 shows the probability of legitimacy with a situation where a TE start to interact negatively after 5th continuous positive direct interactions with TR. Similarly, Fig.3 shows the probability of legitimacy where a TE start positive interaction after 5th continuous negative direct interactions with TR.

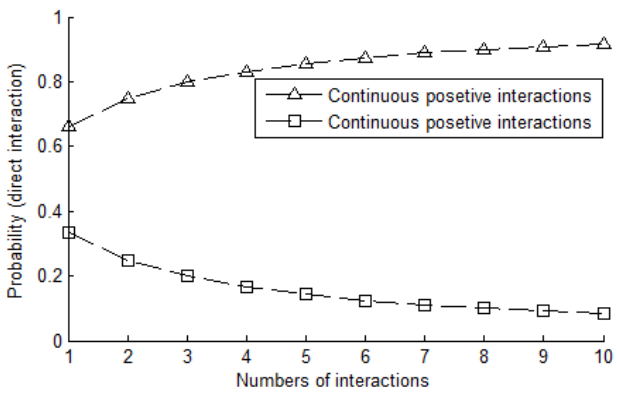


Fig. 1. Simulation Results

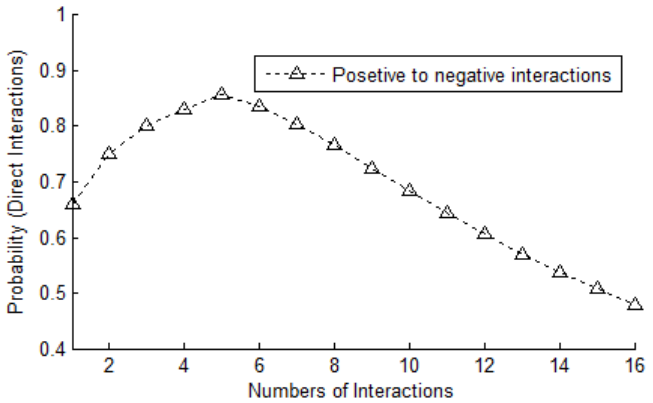


Fig. 2. Simulation Result

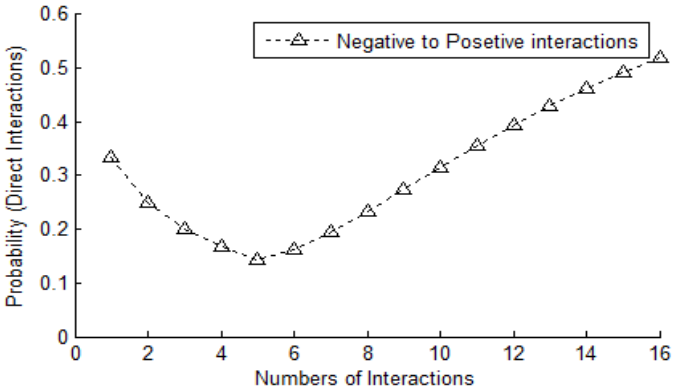


Fig. 3. Simulation Results

V. CONCLUSION

Our proposed Probabilistic Trust Management (PTM) in wireless communication system efficiently defends against malicious behaviour of compromised wireless entity or any other malevolent entity. Proposed PTM isolates malicious behaviour from system by isolating compromised or any other malevolent wireless entity from communication. Our PTM efficiently isolates malicious behaviour depending on direct indirect interactions with their timing constraint and the consideration of legitimacy level of interactions. Our Simulation results have shown the efficient performance of proposed PTM in wireless communication system.

REFERENCES

[1] Daoxi Xiu and Zhaoyu Liu, "A Formal Definition for Trust in Distributed System", SpingerLink, Lecture Notes in Computer Science,Volume 3650/2005, 482-489, 2005.

[2] Dorri Nogoorani, S. , Sharif Univ, Tehran, Iran ; Jalili, R. "Uncertainty in Probabilistic Trust Models", 26th International Conference on Advanced Information Networking and Applications (AINA), 2012 IEEE .

[3] George Vogiatzis, Ian MacGillivray and Maria Chli "A Probabilistic Model for Trust and Reputation", International Foundation for Autonomous Agents and Multiagent Systems , 9th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2010) , May, 10–14, 2010, Toronto, Canada, page:225-232.

[4] G.H. Nguyen , P. Chatalic , M.C. Rousset "A Probabilistic Trust Model for Semantic Peer to Peer Systems", DAMAP'08, March 25th, 2008, Nantes, France. Copyright 2008 ACM 978-1-59593-697-8.

[5] Haimei Xu , Chongqing Commun. Inst., Chongqing, China , Yulin Liu , Shouqing Qi and Yanjun Shi, "A novel trust model based on probability and statistics for Peer to Peer networks", IEEE, International conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering (QR2MSE), 978-1-4799-1014-4 , page: 2047 – 2050, 2013.

[6] Hyun-Kyo Oh, Jin-Woo Kim, Sang-Wook Kim and Kichun Lee "A Probability-Based Trust Prediction Model Using Trust-Message Passing", WWW 2013 Companion, May 13–17, 2013, ACM 978-1-4503-2038-2/13/05.

[7] Yan Lindsay Sun , Wei Yu , Zhu Han , K. J. Ray Liu ,"Information Theoretic Framework of Trust Modeling and Evaluation for Ad Hoc Networks", IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, VOL. 24, NO. 2, FEBRUARY 2006.

Three-order Microstrip Bandpass Filter with Two Transmission Zeros for WLAN/WiMAX Applications

Biswajit Pal,[1,*] and Santanu Dwari[1]

[1] Electronics Engineering, Indian School of Mines
Dhanbad, Jharkhand- 826004, India
[*]biswajit.rpe@gmail.com

Abstract—**A simple novel compact three-order microstrip bandpass filter is proposed using half wavelength microstrip line resonators coupled through parallel stub lines. The filter has two signal paths, one through the parallel coupled lines of resonators and other through the electrical coupling between feed lines for which the filter shows two transmission zeros. The proposed filter has a bandwidth of 1GHz covering WLAN 5.2/5.8 GHz and WiMAX 5.5 GHz bands. The filter is compact and simple structure having very low insertion loss in the passband makes it suitable for wireless application.**

Index Terms— **Bandpass filter, Transmission Zero, Wireless Local Area Network (WLAN) and Worldwide Interoperability for Microwave Access (WiMAX).**

I. Introduction

In modern communication systems planner microstrip bandpass filter (BPF) are widely studied and used because of their low cost, simple structure and easy integration. Different types of compact bandpass filter have been proposed [1-11]. Dual behaviour resonators based on stepped impedance and different-length stub is used for dual transmission zero, given in [1]. In [2], radial-line stubs are used for multimode BPF design. In [3] additional transmission zeros are generated using double split-end quarter-wave-length resonator. It is possible to increase transmission pole with the help of square-loop resonator [4]. The novel compact stub coupled microstrip square open loop BPF is proposed for single pole with dual transmission zeros [5]. Open stub inverter between quarter wave length resonator [6], half wavelength microstrip line resonator and end coupling resonator [7], a half-wavelength transmission line resonator with a pair of tap-connected open-ended stubs [8], dual-mode stub-loaded resonators [9], are used for multiple pole BPF with double transmission zeros. Two order BPF with three finite zero point is achieved in [10].

In this paper, a novel compact three-pole microstrip fed bandpass filter with two transmission zeros is proposed. The proposed BPF uses three parallel resonators coupled by stub lines. Three resonators are used to implement three poles in passband which improves the filter characteristics. The dual transmission zeros are generated by two different signal paths one via parallel resonators and other through the direct signal path by electrical coupling in feed lines. The proposed microstrip band pass filter is covering 10 dB bandwidth of 1.02GHz, starting from 4.87 to 5.89 GHz. The filter can be used for WLAN 5.2/5.8 GHz and WiMAX 5.5 GHz applications. The operating bandwidth covers WLAN bands of 5.15 – 5.35 GHz and 5.725 – 5.825 GHz (IEEE 802.11a) and WiMAX band of 5.25 – 5.85 GHz (IEEE 802.16), have been exempted from licensing. The WLAN 5.2 GHz band is restricted to indoor operation with a maximum EIRP (Effective Isotropic Radiated Power) of 200 mW and the WLAN 5.8 GHz band could be used for both indoor and outdoor applications with EIRP up to 4 watt. Recently, the World Radio communication Conference 2003 has approved a new co-primary allocation in the 5.47 – 5.725 GHz band for mobile service for the implementation of wireless access systems. It is also possible for WiMAX systems to operate in the 5.5GHz band on an uncoordinated and unprotected basis.

CST microwave studio is used for all the simulation and optimization purposes. The proposed filter is designed on a $25\times28mm^2$ low-cost FR4 substrate with thickness of 1.59 mm, dielectric constant of 4.4 and the loss tangent is 0.02. The filter modelling and its results are described in section II and the paper is concluded in Section III.

II. Filter Modelling and Results

The microstrip filter loaded with a single $\lambda/2$ resonator is shown in figure 1. The signal is coupled to the resonator through the parallel stubs from the feed line. The signal is also directly coupled through electrical coupling via gap g_3 in the feed line. The simulated s-parameters are given in figure 2. It is a single pole bandpass filter with dual transmission zero. The pole of the proposed filter can be increased by increasing the no of resonators. The optimized parameters are W_1= 1 mm, W_2=0.8 mm, W_3=2.9 mm, g_1=0.8 mm g_2=0.5 mm, g_3=1.6 mm, L_1=14.5 mm, L_2=8.3 mm.

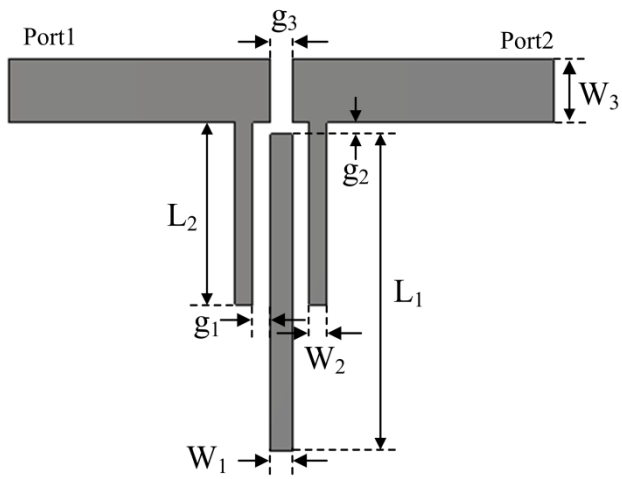


Fig. 1 Single pole bandpass filter.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

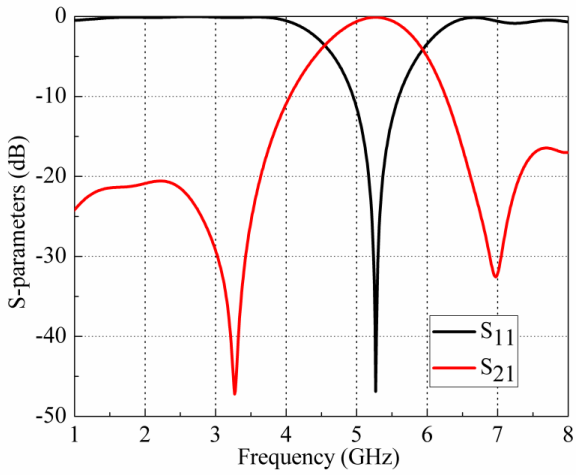


Fig. 2 S-parameters of single pole bandpass filter.

A dual-pole bandpass filter is achieved by using two $\lambda/2$ resonators in place of one used in single pole bandpass filter. It improves the filter bandwidth with sharp out of band rejection. The dual-pole bandpass filter is shown in figure 3 and its s-parameters is given in figure 4. The insertion loss is near to 0 dB level and return loss below -20dB level in the passband. Two transmission zeros are realised at 3.32 GHz and 6.83 GHz. The filter covers the 10 dB bandwidth from 5.03 to 6 GHz. The optimized parameters are W_1= 1 mm, W_2=0.8 mm, W_3=2.9 mm, g_1=0.8 mm g_2=0.5 mm, g_3=1.6 mm, g_4=1.4 mm, L_1=13.8 mm, L_2=8 mm.

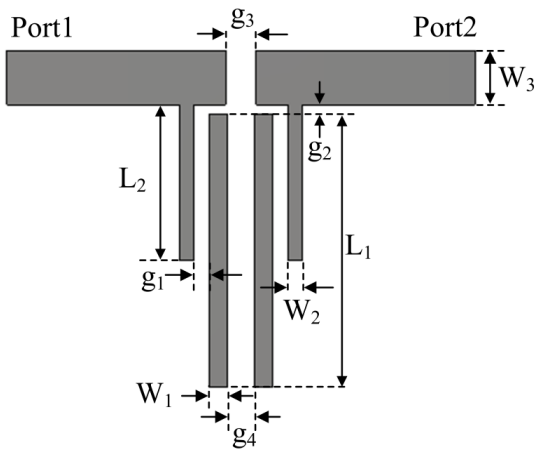


Fig. 3 Dual-pole bandpass filter.

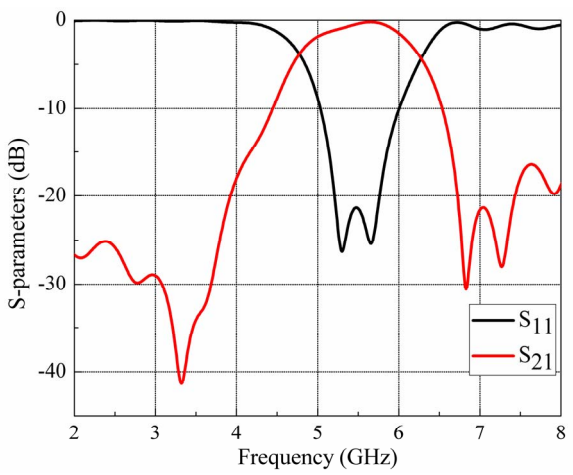


Fig. 4 S-parameters of dual-pole bandpass filter.

The filter performance can still be improved by increasing the no of $\lambda/2$ resonators. A three-pole bandpass filter is achieved by using three $\lambda/2$ resonators in the structure, shown in figure 5. The insertion loss is near to 0 dB level and return loss below -20dB level in the passband. Two transmission zeros are realised at 3.49 GHz and 6.3 GHz. The filter covers the 10 dB bandwidth from 4.87 to 5.89 GHz. So, the proposed triple-pole bandpass filter shows that the transmission zeros come closer to the passband, thus improves the sharp out of band rejection. A comparative study of dual-pole and triple-pole bandpass filter is shown in figure 6. The final optimized parameters for the proposed filter are W_1= 1 mm, W_2=0.8 mm, W_3=2.9 mm, W_4=1.3 mm, g_1=0.8 mm g_2=0.5 mm, g_3=1.6 mm, g_4=1.4 mm, L_1=13.8 mm, L_2=8 mm.

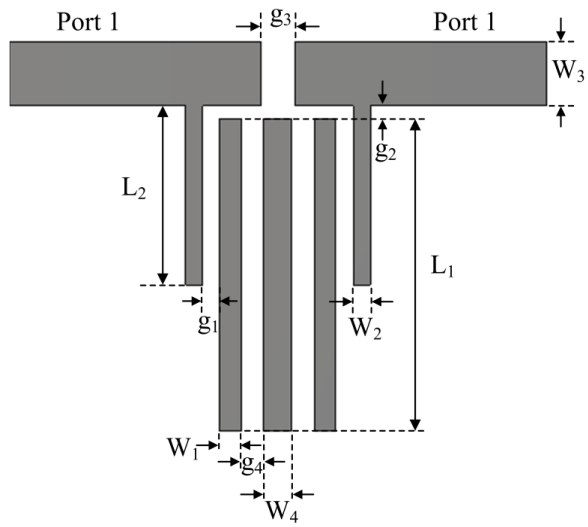


Fig. 5 Proposed three-pole bandpass filter.

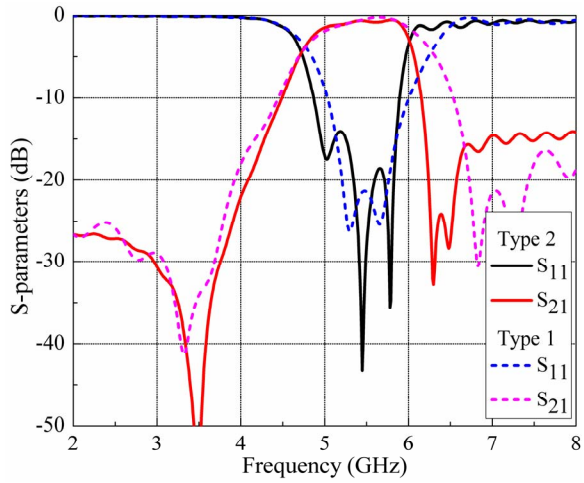


Fig. 6 Comparison between dual-pole (Type 1) and proposed triple-pole (Type 2) bandpass filter.

III. Conclusions

A new compact low cost single band bandpass filter with dual transmission zero is introduced here. The transmission zeros are controllable by the filter poles. The filter poles are increased by increasing the no of $\lambda/2$ resonators. The proposed simple structure is suitable for the use in WLAN and WiMAX applications.

REFERENCES

[1] C. Quendo, E. Rius and C. Person, "Narrow Bandpass Filters Using Dual-Behavior Resonators Based on Stepped-Impedance Stubs and Different-Length Stubs" *IEEE transactions on microwave theory and techniques*, vol. 52, no. 3, pp. 1034-1044, 2004.

[2] L. Zhang, Z. Y. Yu and S. G. Mo, "Novel Planar Multimode Bandpass Filters", *Progress In Electromagnetic Research*, PIER 101, pp 33-42, 2010.

[3] K. U. yen, E. J.Wollack, T. A. Doiron, J. Papapolymerou and Joy Laskar, "A Planar Bandpass Filter Design with Wide Stopband Using Double Split-End Stepped-Impedance Resonators", *IEEE transactions on microwave theory and techniques*, vol. 54, no. 3, pp. 1237-1244, 2006.

[4] K. Yeo, M. J. Lancaster, J. Hong, "The Design of Microstrip Six-Pole Quasi-Elliptic Filter with Linear Phase Response Using Extracted-Pole Technique", *IEEE transactions on microwave theory and techniques*, vol. 49, no. 2, pp. 321-327, 2001.

[5] R. K. Maharjan, B. Shrestha and N.Y. Kim, "Compact Microstrip square open loop bandpass filter using open stub," *IEEE microwave wireless Component letters*, vol 48, no 6, 2012.

[6] J. Lee, J. Cho and S. Yun, "New Compact Bandpass Filter Using Microstrip Resonators with Open Stub Inverter" *IEEE microwave and guided wave Letters,* vol. 10, no. 12, pp 526-527, 2000.

[7] C. Jianxin, Y. Mengxia, X. Jun and X. Quan, "Compact microstrip bandpass filter with two transmission zeros," *IEEE microwave wireless component letters,* vol. 40, no. 5, 2004.

[8] L. Zhu and W. Menzel, "compact Microstrip Band pass Filter with Two Transmission Zeros using a stub-Tappered Half-Wavelength Line Resonator", *IEEE microwave and wireless components letters*, vol. 13, no. 1, pp. 16-18, 2003.

[9] A. Torabi, K. Forooraghi, O. Manoochehri and S. Abbasiniazare, "Compact Microstrip Bandpass Filters Using Dual-Mode Stub loaded Resonators and Capacitive/ Inductive Source-Load Coupling", *Microwave and optical technology letters*, vol. 53, no. 9, pp. 2111-2116, 2011.

[10] K. Ma, J. Ma, M. Do and K. Yeo, "Compact two-order bandpass filter with three finite zero points" *IEEE microwave wireless component letters*, vol. 41, no.15, 2005.

[11] J. S. Hong and M. J. Lancaster, Microstrip Filter for RF/Microwave Application. New York: Wiley, 2001.

Optimal Scheduling in MU-MIMO Uplink Cellular Networks

Md. Zobaarul Haque, Mirza Tahmid Islam and Md. Forkan Uddin*
Dept. of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
*mforkanuddin@eee.buet.ac.bd

***Abstract*—In this paper, the optimal scheduling problem is studied in multi-user MIMO (MU-MIMO) single-hop uplink cellular networks. An optimization framework is developed to compute the optimal max-min throughput and the optimal scheduling of the users. The optimization problem is found to be a very complex linear program. The optimization problem is solved for several medium size cellular networks using an optimization tool. The numerical results show that MU-MIMO technique provides a significant throughput gain over SU-MIMO technique. To avoid the complexity of computation and implementation, a user grouping based simple scheduling algorithm is also proposed for MU-MIMO cellular networks. The performance of the proposed algorithm is evaluated and found to be very efficient.**

***Index Terms*—Multi-user MIMO, scheduling, throughput, optimization.**

I. Introduction

Multiple-input and multiple-output (MIMO) communication technique has made the most significant breakthroughs in terms of channel capacity and reliability in modern wireless communication [1]. Multiple antennas at transmitter and receiver can increase the channel capacity by transmitting independent data streams over multiple spatial channels from transmitter to receiver [2]. If the number of transmitting antennas is M and the number of receiving antennas is N, it is possible to achieve $min(M, N)$-time capacity of single transmitting and receiving antenna configuration [3]. Due to the advantages of the MIMO communication, it has been adopted in the current standards of cellular and other wireless networks.

Throughput is an important performance metric in wireless cellular networks to support voice, video, and data services. The throughput performance in a MIMO cellular network significantly depends on the scheduling algorithm and MIMO technique. MIMO technique can be classified as: single user MIMO (SU-MIMO) and multi-user MIMO (MU-MIMO) and further, MU-MIMO technique can be classified as MU-MIMO with single antenna transmission (MU-MIMO-SA) and MU-MIMO with combined single and multi-antenna transmission (MU-MIMO-SA-MA) [4]. In [5], it is shown that the MU-MIMO-SA scheme provides better throughput than the SU-MIMO scheme in random access based wireless local area networks. In [4], joint routing and scheduling problem is studied in uplink wireless mesh networks and it is found that (i) the MU-MIMO techniques provide significantly higher max-min throughput than the SU-MIMO technique and (ii) the MU-MIMO-SA-MA scheme provides better max-min throughput than the MU-MIMO-SA scheme. The scheduling problem of MU-MIMO cellular networks has been studied in [6]- [9] with various aspects. In [6], a fair scheduling algorithm is proposed for uplink MIMO network. An interference-aware optimal user selection scheme is proposed in [7] for MIMO system in multicell environment. In [8], an scheduling algorithm is proposed for heterogeneous users in MIMO downlink network. In [9], users are scheduled in pairs to achieve the optimal throughput where the resource will be underutilized if the number of base station (BS) antennas are more than two. In the available studies on MU-MIMO scheduling in wireless cellular networks, there is no clear indication on which MIMO scheme is better for single hop cellular networks and on how to perform optimal scheduling.

Usually, throughput-optimal scheduling problem of any wireless network is a very complex linear program (LP) [10]. The complexity of the problem increases significantly with the number of nodes in the network. The optimal solution of the scheduling problem may be achievable for a medium size network by using optimization tools and different optimization techniques, e.g., column generation technique [10]. However, it is very difficult to implement these approaches in practice. Low-complexity scheduling algorithm is necessary for practical solution.

In this paper, single cell uplink MIMO cellular networks with multiple transmission rates is considered to study the optimal scheduling problem. The contributions in this paper are as follows:

- An optimization framework is developed to compute the optimal max-min throughput and the optimal scheduling of the users of a network with the SU-MIMO, MU-MIMO-SA and MU-MIMO-SA-MA schemes.
- The optimization problem is solved numerically using the optimization tool MINOS [11] for several medium size networks under the three different MIMO schemes and the throughput performance of the MIMO schemes is compared.
- Due to the limitation of installing a powerful computing device at a BS, it is difficult to implement the tool dependent based solution approach of the optimization problem in practice. For practical implementation, a user grouping based simple scheduling algorithm is proposed for MU-MIMO cellular networks. The performance of the proposed algorithm is evaluated and found to be very efficient.

The rest of the paper is organized as follows. The system model is presented in detail in Section II. The optimization problem of optimal scheduling is formulated in Section III. In section IV, the numerical results are presented and the various engineering insights are provided. The grouping based simple scheduling algorithm is described and analyzed in Section V. The paper is concluded in Section VI.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

II. System Model

A. Network Topology

A single-hop cellular network consisting of U stationary users and a BS is considered. The set of users $\mathcal{U}$ is given as $\mathcal{U} = \{1, 2, ..., U\}$. The transmission power of each user is P_t. The number of antennas at the BS and at each user is assumed to be N. The antennas are omni-directional and cannot transmit and receive simultaneously. Only uplink communication is considered in the network, i.e., only the users transmit data to the BS.

B. MIMO schemes

1) SU-MIMO: In this scheme, only one user communicates with the BS at a time through its N antennas. The transmitter first encodes the input bits with one encoding block, generates N data streams and transmits through N antennas to the BS. The transmission power at each antenna for each data stream is P_t/N.

2) MU-MIMO-SA: In this scheme, a user transmits using only one of its antennas. So the BS can successfully retrieve data streams from N users at a time. The transmission power at the transmitting antenna is P_t.

3) MU-MIMO-SA-MA: In this scheme, a user can transmit through one or more of its antennas [4]. For a transmission, in this scheme, a transmitter encodes the input bits with one coding block, generates N' data streams, where $1 \leq N' \leq N$ is the number of antennas to be used in transmission, and transmits in parallel with the N' antennas. For this scheme, the transmission power at each antenna of the transmitter is P_t/N'.

C. Physical Layer Model

The number of simultaneous data stream transmission is assumed to be less than or equal to the number of antennas in the BS. Due to its simplicity, zero forcing (ZF) MIMO decoding technique is considered at the BS. Hence, the success or failure of a data stream transmission depends on only to the signal to noise ratio (SNR) of the data stream, not on the signals of the other data streams and the SNR must not be less than a threshold for successful reception [4], [5]. The SNR threshold depends on the modulation and coding scheme, i.e., data rate of the data stream. Let β be the SNR threshold for a data stream i. Thus, the data stream i is successful if [4], [5]

$$\frac{g_i E\left[|s_i|^2\right]}{N_0} \geq \beta \tag{1}$$

where, g_i is the channel gain between the user of data stream i and the BS, $E[|s_i|^2$ is the transmitted power of the received data signal s_i of the data stream i and N_0 is the background noise power. The channel gain between the user of data stream i and the BS is given by $g_i = \left(\frac{d_i}{d_0}\right)^{-\eta}$, where, d_i is the distance between the user of data stream i and the BS, η is the path loss exponent, and d_0 is a reference distance in the far field of the antenna.

D. Rate Selection for a Data Stream

It is assumed that a user can transmit with R different rates. The set of the rates $\mathcal{R}$ is given by $\mathcal{R} = \{r_1, r_2, ..., r_R\}$. The set of the required SNR's for the available data rates is given by $\mathcal{B} = \{\beta_1, \beta_2, ..., \beta_R\}$, where β_j is the SNR threshold for the rate r_j. It is assumed that a user selects the highest possible rate for a data stream based on the SNR of the data stream at the BS, i.e., the rate of a data stream i is r_j if

$$\beta_j \leq \frac{g_i E\left[|s_i|^2\right]}{N_0} \leq \beta_{j+1}. \tag{2}$$

The value of $E\left[|s_i|^2\right]$ in (1) and (2) depends on the MIMO scheme used in the system. It is equal to P_t/N and P_t for the SU-MIMO and MU-MIMO-SA schemes, respectively. For the SU-MIMO-SA-MA scheme, it depends on the number of antennas to be used for transmission.

E. Scheduling

The set of users that can transmit simultaneously without any collision is defined as an independent set (ISET) and denoted by v. For the SU-MIMO scheme, an ISET consists of only one user and there are U ISETs in the cellular network. For the MU-MIMO-SA scheme, a set of users $v \subset \mathcal{U}$ is an ISET if $|v| \leq N$. For the MU-MIMO-SA-MA scheme, a set of users $v \subset \mathcal{U}$ is an ISET if $\sum_{u \in v} {M'}_u^v \leq N$, where ${M'}_u^v$ is the number of data streams to be transmitted by the user u in ISET v. For scheduling, the users of an ISET will be allowed to transmit for a given time duration in a regular interval.

III. Problem Formulation

In this section, the optimization problem is formulated for the SU-MIMO and MU-MIMO schemes.

A. Optimal Scheduling Problem

Let the set of all the ISETs for the network is $\mathcal{I}$ and α_v be the fraction of time that the ISET v will be scheduled by maintaining $\sum_{v \in \mathcal{I}} \alpha_v = 1$. Denote the vector of α_v's by $\boldsymbol{\alpha}$. Let $\mathcal{I}_u \subset I$ be the set of ISETs containing user u. Let the throughput of user u is λ_u and the vector of λ_u's is $\boldsymbol{\lambda}$. The throughput of a user u can be expressed as $\lambda_u = \sum_{v \in \mathcal{I}_u} C_{u,v} \alpha_v$, where $C_{u,v}$ is the selected rate by the user u for ISET v. The value of $C_{u,v}$ can be calculated as $C_{u,v} = {M'}_u^v R_{u,v}$, where $R_{u,v}$ is the data rate of each stream of user u which can be calculated from (2).

Let λ be the minimum throughput of the users for the cellular network. The optimization problem to maximize λ with any MIMO scheme can be expressed as

$$\max_{\boldsymbol{\alpha}, \boldsymbol{\lambda}} \quad \lambda \tag{3}$$

$$\lambda \leq \lambda_u \quad \forall u \in \mathcal{U} \tag{4}$$

$$\lambda_u = \sum_{v \in \mathcal{I}_u} C_{u,v} \alpha_v \quad \forall u \in \mathcal{U} \tag{5}$$

$$\sum_{v \in \mathcal{I}} \alpha_v = 1 \tag{6}$$

$$0 \leq \boldsymbol{\alpha} \leq 1 \quad ; \quad 0 \leq \boldsymbol{\lambda} \tag{7}$$

The objective function in (3) and the constraints in (4) ensures that the minimum throughput of the users is maximized. The constraints in (5) represent the rates of the users. The constraint in (6) relate the fraction of times of the different ISETs and it states that ISETs must be scheduled over a unit period of time. The constraints in (7) define the range of the variables.

The problem (3)-(7) is a LP and the complexity of the problem mainly depends on the number of ISETs in the network. The number of ISETs in a network depends on the number of users and MIMO scheme of the network. For the SU-MIMO

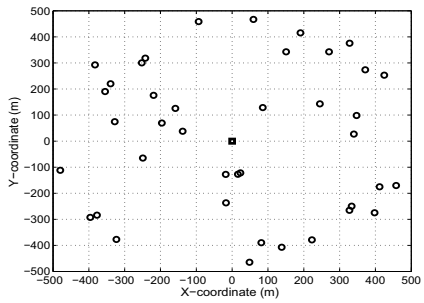


Fig. 1. The positions of the nodes in Network-1.

TABLE I
DATA RATES AND THE CORRESPONDING SNR THRESHOLDS.

Modulation	Coding rate	Data Rate	SNR Threshold (dB)
BPSK	1/2	1	6.4
QPSK	1/2	2	9.4
	3/4	3	11.2
16-QAM	1/2	4	16.4
	3/4	6	18.2

scheme, the number of ISETs is only equal to the number of users and the optimal solution can be obtained easily with some calculations (described later) without any optimization tool. On the other hand, the number of ISETs for the MU-MIMO-SA and MU-MIMO-SA-MA schemes are $\sum_{n=1}^{N}\binom{U}{n}$ and $\sum_{n=1}^{N}\binom{U}{n}\binom{N-1}{n-1}$, respectively. Thus, the complexity of the optimization problem is $O(U^N)$ for the both MU-MIMO schemes.

B. Solution of the Problem for SU-MIMO Scheme

Let j and k are two different users in the network. In the SU-MIMO scheme, the number of ISETs for a user is only one and the ISET contains only the user. Let the fraction of times to be scheduled by users j and k are α_j and α_k, respectively and the data transmission rates of users j and k are C_j and C_k, respectively. Since the minimum throughput need to be maximized, it is assumed that $\lambda^* = \lambda_j = \lambda_k$. Thus, $\alpha_j^* C_j = \alpha_k^* C_k$. Since $\sum_{i\in\mathcal{U}} \alpha_i^* = 1$, it can be shown that $\alpha_k^* = \frac{1}{1+C_k \sum_{i\in\mathcal{U}\setminus k} \frac{1}{C_i}}$. Then the max-min throughput of each user can be calculated as $\lambda^* = C_k \alpha_k^*$.

IV. NUMERICAL RESULTS

A. Networks and Parameters

Twenty networks are considered each consisting of 40 users and a BS. The BS is located at the centre surrounded by the users. The positions of the users are determined by a uniform random distribution. One of the networks (Network-1) is shown in Fig. 1 where the BS is labeled as a box. It is assumed that a user can use five different modulation and coding schemes. The normalized data rates and the SNR thresholds for the different modulation and coding schemes are taken from [10] and are illustrated in Table I. The number of antennas at the BS and at each user is assumed to be 4. The noise power (N_0), path loss exponent (η) and reference distance (d_0) are considered as -100 dBm, 3 and 1m, respectively.

B. Advantages of MU-MIMO Schemes

For the Network-1, the optimal max-min throughput is computed by solving the optimization problem (3)-(7) using the optimization tool MINOS [11] for different transmit power levels under the different MIMO schemes. The max-min throughputs and the throughput gain of the MU-MIMO schemes with respect to the SU-MIMO scheme for Network-1 are shown in Fig. 2 for different transmission power levels. The results show that the MU-MIMO-SA and MU-MIMO-SA-MA schemes provide the same max-min throughput for any transmission power. The same result is found for the other 19 networks. Thus, the MU-MIMO-SA scheme is preferable in cellular networks to avoid the complexity of the MU-MIMO-SA-MA scheme. It is observed that the MU-MIMO schemes improves max-min throughput significantly at a low transmission power. For Network-1, the MU-MIMO schemes provide upto 138.3% throughput gain over the SU-MIMO scheme. One of the important advantages of the MU-MIMO schemes is that the network becomes connected approximately 6 dB lower transmission power than the SU-MIMO scheme.

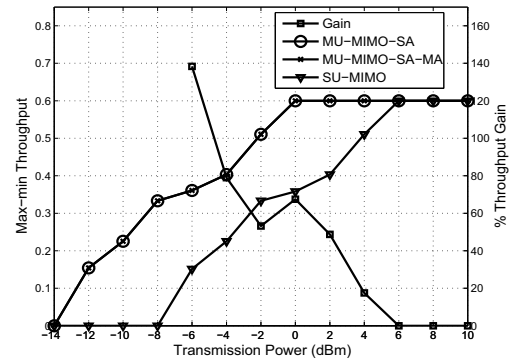


Fig. 2. Throughput comparison of the SU-MIMO, MU-MIMO-SA and MU-MIMO-SA-MA schemes for Network-1.

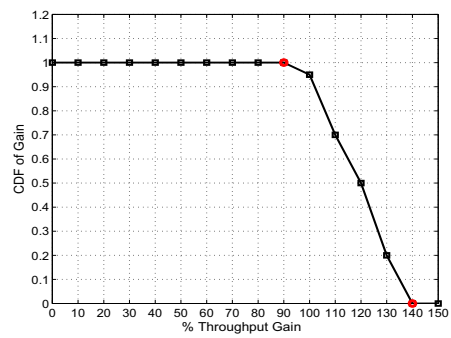


Fig. 3. Cumulative distribution function of gains of the MU- MIMO scheme over SU-MIMO scheme.

The optimal max-min throughputs for the other 19 networks are also computed at $P_t = -6$ dBm and the throughput gain of the MU-MIMO schemes over the SU-MIMO scheme are determined. The cumulative distribution function (CDF) of throughput gain for the 20 networks at $P_t = -6$ dBm is shown in Fig. 3. The results show that in all the networks the throughput gain of the MU-MIMO schemes is more than 90% and less than 140%. The average throughput gain is found to be 118.63%. Thus, MU-MIMO technique provide a significant benefit over SU-MIMO technique in term of max-min throughput.

V. Grouping Based Scheduling

It is difficult to apply the optimization tool based optimal scheduling approach in practice due to the limitation of installing a powerful computing device at a BS. However, the optimization tool is not able to provide a solution for a large network. In this section, a user grouping based simple scheduling algorithm is proposed and the performance of the algorithm is evaluated.

A. Rate Based User Grouping

The number of groups of the users is assumed to be R with R different data rates. The distances of the users and their data rates are assumed to be known to the BS. The sets of users having data rates $r_1, r_2, \ldots$, and r_R are denoted by $\mathcal{U}_1, \mathcal{U}_2, \ldots$, and $\mathcal{U}_R$, respectively. The number of users in group i is given by U_i and $U_i = |\mathcal{U}_i|$.

B. Grouping Based Scheduling and Throughput

It is assumed that at a time only the users of a group are scheduled. An ISET of group i is denoted by v^i. It is also assumed that an ISET contains N users, i.e., $|v^i| = N$. Thus, the number of ISETs in group i is $\binom{U_i}{N}$. Let σ_{v^i} be the fraction of time that the ISET v^i to be scheduled. The throughputs of the users of a group are assumed to be equal and hence, the fraction of times for all the ISETs of a group are equal, i.e., $\sigma_{v_1^i} = \sigma_{v_2^i}$. In the following, notation σ_i is used instead of σ_{v^i} for simplicity of presentation. Since all the ISETs must be scheduled within a unit period of time,

$$\sum_{i=1}^{R} \binom{U_i}{N} \sigma_i = 1. \tag{8}$$

The throughput of a user of group i is denoted by λ_i^s. The number of ISETs that contain the user $u \in \mathcal{U}_i$ of group i is $\binom{U_i-1}{N-1}$. Thus, the throughput of each of the users of group i is given as

$$\lambda_i^s = \binom{U_i - 1}{N - 1} \sigma_i C^i \tag{9}$$

where, $C^i = r_i$ is the data transmission rate of a user of group i. Thus, the objective of the scheduling algorithm is to determine all the values of λ_i^s's and σ_i's.

C. Scheduling Algorithm

Since the minimum throughput of the users need to be maximized, the throughputs of all the users are assumed to be equal, i.e.,

$$\lambda_i^s = \lambda_j^s. \tag{10}$$

for two groups i and j. Using (9) in (10), it can be obtained that

$$\binom{U_i - 1}{N - 1} \sigma_i r_i = \binom{U_j - 1}{N - 1} \sigma_j r_j \tag{11}$$

By replacing all the σ_j's in (8) in term of σ_i using (11), it can be shown that

$$\sigma_i^* = \frac{N}{\binom{U_i-1}{N-1}\{U_i + r_i \sum_{j=1, j\neq i}^{R} \frac{U_j}{r_j}\}}. \tag{12}$$

Using (12) in (9), the max-min throughput of the users can be calculated. Thus, scheduler can calculate the scheduling parameters and the max-min throughput of the users by knowing only the number of users in different groups.

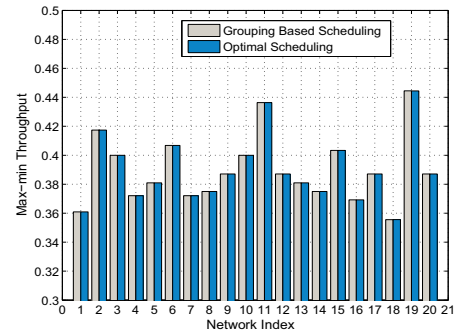


Fig. 4. Throughput comparison of grouping based scheduling algorithm and optimal scheduling.

D. Performance Evaluation of the Algorithm

To evaluate the performance of the simple scheduling algorithm, the max-min throughputs of the 20 networks are determined using the algorithm. The throughput obtained by the simple algorithm and the optimal throughput are shown in Fig. 4. The results show that the simple algorithm can achieve the optimal throughput for all the 20 networks. Thus, the proposed simple algorithm is very efficient.

VI. Conclusion

In this paper, the optimal scheduling problem in single-hop uplink MIMO cellular networks is studied under different MIMO schemes. By formulating an optimization problem and solving the problem for several medium size cellular networks, it is demonstrated that MU-MIMO technique provides a significant throughput gain over SU-MIMO technique in cellular networks. For practical implementation, a very simple and efficient optimal scheduling algorithm is also developed for MU-MIMO cellular networks.

References

[1] H. Blcskei, D. Gesbert, C. Papadias, A. J. Van Der Veen, and A. M. Cipriano, "Space-time wireless systems: from array processing to MIMO communications," Cambridge Univ. Press, 2006.

[2] T. Kailath and A. Paulraj, "Increasing capacity in wireless broadcast systems using distributed transmission/directional reception," US Patent 5345599, Sept. 1994.

[3] K. Raoof, M. Khalighi, and N. Prayongpun, "MIMO systems: Principles, iterative techniques and advanced polarization," in *Adaptive Signal Processing for Wireless Communication*, CRC Press, vol. 2, pp. 95-134, Aug. 2008.

[4] M. F. Uddin, "Joint routing and scheduling in fixed multi-user mimo wireless multi-hop networks," in *Proc. IEEE/CIC ICCC*, pp. 745-751, 2013.

[5] H. Jin, B. C. Jung, and D. K. Sung, "A tradeoff between single-user and multi-user MIMO schemes in multi-rate uplink WLANs," *IEEE Trans. on Wireless Commun.*, vol. 10, no. 10, pp. 3332-3342, 2011.

[6] P. Lee and J. H. Lee, "A fair scheduling algorithm for uplink multiuser MIMO-OFDM system with MMSE-SIC detection," in *Proc. IEEE APWCS*, Aug. 2008.

[7] B. O. Lee, O.-S. Shin, and K. B. Lee, "Distributed user selection scheme for uplink multiuser MIMO systems in a multicell environment," *EURASIP J. Wireless Comm. and Networking*, vol. 202, no. 1, pp. 1-10, Jun. 2012.

[8] X. Yi and E. K. Au, "User scheduling for heterogeneous multiuser mimo systems: A subspace viewpoint," *IEEE Trans. on Vehicular Technology*, vol. 60, no. 8, pp. 4004-4013, 2011.

[9] E. Viterbo and A. Hottinen, "Optimal user pairing for multiuser MIMO," in *Proc. IEEE International Symposium on Spread Spectrum Techniques and Applications (ISSSTA)*, pp. 242-246, Aug. 2008.

[10] J. Luo, C. Rosenberg, and A. Girard, "Engineering wireless mesh networks: Joint scheduling, routing, power control, and rate adaptation," *IEEE/ACM Trans. on Networking*, vol. 18, no. 5, pp. 1387-1400, 2010.

[11] Stanford Business Software, Inc., Mountain View, CA, USA. "MINOS 5.51," [Online]. Available: http://www.sbsi-sol-optimize.com/asp/sol_product_minos.htm

Automatic Transmission Power Control Management Scheme for Mitigating Co-Layer Interference in Femtocell Network

Hossain Mohammad Mahbub, Chowdhury Ahad Ibne Alamgir, Atiqur Rahman

ECE Department, North South University, Bangladesh

mahbub_eee091@outlook.com,chowdhury910@northsouth.edu,atiqur@northsouth.edu

Abstract– **A significant proportion of data traffic originates indoor. Poor signal reception caused by penetration losses through walls severely obstructs the operation of indoor data service in macrocell network. So, the femtocell is a worthwhile solution to improve indoor coverage, increase network capacity and also to enhance user throughput. The random deployment of femtocell inside residence and enterprise environment will cause co-layer and cross-layer interference. We have executed our proposed model in two steps to mitigate co-layer interference. We use automatic transmission power control management scheme. In proposed model, we work with reference femtocell and layer-1 femtocells (HNB$_s$) to regulate its transmission power. It helps to achieve optimum SINR and data rate value. Simulation results validate our proposed scheme, which can effectively control the transmit power of reference femtocell and layer-1 femtocells. We have been considered optimum SINR and throughput as performance indicator in our proposed scheme.**

Keywords– **Reference Femtocell, Automatic Transmission Power Control Management, SINR, Layer-1 Femtocell.**

I. INTRODUCTION

Femtocell is an advancing technology to increase the indoor coverage area and handle the rising demand of indoor multimedia traffic, since more than 80% of wireless communication is reported to be originated in residence area [1]. Femtocell is a low-power wireless access point, which works on licensed spectrum to connect mobile devices to a mobile operator's network using residential DSL or cable broadband connections [2]. There is a possibility of creating interference between femtocells by deploying large-scale femtocells in indoor environment. The overall voice quality and throughput of the network will decrease to user level, if co-layer interference will occur. Main purpose of our proposed model is to improve channel quality by reducing co-layer interference between femtocells and finally we decided to use Automatic Transmission Power Control Management (ATPCM) mechanism. Using this method, we adjusted the transmitted power of femtocell in accordance with the interference level from surrounding femtocells. In this paper, we have been worked on minimizing the power consumption as well as maintaining the SINR requirement for our users. In order to ensure QoS to user end, we have worked on improving SINR for communication channel, better data rate to the user end and regulating transmission power of femtocell.

We organize the paper as follows: In Section II we describe our proposed model, we show simulation parameters and analysis in Section III and we conclude with a summary in Section IV.

II. PROPOSED MODEL

In our research work, co-layer interference mitigation is being taken as main concern. In this section, we will discuss in details about our proposed scheme. We consider a scenario where we have r number of femtocell subscribers for reference femtocell and q number of femtocell base stations in layer-1. Based on these, general SINR formula [3] for communication channel of reference femtocell is written below:

$$\psi_{r,q}^{(i)} = \frac{T_{signal,r}^{(i)}}{n_0 \times \Delta f + \sum_{U=1}^{q} \mathbf{I}_{signal,r*}^{(U)}}$$

$$\Rightarrow \psi_{r,q}^{(i)} = \frac{P_r^{(i)} \times G_r^{(i)}}{n_0 \times \Delta f + \sum_{U=1}^{q} \mathbf{P}_{r*}^{(U)} \times \mathbf{G}_{r*}^{(U)}} \quad (1)$$

Where, $\Psi_{r,q}^{(i)}$ is the signal-to-interference-plus-noise ratio for i^{th} reference femtocell to femtocell subscriber, r with respect to q^{th} number of layer-1 femtocell; $T_{signal,r}^{(i)}$ is transmitted signal from transmitter of reference femtocell; $P_r^{(i)}$ is assigned transmission power of reference femtocell base station to user, r; $G_r^{(i)}$ is the gain for reference femtocell access point to user, r; $\mathbf{P}_{r*}^{(U)}$ is the corresponding transmission power of layer-1 femtocell base station to user, r*; $\mathbf{G}_{r*}^{(U)}$ is the gain for layer-1 FAP to user, r*; n_o is the spectral efficiency of white noise power; Δf is the sub-carrier spacing.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

A. Automatic Transmission Power Control Management (ATPCM) Mechanism

Automatic Transmission Power Control Management Approach is a dynamic interference reduction technique which can be used to mitigate co-layer interference where the femtocells provide good service to user in accordance with the interference signal from neighboring femtocells. In phase I, initially the transmitted power and gain of reference femtocell is set. When the SINR of reference or layer-1 femtocell goes below our defined standard level, then this information is sent to Femtocell Service Manager (FSM). Because, the FSM has administrative privilege to control the Femtocell Access Point (FAP). After receiving these informations from FAP, then FSM sends its command to FAP for improving SINR of communication channel by deploying effective interference mitigation algorithm for a certain femtocell deployment scenario. To reduce interference, transmitter of femtocells mainly adjust its transmission power with a proper procedure which can be accomplished by using Eq. (3) and Eq. (6). The transmission power of all layer-1 femtocells will adjust accordingly.

For developing path loss formula for femtocell in indoor environment, few approximations has been taken. Our Path loss formula [4] is expressed as:

$$PL(\mathrm{dB}) = 38.5 + 20\log_{10}(\mathbf{D}) + \text{Penetration loss} \qquad (2)$$

Where, **D** (in meter) displays the distance between two transmitters of two femtocells or the distance between femtocell and subscriber equipment (SE). Penetration loss is indoor penetration loss through the wall, which is supposed to be 17dB in our proposed scheme [4].

In paper [5], authors considered 15dB SINR as the minimum threshold value for communication channel.

Fig. 1 shows algorithm I flow chart for reference femtocell. In this scheme, the distance between reference femtocell and layer-1 femtocell is specified by us, which is between 14m and 25m. All femtocells are working in close access scenario. In phase I, reference femtocell has worked according to algorithm I and in phase II, the layer-1 femtocells has worked according to algorithm II (Fig. 3). Calculation of SINR for reference femtocell can be accomplished by using Eq. (1).

The adjusted SINR for reference femtocell has been calculated using the formula written below:

$$\psi_{r,q}^{(\mathrm{i})} = \frac{\dfrac{P_r^{(\mathrm{i})} \times \sum_{U=1}^{q} G_{\mathrm{r}*}^{(U)}}{q}}{n_0 \times \Delta\mathrm{f} + \sum_{U=1}^{q} \mathbf{P}_{\mathrm{r}*}^{(U)} \times \mathbf{G}_{\mathrm{r}*}^{(U)}} \qquad (3)$$

Here, $\Psi_{r,q}^{(i)}$ is adjusted signal-to-interference-plus-noise ratio of ith reference femtocell to femtocell subscribers, r with respect to q^{th} layer-1 femtocell.

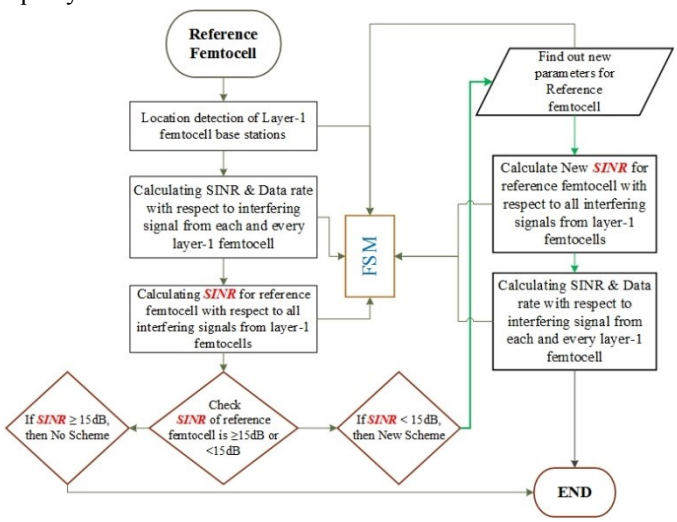


Fig. 1: Algorithm I Flow-chart for Reference Femtocell in Phase I

Data-rate for reference femtocell users in phase I is written below:

$$C_{r,q}^{(i)} = \Delta\mathrm{f} \times \log_2(1 + \alpha * \psi_{r,q}^{(i)}) \qquad (4)$$

Where, α is a constant of Bit Error Rate (BER) and can be defined, α = –1.5/ln (5 BER). So, the setting of BER is 10^{-6} . $C_{r,q}^{(i)}$ is the data rate for reference femtocell with respect to layer-1 femtocell or femtocells.

In phase II, our goal is to increase channel quality by improving signal-to-interference-plus-noise ratio. In proposed model, a femtocell scenario with 12 layer-1 femtocells is considered. Let's about for a layer-1 femtocell, the remaining eleven layer-1 femtocell signals are interference signal. For further development of proposed model, we need to know the distance between layer-1 femtocells. To calculate these distances we have introduced a formula (Fig. 2) which is named as 'Distance Calculating Formula: Femtocell' [4]:

$$\mathrm{D}_{differ} = \sqrt{d_3^2 + d_6^2 - 2 \times d_3 \times d_6 \times \cos(\delta)} \qquad (5)$$

Where, δ is the angle between the length of two sides d_3 and d_6 and can have value from 0 to 360 degrees. d_6 and d_3 is two distances between reference femtocell and layer-1 femtocells, D_{differ} is the distance between two layer-1 femtocells.

The adjusted SINR formula for layer-1 femtocells has been calculated using the formula written below:

$$\varsigma_{t,L}^{\mathrm{K}} = \frac{P_t^{(\mathrm{K})} \times [\max\langle \mathbf{G}_t^{(L)} \rangle]}{n_o \times \Delta\mathrm{f} + \sum_{L \to N^{rem_\mathrm{layer1femto}}} \mathbf{P}_t^{(L)} \times \mathbf{G}_t^{(L)}}$$

$$\Rightarrow \varsigma_{t,L}^{K} = \frac{T_{signal_Layer1,t}^{(K)}}{n_o \times \Delta f + \sum_{L \to N^{rem_layer1femto}} \mathbf{P}_t^{(L)} \times \mathbf{G}_t^{(L)}} \qquad (6)$$

Where, $\varsigma_{t,L}^{K}$ is adjusted signal-to-interference-plus-noise ratio for K^{th} layer-1 femtocell to femtocell users, t with respect to L^{th} number of remaining layer-1 femtocells; $T_{signal_Layer1,t}^{(K)}$ is transmitted signal of Kth layer-1 femtocell to femtocell users, t; $\mathbf{G}_t^{(L)}$ is the corresponding gain of Lth layer-1 femtocell base station to users, *t;* $\mathbf{P}_t^{(L)}$ is the corresponding transmitted power of Lth layer-1 femtocell access point to users, t; $N^{rem_layer1femto}$ is the set of all remaining layer-1 femtocells.

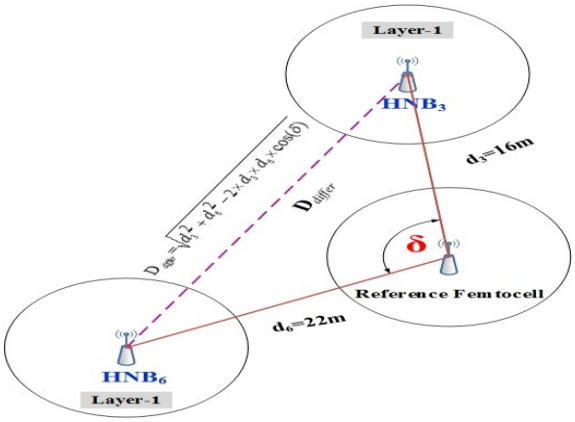


Fig. 2: 'Distance Calculating Formula: Femtocell' in Phase II

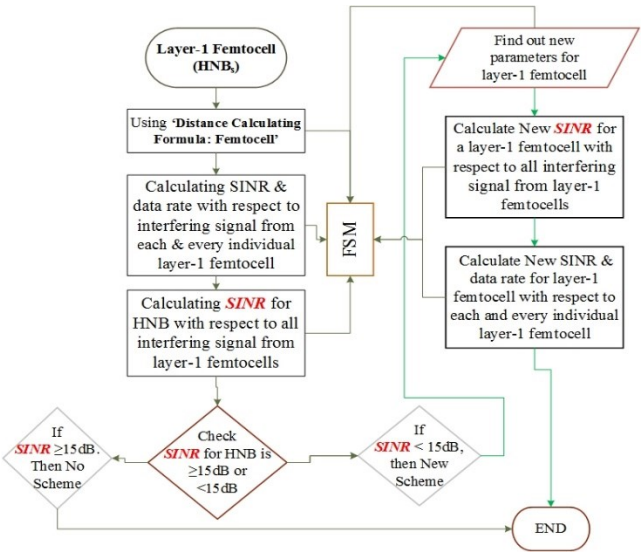


Fig. 3: Algorithm II Flow-chart for Layer-1 Femtocells in Phase II

In phase II, we calculate the data rate for femtocell users by using the formula [Eq. (7)] is presented below:

$$C_{t,L}^{K} = \Delta f \times \log_2[1 + \alpha * \varsigma_{t,L}^{K}] \qquad (7)$$

Where, $C_{t,L}^{K}$ is the throughput of serving K^{th} layer-1 femtocell to users, t with respect to remaining L^{th} number layer-1 femtocells. Fig. 4 and Fig. 5 has shown the femtocell scenario before and after our proposed model (applicable for both phase I and II), which we want to achieve by applying our Automatic Transmission Power Control Management (ATPCM) scheme.

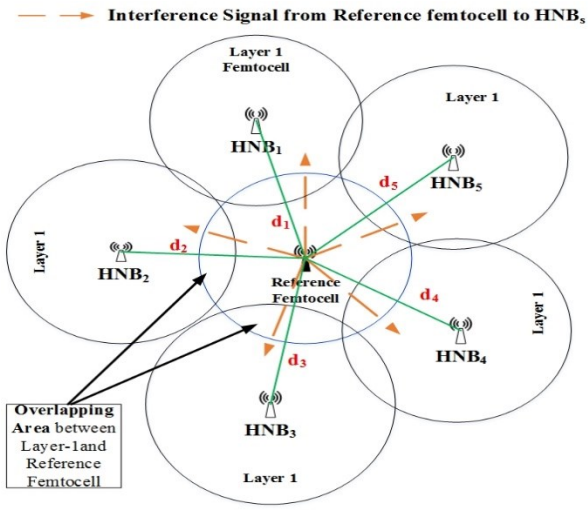


Fig. 4: Femtocell Network before Proposed Scheme

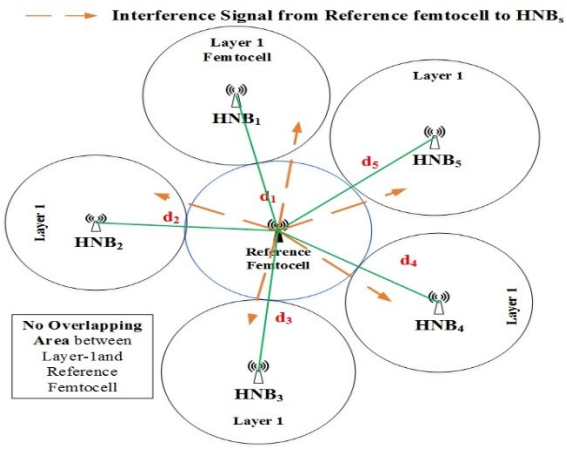


Fig. 5: Femtocell Scenario after Proposed Scheme

III. SIMULATION PARAMETERS and ANALYSIS

The simulation parameters for reference femtocell and layer-1 femtocells are given below:

TABLE I. Simulation Parameter for Reference femtocell and Layer-1 Femtocell

Parameter	**Value (Reference Femtocell)**		**Value (Layer 1 Femtocell)**
	Reference Femtocell	Layer 1 Femtocell	Layer-1 Femtocell
No. of Cells	1	8	8
No. of Users	8	64	64
Channel Bandwidth	15 MHz		15 MHz
Sub Carrier Spacing	15 KHz		15 KHz
White Noise Spectral Density	-174 dBm/Hz		-174 dBm/Hz

Before applying our proposed model, the gain of reference femtocell is calculated using $\mathbf{G}_{reference}=10^{(PL/10)}$. We have considered that the primary transmission power for all femtocells is 20mW and antenna pattern of all femtocells is Omni-directional.

From Fig. 6, we can be perceived that our proposed model has shown noteworthy improvement in femtocell scenario which is considered here. Before applying the scheme there were SINR below 15dB for a reference femtocell, whereas after applying the model SINR of a reference femtocell has achieved 2.60dB improvement.

In addition, data rate has increased successfully for a reference femtocell by our proposed model. Before the proposed model, it was 1.62×10^4 bps (bits per second). After the scheme, it has improved to 1.82×10^4 bps to provide better throughput [Fig. 7].

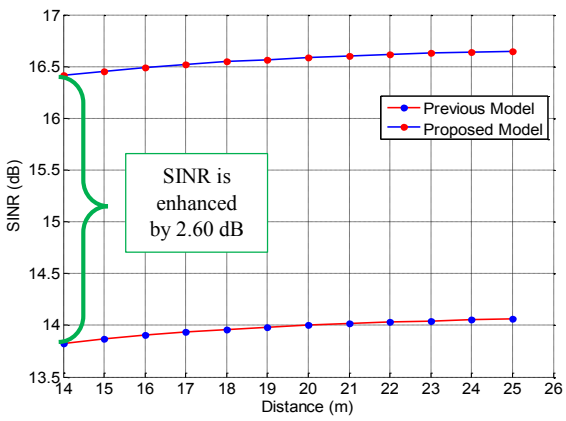


Fig. 6: SINR Improvement graph for a Reference Femtocell

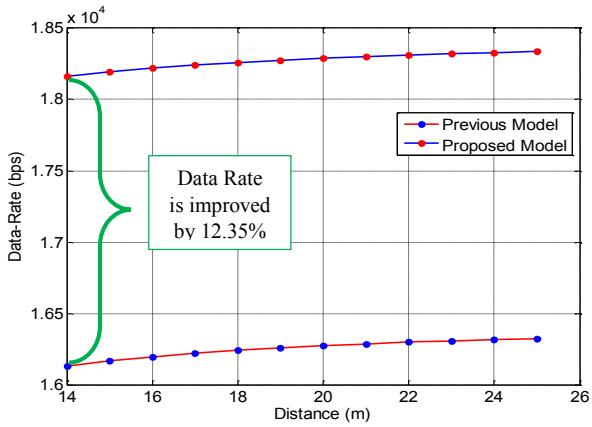


Fig. 7: Data Rate Improvement graph for a Reference Femtocell

From Figure 8, it is showed that firstly the values of SINR for a layer-1 femtocell with respect to other's layer-1 femtocell are less than 15dB, on the other hand after applying the proposed model the values of SINR has shifted above 15dB. From this result, we can say that channel quality has improved.

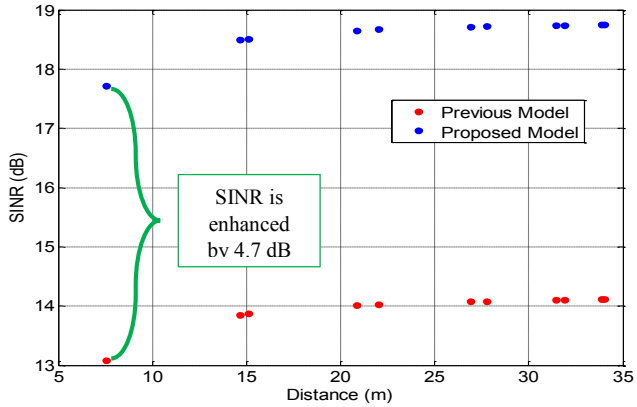


Fig. 8: SINR Improvement graph for a Layer-1 Femtocell

IV. CONCLUSION

In this paper, we have been worked to mitigate co-layer interference by using our proposed power control algorithm. Interference issue is one of the main concern for femtocell network because of its unplanned deployment in macrocell network. After using this scheme, we have developed the channel quality and data rate for communication channel. In first phase, we succeed almost 2.6dB SINR improvement in communication channel for a reference femtocell. This scheme improves the data rate around 12.35% for reference femtocell. Simulation results determine the usefulness of our model for achieving threshold SINR and data rate in communication channel. It reflects that we have effectively reduced co-layer interference in femtocell network by achieving optimum SINR and increased throughput.

REFERENCES

[1] Chandrasekhar, J. Andrews, and A. Gatherer, "Femtocell network: a survey", *IEEE Commun. Mag.,* vol.46, no.9, pp.59-67, Sep. 2008.

[2] Saquib, N., Hossain, E., Le, L. B., Kim, D. I., "Interference management in OFDMA femtocell networks: issues and approaches", *Wireless Communications, IEEE*, vol.19, no.3, pp.86-95, June 2012.

[3] Cao, G., Yang, D., Ye, X., & Zhang, Xin., "A Downlink Joint Power Control and Resource Allocation Scheme for Co-Channel Macrocell-Femtocell Networks", *Wireless Communications and Networking Conference (WCNC),* pp.281-286, 28-31 march 2011.

[4] Saleem, S., & King, H., "Avoidance of Co-tier Interference between femtocells with Different Access Modes", *International Journal of Information and Communication Technology Research,* vol. 2, no.8, pp.617-626, august 2012.

[5] Gutierrez-Estevez, D. M., Canberk. B., & Akyildiz, I. F. (2012). "Spatio-Temporal Estimation for Interference Management in Femtocell Networks," *IEEE 23rd International Symposium on Personal Indoor and Mobile Radio Communication*, pp. 1137-1142.

Braille Thermometer: Thermometer for Deaf-blinds

Design and Implementation of Braille Thermometer and Its Performance Comparison with Clinical Thermometer

Md. Didarul Islam,[1] Abdullah Saif Mondol[2]
Department of Electrical and Electronic Engineering[1], Faculty of technology and bionics[2]
Uttara University[1], Rhine – Waal University of applied sciences[2]
Dhaka, Bangladesh[1]; Kleve, Germany[2]
*s102bueeet@yahoo.com

Abstract— **This paper introduces a new clinical thermometer named as 'Braille Thermometer' which will be used for measuring the body temperature of patients who are blind and also have hearing problem (deaf-blind). It is actually a microcontroller based device which collects temperature through a temperature sensor and indicates this temperature in a dial which is marked in Braille numbers. In this paper, the temperature readings of the Braille thermometer are also crosschecked with the readings of a clinical thermometer where input temperature was taken from the same source.**

***Index Terms*—Braille numbers, thermometer, deaf-blind, microcontroller, temperature sensor, stepper motor.**

INTRODUCTION

Clinical/ medical thermometer is available in the market. But these things are of no use to the blind people. Modern science has created audible thermometers too, which will tell the temperature aloud. Number of those people is not negligible who is blind and also have hearing problem at the same time. Considering people who have such condition, known as deaf-blinds, Braille thermometer is a great attempt. Patient can know their body temperature with the help of Braille numbers. This thermometer can indicate temperature from 98°F to 108°F (only integer temperatures, not fractions). At the end of this paper, the temperature input and corresponding indication is plotted with the help of MATLAB.

RESEARCH METHODOLOGY

In this Braille thermometer, human body temperature is measured in Fahrenheit. But the temperature sensor used here, calculates everything in Celsius. So, the following formula has been used to convert the temperature from degree Celsius to degree Fahrenheit [1]:

$$T_C = \frac{5}{9}(T_F - 32)$$

This temperature sensor has the following relation between input temperature and output voltage:
For 1°C input temperature, the output voltage will be 10 mV; which is discussed in the Modeling & Design section.
In order to get the desired output, corresponding to a specific input signal, the microcontroller used here needs to convert analog signal into digital signal. The conversion result can be formulated as follows [2]:

$$ADC = \frac{V_{IN} \times 1024}{V_{REF}} \quad (1)$$

Where V_{IN} is the input voltage and V_{REF} is the selected reference voltage.

MODELING & DESIGN

The model of the Braille thermometer is constructed with six major components. These components are:

1. 5V fixed DC supply.
2. Temperature sensor.
3. Microcontroller.
4. Motor driver.
5. Stepper motor.
6. Braille dial.

These major sections and their interconnections are illustrated in Fig. 1:

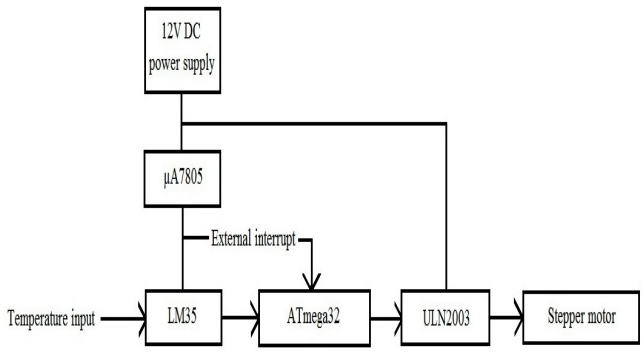


Fig. 1 Block diagram of the Braille thermometer.

In this Modeling and Design section, the entire component's power supply, input and output connections are described. Also the signal transferring between two segments is explained sequentially.

Fixed 5V DC Supply

For both ATmega32 microcontroller and temperature sensor LM35 to operate, a constant 5V DC supply is required. This is done with the help of µA7805 IC. If an input DC voltage greater than 5V is given between pin 1 and pin 2 (ground), with the following circuit arrangement, then a constant output of 5V is found between pin 3 and pin 2 of µA7805 IC [3]:

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

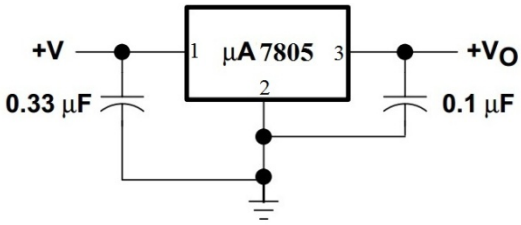


Fig. 2 Fixed 5V DC output regulator using µA7805 IC.

Temperature Sensor

LM35 IC is used as the temperature sensor for Braille thermometer which can detect temperature from -55°C to 150°C [4]. Here temperature is being measured from 98°F to 108°F, in Celsius scale 36.67°C to 42.22°C, which is quite within the range of the temperature sensor.

The constant 5V which acts as the supply voltage for this IC, is connected between pin 1 and pin 3 (ground). LM35 is an analog sensor which converts the surrounding temperature to a proportional analog voltage [5]. Now an output voltage corresponding to room temperature can be found between pin 2 and pin 3. If anyone holds the LM35 IC with his hand, output voltage will change according to his/her body temperature. From LM35 datasheet, few input temperature and output voltage relationships can be found like below [4]:

TABLE I
FEW TEMPWRATURE INPUTS AND THEIR COORESPONDING OUTPUT VOLTAGES FOR LM35 TEMPERATURE SENSOR.

Input temperature (°C)	Output voltage (mV)
-55	-550
25	250
150	1500

Microcontroller

Taking an input signal and converting this signal into an desired output signal, microcontroller is required. ATmega32 microcontroller is chosen for that purpose, considering its connection and coding simplicity.

Now, the output voltage of the LM35 temperature sensor will act as an input signal to the ATmega32 microcontroller's Analog-to-Digital Converter. For this microcontroller, Port A (pin 33 to 40) is the ADC [2]. Pin 40 is selected as the input pin of the microcontroller. The operating principle of the ADC is illustrated in equation (1). 5V fixed DC supply is selected as the reference voltage of the Analog-to-Digital Converter, which is connected to the pin 32 [2] of ATmega32. When the microcontroller gets an external interrupt signal, it will give output signal to its selected output port (Port C chosen here). Pin 16 (INT0) of ATmega32 microcontroller is used here to accomplish that job [2], where with the help of a push button an external interrupt is given whenever the output of the microcontroller is required. The microcontroller code was written and converted to HEX file with AVR studio software. The HEX file was loaded in the microcontroller with the help of PonyProg software, while the microcontroller was connected to a computer's parallel port with a DB 25 connector [6]. The necessary hardware required to load the program in ATmega32 microcontroller is shown in Fig. 3.

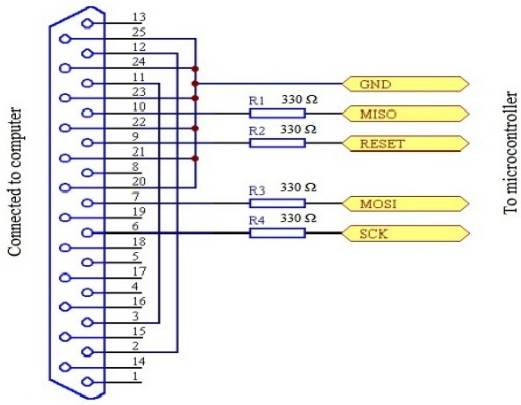


Fig. 3 Hardware used to load program in ATmega32.

The output of the microcontroller is connected to the motor driver.

Motor Driver

ULN2003 motor controller is a Darlington array IC which is generally used for interfacing between low level logic circuitry and power loads. Motor driver is necessary here, because stepper motor cannot be directly operated from a microcontroller. In order to drive a stepper motor, high current is required, which is supplied by that IC. This particular IC has seven drivers. ULN2003 IC's Pin 9 is provided a fixed 12V DC supply and pin 8 of is connected to ground [7].

Stepper Motor

It will be a convenient and feasible design if a stepper motor of 30° step angle is used here. If the step angle is a factor of 30, then it will work too. Stepping motor M42SP-6NPK is used here whose step angle is 7.5° [8]. So, in order to give a 30° rotation, the motor will have to move 4 steps.

Stepper motor M42SP-6NPK has four wires of color Red, Orange, Blue and Yellow. If a square pulse is given between the Red and the Orange wire, then the motor will rotate in clockwise direction. Whereas a square pulse between the Blue and the Yellow wire will rotate the motor in counter clockwise direction.

Braille Dial

The whole motor, indicator setup is mounted on a dial where temperature is written in Braille numbers from 98°F to 108°F.

The temperature indicating Braille dial has 12 sections. These sections are:

98°F	99°F	100°F	101°F	102°F	103°F
104°F	105°F	106°F	107°F	108°F	START

Their corresponding Braille numbers are as follows [9]:

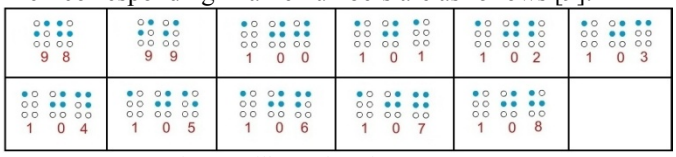


Fig. 4 Braille numbers from 98 to 108.

The Braille thermometer with its power supply and the program loader is shown in Fig. 5:

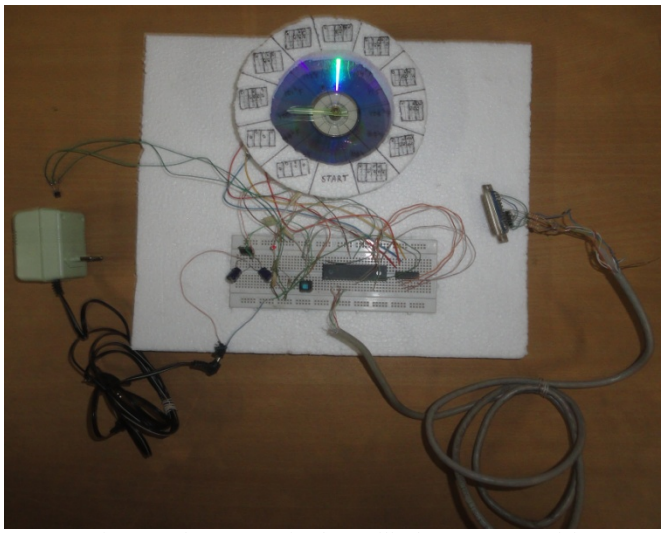


Fig. 5 Hardware setup for the Braille thermometer model.

OPERATING PRINCIPLE

Firstly, the overall circuit will have to be powered up with a 12V DC power supply. From here a fixed 5V DC supply will be generated. In this setup, the temperature sensor and the microcontroller will require 5V fixed DC supply and the motor driver needs 12V DC voltage. Then, when a person holds the temperature sensor, it gives the output voltage corresponding to the input temperature as discussed in the Research Methodology section. Human body temperature is an analog signal and corresponding output voltage of LM35 is also an analog signal. So, the output signal of LM35 will have to be converted to digital signal, since microcontroller deals only with digital signals in its algorithm part. That's why output of the LM35 temperature sensor is connected to the ADC (Analog to Digital Converter) of ATmega32. When the microcontroller's ADC gets an analog input voltage, it converts the analog signal into digital signal as mentioned in equation (1). With an appropriate microcontroller code, ATmega32 can produce desired output voltage at selected output pins corresponding to a particular input. The coding part describes the relation between input temperature and the corresponding number of pulses generated in the output pins to rotate the stepper motor specific degrees. By connecting the bidirectional stepper motor to the microcontroller's output port, the stepper motor can be rotated from 0° to 360° both clockwise and anticlockwise. The stepper motor will have an indicator needle connected to its rotor. Patient will have to follow the indicator with hand and read the corresponding temperature by touching the Braille numbers on the dial.

But, as mentioned earlier, stepper motor cannot be directly operated from microcontroller. In order to do that stepper motor should be connected with microcontroller through a Darlington array IC like ULN2003.

Initially, the stepper motor's indicator will be fixed at a position named as 'START'. Now according to the temperature sensor's input the stepper motor will rotate clockwise few degrees, stay there for 30 seconds and again rotate anticlockwise those particular degrees. That means after waiting 30 seconds, the indicator will return to its initial position. The complete algorithm of Braille thermometer's operation is chronologically explained through a flowchart:

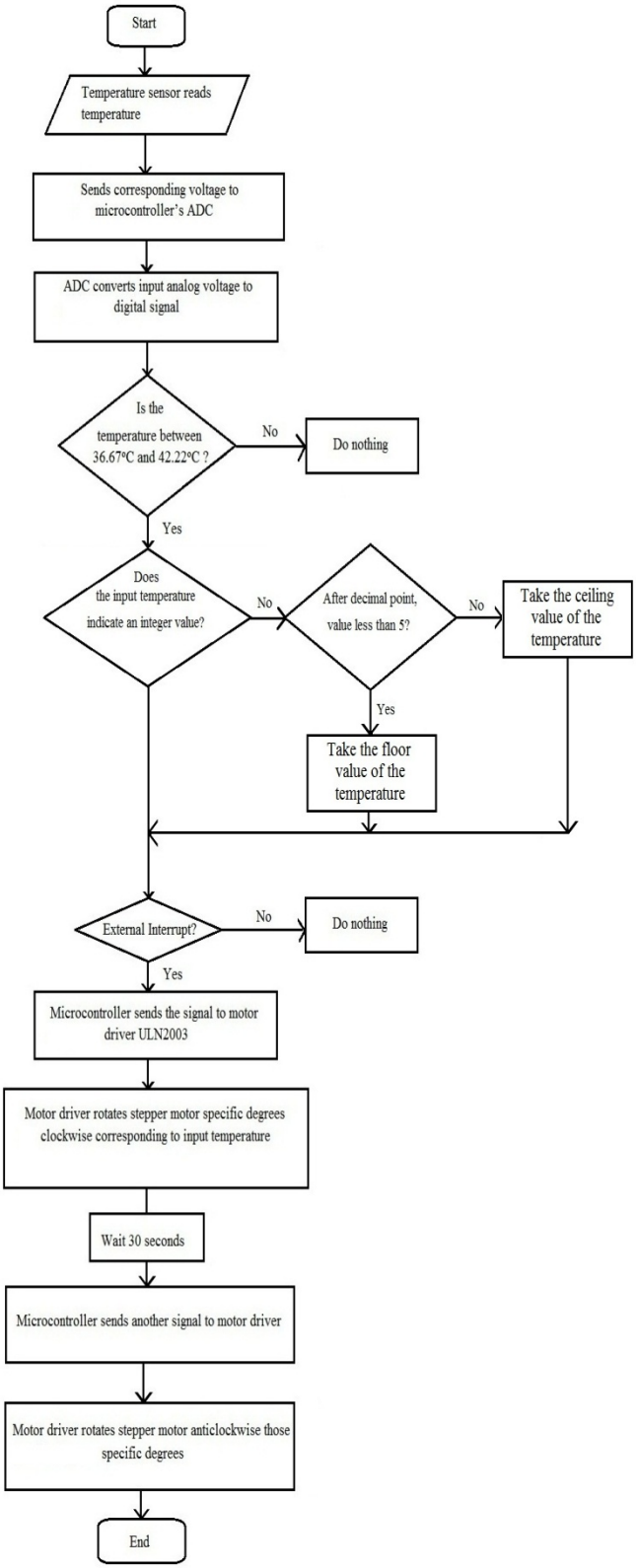


Fig. 6 Flowchart for the operating principle of Braille thermometer.

RESULTS & DISCUSSIONS

In order to check the Braille thermometer's performance, its temperature sensor was heated from 98°F to 108°F with a temperature variable heater. For a particular temperature, Braille thermometer's reading was taken. At the same time, the heater's temperature was also measured by a clinical thermometer. From this comparison it was observed that, for an integer temperature, the reading of the Braille thermometer and the reading of the clinical thermometer was same. Fig. 7 shows how Braille thermometer and clinical thermometer are reading temperature of a heater at the same time.

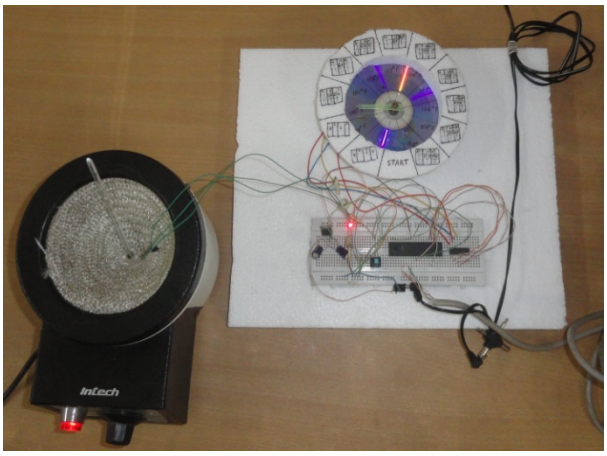


Fig. 7 Temperature reading comparison between Braille thermometer and clinical thermometer.

For 98°F to 108°F temperature, the Braille thermometer's stepper motor rotation was recorded and a temperature versus rotation graph was plotted using MATLAB.

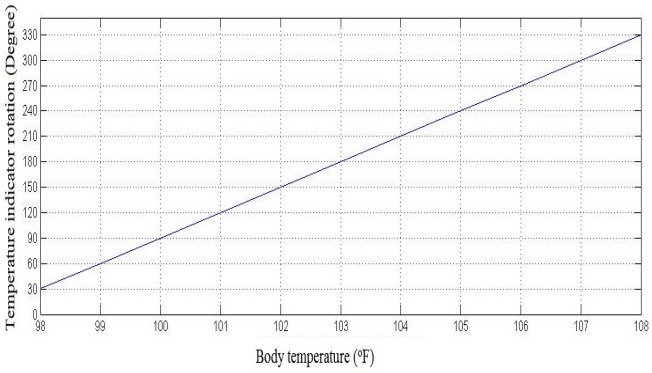


Fig. 8 Graph between temperature and corresponding motor rotation.

Here only integer temperature has been used for calibrating the Braille thermometer. Whenever a fractional temperature is given as input, the microcontroller follows a simple algorithm to convert the fractional temperature into an integer one. If, after the decimal point, the temperature is less than 5, the microcontroller will consider its floor value (e.g. for 100.25°F it will consider it as 100°F). But after the decimal point, if the temperature is equal or greater than 5, the microcontroller will convert it into the ceiling value (e.g. 100.5°F will be interpreted as 101°F).

CONCLUSION & FUTURE WORK

From the previous comparisons and observations it can be concluded that, this Braille thermometer model is only perfect for integer values of body temperature. This is not completely a practical expectation, because, for many medical applications, a little bit temperature change may hold a large significance. Considering this single limitation, Braille thermometer can be an excellent device to measure a deaf-blind patient's body temperature by himself/herself.

For future work/research, Braille thermometer can be constructed in such a way that it will indicate fractional temperatures (e.g. one digit after decimal point) too. But on a single dial, accommodating so many (101 values, if all the values between 98°F to 108°F is show) Braille numbers is not easy to manage. On the other hand, for a deaf-blind patient, it will be very difficult to pinpoint his/her temperature. The idea of solving this problem is: for fractional temperatures, the patient will read the integer part of the temperature from the Braille dial as this model does. And for measuring the fractional part, there will be another vibrating motor, which will vibrate according to the value of the fractional temperature (e.g. for 0.5°F the vibrating motor will vibrate 5 times). In this research work, the circuitry was powered by an adapter. Light weight rechargeable battery can be used to reduce the weight, which will be convenient for carrying too. For professional construction, instead of breadboard; veroboard or Printed Circuit Board (PCB) can be used for wire connections. Finally, the motor used (M42SP-6NPK) has a weight of 96 grams. This can be replaced by a motor of small size and weight (e.g. stepper motor PN 16305. Its weight is around 9 grams and has a 30° step angle [10]). Thus Braille thermometer can be made more accurate, compact, portable and easy to use.

REFERENCES

[1] David Halliday, Robert Resnick, Jeart Walker, "Fundamentals of Physics," 7th edition, pp. 480

[2] Information available at: http://www.atmel.com/images/doc2503.pdf Accessed on: 6th June 2014

[3] Information available at: http://www.ti.com/lit/ds/symlink/ua7805.pdf Accessed on: 7th June 2014

[4] Information available at: http://www.ti.com/lit/ds/symlink/lm35.pdf Accessed on: 7th June 2014

[5] Mahmud, K. ,Alam, M.S. , Ghosh, R : "Design of digital thermometer based on PIC16F77A single chip microcontroller"-2013 3rd International Conference on Consumer Electronics, Communications and Networks (CECNet), IEEE, pp 246-249,ISBN:978-1-4799-2859-0,November 2013

[6] Information available at: http://www.qsl.net/ba1fb/avrisp.gif Accessed on: 6th June 2014

[7] Information available at: http://www.ti.com/lit/ds/symlink/uln2003a.pdf Accessed on: 7th June 2014

[8] Information available at: http://www.datasheetarchive.com/dl/Datasheets-SL3/DSASL0049076.pdf Accessed on: 7th June 2014

[9] Information available at: http://www.ridbcrenwickcentre.com/louisbraille/facts/braille-punctuation/ Accessed on: 6th June 2014

[10] Information Available at: http://www.aeroflex.com/ams/motion/datasheets/stepper_30_degree.pdf Accessed on: 7th June 2014

A Field-Referenced Surface Electrode for Biosignal Detection

Sazia Afreen Eliza,[1,*] Alfred LeBron Paige,[2] and Mohammad Rafiqul Haider[1,*]
[1]Department of Electrical and Computer Engineering, University of Alabama at Birmingham,
Birmingham, AL, USA
[2]Department of Neurology, University of Iowa
Iowa City, IA, USA
*{seliza, mrhaider}@uab.edu

***Abstract*—Biopotential measurements for clinical diagnostics and brain-machine interface are usually carried out by placing multiple electrodes to the target areas. Due to the non-invasiveness, ease of deployment, and relatively simpler manufacturing, surface electrodes have found immense number of applications. This work presents a field-referenced surface electrode for the measurement of surface potentials. Each electrode can be addressed by applying a constant reference signal on the field plate and the corresponding induced potential is measured from the sensing plate. The reference-biased field plate shields the sensing plate from tangential field lines and interference from the neighboring electrodes. Due to the presence of the reference signal, the electric field originating from a biological source crowds towards the referenced-field plate and impacts the induced potential in the sensing plate. Simulation results indicate the better selectivity performance of the proposed structure.**

***Index Terms*—Biosignal, surface electrode, selectivity.**

I. Introduction

The use of high-resolution neural recordings from the cortical surface for basic research and clinical medicine has been shown for a wide range of applications. To localize the seizure focus, neurosurgeons apply surface electrodes to measure the EEG or ECoG signals during the epileptic event. However, the subdural electrodes in use clinically, for the diagnosis and treatment of epilepsy, are much larger ($\sim 3\ mm$ in diameter) and have large interspacing ($\sim 10\ mm$) because of the clinical need to record from large areas of the brain surface ($80\ mm \times 80\ mm$) for the localization of seizure-originating brain regions [1]. Spatial spectral analysis of electrocorticograms from the superior temporal gyrus and motor cortex has revealed that electrode spacing should be 1.25 mm or closer in humans to sufficiently capture the rich spatial information that is available [2]. Motor control signals [3] and spoken words [4] can be decoded with substantially improved performance using electrodes spaced 1 mm apart or less. Therefore, for high-resolution brain mapping, a dense electrode array with high spatial-resolution is inevitable.

Much of the existing research in electrode technology has focused on penetrating electrode arrays, such as the Utah array [5], which can provide a high-resolution interface to a small area of cortex [6]. However, the penetrating microelectrodes suffer from reduced lifetime [7] (6-12 months) and may cause hemorrhage and inflammatory tissue responses from the immediate insertion [8],[9] or over long periods of time, because of the inability of the rigid penetrating electrodes to flex and move as the brain pulses, swells and contracts [10].

Conventional disc electrodes are currently being used for surface potential measurements for EEG, EMG and ECG signal acquisitions. The recorded signals have reference electrode problems as idealized references are not available. Surface Laplacian mapping has been shown to enhance the high spatial frequency components and spatial selectivity of the electrical activity located close to the point of interest. These unique characteristics can be derived as the surface Laplacian being the second spatial derivative of the surface potentials. To increase the spatial selectivity, researchers introduced a bipolar electrode configuration and calculated the induced potential using a five-point finite difference method (FPM). Others applied a quasi-bipolar tri-polar with two elements shorted concentric electrode configuration. To further increase the spatial resolution, a tri-polar concentric electrode configuration for approximating the analytical Laplacian based on a nine-point finite difference method (NPM) has been reported [11]. High-resolution surface electrodes over a large brain area had also been impractical owing to the complexity of connecting thousands of wires in the small intracranial space. Recently a 360-channel active electrode array capable of sampling a five-fold larger region of brain ($10\ mm \times 9\ mm$) than the previous methods with high spatial resolution ($\sim$500 μm spacing) and high temporal resolution ($>$ 10 kS/sec) and with reduced number of interconnecting wires has been reported [1].

Our work reports a field-referenced passive electrode for the measurement of surface potentials. The proposed electrode manifests a field plate around a sensing plate separated by a dielectric layer. Due to the presence of a reference signal on the field plate, the source electric field crowds towards the electrode and affects the induced potential on the sensing plate. The higher the field crowding the closer is the signal source proximity. The reference signal biased field-plate also provides measured signal robustness to interferences and results in better signal selectivity.

The description of our work is as follows: In section II, we have discussed our proposed electrode structure. Section III describes the electrode model for the finite element simulation setup and Section IV shows the simulation results for the electric field distribution and the induced potentials on the electrodes . A comparative study of induced Laplacian potentials has been shown between the proposed electrode and the conventional disc electrode with the changes of the position of an electrostatic source. Section VI summarizes the advantages of our proposed electrode structure and the future work.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

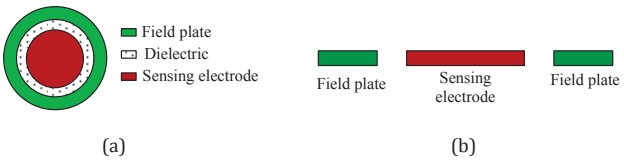


Fig. 1. Schematic view of the proposed electrode. (a) Top view where the field plate encircles the sensing plate separated by a dielectric layer, and (b) Side view of the electrode.

II. Proposed Field-Referenced Electrode

Fig. 1(a) shows the schematic view of a field-addressable electrode. It manifests a center sensing plate and an annular field plate. The proposed electrode is similar to a bipolar electrode; however, unlike the bipolar electrode the field plate in the proposed electrode acts as a reference line to pull the electric field towards the respective sensing plate. The field plate and the sensing plate are isolated by a dielectric layer. Fig. 1(b) shows the side view of the proposed electrode. In this arrangement, measurement of induced potential from the sensing plate is carried out when there is a reference signal in the field plate. The field plate provides not only reference signal location but also electrical shielding from crosstalk and interference. The electric field crowding occurs due to the presence of reference signal and enhances the induced potential of the sensing plate.

III. Finite Element Simulation

The proposed electrode structure has been simulated using Maxwell 2D simulator in electrostatic mode. The simulation has been carried out using adaptive mesh analysis with percentage refinement per pass of 15, percentage error of 0.1, and number of passes of 10. The model and the corresponding simulation results are depicted below:

A. Simulation Model

A nine electrode linear array system as shown in Fig. 2 is simulated using finite element analysis software, Maxwell of Ansoft. Each electrode consists of an inner disc to sense the potential and an outer ring to connect with the reference signal. In 2D, each electrode is represented by three rectangular segments. The inner segment is $4\ mm \times 1\ mm$ and the reference segments are $2\ mm \times 1\ mm$ each. The spacing between the electrodes is $4\ mm$. A constant positive charge is placed $10\ mm$ beneath the array. We notice maximum electric field crowding towards a reference electrode when the electric source is closest to that electrode. As we move the source position along the array, the crowding decreases and less potential is induced on the control electrode. The proximity of the source can be identified from the induced potential profile of the array. Au metal electrodes and dielectric

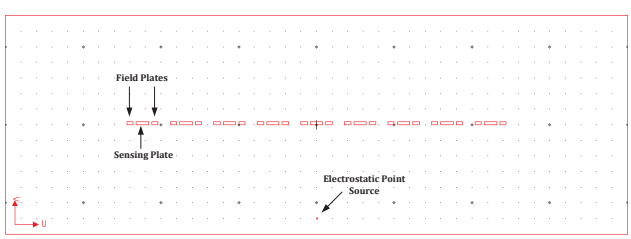


Fig. 2. Simulation of a 9 electrode linear array system using Maxwell 2D.

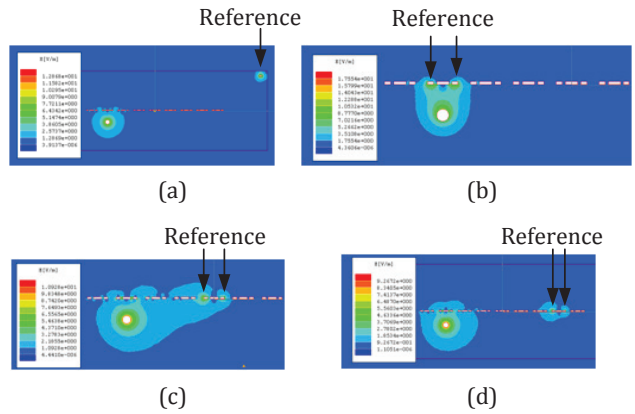


Fig. 3. Electric field distribution as the reference signal is applied to different electrodes along with a distant reference signal at the top right corner. (a) No reference at the electrodes, (b) Reference at the 2nd electrode (c) Reference at the 5th electrode, and (d) Reference at the 8th electrode.

medium are used for simulating the field propagation from a 1 pC of electrostatic source.

IV. Simulation Results

At a time, the plates of a single electrode are ground referenced while the field plates of rest of the electrodes are left floating. A 1 pC floating charge is used as a source and simulation is performed under electrostatic mode. The electric field crowding and the corresponding electric potential is measured from the sensing plates of all the electrodes. The reference location is then shifted to the next electrode and electric potentials on all the center sensing plates are measured. The process is continued for all the reference locations. Fig. 3 shows the electric field distribution for different locations of the reference signals. In Fig. 3(a) conventional disc electrode setup is considered where the reference electrode is far away (top right corner) from the electrode setup. The biosignal source is placed beneath the 2nd electrode. In this case, the electric field spreads in a uniform manner and more than one electrode receives almost equal field strength that reduces the selectivity of the electrode regarding the source location. In Figs 3(b), 3(c) and 3(d), the proposed electrode setup is considered with reference location on the field plates of the electrodes 2nd, 5th, and 8th, respectively. As seen in Fig 3(b) all the electric field is linked to the second electrode due to the close proximity of the source (source is beneath the 2nd electrode) and the reference location (reference is on the field plates of the 2nd electrode). This electric field crowding due to the presence of reference signal provides better selectivity compared to the conventional disc electrode setup with reference location far-away from the electrodes. For the identical source location (beneath the 2nd electrode) Figs 3c and 3(d) show the field crowding with the reference locations on the 5th and 8th electrodes, respectively. The relative strength of the electric field crowding and the corresponding induced potential in the sensing plate indicate the proximity of the biosignal source from that particular field-referenced electrode.

Figs 4 and 5 show comparison of induced potentials between the proposed electrode array and the conventional disc electrode array. In the first case (Fig. 4) the source is kept close to the 5th electrode ($0.6\ mm$ beneath the 5th electrode). For the disc electrode array, induced potentials are measured at the center of each electrode. For the proposed structure, the

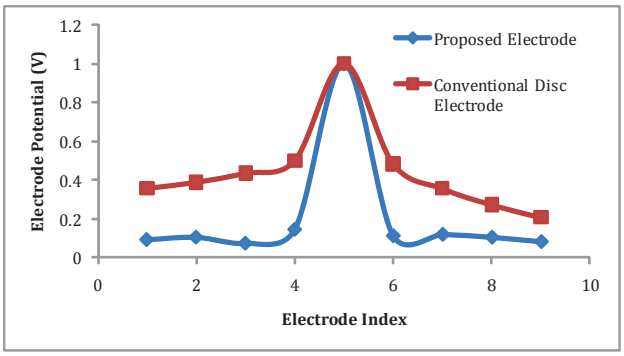


Fig. 4. Induced electrode potential with source position 0.6 mm down the 5th electrode.

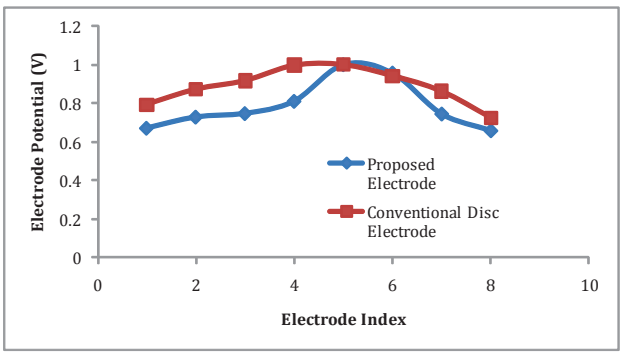


Fig. 5. Induced electrode potential with source position 3 cm down the 5th electrode.

reference is applied on the field plate of a single electrode at a time and corresponding center disc potential is measured. Since the source is closest to the 5th electrode, potential on the 5th disc has the highest value compared to the potentials on the other electrodes for both the structures (conventional and proposed), which can be seen from Fig. 4. However, the proposed structure demonstrates the higher selectivity to determine the location of the electrostatic source. In the second case, we move the source 3 cm down the 5th electrode. Fig. 5 shows that the induced potential curves for both the structure flatten because of the larger distance of the source from the electrode array. As we move the source from the closest to the farthest of the 5th electrode, the proposed electrode structure shows greater selectivity compared to the traditional one. For the comparison of selectivity, normalized induced potential values are used where the induced potential of each electrode is divided by the maximum of the induced potentials for that particular electrode setup. From Figs 4 and 5, the relatively smaller width of the induced potential curve for the proposed electrode indicates the superior selectivity feature.

V. Electrode Array Fabricated on PCB

A prototype electrode-array is fabricated on a printed circuit board and tested using a bench-top test setup. Fig. 6(a) shows a conventional electrode array used as the source electrode and Fig. 6(b) depicts the 4-by-4 field-addressable electrode array. For each electrode, the diameter of the center sensing disc is 3 mm and the annular field plate is 1 mm in width and 1 mm apart from the disc. The sensing discs are connected as row wise forming 4 sense lines and the annular rings of the electrodes are connected column wise forming 4 control

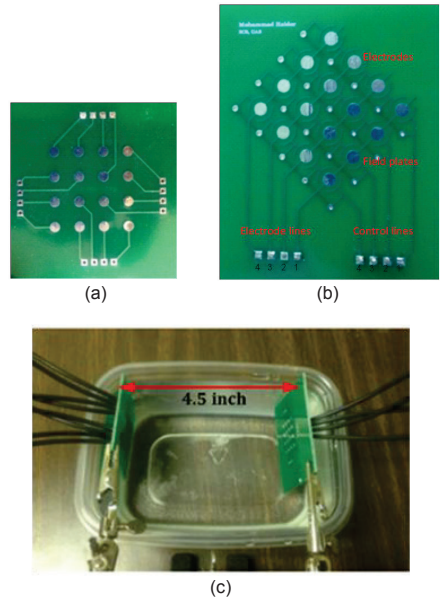


Fig. 6. Test setup (a) source electrode array, (b) proposed field-addressable electrode array, and (c) test setup where source and proposed electrode arrays are immersed in saline water and potentials are measured to determine the proximity of the source point.

lines as depicted in Fig 6(b). Fig. 6(c) shows the test setup. A single control line is activated at a time and potentials are measured along all the sense lines (1, 2, 3, and 4) and this process is repeated for all the control lines (1, 2, 3, and 4). Fig. 7 shows variation of measured electric potentials on different electrode sense lines (1, 2, and 3) with the variation of activating different control lines (1, 2, and 3). The measured results indicate the proximity of the sensing electrodes to the signal source. A zero potential is used as the control field and the induced surface potential is the lowest along the control field line. From Fig. 7, electrode with index 33 is the closest to the source. Measurement results show that the electrode array can be properly addressed by the control/reference electric field lines and the induced electric potential relates to the

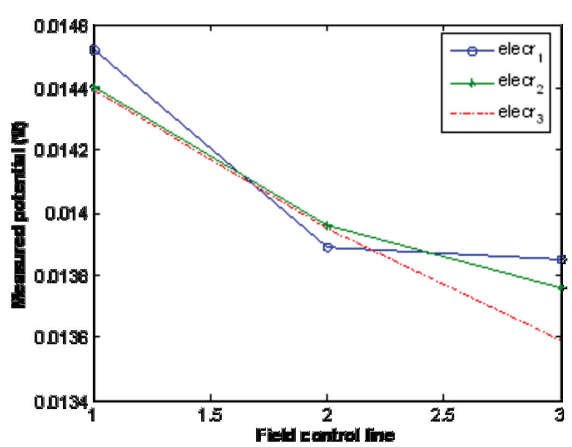


Fig. 7. Variation of measured electric potentials on different electrode sense lines with the variation of activating different control lines. Since the control voltage used is zero potential, the lower the induced potential the better is proximity towards the source electrode. Here electrode with index 33 is closest to the source.

proximity of the source.

VI. Conclusion

We propose a voltage referenced electrode structure for the determination of the proximity of an electrostatic source. The proposed electrode structure manifests a sensing plate and a field plate separated by a dielectric layer. The field plate is utilized to apply the reference bias to the electrode. The proposed electrode works based on the electric field crowding due to the presence of the reference signal. The relative strength of the electric field crowding indicates the proximity of the electrostatic source. The higher the electric field crowding the closer is the proximity of the biosignal source. The proposed structure has been simulated using a 2D finite Element simulation software package. Simulation results indicate the variation of electric field crowding with the variations of reference locations. The proposed structure has been compared with the conventional disc electrode structure and simulation results indicate the better selectivity of the proposed electrode structure under varying depths of the source locations. A prototype electrode array has been fabricated on PCB and the measured results validate the field-addressable feature of the proposed electrode array.

References

[1] J. Viventi, D. H. Kim, L. Vigeland, E. S. Frechette, J. A. Blamco, Y. S. Kim, A. E. Avrin, V. R. Tiruvadi, S. W. Hwang, A. C. Vanleer, D. F. Wulsin, K. Davis, C. E. Gelber, L. Palmer, J. V. der Spigel, J. Wu, J. Xiao, Y. Huang, D. Contreras, J. A. Rogers, and B. Litt, "Flexible, foldable, actively multiplexed, high-density electrode array for mapping brain activity in vivo," *Nature Neuroscience*, 2011 Nov 13;14(12):1599-605. doi: 10.1038/nn.2973.

[2] W. J. Freeman, "Spatial spectral analysis of human electrocorticograms including the alpha and gamma bands," *J. Neurosci. Methods*, 95, 111121 (2000).

[3] S. S. Kellis, P. A. House, K. E.Thomson, R. Brown, and B. Greger, "Human neocortical electrical activity recorded on nonpenetrating microwire arrays: applicability for neuroprostheses," *Neurosurg. Focus*, 27, E9 (2009).

[4] S. Kellis, K. Miller, K. Thomson, R. Brown, P. House, and B. Greger, "Decoding spoken words using local field potentials recorded from the cortical surface," *J. Neural Eng.* 7, 056007 (2010).

[5] P. K. Campbell, K. E. Jones, R. J.Huber, K. W. Horch, and R. A. Normann, "A silicon-based, three-dimensional neural interface: manufacturing processes for an intracortical electrode array," *IEEE Trans. Biomed. Eng.* 38, 758768 (1991).

[6] L. R. Hochberg, M. D. Serruya, G. M. Friehs, J. A. Mukand, M. Saleh, A. H. Caplan, A. Branner, D. Chen, R. D. Penn, and J. P. Donoghue, "Neuronal ensemble control of prosthetic devices by a human with tetraplegia," *Nature*, 442, 164171 (2006).

[7] S. I. Ryu, and K. V. Shenoy, "Human cortical prostheses: lost in translation?" *Neurosurg. Focus*, 27, E5 (2009).

[8] V. S. Polikov, P. A. Tresco, and W. M. Reichert, "Response of brain tissue to chronically implanted neural electrodes," *J. Neurosci. Methods*, 148, 118 (2005).

[9] S. Schmidt, K. Horch, and R. Normann, "Biocompatibility of silicon-based electrode arrays implanted in feline cortical tissue," *J. Biomed. Mater. Res.* 27, 13931399 (1993).

[10] R. W. Griffith and D. R. Humphrey, "Long-term gliosis around chronically implanted platinum electrodes in the Rhesus macaque motor cortex," *Neurosci. Lett.* 406, 8186 (2006).

[11] W. G. Besio, K. Koka, R. Akula, and W. Dai, "Tri-polar concentric ring electrode development for laplacian electroencephalography," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 5, pp. 926 – 933, May 2006.

Development of a Novel Prepaid Gas Metering System

Mohammad Imam Hasan Bin Asad,* Md. Liakot Ali, and Md. Saiful Islam

Institute of Information and Communication Technology (IICT)
Bangladesh University of Engineering and Technology (BUET), Dhaka-1000, Bangladesh
*ihasan@iict.buet.ac.bd

Abstract—**Energy saving by mitigating the waste of energy and save the world from global warming is a very common concern of this day. This paper presents the design and implementation of a prototype novel contactless smartcard based prepaid metering system for domestic purpose that technologically forces the customers to stop the misuses of natural gas. The proposed metering system addresses all the drawbacks of our previously developed and deployed 4500 meters in a pilot site at Dhaka. The functionality of the meter has been tested in the laboratory and demonstrated to the utility company (KGDCL). The proposed meter is more efficient in terms of power consumption, more secured and more user friendly in terms of vending the smart card and top-up the meter than those of its earlier version.**

Index Terms—**Prepaid metering system, Contactless smart card, NFC.**

I. Introduction

To minimize the waste of energy is a very common concern of this day all over the world since the wasted energy in turn produces carbon di oxide (CO_2) and sulphur di oxide (SO_2) and creates global warming [1-2]. One of the main resources of world's energy supply is natural gas. This is more accurate for Bangladesh where it is the most found natural resource. Gas is widely used in industrial, commercial, power, transportation and domestic sector. In most cases where consumption is higher gas meters are implemented in customer premises. But some times for domestic customers are not given any gas meter and fixed monthly rate is used for them. In this case customer pays a fixed amount for whatever the amount of gas he uses. Hence there is a high probability of misuses of gas. In most cases customers pay very little attention for the proper use of gas like many of them use gas for drying clothes, keeping room warm, using inefficient appliances etc. Again there are some dishonest customers who indulge in gas pilferage. And as for this system there is no way to know the actual uses for a particular customer, it would not be possible for the utility company to know whether there is any system loss in the customer end. Though the consumption of an individual is not high but as there are many customers in this sector. If wastage and misuse of gas can be prevented then a considerable amount of gas can be saved for future use. Again in the traditional meter and postpaid billing system utility company has to allocate manpower for meter reading and sending the bills. Prepaid gas metering system may be an effective solution for the problems mentioned above. In this case as the customer has to pay only for the amount he uses, he would be more concern of wastage of gas. The proposed system includes a 2-way dataflow system. Through this system utility companies can check the vending profile of customers along with the consumption pattern to identify any irregularity or problem at the meter end take necessary actions.

Some works have been done on smart meter and prepaid metering system [3-14]. In [8] consumption measurement is done optically which is not power efficient. Along with this the valve control system also consumes a considerable amount of power. These problems were solved any another work [9]. In 2010 smart card based prepaid metering system was designed and developed by IICT, BUET for TITAS Gas Transmission and Distribution Limited, one of the famous gas distribution company in Bangladesh. In most cases the meters were setup outside of house, hence, they were vulnerable to the weather. In that design contact smart card was used for setup and recharging the meter module. As the meter couldn't be sealed perfectly due to the card slot through which contact card was inserted, water and dust may enter easily to the meter. This may degrade the performance and damage the electrical module or part of it. The meter was run by four AA size battery so the power consumption was little bit higher. There were some other limitations also at the software end like each customer was assigned to a zone and he can only buy credit from a specific place or office of that zone. This restricts the flexibility of the customer. All these limitations are analysed and solved in the proposed design. Earlier version of this paper was presented in the conference 'Industrial Engineering, Operation and Maintenance (IEOM 2014)' [16]. Since the audiences of the conference were related to the industry and manufacturing, the different attractive features of the system and its potentiality in commercialization were highlighted there. So it is worth to explore the hardware and software aspects of the proposed prepaid metering system for another group of audiences who are related to electrical, electronic and computer engineering. The next sections of the paper describes the overall design architecture, hardware design, firmware design, software design, test results and conclusion.

II. Design Architecture

The prepaid metering system includes a 2-way dataflow process via contactless smart card. Fig. 1 depicts the block diagram of prepaid gas metering system.

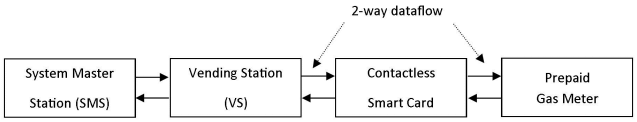


Fig. 1 Contactless smart card based prepaid metering system

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

According to the design the prepaid meters will be installed in customer premises, there will be necessary authorized vending stations all over the service area and the server will be installed in a secure place provided by the utility company. Vending Station (VS) is simply laptop or desktop computer with internet connection. Customers can purchase credit from vending stations. System Master Station (SMS) is the central application server along with database server and other necessary equipment necessary for security purpose where all information is processed.

Whenever a new customer requests gas line from a utility company, first a meter is assigned, set up and installed in his house. Then he is assigned a contactless customer card. With that card he can purchase credit from authorized vending stations and also recharge the meter whenever necessary. When the customer purchases credit, the credit information is written in the card. After that when he recharges the meter with the card, first the meter validates the card and then update the credit information and take action if necessary. Then the meter writes some important information about the meter to the card like consumption, max consumption rate and other health parameters. When the customer goes to the vending station for another purchase, the information stored in the card is uploaded to the central server. In this way utility company can monitor the consumption trend, health of the meter, check whether there is any fault at the meter end take necessary action.

III. HARDWARE DESIGN

For the overall design a G1.6 gas meter is used as base meter with protection of IP65 and our designed electronic board is attached to the meter. In the previous design IP56 base meter was used. The new meter is more suitable for the local weather condition. Fig. 2 shows the block diagram of the hardware design. For design and debugging purpose the overall design is divided in several modules. They are described in following sections.

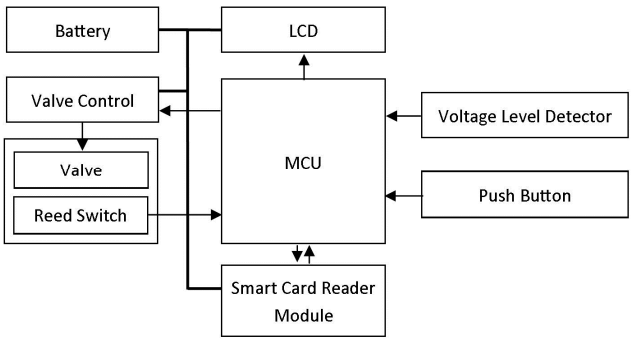


Fig 2. Block diagram of hardware design

A. MCU

The control unit of the design is the MCU module. For the design we use ATMEGA64A, a high-performance, low-power Atmel 8-bit AVR RISC-based microcontroller combines 64KB ISP flash memory with read-while-write capabilities, 2KB EEPROM, 4KB SRAM [15]. It has 53 general purpose I/O lines, 32 general purpose working registers, a real time counter, four flexible timer/counters with compare modes and PWM, two USARTs, a byte-oriented 2-wire serial interface, an 8-channel/10-bit A/D converter. It is a low power microcontroller which consumes low power both in active and sleep mode. This device can operate 2.2 - 5 V. We have designed out electronic module to be run on 3.5 V. In the previous design ATMEGA88 was used and the electronic board was designed for 5 V. Hence, the power requirement reduces for the proposed system.

B. Pulse sense module

The base meter is a gas meter with 8 mechanical wheels that shows the usage among which 3 wheels are for fractional digits. Along with the mechanical wheels it includes a pulse generating system with magnetically working reed switches. On the rightmost wheel a magnetic contact is fixed. Two other magnetic contacts are placed on the two sides of the wheel and connected to the reed switches. Whenever the wheel rotates and the magnetic contact comes near to any of magnetic contact of reed switches, an electronic pulse is generated with the corresponding reed switch. Those two switches are connected to the MCU with two interrupts. For each rotation of the rightmost wheel two electronic pulses are generated alternatively. The consumption is calculated from the received pulses.

C. Valve control module

The base meter also includes a DC brush motor to control its valve. The brush motor is connected to the MCU through a motor control module. To control the motor an H-Bridge driver IC Sanyo LB1836M is used [16]. MCU sends signal to the LB1836 for rotating the motor to the desired direction to control the gas flow.

D. Display module

A display module with SED1335 is used for displaying necessary information like consumption, remaining balance, recharge information, running condition, error etc. To minimize power requirement in normal condition the display module is kept off. Whenever customer press a button mounted on the board it is turned on for a short period of time and it slides necessary information for customer. It also includes icon for valve whether it is closed or open, battery condition and other operational information. If customer doesn't press the button again within a predefined time it is turned off again. It also shows recharge or setup information while recharging or setting up the meter for confirmation of the corresponding processes.

E. Smart card reader module

A smart card reader module is used to communicate with the contactless smart card. The module uses UART protocol to communicate with the MCU. The prepaid system is designed with Near Field Communication (NFC) standard to communicate with the card from meter and vending station [17]. It is based on Radio Frequency Identification (RFID) standards including ISO/IEC 14443. In the design Mifare 1k card is used with the range of 6 inches [18]. It uses contactless transmission for data and energy and no battery is required for that. Each card has a unique serial number and uses mutual three pass authentication (ISO/IEC DIS9798-2). Unlike the display module the smart card or NFC module is kept off for saving power. When the push button is pressed the module becomes active and searches for valid NFC card along with displaying information. If the module finds a valid card then it stops searching and the MCU performs the operation defined in the card. In case of customer card it only updates the credit and billing rate. Some other card types are also defined to set up or to change meter parameter. Previously contact smart card was used and as the insertion

point of the card in the meter, it was vulnerable to the weather. By using contactless smart card we can eliminate the probability and the security of the card is also stronger than the previous one.

F. Voltage level detector module

A voltage level indicator is included in the design. It detects the voltage level and whenever the voltage goes below a tolerable range, MCU saves all information to the EEPROM, lose the valve and shut down normal operation. This module serves another important operation in case of sudden power down i.e. battery is removed from the electronic board.

G. Power module

The whole circuit is designed to run with 3.5 V. To achieve that 3 AA size batteries are used in the batter module which are available anywhere all over the world.

IV. FIRMWARE DESIGN

The firmware is designed and implemented in two layers. The bottom layer is the hardware layer which consists all individual module drivers. The upper layer is the main firmware layer where operational algorithm is implemented. Fig. 3 shows the simplified firmware design.

At the startup MCU loads necessary information from the EEPROM and initializes the meter and peripherals. Then it goes to SLEEP mode to reduce the power consumption. It discontinues its SLEEP mode if any interrupt occurs. Three interrupt types are defined in the hardware: pulse interrupt, push button interrupt and low voltage interrupt. Whenever MCU gets an interrupt it goes to ACTIVE mode and take necessary action depending upon the interrupt and then goes back to the SLEEP mode again.

When customer uses gas the wheel of the meter rotates and MCU gets pulse to its interrupts. For each rotation of the rightmost wheel MCU gets two interrupts. Sometime reed switches may miss one or two pulses. An algorithm is implemented to calculate the usage from the alternate pulses considering the misses. Again sometime reed switch becomes faulty and in that case it is discarded from calculation. After detecting pulse MCU counts usage and deducts credit depending upon the active gas rate. Then it updates the valve status. If balance goes down below a certain level MCU shuts down the valve. In normal condition after getting one pulse if MCU doesn't get any pulse within a certain period of time, it goes to the SLEEP mode.

If the push button is pressed MCU goes to the ACTIVE mode. It activates the display module and slides necessary information like consumption, remaining balance, active billing rate and displays operational icons for a certain period of time. In the meantime MCU searches for valid NFC card. If MCU gets any valid NFC card, it shows a CARD BUSY icon. MCU checks the card type and takes certain action depending upon the card type like recharging credit or changing parameter. Then MCU uploads necessary information like gas usage, health condition. MCU also displays corresponding information in the display module. After successful operation of the card MCU shows REMOVE CARD to indicate that the customer can remove the card. . If MCU doesn't get any valid card or customer doesn't press the button within predefined time, MCU goes back to the SLEEP mode again.

If any low voltage is detected either by battery dying out or sudden removal of battery the MCU wakes up from SLEEP mode, saves all information to the EEPROM, shuts down the valve and stops all normal operation. The MCU includes two LEDs for notification purpose. Most common usage will be for balance condition. Multiple level of balances are defined in the design. If the credit goes down below a certain level then an indicator LED will be blinking to notify the customer. Another important feature in the design is the negative credit. In this case the meter can run even in the negative balance up to a negative balance threshold set by the utility company. This feature is very useful if the credit runs out after office hour or in holidays when customer can't buy credit from vending stations. The meter will run until the balance reaches the negative balance then it will shut down the valve and restrict the gas flow. In case of balance run out or any error like reed switch fault or memory fault the LED indicators are used to notify the customer.

V. SOFTWARE DESIGN

The software of the prepaid system consists two parts: 1. SMS which runs an application that communicates with the vending stations and the database 2. Vending station where customer purchases credit. For redundancy a backup database is also included which automatically syncs with the main database. The vending stations may also run in offline mode i.e. without internet connection. This feature is introduced because sometimes the internet connection becomes unstable due to optical fiber cut or other reasons. In that case whenever the vending station goes online all offline transaction are uploaded to the main database. The application server is a windows server and the central and the backup databases are MySQL on linux machines. Fig. 3 shows the vending window of the vending software running on windows platform.

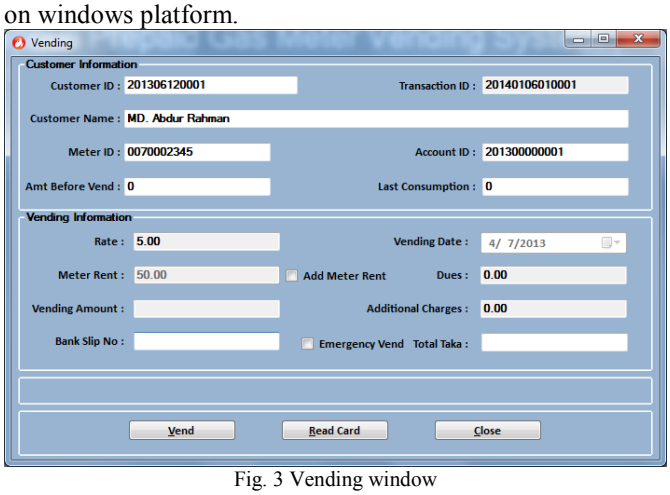


Fig. 3 Vending window

VI. TEST RESULTS OF THE METER

We have implemented our design in our laboratory. We have chosen a gas meter from many which is suitable for our environment and pressure condition of the local utility companies. We have attached out designed board to the meter. The software part was implemented with a central server with backup facility and multiple vending stations. We tested our gas meter and software with all necessary operations regarding prepaid metering system like setting up a meter, card assignment, buying credit from vending station, recharging the card and lastly uploading meter information to

the SMS through contactless smart card. Fig. 4 shows the designed board.

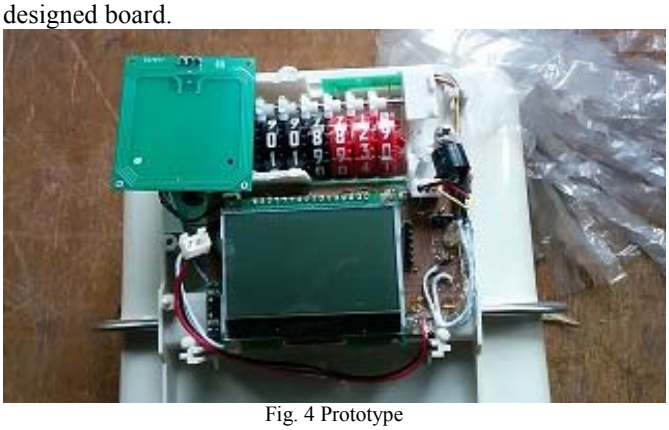
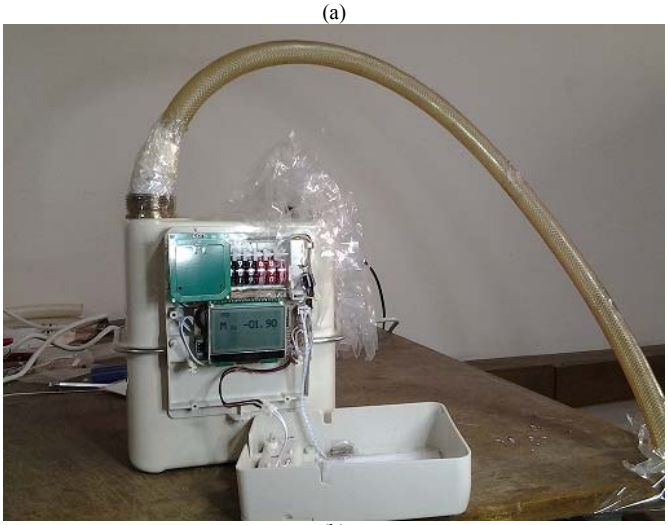


Fig. 4 Prototype

Fig. 5 shows the operation of the meter.

(a)



(b)

Fig 5. Operation in (a) normal condition (b) negative balance

VII. CONCLUSION

In this paper a novel prepaid gas metering system is presented. The concept and design of the metering system is briefly discussed. The design was implemented and the prototype was tested successfully. A demonstration was given in KGDCL regarding the metering system with its different features from which both company and customers may be benefitted. After demonstration KGDCL showed interest to install our designed meter for 20000 domestic customers. There are some places where we can do some improvement. In the next step we will introduce a simple portable vending machine based on embedded system. This will replace the costly laptop or desktop. In that case along with the cost issue utility company and customers may enjoy more flexibility. As nowadays NFC is becoming more popular in the cell phones, we can build smart phone application though which customer may buy credit and also recharge the meter. This will make the system more flexible.

ACKNOWLEDGMENT

The authors gratefully acknowledge to Bangladesh University of Engineering and Technology (BUET) for allowing us to conduct the proposed research using all the facilities in the Centre of Embedded System, Institute of Information and Communication Technology (IICT). We are also grateful to Karnaphuli Gas Distribution Company Limited (KGDCL) for funding 4 lakh taka to develop the system.

REFERENCES

[1] (2014) Efficient energy use on Wikipedia. [Online]. Available: http://en.wikipedia.org/wiki/Efficient_energy_use

[2] (2014) Green Technology website. [Online]. Available: www.green-technology.org/

[3] S. K. Ali Zaidi, H. Masroor, S. R. Ashraf, and A. Hassan, "Design and Implementation of Low Cost Electronic Prepaid Energy Meter," IEEE International Multitopic Conference, pp. 548 – 552, 2008.

[4] N. Anglani, E. Bassi, F. Benzi, L. Frosini, and T. Traino, "Energy smart meters integration in favor of the end user," International Conference on Smart Measurements for Future Grids, pp. 16-21, 2011.

[5] J. Cowburn, "Paying for energy the smart way," IEEE Review, Volume. 47, Issue. 4, pp. 17-20, July, 2001.

[6] G. Deconinck, B. Delvaux, K. De Craemer, Z. Qiu, and R. Belmans, "Smart meters from the angles of consumer protection and public service obligations," International Conference on Intelligent System Application to Power Systems, pp. 1-6, 2011.

[7] L. Gavra, M. S. Crainic, P. Pupsa, and G. Popa, "Residential smart gas meters," International Symposium on Electronics and Telecommunications, pp. 37 – 40, 2012.

[8] M. F. Khan, A. Zoha, and R. L. Ali, "Design and Implementation of Smart Billing and Automated Meter Reading System for Utility Gas," International Conference on Information and Emerging Technologies, pp. 1-6, 2007.

[9] T. H. Khan, T. K. Paul, G. M. Shahabuddin, K. Wahid, A. H. Chowdhury, S. M. L. Kabir, "Towards design of a smart prepaid gas metering system," International Conferenece on Innovations in Information Technology, pp. 55-59, 2009.

[10] B. H. Kwan, and M. Moghavvemi, "PIC based smart card prepayment system," Student Conference on Research and Development, pp. 440 – 443, 2002.

[11] M. McConnachie, "Prepayment advances with smart-card solutions that cover electricity, gas and water," *International Conference on Metering and Tariffs for Energy Supply*, pp. 248 – 250, 1999.

[12] F. MacDonald, "Smart metering: The importance of creating a shared vision - start with a clear view of the end game," IET Conference Publications, pp. 1-15, 2009.

[13] M. S. Parvin, and S. M. L. Kabir, "Standard and interoperable database for pre-paid electricity metering systems in Bangladesh," International Conference on Informatics, Electronics & Vision, pp. 356 – 361, 2012.

[14] M. L. Ali, M. I. H. B. Asad, and M. S. Islam, "Contactless Smart Card Based Prepaid Gas Metering System," Proceedings of the 2014 International Conference on Industrial Engineering and Operations Management (IEOM 2014), Jan 7-9, 2014.

[15] "ATMEGA64 data sheet," Atmel Corporation, USA.

[16] (2014) H Bridge on Wikipedia. [Online]. Available: http://en.wikipedia.org/wiki/H_bridge

[17] (2014) NFC on Wikipedia. [Online]. Available: http://en.wikipedia.org/wiki/Near_field_communication

[18] (2014) Mifare 1k datasheet. [Online]. Available: http://www.mifare.net/products/mifare-smartcard-ic-s/mifare-1k/

Matching of Artificial Limb with Residual Human Limb by Capacitance Measurement Technique

Ashraful Haque*, Sheikh Jawad Ilham, Q. Ahsan
Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology, Dhaka–1205, Bangladesh
Email: *ashrafulhaque99@gmail.com

***Abstract*—Prosthetic amputee rehabilitation is a very common phenomenon in medical science. Artificial limb used as a substitution of a missing body part to restore its function is generally known as prosthesis. The most rudimentary requirement is to design the artificial limb such that it appears as natural as the missing body part in terms of functionality and comfort. A method of designing artificial limb is proposed after a detailed analysis of the existing prosthesis mechanisms, their evolution and drawbacks. The proposed technique presents a prosthetic replacement essentially with a complete negative impression of the residual limb and an exact matching customarily at the contact surface. The technique exploits the capability of measuring the capacitance in the range of picofarad (10^{-12} Farad) to acquire varying surface data of the residual limb for its exact matching with the artificial part. A micro-controller based capacitance measuring equipment is developed and it is applied to measure the unevenness of a surface. The accuracy is validated through image processing.**

***Index Terms*—Artificial human limb, Micro-controller based capacitance meter, Image processing.**

I. Introduction

Lower extremity amputation is a widespread problem prevailing at regions with higher rate of road accidents, natural disasters like earthquakes, cyclones and by wars. Amputee rehabilitation can be accomplished by an artificial limb fitted with the residual part to achieve cosmetic and functional capabilities to the prior level. Major concerns of rehabilitation process are to ensure that, the prosthetic limb appears as natural as missing body part irrespective of upper or lower extremity amputation and at the same time restitution of functionality with proper attachment to the residual limb.

A socket holds the prosthetic limb in place with the residual limb. The socket is considered as a major element of a artificial human limb. Socket must be carefully designed to fit the unique geometry of the residual limb. Existing technologies use different kinds of sockets to hold the artificial human limb in place with the residual body part. Existing technologies include: Patellar Tendon Bearing (PTB) Socket, Total Surface Bearing (TSB) Socket, and Hydro-static Weight Bearing (HWB) Socket. Fist two of them are shown in Fig. 1.

The residual limb can take any random shape after amputation illustrated at Fig. 2. On the other hand almost constant shapes of existing sockets result into mismatch of socket with the residual limb. The drawbacks of the existing technologies mainly arise from two phenomena with a common root of mismatch of a socket with the residual limb.

Again amputee using conventional passive prosthesis generates problems during locomotion. While walking, there is an angle between knee and toe [1], [2]. If the socket is not exactly matched it will create pressure on the contact surface. L. E. Pezzin et. al. [3] shows that nearly one third of the person using prosthetic limb devices expresses dissatisfaction with the comfort of their prosthetic limb. They are not comfortable with the fitting of limb with their body part while walking or standing. If the contact surface of the prosthetic limb does not exactly matches with the residual limb it creates pressure on the residual body part and creates discomfort. As pressure increases discomfort increases and this pressure can also cause serious damage to the soft tissues of residual body part [4]. The paper propounds a method for acquiring 3-D image of the negative impression of the residual limb and that can be incorporated as the top surface of the socket to the residual limb to eliminate almost all the problems arisen from gaps. By

Fig. 1. PTB and TSB Socket.

(a) Stump 1

(b) Stump 2

(c) Stump 3

Fig. 2. Trans-tibial Amputees with Different Stump Shapes.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

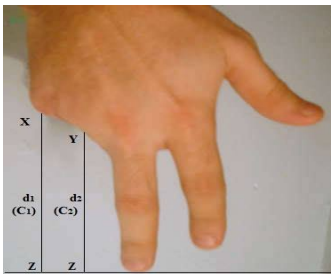


Fig. 3. Translation of Capacitance into Distance.

knowing the accurate force profile the exactly matched contact surface can be synthesized with localized pressure zoning system to be commensurate with force carrying capabilities of the stump for the removal of non uniform pressure related problems.

Traditionally, design of these sockets is done manually and plaster of Paris is used to capture the geometry of residual limb. Early in the 1970, VAPC knee caliper was used for anterior-to-posterior knee measurement and after placing socket in place, heat was used to modify the socket for better adjustment of contact area of artificial limb and residual limb [5]. But this process has serious error because of manual measurement and because of use of heat for exact matching of limb. Excessive heat can damage the socket and cause mismatch in the contact surface. Sand casting techniques were later introduced for designing of sockets [6]. In this technique residual limb is dressed with socks and inserted into liquidized silica with vacuum applied. After limb is totally inserted into the silica and taken out, vacuum is released and a mold of residual limb is thus formed. This mold is then used to design the sockets. However, if the surface area of the limb is considerably high then this technique will be unable to provide satisfactory result, as it could not provide accurate geometry of the uneven surface. It requires a complete new mold to be made every time there is a mismatch of socket and body part. Recent technologies include the use of Rapid Prototyping (RP) technology that creates physical models directly from three-dimensional data received from the computer [7], [8]. This technology is also known as 3-D printing [7]. But it requires the use of a 3-D printer which is still quite expensive and as a result the pricing of sockets increases. The consistency of these sockets is yet to be studied [7].

This paper proposes the designing of an artificial limb based on the capacitance measurement technique for matching of the artificial human limb with the residual body part. Proposed method does not require any manual measurement and also it eliminates the use of mold and plaster of Paris for making sockets. This paper also proposes a method for error checking while designing artificial human limb through image processing. Therefore, it eliminates the necessity of developing more than one socket for final matching of an artificial limb.

II. Methodology

Body part where artificial human limb is needed to be attached is, generally, an uneven surface. If a reference line is picked, each point of the surface has different distance from that. Fig. 3 shows the difference of only two points of the surface of the body part from a certain reference line as an example. Similarly, each point on the surface has different distance from the reference line.

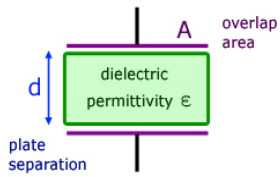


Fig. 4. Parallel Plate Capacitor.

Precise calculation is needed to measure these d_1, d_2 distances. As the difference between these points is not too large, conventional methods for measuring d_1, d_2 is not applicable here. More precise calculation is required for this measurement. The capacitance of a parallel-plate capacitor of Fig. 4 constructed of two parallel plates both of area A separated by a distance d may be expressed as,

$$C = \frac{\epsilon_0 \times \epsilon_r \times A}{d} \tag{1}$$

Where,
C = Capacitance in farads
A = Area of overlap of the two plates
ϵ_r = Dielectric constant of the material between the plates
ϵ_0 = Dielectric constant of vacuum
d = Separation between the two plates in meters.

Capacitance is proportional to the area of overlap and inversely proportional to the separation between conducting sheets. The closer the sheets are, the greater the capacitance. By rearranging (1) it can be written as,

$$d = \frac{\epsilon_0 \times \epsilon_r \times A}{C} \tag{2}$$

Now if the one plate at X and one plate at Z are placed it results C_1 and for one plate at Y and one plate at Z the resulting capacitance is C_2. From (2), d_1 and d_2 are obtained.

Similarly by placing one plate at different points on the limb and one plate on the reference line different capacitance values can be measured and in turn the distance of each point of the residual body part from a reference line, may be obtained.

This methodology can also be applied to a 3-D object where XY plane will be taken as reference plane and distances of different points of the surface will be calculated in a similar way.If the 3-D surface pattern of the residual body part is as Fig. 5 then through capacitance sensing method the distances d_1, d_2, d_3 etc. can be measured to reconstruct the 3-D negative impression of the residual body part.

Since, the distance of each point along the surface of the residual body part varies very slightly, it is very difficult to calculate the distance of each point through conventional measurement technique. Equation (1) shows that there is an inverse relation between capacitance and distance between parallel plates. So even if the difference of distance from one point to the next point is very small the change of the value of capacitance will be easily distinguishable. This is the main advantage of proposed technique that is, accurately measured capacitance even in the range of picofarad can easily be translated into distance of each point of residual body part from a reference point. These distances can be then used to generate an accurate negative image of the residual body part. This negative image will then be used to make exactly matched artificial body parts.

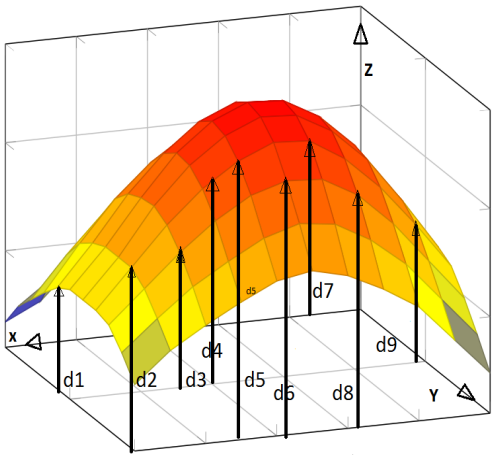


Fig. 5. 3-D Image Construction.

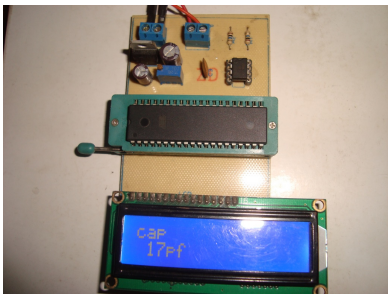


Fig. 6. Developed Capacitance Sensing Device.

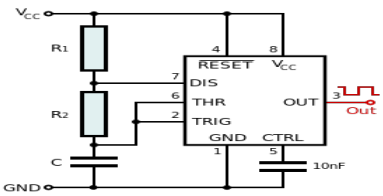


Fig. 7. Astable Multivibrator Circuit.

III. Sensor Design

Existing capacitance meter uses charging and discharging principle of capacitor to measure capacitance [9]. But charging-discharging technique can not distinguish between small amount of capacitance change. A prototype capacitance meter is developed, shown at Fig. 6.

For the capacitance meter 555 timer IC, ATmega16 microcontroller, LCD display, 7805 IC for providing power to microcontroller and LCD display are used. If the capacitance formed by different points of surface of the body is connected to the 555 timer IC in the following manner it forms an astable multivibrator shown at Fig. 7.

The output is a pulse whose frequency is F.

$$F = \frac{1}{ln(2) \times C \times (R_1 + 2 \times R_2)} \tag{3}$$

Here, C is the capacitance that is formed by placing one plate at a reference point and one plate at the surface of the body part.

This output pulse is fed to the microprocessor. Microprocessor determines the capacitance and translates these values to distances.

TABLE I
Error Percentage Calculation of Distances Measured By the Device with Distances Acquired from Image Processing

Reading No.	C_Standard (pF)	C_Measured (pF)	Error (%)
1	15	15	0
2	22	21	4.5
3	30	29	3.33
4	45	43	4.4
5	47	45	4.25
6	60	58	3.34
7	100	98	2
8	120	117	2.5
9	150	147	2

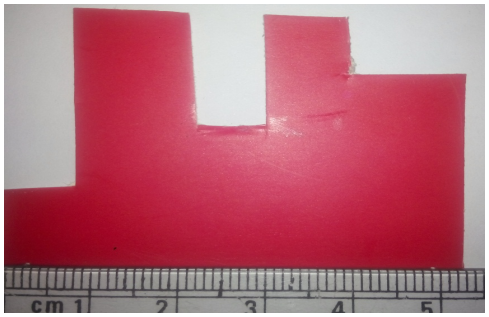


Fig. 8. Captured Image with Centimeter Scale

IV. Data Collection

Readings of different valued standard capacitors were taken using the developed capacitance meter. Table I shows the actual capacitance, measured capacitance and error. The table clearly shows that the error is within 5 percent of the standard value.

V. Result

Distances at different points corresponding to a reference point are determined using the developed capacitance meter. Validation of these measured values are done by image processing in several steps.

- First the image of the object is taken with a centimeter scale lying on the same plane with the intended surface of the object illustrated at Fig. 8.
- The image is cropped properly to get real distance per pixel of the image of that particular surface. The image is then read and coded into grey scale using MATLAB built-in function. Then the edge of that edge detected image is acquired using several MATLAB built-in functions and the best suited one is taken from them to be read into a matrix as white dots in black background with elements either 1 or 0 illustrated at Fig. 9 with inverted color.
- Then from the reference row the index of the last white dot of each individual column is read and plotted which gives the exact pixel distance of that particular point from the reference line illustrated at Fig.10.
- By using real distance per pixel constant obtained earlier, the scale of that plot is changed into real distance. Thus the real distance is measured from the image of the object shown at Fig. 11.

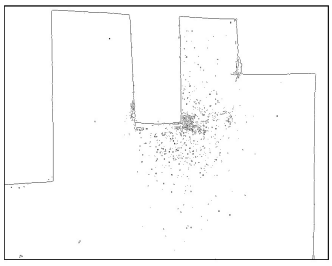


Fig. 9. Edge Detected Image by Prewitt Method.

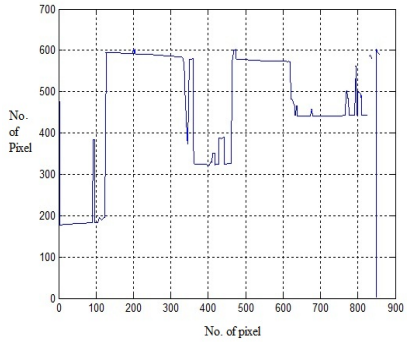


Fig. 10. Reconstructed Image in Pixel Distance.

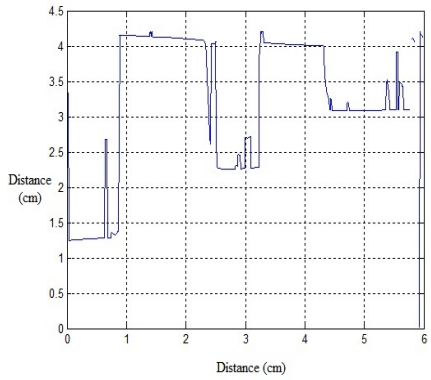


Fig. 11. Reconstructed Image in Real Distance.

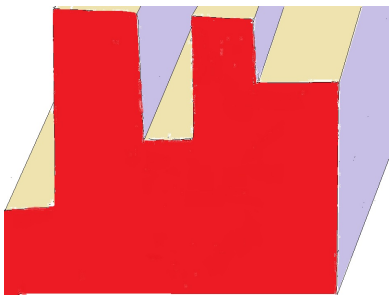


Fig. 12. 3-D view of Sample Object.

- This validation process can also be implemented in case of a 3-D object illustrated at Fig. 12, where individual images of different surfaces with same arrangement, stated earlier, are recorded.

TABLE II
ERROR PERCENTAGE CALCULATION OF DISTANCES MEASURED BY THE DEVICE WITH DISTANCES ACQUIRED FROM IMAGE PROCESSING

Reading No.	d_Actual (cm)	d_Measured (cm)	Error (%)
1	1.2	1.05	5
2	3.7	3.8	2.7
3	2.2	2.25	2.3
4	3.8	4	5.3
5	2.5	2.6	4

The actual distances of the edges are compared with the actual data from the image processing in Table II. The table clearly shows that the difference is not significant.

VI. CONCLUSIONS

Prosthetic amputee rehabilitation success depends profoundly on proper designing of contact surface which is the fundamental part of an artificial limb having a direct contact with the residual limb. Without proper fitting, overall function of the prosthesis as well as comfort to the patient is hampered. Thus it is imperative that, technique used here, to produce limb sockets is accurate, repeatable, and cost-effective and has high patient satisfaction. The unevenness of surface can vary from picometers to several inches. Therefore, precise determination is needed. Exactly matched surface with proper force profile is the perquisite for restoring cosmetic and functional purposes of an artificial human limb. By taking into consideration, all the limitations and drawbacks of existing technologies, this paper proposes a capacitance measurement technique along with the validation through image processing for error checking for matching of artificial human limb.

REFERENCES

[1] S. K. Au, J. Weber, and H. Herr, "Biomechanical design of a powered ankle-foot prosthesis," *IEEE Transactions on Robotics*, vol. 25, pp. 51–66, 2009.

[2] A. H. Hansen, D. S. Childress, S. C. Miff, S. A. Gard, and K. P. Mesplay, "The human ankle during walking : implications for design of biometric ankle prostheses," *Journal of Biomechanics*, vol. 37, pp. 1467–1474, 2004.

[3] L. E. Pezzin, T. R. Dillingham, E. J. Mackenzie, P. Ephraim, and P. Rossback, "Use and satisfaction with prosthetic limb devices and related services," *American Journal of Physical Medicine and Rehabilitation*, vol. 80, pp. 563–571, 2001.

[4] E. S. Neumann, "Measurement of socket discomfort-part i: Pressure sensation," *Journal of Prosthetics and Orthotic*, vol. 13, pp. 99–110, 2001.

[5] A. Staros and H. F. Gardner, "Direct forming of below-knee ptb sockets with a thermoplastic material," *Artificial Limbs*, vol. 14, pp. 57–64, 1970.

[6] J. S. Jensen, P. Poetsma, and N. Thanh, "Sand-casting technique for trans tribal prostheses," *Prosthetics and Orthotics International*, vol. 29, pp. 165–175, 2005.

[7] N. Herbert, D. Simpson, W. D. Spence, and W. Ion, "A preliminary investigation into the development of 3-d printing prosthetic sockets," *Journal of Rehabilitation and Research and developmen*, vol. 40, pp. 141–146, 2005.

[8] T. Krouskop, A. Muilenberg, D. Doughtery, and D. Winningham, "A preliminary investigation into the development of 3-d printing prosthetic sockets," *Journal of Rehabilitation and Research and developmen*, vol. 24, pp. 31–38, 1987.

[9] M. A. Kumar, B. R. Murthy, C. V. Ramana, V. S. Kumar, M. Moodley, and K. V. Madhav, "Microcontroller based capacitance meter," *International Journal of Computer Application and Engineering Technology*, vol. 1, pp. 41–48, 2012.

On Site Oil Test for a Transformer Using Sweep Frequency Response Analysis

Al-Amin Ahmed Simon, Md. Mahbubul Huq Riad, Erfan Ahmed and S. Shahnawaz Ahmed

Department of Electrical and Electronic Engineering,
Bangladesh University of Engineering and Technology
Dhaka-1000, Bangladesh
alamin067@gmail.com

***Abstract*— Oil filled transformers are widely used to step up or step down voltages in power transmission and distribution systems. A periodic maintenance of the transformers includes checking oil level, chemical analysis and dielectric strength (break down voltage) test taking oil sample and sending it to laboratory. Sometimes it becomes impossible to check the oil level without opening the transformer. In this paper the method of sweep frequency response analysis (SFRA) has been proposed for on-site and instant test of oil condition without requiring the transformer to be opened and sending oil sample to a remote laboratory. Notably SFRA has long been used to diagnose transformer winding faults such as mechanically displaced and shorted or open turns. So this paper proposes application of SFRA for the first time ever for spot monitoring of certain conditions such as oil level and contamination in emergency situations. The experimental results obtained from SFRA of a healthy transformer and two transformers having oil related problems were compared. In this way a guideline regarding the frequency range where a deviation from the response of the healthy transformer may suggest oil related problem, has been formulated. The conclusion has also been supported through a PSpice simulation.**

***Index Terms*—Frequency response, Transformer oil, On site oil monitoring**

I. INTRODUCTION

Basically in sweep frequency response analysis (SFRA) a frequency sweep of sinusoidal signals with low amplitude (less than 10V) and commonly in the range of 50 Hz to 1 MHz is injected [1] to a transformer's terminals. For each frequency the injected signal is measured and the output response is measured at another terminal. Then the response is plotted against the entire frequency range in terms of 20 $\log_{10}(V_{out}/V_{in})$ dB (decibels) along the Y-axis and frequency in Hz (Hertz) along the X-axis (in logarithmic scale). The plot is compared with that for a healthy transformer and the results are interpreted [1-3] depending upon the range of frequencies where the deviation occurs.

However, it appears that while the guidelines [1-3] for interpreting the mechanical deformations and electrical faults in windings using SFRA have been reported, no application of SFRA for monitoring the oil conditions such as reduced oil level and oil contamination has ever been reported. Till now oil level and contaminations are cheeked respectively through physical inspection (if a conservator is provided) and dielectric test of collected sample. Sometimes it becomes impossible to check the oil level without opening the transformer.

In this paper a new application of SFRA has been proposed for onsite monitoring of oil level and contamination without collecting oil sample and sending it to a laboratory. For this three 11/0.4 kV oil type transformers respectively a 250 kVA healthy one, another 250 kVA transformer with almost empty tank and a 500 kVA transformer containing moisture contaminated oil were subjected to SFR analysis. In this way a guideline has been formulated regarding the frequency range where a deviation from the response of the healthy transformer may suggest oil related problem. An attempt has also been made for PSpice simulation to support the findings.

II. CONNECTION FOR OIL TEST BY SFR

SFR test can be applied in various ways depending upon the purposes. For winding related faults the connection [1-3] is as in Fig.1.

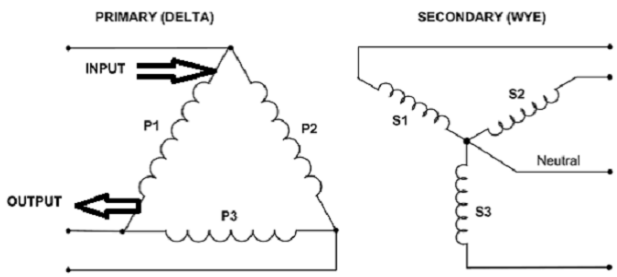


Fig. 1 End to end connection for winding related fault diagnosis

But for SFR test of oil that exists in between the high voltage and low voltage windings, the connection in Fig. 2 has been used in this work. Here for a Δ-Y transformer the input voltage is applied between a high voltage (HV) winding and the neutral of the low voltage (LV) windings and the output voltage is measured between corresponding LV winding and the neutral of LV windings.

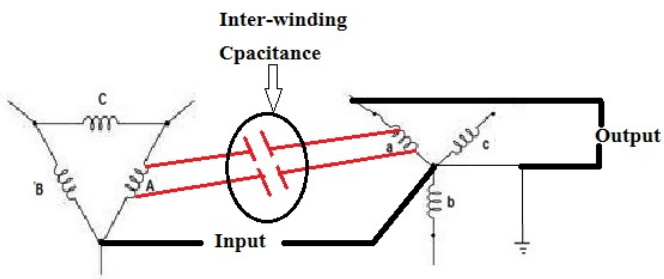


Fig. 2 End to neutral connection for proposed oil related fault diagnosis

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

III. Comparison between Healthy Transformer and a Transformer with Reduced Oil Level

SFR test was done in the range of 100 Hz to 1 MHz using the connection shown in Fig. 2 for the phase 'A' winding and its corresponding LV winding for a 250kVA transformer which was known to have its tank empty of oil. The same was done for another healthy transformer of 250 kVA size. The SFR plots for the two are compared in Fig. 3. The analysis of the plots shows that the deviation from the healthy one's response due to reduced (empty) oil level is confined mostly within the 60 kHz to 300 kHz range.

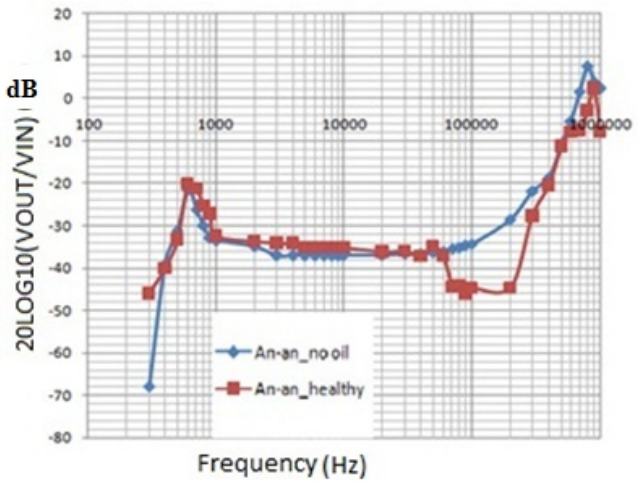


Fig.3 Comparison of SFR plots for two 250 kVA transformers (one healthy and the other with empty tank) using end to neutral connection

Since it was known that both transformers are free from winding related faults so to support the Fig. 3 response due to reduced oil level, the same two transformers were subjected to SFR test using end to end connection setup of Fig.1. The response plots are shown in Fig. 4 which clearly shows that the deviation between the two responses is uniform throughout the frequency spectrum and thus excludes the possibility of winding faults. This end to end test was repeated for every other winding of the two transformers and the responses were found to be similar as in Fig. 4.

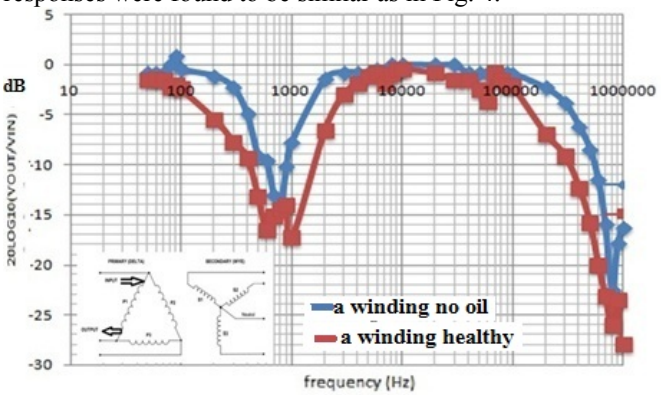


Fig.4 Comparison of SFR plots for the two 250 kVA transformers (one healthy and the other with empty tank) using end to end connection

IV. Comparison BETWEEN HEALTHY transformer and a Transformer with Contaminated Oil

SFR test was done in the range of 100 Hz to 1 MHz using the connection shown in Fig. 2 for the phase 'B' winding and its corresponding LV winding for a 500 kVA transformer which was known to have its oil contaminated by moisture. The same was done for the healthy transformer of 250 kVA size. The SFR plots for the two are compared in Fig. 5. The analysis of the plots shows that the deviation from the healthy one's response due to oil contamination is confined also within the 60 kHz to 300 kHz range.

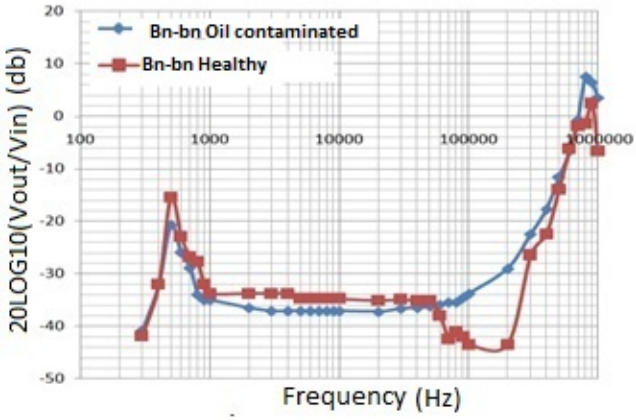


Fig.5: Comparison of SFR plots for a 250 kVA healthy transformer and a 500 kVA transformer with contaminated oil using end to neutral connection

As before the SFR test was also done for these transformers using end to end connection and the responses plotted in Fig. 6 which shows a uniform deviation throughout thre frequency spectrum excluding the possibilities of winding related problems.

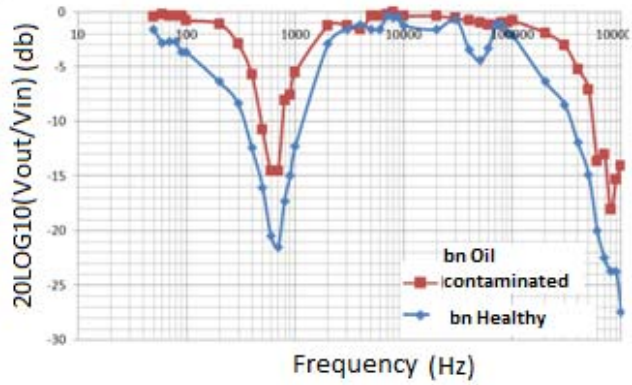


Fig.6 Comparison of SFR plots for the 250 kVA healthy and the 500 kVA transformer with contaminated oil using end to end connection

It should be noted that when transformer oil is contaminated, or the oil level is reduced the capacitance between two sides (HV and LV) reduces. This is because $C = \varepsilon A/d$ so that with the reduced oil level or contamination in oil the permittivity ε as well as the capacitance decreases. Consequently the ratio of output to input voltage increases and the SFR curves move upward relative to the healthy transformer response curves in the range 60 kHz and above as observed in Figs. 3 and 5.

V. PSpice Circuit Simulation

The equivalent circuit for a turn [3] of a transformer in the sweep frequency range has been used in this work to synthesize a circuit representing 10 turns in a winding of the Δ side while 5 turns in a winding of the Y side of a generic transformer. The inter-winding capacitances with typical

values have been added in the present work to represent oil conditions. Then a PSpice simulation was done for the input –output response of this circuit in 10 Hz to 1 MHz range considering end to neutral connection.

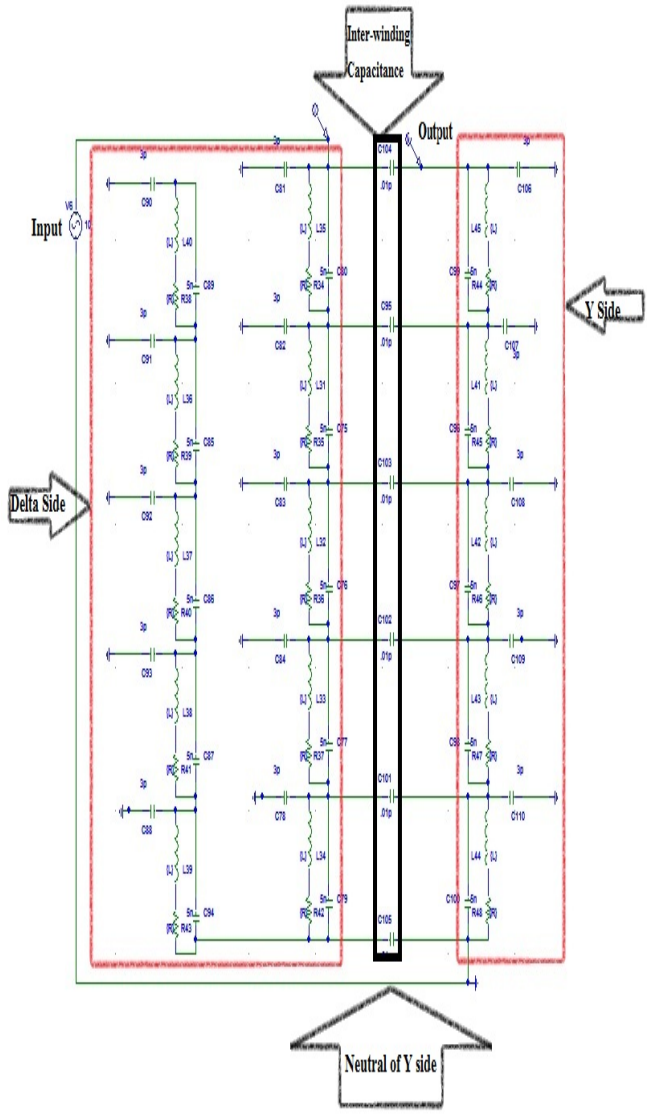


Fig.7 Equivalent circuit for PSpice simulation

Fig. 8 shows the SFR curves obtained from PSpice simulation for various oil conditions which clearly supports one of the the experimental finding that more the oil related problem more is the upward shift in the SFR. However, exact matching of simulation and experimental response curves is not expected. This is because it is difficult to synthesize an equivalent circuit considering actual number of turns in HV and LV windings and exact values of the circuit parameters.

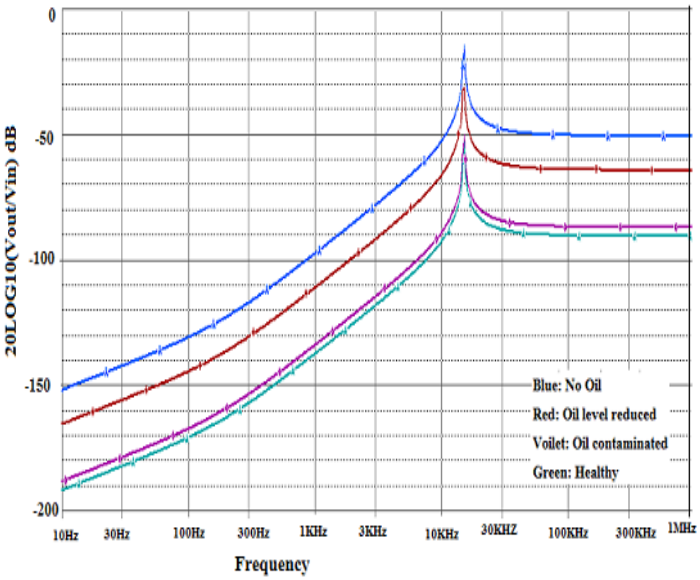


Fig.8 SFR responses from PSpice simulation

VI. Conclusion

Use of sweep frequency response (SFR) for on-site and instant test of oil level and oil contamination in power/distribution transformers during emergency situations has been proposed. For this an end to neutral connection set up is to be used. Based on the experiments conducted on a healthy transformer and two other transformers respectively with reduced oil level and degraded oil, it is suggested that the SFR curves due to oil related problems will shift above that for a healthy transformer usually in the 60 kHz to 300 KHz. A PSpice simulation of the equivalent circuit for the set up of a generic transformer with different oil conditions has also been done and this also supports the basic experimental observation.

Acknowledgments

This research was conducted at the CERS (Central Equipment and Repairing Shop), a unit of Bangladesh Power Development Board, located at Tongi, Gazipur, Bangladesh. All the personnel of CERS are thankfully acknowledged for their help and support.

References

[1] E.P. Dick and C.C. Erven, “Transformer Diagnostic Testing by Frequency Response Analysis” IEEE TRANS. PAS, VOL. 97, PP.2144-2153, NOV./ DEC 1978.

[2] Steven D. Mitchell, James S. Welsh, “Modeling Power Transformers to Support the Interpretation of Frequency-Response Analysis”, IEEE TRANSACTIONS ON POWER DELIVERY, VOL. 26, NO. 4, PP.2705-2717, OCTOBER 2011.

[3] Zhongdong Wang, Jie Li, Dahlina M. Sofian, “Interpretation of Transformer FRA Responses—Part I: Influence of Winding Structure”, IEEE TRANSACTIONS ON POWER DELIVERY, VOL. 24, NO. 2, PP.703-710, APRIL 2009.

Detection of Impurities in Water by Measuring Capacitance

Jamil Wahid and Q.Ahsan

Department of Electrical & Electronic Engineering, Bangladesh University of Engineering & Technology (BUET)
Dhaka-1000, Bangladesh.

jamil.wahid08@gmail.com

Abstract—**Detection of unhygienic impurities in water is one of the major concerns of human society.The detection of impurity in water is not only important for drinking water but also for water in industrial use. The type of impurity to be detected varies with the variation of the use of water. A low cost parallel plate based capacitive sensor is designed and implemented for the experimental setup. The detection circuit proposed is a RC oscillator with a time constant proportional to the capacitance. The resulting time constant is measured by Arduino Uno microcontroller board. Simulations of the oscillator circuit are performed using Proteus. This paper investigates the impurities with different concentrations of Sodium Chloride (NaCl), Sugar($C_6H_{12}O6$) , Ferrous Sulphate ($FeSO_4$) and Copper Sulphate ($CuSO_4$). Empirical relationships are developed to identify the type of impurity and its concentration.**

Index Terms—**Distilled Water, Capacitance Measurement, Electrolyte Impurities.**

I. Introduction

Water is one of the most essential factors for sustaining the life of all living organisms. Apart from drinking water to survive, there are various industries and chemical and biological laboratories where highly purified water is essential. Besides this, the uses of impurity free water include- coolant in nuclear reactors, boilers, automotive cooling, cleansing of surgical apparatus.There has been a great deal of interest in the detection of impurities in distilled water.

It is very difficult to determine the total amount of impurity in a sample by direct chemical analysis. Moreover, it is impossible to produce reproducible results by evaporating the water sample to dryness and weighing the residue as some of the present minerals, mainly chloride is lost during final stages of heating [1]. Hence, the measurement of impurities, mainly saline has been done using both direct method and indirect methods [2]. Most of these measurement methods work on the basis of change of conductivity [3].

Modern conductivity sensors are basically of two types: contact type and noncontact type sensors. Contact type sensors let the solution under test flow through them and sense the conductivity by measuring the current flow through the circuit [4]. In spite of the simplicity of design, they suffer a few significant drawbacks. Since there is direct contact between the solution and probe, it risks corrosion and fouling effects. Space charge polarization effects are also pronounced since same probes are used, both for applying voltage and measuring the resultant current. Although these problems are minimized by using 4 probe sensors, the issue of fouling still exits.

Non-contact type sensors have no contact between the sensing device and the analyte. This gives them advantage over the contact type sensors. But long term stability is dictated by polarization and fouling .The measuring principal in non-contact type sensors is mainly inductive. It consists of two coils. The primary coil is a single coil whereas, the surrounding liquid forms the secondary liquid. An ac current of known frequency and amplitude drives the primary. When this coil is brought near water, transformer action ceases and a eddy current, directly proportional to the conducting ability of the medium, is generated. This kind of sensor has not yet been applied to the determination of impurity in water [5], but has been used in non-invasive electromagnetic conductivity sensing in biomedical research [6].

Non-contact, inductive conductivity sensors, specially double transformer based ones are widely used because of their inherent advantage of not being prone to fouling and polarizing [7],[8]. But, inductive sensors face problems from external fields. The measuring field being used for inductive sensing method is exposed to surrounding media. However, the whole of this media is not coupled inductively. The contributions of these external fields can be as much as up to 20% [9]. This leads to proximity errors. Even the anti-fouling coating can change the geometry of the sensor, resulting in errors. Another problem associated with this method, is its heavy dependence on temperature and pressure variation [5].

All these problems associated with the inductive sensing methods can be lessened to great extent using the capacitive methods for conduction sensing. This paper uses the parallel plate capacitive configuration because of its simplicity of design, inherent confinement of the measuring field between the two plates and easy detection of a small variation of capacitance even in the range of Pico farad (10^{-12}). Only a small field is exposed to the external medium which gives rise to fringing effects. The fringing effects were dealt with by the use of Kelvin guard rings [10].

Four of the most common salts dissolved in water have been used as sample impurities. They are - Sodium Chloride (NaCl), Sugar($C_6H_{12}O_6$) , Ferrous Sulphate ($FeSO_4$) and Copper Sulphate ($CuSO_4$).

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

II. Methodology

A. Sensor Design

The sensor consists of two parallel conducting (copper) plates. One of the plates has a dimension of 2inch×1inch and the other plate is built with slightly larger dimensions (2.1 inch×1.1 inch)for purposes of alignment. The sensor plates are constructed on copper clad substrate boards. Since our experiments were conducted on water and water conducts electricity, we insulated the copper plates by coating it with varnish oil. Contact wires were soldered out for connection purposes. The capacitance of a parallel plate capacitor is given by (1).

$$C = \frac{\varepsilon A}{d} \tag{1}$$

Where, ε is the permittivity of the dielectric medium between two plates. A is electrode area and d is the distance between electrodes. However, in practice, the measured capacitance of a capacitor is generally higher than the calculated capacitance if value of d is comparable to that of A. This increase is attributed to the fringing fields at the plate edges. It is difficult to calculate the fringe field. So, Kelvin guard electrodes (ring) have been used to minimize the fringing effect and also protection from external interference [12]. It is sufficient to use this ring around just one of the electrodes. A typical arrangement for Kelvin guard ring is as shown in Fig. 1

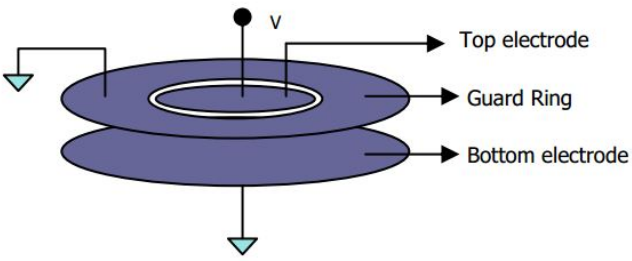


Fig.1 Capacitive Sensor with Kelvin Guard Ring

B. Detection Circuitry

The detection circuit used for this experiment used auto range detection schemes for two different ranges of capacitors. For small capacitors (1pF – 1nF), a voltage divider was formed between the test capacitor and a known capacitor (30pF) as shown in Fig. 2

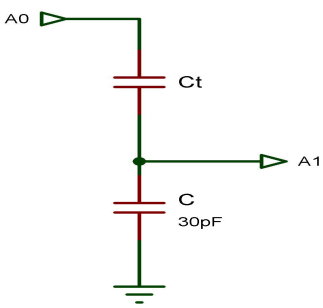


Fig.2 Sensor Circuitry for higher concentration of impurities

Initially, both capacitors are discharged with A0 at zero volts. Then a 5volt is applied at A0. The voltage at A1 will settle down to 1% of its final value within 30ns. The value of Ct is calculated from (2).

$$Ct = \frac{V_{A1}}{V_{A0} - V_{A1}} C \tag{2}$$

For the higher range of capacitors, RC circuit shown in Fig.3 was used.

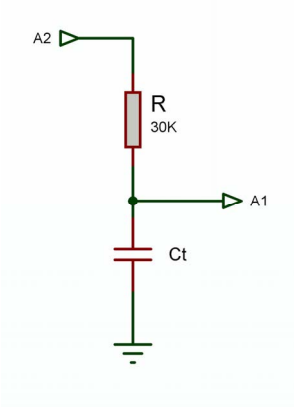


Fig.3 Sensor Circuitry for lower concentration of impurities.

The internal pull up resistance of the Arduino Uno board was used as the resistor of the RC network. It has a value of around 30kΩ. A 5volt was placed at A2 and the voltage at A1 was sampled along with the timestamps. Since, "Analog-Read" function takes longer response time, the "Digital Read" function was used to sense when the voltage at A1 was 2.5 volt. After that, "Analog-Read" was used. The value of test capacitance was calculated from (3).

$$Ct = \frac{-t}{R \ln\left(1 - \frac{V_{A1}}{V_{A2}}\right)} \tag{3}$$

C. Sample and Solution Preparation

The experiment was carried out on solutions with up to 20gm of impurity- Sodium Chloride (NaCl), Sugar ($C_6H_{12}O_6$) , Ferrous Sulphate ($FeSO_4$) and Copper Sulphate ($CuSo_4$) dissolved in 1 Litre of distilled water. A micro pipette was used to measure the volume of water. Proper analysis was carried out to ensure a decrement of 1gm of dissolved impurity for each iteration.

D. Measurement technique

At first the sample solution was prepared. The sensor circuit was immerged into the solution. The resulting capacitance was shown on the 16×2 alphanumeric display. 3 repetitions were done and the average was taken as the corresponding reading.

III. Results

A. Accuracy of the Sensor

Readings were taken for standard value capacitors and in each case the readings were within a tolerance limit of 6%. The reading of the sensor for various ranges of capacitors are listed below in Table 1.

TABLE I
ACCURACY OF SENSOR

Capacitor Model	Real Value	Sensor Output	Error
15J	15pF	15.461pF	3.07%
22J	22pF	23.103pF	5.01%
47J	47pF	48.532pF	3.26%
101J	100pF	102.000pF	2.00%
471K	470pF	478.776pF	1.87%
102K	1.000nF	1.053nF	5.30%
222K	2.22nF	2.34nF	5.40%
392K	3.92nF	3.853nF	-1.71%
472K	4.720nF	4.884nF	3.47%
103K	10nF	10.50nF	5.00%
104K	100nF	104.75nF	4.75%
105K	1uF	0.986uF	1.40%

As we can see, over the large range of capacitance values from 15pF to 1uF, the maximum error is below 6%,which is be accurate enough to serve our purpose.

B. Readings for different impurities

1) *Sodium Chloride(NaCl)* :When sodium chloride was added as the impurity, the measured capacitance was found to decrease with increasing concentration. The value of measured capacitance against the amount of dissolved impurities are plotted in Fig.4

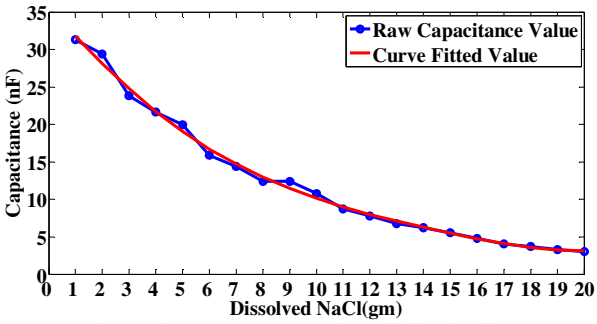


Fig.4 Measured Capacitance Vs Amount of NaCl dissolved in water

2) *Ferrous Sulphate($FeSO_4$):* The observed capacitance decreased for the$FeSO_4$ solution with increasing concentration of impurity .The graph is shown in Fig.5

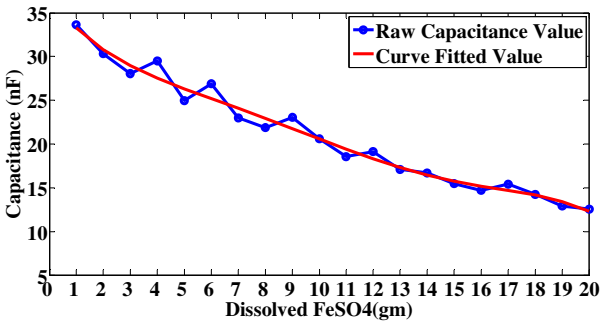


Fig.5 Measured Capacitance Vs Amount of $FeSO_4$ dissolved in water

3) *Copper Sulphate($CuSO_4$):* Similar to the previous two impurities, the capacitance decreases with increased amount of dissolved impurities. The graph for $CuSO_4$ is shown in Fig.6

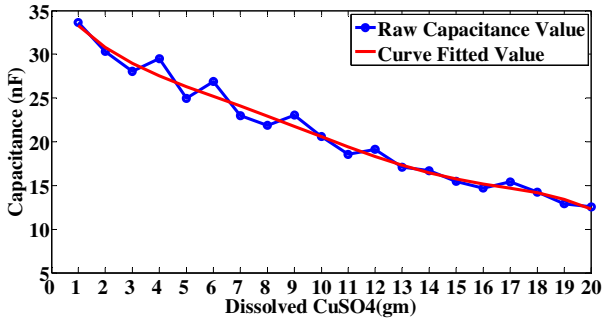


Fig.6 Measured Capacitance Vs Amount of $CuSO_4$ dissolved in water

4) *Sugar ($C_6H_{12}O_6$):* Similar to the previous two impurities, the capacitance decreases with increased amount of dissolved impurities. The graph for $CuSO_4$ is shown in Fig.7

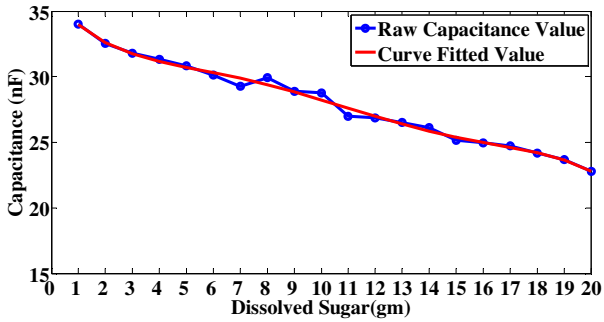


Fig.7 Measured Capacitance Vs Amount of Sugar dissolved in water

C. Data Analysis

If we compare the capacitance vs dissolved impurity graph for the different materials, as shown in Fig. 8, it is evident that for all the cases, the capacitance of distilled water decreases with increasing amount of dissolved impurity.

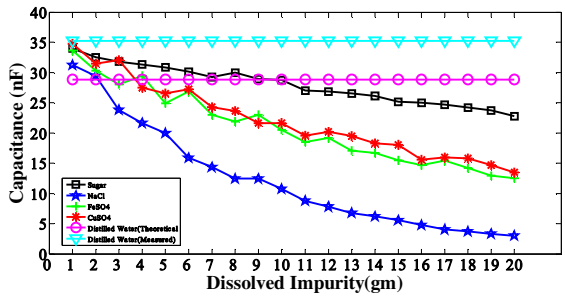


Fig. Comparison of capacitance vs dissolved impurity for different impurities.

This phenomenon can be well explained by *Electrolytic Dissociation Theory in Solutions*. Even a very small concentration of electrolyte usually causes a significant change in the conductivity of the liquid. The dielectric behaviour of liquid medium respond to dissolving of impurities in two probable ways. Firstly, association of ions

may result in generation of ion pairs or similar species. These species of appreciable dipole moment will be expected to make their own contribution to the dielectric polarization process and also the dispersion. Secondly, ions can influence the solvent's molecular interaction by its own strong localized field. A straight forward consequence would be imparting its own characteristics to the solution.

It is the second aspect, that is mostly responsible for the analysis of the dielectric property[13]. The electrostatic field in the immediate proximity of the dissolved impurity ions interact with the polar water molecules in such a way that appreciable number of water molecules are "frozen" around each of the solute ions. The hydrated salt ions are bulkier entity than a non-hydrated ion. Moreover, the hydrated sheath around the dipoles prevents them from orienting themselves with the external electric field. This is the principal cause of reduction in permittivity which results in reduction of capacitance due to addition of impurity.

Another observation from the comparison shown in Fig.8 is that although capacitance decreases for all the impurities, the rate of change of capacitance is not the same for all the impurities. Of the considered impurities, NaCl had the largest rate of decreasing of capacitance while Sugar had the least. NaCl is an ionic compound where as Sugar is a covalent compound where polarity arises due to the hydroxyl bonds (-OH) present. So, the contribution to the local field around the hydrated ions is much more for NaCl than for Sugar. That is why, change of capacitance is most for NaCl and least for Sugar. The total decrease in capacitance for the diffent materials are given.

IV. CONCLUSIONS

A parallel plate capacitive sensor was fabricated to detect the type and concentration of impurities in water by capacitance measurement. Data analysis is carried out to curve fit the acquired raw data in order to develop empirical relations among the variables. Experimental results show, specific trends of reduction of capacitance with the concentration of impurities. This approach could be extended to the detection of impurities in non-binary solutions, containing more than one kind of impurities.

.

REFERENCES

[1] (2005) “Chapter6: Temperature, Salinity and Density”. [Online]. Available: http://oceanworld.tamu.edu/resources/ocng_textbook/chapter06/chapter06_01.htm.

[2] Barnes H., Apparatus and Methods of Oceanography – Part One: Chemical, George Allen and Unwin Ltd., 1959.

[3] Tucker M.J., “Measurement in Oceanography”, J. Phys. E: Sci. Instrum.4 405-413, 1971.

[4] Sudduth, K. A., N. R. Kitchen, G. A. Bollero, D. G. Bullock, and W. J. Wiebold. "Comparison of electromagnetic induction and direct sensing of soil electrical conductivity." Agronomy Journal 95, no. 3 (2003): 472-482.

[5] Striggow K. and Dankert K., “The Exact Theory of Inductive Conductivity Sensors for Oceanographic Application”, IEEE Journal of Oceanic Engineering, Vol. OE-10, No.2, 1985.

[6] Hart L.W., Ko H.W., Meyer J.H, Vasholz D.P., Joseph R.I., “A Non-invasive Electromagnetic Conductivity Sensor for Biomedical Applications”, IEEE Transactions on Biomedical Engineering, Vol.35, No.12, December 1988.

[7] Guardo, Robert, Guy Charron, Yves Goussard, and Pierre Savard. "Contactless measurement of thoracic conductivity changes by magnetic induction." In *Engineering in Medicine and Biology Society, 1997. Proceedings of the 19th Annual International Conference of the IEEE*, vol. 6, pp. 2450-2453. IEEE, 1997.

[8] Natarajan, Saravana P., Jeremy Huffman, Thomas M. Weller, and David P. Fries. "Contact-less toroidal fluid conductivity sensor based on RF detection." In *Sensors, 2004. Proceedings of IEEE*, pp. 304-307. IEEE, 2004.

[9] Langereis G.R., “An Integrated Sensor System for Monitoring Washing Purposes”, M. Eng thesis, University of Twente, Netherlands, 1999.

[10] Heerens, W. Chr. "Multi-terminal capacitor sensors." *Journal of Physics E: Scientific Instruments* 15, no. 1 (1982): 137.

[11] Tiwari, G. N., H. N. Singh, and Rajesh Tripathi. "Present status of solar distillation." Solar energy 75, no. 5 (2003): 367-373.

[12] Bousse, Luc, N. F. De Rooij, and Piet Bergveld. "Operation of chemically sensitive field-effect sensors as a function of the insulator-electrolyte interface." *Electron Devices, IEEE Transactions on* 30.10 (1983): 1263-1270.

[13] Hill N.E., Vaughan W.E., Price A.H. and Davies M., Dielectric Properties and Molecular Behavior, Van Nostrand Reinhold Company, London, 1969.

A Two Dimensional Analytical Model of Drain to Source Current and Subthreshold Slope of a Triple Material Double Gate MOSFET

M. A. Mahmud and S. Subrina*, Member, *IEEE*

Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology
Dhaka-1000, Bangladesh

*samiasubrina@eee.buet.ac.bd

Abstract—**A two dimensional analytical model of drain to source current as well as subthreshold slope of a triple material double gate MOSFET has been developed in this work. Basic drift-diffusion equation has been used to derive the drain to source current model. An expression of pinch-off voltage has been derived for modeling drain current in saturation region. The current in the device has been studied as function of drain to source voltage and gate voltage as well. In the work, leakage current in zero gate bias condition has also been presented. Variations in subthreshold slope of the device owing to change in device parameters have been lifted up. Our model results have been verified with the simulation data obtained by using a professional numerical device simulator.**

Index Terms **— Drain to source current, gate engineering, leakage current, pinch off voltage, subthreshold slope.**

I. Introduction

Short channel effects (SCE) originating from downscaling of electronic devices need to be mitigated for optimum performance of the devices. Triple material double gate MOSFET is a potential candidate for microelectronic industry due to its ability of suppression of various short channel effects (SCE) compared to traditional single or double gate single material MOSFET structures. Double gate structure is advantageous over the traditional single gate structure because of its higher current drive capability, off-current reduction and better scaling and effective short channel effect handling capability [1]-[4]. In addition, as a form of gate engineering, both the upper and lower gate electrodes of this structure consist of three metals having different work functions. The metal close to the source having the highest work function while that close to drain having the lowest. It gives a two-step surface potential profile that suppresses Drain induced barrier lowering (DIBL) effectively. This gate engineering also allows to have the same threshold voltage for the device with a reduced doping concentration in the channel region. This results in better immunity against mobility degradation and higher transconductance. Some theoretical and experimental works on the TM gates have already been published [5]-[6]. Tiwari *et al*. presented a two dimensional analytical model of surface potential and threshold voltage of a triple material device [6]. But no model for drain to source current and subthreshold slope of this structure was reported in the literature.

In this work, a two dimensional analytical model of drain to source current and subthreshold slope of the aforesaid structure along with results from a commercially available 2D device simulator is presented. All the analytical model results show good agreement with the results obtained from 2D device simulator that justify the validity of our model.

II. Analytical Modeling

This section presents the modeling of two dimensional drain to source current and subthreshold slope of a triple material dual gate MOSFET (TMDG MOSFET). A schematic structure of the device is shown in Fig. 1. The x- and y- axes of the 2D device structure are considered to be along the front oxide-semiconductor interface and the source terminal-channel interface, respectively. The total channel length, L is divided into 3 regions. The front and the back gate electrodes of the device are symmetric and composed of three different materials having work functions of Φ_{m1}, Φ_{m2} and Φ_{m3} ($\Phi_{m1} > \Phi_{m2} > \Phi_{m3}$) deposited over lengths L_1, L_2 and L_3, respectively. The front and the back gate terminals are shorted to a fixed potential, V_{GS}. The drain bias is V_{DS} and the source terminal is grounded. A layer of SiO_2 has been used with both front and back gate terminals.

A. Drain to Source Current Model

For modeling of drain to source current, mid-channel potential, Φ_{cj} has been considered instead of surface potential in basic drift and diffusion current equation, because considering surface potential in this case assumes punch-through current to flow only at surface, which is invalid for double gate fully depleted SOI MOSFET [7, 8]. Drain to source current can be expressed as a combination of drift and

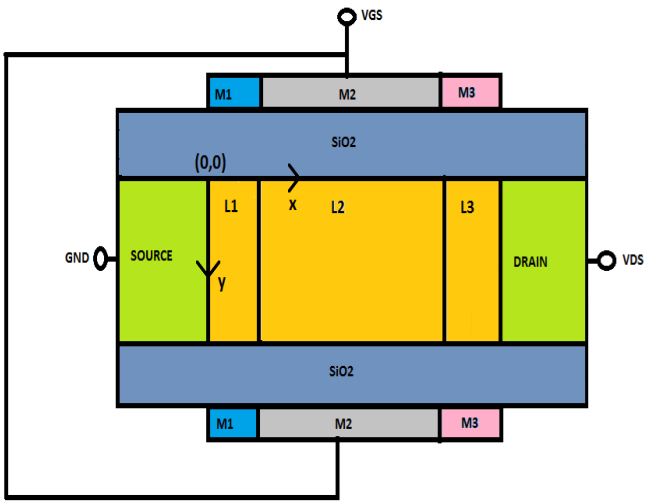


Fig. 1. Schematic diagram of triple material double gate MOSFET

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

diffusion current [9] for all the regions which can be given as:

$$I_{DRIFT_j} = W \cdot \mu_j(x) \cdot Q_{Ij}(x) \frac{d\Phi_{cj}(x)}{dx} \quad (1)$$

$$I_{DIFF_j} = W \cdot \mu_j(x) \cdot V_T \frac{dQ_{Ij}(x)}{dx} \quad (2)$$

where, $W, V_T, \mu_j(x)$ and $Q_{Ij}(x)$ are channel width, thermal voltage, carrier mobility and inversion layer charge, respectively for j=1,2,3 and are given by:

$$\mu_j(x) = \frac{\mu_0}{\sqrt{\left[1+\left\{\frac{E_j(x)}{E_c}\right\}^2\right]}} \quad (3)$$

$$Q_{Ij}(x) = 2\{Q_{sj}(x) - Q_{Dj}\} \quad (4)$$

where, μ_0, $E_j(x)$, E_c, $Q_{sj}(x)$ and Q_{Dj} are low carrier mobility, electric field, critical electric field, surface charge and depletion layer charge, respectively [10]-[11] and are given by:

$$E_j(x) = \frac{C_{ox}\left[\Phi_{cj}(x) - V_{GS} + V_{fbj}\right]}{\in_{si}} \quad (5)$$

$$E_c = 6.01\times10^2 \cdot T^{\frac{3}{2}} \quad (6)$$

$$Q_{sj}(x) = -C_{ox}\left[V_{GS} - V_{fbj} - \Phi_{cj}(x)\right] \quad (7)$$

$$Q_{Dj} = C_{ox}\left(\varphi_F - V_{th} + V_{fbj}\right) \quad (8)$$

where $C_{ox} = \frac{\in_{sio2}}{t_{ox}}$ is oxide capacitance, $\in_{si}$ and $\in_{sio2}$ are permittivities of Si and SiO_2 respectively , $\varphi_F = V_T \ln\left(\frac{N_A}{n_i}\right)$ is Fermi potential, N_A is acceptor doping concentration in all the three regions of the channel and n_i is the intrinsic carrier concentration of silicon. V_{th}, V_{fbj} and Φ_{cj} are threshold voltage, flatband voltage and mid-channel potential, respectively and are given by Tiwari *et al*. [6].

Now, integrating both sides of the Eq. 1 and using Eqs. 3-4 and 7-8,

$$\int_{L^*_{j-1}}^{L^*_j} I_{DRIFT_j}\,dx = \int_{\Psi_{cj-1}}^{\Psi_{cj}} \frac{W\mu_0\left[-2C_{ox}\left\{V_{GS} - \Phi_{cj}(x) + \varphi_F - V_{th}\right\}\right]}{\sqrt{\left[1+\left\{\frac{E_j(x)}{E_c}\right\}^2\right]}} d\Phi_{cj} \quad (9)$$

where,

$$\begin{cases} L^*_0 = 0, \\ L^*_1 = L_1, \\ L^*_2 = L_1 + L_2, \\ L^*_3 = L_1 + L_2 + L_3, \end{cases}$$

and

$$\begin{cases} \Psi_{c0} = \Psi_{cj}(x)\big|_{x=L^*_0} \\ \Psi_{c1} = \Psi_{cj}(x)\big|_{x=L^*_1} \\ \Psi_{c2} = \Psi_{cj}(x)\big|_{x=L^*_2} \\ \Psi_{c3} = \Psi_{cj}(x)\big|_{x=L^*_3} \end{cases}$$

and j=1,2,3

From Eqs. 5, 7 and 9, we get,

$$I_{DRIFT_j} = \frac{S_4}{L_j - L_{j-1}}\left[(S_1 + S_3)\ln\left\{\frac{\left(S_3 + \Phi_{cj}\right) + \sqrt{S_2^2 + \left(S_3 + \Phi_{cj}\right)^2}}{\left(S_3 + \Phi_{cj-1}\right) + \sqrt{S_2^2 + \left(S_3 + \Phi_{cj-1}\right)^2}}\right\} - \left\{\sqrt{S_2^2 + \left(S_3 + \Phi_{cj}\right)^2} - \sqrt{S_2^2 + \left(S_3 + \Phi_{cj-1}\right)^2}\right\}\right] \quad (10)$$

where,

$$S_1 = V_{GS} + \varphi_F - V_{th} \quad (11)$$

$$S_2 = \frac{\in_{si} E_c}{C_{ox}} \quad (12)$$

$$S_3 = V_{fbj} - V_{GS} \quad (13)$$

$$S_4 = -2\in_{si} \mu_0 W E_c \quad (14)$$

Now, integrating both sides of the Eq. 2 and using Eqs. 3-6,

$$\int_{L^*_{j-1}}^{L^*_j} I_{DIFF_j}\,dx = \int_{Q_{Ij-1}}^{Q_{Ij}} \frac{W\mu_0 V_T}{\sqrt{\left[1+\left\{\frac{E_j(x)}{E_c}\right\}^2\right]}} dQ_{Ij} \quad (15)$$

where,

$$\begin{cases} Q_{I0} = Q_{Ij}(x)\big|_{x=L^*_0} \\ Q_{I1} = Q_{Ij}(x)\big|_{x=L^*_1} \\ Q_{I2} = Q_{Ij}(x)\big|_{x=L^*_2} \\ Q_{I3} = Q_{Ij}(x)\big|_{x=L^*_3} \end{cases}$$

and j=1,2,3

From Eqs. 4-6 and 15, we get,

$$I_{DIFF_j} = \frac{S_7}{L_j - L_{j-1}}\left[\ln\left\{\frac{\left(S_5 + Q_{Ij}\right) + \sqrt{S_6^2 + \left(S_5 + Q_{Ij}\right)^2}}{\left(S_5 + Q_{Ij-1}\right) + \sqrt{S_6^2 + \left(S_5 + Q_{Ij-1}\right)^2}}\right\}\right] \quad (16)$$

where,

$$S_5 = 2C_{ox}\left[V_{fbj} + \varphi_F - V_{th}\right] \quad (17)$$

$$S_6 = 2\in_{si} E_c \quad (18)$$

$$S_7 = S_6 W V_T \mu_0 \quad (19)$$

So, the drain to source current can be expressed as,

$$I_{DS} = \sum_{j=1}^{3}\left[I_{DRIFT_j} + I_{DIFF_j}\right] \quad (20)$$

Now, the expression in Eq. 20 is valid for all values of drain to source voltage below pinch off i.e. where the inversion layer charge becomes zero. With the gradual increase of drain to source voltage $V_{DS,}$ at certain value, pinch off occurs at drain end of the channel and this voltage is called Pinch off voltage, V_{DP}. Above V_{DP}, the drain to source current remains constant. Now, the expression of V_{DP} can be obtained by equating inversion layer charge in Eq. 4 at L^*_3 to zero and using Φ_{cj} and Eqs., 4, 7- 8 and 11:

$$V_{DP} = V_{GS} - V_{th} + \varphi_F + \frac{\beta_3}{\eta^2} - A^*_3 e^{\eta L^*_3} - B^*_3 e^{-\eta L^*_3} \quad (21)$$

where, β_3 and η are constants and A^*_3 and B^*_3 are values of

constants A_3 and B_3 [6] at $V_{GS} = V_{th}$ and can be given by a common expression:

$$A^*_j = A_j - \frac{V_{DP}}{2\sinh(\eta L)} \tag{22}$$

$$B^*_j = B_j + \frac{V_{DP}}{2\sinh(\eta L)} \tag{23}$$

So, for $V_{DS} \geq V_{DP}$, drift current component,

$$I^*_{DRIFT_j} = \frac{S_4}{L_j - L_{j-1}}[(\Phi_{c3} + S_3)\ln\left\{\frac{(S_3 + \Phi^*_{cj}) + \sqrt{S_2^2 + (S_3 + \Phi^*_{cj})^2}}{(S_3 + \Phi^*_{cj-1}) + \sqrt{S_2^2 + (S_3 + \Phi^*_{cj-1})^2}}\right\}$$

$$-\left\{\sqrt{S_2^2 + (S_3 + \Phi^*_{cj})^2} - \sqrt{S_2^2 + (S_3 + \Phi^*_{cj-1})^2}\right\}] \tag{24}$$

where,

$$\Phi^*_{cj} = A^*_j e^{\eta L^*_j} + B^*_j e^{-\eta L^*_j} - \frac{\beta_j}{\eta^2} + V_{DP} \tag{25}$$

$$\Phi^*_{cj-1} = A^*_j e^{\eta L^*_{j-1}} + B^*_j e^{-\eta L^*_{j-1}} - \frac{\beta_j}{\eta^2} + V_{DP} \tag{26}$$

Also, for for $V_{DS} \geq V_{DP}$, diffusion current component,

$$I^*_{DIFF_j} = \begin{cases} \frac{S_7}{L_j - L_{j-1}}\left[\ln\left\{\frac{(S_5 + Q^*_{Ij}) + \sqrt{S_6^2 + (S_5 + Q^*_{Ij})^2}}{(S_5 + Q^*_{Ij-1}) + \sqrt{S_6^2 + (S_5 + Q^*_{Ij-1})^2}}\right\}\right]; j=1,2 \\ \frac{S_7}{L_j - L_{j-1}}\left[\ln\left\{\frac{S_5 + \sqrt{S_6^2 + S_5^2}}{(S_5 + Q^*_{I2}) + \sqrt{S_6^2 + (S_5 + Q^*_{I2})^2}}\right\}\right]; j=3 \end{cases} \tag{27}$$

where,

$$Q^*_{Ij} = -2C_{ox}(S_1 - \Phi^*_{cj}) \tag{28}$$

$$Q^*_{Ij-1} = -2C_{ox}(S_1 - \Phi^*_{cj-1}) \tag{29}$$

So, for $V_{DS} \geq V_{DP}$, drain to source current can be expressed :

$$I^*_{DS} = \sum_{j=1}^{3}\left[I^*_{DRIFT_j} + I^*_{DIFF_j}\right] \tag{30}$$

Finally, the drain to source current can be expressed as :

$$I_{DS,tot} = \begin{cases} I_{DS}; V_{DS} < V_{DP} \\ I^*_{DS}; V_{DS} \geq V_{DP} \end{cases} \tag{31}$$

B. Subthreshold Slope

Subthreshold slope can be obtained using the following equation :

$$SS = \frac{dV_{GS}}{d\log[I_{DS}]} \tag{32}$$

III. Results

In this section, the results obtained in the analytical modeling section have been represented graphically along with the results obtained from 2D device simulator for model validity. Work functions used in this analysis are φ_{m1} =4.8eV, φ_{m2} =4.6eV and φ_{m3} =4.4eV. Acceptor doping concentration, N_A of 10^{16} cm^{-3} in all the three regions of the channel and drain and source region (heavily doped with donor atoms) concentration of 10^{20} cm^{-3} have been used in our work. Total channel length, L of the device has been considered to be 120nm. Ratio of the length of the channel regions, $L_1 : L_2 : L_3$ has been kept 1:1:1. In our analysis, channel thickness, t_{si} has been considered 20 nm. Oxide thickness, t_{ox} is kept 3.5nm. It is reported in that the QME (Quantum Mechanical Effects) become negligible if the channel thickness, t_{si} is over 5nm

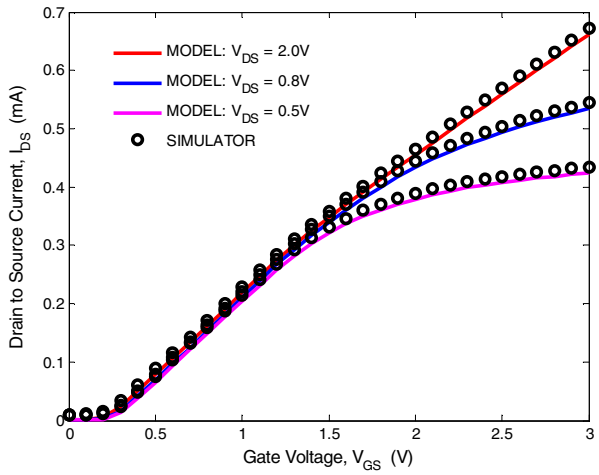


Fig. 2. Drain to source current as a function of gate voltage for different values of drain to source voltage (V_{DS}).

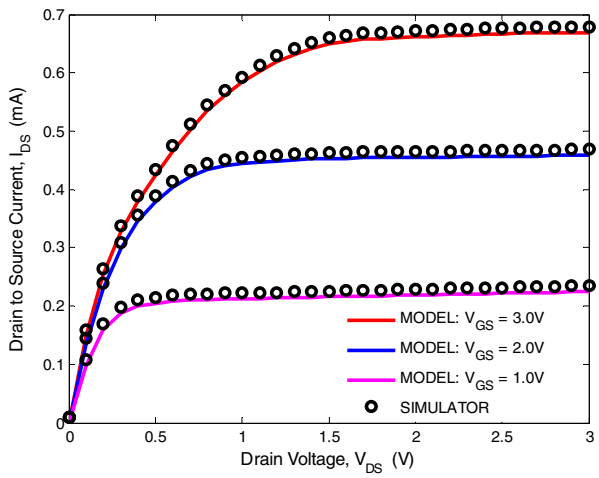


Fig. 3. Drain to source current variation against drain voltage for gate voltgae,V_{GS}= 3V, 2V and 1V

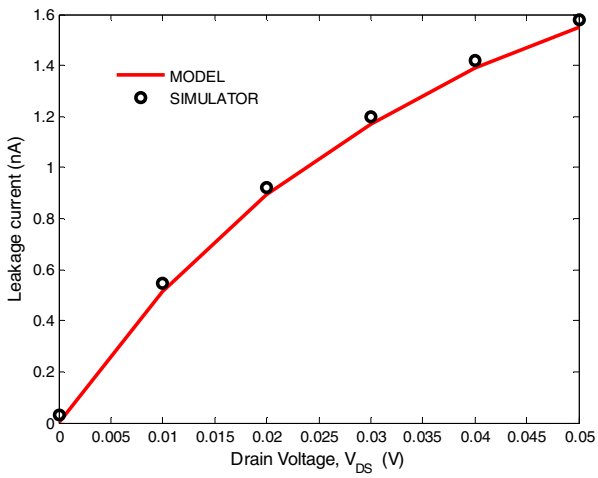


Fig. 4. Leakage current in the device as a function of drain voltage while V_{GS} = 0V

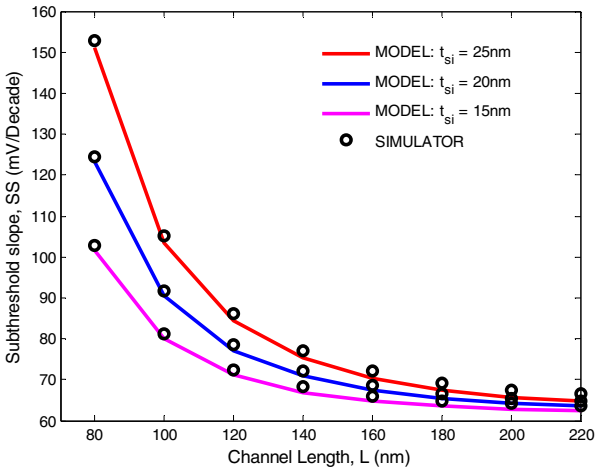


Fig. 5. Subthreshold slope variation with channel length for different values of Si-film thickness (t_{Si}) while V_{DS}=0.1V.

[12]. Since, channel thickness used in our analysis is above it, we have not included QME in our modeling.

Fig. 2 shows the drain to source current with varying gate voltages for three different drain voltages. In the subthreshold region, negligible leakage current flows through the channel. Once the threshold voltage is achieved, the device is turned on and drain to source current increases with increasing gate voltage and drain voltage as well. Fig. 3 presents the plot of drain current with drain voltage for three different values of gate voltage. It is obvious from the figure that, at a certain value of V_{DS} (pinch off voltage ,V_{DP}), pinch off occurs at the drain end and current remains constant with increasing drain voltage. Fig. 4 illustrates leakage current with low drain bias for V_{GS} =0V. Negligible leakage current (in nA) is observed under this condition. Fig. 5 lifts up the subthreshold slope of the device with varied channel length and Si film thickness. It is obvious from the figure that subthreshold slope is higher for device with thicker silicon channel thickness and shorter channel length. With the increase in silicon channel thickness, there is a rise in depletion layer capacitance in a fully depleted MOSFET. Since, subthreshold slope is proportional to the ratio of depletion layer capacitance to gate oxide capacitance [13], thicker channel MOSFETs exhibit higher subthreshold slope. The analytical model results show good agreement with the numerical results from the 2D device simulator.

IV. Conclusion

A two dimensional analytical model of drain to source current and subthreshold slope of triple material double gate MOSFET have been presented in this paper. Dependence of device subthreshold slope on device parameters like channel length and silicon channel thickness has also been explored. The model results have been verified with the simulation results as well.

References

[1] T. Ernst, S. Cristoloveanu, G. Ghibaudo, T. Ouisee, S. Horiguchi, Y. Ono, Y. Takahashi, and K. Murase, "Ultimately thin Double-gate SOI MOSFETs," *IEEE trans. Electron Devices*, vol. 50, no. 3, pp. 830-838, March 2003.

[2] K. Suzuki, Y. Tosaka, T. Tanaka, H. Horie, and Y. Arimoto, "Scaling theory of Double-gate SOI MOSFETs," *IEEE trans. Electron Devices*, vol. 40, no. 12, pp. 2326-2329, December 1993.

[3] H. Lu, and Y. Taur, "An analytical potential model for symmetric and asymmetric DG MOSFETs," IEEE trans. Electron Devices, vol. 53, no. 5, pp. 1161-1168, May 2006.

[4] B. Ray, and S. Mahapatra, "Modeling of channel potential and subthreshold slope of symmetric double-gate transistor," *IEEE trans. Electron Devices*, vol. 56, no. 2, pp. 260-266, February 2009.

[5] P. Razavi, and A. A. Orouji, "Nanoscale triple material double gate (TM-DG) MOSFET for improving short channel effects," *International conference on advances in electronics and microelectronics*, 2008.

[6] P. K. Tiwari, S. Dubey, M. Singh, and S. Jit, "A two-dimensional analytical model for threshold voltage of short-channel triple –material double gate metal-oxide-semiconductor field effect transistor," *Journal of Applied Physics*, 108, October 2010.

[7] K. Suzuki, T. Tanaka, and Y. Tosaka, "Scaling theory for double-gate SOI MOSFET's ," *IEEE Trans. Electron Devices*, vol. 40, no. 12, pp. 2326-2339, December 1993.

[8] N. D. Arora, *MOSFET models for VLSI circuits simulation*, New York: Springer-Verlag Wien.

[9] Y. Tsividis, *Operation and modelling of the MOS transistor*, 2nd ed., New York: McGraw-Hill, 1989.

[10] C. G. Sodini, P. –K. Ko, and J. L. Moll, "The effect of high fields on MOS device and circuit performance," *IEEE Trans. Electron Devices*, vol. ED-31, no. 10, pp. 1386-1393, October 1984.

[11] A. Akers, "The effect of field dependent mobility on the threshold voltage of a small geometry MOSFET," *Solid state electronics*, vol. 23, pp. 173-175, 1989.

[12] A. Kumar, J. Kedzierski, and S. E. Laux, "Quantam-based simulation analysis of scaling in ultrathin body device structures," *IEEE Trans. Electron Devices*, vol. 52, no. 4, pp. 614-617, April 2005.

[13] A. Godoy, J. A. L-. Villanueva, J. A. J-. Tejada, A. Palma, and F. Gamiz, "Quantam-based simulation analysis of scaling in ultrathin body device structures," *Solid state electronics*, vol. 45, pp. 391-397, 2001.

Effect of Stone-Wales Defects and Edge Roughness on The Switching and Frequency performance of Graphene Nanoribbon-FET

Atanu Kumar Saha,[1,*] Gobinda Saha,[1] and A. B. M. Harun-ur Rashid[1]

[1]Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka-1000, Bangladesh
[*]atanukumars@gmail.com

Abstract— **The novel electronic properties of graphene nanoribbons (GNRs) including purely two-dimensional structure along with its tunable bandgap have led to intense research into possible applications of this material in nanoscale devices. However, as yet, dimensions of its possibilities in practical device levels have remained inconsistent. In this paper we propose a model for GNR-FET that is made from only Armchair GNRs. Our complete NEGF-based simulation reveals its potential for fast digital electronics with On/Off ratio up to 10^3, transconductance of 8.5×10^3 nS/nm which lead to a analog operational frequency up to 3.3THz. The effects of Stone-Wales defects and Edge Roughness in GNRs have been analysed here that shows switching and frequency performance degradation of such GNR-FETs.**

Index Terms—**Armchair GNR, Edge Roughness, GNR-FET, NEGF, Stone-Wales Defects.**

I. Introduction

Graphene, the 'miracle material' of the 21st century is a material with immense possibilities due to its excellent electrical, magnetic, thermal, optical and mechanical properties. These properties, along with the ballistic electronic transport and high carrier mobility [1] cause this material to be promising candidate for nanoelectronics applications. For graphene, energy dispersion relation is linear at Dirac-points [2], so electrons in graphene behave as massless fermions [3] and travel through the lattice with long mean free path, which ensures higher electron mobility. As graphene is a zero bandgap material it seems unsuitable for transistor applications. However energy gap can be opened up by means of lateral confinement [4], which is realized by etching the graphene sheet in narrow strips, called graphene nanoribbons (GNRs). GNRs have opened a prospect for GNR based FETs as their energy gap is inversely proportional to their width [5] and depending on this width they can be semiconducting or metallic. The ultrathin structure of GNRs leads to perfect electrostatic control of gate over channel which is a key advantage over conventional semiconductors. Extensive research is now going on for fabricating GNRs with narrow width and smooth edges. Patterned GNRs have been exhibited experimentally with ~10nm width [6]-[7]. However, due to the embryonic stage of this new field of research, state-of-the-art etching techniques are far from atomistic resolution, so edge roughness can play an important role on overall performance of the GNR-FET devices [8]. Computer simulations can be very useful to provide physical insights of such phenomena and determine performance of GNR-FETs with ultrathin width. GNR-FET has been simulated previously using scattering matrix approach based on effective mass Hamiltonian [9], which fails to incorporate scattering due to nonidealities in quantum regime. Whereas quantum simulations based on a tight-binding approach [10] have been used to evaluate GNR-FET characteristics incorporating such nonideal phenomena.

In this paper, *I-V* characteristics of GNR-FET made from only Armchair GNRs are studied in order to establish its switching and frequency performance that can be expected if technological challenges are met. The electronic transport is analyzed using non-equilibrium Green's function (NEGF) formalism [11]. Different types of nonidealities have been investigated. In particular, we have studied the effect of Stone-Wales (SW) defects in the channel and Edge Roughness on the switching and frequency performance of the device. Our simulation study reveals that performance of GNR-FET severely degrades due to such defects.

II. Device Configuration and Simulation

FET operation requires a gate controlled channel region and two electrode regions to serve as source and drain. Depending on the chirality of graphene nanoribbon it can be Armchair or Zig-zag. Armchair GNR (AGNR) and Zig-zag GNR (ZGNR) show semiconducting or metallic behaviour depending on the number of carbon atoms along its width [12]. GNR-FET can be made from one Zig-zag semiconducting GNR channel and two Armchair metallic GNR leads [13]. Taking the variation of bandgap with the width of Armchair GNR into consideration we propose a GNR-FET that can be made from only Armchair GNRs. In the channel region N=7 (E^g=1.35eV) semiconducting Armchair nanoribon of 3.5nm length is used, where N is the number of carbon atoms along the width of the GNR. This region is then connected with two N=8 ($E^g \approx 0$eV) metallic Armchair nanoribbons, considered as electrodes. Width of the device is ~0.74nm. The structure is shown in Fig. 1. In this device HfO_2 ($\kappa \approx 25$) is used as gate oxide and oxide thickness is 0.4nm. This structure provides excellent interface matching between channel and electrode regions, eliminating the possibility of interface defect. While a contact metal would introduce large contact resistance to the device [14], contact with the metallic Armchair GNR makes smoother metal-semiconductor interface with minimum contact resistance.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

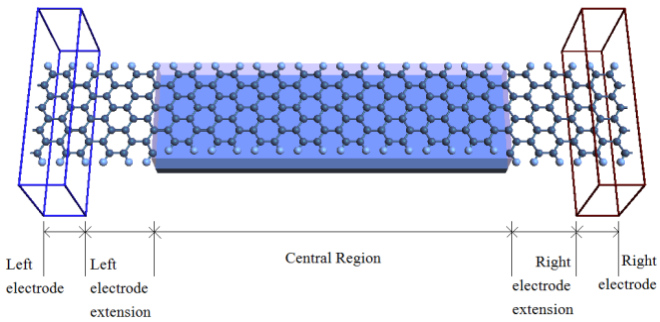


Fig. 1 Device configuration used for NEGF simulation.

In this paper we have used NEGF approach coupled with Poisson's equation to calculate electronic transport in GNR-FET. Here the NEGF formalism is based on tight binding P_Z orbital model which incorporates 3rd nearest neighborhood calibrated through a first-principle calculated *ab-initio* model [15]. The charge distribution described by NEGF is self-consistently coupled with Poisson's equation to incorporate the effect of self-consistent field and solved by multi-grid finite difference method. The edge atoms are considered as hydrogen passivated (σ bond).

III. Results and Discussion

A. Transmission Spectrum

As the first outcome of simulation we get the transmission spectrum of the central region of devices, when both gate and drain bias voltages are zero. Zero energy level is denoted by Fermi energy. Transmission coefficient above Fermi level is for electron and transmission coefficient below Fermi level is for holes. Within the bandgap energy (1.35eV) transmission coefficient is zero. Interesting finding is that, though in the above device channel-electrode interface was thought as perfect matching and the transmission spectrum should be smoothed at all energy level, we can see that transmission coefficient values degrade at some energy level as shown in Fig. 2. It is because interface atoms fail to keep electrons at all the energy levels what central region atom can, along with scattering effects at the interfaces.

B. I-V Characteristics

When a positive gate potential is induced, minimum energy for non-zero electron transmission coefficient comes closer to Fermi energy level, whereas a negative gate potential has opposite effect. So the charge carrier in a two dimensional channel can change from electrons for positive gate voltage to holes for negative gate voltage. But this phenomenon has less effect when we are using semiconducting GNR as channel.

Fig. 3 and Fig. 4 show I-V_{ds} and I-V_{gs} curve for the simulated device. With the increment of gate voltage, conductivity of the channel increases so current increases at constant drain to source voltage. The On/Off ratio is a key parameter to compare performances among different FETs. The Manchester research group demonstrated GFETs with On/Off ratio of ~30 at room temperature [2]. The first top-gated GFET with On/Off ratio < 2 was demonstrated by researchers of AMICA and RWTH Aachen University in 2007 [16]. Here On/Off current ratio for the above device is found to be ~10^3, which seems promising for FET operation.

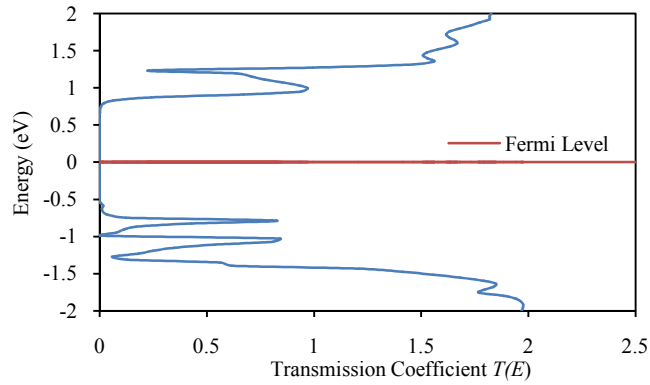


Fig. 2 Transmission spectrum of simulated device at zero gate bias.

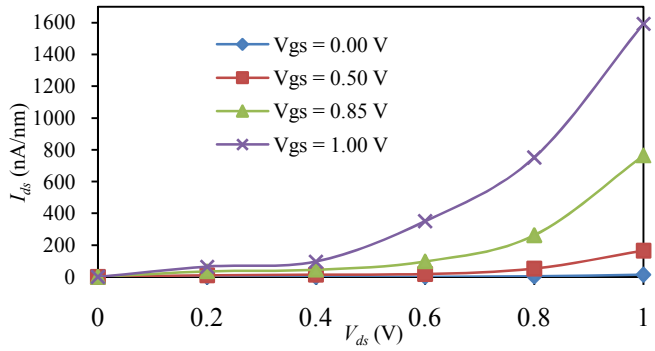


Fig. 3 I-V_{ds} curve at different gate voltages.

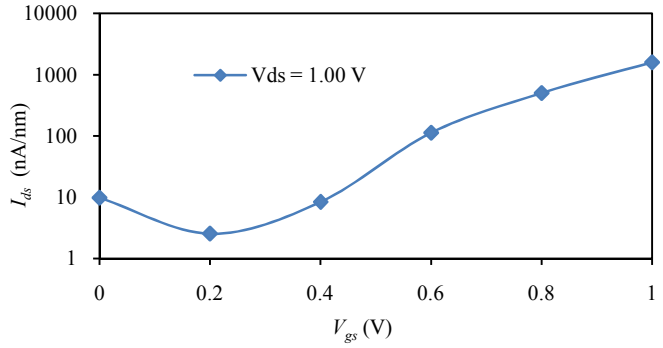


Fig. 4 I-V_{gs} curve at V_{ds}=1V.

C. Frequency Performance

Fig. 5 shows values of transconductance, g_m for different gate voltage at V_{ds}=1V. g_m starts to increase at V_{gs}=0.4V and become saturated after V_{gs}=1V. From this figure we can measure high frequency performance of our device. The cutoff frequency, f_T is related to g_m via the quasi-static approximation as

$$f_T = \frac{g_m}{2\pi C_g} \tag{1}$$

where C_g is the gate capacitance (for this device C_g= 4.1×10^{-10} F/m). Fig. 6 shows the values of f_T at different gate voltages. So for analog application, the device can operate up to 3.3THz. In [10] f_T up to ~4THz was found for defectless GNR-FET. Compared to RF frequency electronics, wider bandwidths, component compactness and improved spatial resolution are possible by GNR-FET, which may open up opportunities for faster electronics.

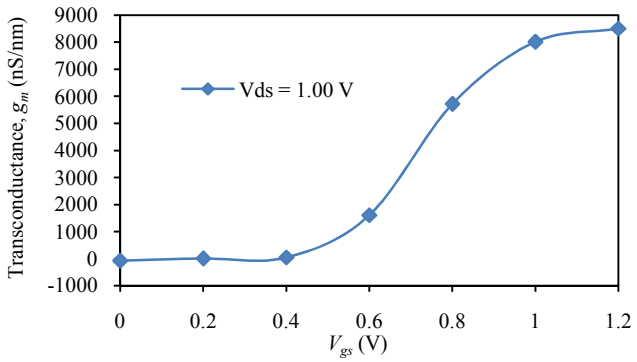


Fig. 5 Transconductance, g_m at different gate voltages for the device.

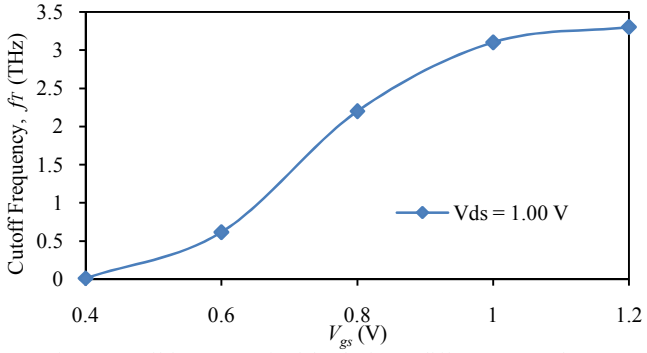


Fig. 6 Cutoff frequency, f_T of the device at different gate voltages.

IV. Defect Analysis In GNR-FET

Fabrication process of graphene nanoribbon may suffer from different types of defects in graphene sheet. Here we will discuss the effects of two such prominent defects on FET performances.

A. Stone-Wales Defects

Stone-Wales (SW) defects [17] are important topological defects in sp^2-bonded carbon materials, playing a central role in the formation, transformation, fracture, and embrittlement of carbon nanotubes, fullerenes, and other carbon nanostructures. SW defects have also been predicted to alter the electronic properties, band structure and density of states of graphene and is likely to impact upon its unique transport properties.

SW defects involve an in-plane 90° rotation of two carbon atoms with respect to the midpoint of the bond. In this transformation four hexagons are changed into two heptagons and two pentagons as shown in Fig. 7(a). Some researchers suggest that SW defects may enhance the tendency of graphene layers to transform into nonplanar nanostructures such as nanotubes and fullerenes and that SW defects may play a role in the intrinsic rumpling of graphene. For this reason we have considered the channel GNR with a single SW defect otherwise planner sheet may transform into nonplanar sheet.

B. Edge Roughness

The edge of the channel nanoribbon may distort if a number of atoms are missing in its edge as shown in Fig. 7(b). Edge roughness or distortion primarily occurs during fabrication process. In such cases the nanoribbon loses its chirality and show inconsistent bandgap behaviour along its length. Such defects can also alter the electronic properties, band structure and have an adverse impact upon its transport properties.

C. Effects of Defects on GNR-FET Performance

Fig. 8 shows the transmission spectrum for defected (Stone-Wales and Edge Roughness) and defectless GNR-FET. Comparing the results it is evident that transmission spectrum is strongly suppressed due to scattering by both types of distortion at higher energies. For edge-distorted ribbon, due to loss of chirality, transmission at bandgap energy range has been observed. From this transmission behaviour *I-V* characteristics of defected GNR-FET are computed (Fig. 9). Current in the device suffering from Stone-Wales defectdegrades significantly due to transmission spectrum suppression. In case of edge distorted GNR-FET higher valueof current is found at lower bias due to transmission at bandgap energy range.

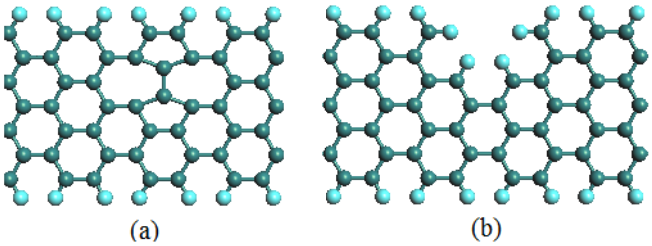


Fig. 7 (a) Stone-Wales defect (b) Edge Roughness.

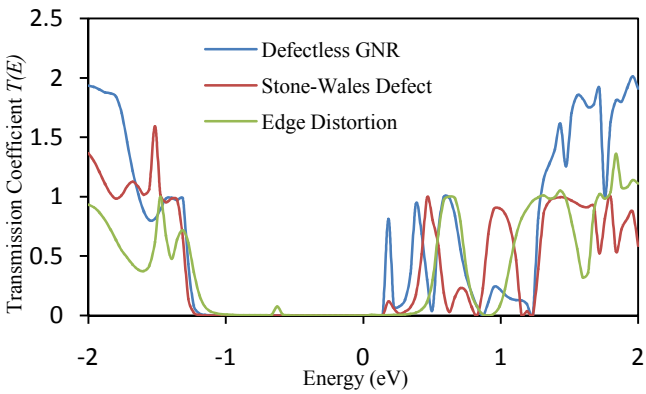


Fig. 8 Transmission spectrum for defected and defectless GNR.

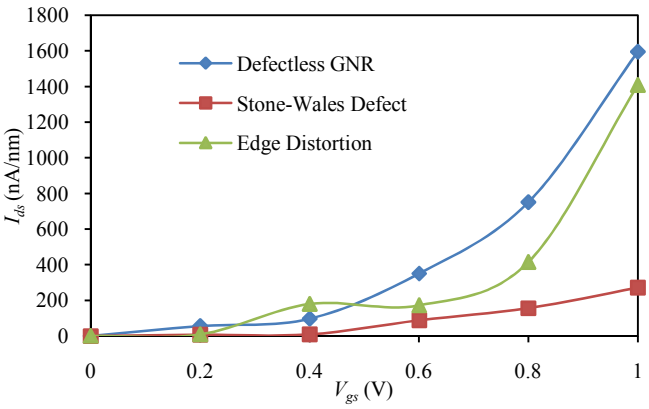


Fig. 9 I-V_{ds} curve of defected and defectless GNR (V_{gs}=1V).

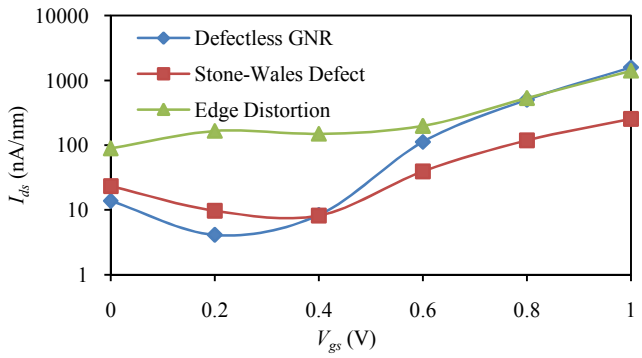


Fig.10 I_{ds}-V_{gs} curve of defected and defectless GNR (V_{ds}=1V).

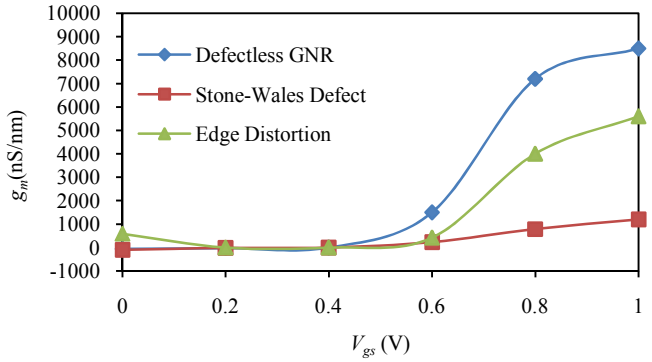


Fig.11 g_m-V_{gs} curve of defected and defectless GNR (V_{ds}=1V).

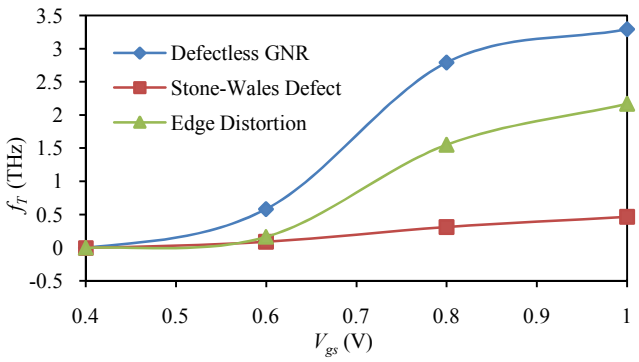


Fig. 12 f_T at different gate voltages for the device (V_{ds}=1V).

Switching performance of the GNR-FET suffering from defects can be determined from Fig. 10. On/Off ratio for SW defected FET and Edge Distorted FET is ~11 and ~16 respectively where for defectless device the ratio is ~10^3, which indicates significant switching performance degradation.

Frequency performance of the defected GNR-FETs can be analysed by first obtaining the transconductance, g_m of the device. Fig. 11 shows values of g_m for different gate voltages at V_{ds}=1V for both defected and defectless GNRs. From the figure it is evident that defect causes significant degradation of the gate voltage control over the channel region. As a result values of g_m reduce significantly for defected GNRs from its defectless case. Cutoff frequency of the device is then obtained using equation (1). Fig. 12 shows the values of f_T at different gate voltages. For analog circuit application, where the defectless device can operate up to 3.3THz, edge distortion limits the operational frequency to 2.2THz causing ~33% reduction in f_T. Stone-Wales defects show severe cutoff frequency degradation by limiting its value to 465GHz causing ~86% degradation. In conclusion we can say that, frequency performance of the GNR devices strongly depends on the fabrication process and introduction of defects causes significant alteration in the cutoff frequencies of the devices.

V. Conclusions

In this paper, we have performed NEGF simulation to demonstrate the intrinsic current-voltage characteristics of ultra-thin GNR-FETs. The FETs made from intrinsic semiconducting Armchair nanoribbons can exhibit very high levels of performance, with On/Off ratio up to ~10^3, transconductance of 8.5×10^3nS/nm which leads to a cutoff frequency up to 3.3THz. Then the effects of Stone-Wales defects and Edge Roughness is analysed. Simulation is performed to show that introduction of such defects during fabrication process may alter the switching performance of the GNR-FETs and degrade the overall frequency performance as well.

References

[1] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, and A. A. Firsov, “Electric field effect in atomically thin carbon films,” *Science*, vol. 306, no. 5696, pp. 666–669, Oct. 2004.

[2] A. K. Geim, “Graphene: Status and Prospects,” *Science*, vol. 324, no. 5934, pp. 1530–1534, Jun. 2009.

[3] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, M. I. Katsnelson, I. V. Grigorieva, S. V. Dubonos and A. A. Firsov, “Two-dimensional gas of massless Dirac fermions in graphene,” *Nature*, vol. 438, no. 7065, pp. 197–200, Nov. 2005.

[4] M. Y. Han, B. Ozyilmaz, Y. B. Zhang, and P. Kim, “Energy band-gap engineering of graphene nanoribbons,” *Phys. Rev. Lett.*, vol. 98, no. 20, p. 206805, May 2007.

[5] Y. W. Son, M. L. Cohen, and S. G. Louie, “Energy gaps in graphene nanoribbons,” *Phys. Rev. Lett.*, vol. 97, no. 21, p. 216803, Nov. 2006.

[6] L. Jiao, L. Zhang, X. Wang, G. Diankov, and H. Dai, “Narrow graphene nanoribbons from carbon nanotubes,” *Nature*, vol. 458, no. 7240, pp. 877–880, Feb. 2009.

[7] X. Li, X. Wang, L. Zhang, S. Lee, and H. Dai, “Chemically Derived, Ultrasmooth Graphene Nanoribbon Semiconductors,” *Science*, vol. 319, no. 5867, pp. 1229–1232, Feb. 2008.

[8] Y. Yoon, and J. Guo, “Effect of edge roughness in graphene nanoribbontransistors,” *Appl. Phys. Lett.*, vol. 91, no. 7, p. 073103, Aug. 2007.

[9] A. K. Saha, G. Saha, M.A.Al Shohel, and A. B. M. H.U. Rashid, “I-V characterization of graphene nanoribbon FET coupling Schrödinger-Poisson equation," in *Proc. ICIEV*, 2013, pp. 1-4.

[10] Y. Yoon, G. Fiori, S. Hong, G. Iannaccone, and J. Guo, “Performance Comparison of Graphene NanoribbonFETs With Schottky Contacts and Doped Reservoirs,” *IEEE Trans. Electron Devices*, vol. 55, no. 9, pp. 2314–2323, Sep. 2008.

[11] S. Datta, “Nanoscale device modeling: The Green’s function method,” *Superlattices Microstruct.*, vol. 28, no. 4, pp. 253–278, Oct. 2000.

[12] L. Yang, C. Park, Y. Son, M. L. Cohen, and S. G. Louie, “Quasiparticle Energies and Band Gaps in Graphene Nanoribbons,” *Phys. Rev. Lett.*, vol. 99, no. 18, p. 186801, Nov 2007.

[13] Q. Yan, B. Huang, J. Yu, F. Zheng, J. Zang, J. Wu, B. Gu, F. Liu, and W. Duan, “Intrinsic Current-Voltage Characteristics of Graphene Nanoribbon Transistors and Effect of Edge Doping,” *Nano Lett.*, vol. 7, no. 6, pp. 1469–1473, Apr. 2007.

[14] F. Xia, V. Perebeinos, Y. Lin, Y. Wu, and P. Avouris, “The origins and limits of metal–graphene junction resistance,” *Nature Nanotech.*, vol. 6, no. 3, pp. 179–184, Feb. 2011

[15] S. Reich, J. Maultzsch, C. Thomsen, and P. Ordejón, “Tight-binding description of graphene,” *Phys. Rev. B*, vol. 66, p. 035412, Jul. 2002.

[16] M. C. Lemme, T. J. Echtermeyer, M. Baus, H. Kurz, "A Graphene Field Effect Device," *IEEE Electron Device Lett.*, vol. 28, no. 4, pp. 282-284, Apr. 2007.

[17] A. J. Stone, and D. J. Wales, "Theoretical studies of icosahedral C60 and some related species," *Chem. Phys. Lett.* vol. 128, no. 5-6, pp. 501-503, Aug. 1986.

Uncoupled Mode Space Approach Towards Transport Modeling of Gate-all-around $In_xGa_{1-x}As$ Nanowire MOSFET

Saeed Uz Zaman Khan*, Md. Shafayat Hossain, Fahim Ur Rahman, Rifat Zaman,
Md. Obaidul Hossen, and Quazi D. M. Khosru

Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology
Dhaka-1205, Bangladesh
*anim.buet@gmail.com

Abstract—**Since the fabrication of first III-V Gate-All-Around (GAA) MOSFET it is under extensive research, as it is one of the potential candidates to replace the state of art tri-gate FinFETs, to continue progressive scaling. In this work, transport characterization of experimentally demonstrated gate-all-around (GAA) $In_xGa_{1-x}As$ nanowire MOSFET in near-ballistic regime is performed using 3D self-consistent Schrödinger-Poisson solver based on Uncoupled Mode Space approach, taking wave function penetration and other quantum mechanical effects into account. The effects of channel length variation on transport characteristics are also examined.**

Index Terms—**Ballistic Transport; Mode Space Approach; Gate-all-around MOSFET; $In_xGa_{1-x}As$ Nanowire; III-V Channel Material.**

I. Introduction

ITRS road map suggests that, to maintain progressive scaling of transistors, extensive research should be done on improving Multi-gate devices and alternate channel structures such as nanowires [1]. Although, 22nm technology Trigate FinFETs are commercially available from 2012, to conform Moore's law transistors must be scaled further to sub-10nm very soon where aggressive short channel effects (SCEs) are prominent. Gate-all-around (GAA) structure with Si nanowire channel has proven most resistant to SCEs [2]-[6] implying that it can be a feasible solution for reaching that goal. Apart from the structure of the device, ITRS roadmap also suggests involvement of III-V materials in the channel for further research [1]. III-V FinFETs [7]-[8] and multi-gate quantum-well FETs [9] provide performance improvements over si-FinFETs in deep submicron regime offering better speed and higher ON/OFF current ratio. Taking the advantages of both GAA architecture and III-V nanowire channel, Gate-all-around $In_xGa_{1-x}As$ nanowire MOSFET with ALD Al_2O_3 as gate dielectric has been experimentally demonstrated by Gu et al. [10] in recent past where better performance than existing FinFETs has been reported. Although, TCAD modeling of transport characteristics is available for this device [11], theoretical transport modeling that concurs with the experimentally reported result is yet to appear in the literature. In this work, rigorous transport characterization of that experimental Gate-all-around $In_xGa_{1-x}As$ nanowire MOSFET is done in ballistic regime and effects of channel length variation is studied using 3D self-consistent simulator based on Uncoupled Mode Space (UMS) approach with a step-like transmission co-efficient.

II. Device Structure

The device used in this study is an $In_{0.53}Ga_{0.47}As$ nanowire FET identical to the experimental demonstration by Gu et al. [10]. It is a gate-all-around device comprising four gates connected together on all four sides of an oxide (Al_2O_3) - $In_{0.53}Ga_{0.47}As$ structure having square-shaped cross-section. The channel material is $In_{0.53}Ga_{0.47}As$ and 10nm thick Al_2O_3 serves as the gate dielectric. The source and drain regions are formed with n+ $In_{0.53}Ga_{0.47}As$. The substrate is 30nm×30nm p+ InP. The doping for p- $In_{0.53}Ga_{0.47}As$ channel is Na=2 × 10^{16} cm^{-3} (Fig. 1).

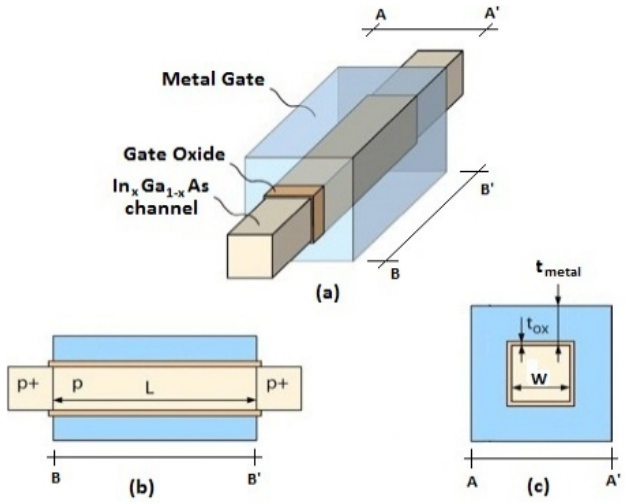


Fig. 1 (a) Schematic 3D view of an inversion-mode GAA n-channel $In_{0.53}Ga_{0.47}As$ (2×10^{16} cm^{-3}) MOSFET with ALD 10nm Al_2O_3/20nm WN gate stacks, a heavily doped wide bandgap InP lies underneath the bottom gate [10]. (b) Lateral cross section showing channel L_{ch}=50nm. (c) Vertical cross section of GAA device under simulation showing fin-width W=30nm, oxide thickness t_{ox}=10nm and metal gate width t_{metal}=20nm.

III. Modeling Approach

A. Self-Consistent Schrödinger- Poisson Solver

The developed self-consistent simulator involves numerical solution of two partial differential equations in a coupled manner, namely Poisson's equation and Schrödinger's equation along with numerical integration. Finite Element Method is used to solve the Partial differential equations (PDE). COMSOL Multiphysics is used as the PDE solver tool which is invoked from MATLAB by using the scripting language of COMSOL. In case of solving Poisson's equation, at the external boundaries of the gates Dirichlet i.e. fixed

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

voltage boundary condition is used and Neumann i.e. continuous electric flux boundary condition is used at all internal boundaries. For Schrödinger's equation, all boundaries are kept as open boundaries to allow wave function penetration.

In the oxide region, charge density (ρ) is zero. For channel region, charge density is obtained from to the following expression.

$$\rho(y, z) = q\sum_i \int_{Ei}^{\infty} D(E)\, f(E)\, |\psi_i\,(y,z)|^2\, dE \qquad (1)$$

Here, i is the number of sub-band, *D(E)* is 1-D density of states and *f(E)* is Fermi-Dirac distribution function with respect to source/drain Fermi level. E_i and ψ_i are Eigen energies and corresponding normalized wave functions obtained from 2D Schrödinger's equation. Ten Eigen energies are considered in charge calculation.

Starting with zero initial charge profile, Poisson's and Schrödinger's equations are solved using mixing coefficient (α) of 0.04 until self-consistency between charge and potential profile is achieved. The potential profile is updated after every iteration using the following expression.

$$V_{final} = V_{new} \times \alpha + V_{old} \times (1-\alpha) \qquad (2)$$

Once the solution converges, the charge distribution inside the device can be determined.

B. Mode Space Approach

A nanowire transistor has a three-dimensional (3D) distribution of electron density and electrostatic potential. As a result, a 3D simulator is required for the simulation of nanowire transistors. The calculation involves a self-consistent solution of a 3D Poisson's equation and a 3D Schrödinger equation with open boundary conditions. Using the finite element method, we solve the 3D Poisson equation to obtain the electrostatic potential. At the same time, we solve the 3D Schrödinger by Uncoupled Mode Space (UMS) approach which provides both computational efficiency and high accuracy as compared with direct real space calculations [12]-[13]. In UMS instead of solving Schrödinger equation for entire device, multiple 2D cross sections are solved and the resultant Eigen energy and wave functions are combined to get the solution for 3D structure [13]. Once self-consistency is achieved, the simulator provides the Eigen energies of different modes and the current is computed by,

$$I_{SD} = \frac{2e}{h}\int_{-\infty}^{+\infty} T(E)(f(\mu_S, E) - f(\mu_D, E))dE \qquad (3)$$

Where, $f(\mu_S, E)$ and $f(\mu_D, E)$ are Fermi-Dirac distribution functions at source and drain end respectively.

For the calculation of electron current, the total transmission coefficient can be written as a summation of the transmission coefficient, $T_m(E)$ for each mode m,

$$T(E) = \sum_{m=1}^{M} T_m\,(E) \qquad (4)$$

The Transmission co-efficient, T(E) can be computed using NEGF formalism and introducing Buttiker Probes the effect of scattering can be calculated[14]-[15]. To avoid computational complexity in this work, only ballistic transport is considered assuming a step-like [16] transmission probability curve.

IV. Results and Discussion

Fig. 2(a) and 2(b) show the 3D potential profile of the device for on and near-threshold conditions respectively. On the other hand, Fig. 3(a) and 3(b) respectively show the electron concentration throughout the channel for those bias conditions. These figures demonstrate the channel formation of the device under study. Since, the cross section of the device is fairly large (30nm x 30nm) the channel is formed near the oxide interface indicating surface inversion, as expected [17].

Fig. 4 depicts, the transfer characteristics and trans-conductance of the device for different bias conditions. From Fig. 5(a) sub-threshold swing (SS) is found to be 285 mV/dec which is greater than the experimentally reported value, 168 mV/dec [10]. On the other hand, threshold voltage is extracted from Fig. 5(b) as -0.7V, which matches with the experimental threshold of -0.68V [10].

Fig. 6(a) shows the output characteristics of the device, revealing DIBL of 64.57mV/V for 50nm channel length. Finally, Fig. 6(b) displays the output characteristics for different channel length, showing sharp increase in DIBL effect with decrease in channel length, as expected. It is to be noted that, the drain current calculated here, is more than twice of the experimental current reported by Gu et al. [10], which is quite reasonable since no scattering is considered in the developed transport model.

V. Conclusions

Using 3D simulation technique, the transport properties of $In_xGa_{1-x}As$ GAAFET is characterized in ballistic regime. Uncoupled Mode Space approach adopted with step-like transmission profile shows consistent results for studying ballistic transport properties of $In_xGa_{1-x}As$ GAAFET. The channel length variation is also performed showing that the decrease in channel length results in lower current but increased DIBL. The result is coherent with the trend of reported experimental data.

Acknowledgment

The authors are grateful to Dept. of EEE, Bangladesh University of Engineering and Technology for providing necessary facilities for carrying out this research.

References

[1] (2013) ITRS website. [Online]. Available: http://www.itrs.net/Links/2013ITRS/Home2013.htm

[2] M. Si., X. Lou, X. Li, J. J. Gu, H. Wu, X. Wang, J. Zhang, R. G. Gordon, and P. D. Ye, "Performance enhancement of gate-all-around InGaAs nanowire MOSFETs by raised source and drain structure," *in Proc. Device Research Conference (DRC)*, 2013, pp. 19-20.

[3] J. -H. Lee, B. -S. Kim, S. -H. Choi, Y. Jang, S. W. Hwang, and D. Whang, "A facile route to Si nanowire gate-all-around field effect transistors with a steep subthreshold slope," *Nanoscale*, vol. 5, no. 19, pp. 8968-8972, 2013.

[4] Y. Tian, R. Huang, Y. Wang, J. Zhuge, R. Wang, J. Liu, X. Zhang, and Y. Wang, "New self-aligned silicon nanowire transistors on bulk substrate fabricated by epi-free compatible CMOS technology: Process integration, experimental characterization of carrier transport and low frequency noise," *in Proc. IEDM 2007*, pp. 895-898.

[5] N. Singh, K. D. Buddharaju, S. K. Manhas, A. Agarwal, S. C. Rustagi, G. Q. Lo, N. Balasubramanian, and D. -L. Kwong, "Si, SiGe nanowire devices by top–down technology and their applications," *IEEE Trans. Electron Device,* vol. 55, no. 11, pp. 3107-3118, 2008.

[6] S. D. Suk, S.-Y. Lee, S.-M. Kim, E.-J. Yoon, M.-S. Kim, M. Li, C. W. Oh, K. H. Yeo, S. H. Kim, D.-S. Shin, K.-H. Lee, H. S. Park, J. N. Han, C. J. Park, J.-B. Park, D.-W. Kim, D. Park, and B.-I. Ryu, "High performance 5 nm radius twin silicon nanowire MOSFET(TSNWFET): Fabrication on bulk Si wafer, characteristics, and reliability," *in Proc. IEDM 2005,* pp. 552-555.

[7] Y. Q. Wu, R. S. Wang, T. Shen, J. J. Gu, and P. D. Ye, "First experimental demonstration of 100 nm inversion-mode InGaAs FinFET through damage-free sidewall etching," *in Proc. IEDM 2009,* pp. 331-334.

[8] H. –C. Chin, C., X. Gong, L.Wang, H. K. Lee, L. Shi, and Y. –C. Yeo, "III–V Multiple-Gate Field-Effect Transistors With High-Mobility Channel and Epi-Controlled Retrograde-Doped Fin," *IEEE Electron Device Lett.*, vol. 32, no. 2, pp.146-148, 2011.

[9] M. Radosavljevic, G. Dewey, J. M. Fastenau, J. Kavalieros, R. Kotlyar, B. Chu-Kung, W. K. Liu et al., "Non-planar, multi-gate InGaAs quantum well field effect transistors with high-k gate dielectric and ultra-scaled gate-to-drain/gate-to-source separation for low power logic applications," *in Proc. IEDM 2010,* p. 126.

[10] J. J. Gu, Y. Q. Liu, Y. Q. Wu, R. Colby, R. G. Gordon, and P. D. Ye, "First experimental demonstration of gate-all-around III–V MOSFETs by top-down approach," *in Proc. IEDM 2011,pp.32.2.1-32.2.4.*

[11] Gu, J. J., Wu, H., Liu, Y., Neal, A. T., Gordon, R. G., & Ye, P. D, "Size-Dependent-Transport Study of Gate-All-Around Nanowire MOSFETs: Impact of Quantum Confinement and Volume Inversion," *IEEE Electron Device Lett*., vol.33, no. 7, pp. 967-969, 2012.

[12] E. Polizzi, and N. Ben Abdallah, "Self-consistent three dimensional models for quantum ballistic transport in open systems," *Physical Review B,* vol. 66, no. 24, p. 245301, 2002.

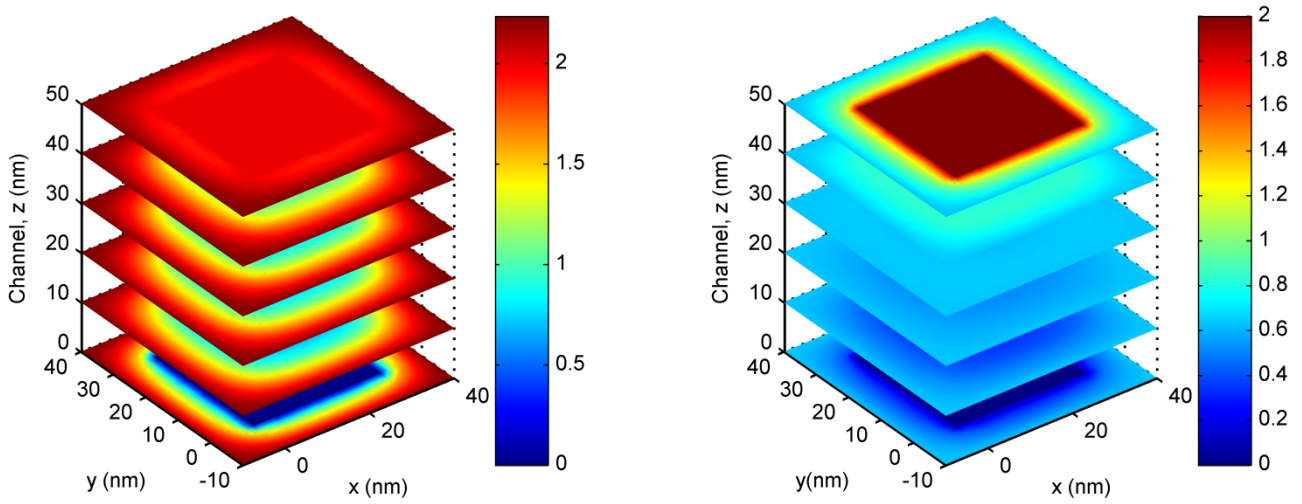
[13] R. Venugopal, Z. Ren, S. Datta, M. S. Lundstrom and D. Jovanovic, "Simulating quantum transport in nanoscale transistors: Real versus mode-space approaches," *J. Appl. Phys.,* vol. 32, no. 7, p. 3730, 2002.

[14] J. Wang, E. Polizzi, and Mark Lundstrom, "A Three-Dimensional quantum simulation of Silicon nanowire transistors with the effective mass approximation," *Journal of Applied Physics,* vol. 96, no. 4, pp. 2192-2203, 2004.

[15] S. Datta, *Quantum Transport: Atom to Transistor*, 1st ed., Cambridge University Press, 2005.

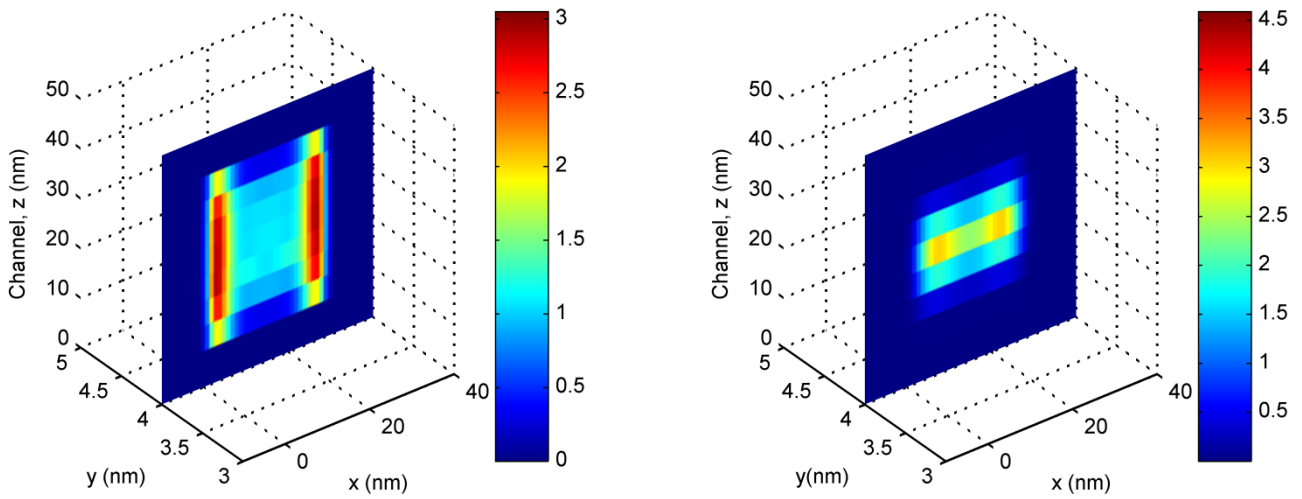
[16] O. Kurniawan, M. F. Ng, W. S. Koh, Z. Y. Leong, and E. Li, "Simplified model for ballistic current–voltage characteristic in cylindrical nanowires," *Microelectronics Journal,* vol. 41, no. 2, pp. 155-161, 2010.

[17] S. R. N. Yun, C. G. Yu, J. T. Park,C. W. Lee, D. Lederer, A. Afzalian, R. Yan, J. P. Colinge, "Quantum Definition of Threshold Voltage in MuGFETs," *In Proc. IEEE Int. Conf. on SOI* 2007, pp.137-138.



(a) (b)

Fig. 2 3D potential profile of the Device at (a) on condition (V_{gs}=1V, V_{ds}=2V) (b) near-threshold condition (V_{gs}=-0.6V, V_{ds}=2V).



(a) (b)

Fig. 3 Across the channel of the device (a) Electron concentration (x10^{25} m^{-3}) at on condition (V_{gs}=1V, V_{ds}=2V) (b) Electron concentration (x10^{20} m^{-3}) near-threshold condition (V_{gs}=-0.6V, V_{ds}=2V).

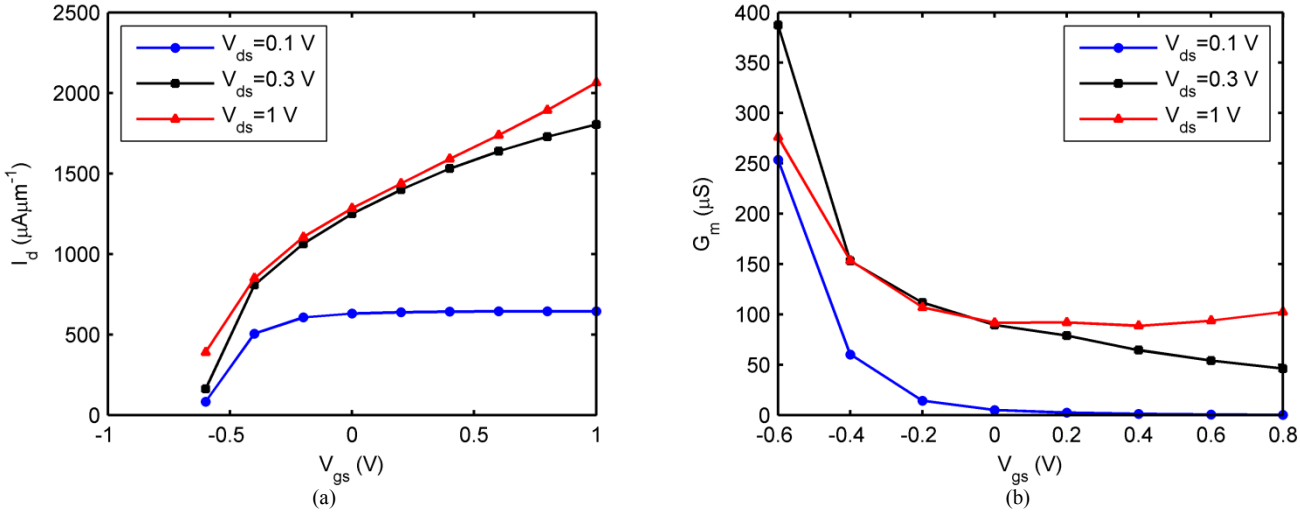


Fig. 4 (a) Transfer Characteristics (b) Transconductance.

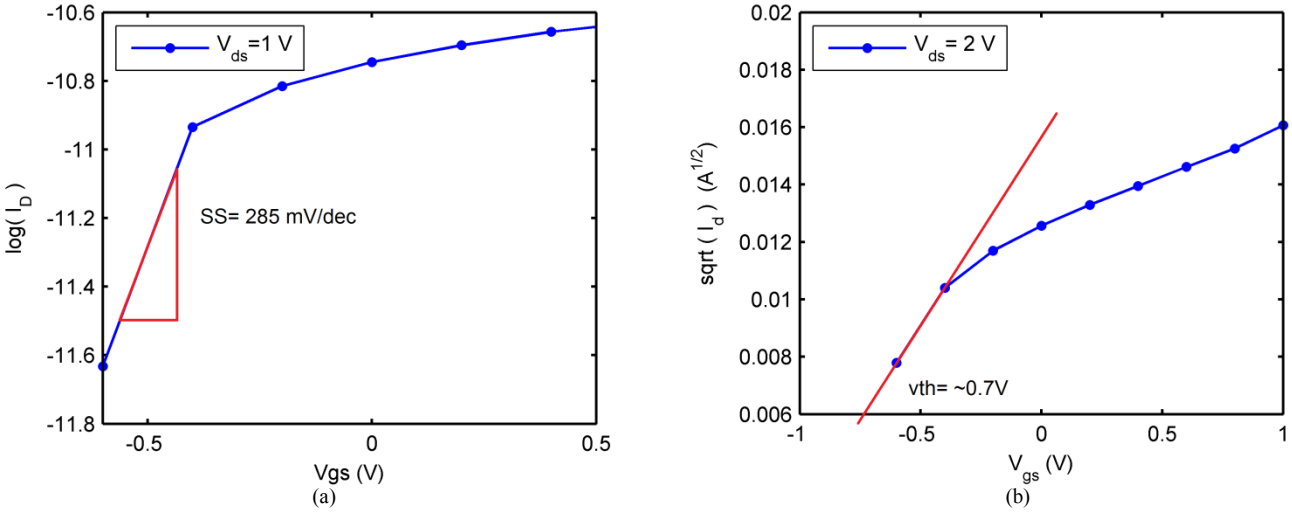


Fig. 5 Extraction of (a) Subthreshold Slope (SS) (b) Threshold Voltage.

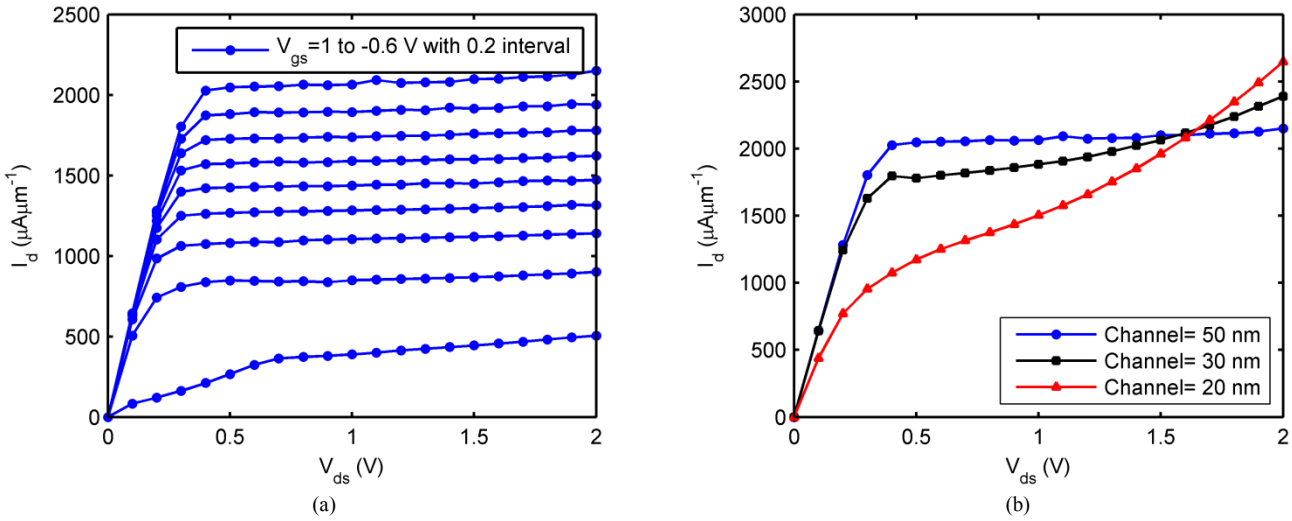


Fig. 6 (a) Output characteristics (DIBL=64.57mV/V) (b) Output characteristics (at V_G=1V) with variation in channel length. DIBL increases remarkably with decreasing channel length.

Thermal Management of FET Devices Using Graphene Heat Spreader

S.M.R. Islam, N. Saquib and S. Subrina*

Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology
Dhaka-1200, Bangladesh

*samiasubrina@eee.buet.ac.bd

Abstract—**Self-heating is a severe problem in electronic devices since local heat generation and temperature rise limit the application of the devices. Here we demonstrate that thermal management of Field Effect Transistor (FET) devices can be substantially improved by introducing alternative heat-escaping channels implemented with graphene. Graphene is a good candidate for heat removal owing to its extremely high thermal conductivity. We simulated heat propagation in different GaAs FET circuits, namely metal oxide semiconductor field effect transistor (MOSFET), metal semiconductor field effect transistor (MESFET) and heterostructure field effect transistor (HFET), without and with graphene lateral heat spreaders. We also simulated the devices with other heat spreaders and compared their performance with that of grapheme. Numerical solutions of the heat propagation equations were obtained using the finite element method. It was found that the incorporation of graphene or few-layer graphene (FLG) layers with proper heat sinks can substantially lower the temperature of the localized hot spots. The maximum temperature in the FETs was studied as function of graphene's thermal conductivity and the thickness of FLG. The developed model and obtained results are important to solve the thermal issues in FET devices.**

Index Terms—**FET devices, graphene, heat conduction, heat spreaders, thermal management.**

I. Introduction

Continuous miniaturization of electronic devices and circuits, increased speed and integration densities resulted in problems with thermal management. As the electronic industry moves toward nanometer designs, one of the most important challenges is growing chip power consumption as well as the issue of self-heating. When we scale the device using constant voltage scaling, there has been a significant increase of drain current as a result of short channel length. It leads to an increased power density which in turn increases the device temperature substantially. As a result, temperature rise in devices can lead to performance degradation and early thermal breakdowns [1]–[3]. So, efficient thermal management plays a significant role in device design for long-term reliability and optimum performance [4]. Since the performance of ULSI circuits depends on temperature, even a small increase in temperature results in reduction of the device lifetime. A possible approach for solving the thermal problem is to find a material with extremely high thermal conductivity. Conventional materials such as Cu or Al used for heat removal has lower thermal conductivity. Hence for efficient heat management an alternative material named graphene, the latest of the discovered allotropes of carbon with extremely high intrinsic thermal conductivity, is a potential candidate. The room temperature thermal conductivity of the suspended graphene is determined to be ranging from 3080 – 5300 W/mK [5]–[6]. Thermal conductivity of graphene revealed strong flake size dependence [6]–[8]. Many studies demonstrated that quality graphene has very high electron mobility (~$2\times10^5 cm^2V^{-1}s^{-1}$), low resistivity (~$10^{-8}$Ohm-m) and very low thermal boundary resistance (~$3.7\times10^{-9}m^2KW^{-1}$) [9]–[12]. The unique properties of graphene make it a promising material for the electronic devices, electrical interconnects and thermal management [5] – [8], [13]. In this work we carry out the feasibility study graphene heat spreader and compare its performance with that of conventional heat spreader for thermal management of FET family.

II. Model of Heat Conduction

To study the feasibility of the use of graphene heat spreader with a view thermal management of FET devices, we simulate heat propagation in GaAs MOSFET, GaAs MESFET and GaAs/AlGaAs HFET with and without graphene heat spreader. The thicknesses of SiC substrateand epitaxial GaAs film are assigned to be 100 μm and 0.2 μm while their thermal conductivities are 490 W/mK and 46 W/mK, respectively [14]. The thickness and thermal conductivity of AlGaAs in HFET is 8 nm and 46 W/mk respectively. The gate oxide layer for MOSFET is of 10 nm thick and its thermal conductivity is 1.4 W/mK. A conventional heat sink is attached to the device structure at the bottom. A graphene heat spreader layer, when used, is connected to the drain region with a side heat sink forming the channel for heat escape as shown in Fig. 1. The heat spreader is insulated from the drain terminal by a thin SiO_2 layer. The arrangement of heat spreader and heat sink is common for MOSFET, MESFET and HFET as shown in the prototype device (Fig. 1).The simulations were carried out with the help of finite element analysis using COMSOL software [15].

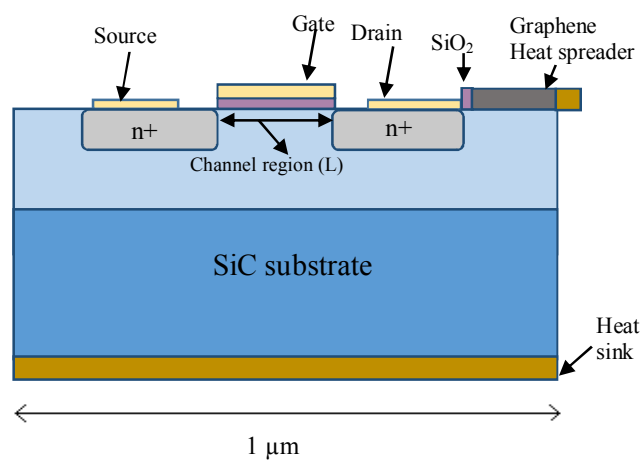


Fig. 1. Schematic view of a prototype FET device (MOSFET) with graphene heat spreader attached to the side heat sink and a main heat sink at the bottom.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

We have modeled the self-heating by solving Joule heating numerically within the device since current in the device channel causes heat generation and the concerned equations can be expressed as

$$E = -\nabla V \quad (1)$$
$$J = \sigma E + J_e \quad (2)$$
$$\nabla . J = Q_j \quad (3)$$

where J is the current density, σ is electrical conductivity, V is the applied voltage and E is the corresponding electric field and Q is the generated heat within the device. The heat transport along graphene layers was assumed to be diffusive since the size of the layer is much larger than the phonon mean free path [6] – [8]. The heat propagation within the device was modeled by solving the Fourier's law of heat conduction.

$$-\nabla \cdot (k \nabla T) = Q \quad (4)$$

where Q is the heat source i.e. the heat energy generated within a unit volume per unit time, T is the absolute temperature and k is the thermal conductivity. The bottom surface of the substrate and the far end of the graphene heat spreader were kept at a constant temperature T = 300K. The external surfaces were modeled in such way that they are kept insulated from environment.

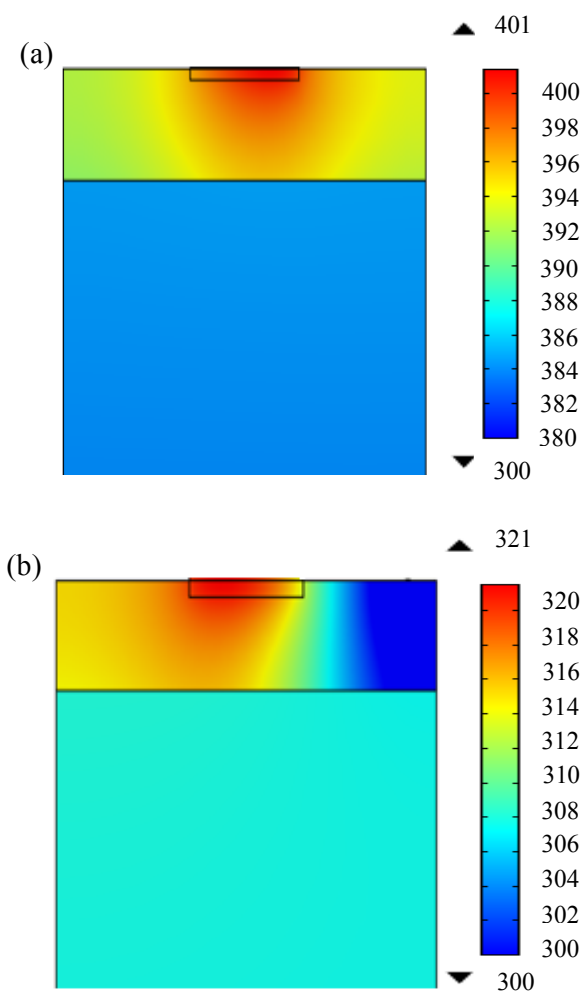


Fig. 2. Temperature profile for GaAs MOSFET (a) without heat spreader(b)with graphene heat spreader for drain voltage V_d=3V,gate voltage V_g=2V.

III. SIMULATION RESULTS

Fig. 2 shows the simulated temperature profiles for the MOSFET without and with the graphene lateral heat spreader. The maximum temperature of MOSFET decreases by 80 K when graphene layer is incorporated into the device for drain voltage of 3V and gate voltage of 2V. The thermal conductivity of single layer graphene (thickness ~ 0.35nm) is considered to be 5000 W/mK. The drop in temperature into the device structure with the heat spreader also depends on the device size and geometry, thickness of the substrate and dissipated power across device.

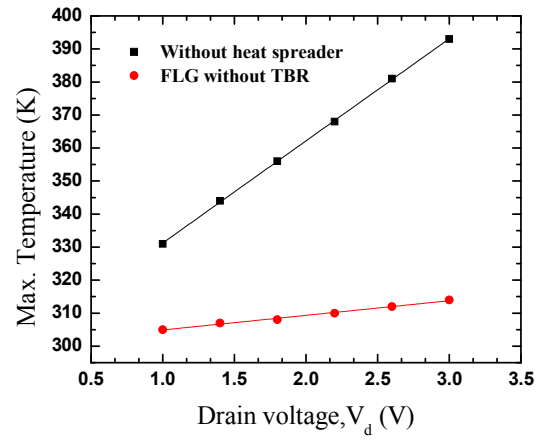


Fig. 3 Maximum temperature in MOSFET as a function of drain voltage without and with graphene heat spreader.

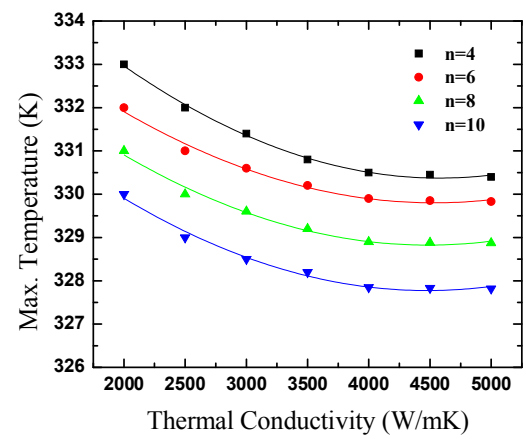


Fig. 4. Maximum temperature in MOSFET as a function of the thermal conductivity of the graphene heat spreader with different number of graphene layers.

Fig. 3 shows the reduced temperature rise as a function of drain voltage while graphene heat spreader is attached to the MOSFET. It shows that at higher drain voltage, higher current causes higher temperature rise. Maximum temperature increases almost linearly with the increase of drain voltage. And the drop in temperature with graphene is higher at higher drain voltage.

Fig. 4 represents the effect of variation of thermal conductivity as well as the thickness of graphene layer on maximum temperature of MOSFET. The maximum temperature is decreased with the increase of thermal conductivity for a given number of graphene atomic layers. So we simulated the device for different no. of graphene layers.

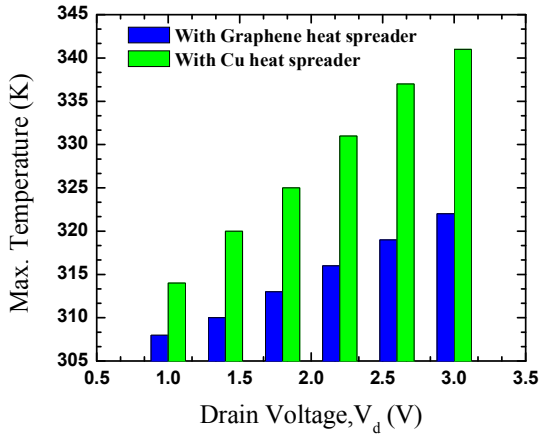


Fig. 5. Performance comparison between the MOSFETs with graphene and copper heat spreaders.

For a given thermal conductivity maximum temperature is reduced significantly with the increase of number of atomic layers of graphene. This improvement occurs at constant thermal conductivity since the higher the number of atomic layers of graphene, the larger the surface area for heat removal. In literature, it is reported that thermal conductivity of FLG decreases with the increase of number of atomic layers [5].In the simulations, the thermal conductivity of FLG was assumed to be ranging from 1000 W/mK to 5000 W/mK.

Fig 5 shows a comparison between the performances of graphene heat spreader with that of Cu. Graphene heat spreader gives the larger temperature reduction compared to conventional heat spreader like Cu heat spreader i.e. Cu shows the least improvement as heat spreader. Based on the dispersion calculations in graphitic materials, the phonon group velocity in graphene is higher than that in Cu, which leads to larger k which in turn causes better the thermal improvement. The thermal conductivity of Cu is reported to be 401W/m.K [16].

Fig. 6 and Fig. 7 indicate the temperature improvement using graphene heat spreader in MESFET and HFET, respectively. The temperature improvement is due to the superior thermal conductivity of graphene.

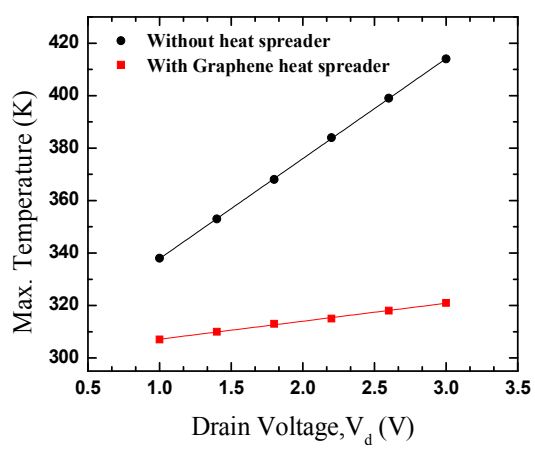


Fig. 6. Maximum temperature in MESFET as a function of drain voltage without and with graphene heat spreader.

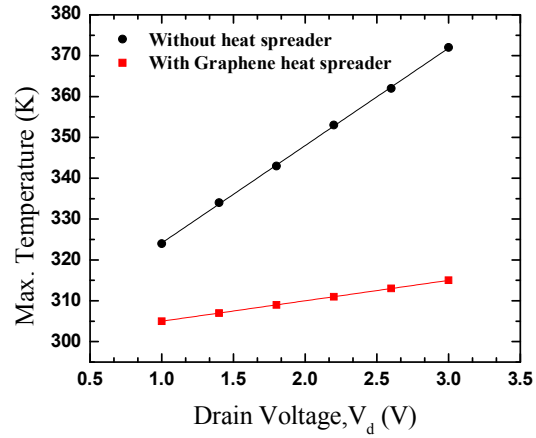


Fig. 7. Maximum temperature in HFET as a function of drain voltage without and with graphene heat spreader.

In Fig. 8 and Fig. 9 we show a comparison between graphene and Cu heat spreader for MESFET and HFET, respectively. In either case graphene outperforms the commonly used copper as heat spreading materials and gives the best result for thermal improvement.

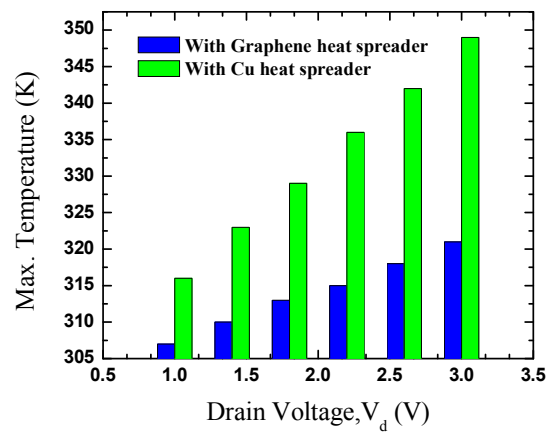


Fig. 8. Performance comparison between the MESFETs with graphene and copper heat spreaders.

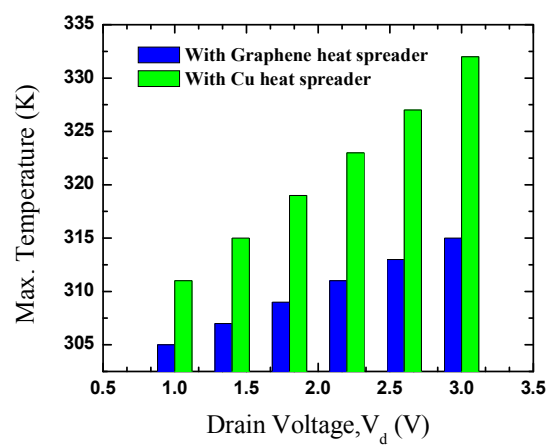


Fig. 9. Performance comparison between the HFETs with graphene and copper heat spreaders.

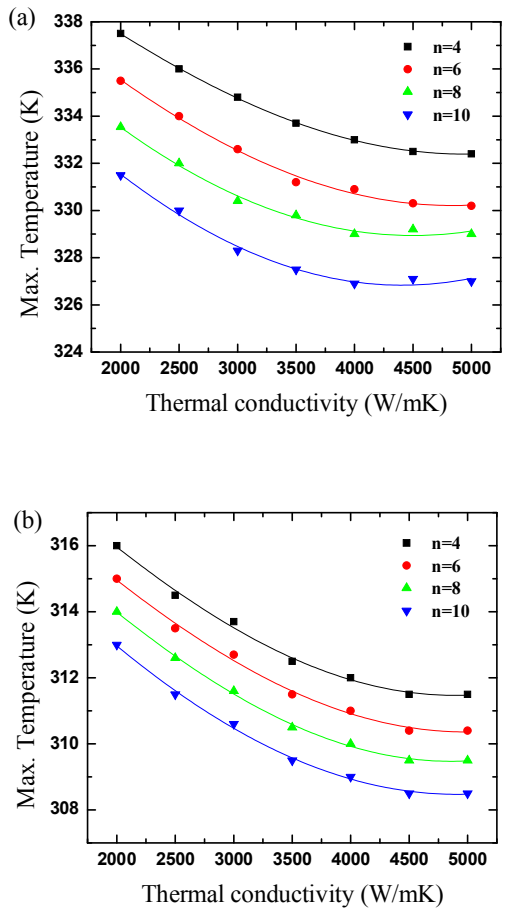


Fig.10. Maximum temperature in (a) MESFET and (b) HFET as a function of the thermal conductivity of the graphene heat spreader with different number of graphene layers.

Fig. 10 represents the effect of variation of thermal conductivity of graphene on maximum temperature of MESFET and HFET. Maximum temperature reduces by a great extent withthe increasing thermal conductivity and thickness of the spreader because it gives larger the surface area for heat removal. The application of few layer graphene is expected to be encouraging since it allows for greater flexibility with fabrication. On the contrary, the thermal conductivity of FLG is lower than that of graphene approaching that of bulk graphite along the basal plane (~2000 W/mK at room temperature). So it requires an optimization in order to achieve the best throughput. The heat generation and its mitigation in the FET family with different heat spreaders for a specific drain voltage, V_d =3V and gate voltage, V_g= 2V, is illustrated in TABLE I.

TABLE I
HEAT MITIGATION IN FET DEVICES WITH DIFFERENT HEAT SPREADERS

Heat spreader	MOSFET Max Temperature (K)	MESFET Max Temperature (K)	HFET Max Temperature (K)
Without heat spreader	401	414	372
Graphene heat spreader	321	323	314
Cu heat spreader	341	348	332

The above table illustrates the thermal improvement for different heat spreaders. Graphene provides the thermal improvement by 20K, 25K and 18K compared to conventional Cu heat spreader in MOSFET, MESFET and HFET, respectively. Since a small thermal improvement increases device lifetime, hence using graphene heat spreader also improves the device reliability. So graphene heat spreader gives an outstanding thermal improvement in FET devices in compare to other conventional heat spreaders.

IV. CONCLUSIONS

Graphene is demonstrated to be a promising heat spreader material for heat dissipation of hot spots with high heat flux. Our results show that the incorporation of graphene heat spreader can cause substantial reduction in temperature of the hot spots compared to other heat spreader i.e. copper. We also illustrate the incorporation of FLG heat spreaders which might be technologically more feasible than single layer graphene. The obtained results would be useful for thermal management of FET as well as electronic devices.

REFERENCES

[1] P. G. Mautry and J. Trager, "Investigation of self-heating in VLSI and ULSI MOSFET's," in *Proc. 1990 IEEE Conf Microelectronic Test Structures,* Mar. 1990, pp. 221-226.

[2] L. I. McDaid, S. Hall, W. Eccleston, and I. C. Alderman, "Monitoring the temperature rise in SOI transistors by measurement of leakage current*," in Proc. 1991 IEEE Int. SOI Conf,* Oct. 1991, pp. 28-29.

[3] L. T. Su, J. E. Chung, D. A. Antoniadis, K. E. Goodson, and M. I. Flik, "Measurement and modeling of self-heating in SOI MOSFETs*," IEEE Trans. Electron Devices,* vol. 41, pp. 69-75, Jan. 1994.

[4] M. Wegmuller, J. P. von der Weid, P. Oberson, and N. Gisin, "High resolution fiber distributed measurements with coherent OFDR," *Proc. ECOC'00*, 2000, paper 11.3.4, p. 109.

[5] A. A. Balandin, S. Ghosh, W. Bao, I. Calizo, D. Teweldebrhan, F. Miao, C. N. Lau, "Superior thermal conductivity of single-layer graphene," *Nano Lett*., vol. 8, pp. 902-907, Mar. 2008.

[6] S. Ghosh, I. Calizo, D. Teweldebrhan, E. P. Pokatilov, D. L.Nika, A. A. Balandin, W. Bao, F. Miao, and C. N. Lau, "Extremely high thermal conductivity of graphene: Prospects for thermal management applications in nanoelectronics," *Appl. Phys. Lett.*, vol. 92, p. 151911, Apr. 2008.

[7] D. L. Nika, E. P. Pokatilov, A. S. Askerov and A. A. Balandin,"Phonon thermal conduction in graphene: Role of Umklapp and edge roughness scattering," *Phys. Rev. B* 79, p. 155413, Apr. 2009.

[8] S. Subrina and D. Kotchetkov, "Simulation of Heat Conduction in Suspended Graphene Flakes of Variable Shapes," J. Nanoelectronics and Optoelectronics, vol. 3, p.1, 2008.

[9] S. V. Morozov, K. S. Novoselov, M. I. Katsnelson, F. Schedin,D. C. Elias, J. A. Jaszczak, and A. K. Geim, "Giant intrinsic carrier mobilities in graphene*," Phys. Rev. Lett.,* vol. 100, p. 016602, Jan. 2008.

[10] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, M. I. Katsnelson, I. V. Grigorieva, S. V. Dubonos, A. A. Firsov, "Two-dimensional gas of Dirac fermions in graphene," *Nature*, vol. 438, pp. 197-200, Nov. 2005.

[11] Y. Zhang, Y.-W. Tan, H. L. Stormer, P. Kim, "Experimental observation of the quantum Hall effect and Berry's phase in graphene," *Nature*, vol. 438, pp. 201-204, Nov. 2005.

[12] J. H. Chen, C. Jang, S. Xiao, M. Ishigami and M. S. Fuhrer*,"*Intrinsic and extrinsic performance limits of graphene devices on SiO2*," Nature Nanotech.,* vol. 3, pp. 206-209, Apr. 2008.

[13] Q. Shao, G. Liu, D. Teweldebrhan, A. A. Balandin*, "*Hightemperature quenching of electrical resistance in grapheme interconnects*," Appl. Phys. Lett.,* vol. 92, p. 202108, May 2008.

[14] M. Shamsa, P. M. Solomon, K. A. Jenkins, A. A. Balandin, W. Haensch*, "*Investigation of thermal crosstalk between SOI FETs by the subthreshold sensing technique*," IEEE Trans. Electron Devices,* vol. 55, pp. 1733-1740, Jul. 2008.

[15] [Online]. Available: http://www.comsol.com.

[16] H. Mayer, in C. A. Neugebauer, J. B. Newkirk and D. A. Vermilyea (eds.), Structure and Properties of Thin Films, Wiley, New York, 1959, p. 225.

Surface Potential Modeling of Junctionless Double Gate MOSFETs using Gradual Depletion Approximation

Shuvro Chowdhury,* Md. Ziaur Rahman Khan
[1]Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology
Dhaka-1000, Bangladesh
*schowdhury@eee.buet.ac.bd

***Abstract*—A new model for long channel Junctionless Double Gate (JL DG) MOSFETs based on gradual depletion inside the channel is presented. This new model uses a regional approach to determine surface potential of JL DG MOSFETs. Because of the gradual depletion approximation surface potential characteristics are found to have smooth transition between fully depleted region and partially depleted region. This smoothness may prove useful for modeling with SPICE like programs. The proposed model shows improved performance over the previous models based on abrupt depletion. Dependences of surface potential on various device and process parameters have also been studied. The validity of the proposed model has been checked with numerical simulations and TCAD tool. Upon comparison, it has been found that the results obtained from proposed model matches quite perfectly with those provided by numerical simulations and TCAD tool.**

***Index Terms*—Junctionless, Double-Gate, Surface potential, Abrupt depletion, Gradual depletion**

I. Introduction

In recent literature, junctionless (JL) double-gate (DG) field effect transistors (JL DGFET) have been reported [1], [2], [3] as promising candidates for future technology nodes. Proper operation of this device requires the device to be fabricated with a high content of the same impurity concentration within the channel and source/drain (S/D) regions. Hence these devices are junctionless. Junctionless FETs exhibit many advantages [4], [5], [6], [7], such as simple and flexible fabrication process, nearly ideal subthreshold slope (SS $\approx$ 60 mV/dec), high ON-OFF current ratio ($I_{\mathrm{ON}}/I_{\mathrm{OFF}} > 10^7$), low S/D series resistance, and small drain-induced barrier lowering. Moreover, the JL transistor shows some peculiar characteristics, like conductance oscillations at low temperature [8] and high temperature behavior [9].

Because of its excellent electrostatic performance, JL DG MOSFET has drawn attention for developing an accurate model to investigate its theoretical foundations, to better understand the behavior of this device, and to elucidate its strengths and weaknesses. Duarte et al. [10] first proposed a simple bulk current model relying on depletion approximation for the JL DGFET but neglected the accumulation condition. Then, Sallese et al. [11] carried out a space-charge-based model for the JL DG MOSFET to calculate the charge density and drain current, but it may cause convergence problems. Subsequently, Duarte et al. [12] proposed a non-piecewise model for the JL DG MOSFET based on parabolic potential approximation by a uniform charge-potential relationship. However, few studies have been carried out to develop a surface-potential-based compact model for the JL DG MOSFET valid in all regions, which includes a more physical and accurate description of the transistor behavior and is commonly accepted in compact modeling applications. Chen et al. [13] has proposed a surface potential based drain current model for JL DG MOSFET. This model is based on abrupt depletion between depleted and neutral regions and shows discontinuities in surface potential and drain current characteristics.

Here in this paper, a *gradual depletion approximation* has been used to determine the surface potential of JL DG MOSFETs. This approximation has made it posible to obtain a *continuous* surface potenatial-gate voltage characteristics. Significant improvement in surface potential modeling has also been achieved through this approximation which has been verified from comparison with numerical simulations.

II. Device description

The n-channel symmetric JL DG MOSFET structure used in this work has been shown in Figure 1. The device has a channel length L, device width W, and, channel thickness of t_{Si}. Gate oxide thickness is t_{ox} for both top and bottom gate oxides. A uniform donor doping concentration N_D within the channel, source and drain regions has been used. The dielectric constants of gate oxide material and channel are ϵ_{ox} and ϵ_{Si} respectively.

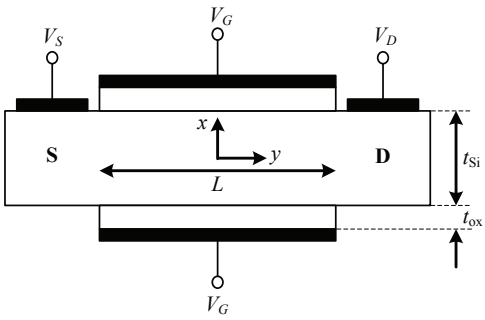


Fig. 1. JL DG MOSFET structure

Using the basic MOS capacitor equation it can be written that

$$V_G = V_{FB} + \phi_{ox} + \phi_s = \phi_{MS} - \frac{Q_{SC}(\phi_s)}{C_{ox}} + \phi_s \quad (1)$$

Here, V_G, ϕ_s, V_{FB}, ϕ_{ox} and ϕ_{MS} are applied gate voltage, surface potential, flatband voltage, voltage drop across oxide and metal-semiconductor workfunction difference respectively. The absence of parasitic oxide charge and interface trap charge density has been implicitly assumed in (1). Q_{SC} is defined as

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

the *space charge density per unit area* and $C_{ox} = \epsilon_{ox}/t_{ox}$ is the *oxide capacitance per unit area*. From hereon, the specification *per unit area* will be omitted when referred to space charge densities and oxide capacitance values.

III. Formulation of working equation

A working equation has been obtained from the Poisson's equation in conjugation with Boltzmann carrier statistics. In an n-type doped channel acceptor doping concentration, $N_A = 0$. For the sake of simplicity it has also been assumed that the hole concentration, $p = 0$. In terms of Boltzmann carrier statistics, electron concentration, n is written as,

$$n = n_0 \exp\left[\frac{\phi - V}{V_T}\right] \tag{2}$$

where, V is known as *quasi Fermi potential*, $V_T = k_B T/q$ is the *thermal voltage*, k_B is Boltzmann's constant, T is the absolute temperature and q is electronic charge. n_0 is the electron concentration at flatband condition where, $\phi = V$. For a heavily doped semiconductor $n_0 \simeq N_D$. With these, normalized Poisson's equation in rectangular coordinates can be written as

$$\frac{d^2u}{dp^2} = \exp(u - v) - 1 \tag{3}$$

where $u = \phi/V_T$ and $v = V/V_T$ are termed as *normalized* electrostatic potential and *normalized* quasi Fermi potential and $p = x/\lambda_D$ with

$$\frac{1}{\lambda_D} = \sqrt{\frac{qN_D}{\epsilon_{si} v_T}} \tag{4}$$

Here, λ_D is known as *Debye length*. (3) is to be solved with appropriate boundary conditions which are given by

$$\left.\frac{du}{dp}\right|_{p=0} = 0 \tag{5}$$

$$\text{and} \qquad u(\pm P) = u_s \tag{6}$$

where $P = \dfrac{t_{si}}{2\lambda_D}$ is the normalized half-channel width. The first boundary condition is due to the symmetry of the device and the second one is due to the presence of Dirichlet boundary at the gate terminals.

IV. Approximate solutions of the working equation

Although (3) has a very simple appearance but it does provide any closed-form solution. This is due to the presence of the -1 term on the right hand side. Numerical techniques can be employed to obtain the exact solution but an approximate closed-form solution may prove very helpful in modeling JL DG MOSFET behavior. That is why in this section, appropriate assumptions have been made to obtain approximate solutions of the working equation stated in (3).

A. Electric field intensity and surface charge

Normalized electric field intensity, $E_n(p)$ is defined as

$$E = -\frac{d\phi}{dp}\frac{dp}{dx} = -\frac{v_T}{\lambda_D}\frac{du}{dp} = -\frac{v_T}{\lambda_D} E_n \tag{7}$$

An expression for $E_n(p)$ can be obtained by integrating (3) with respect to p as

$$E_n(p) = \left(2\left\{e^{[u-v]} - e^{[u(0)-v]} - u + u(0)\right\}\right)^{1/2} \tag{8}$$

Normalized surface electric field intensity, E_{ns} can be found by setting $p = P$ into (8),

$$E_{ns} = \left(2\left\{e^{[u_s-v]} - e^{[u(0)-v]} - u_s + u(0)\right\}\right)^{1/2} \tag{9}$$

$u(0)$ in (8) and (9) is the normalized potential at the center of the channel. Further integration of (8) does not yield any closed-form expression. Therefore approximations are mandatory to obtain any analytical solution.

Surface charge density, Q_{SC} can be expressed as

$$Q_{SC} = -2\epsilon_{Si}\left.\frac{d\phi}{dx}\right|_{x=\frac{t_{Si}}{2}} = -2\left(\frac{\epsilon_{Si} v_T}{\lambda_D}\right) E_{ns} \tag{10}$$

B. Surface potential at accumulation

The proposed model deals only with the solution in depletion region of operation. Therefore it incorporates the solution provided in [13] for accumulation region of operation without further modification. Here the expressions for electric field and surface potential under accumulation is just restated in terms of normalized variables

$$E_{ns} \approx \frac{V_T}{\lambda_D}\sqrt{2\left[\exp(u_s - v) - 1\right]} \tag{11}$$

and

$$\beta v_T\left[\exp(u_s - v) - 1\right] = \left(V_G - V_{FB} - u_s v_T\right)^2 \tag{12}$$

where β is defined as

$$\beta = \frac{2q\epsilon_{Si} N_D}{C_{ox}^2} \tag{13}$$

C. Surface potential at depletion

When the device is biased in depletion or subthreshold conditions i.e., $u_s < v$, the exponential term present on the right hand side of (3) is *linearized* by setting

$$u = u_c + \delta u \tag{14}$$

where u_c is regarded as the potential at the symmetry axis of the device. Assumption of this type is regarded as *gradual depletion approximation* or *first-order approximation* to the exact carrier concentration and has already been applied for JL nanowire FETs [14]. With this approximation (3) is modified as

$$\frac{d^2\delta u}{dp^2} - \exp(u_c - v)\,\delta u = \exp(u_c - v) - 1 \tag{15}$$

which is valid for $0 \le p \le p_0$. The point p_0 is defined as the point where

$$\delta u(p_0) = -1 \tag{16}$$

Let $\alpha = \exp(u_c - v)$ and using this definition (15) can be modified as

$$\frac{d^2\delta u}{dp^2} - \alpha\delta u = \alpha - 1 \tag{17}$$

(17) is to be solved with the following initial conditions

$$\delta u(0) = 0 \tag{18}$$

$$\text{and} \qquad \frac{d\delta u}{dp}(0) = 0 \tag{19}$$

The solution of (17) can be found as

$$\delta u(p) = \frac{\alpha - 1}{\alpha}\left[\cosh\left(\sqrt{\alpha}\,p\right) - 1\right]; \qquad 0 \le p \le p_0 \tag{20}$$

From (16) and (20), the value of p_0 can be determined, i.e.,

$$p_0 = \frac{1}{\sqrt{\alpha}} \cosh^{-1}\left(\frac{1}{1-\alpha}\right) \tag{21}$$

For the region $p_0 < p < P$, the depletion approximation is used instead of (17) i.e.,

$$\frac{d^2\delta u}{dp^2} = -1 \tag{22}$$

The first initial condition required to solve (22) has already been mentioned in (16). The second initial condition can be written as

$$\left.\frac{d\delta u}{dp}(p_0)\right|_{0<p<p_0} = \left.\frac{d\delta u}{dp}(p_0)\right|_{p>p_0} \tag{23}$$

The final solution of (22) can then be obtained as

$$\begin{aligned}\delta u(p) &= -1 - \frac{p^2 - p_0^2}{2} \\ &+ \left[p_0 + \frac{\alpha-1}{\sqrt{\alpha}}\sinh\left(\sqrt{\alpha}\,p_0\right)\right](p - p_0)\end{aligned} \tag{24}$$

which is valid for the region $p_0 < p < P$. Since surface potential, $u_s = u(P)$, expressions for u_s under depletion mode of operation can be obtained by setting $p = P$ into (20) and (24).

$$u_s = u_c + \frac{\alpha-1}{\alpha}\left[\cosh\left(\sqrt{\alpha}P\right) - 1\right] \tag{25}$$

for $p_0 > P$, and

$$\begin{aligned}u_s = u_c &- 1 - \frac{P^2 - p_0^2}{2} \\ &+ \left[p_0 + \frac{\alpha-1}{\sqrt{\alpha}}\sinh\left(\sqrt{\alpha}p_0\right)\right](P - p_0)\end{aligned} \tag{26}$$

for $p_0 < P$.

Finally Q_{SC} can be determined as

$$Q_{SC} = -\frac{\epsilon_{si} v_T}{\lambda_D}\left[\frac{\alpha-1}{\sqrt{\alpha}}\sinh\left(\sqrt{\alpha}P\right)\right] \tag{27}$$

for $p_0 > P$, or

$$Q_{SC} = \frac{\epsilon_{si} v_T}{\lambda_D}\left[P - p_0 + \frac{1-\alpha}{\sqrt{\alpha}}\sinh\left(\sqrt{\alpha}p_0\right)\right] \tag{28}$$

for $p_0 < P$.

V. Results and Discussion

The symmetric device structure shown in Fig. 1 has been simulated with the proposed model and the results have been verified against the results obtained from numerial simualtions using MATLAB and SILVACO ATLAS TCAD tool [15]. The results have also been compared with the abrupt model proposed in [13]. For simulation purpose, the device has been assumed to have $L = 1$ μm, $W = 1$ μm, $t_{ox} = 8$ nm, $t_{Si} = 10$ nm, $N_D = 10^{19}$ cm^{-3} and $\phi_{MS} = 1.12$ eV. Silicon ($\epsilon_{Si} = 11.7$) and SiO_2 ($\epsilon_{ox} = 3.9$) have been used as channel and oxide material respectively.

Fig. 2 demonstrates the validity of the proposed model by comparing the results with those from ATLAS and MATLAB simulation. This figure shows that the proposed model provides better performance for the whole range of gate voltage from deep depletion to strong accumulation. Surface potential obtained from abrupt depletion approximation based model fails to match with numerical solutions in partial depletion region where it simply deviates from the actual solution. In abrupt-depletion approximation based model *discontinuity in transition* from deep depletion to partial depletion around the threshold voltage is readily observable from the figure.

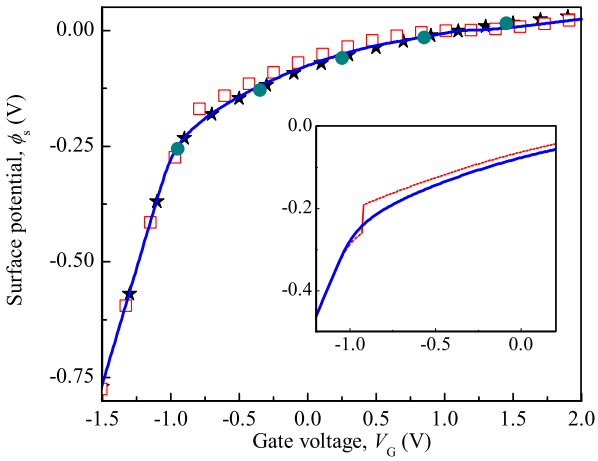


Fig. 2. Surface potential of the JL DG MOSFET plotted against the gate voltage. (Stars) SILVACO ATLAS simulation. (Squares) Abrupt model. (Circles) MATLAB simulation. (Line) First-order model. (Inset) Expanded view of transition from full depletion to partial depletion - solid line corresponds to gradual depletion model and dashed line represents abrupt depletion model.

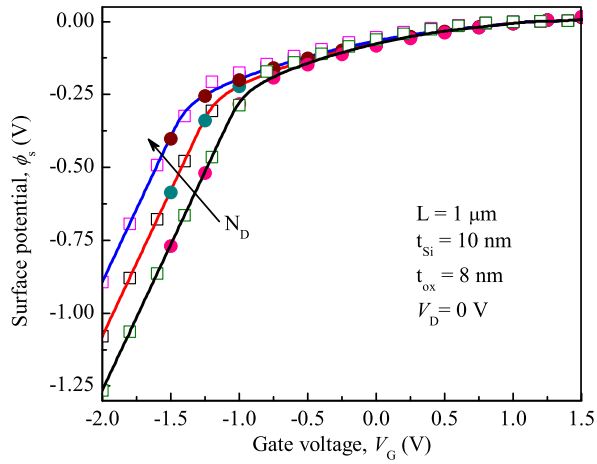


Fig. 3. Surface potential of the JL DG MOSFET plotted against the gate voltage for N_D ranging from 10^{19} cm^{-3} to 1.2×10^{19} cm^{-3} in steps of 1×10^{18} cm^{-3}. (Solid symbols) MATLAB simulation. (Hollow symbols) Abrupt model. (Lines) Gradual depletion model.

The effects of device and process parameters on surface potential characteristics have been illustrated in Fig. 3 to 6. It is readily evident that the variation in gate oxide thickness, channel thickness, doping concentration and quasi-Fermi potential affects the surface potential-gate voltage characteristics significantly.

With the increase in doping concentration surface potential increases in the subthreshoold region and it saturates to a constant value in the accumulation region as evident from Fig. 3. This is because denisty of free electrons increases with the increase in doping concentration.

From Fig. 4 it can be observed that increasing gate oxide thickness shifts the subthreshold region of the characteristics towards left which means an increase in surface potential. The surface potential in accumulation also slowly reduces with increasing oxide thickness. As gate oxide thickness continues to increase, more gate voltage drops across the gate oxide and less potential drops inside the channel.

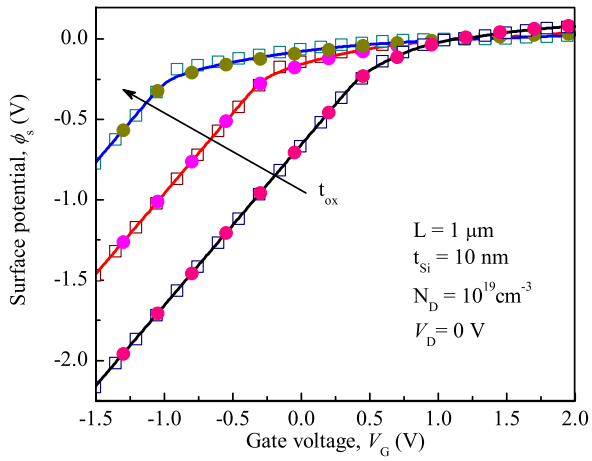


Fig. 4. Surface potential of the JL DG MOSFET plotted against the gate voltage for t_{ox} ranging from 2 nm to 8 nm in steps of 3 nm. (Solid symbols) MATLAB simulation. (Hollow symbols) Abrupt model. (Lines) Gradual depletion model.

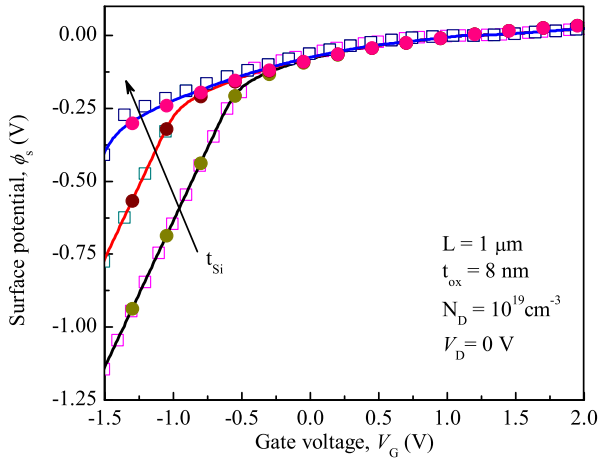


Fig. 5. Surface potential of the JL DG MOSFET plotted against the gate voltage for t_{Si} ranging from 8 nm to 12 nm in steps of 2 nm. (Solid symbols) MATLAB simulation. (Hollow symbols) Abrupt model. (Lines) Gradual depletion model.

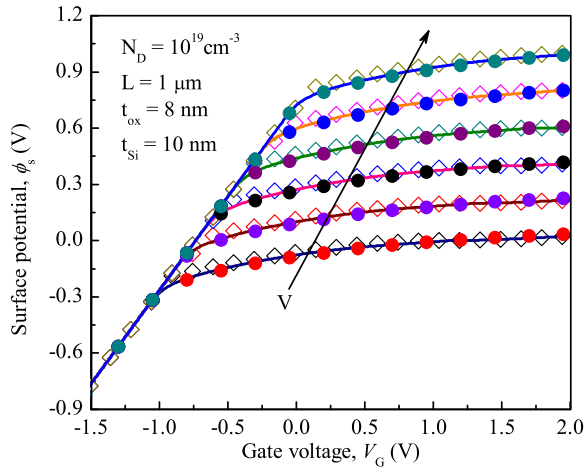


Fig. 6. Surface potential of the JL DG MOSFET plotted against the gate voltage for V ranging from 0 V to 1 V in steps of 0.2 V. (Solid symbols) MATLAB simulation. (Hollow symbols) Abrupt model. (Lines) Gradual depletion model.

Fig. 5 shows that increasing channel thickness also increases surface potential. which is due to the fact that a thicker channel is populated with more free electrons than a thinner channel.

Increasing quasi-Fermi level does not affect the subthreshold region but increases the surface potential value in accumulation region which has been demonstrated in Fig. 6. In all the figures, the discrepancies between the abrupt-depletion based surface potential model and gradual-depletion based model can easily be identified.

VI. Conclusion

An improved analytical model has been proposed to determine the surface potential for a symmetric long-channel JL DG MOSFET. The model has been developed for long-channel devices and hence does not considers short-channel and quantum effects. Possion's equation for the device cannot provide any closed-form solution hence a regional approach has been used for the analysis. The proposed model is valid for all regions of device operation and provides continuous transition from one region to another. This continuity feature may prove very useful for simulation programs and make this model a better candidate in studying the device operation and other charateristics of JL DG MOSFETs.

References

[1] C.-W. Lee, A. Afzalian, N. D. Akhavan, R. Yan, I. Ferain, and J.-P. Colinge, "Junctionless multigate field-effect transistor," *Applied Physics Letters*, vol. 94, no. 5, pp. –, 2009.

[2] J.-P. Colinge, C.-W. Lee, A. Afzalian, N. D. Akhavan, R. Yan, I. Ferain, P. Razavi, B. O'Neill, A. Blake, M. White, A.-M. Kelleher, B. McCarthy, and R. Murphy, "Nanowire transistors without junctions," *Nat Nano*, vol. 5, no. 3, pp. 225–229, March 2010.

[3] I. Ferain, C. A. Colinge, and J.-P. Colinge, "Multigate transistors as the future of classical metal-oxide-semiconductor field-effect transistors," *Nature*, vol. 479, no. 7373, pp. 310–316, 11 2011.

[4] H.-C. Lin, C.-I. Lin, and T.-Y. Huang, "Characteristics of n-Type Junctionless Poly-Si Thin-Film Transistors With an Ultrathin Channel," *Electron Device Letters, IEEE*, vol. 33, no. 1, pp. 53–55, Jan 2012.

[5] C.-W. Lee, A. N. Nazarov, I. Ferain, N. D. Akhavan, R. Yan, P. Razavi, R. Yu, R. T. Doria, and J.-P. Colinge, "Low subthreshold slope in junctionless multigate transistors," *Applied Physics Letters*, vol. 96, no. 10, pp. –, 2010.

[6] C.-J. Su, T.-I. Tsai, Y.-L. Liou, Z.-M. Lin, H.-C. Lin, and T.-S. Chao, "Gate-All-Around Junctionless Transistors With Heavily Doped Polysilicon Nanowire Channels," *Electron Device Letters, IEEE*, vol. 32, no. 4, pp. 521–523, April 2011.

[7] J. Colinge, A. Kranti, R. Yan, C. Lee, I. Ferain, R. Yu, N. D. Akhavan, and P. Razavi, "Junctionless nanowire transistor (jnt): Properties and design guidelines," *Solid-State Electronics*, vol. 6566, no. 0, pp. 33 – 37, 2011, selected Papers from the {ESSDERC} 2010 Conference.

[8] J.-T. Park, J. Y. Kim, C.-W. Lee, and J.-P. Colinge, "Low-temperature conductance oscillations in junctionless nanowire transistors," *Applied Physics Letters*, vol. 97, no. 17, pp. –, 2010.

[9] C.-W. Lee, A. Borne, I. Ferain, A. Afzalian, R. Yan, N. Akhavan, P. Razavi, and J. Colinge, "High-temperature performance of silicon junctionless mosfets," *Electron Devices, IEEE Transactions on*, vol. 57, no. 3, pp. 620–625, March 2010.

[10] J. Duarte, S.-J. Choi, D.-I. Moon, and Y.-K. Choi, "Simple Analytical Bulk Current Model for Long-Channel Double-Gate Junctionless Transistors," *Electron Device Letters, IEEE*, vol. 32, no. 6, pp. 704–706, June 2011.

[11] J. M. Sallese, N. Chevillon, C. Lallement, B. Iniguez, and F. Pregaldiny, "Charge-Based Modeling of Junctionless Double-Gate Field-Effect Transistors," *Electron Devices, IEEE Transactions on*, vol. 58, no. 8, pp. 2628–2637, Aug 2011.

[12] J. Duarte, S.-J. Choi, and Y.-K. Choi, "A Full-Range Drain Current Model for Double-Gate Junctionless Transistors," *Electron Devices, IEEE Transactions on*, vol. 58, no. 12, pp. 4219–4225, Dec 2011.

[13] Z. Chen, Y. Xiao, M. Tang, Y. Xiong, J. Huang, J. Li, X. Gu, and Y. Zhou, "Surface-Potential-Based Drain Current Model for Long-Channel Junctionless Double-Gate MOSFETs," *Electron Devices, IEEE Transactions on*, vol. 59, no. 12, pp. 3292–3298, Dec.

[14] E. Gnani, A. Gnudi, S. Reggiani, and G. Baccarani, "Theory of the Junctionless Nanowire FET," *Electron Devices, IEEE Transactions on*, vol. 58, no. 9, pp. 2903–2910, 2011.

[15] *Atlas Device Simulator User's Manual*, SILVACO Int., Santa Clara, CA (USA), 2005.

Design of an Improved Algorithm for Blind Digital Image Watermarking Using Both Gray-scale and Binary Watermark in DFWHT Domain

Sudip Ghosh,[1,*] Arijit Biswas,[1] Santi P Maity,[2] and Hafizur Rahaman[1,2]

[1]School of VLSI Technology, Indian Institute of Engineering Science and Technology(IIEST)
Howrah, Shibpur, West Bengal, India

[2] Department of Information Technology, Indian Institute of Engineering Science and Technology(IIEST)
Howrah, Shibpur, West Bengal, India

[*]sudipghosh2005@gmail.com

Abstract—**This paper focuses on the design of an improved DFWHT (Discrete Fast Walsh Hadamard Transform) domain robust, blind digital image watermarking algorithm. Both encoding and extraction algorithm have been verified using MATLAB R2013a. Here both grayscale and binary watermarks are possible for embedding and extraction using the same designed algorithm and only grayscale cover image of maximum size(256 x 256) is used. The speciality of this algorithm attracts to the fact that there is no requirement for any other information other than the received watermarked image to extract the watermark at decoding, which makes it blind watermarking. The time complexity of the proposed algorithm is of the order of $O(n^2)$ and it is simple to implement in hardware.**

Index Terms **- Blind Digital Image Watermarking; DFWHT; Robust; time complexity; Binary and Grayscale watermark.**

I. INTRODUCTION

There are various ways by which digital watermarking [1] can be accomplished, such as image, text, video and audio for example. Several data embedding schemes are available for spatial [2] and transform domain (e.g. Discrete Fourier Transform[3], Discrete Cosine Transform[4], Discrete Wavelet Transform[5], Discrete Walsh Transform[6],[7]).Digital Watermark can be classified as visible and invisible watermark. In case of visible watermark the watermark is visible to normal eye such as company logo etc., but for invisible watermark the locations where the watermark is embedded is secret and known only to the authorized person. One of the important characteristics for watermarking is Robustness ,that is how good the watermark is to resist the common signal processing attacks such as Scaling, sharpening and geometrical transforms. The watermark which are not being able to resist the attacks are Fragile in nature, and watermark which can to some extent sustain the attacks are called Semi-fragile. Thus depending on the application for which the watermark is used, watermarking can be of several types. Modern day, to ensure Quality Of Service of a multimedia channel, digital watermarking is also used. For this purpose visible and fragile watermark can be used. For high security data transmission purpose invisible and robust watermark must be used.

For embedding the data, several watermarking techniques are used, but transform domain watermarking are most widely used due to high robustness of the watermark. Among them the most popular for watermarking JPEG and JPEG 2000 images are DCT and Wavelet transform. However, it is been found out that DCT domain transform are much more robust to JPEG but not so robust to JPEG 2000 images. Reverse are true for Wavelet domain transforms. So proper watermarking scheme must be used which shows similar robustness for both JPEG and JPEG 2000 images. It is been reported in [6] that Hadamard and Hartley transforms, due to small values of the standard deviation for the processing noise provide higher resiliency at low quality compression compared to DCT and Wavelet transforms.

In this paper, we have proposed the design of an improved invisible, blind and robust (with respect to sharpening and scaling) image watermarking algorithm in DFWHT [8] domain. The rest of the paper is organized as follows. Section II gives the basics of DFWHT. Section III describes the related works and motivation. Section IV states the proposed image watermark encoding and extraction algorithm in DFWHT domain. Section V focus on the detail description of the watermark encoding and decoding units. Section VI reports the analysis and experimental results. Section VII is the conclusion followed by the references.

II. DISCRETE FAST WALSH HADAMARD TRANSFORM PRELIMINARIES

Let [A] be the original image block and [H] be the Walsh transformation matrix then the Walsh transformed block can be viewed as

$$[Y]= ([H][A][H])/N. \quad (1)$$

Here N is the order of the Hadamard matrix; H is a (N x N) Hadamard matrix of which all element are either +1 or -1. The elements of the Walsh Hadamard Matrix are simple and real valued.

The inverse Walsh Hadamard transform is given by

$$[A]= ([H][Y][H])/N. \quad (2)$$

Here we use the (8x8) Hadamard matrix given by

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{bmatrix}$$

Fig. 1 Order 8 Hadamard matrix.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

Here the first row and first column are DC values, and the rows in which the sign changes are AC values.

III. RELATED WORKS AND MOTIVATION OF PROPOSED WORK

Ho et.al.[6] proposed an algorithm in which the watermark image is transformed by Walsh Hadamard transformation. Then the cover image is partitioned into blocks of size (8 x 8) and the 16 watermark image coefficients values are inserted into cover image block by substituting the lower rows of the blocks which are the high frequency blocks. The scheme stated in [6] has left out the DC coefficient value (that is the first row and column values in the watermark image transformed block) while embedding other AC or high frequency values in a pseudo randomly selected blocks. Thus the disadvantage in [6] is that we have to separately send the DC values. It has been reported in [9] that processing noise increases drastically for high frequency DCT and Wavelet co-efficient. On the other hand the high frequency Hadamard coefficients are much more immune to processing noise. Thus it is quite possible to embed data to the higher frequency bands of the Walsh Hadamard Transformed blocks than to DCT or Wavelet high frequency bands.

Watermarking schemes using Walsh Hadamard Transform is also used in several works. In [10] binary watermark is embedding using spread spectrum techniques in Walsh Hadamard domain. In [11] the grayscale cover image with binary watermark is used which is also embedded using the spread spectrum watermarking and its dual mode architecture has been implemented.

In our work we have proposed an improved algorithm over [6] and here in this paper we embed all the Walsh Hadamard transformed coefficients of the watermark image in the cover image, thus if we do not have any other value to be transmitted, the watermarked image has to be sent only making it blind.

IV. PROPOSED IMAGE WATERMARKING ALGORITHM

BINARY AND GRAYSCALE WATERMARK EMBEDDING

Here a (256 x 256) size cover image is used and a (32 x 32) size grayscale watermark image of maximum pixel value of 15 is used to encode.

A. Partitioning the cover image and Walsh Hadamard transformation

The watermark image of size (32 x 32) is partitioned in (8 x 8) blocks. Then Walsh Transform is applied to all the blocks. Let the Walsh transformed blocks of the watermark image be denoted as 'w'. The transformed blocks are then converted to a linear array of size (1 x 1024).

B. Row and Column selection and corresponding block selection

The random selection of the block for inserting the watermark is inserted in by the user. To insert 8 coefficient values of the watermark image in the lowest row of the cover image selected block the row vector must have 32 random non repeating values as the watermark has 32 rows and 4 random non repeating values for column selection as in each row 8 values are inserted so 4 column select value is required.

The selected row value let be 'm' and column value be 'n' are used to select the first element of the selected block in which the watermark will be embedded. This element can be selected from the original image denoted as 'G', as

$$G(((m-1)*8)+1),((n-1)*8)+1)). \quad (3)$$

The corresponding block thus can be selected by incrementing both row and column values from the initial pixel location, as incrementing the column value as ((m-1)*8+1) to ((m-1)*8+8) and the row value as ((n-1)*8+1) to ((n-1)*8+8). Thus the complete block is selected. This block is Walsh Hadamard transformed.

C. Inserting watermark data

The one dimensional block of Walsh Hadamard transformed data is embedded in the last row of the selected transformed cover image block denoted as 'X" as

$$X(8,1 \text{ to } 8) = w(1, k \text{ to } (k+7)). \quad (4)$$

Here k is used to select the corresponding value to be embedded in the block, where k is initially to 1.

D. Watermarked image generation

The watermarked block is then inserted back into the original image after applying inverse Walsh Hadamard transform on 'X'. We then round up the values and the block is now denoted as 'X1'.This block is then inserted into the region of the image where it was initially taken from. Thus we get the watermarked image denoted as 'G1'.

E. Error values transmitted

For correct decoding of the transmitted watermark the fractional values which are generated due to Walsh transformation are stored in a matrix as

$$E=X1-X. \quad (5)$$

Again let the image right before the transmission be denoted G2.Now it may be possible that due to some reason some values which were actually embedded is altered then this error is also taken into account as

$$E1=G1-G2. \quad (6)$$

Thus the total error value transmitted is

$$Etotal=E1-E. \quad (7)$$

Thus this is the complete encoding method where along with the transformed the error Etotal is transmitted.

BINARY AND GRAYSCALE WATERMARK EXTRACTION

We receive the watermarked image G2 along with the error value Etotal.

A. Generating the decodable block

The received image G2 and error Etotal is used to generate the block from which we can decode the watermark as

$$Y=Etotal + G2. \quad (8)$$

B. Selecting the proper block

From the row and column value which was used at the time of encoding can be used here also to find out the corresponding block.

C. Inverse transform and finding the embedded data

The last row of the corresponding block is stored in a one dimensional array of size (1 x 1024), this is the decoded transformed value of the watermark.

D. Watermark decoding

The linear array is converted to (32 x 32) block which is then inverse Walsh transformed as (8 x 8) block and stored in a (32 x 32) matrix, which is the watermark image. Thus we get the decoded watermark image.

V. DESCRIPTION OF THE WATERMARK ENCODING AND DECODING UNIT

The block diagram of the watermarking method is divided in encoding and decoding blocks. The block diagram of the encoding block is shown in Fig. 2.

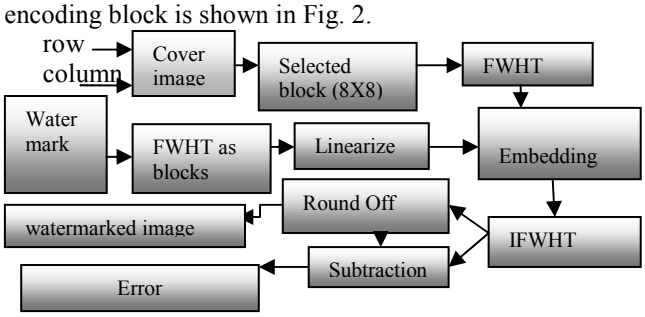


Fig. 2 Block diagram of the watermark encoding block.

The cover image considered here is of size (256 x 256) and the watermark is of size (32 x 32). Here the row and column select values are used to select a particular block from the original image. The row and column select value are inserted by the user at the time of embedding as linear arrays. Then the selected block is Fast Walsh Hadamard Transformed (FWHT). The watermark is also broken in blocks of size (8 x 8) and passed through FWHT block. Equation (4) is used for the Embedding block for performing the watermarking process. After embedding, the block is inverse transformed by the block IFWHT. The inverse transformed values are round off. These blocks are then embedded back in the original cover image to the place which was selected by the row and column select values and thus we get the watermarked image. Thus the Watermarked image and the error value is transmitted along with the image. This is the complete operation of the encoder block. Here the error while transmitting the image is not considered as the process used by the transmitter to transmit the image is not exactly known.

To decode the watermark the decoder block is used .The block diagram for the decoder shown in Fig. 3.

The decoder block gets as input the received watermarked image and the error values. These values are added to get the actual image block generated at the transmitter before rounding off the value for transmission. The row and column select values used at the transmitter is also used here for the selection of the proper block. The selected block is passed through the FWHT block. As the watermark was initially embedded in the last row of the block, the last row is selected and put in a linear array.

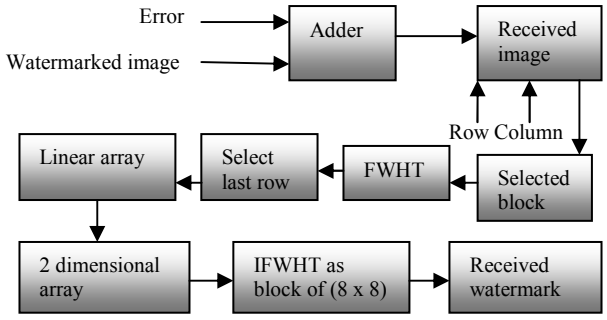


Fig. 3 Block diagram of the watermark decoding block.

The linear array is converted to two dimensional array. The two dimensional array is then inverse transformed block wise then combined to get the original watermark. Thus by using this architecture we can get the original watermark.

VI. ANALYSIS AND EXPERIMENTAL RESULTS

After performing the above stated algorithm in Matlab the following results have been obtained.

For Matlab simulation we used a (256 x 256) cover image and (32 x 32) and (64 x 64) watermark. The operation has been performed on both grayscale and binary watermark and the result obtained is illustrated in Table I. Here the index on which we have stated the quality of watermarking is the Mean Squared Error (MSE) and Peak Signal to Noise Ratio (PSNR).

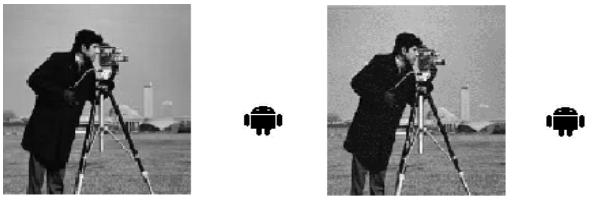
TABLE I
PERFORMANCE METRICS OF THE PROPOSED WATERMARKING ALGORITHM

Type of image	Water mark image size	Encoded image index		
		Normalization value	Peak Signal to Noise Ratio(dB)	Mean Squared Error(MSE)
Gray scale image	32x32	8	46.869	1.3372
		17	52.03	0.472
	64x64	8	42.58	3.588
		17	45.99	1.1134
Binary Image	32x32	1	69.81	0.0065

The normalization factor used here is the amount by which the pixel value is divided by, that is ,if the intensity of a pixel be 255 then after normalization by 17 we get the maximum pixel value as 15. Here are few image examples,

(a) (b) (c) (d)

Fig. 4 Watermark embedding and decoding for (64 x 64) grayscale watemark: (a) is the cover image,(b) is the watermark,(c) is watermarked image,(d) decoded watermark



(a) (b) (c) (d)

Fig. 5 Watermark embedding and decoding for (32 x 32) binary watemark: (a) is the cover image,(b) is the watermark,(c) is watermarked image,(d) decoded watermark

Now we have applied some common watermark attacks such as image sharpening and image scaling to see how the decoded watermark looks like.

TABLE III
DIFFERENT WATRMARK ATTACKS

Original watermark	Difference watermarks after operations	
	Operation	**Received watermarks**
Grayscale	Scaling by 0.75	
	Sharpening	
	Cropping	
Binary	Sharpening + Cropping	

The algorithmic time complexity of different blocks of both encoder and decoder are shown in the Table III as

TABLE IIIII
ALGORITHMIC TIME COMPLEXITY OF DIFFERENT BLOCKS

Unit	Blocks	Time Complexity
Encoder	Block Selection	$O(n^2)$
	Linearization	$O(n^2)$
	Embedder	$O(n^2)$
Decoder	Selection of last row	$O(n)$
	Linear array to 2 dimensional array	$O(n^2)$

The encoder has several blocks, we here select the block with size (n x n). The block selection block uses both row and column values, since both row and column values are to be selected thus we use two loops thus the time complexity is $O(n^2)$.

For linearization again we have to use array of the size 'n^2', Thus here the time complexity is also $O(n^2)$.

The embedder unit has first n.(n-1) elements at the output linearly, but the last row of the matrix is transmitted by logically substituting the lowest row of the original matrix by the transformed watermark image values. Thus here also we perform the operation both row and column wise of the block , hence it has time complexity $O(n^2)$.

In case of the decoder the row selection block is same as the encoder. For selection of the last row from the selected block, we fix the row value to 7 and then the column values are incremented, thus the time complexity here is O(n).

The last row values stored in the linear array is transformed in the 2 dimensional matrix to get the original watermark, thus here also two loops are needed, thus the time complexity is $O(n^2)$.

VII. CONCLUSION

From the above discussion we can finally draw some concluding remarks as: If the error value is not there , that is if the normalization used in the Walsh-Hadamard transform is not taken in consideration and if there is no error while transmission, then the proposed algorithm can be completely blind, which is an improvement over[6] that is we do not need any new values to be transmitted along with the image. The maximum value of the pixel which can be embedded without affecting the quality of the watermarked image is 32, that is each pixel can be represented by 5 bits in binary. The ratio of the of the cover image to the watermark must not be more than 8:1,that is if the cover image be of size (256 x 256) the watermark must be of size maximum (64 x 64) . For higher size of watermark, higher size of cover image is required, else the quality of the watermarked image will degrade. For embedding and decoding unit the time complexity is compatible to be implemented in hardware. In our proposed algorithm, both binary as well as grayscale watermark can be embedded and extracted which is an added advantage.

REFERENCES

[1] S. Katzenbeisser and F. Petitcolas , *Information Hiding Technique for Stenography and Digital Watermarking*, Artech House , London, 2000.

[2] N. Nikolaidis and I. Pitas,"Robust Image Watermarking in the Spatial Domain," Signal Processing, vol. 66, no. 3,pp. 385-403, 1998

[3] Xiu-mei Wen,Wei Zhao, Fan-xing Meng, “Research of digital image watermarking algorithm resisting geometrical attacks in Fourier domain,” International Conference on Computational Intelligence and Security, pp. 265-268, 2009.

[4] C. Pradhan, V. Saxena and A. K. Bisoi, “Non-Blind digital watermarking technique using DCT and cross chaos map,” International Conference on Communications, Devices and Intelligent Systems, pp. 274-277, 2012

[5] Yingli Wang, Xue Bai, Shuang Yan, “Digital image watermarking based on texture block and edge detection in discrete wavelet domain,” International Conference on Sensor Network Security Technology and privacy Communication System, pp. 170-174, 2013.

[6] A. T. S. Ho, J. Shen, S. H. Tan and A. C. Kot, "Digital image-in-image watermarking for copyright protection of satellite images using the Fast Hadamard Transform," IGARRS, pp. 3311-3313, 2002.

[7] Amani Bin Sewaif, Mohammed Al-Mualla and Hussain Al-Ahmad, "Walsh-coded signatures for robust digital image watermarking," IEEE conf. 2004,pp. 431-434.

[8] K. Beauchamp, *Applications of Walsh and Realted Functions with an Introduction to Sequence theory*, Academic Press, London,1984.

[9] M. Ramkumar, A. N. Akansu, "Capacity estimates for data hiding in compressed images," IEEE Transactions on Image Processing, pp. 1252-1263, 2001.

[10] S. P. Maity, M. K. Kundu and S. Maity, "Dual purpose FWT domain spread spectrum image watermarking in real-time,"Elsevier Journal of Computers and Electrical Engineering.Volume 35, Issue 2, March 2009, pp. 415–433

[11] S. Ghosh, S. Talapatra, J. Sharma, N. Chatterjee, Hafizur Rahaman and Santi P Maity, “Dual Mode VLSI Architecture for Spread Spectrum Image Watermarking using Binary Watermark” (ICCCS-2012) from 6-8 October 2012 at NIT Rourkela, India. pp: 784-791

Reduced Feature Sets for Vowel Recognition

Shantanu Sharma[1], Pradip K. Das[1,*]
[1]Department of Computer Science & Engineering
Indian Institute of Technology Guwahati
Assam, INDIA 781039
[*]pkdas@iitg.ernet.in

***Abstract* - The accuracy of speech recognition systems, to a large extent, depend on the feature sets used for representing the speech data. It has been a continuous process to develop feature sets to perform more accurate speech recognition through ASR (Automatic Speech Recognition) systems. Many feature sets and their different combinations have been tried to achieve better accuracy but the feature set providing completely accurate representation is yet to be formulated.**

This paper investigates the generation of reduced sets of MFCCs for vowel recognition. The study is more focused on the generation and behavior of MFCCs for different vowel sounds. The goal is to identify the features that enhance their discriminating capabilities and improve the performance of ASR systems for particular sounds. The results of the analysis show that the proposed reduced features perform well and can be further used to improve the accuracy of ASR systems, specifically the ones in resource constrained devices.

***Index Terms*—Reduced feature sets, MFCCs, Vowel Recognition**

I. Introduction

ONE of the first decisions in any pattern recognition system is the choice of what features to use: How exactly to represent the basic signal that is to be classified, in order to make the classification algorithm's job simple. In the field of speech recognition, the selection of features used for the proper representation of speech data must be capable of identifying and discriminating the underlying patterns in speech with utmost accuracy and reliability.

One of the most powerful speech analysis feature set is LPC (Linear Predictive Coefficients). LPCs are capable of encoding good quality speech at a low bit rate and provide extremely accurate estimates of speech parameters. The most popular feature representation currently used is the Mel-frequency Cepstral Coefficients or MFCCs. Another popular speech feature representation is known as RASTA-PLP, an acronym for Relative Spectral Transform - Perceptual Linear Prediction. PLP was originally proposed by Hynek Hermansky as a way of warping spectra to minimize the differences between speakers while preserving important speech information [4]. RASTA is a separate technique that applies a band-pass filter to the energy in each frequency sub-band in order to smooth over short-term noise variations and to remove any constant offset resulting from static spectral coloration in the speech channel e.g. from a telephone line [5].

In the decade 1980-90 development of speech recognition toolkits was started and from then on several such toolkits have been developed. Most of these toolkits are based on the HMM (Hidden Markov Model) framework. Two popular toolkits are Sphinx and HTK. Both these toolkits are open source and can be configured as per requirements.

In the last decade it has been observed that most of the feature sets like LPCs, PLPCs and MFCCs have reached saturation in providing more accurate results for speech recognition. It inherently stems the requirement of formulation of a better feature set for achieving higher recognition accuracy. In some recent attempts researchers have tried several combinations of different feature sets and have succeeded in achieving slightly better results. Such combinations of feature sets pose more resource requirements in terms of storage and computing power. Such requirements can only be fulfilled with high-end devices but these large feature sets cannot be handled by resource constrained devices because of limited storage, computing ability and power. It raises the requirement that the traditional feature sets need a critical study of the significance of each feature in delegating speech data.

There are many alternate techniques proposed for vowel recognition that are highly accurate, such as hybridization of cepstral analysis and acoustic analysis of waveforms [11] and fuzzy inference rules [12] but these techniques are quite expensive computationally.

This paper is focused towards the study of the most popular feature sets i.e. MFCCs and their significance in representing speech data so that a better feature set can be formed for providing higher accuracy and/or at least comparable accuracy on resource constrained devices.

II. Feature Set For Speech Analysis

There are several feature sets which are used for speech analysis purposes like LPCs (Linear Predictive Coefficients), MFCCs (Mel Frequency Cepstral Coefficients) and PLPCCs (Perceptual Linear Predictive Coefficients). The human ear resolves frequencies non-linearly in the audio spectrum, thus filter bank analysis is more appropriate than Linear Predictive Coding (LPC) analysis since it is a spectrally oriented method. Mel-scale resolves frequencies in the same manner as the human ear. It attempts to mimic sensing and resolution of frequencies similar to that of the human ear. The Mel-scale is a unit of measurement of perceived frequency (pitch) of a tone. FFT-derived *Mel*-scaled cepstral coefficients [6] are the most common form of cepstral analysis. The MFCCs generation and analysis of speech based on MFCCs is similar to [1, 2].

For generation of MFCCs, the HCOPY tool from HTK (Hidden Markov Tool Kit) has been used. HTK is a toolkit for building HMM-based acoustic models as well as statistical back-off n-gram language models for automatic speech recognition. It consists of a set of command line tools written in C language to construct various components of a speech recognition system.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

The following configuration is used for feature extraction using HCOPY:

SOURCEFORMAT = WAV TARGETKIND = MFCC_0_D TARGETRATE = 100000.0 SAVECOMPRESSED = T SAVEWITHCRC = T WINDOWSIZE = 250000.0 USEHAMMING = T PREEMCOEF = 0.97 NUMCHANS = 26 CEPLIFTER = 22 NUMCEPS = 12 ENORMALISE=T	Qualifier: _0 — zeroth coefficient; _D — first differential Recordings are conducted with 16000 samples per second and stored as wave files. Each analysis frame contains 400 samples with step size of 160 samples.

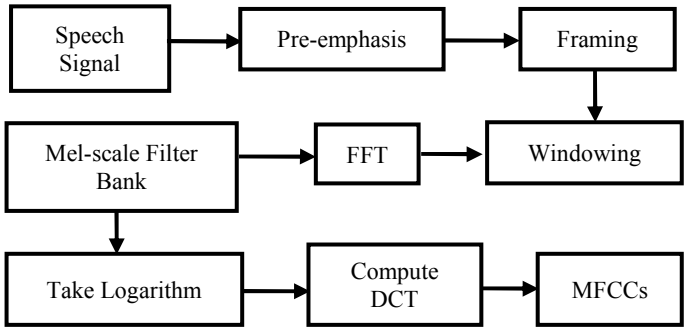


Fig 1.Block Diagram for Generation of MFCCs [3, 10]

The calculation of the MFCC [7] – [9] includes the following steps

1. Pre-emphasis: For pre-emphasis, a high pass finite impulse response (FIR) filter was used, given by the equation below.
$$H(z) = 1 - a * z^{-1} \quad (1)$$
where $0.9 \leq a \leq 1.0$.We take a as 0.97. The signal transformation is done according to the following equation.
$$y(n) = x(n) - a * x(n-1) \quad (2)$$
2. Then frames of 25 ms each were constructed with overlapping of 15 ms and step size of 10 ms.
3. Next the following window was applied on each frame
$$W(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right) \quad (3)$$
where $0 \leq n \leq N-1$ and N is the number of samples in a frame.
4. Let us refer to the time domain signal after applying Hamming window as $s(n)$. After applying framing and Hamming window we obtain $s_i(n)$ where $0 \leq n \leq N-1$ and i the number of frames. Now after calculating the complex DFT for each frame, we get $S_i(k)$ corresponding to the ith time-domain frame.
$$S_i(k) = \sum_{n=0}^{N-1} s_i(n)\, e^{-\frac{i2\pi kn}{N}} \quad (4)$$
where $1 \leq k \leq K$ and K is the length of the DFT. We get the periodogram estimate of the power spectrum $P_i(k)$ for i^{th} speech frame as follows:
$$P_i(k) = \frac{1}{N}|S_i(k)|^2 \quad (5)$$
5. Next Mel-spaced filter-banks were computed. It comprises of 26 triangular filters that are to be applied to the power spectral estimate from the previous step.
Each of these 26 filters is a vector that is applied to power spectral estimate of each speech frame. The values thus obtained, for each filter, are summed up to calculate energies passed through each filter.
6. Then the logarithm of each of the 26 energy values was obtained from the previous step.
7. Then discrete cosine transform (DCT) was applied to these 26 log energies.
8. After the application of DCT we obtain 26 coefficients. From these we choose 13 lower order coefficients which are called Mel Frequency Cepstral Coefficients (MFCCs).

From these 13 MFCCs we drop the 0th coefficient for our analysis and use 12 higher order coefficients.

III. Analysis Of MFCCS

After computing the MFCCs, we conducted a series of initial experiments to analyze the behavior of MFCCs for the five out of eight primary cardinal English vowel sounds (viz. /a/, /e/, /i/, /o/ and /u/) segments.

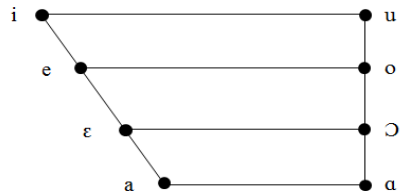


Fig 2. Sets of primary cardinal vowels produced by Daniel Jones

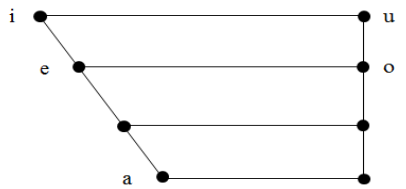


Fig 3. Sets of cardinal vowels taken into account for analysis

The results were further used to determine the importance of each of the MFCCs in representing the vowels in their steady states and to identify the subset of MFCCs that produces more sophisticated delegation of vowels in their steady states as compared to others.

Outline of the Experiment:

- For each speaker in the database, conduct the following:
  1) Collect multiple (10) utterances of each vowel.
  2) Extract the steady state segment of each vowel from every recorded utterance.
  3) Use HCopy software to generate the Mel Frequency Cepstral Coefficients (MFCCs) for the segmented steady state of each utterance.
  4) Collect each MFCC for all the steady state utterances of a vowel in separate files.
  5) Prepare plots of collection of each MFCC for every vowel.
  6) Select the four least varying MFCCs from the set of 12 coefficients.
- From previous results, for each steady state vowel consolidate three most frequent MFCCs.
- Finally, verify the results.

IV. Observations

From the plots of Fig. 2-5 it is observed for vowel /a/ that the MFCCs c_1, c_2, c_4 and c_6 give steady plots while for other

MFCCs (viz. c_3, c_5, c_7, c_8, c_9, c_{10}, c_{11} and c_{12}) the values do not exhibit any regularity.

It is intuitive to examine and verify that the MFCCs that follow this regularity contribute significantly in identifying the vowel while the non-regular MFCCs reduce the recognition of that particular vowel because of their divergent behavior.

This leads us to infer that if we could identify the set of regular MFCCs for each vowel in its steady state, we can develop a superior vowel recognition system with higher recognition rate.

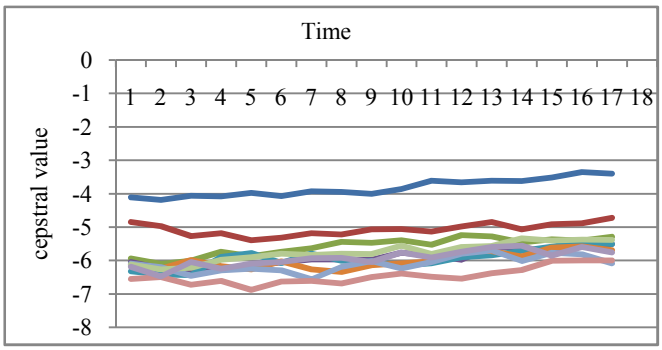


Fig 4. Plot of 2nd MFCC for Vowel /a/

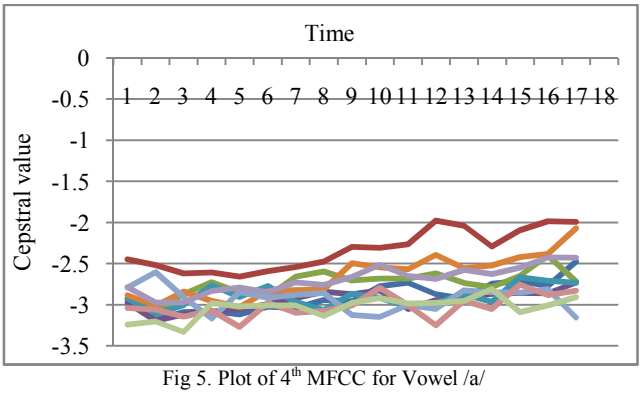


Fig 5. Plot of 4th MFCC for Vowel /a/

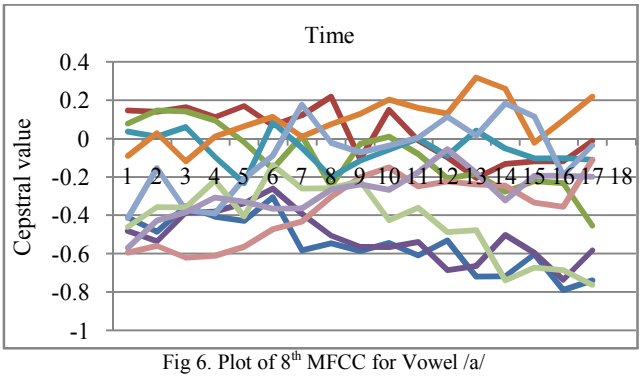


Fig 6. Plot of 8th MFCC for Vowel /a/

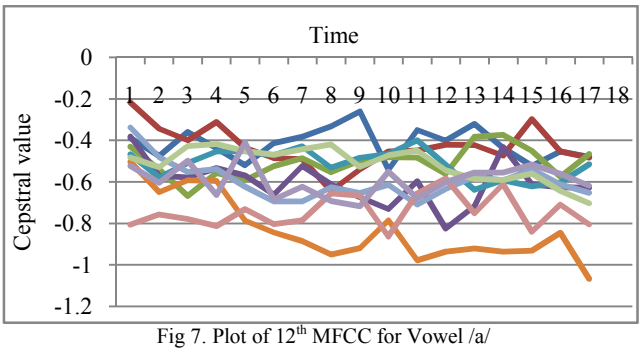


Fig 7. Plot of 12th MFCC for Vowel /a/

Similar analysis is performed for all the five English spoken vowels several times to find out these regular and non-regular set of MFCCs. The experimental results are as follows:

A. Experimental Outcomes:

The following observations were found based on the five English vowels using two recorded data sets and are tabulated in Tables I and II.

TABLE I & II
VOWELS AND REGULAR MFCCS FROM TWO EXPERIMENTS

Vowel	Regular MFCCs
/a/	c_1, c_2, c_4, c_6
/e/	c_1, c_2 , c_3 ,c_5
/i/	c_1, c_3, c_4, c_5
/o/	c_4, c_5, c_6, c_7
/u/	c_3, c_4, c_8, c_{10}

Vowel	Regular MFCCs
/a/	c_2, c_4, c_5,c_6
/e/	c_1, c_2 , c_3 ,c_5
/i/	c_1, c_3, c_5, c_7
/o/	c_2,c_4, c_5, c_6
/u/	c_4, c_5, c_8, c_{10}

From Tables III and IV we can observe that c_8, c_9, c_{11}, c_{12} do not contribute in formation of regular sets of MFCCs and hence might only be contributing in degradation of the recognition results.

TABLE III
CONSOLIDATED SET OF THREE MFCCS FOR EACH VOWEL

Vowel	Regular MFCCs	Vowel	Regular MFCCs
/a/	c_2, c_4, c_6	/o/	c_4, c_5, c_6
/e/	c_1, c_2 , c_3	/u/	c_4, c_5, c_{10}
/i/	c_1, c_3, c_5		

TABLE IV
MFCC-WISE CONTRIBUTION FROM REGULAR SETS

MFCC	Vowel	MFCC	Vowel	MFCC	Vowel
c_1	/e/,/i/	c_5	/e/,/i/,/o/	c_9	
c_2	/a/,/e/	c_6	/a/,/o/	c_{10}	/u/
c_3	/e/,/i/	c_7	/i/	c_{11}	
c_4	/a/,/o/	c_8		c_{12}	

After formation of regular sets we proceed further to verify the results for each of the vowels. For this we take up several test steady state vowel instances and computer their distances from standard steady state vowel instances. First, taking all MFCCs and then using regular set of MFCCs, we compute the distance as the distinctiveness capability of different feature sets. The step-wise verification procedure is as follows:

1) Select 5 instances of steady states of vowels (one for each vowel) having same length and fix them as references i.e. standard steady state vowel instances.
2) Record multiple test utterances of different vowels and generate multiple test instances from steady states for each vowel.
3) Compute distance of test steady state instance with each of the reference steady state, initially taking all 12 MFCCs and then taking the set of 3 regular MFCCs using Euclidian distance:

$$D = \frac{1}{n}\sum_{i=1}^{n}(r_i - t_i)^2 \qquad (6)$$

where D is the distance, r_i is the ith MFCC for reference steady state vowel instance, t_i is the ith MFCC for test steady state vowel instance and n is the number of MFCCs used.

4) After computing the distances for all instances of each test–reference vowel pair, we calculate the average distance values and tabulate them.
5) Compare the results.

TABLE V
DISTANCE MATRIX WITH 12 MFCCS

Test \ Ref	/a/	/e/	/i/	/o/	/u/
/a/ 1	2.49	91.23	134.50	29.22	60.15
/a/ 2	3.76	85.35	128.30	33.14	55.91
/a/ 3	1.19	89.98	125.20	22.76	49.33
/a/ 4	3.54	85.05	125.50	32.85	51.86
/a/ avg	**2.74**	87.90	128.37	29.49	54.31
/e/ 1	81.97	1.72	25.70	55.60	78.42
/e/ 2	82.96	1.25	26.52	56.22	81.38
/e/ 3	79.20	2.44	33.83	56.61	83.87
/e/ 4	87.80	3.15	36.76	61.90	91.61
/e/ avg	82.98	**2.14**	30.70	57.58	83.82
/i/ 1	125.40	27.87	5.21	85.12	76.03
/i/ 2	127.70	28.05	9.74	86.89	77.29
/i/ 3	127.30	27.91	9.48	86.55	77.09
/i/ 4	102.90	22.66	4.21	68.43	56.95
/i/ avg	120.83	26.62	**7.16**	81.74	71.84
/o/ 1	26.13	61.12	88.31	2.17	37.72
/o/ 2	21.90	48.71	75.22	2.19	36.43
/o/ 3	22.44	65.5	92.77	2.17	38.57
/o/ 4	20.40	55.35	81.51	1.39	37.43
/o/ avg	22.71	57.67	84.45	**1.97**	37.53
/u/ 1	43.01	65.53	60.31	27.23	5.91
/u/ 2	39.32	69.93	67.60	23.38	5.77
/u/ 3	38.75	65.81	63.31	25.18	5.20
/u/ 4	41.56	62.47	60.38	30.65	4.64
/u/ avg	40.66	65.93	62.90	26.61	**5.37**

TABLE VI
DISTANCE MATRIX WITH 3 MFCCS

Test \ Ref	/a/	/e/	/i/	/o/	/u/
/a/ 1	2.65	273.90	408.60	28.72	50.02
/a/ 2	7.61	278.40	393.20	39.25	36.28
/a/ 3	1.71	259.20	387.70	40.84	36.17
/a/ 4	5.67	265.20	388.80	30.03	41.10
/a/ avg	**4.40**	269.17	394.57	34.71	40.89
/e/ 1	19.55	3.76	18.60	71.81	54.13
/e/ 2	19.73	2.58	13.55	65.13	52.51
/e/ 3	27.47	6.46	18.89	64.11	56.07
/e/ 4	33.73	8.11	13.32	60.35	46.58
/e/ avg	25.12	**5.22**	16.09	65.35	52.32
/i/ 1	55.53	81.94	9.85	58.63	18.10
/i/ 2	43.70	70.86	18.18	54.70	11.60
/i/ 3	44.40	71.31	18.08	56.47	77.09
/i/ 4	68.23	71.24	5.73	68.43	56.95
/i/ avg	52.96	73.83	**12.95**	60.10	17.76
/o/ 1	29.29	155.60	215.90	1.10	62.72
/o/ 2	18.74	122.20	179.40	1.17	54.89
/o/ 3	33.72	173.40	235.20	2.40	70.28
/o/ 4	20.35	147.10	208.20	1.11	55.90
/o/ avg	25.52	149.58	209.68	**1.44**	60.94
/u/ 1	80.78	234.70	181.60	41.43	5.23
/u/ 2	65.31	209.90	202.40	27.31	7.21
/u/ 3	54.25	215.20	194.40	28.03	7.97
/u/ 4	66.18	200.60	191.90	49.42	8.56
/u/ avg	66.63	215.10	192.58	36.55	**7.24**

From results shown in Tables V and VI it is evident that for different test–reference vowel pairs, distance values obtained through regular MFCCs sets are more than the distances obtained using all 12 MFCCs for most vowels. It is also clear that the distinctiveness of the distance is more as the dissimilar vowels show large dissimilarity.

VI. CONCLUSION

From these preliminary experiments we conclude that specific subsets of features of conventional MFCC feature sets exhibit more uniform behavior for particular vowel sounds than for others. Taking that as a back drop, more experiments are to be conducted to substantiate this fact for several utterances as well as for several speakers so that a speech recognition system with a reduced set may be made possible with at least with same recognition accuracy or better.

ACKNOWLEDGMENT

The authors gratefully acknowledge the funding support provided by DeiTy, New Delhi to carry out this research work.

REFERENCES

[1] Andras Zolnay, “Acoustic Feature Combination for Speech Recognition,” PhD thesis, November 2006.

[2] Chin Kim On, Paulraj M. Pandiyan, Sazali Yaacob and Azali Saudi, “Mel-Frequency Cepstral Coefficient Analysis in Speech Recognition,” in International Conference on Computing & Informatics, ICOCI '06., pp. 1 – 5, 2006,.

[3] H. Kekre, V. Bharadi, A. Sawant, O. Kadam, P. Lanke and R. Lodhiya, “Speaker Recognition using Vector Quantization by MFCC and KMCG Clustering Algorithm,” in International Conference on Communication, Information & Computing Technology (ICCICT), Mumbai, pp. 1 – 5, 2012.

[4] Hynek Hermansky, “Perceptual Linear Predictive (PLP) analysis of Speech,” in Journal of the Acoustical Society of America, vol 87, no. 4, pp. 1738 – 1752, April 1990.

[5] Hynek Hermansky, Nelson Morgan, “RASTA processing of Speech,” in IEEE Transactions on Speech and Audio Processing, vol. 2, no. 4, pp. 578 – 589, October 1994.

[6] Joseph Picone, “Signal Modeling Techniques in Speech Recognition,” in Proceedings of the *IEEE*, vol. 81, Issue: 9, pp. 1215 – 1247, 1993.

[7] M. Hossan, S. Memon and M. Gregory, “A novel approach for MFCC feature extraction,” in 4th International Conference on Signal Processing and Communication Systems (ICSPCS), Gold Coast, QLD, pp. 1 – 5, 2010.

[8] Tze Fen Li and Shui-Ching Chang, “Speech Recognition of Mandarin Syllables using both Linear Predict Coding Cepstra and Mel Frequency Cepstra,” in Proceedings of the 19th Conference on Computational Linguistics and Speech Processing, ROCLING, Taipei, Taiwan, ROC, September 2007.

[9] Utpal Bhattacharjee, “A Comparative Study Of LPCC And MFCC Features for the Recognition of Assamese Phonemes,” in International Journal of Engineering Research & Technology (IJERT), vol. 2 Issue 1, pp. 1 – 6, January 2013.

[10] Yuan Yujin, Zhao Peihua and Zhou Qun, “Research of Speaker Recognition based on combination of LPCC and MFCC,” in IEEE International Conference on Intelligent Computing and Intelligent Systems (ICIS), pp. 765 – 767, 2010.

[11] Hrudaya Ku. Tripathy, Pradip K. Das and A. B. Laha, "High Precision Vowel Recognition based on a Hybrid Approach", in Proceedings of the 1st International Conference on Signal and Image Processing, BVB College of Engineering and Technology, Hubli, Karnataka, India, 7-9 December 2006, pp. 992-996.

[12] Hrudaya Ku. Tripathy, B. K. Tripathy and Pradip K. Das, "A Knowledge based Approach using Fuzzy Inference Rules for Vowel Recognition", Journal of Convergence Information Technology, Volume 3, Number 1, March 31, 2008, pp. 51-56, ISSN: 1975-9320.

Video-Based Affinity Group Detection Using Trajectories of Multiple Subjects

Abdullah Al Masum, Mahady Hasan Rafy, S M Mahbubur Rahman*, *Member*, *IEEE*
Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology
Dhaka-1205, Bangladesh
E-mail: {masum.buet08, mahmud081}@gmail.com, mahbubur@eee.buet.ac.bd

***Abstract*—Affinity detection has been largely motivated by the increasing interest in modelling the social behavior of humans. This paper presents a supervised learning method for affinity detection which is based on an inference obtained from tracking trajectories of the human subjects captured in video sequences. In particular, the proxemic cues of group detection such as the pair-wise similarity of the positional and translational measurements of the tracked people are used in the well-known principal component analysis-based feature extraction process. The existence or non-existence of pair-wise affinities is recognized using the nearest neighbor detector applied on the proposed features and the majority voting-based fusion of decisions. Experiments conducted on surveillance video captured in diverse-type of movements of the subjects show favorable results in terms of accuracy of detecting affinities when compared with the ground truth.**

***Index Terms*—Affinity, social interactions, tracking, video surveillance**

I. Introduction

Social interactions are ubiquitous in our daily life. People organize themselves in groups not only to perform activities, but also to share views, opinions and thoughts. Through the observations of social interactions many behavioral traits of the interlocutors can be inferred [1]. For this reason, in order to find a group, the automatic detection of affinity from movement dynamics of people has become an active research topic over the last few years especially in the area of video surveillance. In general, a group refers to a collection of people those maintain relative closeness of in terms of position or moving trajectory in a given local area.

The seminal work in computer vision on movement of a group of living being may be attributed to Reynolds [2], where it was reported that the motions of flock of birds avoid collisions as well as they maintain matching velocity and remain close to the flock center. The social force model reported in [3] shows that the walking patterns of group of humans have a moving tendency towards a single destination with similar velocity. Further, the subjects in the group maintain certain distance to avoid collision and possess equal strength attractive or repulsive proneness. The perceived personal space and comfort levels defined by sociologists have been incorporated in the Voronoi diagrams to find the voluntary or involuntary groups formed in crowds [4]. The Bayesian nonparametric probabilistic models have also been introduced for group classification by considering the interpersonal distance and velocity as random variables [1].

In many cases, the affinities are interactively used to improve the accuracy of the trajectories of the individuals within a group. For example, a conditional random field is used to find the association of a membership in a group and thereby accuracy of the motion trajectory of that person is improved in an iterative way in [5]. An energy functional comprising a number of terms including the velocity, change of velocity, moving direction, attraction, grouping, and collision are minimized to obtain reasonably accurate motion trajectories of the pedestrian in [6]. In [7] and [8], it has been shown that the learning of the affinities of motion tracklets (fragments of tracks) and thereby linking of these supervised tracklets using the Hungarian or K-means clustering algorithm may increase the accuracy of tracking trajectories. Detection tracklets and trajectory tracklets have also been used to overcome the problems of occlusion while tracking of groups [9]. A more detail about such algorithms those have used the affinities of the tracklets for improving the accuracies of motion trajectories may be found in [10].

The understandings of affinities hidden in the interpersonal distances and moving velocities are not only required for classifying the groups and improving accuracies of motion trajectories, but also for the detection of social events or abnormalities. In [11], the trajectories and histograms of the subjects have been used in the hidden Markov tree model for classification of three types of unusual events, namely, the subject moving in opposite direction, still subject in a moving crowd, and subject moving faster than the crowd. Symmetric and asymmetric events are identified in [12] by considering activity features based on group representative.

In the literature, the hidden affinity information obtained from the positional or translational measurements are often used for classification of groups and events commonly seen in our social life. Even in many cases, these information are interactively used to improve the accuracies of the measured quantities. In our view, the existence of affinity should be detected first, before using such information for the classification of groups and events or for improving the accuracy of the measurements. In this context, this paper presents a novel supervised detection algorithm to identify the existence of pair-wise affinity of two persons in video surveillance. The existence of affinity is detected using the principal component analysis (PCA)-based learning process of the positional and translational features obtained from trajectories. Experiments are conducted on a developed video database to evaluate the performance of the proposed affinity detection algorithm.

The paper is organized as follows. In Section II, the proposed affinity detection algorithm is presented. In Section III, experimental results of affinity detection on the video database are given. Finally, concluding remarks are provided in Section IV.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

II. Affinity Detection Method

The proposed affinity detection method takes the measurements of a tracking algorithm as inputs. The tracking algorithm must be capable of providing frame-by-frame positions of multiple subjects in a video sequence. Let O_k^i $(i \in 1, 2, \cdots, M)$ $(k \in 1, 2, \cdots, N)$ represent a subject in a tracklet of video sequence, where M and N are the number of individuals and number of frames in the tracklet, respectively. The proposed method first extracts the affinity features of pair-wise groups from a suitable set of measurements obtained from training tracklet, then the supervisory feature extraction process is applied to the pair-wise groups in the testing phase. The closeness of the test features with the training features determines the existence of affinity in question. Next, we present the step-by-step procedure of obtaining the pair-wise affinity using the trajectories of multiple subjects. The steps include the pair-wise group selection, construction of affinity vector, feature selection from the tracklets, and feature comparison and fusion for affinity detection.

A. Pair-Wise Affinity Group Selection

Let the affinity of a pair of subjects be denoted as $\mathcal{A}^{ij}$ $(i \neq j)$, the logic value one of which shows the existence of affinity and vice versa. Let $\mathcal{S}_n$ $(n \in 1, 2, \cdots, Z)$ denote a social group formed by a pair of subjects in a scene, where total number of groups in a frame k is given by

$$Z = \binom{M}{2} = \frac{M(M-1)}{2} \tag{1}$$

The number of frames N in a tracklet is chosen in such a way that Z remains same for the entire frames in the tracklets with the exceptions due to occlusions. It is noted that although this study concentrates on the pair-wise affinity, the decision on affinity of multiple persons can be reached easily by identifying the common individuals in such pair-wise groups.

B. Construction of Affinity Vector

The proxemic cues such as the positional and translational measurements of humans are used for quantifying affinity [13]. In the proposed method, the features of affinity are obtained from the positions and velocities of x- and y-coordinates measured by the tracker from a video. Let each of four measurements be denoted as ϕ_{km}^i $(m \in X, V_x, Y, V_y)$, where X, V_x, Y, V_y represent the coordinate position and velocity of a subject $\mathcal{O}_k^i$ in x and y-axes, respectively. It is noted that if the frame rate of the video is known, then the velocity of a subject can be measured only from his or her positions in the frames. Since the closeness of these measurements of a pair of subjects indicates the affinity between two, the affinity can be quantified as

$$\gamma_{km}^{ij} = \left\| \phi_{km}^i - \phi_{km}^j \right\|_2 \qquad i \neq j \tag{2}$$

where $\| \cdot \|_2$ denotes the Euclidean norm. The affinity vector of size $(N \times 1)$ for a pair-wise group S_n in the tracklet can be constructed from a measurement-type m as

$$\mathbf{\Gamma}_m^n = \left[\gamma_{1m}, \gamma_{2m}, \cdots, \gamma_{Nm} \right]^T \tag{3}$$

where T represents the transpose of the vector.

C. Feature Selection from Tracklets

Let η $(\eta < Z)$ be the number of pair-wise group for which the existence of affinity is confirmed in a training tracklet. Thus, an affinity existence matrix (AEM) of size $(N \times \eta)$ may be formed from the observed vectors as

$$\mathbf{\Psi}_m^\alpha = \left[\mathbf{\Gamma}_m^1, \mathbf{\Gamma}_m^2, \cdots, \mathbf{\Gamma}_m^\eta \right] \tag{4}$$

In a similar fashion, the affinity non-existence matrix (ANM) of size $(N \times (Z - \eta))$ can be formed as

$$\mathbf{\Psi}_m^\chi = \left[\mathbf{\Gamma}_m^1, \mathbf{\Gamma}_m^2, \cdots, \mathbf{\Gamma}_m^{(Z-\eta)} \right] \tag{5}$$

The positional or translational measurements in the matrices AFM and ANM must go through a dimension reduction process, so that affinity or non-affinity features can be extracted efficiently. In the proposed method, we prefer to use the PCA-based dimension reduction technique, since such method uses the singular value decomposition which is independent of data-type. In the PCA-based approach, the unitary vector $\boldsymbol{W}_m^\alpha$, i.e., the optimal projection axis, for the AFM is sought in such a way that the vector maximizes the generalized total scatter criterion defined by [14]

$$\mathcal{J}(\boldsymbol{W}_m^\alpha) = \boldsymbol{W}_m^\alpha \boldsymbol{C}_m^\alpha [\boldsymbol{W}_m^\alpha]^T \tag{6}$$

where $\boldsymbol{C}_m^\alpha$ is the covariance or scatter matrix of size $(\eta \times \eta)$ given as

$$\boldsymbol{C}_m^\alpha = E\Big([\mathbf{\Psi}_m^\alpha - E(\mathbf{\Psi}_m^\alpha)]^T [\mathbf{\Psi}_m^\alpha - E(\mathbf{\Psi}_m^\alpha)] \Big) \tag{7}$$

with $E(\cdot)$ being the expectation operator. It can be shown that the optimal projection axis $\tilde{\boldsymbol{W}}_m^\alpha$ is the eigenvector of $\boldsymbol{C}_m^\alpha$ corresponding to the largest eigenvalue [14]. In general, it is not enough to have only one optimal projection axis, instead the vectors corresponding to d $(d \ll N, \eta)$ number of eigenvalues are used to form the projection matrix of size $(d \times \eta)$ denoted as $\boldsymbol{P}_m^\alpha$. Finally, the features of affinity groups of training tracklets having size $(d \times N)$ are obtained as

$$\boldsymbol{G}_m^\alpha = \boldsymbol{P}_m^\alpha [\mathbf{\Psi}_m^\alpha]^T \tag{8}$$

In a similar fashion, the projection matrix of non-affinity groups $\boldsymbol{P}_m^\chi$ can be obtained and the features of the corresponding groups in the training tracklets are estimated as

$$\boldsymbol{G}_m^\chi = \boldsymbol{P}_m^\chi [\mathbf{\Psi}_m^\chi]^T \tag{9}$$

D. Feature Comparison and Fusion

At a given instance, the measurement-type m $(m \in X, V_x, Y, V_y)$ may form an observation matrix $\mathbf{\Psi}_m$ for which the belongingness to affinity or non-affinity is to be detected. The features of these observations obtained from the projection matrices corresponding to AFM and ANM, viz., $\boldsymbol{P}_m^\alpha$ and $\boldsymbol{P}_m^\chi$, may be denoted as $\boldsymbol{F}_m^\alpha$ and $\boldsymbol{F}_m^\chi$, respectively. In the proposed method, we prefer to use the cosine angle to measure the feature similarity. The angles between the observed feature and any of the training features in the cases of affinity and non-affinity conditions are given by

$$\theta_m^\alpha = \cos^{-1} \frac{\langle \boldsymbol{F}_m^\alpha, \boldsymbol{G}_m^\alpha \rangle}{\|\boldsymbol{F}_m^\alpha\|_2 \|\boldsymbol{G}_m^\alpha\|_2} \tag{10}$$

$$\theta_m^\chi = \cos^{-1} \frac{\langle \boldsymbol{F}_m^\chi, \boldsymbol{G}_m^\chi \rangle}{\|\boldsymbol{F}_m^\chi\|_2 \|\boldsymbol{G}_m^\chi\|_2} \tag{11}$$

where $\langle \cdot, \cdot \rangle$ represents the inner product. Existence of affinity with a single-type measurement m is detected by taking the

(a) (b)

Fig. 1. First and last frames of a typical video clip comprising variable number of subjects of interest with tracking trajectories. Number of subjects are such that one pair-wise affinity exists in (a) five subjects (b) four subjects.

decision obtained from the observed and training sets that produces the minimum angular distance of the features. Hence, the decision on a group in a tracklet is determined from the observation m as

$$\tilde{\mathcal{A}}_m = \begin{cases} A^1 & \text{if } (\theta_m^{\alpha})_{\min} < (\theta_m^{\chi})_{\min} \\ A^0 & \text{if } (\theta_m^{\alpha})_{\min} \geq (\theta_m^{\chi})_{\min} \end{cases} \tag{12}$$

where A^1 and A^0 correspond to the existence and non-existence of the affinity in the group, $(\theta_m^{\alpha})_{\min}$ and $(\theta_m^{\chi})_{\min}$ are the two minimums of the angular distances of the observed feature from the affinity and non-affinity training features, respectively. Since the existence of affinity in a pair-wise group is detected from four features defined by m, in order to reach to the final decision on the existence of affinity requires a fusion mechanism. In the proposed method, we use a majority voting-based decision fusion to detect the affinity of a pair-wise group given by

$$\hat{\mathcal{A}} = \text{maj}_m(\tilde{\mathcal{A}}_m) \tag{13}$$

where maj$(\cdot)$ provides the majority of the individual feature-based decisions. Due to the fact that four measurements are used in the proposed detection method, a tie may occur in which case the preference on fusing the decisions was assigned to the lateral direction of the motion trajectory of the subjects. In other words, if the flow direction is in y-direction then the preference is X, V_x, Y, and V_y, otherwise for flow of x-direction it is Y, V_y, X, and V_x.

III. Experimental Results

Evaluation of the performance of video-based detection of the affinity in a group is a challenging task. This is mainly due to the absence of well accepted evaluation protocol or publicly available database considering different scenarios. In this context, we have developed a typical video-based affinity detection database consisting of 30 video clips each with a length of at least 6 seconds captured in the vicinity of university area. The frame rate of the clips is 50 fps, wherein the pixel resolution of each frame is 480×640. One third of the clips are captured in such a way that at least three human-subjects are in slow moving condition among which at least one pair-wise affinity is assured. One third of the video clips have subjects with notable variations of the speeds such as strolling, running, and cycling. Trajectories of the subjects in these clips include the uniflow forward or reverse directions and curvature-type. Finally, the third set of clips comprises subjects with random walks in crowds. On average, half of the video clips are captured in the sunny conditions and half in cloudy or low illumination conditions.

The tracking trajectories of the subjects of interests in the video clips are estimated using the Gaussian-scale mixture models of the background and foreground developed in [15]. Fig. 1 shows first and last frames of a typical video clips used in the experiments along with the trajectories of the subjects of interests. From the frames shown in this figure, it can be seen that the number of subjects vary over the period of observation. The variations of the speeds include the walking, running, and cycling. The direction of trajectories are uniflow forward and backward as well as curvature-type. In the figure, the ground truth affinity is shown using green color boundaries of the subjects, whereas non-affinities are shown using red color boundaries.

In the proposed method, an important choice is the number of frames N chosen to form the tracklets. In the experiments, it is observed that for video clips wherein the variations of speed is low for the subjects of interests, a long sequence of frames can be used to form tracklets for increasing the accuracy of affinity detection. On the other hand, for high level of movements in a crowd, a short sequence is preferable to construct the tracklets. It is also noted that computational load of the affinity detection method increases with the length of sequence used in forming the tracklets. As a compromise between overall accuracy and computational complexity, the results of affinity detection presented in this paper use $N = 10$ to form the tracklets in all video clips. In order that the results obtained are statistically robust, the tacklets are formed by randomly chosen N number of chornological sequence in a clip and then accuracy of affinities or non-affinities are detected. The average results of accuracy shown in this paper use 20 number of such randomly selected set of sequences.

Table I shows the average of accurately detected pair-wise affinity or non-affinity groups those exist in three-types of video clips considered in the experiments. Here, the accurately

TABLE I
RESULTS CONCERNING THE ACCURATELY IDENTIFIED AFFINITY OR NON-AFFINITY GROUPS IN THE VIDEO CLIPS BY USING DIFFERENT FEATURES†

Video Sequences	Ground Truth		X		Y		V_x		V_y		Fused	
	A^1	A^0	A^1	A^0	A^1	A^0	A^1	A^0	A^1	A^0	A^1	A^0
Clip Type-1 (Slow Moving)	**25**	**80**	15(60)	80(100)	7(28)	80(100)	20(80)	80(100)	25(100)	75(94)	23(92)	78(98)
Clip Type-2 (Speed Variations)	**25**	**120**	20(80)	120(100)	10(40)	115(96)	18(72)	115(96)	6(24)	120(100)	20(80)	102(85)
Clip Type-3 (Crowded Place)	**50**	**150**	32(64)	150(100)	50(100)	140(93)	20(40)	150(100)	6(12)	150(100)	32(64)	147(98)
Overall Accuracy	**100**	**350**	67(67)	350(100)	67(67)	335(96)	58(58)	345(99)	37(37)	345(99)	75(75)	327(93)

†The values inside the brackets indicate the accuracies in percentage of the detections.

detected groups using four-types of features those are obtained from positional and translational measurements of the trajectories are presented individually as well as the final results obtained from decision fusion are also given. A comparison of detection results with the ground truth reveals the following notable observations:

- Accuracies of detecting affinities decrease with the speed as well as the randomness of the movements of the subjects.
- In a slow moving groups, information obtained from the lateral distance X as well as the translational measurements V_x and V_y are effective for affinity detection.
- In case of speed variations within the subjects, information obtained from the lateral distance X and velocity V_x are more useful than the flow directional measurements.
- If the subjects are in a crowd, positional measurements X and Y are more effective than the translational information.
- The fusion of decisions provides the overall best results on affinity detection, if the types of movements of the subjects are not known to the algorithm. In such a case, however, the positional information plays the significant role as compared to the translational information.
- The proposed features are capable of detecting the non-existence of affinity without any significant deviations.

IV. CONCLUSION

Affinities inferred from the trajectories of video surveillance have potential applications in understanding the social interaction of humans, prior detection of possible events, or even improving the accuracy of tracking. This paper has presented a supervised learning method to automatically detect the existence of affinity, which can latter be used for further inference as necessary. In order to detect the affinities, the proxemic cues such as the positional and translational measurements have been used to obtain the affinity-related features. The PCA-based dimensionality reduction approach and the cosine angle-based similarity comparison method have been adopted to determine the existence of affinity in a pair-wise subjects. Experiments have been carried out on a number of video sequences having subjects with diverse type of movements such as slow moving, notable speed variations, and crowd to investigate the implications of the proposed feature sets on distinguishing between affinity and non-affinity. It is found that the measurements of position are more effective than that of translation of the subjects. Further, if the movement types of the subjects are unknown, a fusion of the feature-based decisions is necessary to reach to the conclusion about the existence of affinity in a group.

REFERENCES

[1] M. Zanotto, L. Bazzani, M. Cristani, and V. Murino, "Online Bayesian non-parametrics for social group detection," in *Proc. British Machine Vision Conference*, Surrey, UK, 2012, pp. 111.1–111.12.

[2] C. W. Reynolds, "Flocks, herds, and schools : A distributed behavioral model," *Computer Graphics*, vol. 21, no. 4, pp. 25–31, 1987.

[3] D. Helbing and P. Molnar, "Social force model for pedestrian dynamics," *Physical Review E*, vol. 51, no. 5, pp. 4282–4286, 1995.

[4] J. C. S. Jacques, A. Braun, J. Soldera, S. R. Musse, and C. R. Jung, "Understanding people motion in video sequences using Voronoi diagrams," *Pattern Analysis and Applications*, vol. 10, no. 4, pp. 321–332, 2007.

[5] S. Pellegrini, A. Ess, and L. V. Gool, "Improving data association by joint modeling of pedestrian trajectories and groupings," in *Lecture Notes in Computer Sciencce : 11th European Conference on Computer Vision*, Crete, Greece, 2010, pp. 452–465.

[6] K. Yamaguchi, A. C. Berg, L. E. Ortiz, and T. L. Berg, "Who are you with and Where are you going?" in *IEEE Conf. Computer Vision and Pattern Recognition*, Providence, RI, 2011, pp. 1345–1352.

[7] B. Yang and R. Nevatia, "Multi-target tracking by online learning of non-linear motion patterns and robust appearance models," in *IEEE Conf. Computer Vision and Pattern Recognition*, Providence, RI, 2012, pp. 1918–1925.

[8] Z. Qin and C. R. Shelton, "Improving multi-target tracking via social grouping," in *IEEE Conf. Computer Vision and Pattern Recognition*, Providence, RI, 2012, pp. 1972–1978.

[9] K. Fragkiadaki, W. Zhang, G. Zhang, and J. Shi, "Two-granularity tracking: Mediating trajectory and detection graphs for tracking under occlusions," in *Lecture Notes in Computer Sciencce : 12th European Conference on Computer Vision*, Florence, Italy, 2012, pp. 552–565.

[10] H. B. Shitrit, J. Berclaz, F. Fleuret, and P. Fua, "Multi-commodity network flow for tracking multiple people," *IEEE Trans. Pattern Analysis and Machine Intelligence*, accepted in 2013, DOI: 10.1109/TPAMI.2013.210.

[11] F. Porikli and T. Haga, "Event detection by eigenvector decomposition using object and frame features," in *IEEE Conf. Computer Vision and Pattern Recognition*, Washington, DC, 2004, p. 114.

[12] W. Lin, M.-T. Sun, R. Poovendran, and Z. Zhang, "Group event detection for video surveillance," in *42nd IEEE International Symposium on Circuits and Systems*, Taipei, Taiwan, 2009, pp. 2830–2833.

[13] E. T. Hall, *The Silent Language*, 1st ed. NY: Ancror Books, 1959.

[14] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, 1st ed. NJ: Prentice-Hall, 1982.

[15] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, vol. 2, Fort Collins, CO, 1999, pp. 246–252.

Discriminative Masking for Non-Cooperative IrisCode Recognition

Bashirul Azam Biswas, Shams Shad Islam Khan, S M Mahbubur Rahman*, *Member*, *IEEE*
Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology
Dhaka-1205, Bangladesh
E-mail: {bashirulazam, shamsshadrafi}@gmail.com, mahbubur@eee.buet.ac.bd

***Abstract*—The reliability of iris-based biometric recognition faces challenges due to non-cooperative image acquisition. This paper presents a novel non-cooperative recognition method by adopting a step-by-step segmentation process of the iris regions and modifying the generated IrisCodes in a way that the bits having significant discriminatory power are selected using an effective masking process. Both the statistical inter- and intraclass consistencies are used to estimate the discriminatory power of each bit position of the IrisCodes. Experimental results on commonly-referred non-cooperative iris image database show the significance of using discriminatory bits to improve the recognition accuracy for both the log-Gabor and discrete cosine transform-based IrisCodes.**

***Index Terms*—Discriminative masking, IrisCodes, non-cooperative iris recognition**

I. Introduction

Due to the increasing demand of security, the deployment of biometric identification systems that use physiological or behaviorial characteristics of humans becomes inevitable. Among different types of physiological biometrics, the increasing interest towards 'iris recognition' system for a large population is well-known. This is mainly due to the fact that the iris patterns such as furrows, freckles, crypts, and coronas possess lifetime permanence and the acquisition of iris images requires a noncontact process.

In 1987, Flom and Safir first developed the iris-based person identification system [1]. In general, an iris recognition system consists of three major components: (a) acquisition and localization of iris regions from eye images, (b) feature extraction and encoding of iris patterns, and (c) feature or code classification. Pioneering work for commercialization of iris recognition system may be attributed to Daugman [2]. It was reported that in case of cooperative acquisition of eye images, the annular region of iris image may be extracted by using a suitable integro-differential operator [2] or the Hough transform [3]. Next, the normalized and size-invariant templates are obtained by mapping the iris regions from the Cartesian coordinates to a nonconcentric coordinate system.

A considerable amount of research studies have been carried out in order to obtain suitable binary features, known as IrisCodes, from the iris templates [4]. Among these, the most common codes are generated from the binary phase of the Gabor filtered coefficients of the iris templates [2]. Ma *et al.* [5] proposed an IrisCode generation method by using the block-based local means and standard deviations of the Gabor filtered coefficients of the templates. The two dimensional (2D) Gauss-Hermite moments of the square-shaped non-overlapping blocks of the iris templates have also been used to obtain the IrisCodes [6]. Monro *et al.* [7] have generated IrisCodes using the zero-crossings of the differences of the 1D discrete cosine transform (DCT) of two suitably grouped and 45° rotated vertical scan lines of the iris templates. Other features used to generate IrisCodes from the rectangular templates include those are obtained from the dual-tree complex wavelets, directionlets, Laplacian and derivative of Gaussian, log-Gabor [8], Gabor energy, ridge energy, geometric moments, and improved scale invariant feature transform. In general, the bits of the IrisCodes contributed from the middle bands of the cooperative iris images are more consistent than the inner bands [9]. It was reported that the existence of inconsistent fragile-bits and their positions vary across genders and types of filters used to generate the codes. In [8], the fragile-bits are obtained by locating the positions for which the bit values change within a class. A class-dependent weight mask to locate the consistent bits in cooperative iris images using a computationally highly expensive optimization technique was reported in [10].

In the literature, most of the iris recognition schemes focus on the IrisCodes obtained from cooperative eye-images. The cooperative environment demands that the users are fully aware of recognition process and actively participate in highly constrained imaging conditions while capturing eye images. Images captured in less constrained conditions may lead to extremely heterogeneous data wherein iris regions may have severe obstructions arises from eyelids, eyelashes, or glasses due to off-angle camera view, reflected light on iris regions caused by illumination variations, poor quality intricate pattern caused by motion blur or out-of-focus camera lens. With a view to encompass the success of iris recognition algorithms in a broader range of applications, increasing interests have been found over the last few years considering non-cooperative iris recognition technique with specific degradation. For example, in [11], the off-axis iris images are corrected to frontal view images using the asymmetry of segmented regions and a suitable map of refractive indexes. The depth-of-field of the iris images has been improved by using suitable modulation transfer function [12] or simulated wavefront-coded image [13] for increasing recognition performance of out-of-focused iris images. Selection of frames having good quality iris regions or fusion of such frames in video sequences have also been attempted to take care about the problems of non-cooperative acquisition [14]. The success of recognition from nonideal iris recognition highly depends on the robust segmentation of the iris regions and the IrisCodes generated from the corresponding templates [15].

This paper presents a novel non-cooperative iris recognition algorithm in which the effective iris regions are isolated

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

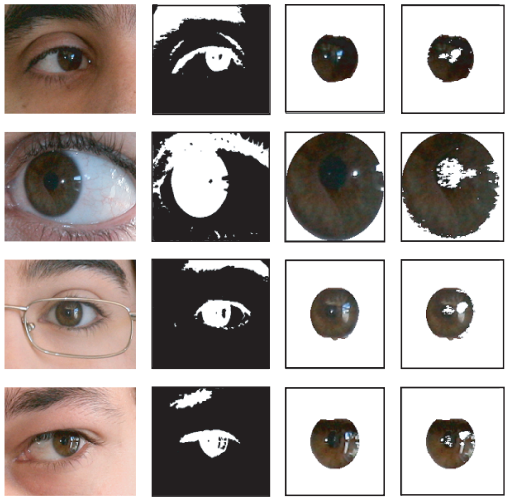


Fig. 1. Visual outputs of segmentation of iris regions from the non-cooperative images. Left most column shows sample images. The second column presents the non-skin regions. The iris and pupil regions are shown in third column. Finally, the iris regions excluding the pupils are shown in the rightmost column.

using a step-by-step segmentation process. Then, the IrisCodes are generated from the templates using the log-Gabor or DCT coefficients of the templates. A discriminative masking approach, which is independent of any kind of degradations, is proposed by estimating the statistical consistency of each bit in the IrisCodes. The significance of the masking is investigated on commonly-referred iris image database, in which images are captured from individuals with very little cooperation.

The paper is organized as follows. In Section II, the method of obtaining IrisCodes by segmentation of iris regions is given. The proposed approach of generation of discriminative mask is presented in Section III. In Section IV, experimental results of recognition on iris image database are given. Finally, concluding remarks are provided in Section V.

II. Generation of IrisCodes from Eye Images

In order to generate the IrisCode, the eye region is required to detect first from the acquired image. Then annular-shaped iris region is to be isolated from the eye image. The iris region is then mapped to a rectangular-shaped iris template, from which the binary IrisCode is generated using suitable feature extraction process. The step-by-step procedures of obtaining the IrisCode from the acquired image are given below.

A. Segmentation of Eye Region

In the non-cooperative iris recognition algorithm, the first task is to isolate the eye regions from the skin color face regions and hairy eyebrows. For such segmentation, statistical feature-based classification is adopted. In particular, the mean and standard deviation calculated from three local neighborhoods of sizes 3×3, 5×5, and 7×7 centered at each pixel of the color images are used as region features. Then, pixels each represented by nine features for skin and non-skin regions are clustered using the fuzzy C-means algorithm. The skin and non-skin areas are isolated by the K-nearest neighbor classifier, wherein the clustered training features are used and the K value is chosen as unity. It is noted that the eyes are the regions of interest that include the iris, pupil, and sclera in the non-skin areas excluding the eyebrows.

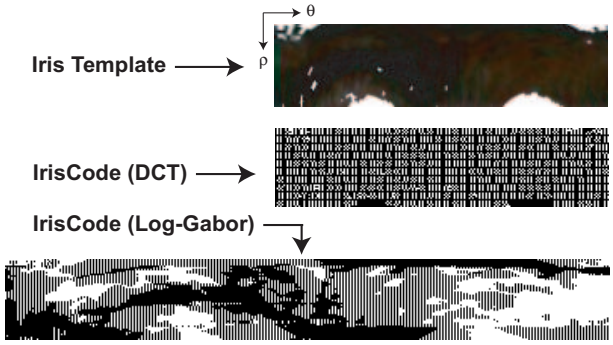


Fig. 2. Visual outputs of typical iris template as well as the corresponding DCT and log-Gabor-based IrisCodes.

B. Segmentation of Iris

Since the chromatic values of the iris regions, pupil inside the iris regions and the eyebrows are very close to each other, they can be treated as non-sclera regions. The sclera and non-sclera regions are segmented using the similar method adopted for classifying the skin and non-skin regions explained previously. The non-sclera regions comprise the circular-type iris regions including the pupil and arch-type eyelashes or eyebrows. In order to obtain the circular edges from the non-sclera regions, we have used the gradients of the pixel intensities obtained from the Canny edge detector and the circular Hough transform [6]. The closed-circular boundary includes the iris region and pupil, which is latter separated using the popular Otsu's thresholding method [16]. Fig. 1 shows the visual outputs of the segmentation of iris regions from typical four samples of images obtained from the database UBIRIS v2.0 [17]. The non-cooperative nature of the acquisition are obvious from the images shown in the left most column. The second column shows the non-skin areas that include the regions of interest. The third column shows the circular patterns inside the eye regions that include the iris as well as the pupil, wherein a zoom-in factor two is applied. Finally, the iris patterns are shown in the right most column by isolating the pupils remaining in the center area. These results clearly reveal that the proposed segmentation method can provide the iris regions very efficiently from the non-cooperative images.

C. Generation of IrisCode

Let the segmented region of an iris image be denoted as 2D continuous signal $I(x, y)$ in a Cartesian coordinate system of (x, y). Let r $(r > 0)$ be the radius of the circular iris/sclera boundary with a center (x_c, y_c) obtained from the Hough transform. According to the homogeneous rubber sheet model [6], the mapping of $I(x, y)$ to a doubly dimensionless and nonconcentric coordinate system of (ρ, θ) $(0 \leq \rho \leq 1)$, $(0 \leq \theta \leq 2\pi)$ can be represented as

$$I(x(\rho, \theta), y(\rho, \theta)) \longrightarrow I(\rho, \theta) \tag{1}$$

with

$$x(\rho, \theta) = (1 - \rho)x_c + \rho x_s(\theta) \tag{2}$$

$$y(\rho, \theta) = (1 - \rho)y_c + \rho y_s(\theta) \tag{3}$$

where the coordinates $(x_s(\theta), y_s(\theta))$ are on the circular iris/sclera boundary at an angle θ. The iris templates generated from this mapping process are ensure to have only the iris patterns obtained in the segmentation stage excluding

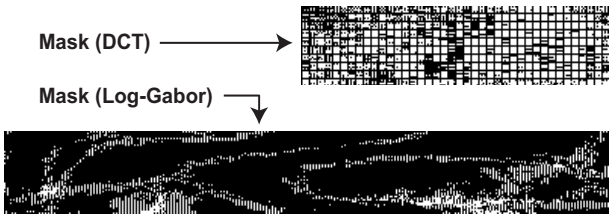


Fig. 3. Visual outputs of typical discriminative masks considering the DCT and log-Gabor-based IrisCodes.

the unnecessary regions such as the pupils, occlusions, and reflections. The iris templates generated may be used to obtain IrisCodes using different strategies adopted in the literature. In this study, we consider two types of IrisCodes obtained from the intensity image of iris templates. The first code is obtained from the quantization of the complex phase of log-Gabor coefficients of templates [8] and the second from the zero-crossings of 1D DCT coefficients of the recommended scan lines in the templates [7] except for choosing the non-overlapping windows both on the horizontal and vertical directions. Fig. 2 shows the visual outputs of a typical iris template and its corresponding codes using the DCT and log-Gabor-based features. It can be seen from this figure that the iris template generated from the proposed method effectively excludes the pupil and other unwanted regions. Due to the real and imaginary parts of complex phase, the log-Gabor-based IrisCode appears as twice the size of the DCT-based one.

III. Discriminative Masking of IrisCodes

It is well known that the random occurrence of each bit in the IrisCodes follows the binomial distribution. In order to recognize an identity using the IrisCodes, a significant number of bits should have close matching within the class. In practice, however, the occurrences of bits in a number of positions within a class appear to be independent when the codes are generated from non-cooperative images. Notably, the inclusion of such inconsistent bits rather acts as an impediment to obtain the minimum within class distance of IrisCodes. In such a case, the discriminant masks can be estimated from each class of the training IrisCodes, so that only the bits in the consistent positions may be used for recognition of a test code. To obtain the class-specific discriminant mask, we determine the within and between class consistencies for each bit position of the codes. Let an IrisCode of size $(M \times N)$ belong to a class label ℓ $(\ell \in 1, 2, \cdots, L)$ in a database of L number of classes. Let μ_{k1}^{ℓ} and μ_{k0}^{ℓ} $(0 \leq \mu_{k1}^{\ell}, \mu_{k0}^{\ell} \leq 1)$ represent the within class mean values of the bits '1' and '0', respectively, for a certain bit position k $(k \in 1, 2, \cdots, MN)$. Since perfectly followed random binomial distribution possesses equal valued mean for both the bits, the deviation of these mean values represents the within class consistency given by

$$\mathcal{W}_k^{\ell} = \left| \mu_{k1}^{\ell} - \mu_{k0}^{\ell} \right| \quad \ell = 1, 2, \cdots, L \quad k = 1, 2, \cdots, MN \tag{4}$$

In a similar fashion, the between class consistency of the bits at the same position can be estimated as

$$\mathcal{B}_k^{\ell} = \frac{1}{2L} \sum_{\ell=1}^{L} \left| (\bar{\mu}_{k1} - \mu_{k1}^{\ell}) - (\bar{\mu}_{k0} - \mu_{k0}^{\ell}) \right| \tag{5}$$
$$\ell = 1, 2, \cdots, L \quad k = 1, 2, \cdots, MN$$

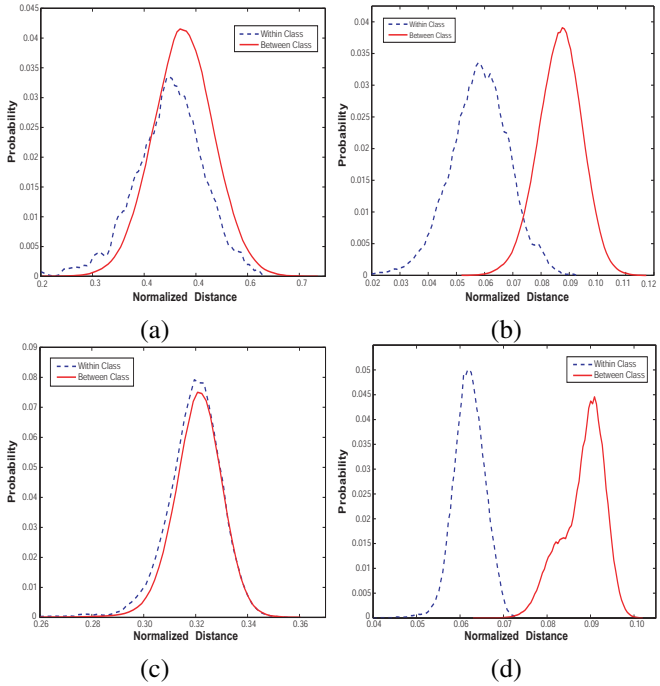


Fig. 4. Histograms of the within- and between-class Hamming distances showing the effect of masking in the IrisCodes. (a) Log-Gabor-based codes without mask [8], (b) log-Gabor-based codes with proposed mask, (c) DCT-based codes without mask [7], and (d) DCT-based codes with proposed mask.

where $\bar{\mu}_{k1}$ and $\bar{\mu}_{k0}$ $(0 \leq \bar{\mu}_{k1}, \bar{\mu}_{k0} \leq 1)$ are the mean values of the equiprobable bits '1' and '0', respectively, considering the entire database. A bit in the IrisCode will be effective for recognition, only if the bit position has a high within-class consistency and a low between-class consistency. In other words, the class-specific discriminant capability of a bit position can be measured as

$$\mathcal{D}_k^{\ell} = \frac{\mathcal{W}_k^{\ell}}{\mathcal{B}_k^{\ell}} \quad \ell = 1, 2, \cdots, L \quad k = 1, 2, \cdots, MN \tag{6}$$

In constructing the discriminative mask, we prefer to use only the first αMN $(0 < \alpha \leq 1)$ number of bit positions in the IrisCodes that have high values of $\mathcal{D}_k^{\ell}$. Fig. 3 shows typical class-specific discriminative masks obtained for the DCT and log-Gabor-based IrisCodes using the proposed method using $\alpha = 0.50$ and 0.20, respectively. From this figure, it can be seen that the significant discrimination exists in bit positions of the radial axis as compared to that of the angular axis for both the cases.

IV. Experimental Results

Extensive experimentations have been carried out to evaluate the clustering performance of the proposed discriminative iris recognition method as compared to the existing ones. In this section, the recognition results are given for a generic set of iris images, which are obtained from a well-known database, namely, UBIRIS v2.0 [17]. The database has 11,102 images each having a size of 800×600. The iris images were collected from two eyes of 261 persons. The database contains 14 types of non-cooperative factors, classified into two major categories: local and global. The local category is comprised of iris obstructions, reflections, off-angle, and partial images, while the global comprises poor focused, motion-blurred, rotated, improper lighting, and out-of-focus iris images. In order to obtain the generic database,

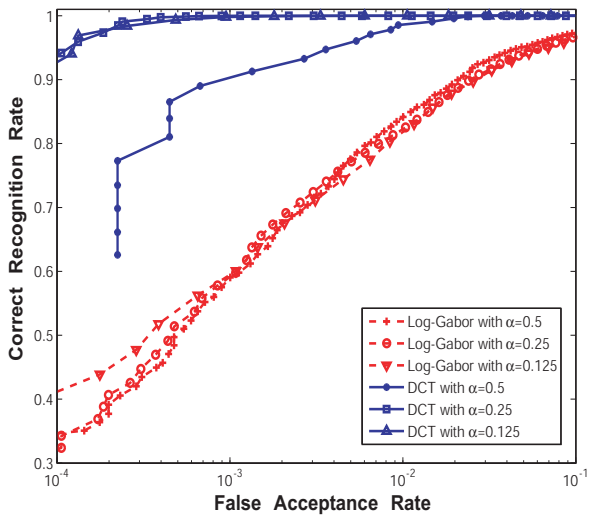


Fig. 5. ROC curves showing the correct recognition rate as a function of false alarm rate with varying α used in the masking of the log-Gabor and DCT-based IrisCodes.

we consider entire set of 261 classes with each having 15 iris images. The significance of using the proposed class-specific masking on the recognition accuracy is evaluated in two ways. First, the intra- and interclass distances are measured and the histograms are constructed to identify the separability of IrisCodes. Second, the receiver operating characteristics (ROC) curves are estimated to identify the probability of correct recognition for a given false alarm rate. In order that the experiments conducted are statistically robust, the histograms and the ROCs curves are estimated from a randomly chosen 50 persons each having 10 iris images and using the leave-one-out method. The average results presented in this paper are obtained by repeating the same process for 20 times.

Fig. 4 shows the histograms depicting the effect of proposed masking on the normalized inter- and intraclass Hamming distances of the log-Gabor [8] and DCT-based [7] IrisCodes with $\alpha = 0.25$. From this figure, it can be seen that in both the cases, the conventional IrisCodes fail to distinguish the classes from the distances estimated. On the other hand, if the proposed mask is applied with the codes, the mode of the histograms shifted significantly showing the capability of recognizing the iris images captured in non-cooperative environments. In the experiments, it is observed that such shifting of the modes is significant for the DCT-based codes than that of the log-Gabor-based codes. Fig. 5 shows the ROC curves obtained from the proposed mask-based IrisCodes depicting the impact of using the reduced number of discriminatory bits on the recognition rate for a given false alarm. From this figure, it is seen that the recognition rate for the log-Gabor-based IrisCodes does not change significantly with the selection of discriminatory bits through the masking process. It is noted that the masking on the DCT-based IrisCodes results in excellent performance on the recognition rate. It can be seen from Fig. 5 that only using the discriminatory bits having a size of 12.5% of the traditional IrisCodes may provide a recognition accuracy of approximately 94% at a very low false alarm rate of 10^{-4}. It is worth mentioning that the proposed discriminative masking method reduces the computational complexity of comparisons of codes significantly with a slight increasing load of enrollment process.

V. Conclusion

Iris recognition from non-cooperative image acquisition is a vision for the future biometric system, as an example, please refer to the science fiction feature film, Minority Report, released in 2002. Traditional IrisCodes designed for cooperative iris images completely fail in recognizing identity in non-cooperative environments. In this context, this paper has first presented a new step-by-step approach of segmentation of iris regions from the eye images captured in such environments. Then a class-specific mask has been introduced by estimating the discrimination capability of each bit of the IrisCodes generated from the DCT and log-Gabor-based coefficients of templates. Experiments have been conducted on a well-known UBIRIS v2.0 database that was designed for investigating the performance of non-cooperative iris recognition. The histograms of the inter- and intraclass distances and the ROC curves obtained from the DCT-based and log-Gabor-based IrisCodes show that the proposed segmentation and masking process are highly effective for non-cooperative recognition, especially for the former than the latter.

References

[1] L. Flom and A. Safir, "Iris recognition system," US Patent 4 641 394, 1987.

[2] J. Daugman, "High confidence visual recognition of persons by a test of statistical independence," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 11, pp. 1148–1161, 1993.

[3] R. P. Wildes, "Iris recognition: An emerging biometric technology," *Proc. IEEE*, vol. 85, no. 9, pp. 1348–1363, 1997.

[4] A. W. K. Kong, D. Zhang, and M. S. Kamel, "An analysis of IrisCode," *IEEE Trans. Image Processing*, vol. 19, no. 2, pp. 522–532, 2010.

[5] L. Ma, T. Tan, Y. Wang, and D. Zhang, "Personal identification based on iris texture analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1519–1533, 2003.

[6] S. M. M. Rahman, M. M. Reza, and Q. M. Z. Hasani, "Low-complexity iris recognition method using 2D Gauss-Hermite moments," in *Proc. 8th International Symposium on Image and Signal Processing and Analysis*, Trieste, Italy, 2013, pp. 135–139.

[7] D. M. Monro, S. Rakshit, and D. Zhang, "DCT-based iris recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 586–595, 2007.

[8] K. Hollingsworth, K. W. Bowyer, and P. J. Flynn, "Improved iris recognition through fusion of Hamming distance and fragile bit distance," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2465–2476, 2011.

[9] K. P. Hollingsworth, K. W. Bowyer, and P. J. Flynn, "The best bits in an iris code," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 6, pp. 964–973, 2009.

[10] W. Dong, Z. Sun, and T. Tan, "Iris matching based on personalized weight map," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 33, no. 9, pp. 1744–1757, 2011.

[11] Y. Si, J. Mei, and H. Gao, "Novel approaches to improve robustness, accuracy and rapidity of iris recognition systems," *IEEE Trans. Industrial Informatics*, vol. 8, no. 1, pp. 110–117, 2012.

[12] B. J. Kang and K. R. Park, "Real-time image restoration for iris recognition systems," *IEEE Trans. Systems Man, and Cybernatics - Part B*, vol. 37, no. 6, pp. 1555–1566, 2007.

[13] V. N. Boddeti and B. V. K. V. Kumar, "Extended-depth-of-field iris recognition using unrestored wavefront-coded imagery," *IEEE Trans. Systems Man, and Cybernatics - Part A*, vol. 40, no. 3, pp. 495–508, 2010.

[14] K. Hollingsworth, T. Peters, K. W. Bowyer, and P. J. Flynn, "Iris recognition using signal-level fusion of frames from video," *IEEE Trans. Information Forensics and Security*, vol. 4, no. 4, pp. 837–848, 2009.

[15] J. Zuo and N. A. Schmid, "On a methodology for robust segmentation of nonideal iris images," *IEEE Trans. Systems Man, and Cybernatics - Part B*, vol. 40, no. 3, pp. 703–718, 2010.

[16] W. Pedrycz, *Knowledge-Based Clustering: From Data to Information Granules*. Hoboken, NJ: Wiley & Sons, 2005.

[17] H. Proenca, S. Filipe, R. Santos, J. Oliveira, and L. A. Alexandre, "The UBIRIS.v2: A database of visible wavelength iris images captured on-the-move and at-a-distance," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 32, no. 8, pp. 1529–1535, 2010.

Significance of Acoustic Features for designing an Emotion Classification System

Sudhakar Kumar,[1,*] Tushar Kanti Das,[2,**] and Rabul Hussain Laskar[2,***]

[1]School of Electronics & Communication Engineering
Shri Mata Vaishno Devi University, Jammu and Kashmir, India

[2]Dept. of Electronics & Communication Engineering
National Institute of Technology Silchar, Silchar, India

[*]2012eec02@smvdu.ac.in [**]tusharkdnits@gmail.com
[***]rabul18@yahoo.com

***Abstract*—This paper reports some of the observations carried out on SUSE database for emotion classification. A comparative study is made to evaluate the performance of Linear Prediction Cepstral Coefficients (LPCCs) and Mel Frequency Cepstral Coefficients (MFCCs) for designing the emotion classification system for word level utterances. The significance of the orders of the coefficients has been carried out during this study. The results obtained using 12th order LPCC and 13th order MFCC are compared with respect to their reduced dimensions of lower orders. A new classification system based on the feature extraction technique using the 2nd, 3rd and 4th order coefficients of both MFCCs and LPCCs is also proposed. This paper compares the accuracy level of both MFCC and LPCC, enabling us to decide which orders of the parameters (both MFCC and LPCC) are more efficient in conveying the emotion for word level utterances. The initial experiments performed at word level utterances reveal that LPCC is more efficient in detecting emotion as compared to MFCC. Further, we noticed that the emotions conveyed in word level utterances are detected more accurately than that in sentence level utterances. The result suggests that word level approach provides better performance for emotion classification as compared to sentence level approach if the system is designed using vocal tract information only.**

***Index Terms*—Emotion, Classification, Vocal tract Characteristics, LPCC, MFCC, LP Residual, Mean Value Distance, Normalized Mean.**

I. INTRODUCTION

In present scenario, emotion classification finds its applications in several areas like human-computer interaction, speaker verification system, software engineering, animated characters, natural sounding voices [1], [2] etc. In any conversation, in addition to the information conveyed through the speech signal, the fashion in which the words are uttered also carries some important non-linguistic information such as the emotional characteristics of the speaker at that instant. Such characteristics may be parameterized to act as an emotion classifier for the speaker. The accurate classification of emotion becomes difficult in this sense that there is hardly any fixed parameter for differentiating among emotions i.e. there is no clear answer as to what the 'correct' emotion is for a given speech [2]. The situation becomes more critical when it comes to classifying similar emotions for different speakers. For example, a '*fellow'* might utter '*hello*' with a welcoming happy attitude as well as it can also mean an irritating kind of statement and certain other kind of emotions may be associated with it. Hence, it is a tough task to design a robust emotion classification system which may accurately detect the emotions of the speakers.

In order to design and train an emotion classification system, several databases (generally sentence based) have been employed like Chinese emotional database, Berlin emotional database [3], [2], SUSAS [4], IITKGP-SESC [5] etc. Generally the features extracted from a large number of speech samples are used for training the emotion classification system. However, as the databases are created by the speech samples of a fixed number of professional speakers (actors/actresses), so the parameterization of a non-professional speaker in an uncontrolled environment is going to be completely different.

Most of the algorithms used in emotion recognition are based on features extracted at sub-segmental (source excitation), segmental (vocal tract system) and supra-segmental (prosody) characteristics [6]. Though the excitation source signal contains some useful emotion oriented information, these features alone are not sufficient to provide appreciably high emotion detection performance. However the performance may be enhanced by combining spectral features. The results provided in the literature suggest that the vocal tract characteristics carry significant information about emotion. Therefore, in this paper, a comparative analysis has been carried out using MFCCs and LPCCs as acoustic features for classification of emotions. Moreover, the significance of the order of the coefficients with respect to emotion classification has also been carried out during this study. The contributions of the paper are: (a) a comparative analysis of LPCC and MFCC in context of emotion classification, (b) the significance of the order of the coefficients (either LPCC or MFCC) in case of emotion classification, (c) the word level versus sentence level utterances in conveying emotion when the system is designed using vocal tract information only.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

The rest of the paper is organized as follows. Section II deals with the database used for this study and different feature extraction techniques for extracting vocal tract information from the speech signal. The emotion modeling technique used in this paper has been discussed in section III. Section IV highlights the experimental observations made during this study. Section V summarizes the results along with the scope for future work.

II. EMOTION CLASSIFICATION USING SUSE DATABASE

A. Database

The SUSE (Speech under Simulated Emotion) database [4] has been used during this study. The emotions considered in the database are neutral, happy, angry, loud, compassion, sad, bore, fear, lombard and question. The database contains the words like break, degree, destination, hot, ten, hello etc. uttered by different speakers (two utterances per speaker) for different words in all the emotions. All speeches collected in this database are sampled at 8 KHz and sampling frequency with 16 bits/sample resolution.

For the system proposed here, only four emotions-anger, lombard, neutral and question-have been used.

B. Feature Extraction Technique

The emotions of a particular speaker lies at sub-segmental (source excitation), segmental (vocal tract system) and supra-segmental (prosody) levels in the speech signal. The literature reveals that vocal tract characteristics are more significant in case of emotion classification. Two acoustics such as LPCC and MFCC are generally used to represent the vocal tract characteristics. These two feature extraction techniques are discussed as:

1) Mel Frequency Cepstral Coefficients:

Non-linear Mel scale filter bank is used to emphasize lower frequency components over higher ones because of the following two reasons:

- Lower frequency components of any speech sample carry more phoneme information.
- Human auditory system processes speech signals in a non-linear fashion.

In speech processing, Mel Frequency Cepstrum Coefficients (MFCC) is a representation of the short term power spectrum of a speech frame using linear cosine transform of log power spectrum on a non-linear Mel frequency scale. The notable difference between the normal cepstrum and MFCC is that in MFCC, the frequency bands spacing in the Mel scale increases gradually as we move from lower to higher frequency regions and gives the perception of a human ear, which approximates our auditory system's response more closely than the linearly spaced frequency bands.

Conversion formula from normal frequency *f (linear scale)* to Mel frequency *m (Mel scale)* is:

$$m = 2595 \log_{10}\left(\frac{f}{700} + 1\right) \tag{1}$$

The various studies suggest that MFCCs provide an appreciable performance while classifying various emotions. The experiments carried out on IITKGP-SESC database suggest that the average emotion classification performance using MFCC features (using block processing approach) is 63.38%, while that using MFCC with Formant features is 68% [7].

2) Linear Prediction Cepstral Coefficients:

In the context of emotion classification, LPCCs are used to capture emotion specific information manifested through vocal tract characteristics of the speaker. In this paper, the 12th order LPCC has been calculated per speech frame of 20 ms using a frame shift of 10ms (20±50%).

The key idea behind LP analysis is that m^{th} speech sample can be predicted by linear combination of its previous n samples i.e.

$$S(m) \approx a_1 S(m-1) + a_2 S(m-2) + \cdots + a_n S(m-n) \tag{2}$$

where $a_1, a_2, a_3, \ldots$ are the constants (known as LP Coefficients) over a speech analysis frame. These coefficients are employed to predict the speech samples. Further the error in prediction $(e(m))$is calculated by

$$(e(m)) = S(m) - \hat{S}(m) = S(m) - \sum_{k=1}^{n} a_k\, S(m-k) \tag{3}$$

where $S(m)$ and $\hat{S}(m)$ are the original and predicted speech signals respectively. Now in order to compute a unique set of predictor coefficients, sum of squared

$$E_m = \sum_p [S_m(m) - \sum_{k=1}^{n} a_k S_m(p-k)]^2 \tag{4}$$

where p is the number of samples in a particular analysis frame. To solve above equation for LPCC, E_m has to be differentiated with respect to each a_k and equating the result to zero as

$$\frac{\partial E_m}{\partial a_k} = 0 \qquad for\ k = 1, 2, 3 \ldots n \tag{5}$$

After evaluating all a_k, we may find Cepstral coefficients using the following recursion:

$$C_0 = \log_e n \tag{6}$$

$$C_p = a_p + \sum_{k=1}^{p-1} \frac{k}{p} C_k\, a_{p-k} \quad for\ 1 < p < n \tag{7}$$

$$C_p = \sum_{k=p-n}^{p-1} \frac{k}{p} C_k\, a_{p-k} \qquad for\ p > n \tag{8}$$

The emotion classifier using LPCC provides better results as compared to MFCC. The experimental observation (performed on IITKGP-SESC) indicate that the average emotion recognition using LPCC features (using block processing approach) is 68%, while that using LPCC with Formant features is 69% [7].

III. EMOTION MODELING TECHNIQUE

Several authors have proposed their algorithms using excitation source, vocal tract system and prosodic features. While using spectral features, more often vector quantization

[8] is performed and consequently LBG algorithm [9], [10] is used for training purposes

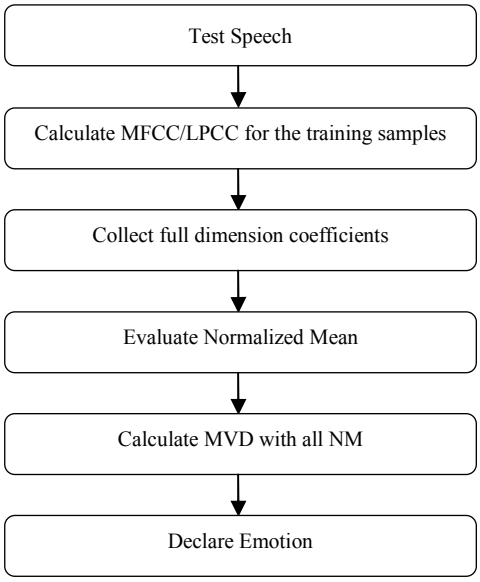


Fig. 2. System Modeling (full dimension)

.

The modeling technique has been devised using both LPCC and MFCC. Initially the technique was formulated using the full dimension of the spectral features (i.e. up to 12th order and 13th order in case of LPCCs and MFCCs respectively). The reduced dimension using various orders has been explored for designing emotion classification system. It has been observed that the feature vector derived using the 2nd, 3rd and 4th order dimension provides an almost equivalent performance as compared to its full dimension.

For training purposes, ten speech samples (regardless of the speaker and the text) for all the four emotions were taken. Initially, using the full dimension MFCC/LPCC, a normalized mean (NM) was set as a reference template for all the four emotions. While testing, when some input speech sample was fed, the mean value distance (MVD) between the full frame coefficients (MFCC/LPCC) of the input sample and the reference template of all the emotions was evaluated. The emotion was declared to be one which showed minimum distance with the input sample.

Later, while using reduced dimension (for both LPCC/MFCC), the same methodology was followed but this time, the normalized mean was set using only the 2nd, 3rd and 4th order coefficients. The flow chart depicting the full dimension model in Fig.2 and reduced dimension in Fig.3.

IV. Experimental Observations

The experimental observations (using full dimension) carried out on the sentence level utterances (IITKGP-SESC) depict that the average emotion recognition using MFCCs is 63.38%, whereas it is 68% using LPCCs [7]. In the proposed system (carried out on SUSE database), the performance of emotion classification using full dimension is provided in table I.

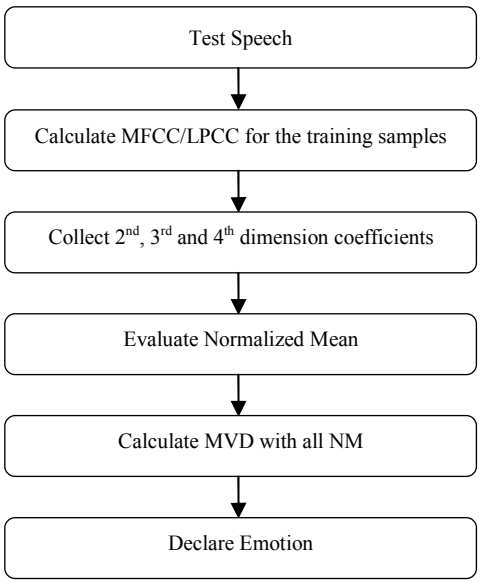


Fig. 3. System Modeling (reduced dimension)

It has been observed table I, in case of full dimension approach, the accuracy of emotion classification system designed using MFCC as acoustic feature is 79.52% (EER 20.48%) and that for LPCC is 83.60% (EER 16.40%).

The experiment has been carried to verify the importance of acoustic dimension of LPCC/MFCC with respect to their contribution in the spoken utterances like words. The framework of the experiment is designed using the reduced dimensions of LPCC/MFCC. Various combinations of the dimensions have been explored during this study. The study reveals that the feature vectors designed using the reduced dimensions provide comparable results as that of the full

TABLE I. Emotion Recognition Performance (In %) Using Full Dimension MFCC/LPCC

Emotion	MFCC Approach		LPCC Approach	
	EER	Accuracy	EER	Accuracy
Anger	20.99	79.01	23.67	76.33
Lombard	14.61	85.39	12.61	87.39
Neutral	18.67	81.33	18.67	81.33
Question	27.68	72.32	10.68	89.32
Average	20.48	79.52	16.40	83.60

TABLE II. Emotion Recognition Performance (in %) using reduced dimension MFCC/LPCC

Emotion	MFCC Approach		LPCC Approach	
	EER	Accuracy	EER	Accuracy
Anger	10.27	89.73	15.00	85.00
Lombard	28.61	71.39	20.00	80.00
Neutral	30.00	70.00	23.89	76.11
Question	31.94	68.06	26.39	73.61
Average	25.20	74.80	21.32	78.68

dimension for emotion classification. The experimental results showed that the performance for reduced dimensions is less than 4-5% as compared to that of full dimensions. It has been observed from table II, the accuracy of reduced MFCC based system is 74.80% (EER 25.20%) and that for LPCC based system is 78.68% (EER 21.32%).

The results obtained from the tables suggest that the acoustic feature derived using lower order coefficients (2nd, 3rd and 4th) carry significant information about the emotion in a spoken utterance. It is apparent that the emotional characteristic may contain in the low frequency regions of the spoken utterances. Further, it has been observed that LPCCs provides a relatively better performance as compared to MFCCs used as acoustic feature to represent the vocal tract transfer function. This may be due to more emphasis given by LPCC feature to the mid frequency bands of the spoken utterance.

The experiment has been carried out on sentence level utterances along with the word level utterances and the results were obtained as shown in table III. The experiment results obtained for sentence level utterances are similar to that of the results obtained with IITKGP-SESC [7].

The results suggest that emotion classification system designed using word level utterances provide better performance as compared to system designed using sentence level utterances. This may due to exploration of vocal tract characteristics during designing the system. However, the performance may be different when other level of acoustic features is used as complementary cues to represent the emotion.

The lower order acoustic features contain significant information about a particular emotion. The new feature set derived using the subspace has lower dimensionality. Therefore, the computational complexity as well as storage requirement may have significantly reduced. This may be helpful in case of implementation of a real time system having large databases.

TABLE III. PERFORMANCE OF SYSTEM BASED ON SENTENCE LEVEL UTTERANCES

Emotion	Emotion Recognition Performance (in %)	
	LPCCs	MFCCs
Anger	53.00	40.00
Lombard	63.00	50.00
Neutral	63.00	60.00
Question	53.00	47.00
Average	58.00	49.25

V. SUMMARY AND CONCLUSIONS

The LPCCs and MFCCs are two of the widely used acoustic features used for emotion recognition. This paper compares the relative performance of the two acoustic features for emotion classification on SUSE database. It exploits the full dimension as well as reduced dimension of both the acoustic features and tries to evaluate the relative performance of reduced dimensional acoustic feature. It has been observed that the lower order coefficients carry significant emotion specific information as compared to the higher order coefficients. It may also be concluded that LPCCs serve as a better acoustic feature as compared to MFCC for designing an emotion classification system. The study also suggests that the system designed using word level utterances provides better performance as that of sentence level utterances, using vocal tract characteristics alone. Rigorous analysis may be made based on this initial study to analyze the significance of acoustic dimensions with respect to emotion. The system may be designed with reduced acoustic dimension complemented by other level of acoustic features (such as sub-segmental and supra-segmental) to provide a robust and satisfactory performance. The high level acoustic features may also be explored to enhance the performance of emotion classification system.

REFERENCES

[1] A. Kolakowska, A. Landowska, M. Szwoch, W.Szwoch and M. R. Wróbel, “Emotion Recognition and its Applications,” in *Human-Computer Systems Interaction: Backgrounds and Applications 3, Advances in Intelligent Systems and Computing*, vol. 300, Springer Switzerland, 2014, pp. 51-62.

[2] S. Casale, A. Russo, G. Scebba, and S. Serrano, "Speech emotion classification using machine learning algorithms," in *IEEE International Conference on Semantic Computing*, Santa Clara, CA, Aug 2008, pp. 158-165.

[3] Z. Xiao, E. Dellandrea, W. Dou and L. Chen, “Feature Extraction and Selection for Speech Emotion Classification,” in *IEEE Conf. on Advanced Video and Signal Based Surveillance*, Como, Italy, Sep 2005, pp. 411-416.

[4] S. Shukla, S. R. M. Prasanna and S. Dandapat, “Speech Recognition under Stress Condition,” in *15th National Conference on Communications*, IIT Guwahati, Jan 2009, pp. 299-302.

[5] S. G. Koolagudi, S. Maity, V. A. Kumar, S. Chakrabarti and K. S. Rao, “IITKGP-SESC: Speech Database for Emotion Analysis,” in *Contemporary Computing, Communications in Computer and Information Science,*vol. 40, Springer Berlin Heidelberg, Aug 2009, pp. 485-492.

[6] S. G. Koolagudi, N. Kumar and K. S. Rao, “Speech emotion recognition using segmental level prosodic level analysis,” in *ICDeCom, Int. Conf. on Devices and Communications,* Mesra, Feb 2011, pp. 1-5.

[7] S. G. Koolagudi and S. R. Krothpalli, “Emotion recognition from speech using sub-syllabic and pitch synchronous spectral features,” in *International Journal of Speech Technology*, vol. 15, issue 4, Springer US, Dec 2012, pp. 495-511.

[8] G. S. Raja and S. Dandapat, “Speaker recognition under stressed condition,” in *International Journal of Speech Technology*, vol. 13, No. 3, Springer US, Sept 2010, pp. 141-161.

[9] N. Sato and Y. Obuchi, “Emotion Recognition using Mel-Frequency Cepstral Coefficients,” in *Information and Media Technologies*, vol. 2, No. 3, 2007, pp. 835-848.

[10] D. R. Porcheron and M. F. Zanuy, “Speaker Recognition with a MLP Classifier and LPCC Codebook,” in *Proc. of ICASSP,* vol. 2, Mar 1999, Phoenix, AZ, pp. 1005-1008.

SVD and Chaotic System Based Watermarking Approach for Recovering Crime Scene Image

Md. Moniruzzaman,[1,*] Md. Abul Kayum Hawlader,[1] Md. Foisal Hossain,[1] and M. A. Rashid[2]

[1]Department of Electronics and Communication Engineering, Khulna University of Engineering and Technology (KUET)
Khulna-9203, Bangladesh

[2]Faculty of Design Arts and Engineering Technology, University Sultan Zainal Abidin (UniSZA)
Kuala Terengganu, Malaysia

[*]mdzamankuet@gmail.com

Abstract—**This paper presents a watermarking approach for recovering crime scene image when the crime scene image has been tampered, based on chaotic system and singular value decomposition (SVD). At first, authentication information has been embedded into the region of information (ROI) of the scrambled crime scene image. The authentication information has been generated by using logistic map, while the scrambled crime scene image has been obtained by two dimensional Arnold's cat map. After then, ROI has been compressed by discrete cosine transform (DCT) and the compressed ROI has been embedded into region of background (ROB) by using singular value decomposition. In extraction stage, the authentication information has been extracted from the ROI of the watermarked crime scene image. If the extracted authentication information is not same as the embedded information, ROI has been extracted from the ROB. The results of the proposed method has been evaluated by using peak signal to noise ratio (PSNR) and mean squared error (MSE) performance parameters. At the same time, the proposed watermarking approach has also been compared with other methods. The proposed method gives better values of MSE and PSNR than the compared methods.**

Index Terms—**Image watermarking, Logistic map, Arnold's cat map, Singular value decomposition (SVD), PSNR, MSE.**

I. Introduction

The authenticity and integrity of an image can be ensured by using digital image watermarking scheme [1]. In many real life applications need exact authentication, such as: for diagnosing the medical images, for detecting the criminals from crime scene images.

Crime scene images consist of two parts- region of information (ROI) and region of background (ROB). For crime images, exact authentication is required in region of information. Any alternation of ROI may cause suspecting a person other than actual criminal. Watermarking is one of the possible solutions to give exact authentication in crime scene images by embedding authentication information, which is usually imperceptible.

The embedded information can be spatial domain based or transform domain based. In spatial domain based watermarking scheme, the authentication information is directly embedded into the pixels gray intensities. The authentication information has been embedded into the coefficients of transformed image in transform domain based watermarking [2]-[4].

The watermark can be classified into two ways as fragile and robust watermark according to its nature. Further, fragile watermark can also be classified into two ways as complete fragile and semi-fragile watermarks. Complete fragile watermarks are used for image tamper detection [5], [6]. Semi-fragile watermarks are commonly used for detecting the modified areas which has been tampered by the attackers and recovering the remaining watermarks from the undamaged areas [7]. Robust watermarks are difficult to remove by the attackers and shows robustness against different types of attacks [8].

In recent years, various watermarking methods have been proposed for recovering the tampered areas, which are applicable in crime detection [9], [10]. In these methods, the input image has been divided into two regions, ROI and ROB. The information of ROI has been embedded into ROB. Therefore, if ROI has been tampered, it can be recovered from ROB.

In this paper, SVD and chaotic system based watermarking approach has been proposed for recovering crime scene image from tampered image. The authentication information generated by logistic map has been embedded into scrambled ROI. After that, compressed ROI has been embedded into ROB. ROI has been extracted in the extraction stage, when the extracted authentication information is not same as the embedded authentication information.

The rest of this paper is organized as follows. The proposed watermarking approach with details has been explained in section II. Section III depicts the experimental results and comparison. Finally, section IV concludes the paper.

II. Proposed Watermarking Approach

In the proposed watermarking approach, gray scale crime scene image has been considered, which consists of region of information (ROI) and region of background (ROB). The proposed scheme designed the algorithm in four stages. In first stage, the authentication information has been embedded into ROI of the image. After that, ROI has been compressed in the second stage. After completing second stage, the compressed ROI has been embedded into ROB of the image in the third stage. Finally, in fourth stage or extraction stage, the ROI has been extracted from the ROB. ROI extraction only occurs when the extracted authentication information is not same as the embedded authentication information.

A. Embedding Authentication Information into ROI

In this stage, the authentication information has been embedded into ROI of crime scene image, depicted in Algorithm 1. The proposed model generates random number sequences as authentication information by using logistic map,

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

which is one of the simplest chaotic maps. The map generates random number sequences or authentication information for the proposed scheme as follows-

$$x_{n+1} = \mu x_n (1 - x_n) \tag{1}$$

where, $0<\mu\leq 4$. The map is highly sensitive to the initial conditions. Different initial values generate different random number sequences, which are shown in Fig. 1.

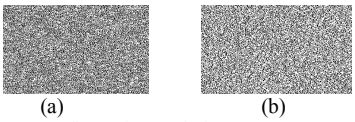
The generated authentication information has been embedded into ROI image, but not in a consecutive way. The generated authentication information has been embedded into the least significant bit (LSB) plane of scrambled ROI image. Therefore, the attacker will not be able to forge the authentication information as the information has been embedded into the scrambled ROI image. ROI image has been scrambled by using 2D Arnold's cat map to improve the security of embedding the authentication information. Without changing pixels gray level intensities, 2D Arnold's cat map shuffles the pixels positions according to the initial conditions and number of iterations. Arnold's cat map shuffles the pixels positions as follows-

$$\begin{bmatrix} x_{n+1} \\ y_{n+1} \end{bmatrix} = \begin{bmatrix} 1 & a \\ b & ab+1 \end{bmatrix} \begin{bmatrix} x_n \\ y_n \end{bmatrix} \bmod N = A \begin{bmatrix} x_n \\ y_n \end{bmatrix} \bmod N \tag{2}$$

where, N is the dimension of the image, a and b are the control parameters and $det\ (A) = 1$. The map is periodic in nature, which means pixels back to their original positions after performing several iterations. The periodic phenomenon of Arnold's cat is shown in Fig. 2. After embedding authentication information into the scrambled ROI image, inverse cat map has been applied to the scrambled ROI image to back the pixels to their original positions.

B. Region of Information (ROI) Compression Using DCT

The ROI of crime scene image has been compressed using discrete cosine transform (DCT), described in Algorithm 2. JPEG image compression standard uses DCT which is a fast transform, widely used and robust method for image compression. At first, ROI image has been divided into several $m_1 \times n_1$ blocks. After then, DCT has been applied to each block to obtain DCT coefficients. The upper half part of DCT coefficients of each $m_1 \times n_1$ block has been taken and the lower half part of each $m_1 \times n_1$ block has been made to zero. After performing this action, inverse discrete cosine transform (IDCT) has been applied to each block to get the compressed ROI crime scene image.



(a) (b)

Fig. 1 Two random configurations of size 128×128, generated by logistic map, initial conditions of (a) μ=3.90 and x(1)=0.4545; (b) μ=3.93 and x(1)=0.4545.

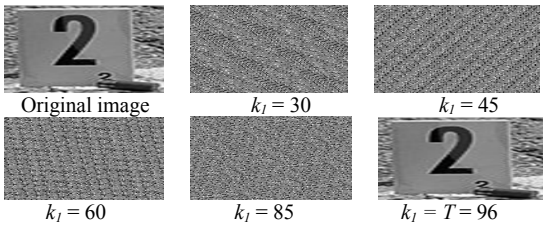


Fig. 2 Periodic phenomenon of Arnold's cat map, shows period (T) at $k_1 = 96$ of 128×128 image.

Algorithm 1 Embedding authentication information

1. Input: ROI of crime scene image
2. Output: ROI crime scene image with authentication information.
3. Scramble ROI using 2D Arnold's cat map.
4. Generate random number sequences as authentication information (*AI*) using logistic map.
5. Embed authentication information into LSB of scrambled ROI image.
6. Apply inverse cat map and obtain ROI image with authentication information.

Algorithm 2 ROI compression

1. Input: ROI image.
2. Output: Compressed ROI image.
3. Extract ROI from the crime scene image.
4. Apply DCT to ROI image to obtain its DCT coefficients.
5. Make the lower energy of DCT coefficients to zero
6. Apply IDCT to obtain compressed ROI image.

C. Embedding ROI into ROB of Crime Scene Image

The information of compressed ROI has been embedded into the ROB of crime scene image, described in Algorithm 3. For this process, singular value decomposition (SVD) [11] has been applied to the ROB of crime scene image. SVD decomposes ROB in orthogonal matrix U_{ROB}, diagonal matrix S_{ROB} and transpose of orthogonal matrix V_{ROB}. The singular values diagonal matrix (S_{ROB}) has been used as a marking space in which ROI information has been embedded. On the other hand, SVD has also been applied to ROI to decompose ROI in U_{ROI}, S_{ROI} and V_{ROI}. After applying SVD to ROI, S_{ROI} has been converted to real numbers in [0 255] domain as follow-

$$\beta = 255 / S^1_{ROI} \tag{3}$$

$$S^i_{c_ROI} = S^i_{ROI} \times \beta \tag{4}$$

where, S_{c_ROI} represents converted S_{ROI}, r is the rank of the matrix S_{ROI} and β is the converting coefficient. S_{ROI} has been converted to S_{c_ROI} to increase the perceptual similarity of watermarked image.

After then, the S_{c_ROI} has been embedded with S_{ROB} as follows-

$$S'_{ROB} = S_{ROB} + \left(\alpha \times k \times S_{c_ROI}\right) \tag{5}$$

where, α and k are the embedding factors. Finally, inverse SVD has been applied to ROB image to obtain watermarked ROB image in which the information of compressed ROI has been embedded. Therefore, watermarked crime scene image consists of authentication information into ROI and the information of ROI into ROB, which are invisible to human eyes.

D. Recovering ROI from ROB of Crime Scene Image

At first, the authentication information has been extracted from the scrambled ROI image. If the ROI image has been tampered, the extracted authentication information will not be same as the embedded information. In that case, the tampered areas can be located by applying logical Exclusive-or (X-or) between embedded and extracted authentication information.

At the same time, tampered areas can also be recovered from ROB of watermarked crime scene image.

For recovering the embedded information of ROI from the watermarked ROB, SVD has been applied to the watermarked ROB to obtain singular values diagonal matrix (S'_{ROB}). The singular values diagonal matrix which was embedded with singular values diagonal matrix of original ROB (S_{ROB}), can be extracted by-

$$(S_{c_ROI})^{ext} = \left(S'_{ROB} - S_{ROB}\right)/\left(\alpha \times k\right) \tag{6}$$

After extraction, $(S_{c_ROI})^{ext}$ has been converted to the original range to obtain the original singular values diagonal matrix (S_{ROI}) of ROI image, which has been obtained by-

$$S_{ROI}^{ext} = (S_{c_ROI}) / \beta \tag{7}$$

Finally, the ROI of crime scene image can be recovered by applying inverse SVD, which can be described by-

$$ROI' = U_{ROI} S_{ROI}^{ext} (V_{ROI})^T \tag{8}$$

where, ROI' is the recovered ROI from the watermarked ROB image, when the image has been tampered by the attackers.

III. Experimental Results and Comparison

In experimental evaluation, various crime scene images of different sizes have been taken as input images and proposed scheme has been applied on them.

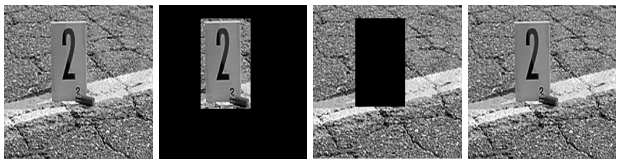
The proposed approach firstly uses Algorithm 1 to embed the authentication information into the region of information of the crime scene image. Using Algorithm 2, ROI image has been compressed by discrete cosine transform (DCT). With the help of singular value decomposition (SVD), compressed ROI has been embedded into ROB by using Algorithm 3.

Algorithm 3 Embedding information of ROI into ROB

1. Input: Crime scene image I, ROB, compressed ROI image.
2. Output: Watermarked crime scene image I_w.
3. Apply SVD to ROB to obtain singular values diagonal matrix (S_{ROB}).
4. Apply SVD to compressed ROI to obtain S_{ROI}.
5. Convert S_{ROI} into the range of [0-255].
6. $S'_{ROB} = S_{ROB} + (\alpha \times k \times S_{c_ROI})$, where S_{c_ROI} represents converted S_{ROI}, α and k are the scaling factors.
7. Apply inverse SVD on ROB to obtain I_w, in which ROI information is embedded into ROB.

Algorithm 4 Recovering information of ROI from ROB

1. Input: Watermarked crime scene image I_w.
2. Output: Recovered ROI from ROB.
3. Extract authentication information (AI') from ROI.
4. **if** $AI' == AI$
5. Do nothing
6. else
7. Apply SVD to ROB to obtain S'_{ROB}.
8. $S_{c_ROI} = (S'_{ROB} \square S_{ROB}) / (\alpha \times k)$.
9. Convert S_{c_ROI} to original range to get S_{ROI}.
10. Obtain recovered ROI by applying inverse SVD.
11. **end**

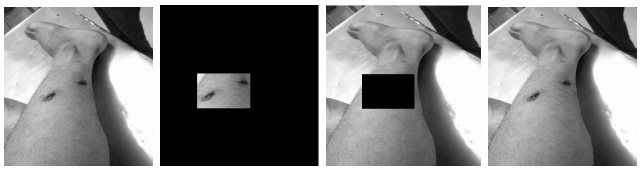


(a) (b) (c) (d)

Fig. 3 Crime scene image: (a) original image, (b) ROI of the image, (c) ROB of the image, (d) watermarked crime scene image

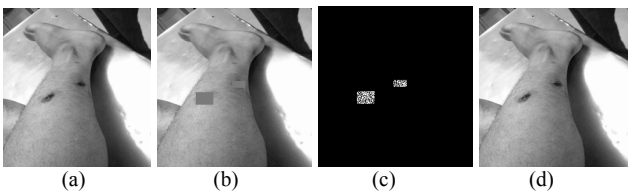
(a) (b) (c) (d)

Fig. 4 (a) watermarked crime scene image, (b) tampered image, (c) the detected tampered areas, (d) the recovered image.



(a) (b) (c) (d)

Fig. 5 Crime scene image: (a) original image, (b) ROI of the image, (c) ROB of the image, (d) watermarked crime scene image



(a) (b) (c) (d)

Fig. 6 (a) watermarked crime scene image, (b) tampered image, (c) the detected tampered areas, (d) the recovered image.

Finally, by applying Algorithm 4, the ROI has been recovered from ROB in the extraction stage.

Fig. 3(a) shows a crime scene image of evidence marker with bullet shell. The ROI of this crime scene image is shown in Fig. 3(b), while Fig. 3(c) shows the ROB of the image. Fig. 3(d) shows the watermarked crime scene image in which the authentication information has been embedded into ROI and the DCT compressed ROI has been embedded into ROB image. The bullet shell that was shown in Fig. 3(a) has been removed from ROI, which is shown in Fig. 4(b). Fig. 4(c) depicts the located tampered areas and Fig. 4(d) shows the recovered bullet shell have been obtained by the proposed approach.

Fig. 5 shows another crime scene image which has been tampered and recovered subsequently in Fig. 6. Fig. 5(b) shows the ROI of crime scene image of Fig. 5(a), while ROB is shown in Fig. 5(c). The watermarked image of Fig. 5(a) is shown in Fig. 5(d). In Fig. 6(b), the ROI of Fig. 5(a) has been tampered by the third parties, in which the spot of attacked leg by bullet has been removed. The tampered areas have been located by the proposed scheme, which is shown in Fig. 6(c). The tampered image has also been recovered by the proposed approach, as shown in Fig. 6(d).

The performance of the proposed approach has been evaluated by using two performance parameters- (1) mean square error (MSE) and (2) peak signal to noise ratio (PSNR).

A. Mean Squared Error (MSE)

MSE is used to test the similarity between two images. The MSE between original image and watermarked image can be obtained by-

$$MSE = \frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}\left(I(x,y)-I'(x,y)\right)^2 \quad (9)$$

where, I and I' are the original image and watermarked image, respectively, having resolution of $M \times N$ pixels.

B. Peak Signal to Noise Ratio (PSNR)

PSNR is usually used to measure the quality of the image which is attacked by low quality of noise. In the proposed watermarking scheme, PSNR has been used to measure the quality of the watermarked image and also the quality of the recovered image. PSNR can be obtained by-

$$PSNR = 10\log_{10}\frac{\max^2}{MSE} \quad (10)$$

where, max is the maximum gray level of the image. For 8-bit image $max = 255$.

TABLE I shows MSE and PSNR values of watermarked image of the proposed watermarking approach with different values of embedding factor. For embedding ROI into ROB using (5) two embedding factors (k and α) have been used. Different values of α and a fixed value of k (k = 0.01) have been used in the proposed watermarking scheme for embedding ROI into ROB. Investigating TABLE I, it can be seen that with the increasing values of α, values of MSE increases and Values of PSNR decreases. But for all values of embedding factor, the proposed approach maintains satisfactory results of MSE and PSNR, which proves that the watermarked image maintains perceptual similarity with good quality. Fig. 7 graphically shows the relationship between α and PSNR values of the proposed method for both given images.

TABLE II depicts the performance comparison between the proposed watermarking method and other two methods for bullet shell image. MSE and PSNR values of watermarked image and recovered image for proposed method and other methods have been shown in TABLE II. From TABLE II, it can be said that the proposed method dominates the other methods in MSE and PSNR values.

TABLE I
MSE AND PSNR VALUES OF WATERMARKED IMAGE

Embedding factor (α)	Bullet Shell		Attacked Leg	
	MSE	PSNR	MSE	PSNR
10	0.59	50.44	0.49	51.23
20	1.17	47.43	0.93	48.46
30	2.66	43.88	2.52	44.12
40	4.38	41.72	4.12	41.98
50	6.83	39.78	6.35	40.10
60	9.71	38.26	8.34	38.92
70	13.15	36.94	11.22	37.63
80	17.10	35.80	15.42	36.25

TABLE II
COMPARISON RESULTS OF MSE AND PSNR VALUES OF WATERMARKED IMAGE AND RECOVERED IMAGE

Methods	Watermarked Image		Recovered Image	
	MSE	PSNR	MSE	PSNR
Method [9]	6.94	39.72	10.52	37.91
Method [10]	4.51	41.59	9.91	38.17
Proposed Method	0.59	50.44	1.17	47.46

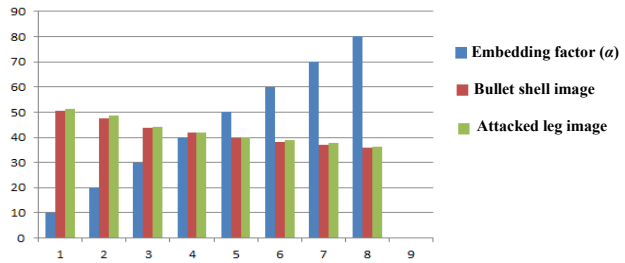


Fig. 7 (a) Relationship between α and PSNR values of proposed method for both images.

IV. CONCLUSION

This paper proposed a watermarking approach for recovering crime scene image from tampered image. Using chaotic system, authentication information has been embedded into ROI and the information of ROI has been embedded into ROB. SVD has been used for embedding process to increase the perceptual similarity. High values of PSNR have been obtained both for watermarked image and recovered image of the proposed watermarking approach. In addition, the presented method has also been compared with other two state-of-arts methods. The proposed scheme dominates the other methods in MSE and PSNR values, which proves that the proposed watermarking approach maintains good qualities of watermarked image and recovered image.

REFERENCES

[1] Li, C.T. and Si, H. "Wavelet-based fragile watermarking scheme for image authentication," J. Electron. Imaging, vol. 16, Issue 1, pp. 17-22, 2007.

[2] G. Sun, Y. Yu, "DWT Based Watermarking Algorithm of Color Images," 2nd IEEE Conf. on Industrial Elec. and App., pp. 1823-1826, May 2007.

[3] N. Kashyap and G. R. Sinha, "Image Watermarking Using 3-Level Discrete Wavelet Transform," International Journal of Modern Education and Computer Science, vol. 3, pp. 50-56, 2012.

[4] W.M. Yang, J. Zheng, "A Watermarking Algorithm Based on Wavelet and Cosine Transform for Color Image," First Int. Workshop on Edu. Tech. and Comp. Sc., pp. 899-903, March 2009.

[5] Monzoy, M.; Salinas, M.; Nakano, M. & Pérez, H., "Fragile Watermarking for Color Image Authentication," 4th Int. Conf. Electrical and Electronic Engineering (ICEEE 2007), pp. 157-160, 2007.

[6] Shao-Hui Liu, Hong-Xun Yao, Wen Gao, Yong-Liang Liu, "An image fragile watermark scheme based on chaotic image pattern and pixel-pairs," Applied Mathematics and Computation, vol. 185, pp. 869–882, 2007.

[7] K. Maeno, Q. Sun, S. F. Chang and M. Suto, "New semi-fragile image authentication watermarking techniques using random bias and nonuniform quantization," IEEE Trans. on multimedia, vol. 8, pp. 32-45, 2006.

[8] Chandra M. B. and Srinivas K. S., "Robust Multiple Image Watermarking Scheme using Discrete Cosine Transform with Multiple Descriptions," International Journal of Computer Theory and Engineering, vol. 1, no. 5, pp. 1793–8201, 2009.

[9] J. H. Chen, W. Y. Chen and C. H. Chen, "Images tamper self-detection and self-recovery using SPIHT techniques," International Journal of Innovative Computing, Information and Control, vol. 9, no. 9, September 2013.

[10] S. D. Lin, J. H. Lin and C. Y. Chen, "A ROI based semi-fragile watermarking for image tamper detection and recovery," International Journal of Innovative Computing, Information and Control, vol. 7, no. 12, December 2011.

[11] N.V. Dharwadkar, B.B. Amberker, A. Gorai, "Non-blind watermarking scheme for color images in RGB space using DWT-SVD," Int. Conf. on Communications and Signal Processing (ICCSP)., pp. 489-493, February 2011.

A Blind Audio Watermarking Method Based on Lifting Wavelet Transform and QR Decomposition

Pranab Kumar Dhar
Department of Computer Science and Engineering
Chittagong University of Engineering and Technology (CUET)
Chittagong-4349, Bangladesh
pranabdhar81@gmail.com

Abstract— **This paper presents a blind audio watermarking method based on lifting wavelet transform (LWT) and QR decomposition (QRD) for audio copyright protection. In our proposed method, initially the original audio is segmented into non-overlapping frames. Watermark information is embedded into the largest element of the upper triangular matrix obtained from the low frequency LWT coefficients of each frame. A blind watermark detection technique is introduced to identify the embedded watermark under various attacks. Simulation results indicate that the proposed watermarking method is highly robust against different attacks. In addition, it has high data payload and provides good imperceptible watermarked sounds. Moreover, it shows superior performance than the state-of-the-art watermarking methods reported recently.**

Index Terms— **copyright protection; lifting wavelet transform; QR decomposition.**

I. Introduction

The recent development of digital world has greatly facilitated the transmission and manipulation of multimedia data such as text, image, audio, and video. Easy access and duplication of multimedia data have led to serious problem for copyright protection. Digital watermarking has drawn extensive attention for protecting multimedia data from unauthorized copying. Therefore, media owners can use this technique to insert some information into their digital contents to show authenticity and ownership. This technique has several applications such as content authentication, copyright protection, data indexing, broadcast monitoring, fingerprinting, medical safety, and so on. In recent years, a significant number of audio watermarking techniques have been reported. A comprehensive survey on audio watermarking can be found in [1]. Most audio watermarking methods utilize either a time domain [2]-[3] or a transform domain such as discrete wavelet transform (DWT) [4]-[5], discrete cosine transform (DCT) [6], fast Fourier transform (FFT) [7], and discrete fractional sine transform (DFST) [8]. Time domain methods are very efficient and easy to implement, however, transform domain methods can provide high robustness. Conventional wavelet transform provides good results for its multi resolution characteristics and perfect reconstruction. However, it is mainly calculated by convolution operation, resulting in high computation. In addition, the generated floating numbers increase storage requirements. As a result, the lifting wavelet transform (LWT) [9] is designed to increase the efficiency and it is now used in audio watermarking [10]. Moreover, some other techniques such singular value decomposition (SVD) [11]-[14], empirical mode decomposition (EMD) [15], patchwork algorithm [16], and audio histogram techniques [17]-[18] are becoming popular in audio watermarking field.

In this paper, a blind audio watermarking method based on LWT and QR decomposition (QRD) is proposed. The main features of the proposed method include (*i*) it utilizes the LWT and QRD jointly, (*ii*) it uses Bernoulli map, containing the chaotic characteristic to enhance the confidentiality of the proposed method, (*iii*) watermark extraction process is blind, (i*v*) subjective and objective evaluations reveal that the proposed method maintains a high audio quality, and (*v*) it achieves a good trade-off among imperceptibility, robustness, and data payload. Simulation results demonstrate that the proposed watermarking method shows high robustness against various attacks such as noise addition, cropping, re-sampling, re-quantization, and MP3 compression. Moreover, it outperforms the state-of-the-art methods [4]-[5], [7], [11], [13], [15]-[18] in terms of imperceptibility, robustness, and data payload. The data payload of the proposed method is 172.39 bps which is relatively higher than that of the state-of-the-art methods.

The rest of this paper is organized as follows. Section II discusses the background information including LWT and QRD. Section III introduces our proposed watermarking method including watermark pre-processing, watermark embedding and detection processes. Section IV presents the experimental results of the proposed method. Lastly, Section V concludes this paper.

II. Background Information

A. Lifting Wavelet Transform

The LWT is designed to reduce the computation time and memory requirement. It has several unique properties in comparison with traditional wavelet: (*i*) it allows an in-place implementation of the fast wavelet transform thus it can be calculated more efficiently and needs less memory space, (*ii*) it is particularly easy to build non-linear wavelet transforms, (*iii*) it has the frequency localization features which overcome the weakness of the traditional wavelet transform. The main principle of the lifting wavelet is to construct a new wavelet with better characteristics based on a simple wavelet. Lifting wavelet scheme consists of three steps: split/merge, prediction, and update. The detailed reasoning and proof of the lifting scheme is given in [9].

B. QR Decomposition

In QRD, the orthogonal-triangular decomposition of a matrix is performed with the following definition:

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

$$[Q,R]=qr(A) \tag{1}$$

where R is an $m\times m$ upper triangular matrix and Q is an $m\times m$ unitary matrix, so that $A=QR$. The columns of Q are obtained from the columns of A by Gram-Schmidt orthogonalization process. One of the good properties of R matrix is that when columns of A have correlation with each other, the absolute values of the elements in the first row of R matrix are probably greater than those in other rows. This property can be utilized for embedding watermark in audio signal.

III. Proposed Watermarking Method

Let $X=\{x(n),\ 1\le n\le L\}$ be an original audio signal with L samples, $W=\{w(i),\ 1\le i\le I\}$ be a binary watermark sequence to be embedded into the original audio signal.

A. Watermark Preprocessing

To enhance the security of the proposed method, watermark sequence should be preprocessed first. The proposed method utilizes a Gaussian map, containing the chaotic characteristics to encrypt the binary watermark sequence for enhancing the confidentiality of the proposed method. It can be defined as follows:

$$y(i+1)=\exp\left(-a(y(i))^2\right)+b \tag{2}$$

where $y(1)\in(0,1)$, a and b are real parameters (map's initial condition). Then $y(i)$ is mapped using the following rule:

$$z(i)=\begin{cases}1 & if\ y(i)>T\\ 0 & otherwise\end{cases} \tag{3}$$

where T is a predefined threshold. Finally $w(i)$ is encrypted by $z(i)$ with the following rule:

$$u(i)=z(i)\oplus w(i),\quad 1\le i\le I \tag{4}$$

where $\oplus$ is the exclusive-or (XOR) operation and $u(i)$ is the encrypted watermark sequence. After applying this chaotic encryption technique, the original watermark is encrypted and can not be found by random search. In this study, the value of $y(1)$, a, and b are used as secret key K.

B. Watermark Embedding Process

The proposed watermark embedding process is described as follows:

1) The original audio signal X is first segmented into non-overlapping frames $F=\{F_1, F_2, F_3,.....,F_I\}$.

2) A two-level LWT is performed on each frame F_i. This operation produces three sets of coefficients D_1, D_2, and A_2, where D_1, D_2 and A_2 represent the detailed and approximate coefficients, respectively.

3) The approximate coefficients A_2 of each frame F_i are rearranged into an $N\times N$ square matrix H_i. This is done by dividing the coefficient set into N segments with N coefficients.

4) QRD is performed to decompose each matrix H_i into two matrices: upper triangular matrix R_i and unitary matrix Q_i. The QRD operation is represented as follows:

$$[Q_i,R_i]=qr(H_i) \tag{5}$$

5) In order to guarantee the robustness and transparency, the proposed method embeds watermark bits into the largest element of each matrix R_i using a quantization function. Let $Y_i=round\left(\frac{R_{i(\max)}}{D}\right)$, where D is a predefined quantization coefficient and $R_{i(\max)}$ is the largest element of each R_i. The embedding equation is given as follows:

$$Y_i'=\begin{cases}Y_i+C-(Y_i \bmod M), & if\ u(i)=1\\ Y_i+C-((Y_i+C) \bmod M), & if\ u(i)=0\end{cases} \tag{6}$$

where $M=2C$, C is an integer, mod is the modulo operation.

6) The modified largest element $R'_{i(\max)}$ of each matrix R_i is calculated using the following equation:

$$R'_{i(\max)}=Y_i'D \tag{7}$$

where R'_i is the upper triangular matrix.

7) Reinsert each modified largest element $R'_{i(\max)}$ into each matrix R_i and inverse QRD is applied to obtain the modified matrix H_i' which is given by

$$H_i'=Q_iR_i' \tag{8}$$

Each matrix H_i' is then reshaped to create the modified approximate coefficients A'_2 of each frame F_i by performing the inverse operation of step 3.

8) After substituting the modified approximate coefficients A'_2 for A_2, a two-level inverse LWT is performed to obtain the watermarked audio frame F_i'.

9) Finally, all watermarked frames are concatenated to calculate the watermarked audio signal X'.

C. Watermark Detection Process

The proposed watermark detection process is described as follows which does not need the original audio signal to extract the watermark.

1) A two-level LWT is performed on each frame F_i^* of the attacked watermarked audio signal.

2) The approximate coefficients A_2^* of each frame F_i are rearranged into an $N\times N$ square matrix H_i^*.

3) QRD is performed on each R_i^* and select the largest element $R^*_{i(\max)}$ from each matrix R_i^* of the attacked watermarked audio frame.

4) Calculate Y_i^* of each $R^*_{i(\max)}$.

5) Encrypted watermark sequence is extracted as follows:

$$u^*(i)=\begin{cases}1 & if\ (Y_i^* \bmod M)=1\\ 0 & otherwise\end{cases} \tag{9}$$

6) Perform chaotic decryption using the secret key K to find the binary watermark sequence with the following rule:

$$w^*(i)=z(i)\oplus u^*(i) \tag{10}$$

IV. Simulation Results

In this section, several experiments were conducted on four different types of 16 bit mono audio signals (Pop, Jazz, Classical, and Speech) sampled at 44.1 kHz to demonstrate the performance of the proposed method. Each audio file contains 262,144 samples (duration 5.94 sec). Each audio signal is divided into frames of size 256 samples. Thus, the total number of frames is 1024. In each frame of audio signal, we embed one bit watermark data. In this study, the selected value for $y(1)$, a, b, T and M are 0.4, 5.90, -0.39, 0.25, and 1, respectively. These parameters have been selected in order to achieve a good trade-off among the conflicting requirements of imperceptibility, robustness, and data payload.

A. Imperceptibility Test

The imperceptibility of the watermarked audio signal is evaluated by using two ways: (*i*) subjective listening test (*ii*) objective test.

1) Subjective Listening Test

An informal subjective listening test was performed to evaluate the perceptual quality of the watermarked audio. Ten listener involved in this listening test and were asked to classify the differences between them, using a five point subjective difference grade (SDG) ranges from 5.0 to 1.0 (imperceptible to very annoying) as shown in Table I. The average SDG (i.e., mean opinion) scores for different watermarked sounds using the proposed method are shown in Table II. From the test results, we observed that the average mean opinion score (MOS) of all watermarked sound using the proposed method is 4.90, indicating that original and watermarked audio signals are perceptually similar.

TABLE I. SUBJECTIVE AND OBJECTIVE DIFFERENCE GRADES

SDG	ODG	Description	Quality
5	0	Imperceptible	Excellent
4	-1	Perceptible, but not annoying	Good
3	-2	Slightly annoying	Fair
2	-3	Annoying	Poor
1	-4	Very annoying	Bad

TABLE II. MOS, ODG AND SNR RESULTS FOR DIFFERENT WATERMARKED SOUNDS

Types of Signal	MOS	ODG	SNR
Pop	4.9	-0.74	34.65
Jazz	5.0	-0.56	35.48
Classical	4.9	-0.79	34.93
Speech	4.8	-0.83	33.26
Average	**4.9**	**-0.73**	**34.58**

2) Objective Test

Objective test was conducted using the objective difference grade (ODG) which is one of the output values obtained from the perceptual evaluation of audio quality (PEAQ) measurement technique specified in ITU-R BS.1387 (International Telecommunication Union-Radio-communication Sector) standard [19]. It corresponds to the subjective grade used in human based audio tests. The ODG ranges from 0.0 to -4.0 (imperceptible to very annoying) as shown in Table I. The objective quality of the watermarked audio signals is calculated in terms of ODG and shown in Table II. We observed that the average ODG value is -0.73, indicating that original and watermarked audio signals are perceptually indistinguishable.

Objective evaluation was also done by calculating the SNR which is given by

$$SNR = 10\log_{10}\frac{\sum_{i=1}^{L} x^2(n)}{\sum_{i=1}^{L}\left[x(n)-x^*(n)\right]^2} \quad (11)$$

where $x(n)$ and $x^*(n)$ are the original and watermarked audio signals in time domain, respectively. According to the International Federation of the Phonographic Industry (IFPI) standard [13], audio watermarking should be imperceptible when SNR is over 20 dB. After embedding watermark information, the SNRs of the watermarked audio signals using the proposed method are above 20 dB, shown in Table II, which satisfied the IFPI standard.

The proposed method was compared with the several recent methods in terms of the SNR and MOS. This comparison is based on the reported results in [4]-[5], [7], [11], [15]-[16] as shown in Table III. From the comparison of results, it is seen that our proposed method outperforms the recent watermarking methods in terms of SNR and MOS, indicating the high transparency of the watermarked audio signals.

TABLE III. SNR AND MOS COMPARISON BETWEEN THE PROPOSED AND SEVERAL RECENT METHODS

Reference	Algorithm	SNR	MOS
[5]	DWT-based energy proportion	17.95	4.15
[4]	Wavelet-based entropy	22.46	4.38
[15]	EMD	24.12	--
[16]	Patchwork	24.95	4.67
[7]	FFT amplitude modification	25.70	--
[11]	DCT-SVD	32.53	4.71
Proposed	**LWT-QRD**	**34.58**	**4.90**

B. Robustness Test

Various signal processing attacks summarized in Table IV were performed to assess the robustness of the proposed method.

TABLE IV. ATTACKS USED IN THIS STUDY FOR WATERMARKED SOUND

Attacks	Description
Noise addition	Additive white Gaussian noise (AWGN) is added to the watermarked audio signal.
Cropping	1000 samples are removed from the beginning of the watermarked signal and then these samples are replaced by the watermarked samples attacked with AWGN.
Re-sampling	The watermarked signal originally sampled at 44.1 kHz is re-sampled at 22.050 kHz and then restored by sampling again at 44.1 kHz.
Re-quantization	The 16 bit watermarked audio signal is quantized down to 8 bits/sample and again re-quantized back to 16 bits/sample.
MP3 Compression	MPEG-1 layer 3 compression with 128 kbps is applied to the watermarked audio signal.

TABLE V. NC AND BER OF THE EXTRACTED WATERMARK FOR DIFFERENT AUDIO SIGNALS

Audio Signal	Attack Type	NC	BER (%)
Pop	No attack	1	0
	Noise addition	0.9919	0.9766
	Cropping	0.9926	0.8789
	Re-sampling	0.9894	1.2695
	Re-quantization	1	0
	MP3 compression	0.9822	2.1484
Jazz	No attack	1	0
	Noise addition	0.9943	0.6836
	Cropping	0.9992	0.0977
	Re-sampling	0.9918	0.9766
	Re-quantization	1	0
	MP3 compression	0.9748	3.0303
Classical	No attack	1	0
	Noise addition	0.9959	0.4883
	Cropping	0.9984	0.1953
	Re-sampling	0.9910	1.0742
	Re-quantization	1	0
	MP3 Compression	0.9721	3.3203
Speech	No attack	1	0
	Noise addition	0.9951	0.5865
	Cropping	0.9976	0.2930
	Re-sampling	0.9902	1.1719
	Re-quantization	1	0
	MP3 Compression	0.9737	3.1250

Normalized correlation (NC) coefficient is used to compare the similarities between the original watermark W and the extracted watermark W^*, which is calculated as:

$$NC\,(W,W^*) = \frac{\sum_{i=1}^{I} w(i)\cdot w^*(i)}{\sqrt{\sum_{i=1}^{I} w(i)\cdot w(i)}\sqrt{\sum_{i=1}^{I} w^*(i)\cdot w^*(i)}} \quad (12)$$

The correlation between W and W^* is very high when NC (W, W^*) is close to 1. On the other hand, the correlation between W and W^* is very low when NC (W, W^*) is close to zero.

TABLE VI. A GENERAL COMPARISON BETWEEN THE PROPOSED AND SEVERAL RECENT METHODS SORTED BY DATA PAYLOAD

Reference	Algorithm	Payload (bps)	Re-sampling BER (%)	Re-quantization BER (%)	MP3 compression BER (%)
Proposed	**LWT-QRD**	**172.39**	**1.27 (22.05 kHz)**	**0 (8 bits/sample)**	**3.32 (128 kbps)**
[4]	Wavelet-based entropy	172.28-86.14	9.1 (22.05 kHz)	--	6.7 (128 kbps)
[5]	DWT-based energy proportion	114.82	6.92 (22.05 kHz)	--	5.71 (80 kbps)
[15]	EMD	46.9-50.3	3 (22.05 kHz)	0 (8 bits/sample)	1 (32 kbps)
[13]	DWT-SVD	45.90	2 (22.05 kHz)	0 (8 bits/sample)	1 (32 kbps)
[11]	DCT-SVD	43	0 (22.05 kHz)	0 (8 bits/sample)	3 (32 kbps)
[18]	Histogram	3	0 (--)	0 (8 bits/sample)	15 (128 kbps)
[19]	DWT-based Histogram	2	0 (16 kHz)	0 (8 bits/sample)	17.50 (64 kbps)
[10]	LWT	--	16.50 (36.750 kHz)	22.09 (8 bits/sample)	51.73 (128 kbps)

The bit error rate (BER) is used to measure the robustness of a watermarking method and is computed as:

$$BER(W,W^*) = \frac{\sum_{i=1}^{I} w(i) \oplus w^*(i)}{I} \tag{13}$$

where $\oplus$ is an exclusive or (XOR) operation.

The proposed method utilizes chaotic encryption to enhance the security. The proposed watermark embedding and detection processes depend on the quantization parameter D and the secret key K. Thus, it is impossible to maliciously detect the watermark without them. Table V shows the robustness results of the proposed method in terms of NC and BER against several attacks for the audio signal 'Pop', 'Jazz', 'Classical', and 'Speech'. The NC values are all above 0.97 and the BER values are below 4%. This clearly indicates a high robustness of the proposed method against different attacks. This is because watermark bits are embedded into the largest element of the upper triangular matrix obtained from the low frequency LWT coefficients of each frame.

Table VI shows a general comparison between the proposed and the several recent methods sorted by data payload, which is based on the reported result in the references [4]-[5], [10]-[11], [13], [15], [18]-[19]. Moreover, re-sampling, re-quantization, and MP3 compression are compared in Table VI. From this comparison, we observed that the proposed method provides higher data payload and lower BER values against several attacks compared with the recent watermarking methods.

V. CONCLUSION

A blind audio watermarking method based on LWT and QRD has been presented in this paper. The performance analysis of the proposed method indicates that it is highly robust against various attacks including noise addition, cropping, re-sampling, re-quantization, and MP3 compression. In addition, audio quality evaluation tests show high imperceptibility of the watermark in the audio signals. Moreover, it has high data payload and provides superior performance than the state-of-the-art audio watermarking methods. These results verify the effectiveness of the proposed watermarking method as a suitable candidate for audio copyright protection.

REFERENCES

[1] S. Katzenbeisser and F. A. P. Petitcolas, *Information Hiding Techniques for Steganography and Digital Watermarking*, Artech House Norwood Mass, USA, 2000.

[2] P. Bassia, I. Pitas, and N. Nikolaidis, "Robust audio watermarking in the time domain," *IEEE Trans. Multimedia*, vol. 3, no. 2, pp. 232-241, 2001.

[3] W. N. Lie and L. C. Chang, "Robust high quality time domain audio watermarking based on low frequency amplitude modification," *IEEE Trans. Multimedia*, vol. 8, no. 1, pp. 46-59, 2006.

[4] S. T. Chen, H. N. Huang, C. J. Chen, K. K. Tseng and S. Y. Tu, "Adaptive audio watermarking via the optimization point of view on the wavelet-based entropy," *Digital Signal Process.*, vol. 23, no. 3, pp. 971-980, 2013.

[5] S. T. Chen, H. N. Huang, C. J. Chen, and G. D. Wu, "Energy-proportion based scheme for audio watermarking," *IET Signal Process.*, vol. 4, no. 5, pp. 576-587, 2010.

[6] I. Cox, J. Killian, F. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Process.*, vol. 6, no. 12, pp. 1673-1687, 1997.

[7] D. Megías, J. Serra-Ruiz, and M. Fallahpour, "Efficient self–synchronized blind audio watermarking system based on time domain and FFT amplitude modification," *Signal Process.*, vol. 90, no. 12, pp. 3078-3092, 2010.

[8] M. Fan and H. Wang, "Chaos-based discrete fractional sine transform domain audio watermarking scheme," *Comp. Elect. Eng.*, vol. 35, no. 3, pp. 506-516, 2009.

[9] W. Sweldens, "The lifting scheme: a custom-design construction of biorthogonal wavelets," *Appl. Computational Harmonic Analysis*, vol. 3, no. 2, pp. 186,-200, 1996.

[10] E. Erçelebi and L. Batakçı, "Audio watermarking scheme based on embedding strategy in low frequency components with a binary image," *Digital Signal Process.*, vol. 19, no. 2, pp. 265-277, 2009.

[11] B.Y. Lei, I. Y. Soon, and Z. Li, "Blind and robust audio watermarking scheme based on SVD-DCT," *Signal Process.*, vol. 91, pp.1973-1984, 2011.

[12] W. Al-Nuaimy, M. A. M. El-Bendary, A. Shafik, F. Shawki, A. E. Abou-El-Azm, N. A. El-Fishawy, S. M. Elhalafawy, S. M. Diab, B. M. Sallam, F. E. A. El-Samie, and H. B. Kazemian, "An SVD audio watermarking approach using chotic encrypted images," *Digital Signal Process.*, vol. 21, no. 6, pp. 764-779, 2011.

[13] V. K. Bhat, I. Sengupta, and A. Das, "An adaptive audio watermarking based on the singular value decomposition in the wavelet domain," *Digital Signal Process.*, vol. 20, no. 6, pp. 1547-1558, 2010

[14] F. E. A. El-Samie, "An efficient singular value decomposition algorithm for digital audio watermarking," *Int. J. Speech Technol.*, vol. 12, no. 1, pp. 27-45, 2009.

[15] K. Khaldi and A. O. Boudraa, "Audio watermarking via EMD," *IEEE Trans. Audio, Speech Lang. Process.*, vol. 21, no. 3, pp. 675-680, 2013.

[16] H. Kang, K. Yamaguchi, B. Kurkoski, K. Yamaguchi, and K. Kobayashi, "Full-index-embedding patchwork algorithm for audio watermarking," *IEICE Trans. Inf. & Syst.*, vol. E91-D, no. 11, pp. 2731-2734, 2008.

[17] S. Xiang and J. Huang, "Histogram based audio watermarking against time scale modification and cropping attacks," *IEEE Trans. Multimedia*, vol. 9, no. 7, pp. 1357-1372, 2007.

[18] S. Xiang, H. J. Kim, and J. Huang, "Audio watermarking robust against time scale modification and MP3 compression," *Signal Process.*, vol. 88, no. 10, pp. 2372-2387, 2008.

[19] T. Thiede, W. C. Treurniet, R. Bitto, C. Schmidmer, T. Sporer, J. G. Beerens, C. Colomes, M. Keyhl, G. Stoll, K. Brandenburg and B. Feiten, "PEAQ - the ITU standard for objective measurement of perceived audio quality," *J. Audio Eng. Soc.*, vol. 48, no. 1/2, pp. 3-29, 2000.

Design of Ternary Reversible Sequential Circuits

Mozammel H A Khan

Department of Computer Science and Engineering, East West University
Aftabnagar, Dhaka 1212, Bangladesh
mhakhan@ewubd.edu

Abstract— **A good number of efforts have been made on design of ternary reversible combinational circuits, but no significant attempt has yet been made to design arbitrary ternary reversible sequential circuits. In this paper, we have proposed a novel design method of ternary reversible sequential circuit using direct feedback of the present state output. We have also shown the way of making the circuit falling-edge triggered and pre-settable using Fredkin gate. For this purpose, for the first time, we have shown realization of Fredkin gate on the top of M-S gates. We have shown design examples of an arbitrary two-trit sequential circuit, three-trit up/down counter, and three-trit universal register.**

Index Terms—**Reversible logic, ternary reversible gates, ternary reversible sequential circuit, TGFSOP expression.**

I. Introduction

Ternary circuit is more compact than binary circuit, since ternary encoded realization of a function requires 0.63 times input lines than binary realization [1]. Ternary reversible/quantum circuits are realizable using linear ion-trap technology [2]. The reversible gates proposed in [2] are commonly known as Muthukrishnan-Stroud (M-S) gates. A good number of design methods for ternary reversible combinational circuits are proposed in the literature [1], [3]-[13]. Among them, ternary Galois field sum of products (TGFSOP) expression based design method [1], [3] has become very promising and practical, since ternary logic function with any number of inputs can be easily represented as TGFSOP expression and any TGFSOP expression can be realized as cascade of ternary Feynman and Toffoli gates. Moreover, ternary Feynman and Toffoli gates can be realized on the top of M-S gates [14]. On the other hand, only one attempt has been made to design ternary reversible sequential circuit [15], where design of only D and T flip-flops are proposed and nothing is stated regarding design of arbitrary sequential circuits using these flip-flops.

In this paper, we have proposed a novel design method for ternary sequential circuits, where the present state outputs are directly fed-back to the next state determination circuit and that circuit is realized using TGFSOP expression as a cascade of ternary one-trit (ternary digit), Feynman and Toffoli gates. We have also developed methods for making the ternary reversible sequential circuit falling-edge triggered and pre-settable. As design example, we have presented designs of an arbitrary sequential circuit with two-trit state, one input, and one output; a three-trit up/down counter; and a three-trit universal register.

The rest of the paper is organized as follows. In section II, we have discussed preliminaries of TGFSOP expression and ternary reversible gates. We have discussed the proposed method of designing ternary reversible sequential circuit in section III. Finally, in section IV, we have concluded the paper.

II. Preliminaries

The basis of TGFSOP expression is the GF(3) operation, which has a set of values $T = \{0, 1, 2\}$ and two binary operations - addition (denoted by $\oplus$) and multiplication (denoted by $\cdot$ or juxtaposition). The GF(3) operations are mod 3 operations. The GF(3) operations are commutative. The multiplication operation is distributive over addition operation.

There are 3! = 6 possible permutations of 0, 1, and 2. Therefore, there are six possible reversible literals of a ternary variable that take the values as shown in Table I. The literal x^{+0} is often written as x.

TABLE I
Ternary Reversible Literals

Variable	Reversible Literal					
x	x^{+0}	x^{+1}	x^{+2}	x^{01}	x^{02}	x^{12}
0	0	1	2	1	2	0
1	1	2	0	0	1	2
2	2	0	1	2	0	1

In [3], we have developed 16 ternary Galois field expansions (TGFE), which are used to expand a ternary logic function $f(x_1, \ldots, x_n)$ on any of the n variables to generate the TGFSOP expression. In [3], Kronecker tree is constructed for TGFSOP minimization. In this paper, we have used pseudo Kronecker tree for TGFSOP minimization as discussed in section III. In general, pseudo Kronecker tree produces more minimization than Kronecker tree.

The one-trit gates are shown in Fig. 1(a), where the output P is the U-transform of the input A and U is any unitary transform from Table I. The cost of one-trit gate is assumed to be one. Ternary M-S gate is shown in Fig. 1(b). The output $P = A$. When $A = 2$, then the output Q is the U-transform of B, otherwise $Q = B$. The cost of M-S gate is assumed to be one. Ternary Feynman gate is shown in Fig. 1(c). The outputs are $P = A$ and $Q = A \oplus B$. Realization of Feynman gate on the top of M-S gate is proposed in [14] and its cost is four. Ternary multiple-controlled (n-trit with $n > 2$) Toffoli gate is shown in Fig. 1(d). Realization of multiple-controlled Toffoli gate on the top of M-S gate is proposed in [14] and its cost is 8 + 2×cost of (n - 1)-trit Toffoli gate. The cost of Toffoli gates for n = 2, 3, 4, 5, 6, 7, 8, 9, 10 are 4, 16, 40, 88, 184, 376, 760, 1528, 3064, respectively.

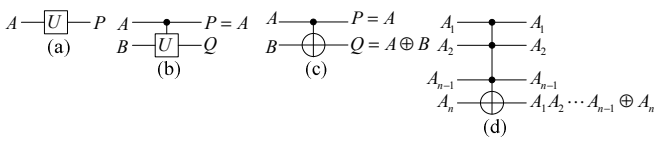


Fig. 1 Ternary (a) one-trit, (b) M-S, (c) Feynman, and (d) multiple-controlled Toffoli gates.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

Ternary Fredkin gate is shown in Fig. 2(a). If $A \neq 2$ (i.e., $A^{02}A^{02} = 1$ and $AA^{01} = 0$), then $Q = B$ and $R = C$. If $A = 2$ (i. e., $A^{02}A^{02} = 0$ and $AA^{01} = 1$), then $Q = C$ and $R = B$. Thus, when $A \neq 2$, then the inputs B and C are passed directly to the outputs and when $A = 2$, then they are swapped at the output. The realization of Fredkin gate on the top of M-S gate is shown in Fig. 2(b) and its cost is 15.

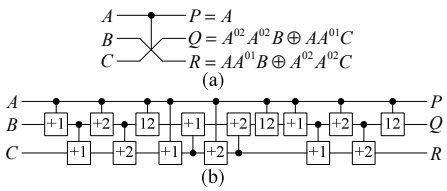


Fig. 2 Ternary (a) Fredkin gate and (b) its realization on the top of M-S gate.

III. Design of Ternary Reversible Sequential Circuit

The basic principle of sequential circuit is that the next state and the output are determined based on the present state and the input. So, if the present state output can be directly fed-back to the next state determination circuit, then ternary reversible sequential circuit can be designed. Let Q and Q^+ be the present state and the next state, respectively. Then Q^+ can be expressed as $Q^+ = Q \oplus Q \oplus Q \oplus Q^+ = Q \oplus Q^*$, where $Q^* = Q \oplus Q \oplus Q^+$ is the modified next state. Now, Q^* is determined as the function of the present state and the input. Let C be the clock. Then the next state is realized in the next state determination circuit as $Q^+ = Q \oplus CC^{01}Q^*$. When $C \neq 2$ (i. e., $CC^{01} = 0$), then $Q^+ = Q$, which provides the feedback of the present state through the next state determination circuit to maintain the present state. When $C = 2$ (i. e. $CC^{01} = 1$), then $Q^+ = Q \oplus Q^*$, which generates the next state in the "Next state determination" circuit. We express the next state function as TGFSOP expression. The TGFSOP expression is minimized using pseudo Kronecker tree, where one of the 16 TGFEs [3] is used for each of the $(3^n - 1)/2$ internal nodes of the tree. For this purpose, we have written a C program to construct pseudo Kronecker tree from the given output vector and to determine the corresponding minimized TGFSOP expression. We also implement the output of the sequential circuit in the "Output determination" circuit using TGFSOP expression of the output as function of present state and input.

The proposed design method is illustrated using an arbitrary sequential circuit with two-trit state $Q1Q0$, one input x, and one output z as shown in the state transition diagram of Fig. 3. The truth table representing the next states $Q1^+Q0^+$, the output z, and the modified next states $Q1^*Q0^*$ is shown in Table II. From the output vector of the modified next state $Q1^*$ of Table II, we have generated the pseudo Kronecker tree for $Q1^*$ as shown in Fig. 4. The number inside each circle is the selected TGFE for that node. The reversible literals corresponding to each of the cofactors of the selected TGFE are written along the edges. Each product term of the generated TGFSOP expression is determined by multiplying the literals along the path from the root to the non-zero constant leaf and the constant itself. Thus, the minimized TGFSOP expression for $Q1^*$ from the tree of Fig. 3 is

$$Q1^* = 2x^{+1}Q1^{+1}Q0^{+1} \oplus x^{12}Q1^{+2}Q0^{+2} \quad (1)$$

The corresponding next state equation for $Q1^+$ is determined from (1) using the formula $Q^+ = Q \oplus CC^{01}Q^*$ as follows. Readers should note that $2C^{01} = C^{+2}$.

$$Q1^+ = Q1 \oplus CC^{+2}x^{+1}Q1^{+1}Q0^{+1} \oplus CC^{01}x^{12}Q1^{+2}Q0^{+2} \quad (2)$$

Similarly, we have determined the equation for the next state $Q0^+$ as follows:

$$Q0^+ = Q0 \oplus CC^{01}Q1Q0 \oplus CC^{+2}xQ1^{+2} \quad (3)$$

From the output vector of the output z from Table II, we have determined the TGFSOP expression for the output z as follows:

$$z = Q0 \oplus xQ1^{01}Q0 \oplus xQ1Q0^{+1} \quad (4)$$

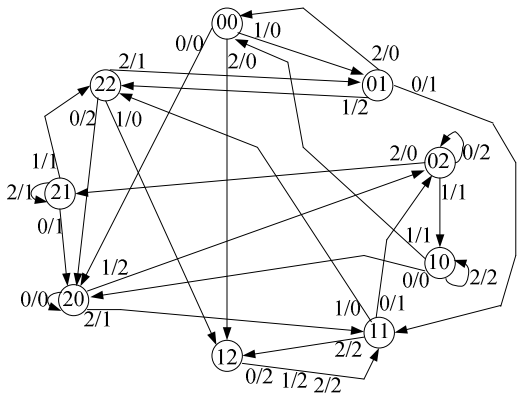


Fig. 3 State transition diagram of an arbitrary sequential circuit.

TABLE II
Truth Table Representing Next States, Output, and Modified Next States of the Sequential Circuit of Fig. 3

x	$Q1$	$Q0$	$Q1^+$	$Q0^+$	z	$Q1^*$	$Q0^*$
0	0	0	2	0	0	2	0
0	0	1	1	1	1	1	0
0	0	2	0	2	2	0	0
0	1	0	2	0	0	1	0
0	1	1	0	2	1	2	1
0	1	2	1	1	2	0	2
0	2	0	2	0	0	0	0
0	2	1	2	0	1	0	2
0	2	2	2	0	2	0	1
1	0	0	0	1	0	0	1
1	0	1	2	2	2	2	1
1	0	2	1	0	1	1	1
1	1	0	0	0	1	2	0
1	1	1	2	2	0	1	1
1	1	2	1	1	2	0	2
1	2	0	0	2	2	1	2
1	2	1	2	2	1	0	1
1	2	2	1	2	0	2	0
2	0	0	1	2	0	1	2
2	0	1	0	0	0	0	2
2	0	2	2	1	0	2	2
2	1	0	1	0	2	0	0
2	1	1	1	2	2	0	1
2	1	2	1	1	2	0	2
2	2	0	1	1	1	2	1
2	2	1	2	1	1	0	0
2	2	2	0	1	1	1	2

The reversible realization of the sequential circuit of Fig. 3 is shown in Fig. 5. Each of the five parts of Fig. 5 is discussed separately.

In the "Feedback" part, the two present states are copied using two Feynman gates and fed-back to the "Next state determination" part of the circuit.

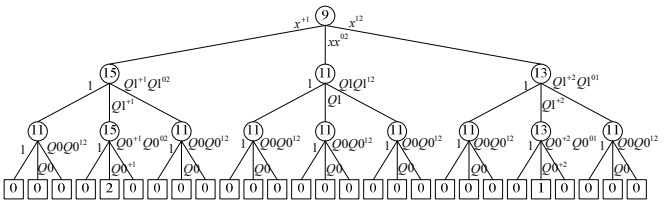


Fig. 4 Pseudo Kronecker tree for minimizing TGFSOP expression for $Q1^*$ of Table II.

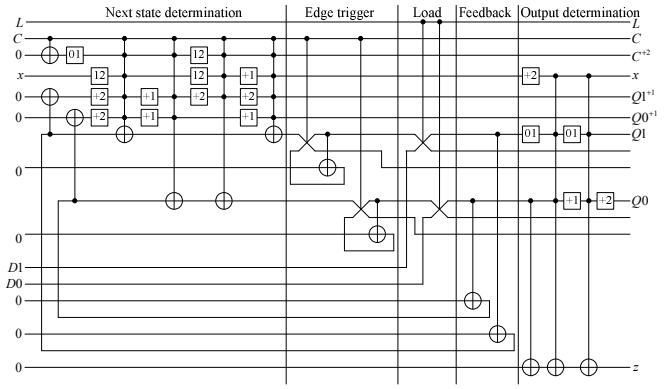


Fig. 5 Reversible realization of the sequential circuit represented by the state transition diagram of Fig. 3.

In the "Load" part, when $L = 2$, then $D1$ and $D0$ data inputs will be loaded to the state output $Q1$ and $Q0$, respectively, through the Fredkin gates and they will become the present state. Then these newly loaded output states will be fed-back to the "Next state determination" part by the "Feedback" part of the circuit. As discussed below, the "Next state determination" part determines the next state when the clock $C = 2$. So, the output state loading should me made asynchronously only when $C \neq 2$ so that the loaded states are fed-back before the next state determination starts. When $L \neq 2$, then the Fredkin gates will pass the two state values provided by the "Edge trigger" part to maintain the output states unchanged.

The next state equations of (2) and (3) are realized in the "Next state determination" part of Fig. 5 as function of the fed-back present states $Q1$ and $Q0$, the input x, and the clock C. When $C \neq 2$ (i. e., $CC^{01} = 0$), then $Q^+ = Q \oplus CC^{01}Q^* = Q$. In this case, the two outputs of the "Next state determination" part will simply be equal to the two fed-back present state values. When $C = 2$ (i. e., $CC^{01} = 1$), then $Q^+ = Q \oplus CC^{01}Q^* = Q \oplus Q^*$. In this case, the "Next state determination" part will determine the two next states.

In the "Edge trigger" part, when $C = 2$, then the Fredkin gate will provide the feedback to maintain the state output unchanged. At the same time, the "Next state determination" part will determine the next state. After that when the clock goes from $C = 2$ to $C \neq 2$, then the Fredkin gates will pass the outputs of the "Next state determination" part to the state outputs through the "Load" part provided $L \neq 2$ is maintained. Thus, the sequential circuit becomes a falling-edge triggered circuit.

In the "Output determination" part, the output z is realized using the TGFSOP expression of (4) as function of the present state $Q1Q0$ and the input x.

The circuit of Fig. 5 has 17 one-trit, eight Feynman, four Fredkin, two four-trit Toffoli, two five-trit Toffoli, and two six-trit Toffoli gates. Thus, its cost is 17×1 + 8×4 + 4×15 + 2×40 + 2×88 + 2×184 = 733. The useful outputs of the circuit are $Q1$, $Q0$, and z. Thus, the circuit has 10 garbage outputs. The state outputs from the "Next state determination" part are older version of the states and are not useful for further use. For this reason, they are considered as garbage outputs.

We have designed a three-trit up/down counter as shown in Fig. 6. The state outputs are $Q2$, $Q1$, and $Q0$. The mode selector input M determines the direction of count. When $M \neq 2$, then it counts in the up direction. When $M = 2$, then it counts in the down direction. The determined next state equations are given below:

$$Q2^+ = Q2 \oplus CC^{+2}Q1^{+1}Q1^{01}Q0^{+1}Q0^{01} \oplus CC^{+2}M^{+1}M^{02}Q1^{+1}Q1^{01}Q0^{+1}Q0^{01} \oplus CC^{+2}M^{+1}M^{02}Q1Q1^{01}Q0Q0^{01} \tag{5}$$

$$Q1^+ = Q1 \oplus CC^{+2}Q0^{+1}Q0^{01} \oplus CC^{01}M^{+1}M^{02}Q0^{+2}Q0^{01} \tag{6}$$

$$Q0^+ = Q0 \oplus CC^{+2} \oplus CC^{01}M^{+1}M^{02} \tag{7}$$

Based on the next state equations (5) to (7), reversible realization of three-trit up/down counter is shown in Fig. 6. The circuit of Fig. 6 has 11 one-trit, 12 Feynman, six Fredkin, one three-trit Toffoli, two five-trit Toffoli, two seven-trit Toffoli, and two nine-trit Toffoli gates. Thus, its cost is 11×1 + 12×4 + 6×15 + 1×16 + 2×88 + 2×376 + 2×1528 = 4149. The circuit has 15 garbage outputs. The state outputs from the "Next state determination" part are older version of the states and are not useful for further use. For this reason, they are considered as garbage outputs.

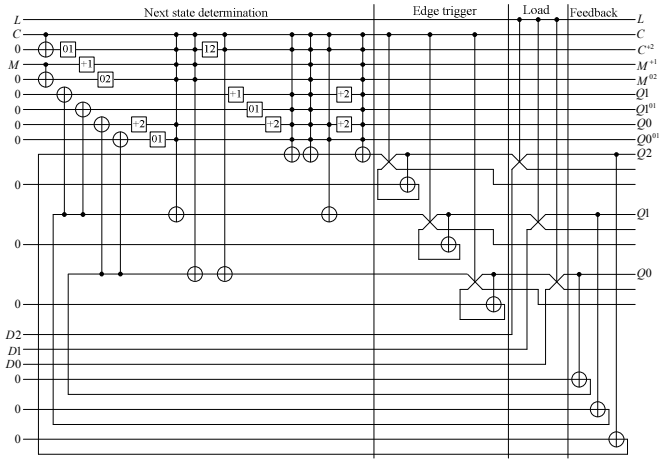


Fig. 6 Reversible realization of three-trit up/down counter.

We have designed a three-trit universal register as shown in Fig. 7. The state outputs are $Q2$, $Q1$, and $Q0$. The "Load" part of the circuit provides asynchronous load facility for parallel input. When $L = 2$, then the parallel input data $D2$, $D1$, and $D0$ are loaded to the state outputs $Q2$, $Q1$, and $Q0$, respectively. Parallel outputs are taken form $Q2$, $Q1$, and $Q0$. The mode selector input M determines the direction of shift. When $M \neq 2$, then the circuit shift the data in the right direction. In this case, the serial input is D_R and the serial output is $Q0$. When $M = 2$, then the circuit shifts the data in the left direction. In this case, the serial input is D_L and the serial output is $Q2$. Thus, the circuit is a three-trit universal register. The determined next state equations are given below:

$$Q2^{+} = Q2 \oplus CC^{01}Q2^{01}Q1 \oplus CC^{01}Q2Q1^{+2} \oplus CC^{01}M^{+1}M^{02}D_R^{01}Q1 \oplus CC^{01}M^{+1}M^{02}D_RQ1^{+2} \quad (8)$$

$$Q1^{+} = Q1 \oplus CC^{01}Q1^{01}Q0 \oplus CC^{01}Q1Q0^{+2} \oplus CC^{01}M^{+1}M^{02}Q2^{01}Q0 \oplus CC^{01}M^{+1}M^{02}Q2Q0^{+2} \quad (9)$$

$$Q0^{+} = Q0 \oplus CC^{01}Q0^{01}D_L \oplus CC^{01}Q0D_L^{+2} \oplus CC^{01}M^{+1}M^{02}Q1^{01}D_L \oplus CC^{01}M^{+1}M^{02}Q1D_L^{+2} \quad (10)$$

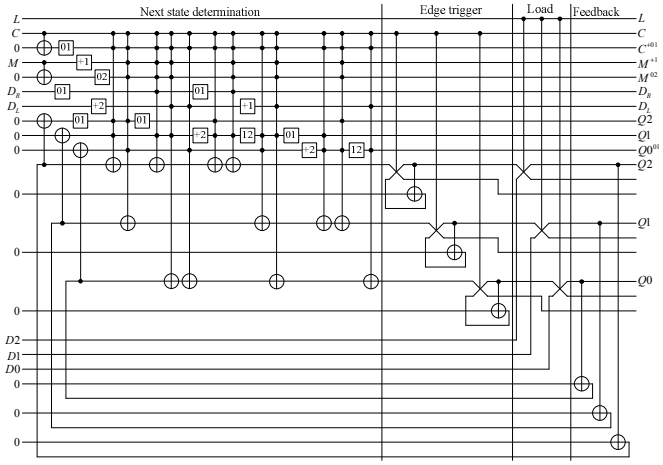


Fig. 7 Reversible realization of three-trit universal register.

Based on the next state equations (8) to (10), reversible realization of three-trit universal register is shown in Fig. 7. The circuit of Fig. 7 has 14 one-trit, 11 Feynman, six Fredkin, six five-trit Toffoli, and six seven-trit Toffoli gates. Thus, its cost is 14×1 + 11×4 + 6×15 + 6×88 + 6×376 = 2932. The circuit has 16 garbage outputs. The state outputs from the "Next state determination" part are older version of the states and are not useful for further use. For this reason, they are considered as garbage outputs.

IV. Conclusions

Ternary circuits are more compact than corresponding binary circuits. Though a good number of efforts have been made in designing ternary reversible combinational circuits, no significant effort has been reported on designing ternary reversible sequential circuits. In this paper, for the first time, we have proposed a novel technique for designing ternary reversible sequential circuit. For this purpose, we have provided direct feedback of the present state output to the next state determination circuit. To accommodate the direct feedback of the present state to the next state determination circuit, a special technique has been developed for determining modified next state. Actual next state is then determined using the fed-back present state and the modified next state equation. Asynchronous load facility and falling-edge triggering is provided using ternary Fredkin gate. For determining the next state equation and the output equation, we have used TGFSOP expression. In the earlier effort [3], TGFSOP expression was minimized using Kronecker tree. As the pseudo Kronecker tree produces more minimized TGFSOP expression, in this paper, we have minimized the TGFSOP expression using pseudo Kronecker tree.

References

[1] M. H. A. Khan, “GFSOP-based ternary quantum logic synthesis,” in *Optics and Photonics for Information Processing IV, Proc. SPIE*, vol.7797, 2010.

[2] A. Muthukrishnan and C. R. Stroud Jr., "Multivalued logic gates for quantum computation," *Phy. Rev. A*, vol. 62, no. 5, 2000, pp. 052309/1-8.

[3] M. H. A. Khan, M. A. Perkowski, M. R. Khan, and P. Kerntopf, “Ternary GFSOP minimization using kronecker decision diagrams and their synthesis with quantum cascades,” *J. Multiple-Valued Logic Soft Computing*, vol. 11, no. 5-6, 2005, pp. 567-602.

[4] M. H. A. Khan, “Design of Reversible/Quantum Ternary Multiplexer and Demultiplexer,” *Engineering Letters*, vol. 13, no. 2, 2006, pp. 65-69.

[5] M. H. A. Khan and M. A. Perkowski, “Quantum ternary parallel adder/subtractor with partially-look-ahead carry", *J. Systems Architecture*, vol. 53, no. 7, 2007, pp. 453-464.

[6] M. H. A. Khan, “Design of reversible/quantum ternary comparator circuits,” *Engineering Letters*, vol. 16, no. 2, 2008, pp. 178-184.

[7] S. Kotiyal, H. Thapliyal, and N. Ranganathan, "Design of a ternary barrel shifter using multiple-valued reversible logic," in *Proc. Conf. Nanotechnology*, 2010, pp. 1104-1108.

[8] R. P. Zadeh and M. Haghparast, "A new reversible/quantum ternary comparator," *Australian J. Basic Applied Sciences*, vol. 5, no. 12, 2011, pp. 2348-2355.

[9] N. Nower and A. R. Chowdhury, "Design and analysis of a compact reversible ternary systolic array," *Intl. J. Computer Electrical Engineering*, vol. 3, no. 6, Dec. 2011, pp. 890-895.

[10] J. E. Rice and R. Rahman, "A modular approach to designing an online testable ternary reversible circuit," *Intl. J. Information Computer Science*, vol. 2, no. 5, July 2013, pp. 65-76.

[11] B. Mondal, P. Sarkar, P. K. Saha, and S. Chakraborty, "Synthesis of balanced ternary reversible logic circuit," in *Proc. Intl. Symp. Multiple-Valued Logic*, 2013, pp. 334-339.

[12] N. M. Nayeem and J. E. Rice, "Design of an online testable ternary circuit from the truth table," in *Reversible Computation 2012*, LNCS 7581, 2013, pp. 152-159.

[13] X. Li, G. Yang, and D. Zheng, "Logic synthesis of ternary quantum circuits with minimal qutrits," *J. Computers*, vol. 8, no. 8, Aug. 2013, pp. 1941-1946.

[14] M. H. A. Khan, “Quantum realization of multiple-valued Feynman and Toffoli gates without ancilla input,” in *Proc. Intl. Symp. Multiple-Valued Logic*, 2009, pp. 103-108.

[15] M. Mohammadi, M. Eshghi, and M. Haghparast, "On design of multipl-valued sequential reversible circuits for nanotechnology based systems," in *Proc. IEEE Region 10 Conf. (TENCON)*, 2008.

SET Based Low Power Consuming Digital Circuits

Ahmed Shariful Alam,[1,*] Md. Abdur Rahman,[1] Abu Hena Md. Mustafa Kamal,[1] and Md Salek Mahmud[2]
[1]Department of EEE, Islamic University of Technology
Gazipur, Bangladesh
[2]Department of EEE, BUBT
Mirpur-2, Bangladesh
*estiak2126@yahoo.com

***Abstract*—Single Electron Transistor (SET) is an ultra small device which can be used as a possible replacement of a MOSFET. Like CMOS technology, CSET (Complementary SET) models can be used to build same digital circuits. Apart from the drawbacks (which have been discussed at the end of this paper), these digital circuits work fine. One of the main advantages of a SET based digital circuit over a MOSFET based digital circuit is its low power consumption. In this paper we have focused on this issue. For an example, we have presented the circuits of NAND gate and NOR gate made up of both MOSFETs and SETs. The PSPICE simulations of these circuits show that the SET built digital circuits consume much less power than those of MOSFETs.**

***Index Terms*—Moore's law, low power consumption, NAND gate, NOR gate, orthodox theory, background charge**

I. Introduction

According to International Technology Roadmap for Semiconductors (ITRS), CMOS technology may be obsolete by 2018 due to its reaching of minimum feature size which is not compatible with the Moore's law. Besides the reduction of power consumption of different electrical and electronic appliances has already been a major issue. Single electron transistor (SET) is a new device which can solve both the problems. This paper emphasized on the low power consumption of SET.

II. MOSFET Based Nand gate and simulated results

Fig. 1 shows the circuit diagram a MOSFET based 2-input NAND gate. Here the supply voltage is taken to be 3.5 V.

All the simulations of MOSFET based circuits (both NAND gate and NOR gate) were done in PSPICE. Plotting of the wave shapes and power consumption calculation were done in MATLAB. The equations which were used in PSPICE can be found in [1]

Fig. 2 shows the PSPICE simulation result of transient analysis of a CMOS NAND gate. Fig. 3 shows the power consumptions across MPU1 and MPU2 at different times respectively and Fig. 4 shows the power consumption across MPD1 and MPD2 at different times respectively.

Average power consumptions across MPU1, MPU2, MPD1 and MPD2 are 9.4545×10^{-6} Watts, 5.4681×10^{-6} Watts, 5.9826×10^{-6} Watts and 3.9908×10^{-6} Watts respectively.

Total average power consumption across the whole circuit = 2.4896×10^{-5} Watts.

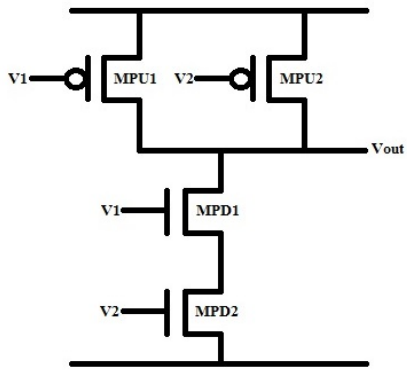


Fig. 1. Circuit diagram of a MOSFET based NAND Gate

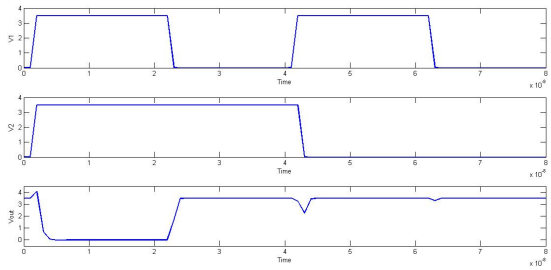


Fig. 2. Transient analysis of a MOSFET based NAND gate

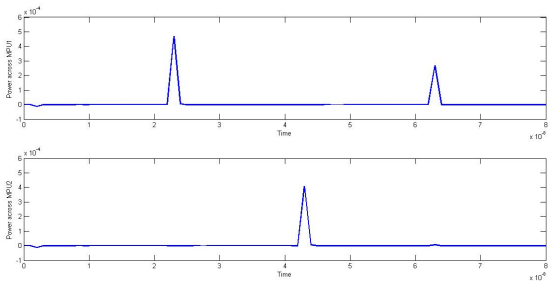


Fig. 3. Power consumptions across MPU1 and MPU2 of a NAND gate

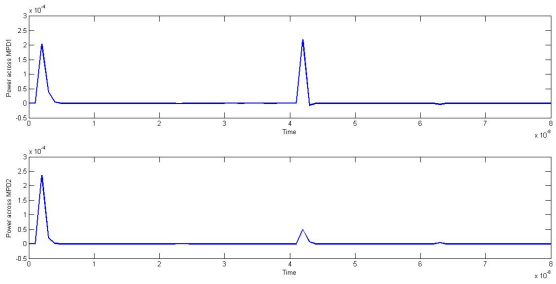


Fig. 4. Power consumptions across MPD1 and MPD2 of a NAND gate

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

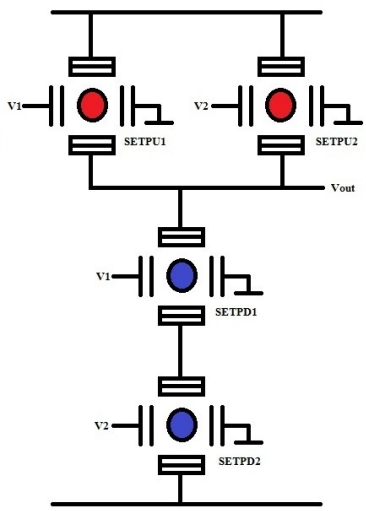


Fig. 5. Circuit diagram of a SET based NAND Gate

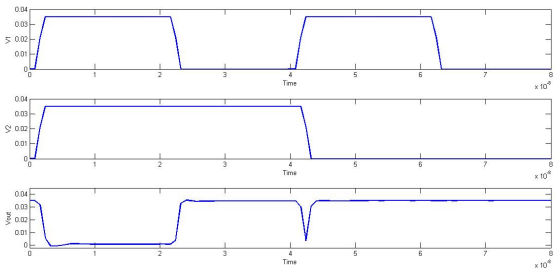


Fig. 6. Transient analysis of a SET based NAND gate

III. SET Based Nand gate and simulated results

Fig. 5 shows the circuit diagram a SET based 2-input NAND gate. Here the supply voltage is taken to be 35mV. The red and blue colored islands of the circuit indicate that the background charges [2] of them are positive and negative respectively. The background charges may have different signs but the absolute values of them are considered same throughout the whole simulation.

The switching characteristics of SET is described in [3]. All the simulations of SET based circuits (both NAND gate and NOR gate) were done in PSPICE. Plotting of the wave shapes and power consumption calculation were done in MATLAB. The equations which were used in PSPICE can be found in [4].

Fig. 6 shows the PSPICE simulation result of transient analysis of a SET based NAND gate. Fig. 7 shows the power consumptions across SETPU1 and SETPU 2 at different times respectively and Fig. 8 shows the power consumption across SETPD 1 and SETPD 2 at different times respectively.

Average power consumptions across SETPU1, SETPU2, SETPD1 and SETPD2 are 2.7457×10^{-12} Watts, 2.8335×10^{-12} Watts, 4.6971×10^{-12} Watts and 2.0068×10^{-12} Watts respectively.

Total average power consumption across the whole circuit = 1.2283×10^{-11} Watts.

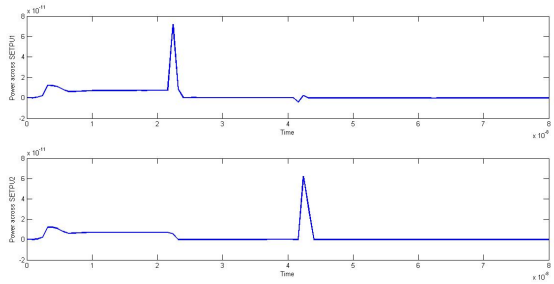


Fig. 7. Power consumptions across SETPU1 and SETPU2 of a NAND gate

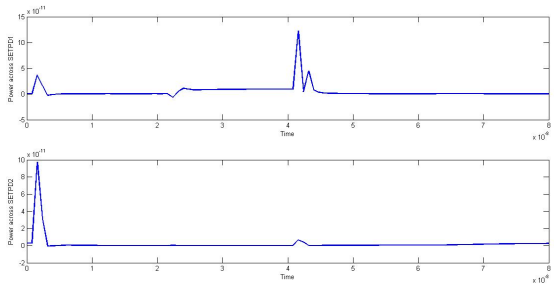


Fig. 8. Power consumptions across SETPD1 and SETPD2 of a NAND gate

IV. MOSFET Based Nor gate and simulated results

Fig. 9 shows the circuit diagram of a MOSFET based 2-input NOR gate. Here the supply voltage is 3.5V as usual for a MOSFET based digital circuit.

Fig. 10 shows the PSPICE simulation results of transient analysis of a CMOS NOR gate. Fig. 11 shows the power consumptions across MPU1 and MPU2 at different times respectively and Fig. 12 shows the power consumption across MPD1 and MPD2 at different times respectively.

Average power consumptions across MPU1, MPU2, MPD1 and MPD2 are 9.4953×10^{-6} Watts, 5.0227×10^{-6} Watts, 4.8594×10^{-6} Watts and 3.5093×10^{-6} Watts respectively.

Total average power consumption across the whole circuit = 2.2887×10^{-5} Watts.

V. SET Based Nor gate and simulated results

Fig. 13 shows the circuit diagram of a SET based 2-input NOR gate. Here the supply voltage is 35mV as usual for a

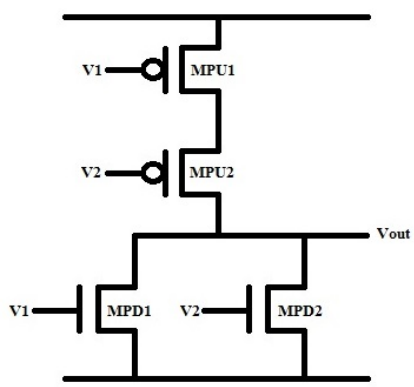


Fig. 9. Circuit diagram of a MOSFET based NOR Gate

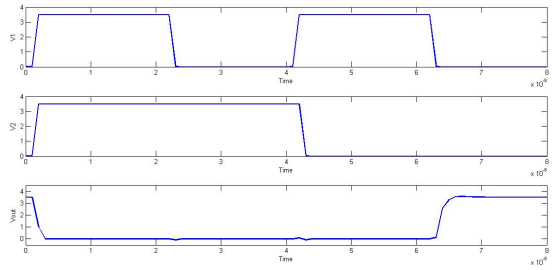


Fig. 10. Transient analysis of a MOSFET based NOR gate

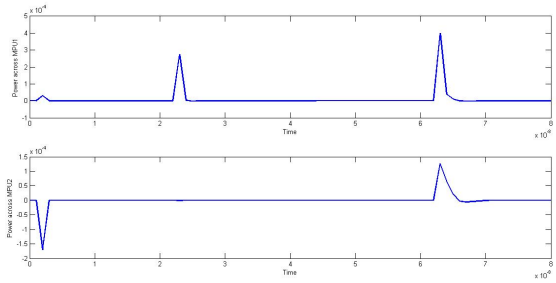


Fig. 11. Power consumptions across MPU1 and MPU2 of a NOR gate

SET based digital circuit.

Fig. 14 shows the PSPICE simulation result of transient analysis of a SET based NOR gate. Fig. 15 shows the power consumptions across SETPU1 and SETPU2 at different times respectively and Fig. 16 shows the power consumption across SETPD1 and SETPD2 at different times respectively.

Average power consumptions across SETPU1, SETPU2, SETPD1 and SETPD2 are 1.7578×10^{-12} Watts, 4.7476×10^{-12} Watts, 2.5310×10^{-12} Watts and 2.3184×10^{-12} Watts respectively.

Total average power consumption across the whole circuit = 1.1355×10^{-11} Watts.

VI. Comparisons of MOSFET and SET based Circuits

From the simulation results it is clear that SET shows almost the same results in transient analysis as the MOSFET does. The main advantage of SET based circuits focused here over MOSFET based circuits is the consumption of less power. The simulation results show that both SET based NAND and NOR gates consume almost **1 million** times less power than those of CMOS circuits.

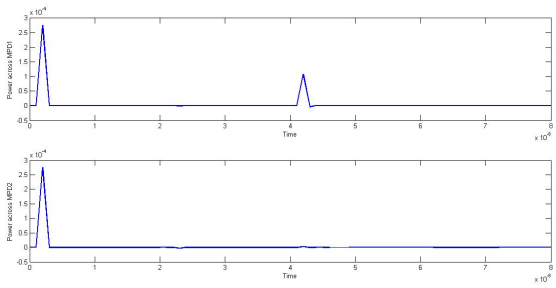


Fig. 12. Power consumptions across MPD1 and MPD2 of a NOR gate

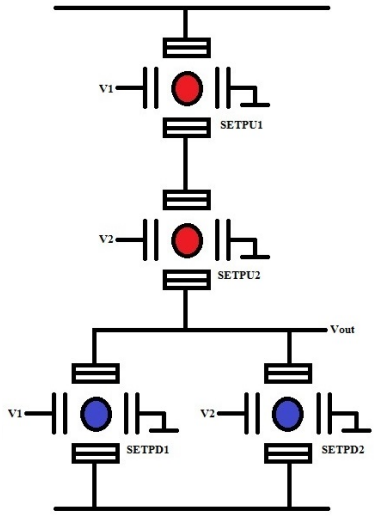


Fig. 13. Circuit diagram of a SET based NOR Gate

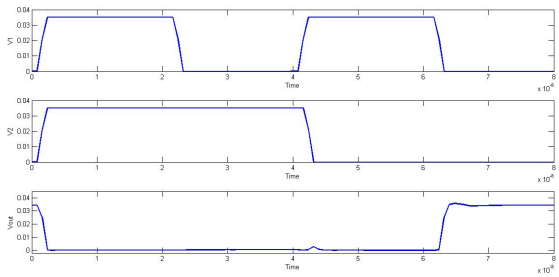


Fig. 14. Transient analysis of a SET based NOR gate

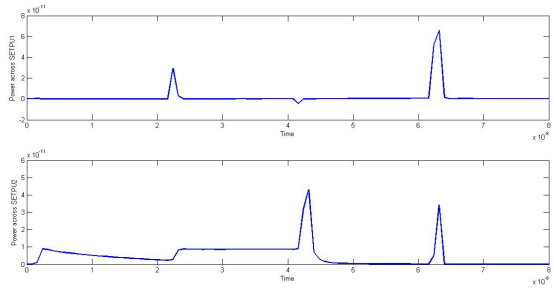


Fig. 15. Power consumptions across SETPU1 and SETPU2 of a NOR gate

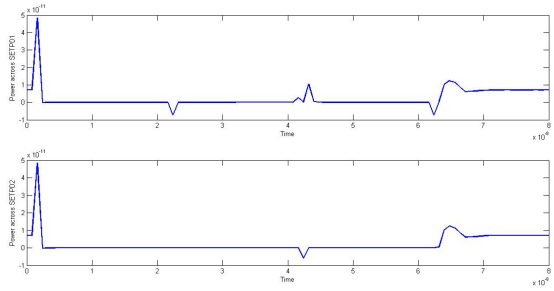


Fig. 16. Power consumptions across SETPD1 and SETPD2 of a NOR gate

VII. Limitations and conclusion

There are some limitations of SET based circuits. In these simulations orthodox theory [5] of single electron tunneling was used where co-tunneling was not considered. The simulations were done assuming the temperature at 27k. In room temperature these simulations do not work. Moreover, SET is very sensitive on background charge issue. We have assumed the background charge of all the SETs in a circuit constant and equal (absolute value). Background charges barely behave like this way. But still SET can be important in coming days' VLSI technologies if these limitations can be recovered.

References

[1] D. A. Pucknell and K. Eshraghian, *Basic VLSI design*, Prentice-Hall, 1995.

[2] J. R. Tucker, "Complementary digital logic based on the Coulomb blockade," *Journal of Applied Physics*, 72, 4399, 1992.

[3] K. K. Likharev, *Single-Electron Devices and Their Applications*, Proceedings of the IEEE, vol. 87, pp. 606 - 632, Aug. 2002.

[4] G. Lientschnig, "Single-Electron and Molecular Devices," Ph.D. thesis, Delft University of Technology, Delft, Netherlands, Nov. 2003.

[5] A. Scholze, "Simulation of single-electron devices," Ph.D. dissertation, Swiss Federal Institute of Technology, Zurich, Switzerland, 2000.

Design of DDR4 SDRAM Controller

Md. Ashraful Islam,[1,*] Md. Yeasin Arafath,[2] and Md. Jahid Hasan[3]

[1]ASIC Design Engineer, Fastrack Design Ltd., Bangladesh
[2]ASIC Design Engineer, Fastrack Design Ltd., Bangladesh
[3]ASIC Design Engineer, Fastrack Design Ltd., Bangladesh
[*1]ashraful@fastrack-design.com, [2]yeasin@fastrack-design.com, [3]jahid@fastrack-design.com

Abstract— **Currently, the demand of memories has been increasing due to its higher speed, lower cost and lower power consumption. The complexity of instructions to control the memory devices increases with the increase in technology. DDR4 is the latest generation family of DDR SDRAM. This memory device provides higher reliability, availability and serviceability than other DDR memories. In this paper, the overall architecture of the DDR4 SDRAM controller is proposed .The detailed design and operation of its individual sub blocks is described. Also the advantages of DDR4 over DDR3 SDRAM are discussed.**

***Index Terms*—DDR4, SDRAM Controller, Rank.**

I. Introduction

DDR4 SDRAM, an abbreviation for Double Data Rate type 4 Synchronous Dynamic Random Access Memory is the latest member of DDR family of technologies mostly for the computing, server, biomedical and embedded applications. Recently DDR3 SDRAM is widely used in PC, Smart phones, Tablets, High-end servers and micro server systems. DDR4 is focusing on data reliability. A large number of server systems are required for increasing high-bandwidth and high capacity of multimedia contents [1]. Hence, a main memory of server systems is one of the vital components, so it should definitely have high performance feature [2]. The increasing demand of faster operating systems as well as big size applications created the need of using high capacity memory. Due to the use of high capacity memory, power consumption is also increased immensely which is one of the major concerns for memory market. DDR4 not only increased its volume but also reduced the power consumption.

The main changes of DDR4 SDRAM are supply voltage, higher data rate, higher densities, more banks, faster bursts access and higher system reliability compared with DDR3 SDRAM.

Figure 1 shows the new features which is introduced in DDR4 SDRAM such as : CRC, C/A parity, data bus inversion (DBI), Command address latency (CAL), per-DRAM addressability , CAL (command address latency), MPR (multi-purpose registers), gear-down mode, FGREF (fine granularity refresh) and temperature controlled refresh. Among these, CRC, C/A parity, data bus inversion (DBI) for solving the data reliability issue. Using Per DRAM addressability mode only selected DRAM programmed is possible in a single Rank.

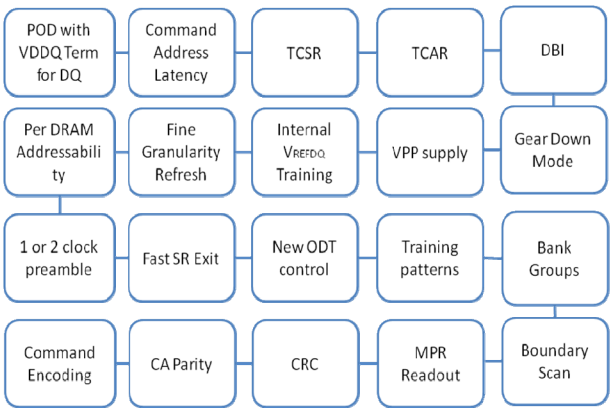


Fig.1 New Features of DDR4 SDRAM

Now Table 1 shows the comparison of DDR4 SDRAM than others.

TABLE 1
COMPARISON TABLE OF DDR3 AND DDR4 SDRAM

Feature	DDR3	DDR4	DDR4 Advantage
Voltage (core and I/O)	1.5V	1.2V	Reduces memory power demand
Prefetch	8n	8n	DDR4 also have 8n prefetch but parallel bank group
Densities	512Mb–8Gb	2Gb–16Gb	Better enablement for large-capacity memory subsystems
Data rate (MT/s)	800, 1066, 1333, 1600, 1866, 2133	1600, 1866, 2133, 2400, 2667, 3200	Migration to higher data bandwidth
Internal banks	8	16	More Banks
Bank groups (BG)	0	4	Faster burst accesses

DDR4 SDRAM controller is used as a bridge to interface with SDRAM memory devices and processors subsystem. It manages bi-directional data flow of memory. The main function of controller is data read and write. It also assists the DRAM to retain the data such as periodic refresh, precharge etc. Data transfer is accomplished with the bi-directional differential data strobe (DQS, DQS#). The DDR4 SDRAM memory device transmits data strobe (DQS) for read operation and controller device transmits DQS for write operation.

This paper is divided by following sectors: sector 1 is introduction. Sector 2 is discussed the functional description of DDR4 SDRAM. Sector 3 is described the overall architecture of DDR4 SDRAM controller and finally the conclusion sector.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

II. FUCNTIONAL DESCRIPTION OF DDR4 SDRAM

DDR4 SDRAM has 8n prefetch with parallel bank for higher data transfer rate. The memory internally configured as 16 banks, 4 bank group with 4 banks for each bank group [4]. Bank is independent memory array. The row and column address is used to locate the requested address of bank in memory. And Page is unique address comprise of bank group, bank and row address. A page size is a total number of columns in a row. Rank architecture is very important to enhance data rate without increasing burst length [3]. Figure 2 shows the organization of memory.

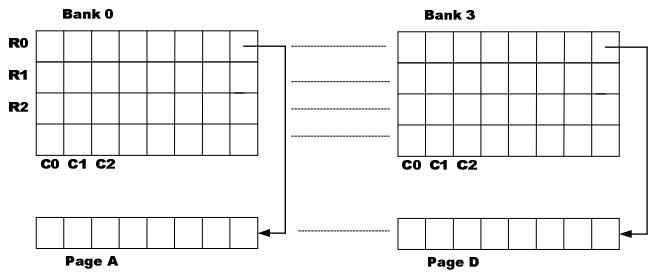


Fig. 2 RAM organization

DDR4 SDRAM read and writes operation is burst oriented. It begins at a specific selected location, and continues for a burst length of eight in a programmed sequence. The memory starts operation begins with an ACTIVATE Command, which is then followed by a Read or Write command. To select the bank and row to be activated, address bits registered is used with the ACTIVATE Command. Also to select the starting column location for the burst operation, address bits registered is used with the Read or Write command [4].

The following state diagram shows in figure 3 is described the possible state transitions and commands to control DDR4 SDRAM [4].

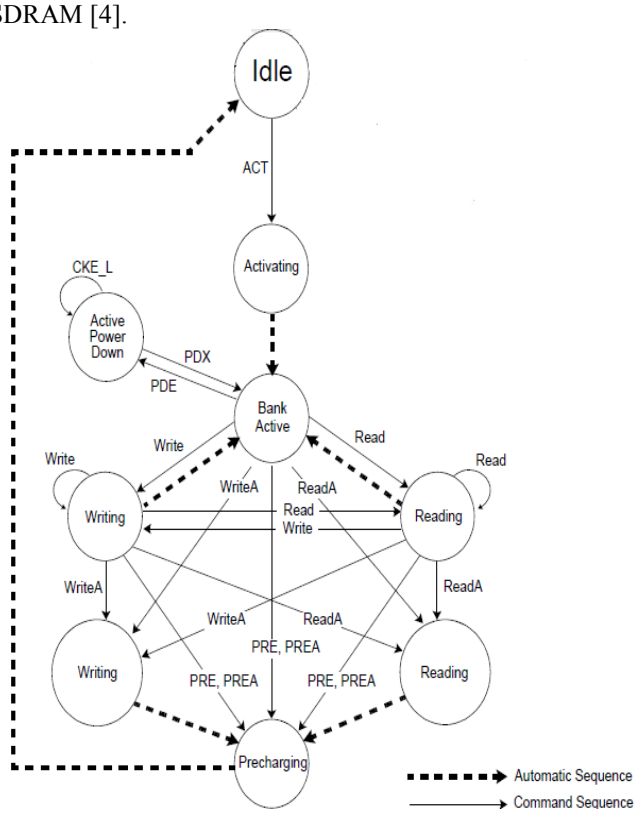


Fig. 3 State diagram of DDR4 SDRAM

III. ARCHITECTURE OF DDR4 SDRAM CONTROLLER

Fig. 4 shows the proposed overall architecture of DDR4 SDRAM controller.

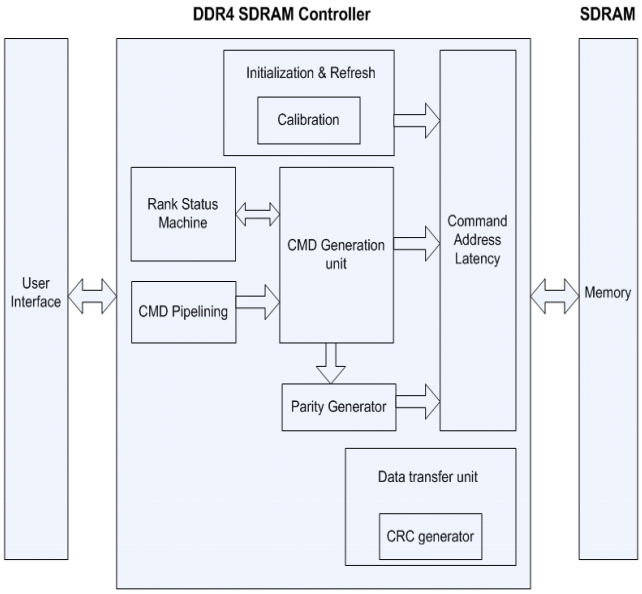


Fig. 4 Architecture of DDR4 SDRAM Controller

A. *Rank status machine*

In the controller we propose a Bank management unit (we called Rank status machine) which tracks the present state of each memory bank in the original DRAM. This provides greater flexibility to perform parallel read write command processing. The rank status machine architecture is shown in Fig.5.

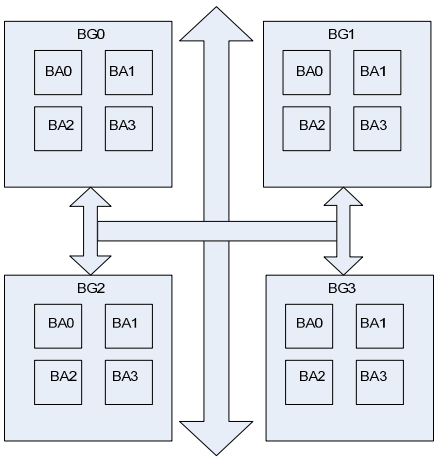


Fig. 5 Rank architecture

There are 4 Bank groups status processor and each Bank group contains 4 banks status processor, so total 16 banks can be managed at one time. Each status processor contains specific timing counter such as tCCD, tRRD, tFAW etc. The status is: bank active, readable, writeable, prechagable and row address. This ensures read-to-write and write-to-read bus turnaround time; active-to-precharge, activate delay between different banks and bank groups. When a command is sent to DDR4; timing constraint's counters corresponding to that command start to count and updated the status upon counting. For an example if a bank "x" in DDR4 is in idle state and a activate command is issued to the DDR4. The timing constraints related to activate command are tRRD (active-to-active command delay), tRCD (active-to-read/write delay),

tRAS (active-to-precharge delay) etc. In rank status machine corresponding bank "x" status processor starts counting of these times i.e. tRRD, tRCD, tRAS etc. after issuing activate command. At beginning the bank "x" status was bank active =0, readable =0, writeable =0, prechagable =0. After the tRCD time counts bank "x" status updated as bank active =1, readable =1, writeable =1, prechagable =0.

DDR4 introduces new specifications for read to read or write to write delay which are called tCCD_S and tCCD_L time for read or writes to banks of different Bank group and same Bank groups respectively.

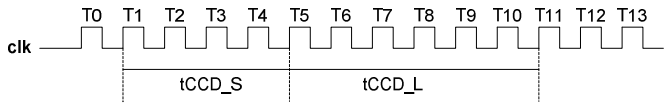


Fig. 6 tCCD variation of rank operation

The full bandwidth of DDR4 cannot be used, until Bank group interleaving access is not employed [5]. To achieve the better performance the controller can performs the bank interleaving operations by monitoring rank status.

B. CMD pipelining

Today's most of system shares the common RAM (Such as multiple processor, GPU, GPIO shares the DRAM). So the controller queues up multiple commands from different peripherals using command pipeline. The command pipeline is used for optimal bandwidth utilization. Command pipeline is used by the Command generation unit to pre-process the target bank for read or write operation such as active and precharge.

C. Command address latency

Command-to-Address Latency (CAL) function offers by DDR4 memory that can be used to save power. Basically, CAL is the delay between enable a Chip select and command/Address in clock cycles which is defined by mode register. Before issuing a command, it provides the DRAM time to active the CMD/ADDR receivers. The receivers can be latched, whenever the command and the address are latched. In case of providing consecutive commands, the receiver will have enabled by DRAM for the period of command sequence [4].

D. Calibration Unit

Calibration Unit performs write leveling, read leveling and ZQ calibration. DDR4 memory used fly-by topology for better signal integrity. To reduce the number of stubs and their length this topology is used. Despite benefiting, it causes skew problem between clock and strobe at every DRAM on the DIMM module. Hence, to solve the skew problem memory controller uses the 'write leveling' feature and feedback from the DDR4 SDRAM to adjust the data strobe to clock relationship [4]. This feature can compensate for unbalanced loading on the board for write and read operations. Calibration unit put the DDR4 into write leveling state by writing mode register MR1 of DDR4. The write strobe DQS is repeatedly delayed in small increments by the Delay Locked Loop (DLL). DDR4 samples the CLK with the rising edge of DQS and provide feedback on DQ. During this protocol, each set of DQ is output with a "0" until the rising edge is detected by the DQS at which time the DQ will be output with a "1". The calibration unit will detect these "1" on the DQ bus and then knows the correct DQS compensation to align the DQS and CLK on the write path. Once all DQS have been adjusted, these compensation values of DLL stored for each DQS for future usage. Then the memory controller sends another MRS command to exit the write level mode.

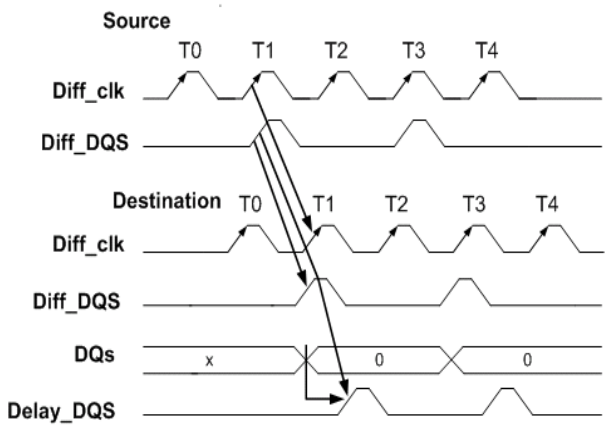


Fig. 7 Write Leveling

As DDR4 write leveling manages the DQS/DQ on write data, the DDR4 read leveling manages the DQS/DQ on read data. DDR4 read leveling is to compensate the imbalanced loading on the read path. First the memory controller puts the DDR4 memory devices into MPR mode by writing to the MR3 register. This puts the DDR4 memory devices into the read leveling mode which outputs a training pattern of consecutive "01" on each memory read command. Since the memory controller knows that data stream, so it will adjust the internal DQS delay to capture the DQ using DQS. Memory controller repeats these calibration reads until read data capture at memory controller is optimized.

ZQ Calibration command is used to calibrate DRAM Ron & ODT values. The ZQ calibration accounts the voltage and temperature variation on DRAM output driver.

E. Command Generation Unit

The command generator accepts user commands from the CMD pipelining unit. When the unit receives a single request, from pipelining unit request is processed immediately. For specified read or write request, Command Generation Unit lookup in the rank status to ensure the timing constraints for each previous memory transaction is met up before start new transaction. It also checks for availability of target bank for read or write operation from the rank status. Bank availability means the requested row in corresponding bank is active. If the requested bank is not active then it generates the necessary precharge and active command to activate the bank in DDR4. The command generation unit halt itself from throwing next command until the timing requirements of DDR4 are met. Upon availability of the bank; the requested read or write command is send to the DDR4 and internal rank status machine so that rank can update itself each time when a new command is throwing to the DDR4. If there are outstanding requests from two or more queues, the Command generation unit process the requests in look-ahead way. While processing

the active request from pipeline if there are any unsatisfied timing constraints which lead the controller to stall the transaction, controller then check the next upcoming request from queue. Command generation unit check the next requests bank address to find out the upcoming requests bank whether it is in active and activated row does match with the requested row address or not. If the condition does not met then command generation unit perform necessary precharge activate sequence to active that bank, so that when the request will be active the request will incur minimum stall time. Such deliberately bank active of future requests utilize the active requests stall time. This opportunistically look-ahead activates improve overall throughput.

F. Data Transfer Unit

Data transfer unit consist a read FIFO and a write FIFO. The read FIFO receives data from the DDR4 and passes that data to the user. The write FIFO receives write data from the user interface and passes that data to the DDR4. Read data from DDR4 arrive at the controller at both positive and negative edge of clock. The burst length of DDR4 is 8. Read FIFO captures these data in 4 clock cycles and provides the user read data in a single clock cycle on next positive edge of clock. Similarly write FIFO loads the user write data in a single clock and send to the DDR4 in next 4 clocks both at positive and negative edge. During the write data transfer to DDR4 it also added up CRC code at trail of each burst.

G. CRC Generator

As speed increased data prone to error. To solve Data reliability issue during WRITE operations, DDR4 included Cyclic Redundancy Check (CRC) along with DBI. DDR4 uses an 8-bit CRC header error control. The CRC polynomial used by DDR4 is the ATM-8 HEC, X^8+X^2+X^1+1.

Figure 8 shows CRC generator. DRAM generates checksum in each write burst per Data Strobe lane. CRC use 72 bits of data and all unallocated bits are 1s. DRAM compares the checksum. If two checksums do not match then DRAM flags an error.

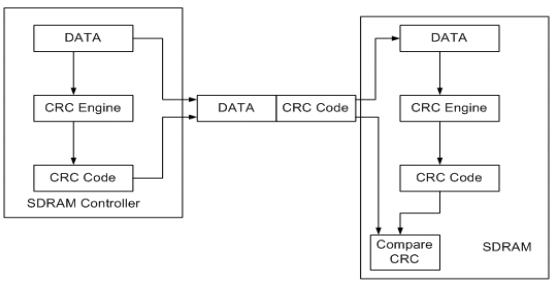


Fig. 8 CRC Generator

H. Initialization and Refresh Unit

Initialization and Refresh Unit performs the power up initialization and periodic refresh. After power up or reset the DDR4 must be initialized before start read or write transaction. The reset pin of DDR4 is kept low by the controller for 100ns to begin the initialization process. To active DDR4 clock, reset pin set to high and CKE pin set to low for 5100us. Then CKE pin drive to high. After tXPR period of time when DRAM is initialized Mode Register Set commands are issued for proper mode of operation such as burst length, additive latency, cas latency etc. After setting the mode register the Initialization and refresh unit switch on the calibration unit to perform the read and write leveling. Upon completion of coalition the DDR4 is ready for normal operation. It has a tREFI counter, which counts DDR4 refresh interval. On each tREFI count the controller issues refresh command to DDR4.

IV. CONCLUSIONS

The proposed architecture is designed and implemented according to DDR4 SDRAM JEDEC specification. DDR4 has reduced power consumption whereas it transfer rate is more above of its predecessor i.e. DDR3, DDR2, DDR. Its application will be more widely used from low power mobile computing device to high density servers. To enhance the throughput and reliability the controller should be aware of both parallel bank operation and error detection. For that reason we proposed parallel bank group and bank status machine co-ordinated by rank and separate data transfer unit with CRC generator. For better performance upon notifying error status from DDR4 the controller should be able to handle the ongoing transaction and suspended transaction traffic. This is the upcoming challenge for DDR4 controller which we will try in future.

REFERENCES

[1] R. Ramakrishnan, “CAP and cloud data management,” *Computer*, vol. 45, no. 2, pp. 43–49, Feb. 2012.

[2] M. E. Tolentino, J. Turner, and K. W. Cameron, “Memory MISER: Improving main memory energy efficiency in servers,” *IEEE Trans. Comput.*, vol. 58, no. 3, pp. 336–350, Mar. 2009.

[3] T.-Y. Oh *et al.*, “A 7 Gb/s/pin GDDR5 SDRAM with 2.5 ns bank-to bank active time and no bank-group restriction,” in *IEEE ISSCC Dig. Tech. Papers*, 2010, pp. 434–435.

[4] *DDR4 SDRAM Specification (JESD79-4)*, JEDEC Standard, JEDEC solid state technology association, September 2012.

[5] Kyomin Sohn, Taesik Na, Indal Song, Yong Shim, Wonil Bae, Sanghee Kang, Dongsu Lee, Hangyun Jung, Seokhun Hyun, Hanki Jeoung, Ki-Won Lee, Jun-Seok Park, Jongeun Lee, Byunghyun Lee, Inwoo Jun, Juseop Park, Junghwan Park, Hundai Choi, Sanghee Kim, Haeyoung Chung, Young Choi, Dae-Hee Jung, Byungchul Kim, Jung Hwan Choi, Seong-Jin Jang, Chi-Wook Kim, Jung-Bae Lee, and Joo Sun Choi “A 1.2 V 30 nm 3.2 Gb/s/pin 4 Gb DDR4 SDRAM With Dual-Error Detection and PVT-Tolerant” *IEEE JOURNAL OF SOLID-STATE CIRCUITS*, VOL. 48, NO. 1, JANUARY 2013.

A Graph Rewriting Approach to Replace Asynchronous RAMs in Circuits with Cycles for FPGAs

Md. Nazrul Islam Mondal[1*] Md. Shahid Uz Zaman[2] and Biprodip Pal[3]
[1,2,3] Dept. of Computer Science and Engineering, Rajshahi University of Engineering and Technology, Bangladesh
*nimbd@yahoo.com

***Abstract*—Circuit design that minimizes the number of clock cycles is easy if we use asynchronous read operations. However, most of FPGAs support synchronous read operations, but do not support asynchronous read operations. It is one of the main difficulties for users to implement parallel and hardware algorithms in FPGAs. The main contribution of this paper is to provide one of the potent approaches to resolve this problem. We assume that a circuit which includes cycles using asynchronous RAMs designed by a non-expert or quickly designed by an expert is given. Our goal is to convert this circuit with asynchronous RAMs into an equivalent synchronous ones. The resulting circuit with synchronous RAMs can be embedded into the FPGAs.**

***Index Terms*—FPGA, Block RAMs, asynchronous read operations, rewriting algorithm.**

I. Introduction

An *FPGA*is a programmable VLSI (Very Large Scale Integration) in which a hardware designed by users can be embedded quickly. Typical FPGAs consist of an array of programmable logic blocks (slices), memory blocks, and programmable interconnects between them. Some circuit implementations using FPGAs are described in the papers [4], [5], [6], [7] to accelerate computation.

In our previous paper [1], we have presented a circuit rewriting approach for circuits represented by a directed acyclic graph (DAG), which has no directed cycle. More specifically, presented approach presented in the paper [1] has a strict restriction in terms of input circuits. It works only for a circuit whose underlying graph is a DAG. Although most of practical circuits have cycles, it can not handle such circuits.

To resolve this problem, this paper modifies the circuit rewriting algorithm in the paper [1] to process practical circuits with cycles. More specifically, our new circuit rewriting algorithm can convert any circuit represented by a directed reachable graph (DRG) having cycles.

The main contribution of this paper is to present a circuit rewriting approach that converts *an asynchronous circuit* consisting

> *Combinational Circuits (CCs), Registers (Rs), and RAMs with asynchronous read operations (ARAMs)*

into *an equivalent synchronous circuit* consisting

> *Combinational circuits (CCs), Registers (Rs), and RAMs with synchronous read operations (SRAMs).*

The outlines of our work are as follows:

1) We use *a negative register* (NR) which is originally introduced in our previous paper [2].
2) We define simple *six rules* that rewrite a circuit.
3) The rewriting algorithm just repeats applying these rules until no more rules can be applied. When the rewriting algorithm terminates, we have an equivalent ARAM-free circuit to the original circuit.

The paper organization is described to the following Sections.

II. Random Access Memory (RAM)

A RAM is an array of memory where information can be stored until power is switched off. It has b-bit data input D, e-bit address input A and c-bit data output Q, it can store 2^e words such as $M[0]$, $M[1]$, ..., $M[2^e - 1]$ with c bits each, shown in Fig 1.

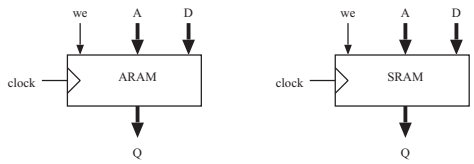


Fig. 1. An ARAM and an SRAM.

A RAM can support Asynchronous read, Synchronous read and Synchronous write operations as illustrated details in our previous paper [1].

For the reader's benefit, we need to describe two types of RAM. However, we refer readers to our previous paper [1] for the page limitation.

In this work, we consider Write After Read (WAR) mode for data handling of the memory blocks. The Fig 2 for the illustrations of the WAR and RAW. The Fig. 2 shows a timing diagram of the ARAM, SRAM, Register (R) and Negative Register (NR). Initially global reset is 1 and it drops to 0 just before time 0. We assume that write enable *we* is high for the first several clock cycles from the beginning (initially *we* is 1 and drops to 0 before the fourth rising clock edge). Data 11, 12, 13, −, −, − and 1, 2, 3, 1, 2, 3 are given to the input data port D and address port A respectively. The dash (-) line represents any data which is not necessary in our case. For simplicity, we have used the written data values of $M[1]$, $M[2]$, $M[3]$ for an SRAM to read again at the latest rising clock edge at time 3, 4, 5 respectively. Note that, M[1], M[2] and M[3] represent data value of memory location 1, 2 and 3 respectively. We have also used the written data values of $M[1]$, $M[2]$, $M[3]$ for an ARAM to read immediately at time 3, 4, 5 respectively. The edges at time 0, 1, 2 of the clock represent the latest rising edges for the stored data values of $M[1]$, $M[2]$, $M[3]$ respectively of an SRAM and ARAM. On

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

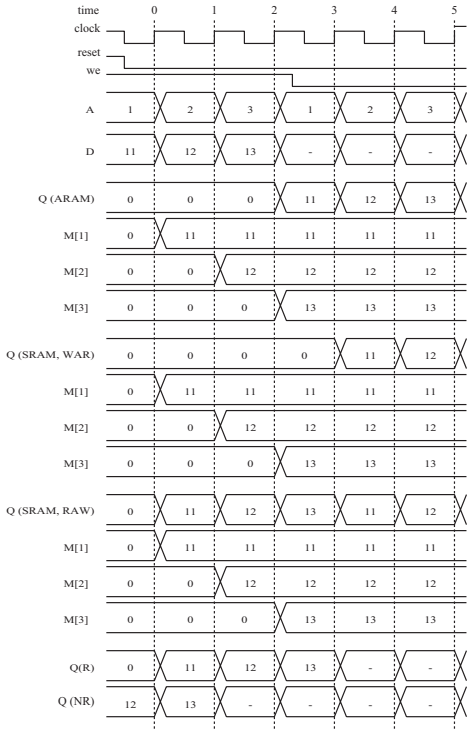


Fig. 2. A timing chart of an ARAM, an SRAM, a register (R) and a negative register (NR).

the other hand, the edges at time 0, 1, 2, 3, 4 of the clock represent the latest rising edges for the output data sequence of 0, 0, 0, 11, 12 respectively of an SRAM when it follows WAR. We assume that the stored values of M[a] are initialized by 0 of an SRAM and an ARAM. For the case of an ARAM, the data values of $M[1]$, $M[2]$, $M[3]$, $M[1]$, $M[2]$, $M[3]$ correspond to 0, 0, 0, 11, 12, 13 are taken respectively at time 0, 1, 2, 3, 4, 5 from the output port Q immediately due to the asynchronous read operations. Therefore, the output sequence of an ARAM, $[Q(ARAM)]$ is: 0, 0, 0, 11, 12, 13. According to the WAR, the output sequence of an SRAM, $[Q(SRAM, WAR)]$ is: 0, 0, 0, 0, 11, 12 at time 0, 1, 2, 3, 4, 5 respectively. On the other hand, according to the RAW, the output sequence of an SRAM, $[Q(SRAM, RAW)]$ is: 0, 11, 12, 13, 11, 12 at time 0, 1, 2, 3, 4, 5 respectively. Note that the output value of an SRAM at time 0 is initialized by 0. The stored data of $M[1]$, $M[2]$ and $M[3]$ of the time 0, 1, 2, 3, 4, 5 are the same for an ARAM and an SRAM which are $M[1]$: 0, 11, 11, 11, 11, 11; $M[2]$: 0, 0, 12, 12, 12, 12 and $M[3]$: 0, 0, 0, 13, 13, 13. The output of R is 0 at time 0. Also at time 1, 2, 3, 4, 5; the value of output R is 11, 12, 13, -, - respectively. The value of output NR is: 12, 13, $-$,$-$, $-$, $-$ of the time 0, 1, 2, 3, 4, 5 respectively.

III. Circuits and Their Equivalence

Let us consider a synchronous sequential circuit that consists of input ports, output ports, combinational circuits (CCs), registers (Rs), Random Access Memories (RAMs), a global clock input (clock), a global reset input (reset) and a write enable input (we).

Since there is a page limitation, we refer readers to our previous paper [2], [1] for details about the behavior of combinational circuits (CCs) and Registers (Rs), SRAM, ARAM and circuit equivalence.

In this paper, we assume that a fully synchronous circuit as illustrated in Fig. 3. We also assume that a circuit has loop or cycle.

Now we are in position to describe the equivalence of two circuits. Because of page limitation, we refer readers to our previous paper [1] for details about it.

Suppose that a circuit X with ARAMs is given. The main contribution of this paper is to show

- a necessary condition such that an ARAM-free circuit, Y can be generated, which is equivalent to X, and
- an algorithm to derive Y if the necessary condition is satisfied.

IV. Circuit Graph and Rewriting Rules

We simply use a directed graph to denote the interconnections of a fully synchronous circuit. We call such graph *as a circuit graph*. A circuit graph consists of a set of nodes and a set of directed edges connecting two nodes. Each node is labeled by either I (Input port), O (Output port), CC (Combinational Circuit), R (Register), NR (Negative Register), ARAM, or SRAM. The Fig. 3 illustrates an example of a directed graph.

Let us define rules to rewrite a circuit graph. The readers should refer to Fig 4 for illustrating the rules, where P and S denote predecessor and successor nodes respectively. The nodes between predecessor and successor nodes are rewritten as follows:

Rule 0 ARAM node is rewritten into SRAM+NR.

Rule 1 Adjacent R and NR nodes are rewritten into NULL circuit, that is, they are removed.

Rule 2 R+SRAM is rewritten into SRAM+R. More specifically, if each incoming edge of an SRAM node is connected to a R node, then all the Rs are removed to the outgoing edge of the SRAM node.

Rule 3 If all the incoming edges of a CC node are connected to a R node, then all the Rs are removed to all the outgoing edges of the CC node.

Rule 4 NR+SRAM is rewritten into SRAM+NR. More specifically, if one of the incoming edges of an SRAM node is connected to an NR node, then the NR node is removed, a R node is added to all other incoming edges, and the NR node is moved to the outgoing edge of the SRAM node.

Rule 5 If one of the incoming edges of a CC node is connected to an NR node, then the NR node is removed, a R node is added to all the other incoming edges, and the NR node is moved to all the outgoing edges of the CC node.

Let us confirm that, after applying one of the rewriting rules, an original circuit and the resulting circuit are an equivalent. Let a_i, b_i, c_i, d_i, a_i (address data) and *we* ($i \geq 0$) denote inputs given from the predecessor node at time i.

Rule 0 Both ARAM and SRAM+NR have the same output sequences and stored data.
ARAM/SRAM+NR(output sequence): $\langle 0, 0, 0, 11, 12, 13, \rangle$
ARAM/SRAM+NR (stored data of M[1], M[2] and M[3]):

- **M[1]:** $\langle 0, 11, 11, 11, 11, 11, \rangle$
- **M[2]:** $\langle 0, 0, 12, 12, 12, 12, \rangle$ and

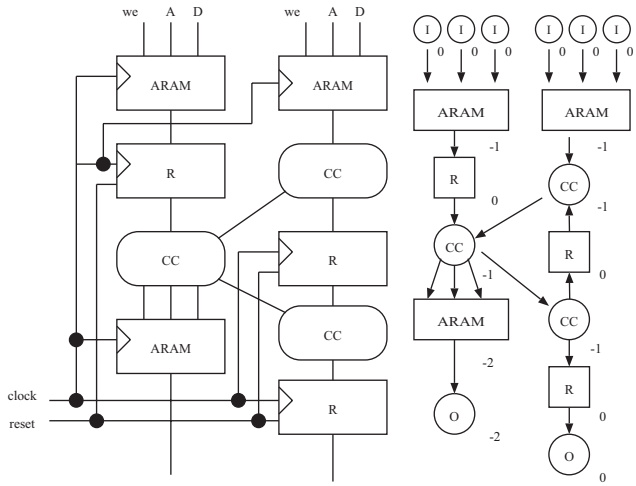


Fig. 3. An example of a fully synchronous circuit and the corresponding circuit graph with potentiality.

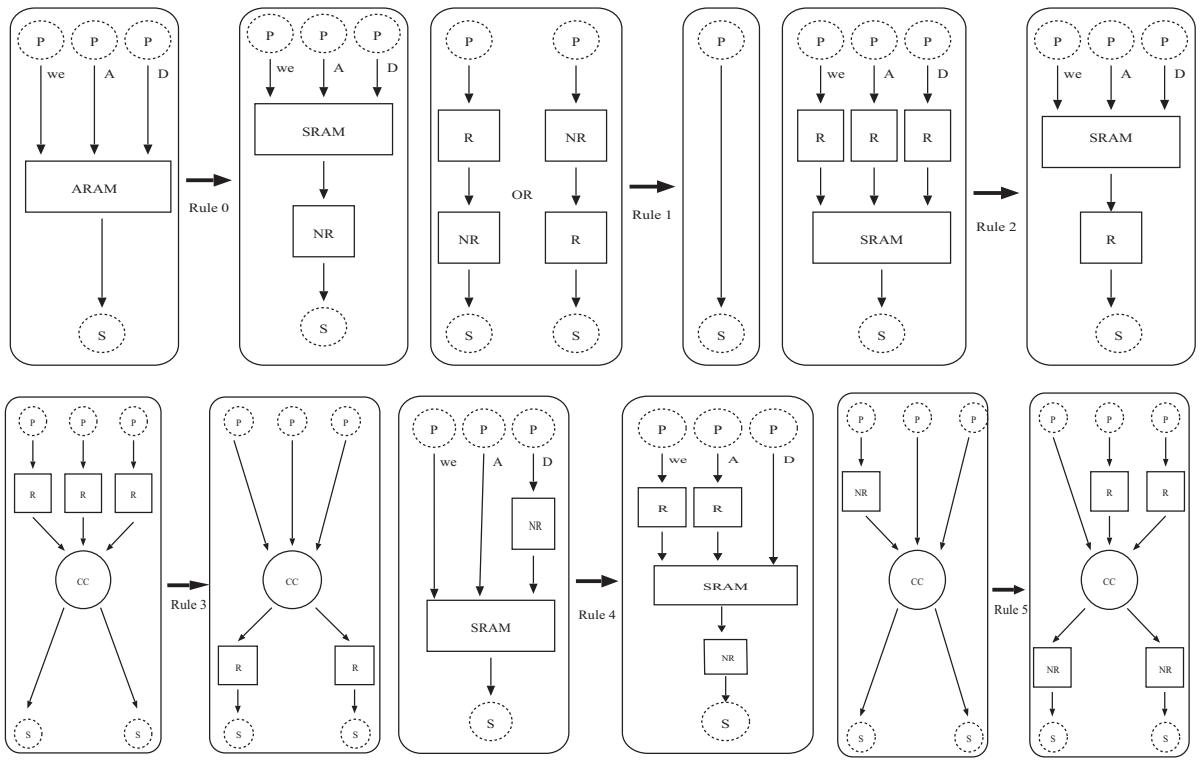


Fig. 4. Rules to rewrite a circuit graph.

- **M[3]:** $\langle 0, 0, 0, 13, 13, 13, \rangle$.

Thus they are an equivalent.

Rule 1 R+NR and NR+R have the following output sequences $\langle 11, 12, 13, -, -, -, \rangle$ and $\langle 0, 12, 13, -, -, -, \rangle$, respectively. Also, NULL circuit has the following output sequence $\langle 11, 12, 13, -, -, -, \rangle$. Thus, they are an equivalent.

Rule 2 R+SRAM has the following output sequence and stored data.

R+SRAM(output sequence): $\langle 0, 0, 0, 0, 0, 11, \rangle$

R+SRAM(stored data of M[1], M[2] and M[3]):

- **M[1]:** $\langle 0, 0, 11, 11, 11, 11, \rangle$
- **M[2]:** $\langle 0, 0, 0, 12, 12, 12, \rangle$ and
- **M[3]:** $\langle 0, 0, 0, 0, 13, 13, \rangle$.

SRAM+R has the following output sequence and stored data.

SRAM+R(output sequence): $\langle 0, 0, 0, 0, 0, 11, \rangle$

SRAM+R(stored data of M[1], M[2] and M[3]):

- **M[1]:** $\langle 0, 11, 11, 11, 11, 11, \rangle$
- **M[2]:** $\langle 0, 0, 12, 12, 12, 12, \rangle$ and
- **M[3]:** $\langle 0, 0, 0, 13, 13, 13, \rangle$.

Thus they are an equivalent.

Rule 3 The output sequences of the left-hand side of the rule are $\langle f(a_1, b_0, c_0),\ f(a_2, b_1, c_1),\ f(a_3, b_2, c_2),\ \ldots \rangle$ and $\langle g(a_1, b_0, c_0),\ g(a_2, b_1, c_1),\ g(a_3, b_2, c_2),\ \ldots \rangle$. Those of the right-hand side are $\langle f(a_1, b_0, c_0),\ f(a_2, b_1, c_1),\ f(a_3, b_2, c_2),\ \ldots \rangle$ and $\langle g(a_1, b_0, c_0),\ g(a_2, b_1, c_1),\ g(a_3, b_2, c_2),\ \ldots \rangle$. Thus, they are an equivalent.

Rule 4 NR+SRAM has the following output sequence and

stored data.
NR+SRAM(output sequence): $\langle 0,0,0,0,12,13,\rangle$
NR+SRAM(stored data of M[1], M[2] and M[3]):

- **M[1]:** $\langle 0,12,12,12,12,12,\rangle$
- **M[2]:** $\langle 0,0,13,13,13,13,\rangle$ and
- **M[3]:** $\langle 0,0,0,-,-,-,\rangle$.

SRAM+NR has the following output sequence and stored data.
SRAM+NR(output sequence): $\langle 0,0,0,0,12,13,\rangle$
SRAM+NR(stored data of M[1], M[2] and M[3]):

- **M[1]:** $\langle 0,0,12,12,12,12,\rangle$
- **M[2]:** $\langle 0,0,0,13,13,13,\rangle$ and
- **M[3]:** $\langle 0,0,0,0,-,-,\rangle$.

Thus they are an equivalent.
For the Rule 4, we consider that an NR is connected to data input port, *D* only. If an NR is connected to the address input port, A or write enable input, *we*, an equivalency of the Rule 4 can be shown in the similar way.

Rule 5 The output sequences of the left-hand side of the rule are $\langle f(0,0,0),\ f(a_0,b_0,c_0),\ f(a_1,b_1,c_1),\ \ldots\rangle$ and $\langle g(0,0,0),\ g(a_0,b_0,c_0),\ g(a_1,b_1,c_1),\ \ldots\rangle$. Those of the right-hand side are $\langle 0,\ f(a_0,b_0,c_0),\ f(a_1,b_1,c_1),\ \ldots\rangle$ and $\langle 0,\ g(a_0,b_0,c_0),\ g(a_1,b_1,c_1),\ \ldots\rangle$. Thus, they are an equivalent.

We are now in position to describe the rewriting algorithm. Suppose that an input circuit graph has nodes with labels *I*, *O*, *R*, *ARAM*, *SRAM*, and *CC*. The following rewriting algorithm generates a circuit graph equivalent to the original circuit graph.

> *Find a minimum i such that Rule i can be applied to the current circuit graph. Rewrite the circuit graph using such Rule i. This rewriting procedure is repeated until no more rewriting is possible.*

V. Behavior of Our Circuit Rewriting Algorithm

Let us observe the behavior of the rewriting algorithm.

- First, the rewriting algorithm repeats the applying Rule 0 to all ARAM nodes until all ARAM nodes are rewritten into SRAM+NR. After that, Rule 0 is never be applied.
- Rule 1 is applied and adjacent R and NR nodes are removed whenever possible.
- R or NR node is moved toward the output nodes using Rule 2, Rule 3, Rule 4 and Rule 5 whenever possible.

In our previous paper [1], we have shown one of the applications of our rewriting algorithm. Since application example is similar for rewriting algorithm, presented in this paper, hence we are not incorporating it.

For the purpose of clarifying the condition such that the rewriting algorithm can generate NR-free circuit graph. We define *the potentiality of the nodes* in a circuit graph. Suppose that a node v of a circuit graph has k (≥ 0) incoming edges such as $(u_1,v),(u_2,v),\ldots,(u_k,v)$. Let us define *the potentiality* $p(v)$ of a node v as follows:

- If v is I, then $p(v)=0$.
- If v is O, then $p(v)=p(u_1)$.
- If v is SRAM, then $p(v)=\min(p(u_1),p(u_2),p(u_3))$
- If v is ARAM, then $p(v)=\min(p(u_1),p(u_2),p(u_3))-1$.
- If v is NR ,then $p(v)=p(u_1)-1$.
- If v is R then $p(v)=p(u_1)+1$.
- If v is CC, then $p(v)=\min(p(u_1),p(u_2),\ldots,p(u_k))$.

The Fig 3 shows the potentiality of each node. We have the following theorem.

Theorem 1: All O nodes and all cycles of a circuit graph have non-negative potentiality, if and only if our rewriting algorithm generates an ARAM-free and NR-free circuit graph, equivalent to the original circuit graph.

In other words, we can determine a fully synchronous circuit that can be converted into an ARAM-free circuit by evaluating the potentiality of all O nodes and all cycles of the corresponding circuit graph. Also, the potentiality of all O nodes and all cycles are non-negative, our rewriting algorithm generates an ARAM-free and NR-free circuit graph, and the corresponding fully synchronous circuit is ARAM-free and equivalent to the original fully synchronous circuit.

VI. Proof of Theorem 1

The main purpose of this section is to show a proof of Theorem 1. Hence, we need to show several lemmas for a proof of Theorem 1. However, we are not incorporating the proof of Theorem 1 in this paper because of the page limitation. The reader may think that this paper is useless without proof of Theorem 1. In this case, we refer the readers to the proof part of our previous papers [3], [1] which is the similar for this paper.

VII. Conclusions

In this paper, we have presented a rewriting algorithm and six rewriting rules to obtain the equivalent circuits with Synchronous RAMs (SRAMs) for the circuits with Asynchronous RAMs (ARAMs) including cycles. Using our rewriting algorithm, any sequential circuit with ARAMs represented by a directed reachable graph (DRG) can be converted into an equivalent fully synchronous sequential circuit with no ARAMs to support the current FPGA architecture. Note that, rewriting algorithm presented in our previous paper [1] was only capable to convert DAG circuits (no cycle) with ARAMs into an equivalent circuit with SRAMs for implementing in modern FPGAs. Hence, there was a strict limitation to handle practical circuits with cycles. However, our algorithm presented in this paper can do it automatically.

References

[1] M.N.I Mondal, K. Nakano, and Y. Ito, "An Algorithm to Obtain Circuits with Synchronous RAMs," *Int'l Journal of Communication and Computer (JCC)*, vol. 9, pp. 547–559, May 2012.

[2] M.N.I Mondal, K. Nakano, and Y. Ito, " A Graph Rewriting Approach for Converting Asynchronous ROMs into Synchronous Ones," *IEICE Transaction on Information and Systems*, vol. E94-D, pp. 2378–2388, Dec. 2011.

[3] M.N.I Mondal, K. Nakano, and Y. Ito, " A Rewriting Approach to Replace Asynchronous ROMs with Synchronous Ones for the Circuits with Cycles," *Int'l Journal of Networking and Computing (IJNC)*, vol. 2, pp. 269–290, July 2012.

[4] J. Bordim, Y. Ito and K. Nakano, "Accelerating the CKY parsing using FPGAs,"*IEICE Transactions on Information and Systems*, vol. E86-D, pp. 803–810, Dec. 2003.

[5] J. Bordim, Y. Ito and K. Nakano, "Instant specific solutions to accelerate the CKY parsing for large context-free grammars,"*International Journal on Foundations of Computer Science*, pp. 403–416, Dec. 2004.

[6] K. Nakano and Y. Yamagishi, "Hardware n choose k counters with applications to the partial exhaustive search," *IEICE Transactions on Information and Systems*, 2005.

[7] Y. Ito and K. Nakano, "A hardware-software cooperative approach for the exhaustive verification of the collatz conjecture,"*In Proc. of International Symposium on Parallel and Distributed Processing with Applications*, pp. 63–70, 2009.

Resonant Frequency Analysis of Graphene Nanoribbon Based VLSI Interconnect System

Asif Hassan[1, *], Turja Nandy[2], Minhaz Ibne Abedin[3], Md. Akramul Islam[4], Arin Dutta[5]
Elcetronics and Communication Engineering, Khulna University of Engineering and Technology, Bangladesh[1]
Applied Physics, Electronics and Communication Engineering, University of Dhaka, Bangladesh[2, 5]
Electrical and Electronics Engineering, Chittagong University of Engineering and Technology, Bangladesh[3, 4]
[*]asif08ece@live.com

Abstract—**Graphene nanoribbons (GNRs) are widely used as a channel of nanoscale based transistor. Now it is investigated in very large scale integration (VLSI) based interconnect system because of its higher degree of conductivity. It is seen that the RLC (resistor inductor and capacitor) equivalent circuit of GNR interconnect system has four important parameters like classical capacitance, quantum capacitance, magnetic inductance and kinetic inductance. These parameters have an important relation with the structure (here width) of GNR. Another important parameter is the resonant frequency which is normally calculated from the RLC model of a system. In this paper we have analyzed these four parameters for varying width of GNR. Then the resonant frequency response of GNR based interconnect system has been observed by varying width of GNR. Moreover, we have analyzed our work on resonant frequency by considering the two degeneracy factors of quantum capacitance.**

Keywords—Graphene nanoribbon, classical capacitance, quantum capacitanc, magnetic inductance , kinetic inductance.

I. INTRODUCTION

Graphene is a flat monolayer of carbon atoms tightly packed into two dimensional honeycomb lattice [1], has appeared as one of the most potential candidate in fundamental research [2].When it is patterned into a nanoscale ribbon then it is called graphene nanoribbon (GNRs) have been studied for interconnects [3]-[5]. It is showed that traditional copper based interconnect suffers serious problems due to increased resistivity and susceptibility to electro-migration [6]. Otherwise GNRs have larger mean free path and higher amount of thermal conductivity it lags the earlier Cu based interconnect [7].

Depending on the geometrical structure GNRs can be either semiconducting or metallic [8]. Here the basic geometrical parameter is number of dimerline (N) which is also called number of hexagonal ring in GNR. And from the equivalent circuit model of GNRs structure, it is observed that an RLC model and interconnect system is formed. It is also seen that RLC model can be splitted into some important parameters like classical capacitance, quantum capacitance, magnetic inductance and kinetic inductance Upon this GNR based interconnect the resonant frequency can be observed which will be useful hint of GNR based VLSI technology also in microprocessor.

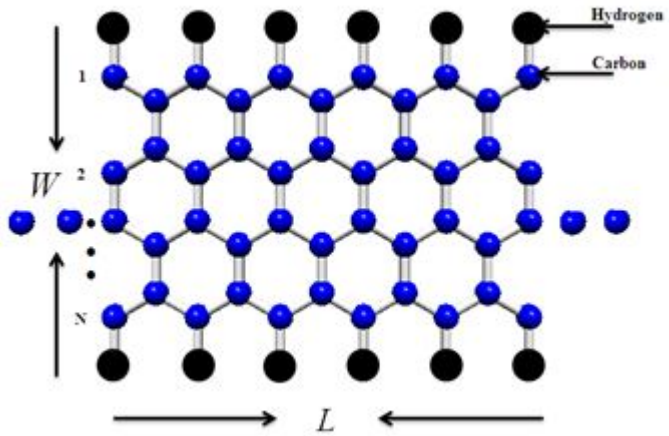


Figure 1. Schematic of a GNR. The blue circles denote carbon atoms termintaing by black circle hydrogen atoms. The number of dimmer line, width and length of GNR are represented by N , W and L respectively.

In this paper we will calculate some parameter of RLC interconnect system like resonant frequency, classical capacitance, quantum capacitance for two degeneracy factors, magnetic inductance and kinetic inductance for varying width of GNR from corresponding number of dimer line N.

II. THEORY

As stated earlier the geometrical structure which is width of GNR can be found from N. And sample which is made by this geometry shows semiconducting property in one condition and metallic property for other. Now the width of semiconducting GNR which is usually called as armchair GNR is-

$$W_{ac} = (N-1)\frac{\sqrt{3}}{2}a \qquad (1)$$

where $a = 1.42$ A^0 is the nearest neighbor distance.

The simple electrical circuit model of monolayer Graphene Nanoribbon is like RLC circuit which is the first concern in Graphene Nanoribbon based interconnect system.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

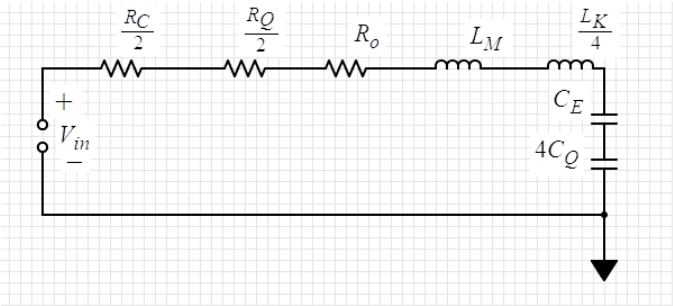


Figure 2. Schematic of a RLC model of monolayer GNR interconnect system.

In Fig. 2, R_C , R_Q and R_o denotes respectively the imperfect contact resistance quantum resistance and ohmic resistance. Also L_M , L_K denotes respectively magnetic inductance and kinetic inductance and C_E , C_Q denotes classical capacitance and quantum capacitance.

The frequency at which a system here in RLC circuit, the system or electric signal of the system oscillates with greater amplitude than that frequency is called resonant frequency of that system. From the above Fig. 2, we can calculate the resonant frequency excluding only ($\frac{R_C}{2}$, $\frac{R_Q}{2}$ and R_o) resistive part-

$$2\pi f_r L_M + 2\pi f_r \frac{L_K}{4} - \frac{1}{2\pi f_r C_E} - \frac{4}{2\pi f_r C_Q} = 0 \tag{2}$$

So

$$f_r = \frac{1}{2\pi} \frac{\sqrt{\left(\frac{1}{C_E} + \frac{4}{C_Q}\right)}}{\sqrt{L_M + \frac{L_K}{4}}} \tag{3}$$

For calculating other paprameter we have used their established relation with the width of GNR.

The classical capacitance is calculated from the device geometry like width of GNR, dielectric constant of insulator etc. Now the classical capacitance become-

$$C_E = N_G \varepsilon_o k \left(\frac{W_{ac}}{t_{ox}} + \alpha \right) \tag{4}$$

where N_G is the number of gates, k is the relative dielectric constant of the gate insulator, ε_0 is the permittivity of air, t_{ox} represents the gate insulator (oxide) thickness and $\alpha = 1$ is a dimensionless fitting parameter[9].

As the device is controlled by a source voltage the classical capacitance cannot be well enough to investigate the electronic behavior of GNR based device. So another capacitance must be included for better understanding of observing the device behavior like quantum capacitance which is [10]

$$C_Q = \frac{e^2}{3ta\pi} \left(\frac{x + \frac{E_G}{2k_B T}}{\sqrt{x + x\frac{E_G}{k_B T}}} \right) f(E) \tag{5}$$

Where

$$x = E - \frac{E_G}{2}$$

$$E_G = -\frac{2\pi(\gamma_1 - 2\gamma_3)}{\sqrt{3}(\frac{2}{\sqrt{3}}\frac{W_{ac}}{a} + 2)} + \frac{6(\gamma_3 + \Delta_{\gamma_1})}{\left(\frac{2}{\sqrt{3}}\frac{W_{ac}}{a} + 2\right)} \tag{6}$$

where γ_1 = - 3.2 eV is the first nearest neighbor hopping parameter, γ_3 = - 0.3 eV is the third nearest neighbor hopping parameter, Δ_{γ_1} = - 0.2 eV is the correction of γ_1 [11]. $k_B = 1.38 \times 10^{-23} J/K$ (Boltzmann constant), $e = 1.6 \times 10^{-19}$ Columb, T = Room temperature, t = 2.7 eV, hopping integral of perfect GNR.

$$f(E) = \frac{1}{\exp\left(\frac{E - E_F}{k_B T}\right) + 1}$$

If GNR based device is tuned by a supply voltage, degeneracy can be determined from the position of Fermi level. In degenerate regime, the Fermi level is located within or close to the conduction band so for this condition becomes $E - E_F < 3k_B T$ and therefore, the Fermi function can be approximated as $f(E)$ = 1. On the other hand, for a nondegenerate condition when the Fermi level is located inside the GNR bandgap, the condition is given by $E - E_F > 3k_B T$. So Fermi function becomes $f(E) = \exp[(E_F - E)/k_B T]$.

Now the quantum capacitance in degenerate regime can be written as-

$$C_{Q(\deg)} = \frac{e^2}{3ta\pi} \left(\frac{x + \frac{E_G}{2k_B T}}{\sqrt{x + x\frac{E_G}{k_B T}}} \right) \tag{7}$$

Another the quantum capacitance in nondegenerate regime can be written as-

$$C_{Q(non\deg)} = \frac{e^2}{3ta\pi} \left(\frac{x + \frac{E_G}{2k_B T}}{\sqrt{x + x\frac{E_G}{k_B T}}} \right) \exp(\eta) \tag{8}$$

where $$\eta = \frac{E_C - E_F}{k_B T} \quad (9)$$

Kinetic inductance arise from the energy stored in the magnetic field due to a current. Now to compute the kinetic inductance of GNR by extending the derivation of kinetic inductance of carbon nanotubes [12].The kinetic inductance in relation with GNR width is [13]-

$$L_K = \frac{\pi^3 \hbar^2 L}{E_F e^2 W_{ac}} \quad (10)$$

Where L is the length of a monolayer GNR

Otherwise magnetic inductance arises from the energy stored in the motion of the charge carriers. And for one layer of GNR the magnetic inductance becomes [7]-

$$L_M = \frac{\mu d}{W_{ac}} \quad (11)$$

Where $\mu = \mu_o = 4nH\,cm^{-1}$ is the permeability of GNR, d=5nm distance of another materials.

III. Results and Discussion

The width of the GNRs is derived from the number of dimer lines N which is obtained from equation (1) .If we look into the equation (1) it can be said easily that there is a linear relation between the dimer line and width of GNR. From this calculated width we can also measure the bandgap energy using the equation (6).Here also it can be said that there will be inverse relation between the width of GNR and bandgap energy.

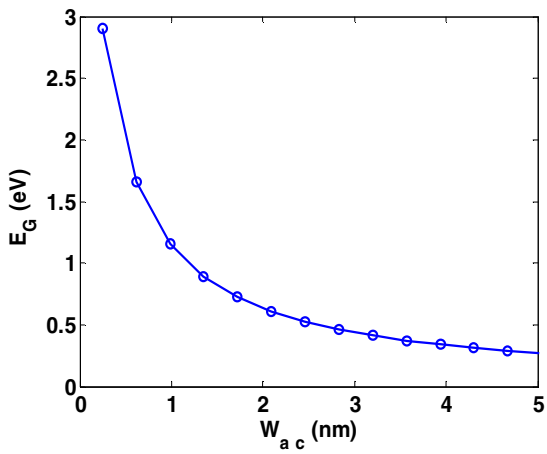


Figure 3. Bandgap Energy (E_G) vs. width (W_{ac}) of GNR .

Here, at first we can observe the band gap energy for the above calculated width of GNR which will be useful in calculating the quantum capacitance. From the above Fig.3 it is seen that there is an inverse relation between the band gap energy and width of GNR. As the width is increased the band gap is decreased.

Now we will observe the classical capacitance. From the equation (4) it can be also said that there will be a linear relation between width of GNR and classical capacitance which will be not sufficient. So we will observe the response of classical capacitance along with quantum capacitance in i)degenerate regime and ii) nondegenerate regime for corresponding changing the width of GNR.

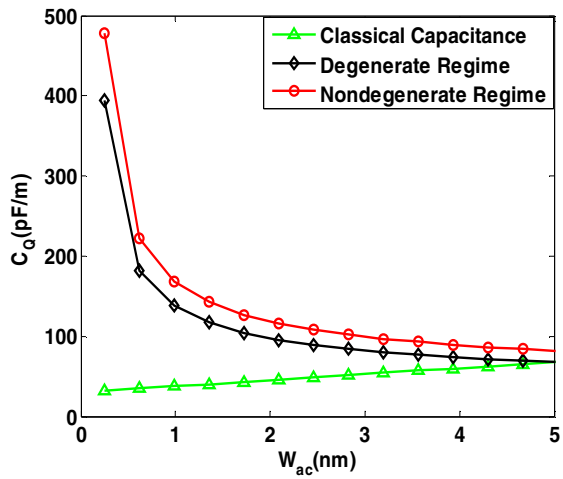


Figure 4.Quantum capacitance in two regimes vs. width (W_{ac}) of GNR.

From the Fig.4 it is seen that classical capacitance gives linear response with width of GNR and the values are varied from 31 pF/m to 73 pF/m. Also it is seen that there is an inverse relation between quantum capacitance in degenerate, nondegenerate regime and width of A-GNR. Here we have analyzed that the quantum capacitance value for a fixed Fermi energy which is close to conduction (for this two regime works simultaneously). As the Fermi level close to conduction band and it is seen from equation (9) that there is a relation of conduction band energy in nondegenerate regime. So quantum capacitance in this regime will be greater value than in degenerate regime. Quantum capacitance in degenerate regime is calculated as 394.09 pF/m to 64.57pF/m. And quantum capacitance in nondegenerate regime is calculated as 478.36 pF/m to 78.38 pF/m.

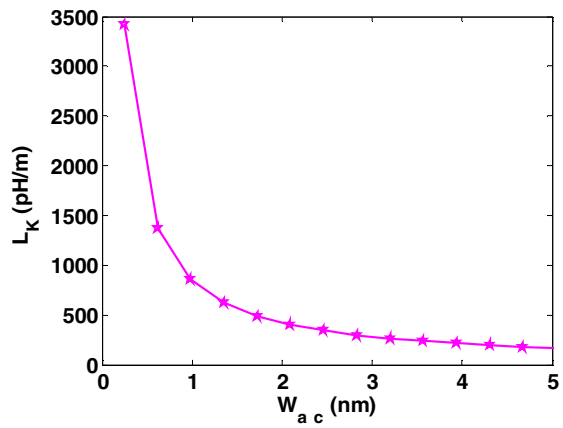


Figure 5. Kinetic inductance (L_K) vs. width (W_{ac}) of GNR

Now we will observe the kinetic inductance for varying width of GNR. From the Fig.5 it is seen the kinetic inductance is decreasing for increasing width of GNR width. It means enlargement of width of GNR reduces kinetic energy of charge carrier. Here it is calculated for 3415 pH/m to 1453 pH/m.

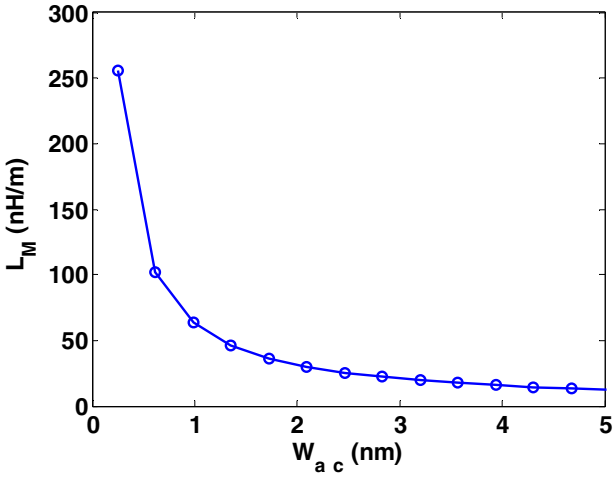


Figure 6.Magnetic inductance (L_M)vs. width (W_{ac}) of GNR

From the Fig.6 it is also observed that increasing the width of GNR reduces the energy stored as magnetic field. It is due to the current through the GNR channel. For this magnetic inductance is greater than kinetic inductance. Because we know that current in a device is formed from integer number of charge carrier (electron) movements. Here this inductance is calculated for 255.46 nH/m to 10.87 nH/m.

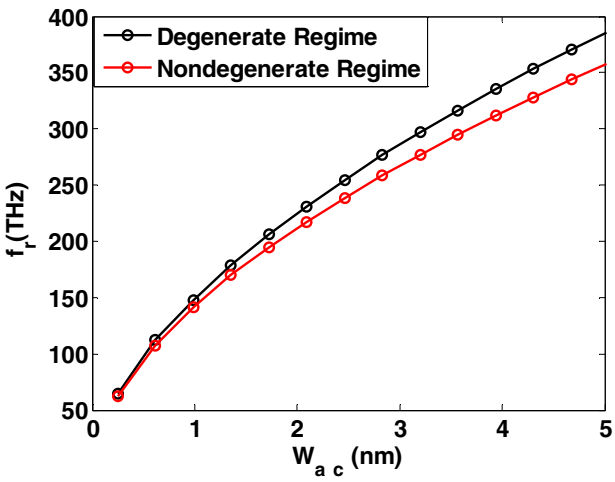


Figure 7. Resonant frequency (f_r)vs. width (W_{ac}) of A-GNR

From Fig.7 we have seen that the resonant frequency in two regime works in two ways. It can be described when the device structure is decreasing, the resonant frequency traces 387.38 THz to 62.60 THz. After that the frequency traces 63.99 THz to 418.81 THz when the device structure is increased.

IV. Conclusion

The RLC interconnect system is one of the important phenomena in microchip, microprocessor, VLSI technology. In this paper we have first observed equivalent circuit of monolayer GNR. Then we have seen that the system can be splitted into some other important parameter like classical capacitance, quantum capacitance, magnetic inductance, kinetic inductance. Then we have analyzed these parameters for varying the structure (here width) of GNR. Quantum capacitance, magnetic inductance, kinetic inductance parameter shows inverse relation with the width of GNR. At last we have observed that the resonant frequency of the system for considering the two regime of quantum capacitance by varying width of GNR. Here we have seen that resonant frequency is larger in degenerate regime than in nondegenerate regime. So in the fabrication it will be fruitful if there is a compromising between width of GNR and resonant frequency of GNR based interconnect system.

References

[1] A. K. Geim and K. S. Novoselov, "The rise of graphene", *Nat. Mat.*, vol. 6,pp. 183–191, 2007.

[2] K. S. Novoselov , "Electric Field Effect in Atomically Thin Carbon Films," *Scie.* ,vol.306, pp.666–669,2004

[3] Q. Shao , G. Liu , D. Teweldebrhan and A. A. Balandin "High-temperature quenching of electrical resistance in graphene interconnects," *Appl. Phys. Lett.*, vol. 92, pp.202108.1-3, 2008

[4] A. Naeemi and J. D. Meindl "Conductance modeling for graphene nanoribbon interconnects", *IEEE Electron Device Lett.*, vol. 28, pp.428 -431 2007

[5] A. Naeemi and J. D. Meindl "Performance benchmarking for graphene nanoribbon, carbon nanotube, and Cu interconnects," *Proc. IEEE Int. Interconnect Technol. Conf.*, 2008, pp.183 -185 .

[6] International Technology Roadmap for Semiconductors (ITRS) reports, 2006. [Online]. Available: http://www.itrs.net/reports.html

[7] S. Haji Nasiri, M. K. Moravvej-Farshi and R. Faez, "Time Domain Analysis of Graphene Nanoribbon Interconnects Based on Transmission Line Model,"*Ira. Jour. of Elect. & Electro.Engi.*, Vol. 8,pp.37-44, 2012.

[8] D. A. Areshkin , D. Gunlycke and C. T. White "Ballistic transport in graphene nanostrips in the presence of disorder: Importance of edge effects," *Nano Lett.,* vol. 7, pp.204 -210, 2007.

[9] J. Guo, Y. Yoon and Y. Ouyang, "Gate Electrostatics and Quantum Capacitance of Graphene Nanoribbons," *Nano Lett.*, vol. 7, no. 7, pp. 1935-1940, 2007.

[10] Z. Johari, N. A. Amin, M. T. Ahmadi, D. C. Y. Chek, S. M. Mousavi and R. Ismail, "Modeling of Quantum Capacitance in Graphene Nanoribbon," *Int'l Conf. Enabling Science And Nanotechnology,* vol. 1341, 2010, pp. 384-387.

[11] D. Gunlycke and C. T. White, "Tight-binding Energy Dispersions of Armchair-edge Graphene Nanostrips," *Phys. Rev. B*, vol. 77, no. 11, p. 115116, 2008.

[12] Burke P., "Luttinger Liquid Theory as a Model of the Gigahertz Electrical Properties of Carbon Nanotubes," *IEEE Trans. on Nano.* , vol. 1, pp.129-144,2002.

[13] J. Chauhan and J.Guo, "Assessment of high-frequency performance limits of graphene field-effect transistors,"*Nano Res.,*Vol. 4, pp. 571-579,2011.

Low Power Dissipation Logic Inverter Design Using Atomic-Level Width Controlled GNR-FETs

Gobinda Saha,[1,*] Atanu Kumar Saha,[1] and A. B. M. Harun-ur Rashid[1]

[1]Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka-1000, Bangladesh
[*]gobinda13@yahoo.com

Abstract— **In this paper we propose a digital logic inverter based on Graphene Nanoribbon Field Effect Transistor (GNR-FET) using ratioed circuit topology. Different types of GNR-FET architectures have been analysed through their current-voltage characteristics to find their probable applications like interconnect, on-chip resistance etc. Here electronic transport calculation is based on NEGF approach which is coupled with Poisson's equation to incorporate the effects of self-consistent field. Our all-graphene inverter shows output voltage swing of ~0.8V for an input voltage swing of 1V having a noise margin of 0.2V with a rise time of ~14.4ps and fall time of ~1.78ps. When it operates in 10GHz and 80% duty cycle the total power consumption is ~100nW which seems promising for low power logic design.**

Index Terms—**Graphene nanoribbon (GNR), Logic inverter, GNR-FET, Non-equilibrium Green's Function (NEGF), Low power, Noise margin.**

I. Introduction

Following Moore's law [1], semiconductor based electronics gone through a very dynamic evolution over decades. Key to this success was the continuous scaling of silicon Metal Oxide Semiconductor Field Effect Transistor (Si MOSFET). Today Si MOSFETs with 20nm gates are in mass production, and for the year 2020, the International Technology Roadmap for Semiconductor (ITRS) requires 10nm MOSFETs [2]. Unfortunately it becomes more and more difficult to further scale the Si MOSFET and simultaneously achieve the needed device performance.

Recently, Graphene, a 2-D carbon-based material, has attracted a lot of attention due to its unique electronic transport properties such as a high Fermi velocity, outstanding carrier mobility, and a high carrier saturation velocity [3]. These characteristics make graphene an excellent candidate for advanced application in future electronics. But due to the zero bandgap nature of unpatterned graphene sheets Graphene-FET (GFETs) suffer from a high off current which makes it unsuitable for digital logic circuit application whereas analog applications of GFETs is potentially emerging [4]-[5]. However lateral confinement of graphene sheet following some particular pattern (chirality) opens up a bandgap (E^g <1.4eV), called graphene nanoribbon (GNR). GNRs have opened a prospect for GNR based FETs as their energy gap is inversely proportional to their width [6] and depending on this width they can be semiconducting or metallic. The ultrathin structure of GNRs leads to perfect electrostatic control of gate over channel which is a key advantage over conventional semiconductors. Extensive research is now going on for fabricating GNRs with narrow width and smooth edges. Patterned GNRs have been exhibited experimentally with ~10nm width [7]-[8]. So, computer simulations can be very useful to provide physical insights and determine performance of GNR-FETs with ultrathin (~1nm) width.

In this paper we have analyzed different GNR-FET architectures where electronic transport is modeled by Non-equilibrium Green's Function (NEGF) formalism which incorporates electron-electron interaction through self-consistent field. Taking accounts the analysis of transmission spectrums and *I-V* characteristics of the devices in a specified voltage range (0V to 1V) suitable applications of different GNR-FETs are proposed. Finally we design a digital logic inverter with necessary optimization and analyzed its different performance parameters such as transfer characteristics, power consumption, noise margin and then operated by a 10GHz, 80% duty cycle input to observe rise time and fall time.

II. Device Configuration and Types

Graphene nanoribbon can be Armchair (A-GNR) or Zig-zag (Z-GNR) depending on the chirality. Depending on the number of carbon atoms along its width Armchair GNR and Zig-zag GNR show semiconducting or metallic behaviour [9]. Taking the variation of bandgap with the width of armchair GNR into consideration we propose four types of GNR-FETs that can be made from only Armchair GNRs.

Fig. 1 shows channel and source-drain regions of four types of GNR-FETs proposed in this paper. For Type-I device in Fig. 1(a), N=7 (E^g = 1.35eV) semiconducting armchair nanoribbon of 3.5nm length is used in the channel region, where N is the number of carbon atoms along the width of armchair GNR. This region is then connected with two N=8 ($E^g \approx$ 0eV) metallic armchair nanoribbons, considered as electrodes. Width of this device is ~0.74nm. This proposed structure provides excellent interface matching between channel and electrode regions thus eliminating the possibility of interface defects, reducing the possibility of impurity scattering. While a contact metal would introduce large contact resistance to the device [10], contact with the metallic armchair GNR makes smoother metal-semiconductor interface with minimum contact resistance. Fig. 1(b), 1(c), 1(d) shows other combinations of GNRs considered in this paper. Type-II and Type-III devices are solely made of metallic GNR (N=11) and semiconducting GNR (N=7) respectively. In Type-IV device top-gated channel is a cascade of metallic (N=8)-semiconducting (N=7)-metallic (N=8) GNR. The detailed behaviour of all types of devices under potential will be discussed in the subsequent sections.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

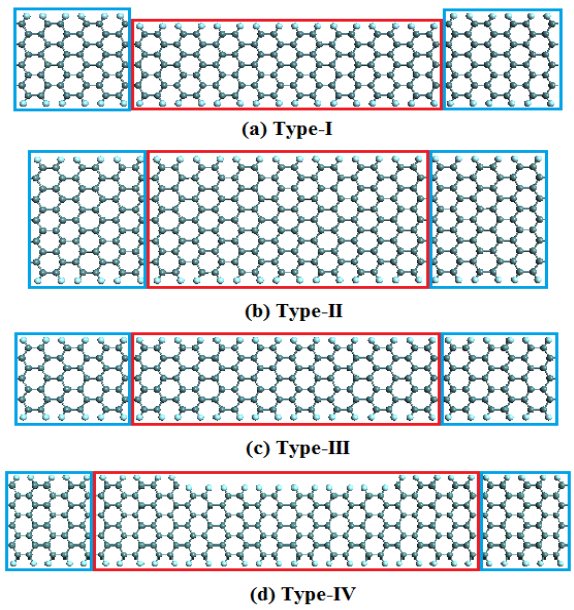


Fig. 1 Different types of device configuration (Red boxes are top gated channel region and blue boxes are electrodes).

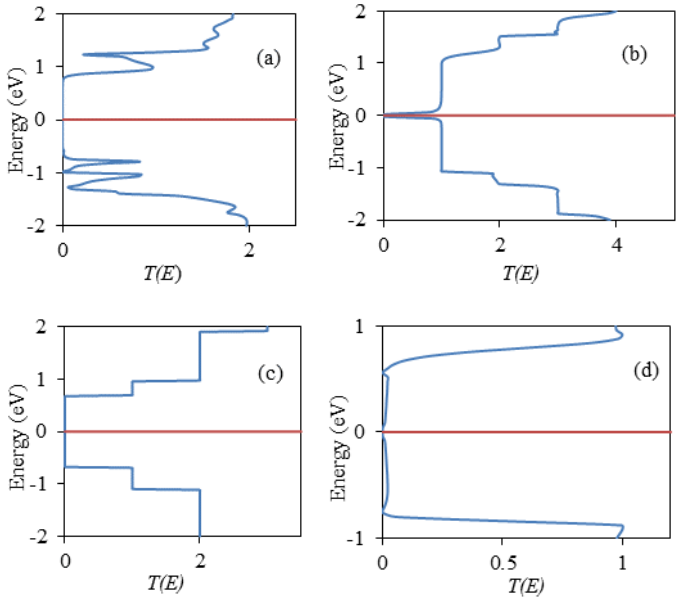


Fig. 2 Transmission Spectrum, *T(E)* of (a) Type-I device (b) Type-II device (c) Type-III device (d) Type-IV device at zero gate bias.

III. Simulation and Results

In this paper we have used non-equilibrium Green's functions (NEGF) approach coupled with Poisson's equation to calculate electronic transport in GNR-FET. Here the NEGF formalism [11] is based on tight binding P_Z orbital which incorporates 3rd nearest neighborhood calibrated through a first-principle calculated *ab-initio* model [12]. The charge distribution described by NEGF is self-consistently coupled with electrostatics to incorporate the effect of self-consistent field and solved by multi-grid finite difference method. The edge atoms are considered as hydrogen passivated (σ bond) and HfO_2 ($\kappa \approx 25$) is used as gate oxide with oxide thickness of ~0.4nm for minimized gate leakage.

A. Transmission Spectrum

Transmission spectrum for Type-I device is shown in Fig. 2(a). In the bandgap (1.35eV) energy range it has zero transmission coefficient, *T(E)*.

On the other hand Type-II device in Fig. 2(b) shows zero bandgap ensuring metallic behaviour where a small bias voltage can produce very high current but having less impact of top-gate voltage upon this current. Whereas in Type-III device transmission spectrum may seem as 2D semi-conductor like as shown in Fig. 2(c), but current may degrade a lot due to very high contact resistance and schottkey barrier due its relatively high bandgap.

Type -IV device is in general sense a cascade of Type-III and Type-II device. Its channel is formed by both semiconducting (N=7) and metallic (N=8) GNR, where for metallic region transmission coefficient is non-zero in the bandgap region of semiconducting GNR as shown in Fig. 2(d), which may give rise to piecewise linear resistance behaviour and this resistance is controllable by top-gate voltage. By varying the relative length of semiconducting and metallic GNR it is possible to construct different resistance (in mega ohm range) at fixed gate potential.

So we can come in a decision that Type-I device may be chosen for better performance as FET, Type-II for interconnect, Type-IV for on-chip resistance.

B. I-V Characteristics

From the transmission spectrums it is evident that gate voltage and drain-source bias voltage will have larger impact on the current in the channel region for Type-I device than any other three types of devices. When a positive gate potential is applied, minimum energy for non-zero electron transmission coefficient comes closer to Fermi energy level which in turns gives rise to current, I_{ds}. The current-voltage characteristics, I_{ds}-V_{ds} curve for Type-I device for different gate potential is shown in Fig. 3. The current-voltage characteristics of Type-II device is shown in Fig. 4. This device shows resistor like behaviour (almost linear *I-V* characteristics) with resistance in kilo-ohm (KΩ) range.

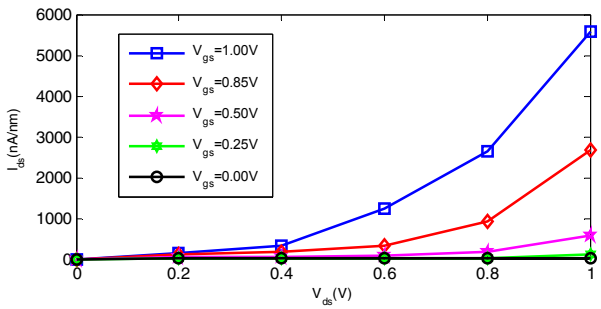


Fig. 3 I-V_{ds} curve at different gate voltage for Type-I device.

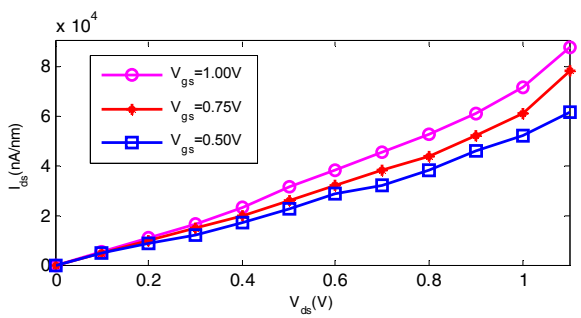


Fig. 4 I-V_{ds} curve at different gate voltage for Type-II device.

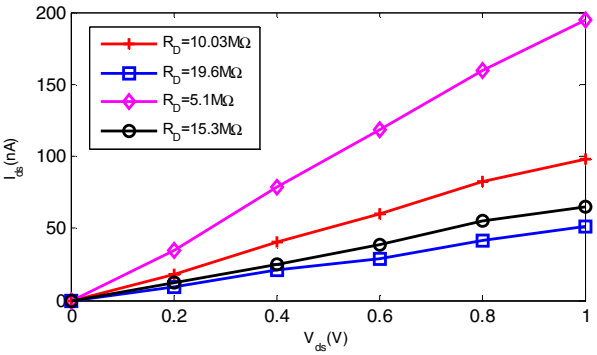
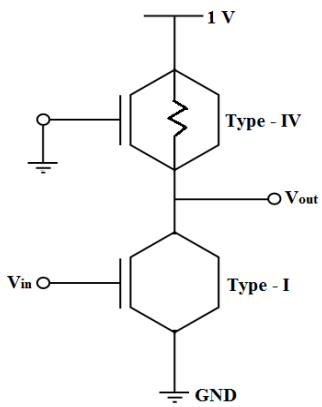


Fig. 5 *I-V*$_{ds}$ curve of Type-IV device at V_{gs}=0V.

Fig. 5 shows I_{ds}-V_{ds} curves of Type-IV device. This device also shows linear resistor like *I-V* characteristics in a specified voltage range (0V to 1V) with resistance in the mega-ohm (MΩ) range. Using this property on-chip resistance can be modelled. In this paper, relative length of semiconducting and metallic GNR in the channel region was varied to construct different values of resistance.

IV. Digital Logic Inverter Design

Implementation of digital logic inverter or switching circuit using GNR-FET is somewhat challenging due to its unconventional behaviour (absence of current saturation) with respect to MOSFET. Due to absence of complementary GNR-FET pairs, the logic architecture proposed here is based on ratioed circuit topology. From the previous discussions it is evident that Type-I GNR-FET is suitable as pull-down network. Whereas pull up network is made by Type-IV GNR-FET because of its high resistance, R_D and Type-III GNR-FET is used as interconnect between pull-up and pull-down network. To obtain desired inverter transfer characteristics an optimum value of R_D is required. A schematic diagram and its layout diagram of our designed logic inverter are shown in Fig. 6 and Fig. 7 respectively.



V_{in}	0V	1V
V_{out}	0.90V	0.11V
I_{static}	8.63nA	88.6nA
P_{static}	8.63nW	88.6nW

Fig. 6 Inverter using Type-I and Type-IV GNR-FETs.

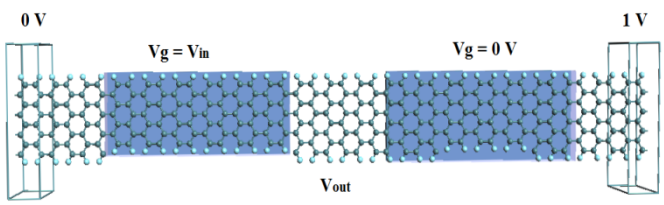


Fig. 7 Layout diagram of a GNR inverter.

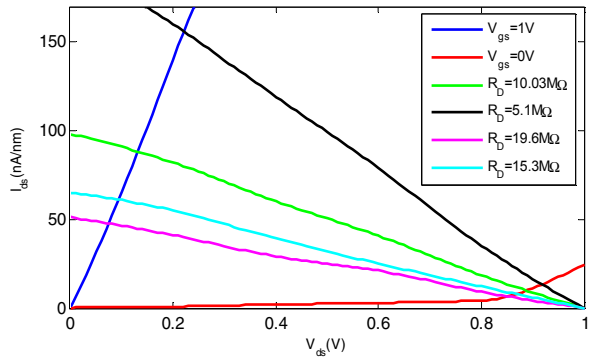


Fig. 8 I_{ds}-V_{ds} curve of Type-IV device with different load, R_D lines.

A. Optimization of R_D

The model of logic inverter proposed in this paper is based on the I_{ds}-V_{ds} curves of Type-I device and Type-IV device. In Fig. 8, *I-V* characteristics for both devices are shown in the same plot. From the figure we can see that variation of resistance, R_D of the pull up network would cause significant alteration in the output logic level of the inverter. When R_D is ~10MΩ, output voltage, V_{out} =0.9V and V_{out}=0.11V for input voltage, V_{in} =0V (Logic 0) and V_{in}=1V (Logic 1) respectively. Decreasing R_D (i.e. R_D~5.1MΩ) would improve inverter performance in Logic 0 input but degrade the performance in case of Logic 1 input. On the other hand if R_D is increased (i.e. R_D~15.3MΩ or R_D~19.6MΩ), performance in Logic 1 input would improve but performance in Logic 0 input would fall. For this reason, we selected the value of R_D at which the difference of output voltage logic levels is maximum. In our simulation optimum value of R_D was found to be ~10MΩ.

B. Power Dissipation and Delay

Static power dissipation in the designed inverter can be found by first calculating the static current, I_{static} of the device during different input logic level. As listed in Fig. 6 I_{static}=88.61nA for input Logic 1 and I_{static}=8.63nA for input Logic 0. So, static power dissipation is ~88.61nW for input Logic 1 and ~8.63nW for input Logic 0. Compared to other available technology these values are very low. Therefore, this device may become a viable option to the circuit design where power dissipation poses major headache.

If V_{out} is fed to operate a second inverter stage then the gate capacitance of the second stage will imposes a rise/fall time delay upon V_{out}. From the device configuration, gate capacitance, C_g of Type-I device is found to be 4.14×10^{-10} F/m. Using RC delay model rise time and fall time of the inverter is found to be ~14.4ps and ~1.78ps. The big difference between the rise and the fall time is due to the ratioed circuit topology where a large current flows during fall time enabling faster discharge of capacitor whereas very little current flows during rise time resulting in slower capacitor charging. Equal rise time and fall time for the device can be achieved by sacrificing the optimized value of R_D causing output logic voltage degradation.

C. Transfer Characteristics

Output input voltage relationship is summarized by voltage transfer characteristics shown in Fig. 9. This curve gives steady-state value of V_{out} for each value of V_{in}. When V_{in}=0V very small current (~8.63nA) flows through pull down transistor. As a result V_{out} is slightly less than to V_{DD} (1V). This voltage is referred as V_{OH} and its value is 0.9V.When V_{in}

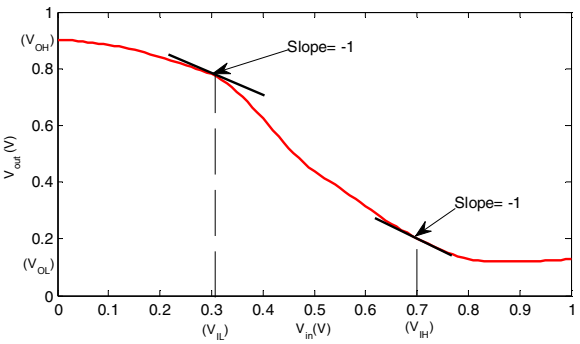


Fig. 9 Voltage transfer characteristics of the inverter.

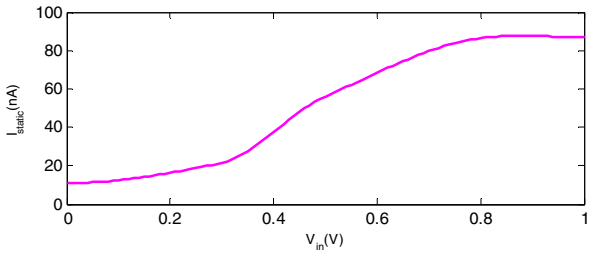


Fig. 10 I_{static}-V_{in} curve of the inverter.

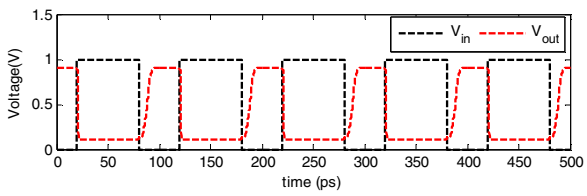


Fig. 11 Output voltage of the inverter when simulated with a 10GHz, 80% duty cycle input voltage signal.

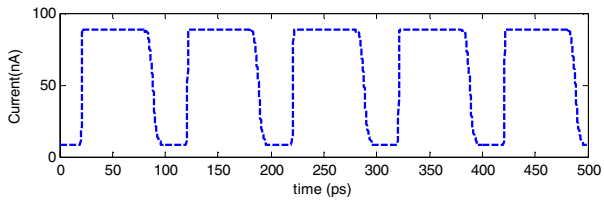


Fig. 12 Current through the inverter.

is 1V, current through pull down transistor increases significantly but still there is some voltage drop across the device. As a result V_{out} is slightly greater than to *GND* (0V). This voltage is referred as V_{OL} and its value is 0.11V.

Fig. 9 also shows the points where V_{out} begins to change from high to low and vice versa. The voltage V_{IL} indicates the points where V_{out} is high and slope of the curve is -1. This is the maximum input voltage level that inverter will interpret as low. Similarly V_{IH} is the minimum input voltage level that inverter will interpret as high. In the curve it indicates the points where V_{out} is low and slope of the curve is also -1.In our simulation we found V_{IL}=0.31V and V_{IH}=0.7V.

Variation of static current due to V_{in} is plotted in Fig. 10. Fig.11 shows the simulated output voltage of the inverter circuit when a sample input signal of 10GHz with 80% duty cycle is given at V_{in}. Current through this device is plotted in Fig. 12. From this figure we found that total power dissipation in the device for the above input signal is ~100nW. This result is very promising compared to complementary graphene inverter [13], where a static power dissipation is found to be ~0.77mW with maximum allowable operating frequency ~4.5GHz.

D. Noise Margin

Noise margin is defined as the ability to tolerate noise without altering the correct operation of the circuit. It is an important design parameter for logic circuits. Low noise margin, NM_L and high noise margin, NM_H are defined by the following equations

$$NM_L = V_{IL} - V_{OL} \quad (1)$$

$$NM_H = V_{OH} - V_{IH} \quad (2)$$

Using equations (1) and (2) values of noise margin were found to be NM_L=0.2V and NM_H=0.2V which is low compared to existing CMOS technology.

V. Conclusions

In this paper, we have performed NEGF simulation to demonstrate the intrinsic current-voltage characteristics of ultra-thin GNR-FETs. Different scattering phenomenon (phonon scattering, lattice mismatch with dielectric) are treated ideally because of its low voltage application. These results are used to design an all graphene digital logic inverter. Then we derived different performance parameters for ultra-fast logical operation which shows sufficient low power consumption (~100nW) even at high operating frequency (10GHz). In conclusion, our work may lay the basis for establishing the realistic operating range of GNR-FET based nanoelectronic logic circuit design, providing useful guidelines for future experimental exploration.

References

[1] G. E. Moore, "No exponential is forever: But 'Forever' can be delayed," in *Dig. Tech Papers Int. Solid-State Circuits Conf.*, 2003, pp. 20-23.

[2] ITRS – The International Technology Roadmap for Semiconductors. [Online]. Available: http://www.itrs.net.

[3] V. E. Dorgan, M. Bae, and E. Pop, "Mobility and saturation velocity in graphene on SiO_2," *Appl. Phys. Lett.*, vol. 97, no. 8, p. 082112, Aug. 2010.

[4] Y.-M. Lin, K. A. Jenkins, A. Valdes-Garcia, J. P. Small, D. B. Farmer, and P. Avouris, "Operation of graphene transistors at gigahertz frequencies," *Nano Lett.*, vol. 9, pp. 422–426, Jan. 2009.

[5] Y. Q. Wu, Y.-M. Lin, K. A. Jenkins, J. A. Ott, C. Dimitrakopoulos, D. B. Farmer, F. Xia, A. Grill, D. A. Antoniadis, and Ph. Avouris, ''RF performance of short channel graphene field-effect transistor,'' in *Int. Electron Devices Meeting Tech. Dig.*, 2010.

[6] Y. W. Son, M. L. Cohen, and S. G. Louie, "Energy gaps in graphene nanoribbons," *Phys. Rev. Lett.*, vol. 97, no. 21, p. 216803, Nov. 2006. .

[7] X. Li, X. Wang, L. Zhang, S. Lee, and H. Dai, "Chemically Derived, Ultrasmooth Graphene Nanoribbon Semiconductors," *Science*, vol. 319, no. 5867, pp. 1229–1232, Feb. 2008.

[8] L. Jiao, L. Zhang, X. Wang, G. Diankov, and H. Dai, "Narrow graphene nanoribbons from carbon nanotubes," *Nature*, vol. 458, no. 7240, pp. 877–880, Feb. 2009.

[9] L. Yang, C. Park, Y. Son, M. L. Cohen, and S. G. Louie, "Quasiparticle Energies and Band Gaps in Graphene Nanoribbons," *Phys. Rev. Lett.*, vol. 99, no. 18, p. 186801, Nov 2007.

[10] F. Xia, V. Perebeinos, Y. Lin, Y. Wu, and P. Avouris, "The origins and limits of metal–graphene junction resistance," *Nature Nanotech.*, vol. 6, no. 3, pp. 179–184, Feb. 2011.

[11] S. Datta, "Nanoscale device modeling: The Green's function method," *Superlattices Microstruct.*, vol. 28, no. 4, pp. 253–278, Oct. 2000.

[12] S. Reich, J. Maultzsch, C. Thomsen, and P. Ordejón, "Tight-binding description of graphene," *Phys. Rev. B*, vol. 66, p. 035412, Jul. 2002.

[13] F. Traversi, V. Russo, and R. Sordan, "Integrated complementary graphene inverter," *Appl. Phys. Lett.*, vol. 94, no. 22, p. 223312, Jun. 2009.

An efficient design of FSM based 32-bit unsigned high-speed pipelined multiplier using Verilog HDL

Abdullah-Al-Kafi, [1,*] Atul Rahman [2], Bushra Mahjabeen [3], Mahmudur Rahman [4]

[1]ASIC Design Engineer, Fastrack Anontex Limited
Dhaka, Bangladesh

[2]ASIC Design Engineer, Fastrack Anontex Limited
Dhaka, Bangladesh

[3]ASIC Design Engineer, Fastrack Anontex Limited
Dhaka, Bangladesh

[4]ASIC Design Engineer, Fastrack Anontex Limited
Dhaka, Bangladesh

[*]kafi.eee07@gmail.com

Abstract— **This paper shows a new methodology to design hardware of 32-bit unsigned pipelined multiplier. The proposed hardware design is based on Finite State Machine (FSM) for reducing hardware resources and proliferating maximum frequency. Our suggested pipelined multiplier design contains only four 40-bit full adders to complete 64-bit 32 partial products addition. The synthesis report of the 32-bit pipelined multiplier shows that the usage of the logical resources of FPGA is significantly less than the earlier 32-bit multiplier design. Moreover, the proposed pipelined multiplier hardware doesn't use any DSP or dedicated multiplier block of FPGA for the multiplication process. So, our proposed design is technology independent and it gives uttermost performance both in FPGA and ASIC. The maximum frequency achieved for operation latency of 9-clock cycle is 326.243 MHz for computing 32-bit×32-bit unsigned multiplication and the pipelined multiplier hardware design is tested on Xilinx Virtex-6 XC6VLX75T-3-FF484 FPGA.**

Index Terms— **FSM, Maximum frequency, Pipeline, FPGA and ASIC.**

I. INTRODUCTION

Multiplier circuit is an important block in digital system design such as Digital Signal Processing (DSP) block, Microprocessor, Microcontroller, floating point multiplier, sign multiplier, Multiply-Accumulator (MAC) block etc. The main challenge to design 32-bit multiplier is the addition of large number of partial product. So, efficient hardware design of multiplier has been developed frequently in order to decrease hardware resources and increase maximum frequency. [1-2, 5]

There are mainly two types of hardware implementation method of multiplier, combinational and pipelined implementation method. In combinational implementation, multiplier uses adder tree to add the partial products. As a result combinational multiplier increases both hardware resources and critical path delay. So, in large multiplier implementation combinational method is almost obsolete in digital circuit design. On the other hand, pipelined multiplier is reduced hardware resources and increase maximum frequency by using different pipelined stage. Although, pipelined multiplier reduces hardware resources, but it doesn't use the hardware evenly to complete the multiplication process due to lack of pipelined controller. As a result, the hardware implementations of this multiplier have less maximum frequency, higher clock cycle and higher usage of hardware resources.

Various techniques on implementing 32-bit unsigned multiplier based on pipelined architecture and combinational architecture can be found in the literature. A 32-bit unsigned multiplier is implemented in FPGA using Vedic algorithm and maximum combinational path delay achieved was 22.829 ns. [6] Moreover, the synthesis report shows that the design uses 1188 slices, 2054 4-input LUT and 128 IO buffer of the FPGA. Another implementation of unsigned multiplier uses 659.292 ns to complete the multiplication process. A pipelined multiplier implementation was proposed to reduce the usage of hardware resources and increase maximum frequency for implementing in FPGA and the designed FPGA based 32-bit multiplier's maximum frequency was 319.488 MHz (critical path delay 3.13 ns) [5] with operation latency of 16 clock cycle. The usage of hardware resources ware less compared to that of the counterpart at that time. At the same time, many designer uses DSP block or dedicated multiplier blocks to implement multiplier in FPGA to increase maximum frequency. Altera have Library Parameterized Module (LPM) multiplier where a 32-bit multiplier uses only four 18-bit DSP block for implementing the multiplier and maximum frequency achieved was 454 MHz at 3 output latency [7]. But this multiplier design is useless in implementing in ASIC because ASIC synthesis library does not have such DSP or dedicated multiplier block. So we need an efficient design to implement multiplier in ASIC. [4-5]

In this paper, we have proposed a new novel approach to efficiently design hardware of 32-bit unsigned pipelined multiplier. For the testing purpose we have tested our design in FPGA. The result of the synthesizing report supersedes previously proposed hardware design's synthesize report in terms of less clock cycle, maximum operating frequency and less usage of hardware resources. We have used shift and add algorithm for multiplication process. Our proposed multiplier design uses only four 40-bit adders (four 36-bit adder and four 4-bit adder), Mux (four 9-to-1 Mux and four 8-to-1 Mux) and Demux (two 1-to-9 Demux and two 1-to-8 Demux) for

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

partial product addition. For controlling the Mux and Demux we have used simple Finite State Machine (FSM) which has nine stages to complete the multiplication. Our design ensures maximum usage of the hardware resources in every clock cycle. Pipelining hazard, formation of combinational loops and unwanted latches are completely removed during the hardware design process.

II. Proposed Hardware Design Methodology and Architecture for 32-bit Pipelined Multiplier

A. Design Method

In this paper, we have applied simple shift and add algorithm for the multiplication process. In this process n-bit multiplicand and n-bit multiplier will generate 2n-bit product. To complete this process, it will generate n number of partial product and each partial product will shift left to one bit. Now, 2n-bit product will be generated by adding n number of partial product. Let's illustrate this method using a 4-bit multiplication.

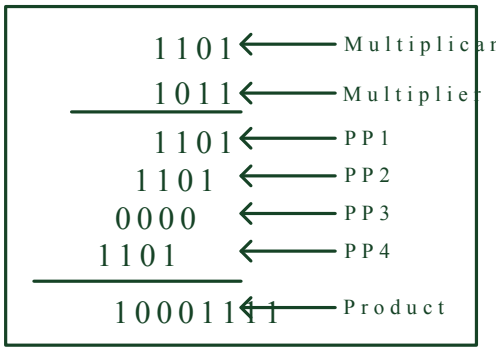


Fig. 1 Multiplication example using shifts and add algorithm.

This direct partial product addition process is useful for small multiplier but for larger multiplier this asynchronous addition process is futile in digital circuit. For example, a 32-bit multiplier will generate 32 partial products and each partial product will be 64-bit. So, 32 64-bit asynchronous additions will raise the critical path delay as well as hardware resources. To solve this problem we have designed FSM based pipelined multiplier.

B. Architecture Operation

In our proposed multiplier circuit, we have reduced adder number and added the partial product sequentially to increase maximum operating frequency and reduce hardware resources. To begin with, we have 32 32-bit partial products. So, we can add the adjacent partial products to reduce to adder width. At first stage, we have to change the 32-bit partial product to 34-bit partial product. So there are 16 34-bit additions and that will generate 16 34-bit partial products for second stage. In second stage, we have to change the 34-bit partial product to 36-bit partial product. So there are 8 36-bit additions and that will generate 8 36-bit partial products for third stage. In third stage, we have to again change the 36-bit partial product to 40-bit partial product. So there are 4 40-bit additions and that will generate 4 40-bit partial products for forth stage. In forth stage, we have to modify 40-bit partial product to 48-bit partial product. So there are 2 48-bit additions and that will generate 2 48-bit partial products for fifth stage. In final stage, we have to modify 48-bit partial product to 64-bit partial product. This two 64-bit partial product addition will give 64-bit product. Our proposed multiplier design can be divided into three major parts and these are partial product generator block, Datapath block and FSM block.

1) *Partial Product generator block*: This block generate partial product from multiplicand and multiplier. As we have seen that there are 31 different additions process among them 24 additions (16 34-bit additions and 8 36 bit additions) can be done by 36 bit adder. So we have generate the partial product in 36-bit to reduce the partial product modification process in verilog coding.

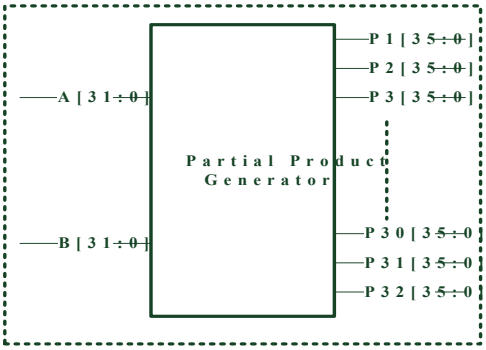


Fig. 2 Partial Product Generator block diagram

2) *Datapath block:* This block handles all the addition of the partial products. This block contains 4 40-bit adders (4 36-bit adder and 4 4-bit adder), 4 9-to-1 Mux and 4 8-to-1 Mux, 2 1-to-9 Demux and 2 1-to-8 Demux.

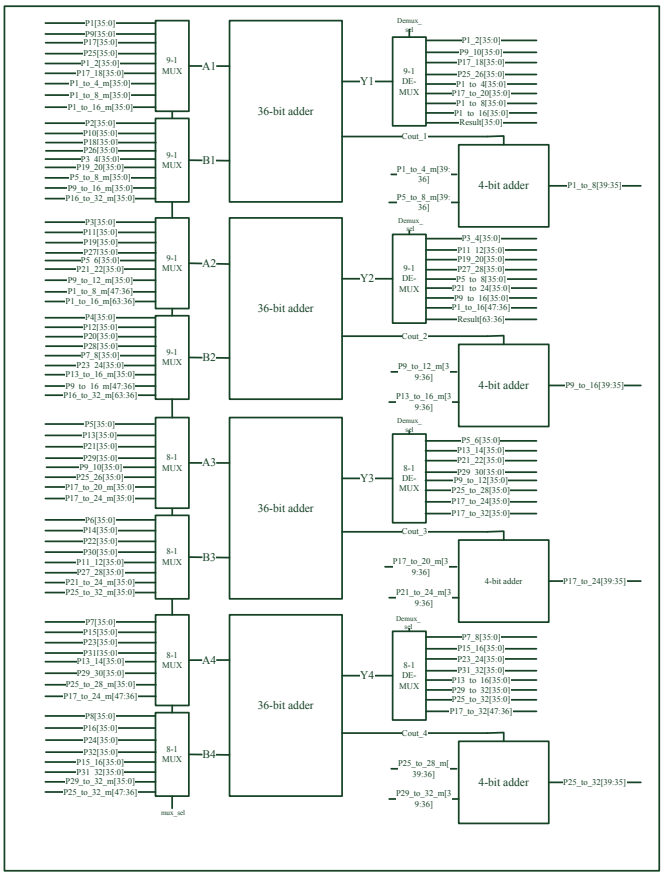


Fig. 3 Datapath block diagram

All the outputs of partial product generator block are assigned as inputs in the Datapath block. There are 32 36-bit partial products and 16 36-bit additions. So it needs 4 clock cycles to complete these 16 additions. And, these 16 36-bit partial products are stored in a register. Then, there are 16 36-bit partial products need 8 36-bit additions. So it needs 2 clock cycles to complete these 8 additions. This 8 addition will generate 8 36-bit partial products. After that, these 8 36-bit partial products (P1_to_4, P5_to_8, P9_to_12, P13_to_16, P17_to_20, P21_to_24, P25_to_28, and P29_to_32) are modified into 40-bit partial product (P1_to_4m, P5_to_8m, P9_to_12m, P13_to_16m, P17_to_20m, P21_to_24m, P25_to_28m, and P29_to_32m). Now, there are 8 40-bit partial products need 4 40-bit additions. And it can be done in one clock cycle where LSB 36 bits are assigned in 36-bit adder and MSB 4 bits are assigned in the 4-bit adder. If any carry generated in the 36-bit adder that will propagate to the 4-bit adder and 40-bit result will be concatenated from the two adder output. Then, there are 4 40-bit partial products (P1_to_8, P9_to_16, P17_to_24, and P25_to_32) and they are modified into 48-bit partial product (P1_to_8m, P9_to_16m, P17_to_24m, and P25_to_32m). Then first 2 48-bit partial products are assigned in the first 2 adder. In first adder LSB 36-bits are assigned and in second adder MSB 12 bits are assigned. If any carry generated in the first adder that will propagate to the second adder and 48-bit result will be concatenated from the two adder output. Same process will be followed for the last two partial products. Then, there are two 48-bit partial products (P1_to_16, P17_to_32) and they are modified into 64-bit partial product (P1_to_16m, P17_to_32m). First 36 bits are assigned in the first adder and remained 28 bits are assigned in the second adder. The result will be concatenated output of the two adders.

```
assign P1_to_4m   = {4'h0, P1_to_
assign P5_to_8m   = { P5_to_8, 4'h
assign P9_to_12m  = {4'h0, P9_to_
assign P13_to_16m = { P13_to_16
assign P17_to_20m = {4'h0, P17_to
assign P21_to_24m = { P21_to_24
assign P25_to_28m = {4'h0, P25_to
assign P29_to_32m = { P29_to_32

assign P1_to_8m  = {8'h00, P1_to_
assign P9_to_16m  = {P9_to_16, 8'
assign P17_to_24m  = {8'h00, P17_
assign P25_to_32m  = {P25_to_32

assign P1_to_16m = {16'h0000, P1_
assign P17_to_32m = {P17_to_32,
```

Fig. 4 Verilog code for partial product modification

3) *Finite State Machine (FSM) block:* This block generates Mux_sel and Demux_sel input for the Datapath block. This block can be divided into 2 parts Present State Register (PSR) and Next State Output Logic (NOSL). PSR is a sequential logic and it gives output in the positive edge clock. NOSL is a combinational circuit it gives the output of the FSM. This FSM contains 9 state and gives Mux_sel [3:0] and Demux_sel [3:0] outputs.

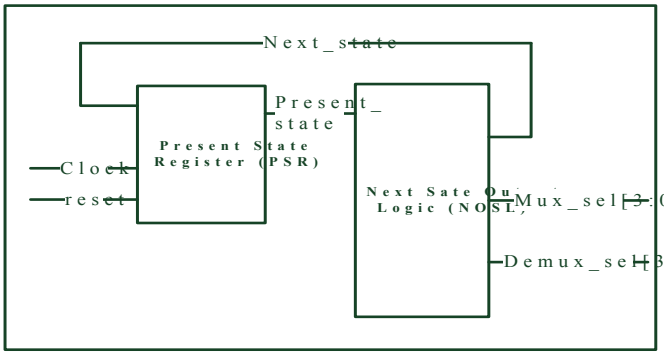


Fig. 5 Finite State Machine (FSM) block diagram

III. Result and Discussion

A. *Synthesis report of pipelined multiplier*

The proposed 32-bit unsigned pipelined multiplier is implemented on Xilinx Virtex-6 XC6VLX75T-3-FF484 FPGA. The results of the synthesizing reports of pipelined multiplier are shown in the Table I and Table I. Table II shows the advance level HDL synthesis report and Table II shows Low level HDL synthesis report.

TABLE I
Advance HDL Level synthesis report for 32-bit pipelined multiplier in Xilinx Virtex-6 XC6VLX75T-3-FF484 FPGA

37-bit adder	4-bit adder	Number of Flip-Flop	Number of Multiplexer	Number of FSM
4	4	1132	1005	1

There are 1005 different types of multiplexers among them 72 are 1-bit 12-to-1 multiplexer, 781 are 1-bit 2-to-1 multiplexer, 72 are 1-bit 8-to-1 multiplexer, 2 are 36-bit 12-to-1 multiplexer, 71 are 36-bit 2-to-1 multiplexer, 2 are 36-bit 8-to-1 multiplexer, 4 are 48-bit 2-to-1 multiplexer, and one 64-bit 2-to-1 multiplexer.

TABLE II
Low Level synthesis report for 32-bit pipelined multiplier in Xilinx Virtex-6 XC6VLX75T-3-FF484 FPGA

Clock Cycle	Maximum Frequency (MHz)	No. of Slice Registers	No. of LUT FF Pairs	No. of occupied slices	No. of Slice LUTs
9	326.243	1061	1790	1059	1788

Synthesis report shows that multiplier can support maximum 326.243 MHz and the multiplier circuit used only 1059 slices among 1790 slices.

B. *Simulation result:*

Simulation is done in Modelsim Simulator 10.0c. The following figure 6 shows the functional simulation of pipelined multiplier. It shows that after assigned inputs, it takes nine clock cycles for the result. At first reset input is high for one clock cycle and after that it takes nine clock cycles for getting the result.

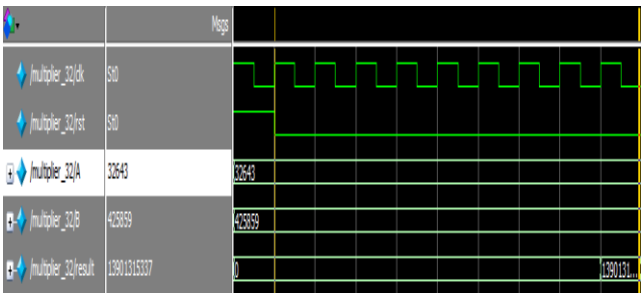


Fig. 5 Timing Diagram of Functional Simulation of 32-bit pipelined multiplier

C. *Compare the result with other design:*

Our proposed FSM based Pipelined design gives better output latency and lower critical path delay or lower minimum period than other previous design. Table III shows the comparison of critical path delay and output latency with other design. It clearly shows that our proposed design gives improved result than other design.

TABLE III
COMPARISON OF CRITICAL PATH DELAY AND OUTPUT LATENCY WITH PROPOSED PIPELINED DESIGN AND OTHER DESIGN

32-bit Multiplier Design	Output latency (Clock Cycle)	Critical Path Delay (ns)
Paper-5	32	3.13
Paper-6	16	22.829
Proposed Multiplier design	**9**	**3.065**

IV. CONCLUSIONS

In the presented paper, we have designed a pipelined multiplier which gives better performance than other pipelined design. The main advantage of this design is it used less output latency to calculate the result than other design. We can increase the output latency by adding more adders. Similarly, we could increase maximum frequency of the multiplier by designing the adder in gate level. Another benefit of this design is that it technology independent. This design gives maximum performance both in FPGA and ASIC.

REFERENCES

[1] V.Vijayalakshmil, R.Seshadd, Dr.S.Ramakrishnan, *Design and Implementation of 32-bit unsigned multiplierusing CLAA and CSLA*, IEEE, 2013.

[2] Raahemifar K. ,Ahmadi M. , *Fast 32-bit digital multiplier Circuits and Systems*, 2000. Proceedings. ISCAS 2000 geneva. The 2000 IEEE International Symposium

[3] Kiwon Choi, Minkyu Song, *Design of a high performance 32x32-bit mulltiplier with a novel sign select Booth encoder Circuits and Systems*, 2001. ISCAS 2001. The 2001 IEEE International Symposium

[4] Ma Yan , Wang De-li , *The design of an architecture-aware 32- bit signed/unsigned multiplier*, Computer Engineering and Technology (IC- CET), 2010 2nd International Conference

[5] Qingzheng Li Guixuan Liang Bermak, A. *A High-speed 32-bit Signed/Unsigned Pipelined Multiplier* Electronic Design, Test and Ap- plication 2010. DELTA 10. Fifth IEEE International Symposium

[6] Syed Shahzad Hussain Shah, Muhammad Naseem Majoka, and Gulistan Raja: *Design And Implementation of 32-bit VEDIC multiplier on FPGA.* First International Conference on Modern Communication & Computing Technologies (MCCT'14)

[7] http://www.altera.com/literature/ug/ug_lpm_alt_mfug.pdf

Low Power High Speed Ternary Content Addressable Memory Design using 8 MOSFETs and 4 Memristors – Hybrid Structure.

Shawana Tabassum,[1]* Farhana Parveen,[1] and A.B.M. Harun-ur Rashid[1]

[1] Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology
Dhaka-1000, Bangladesh
*shawana.tabassum@gmail.com

Abstract—**Ternary Content Addressable Memory (TCAM) compares a stored word in its memory with a search word and returns the address of the stored word. In this paper we have proposed a memristor based TCAM cell which has only 8 transistors and 4 memristors whereas a conventional transistor based TCAM cell consists of 16 transistors. The average search energy per bit of the proposed Memristor based TCAM cell is almost 70.9% smaller and the search time in the proposed cell is almost 71% smaller than the corresponding quantities in a 16T TCAM cell. The voltage margin of our proposed cell is 138.9% higher than that in a 16T cell. Moreover, the required area for layout design of the proposed cell is almost 64% smaller than that of the 16T cell. Hence, our proposed Transistor-memristor hybrid structure shows superior performance in terms of speed, energy consumption, voltage margin and area occupancy.**

Index Terms—**TCAM, Memristor, Hysteresis.**

I. Introduction

There are three fundamental passive circuit elements: the resistor, the capacitor and the inductor. But Leon Chua predicted the existence of a fourth fundamental device, which he called a memristor[1]. Memristor has a non-volatile nature which means it can store data even when no power supply is applied to the circuit. Because of its long-term memory and suitability for nanoscale design, memristor can be used as a non-volatile memory device in future semiconductor memory chips. Memristor is basically a two terminal device. It can be formed from a thin semiconductor film [2], [3] with one layer of insulating TiO_2 and one layer of oxygen deficient TiO_{2-x}. The film resides between two metal contacts. The resistance of the doped layer of TiO_{2-x} is lower compared to the resistance of undoped layer of TiO_2.The ionic motion of oxygen vacancies (+2) [4] results in this resistance change. Fig. 1 shows the doped and undoped layers of a memristor. The memristor has a size D. The lengths of doped and undoped regions are *w* and (D-w) respectively.

As a linear circuit element, Memristance M is constant and memristor behaves like a resistor. The most basic mathematical definition of a current-controlled memristor for circuit analysis is the differential form

$$v = R(w)i \tag{1}$$

$$\frac{dw}{dt} = i \tag{2}$$

Here w is the length of doped region and R(w) is memristance as a function of w.

The application of an external bias v(t) across the device moves the boundary between the two regions by causing the

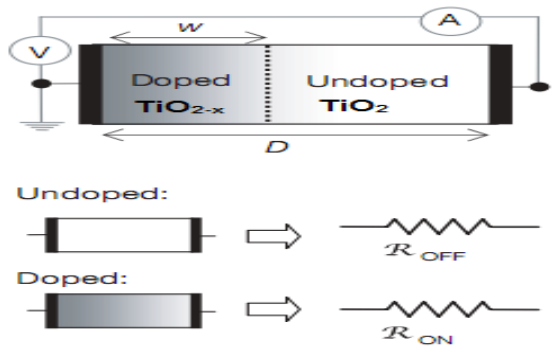


Fig. 1 A single memristor

charged dopants to drift. The total resistance of the memristor is a sum of the resistances of the doped and undoped regions,

$$R_{MEM}\left(\frac{w(t)}{D}\right) = R_{ON}\left(\frac{w(t)}{D}\right) + R_{OFF}\left(1-\frac{w(t)}{D}\right) \tag{3}$$

Here R_{ON} is the resistance of the memristor when it is completely doped and R_{OFF} is the resistance of the memristor when it is completely undoped.

When a positive voltage is applied to the left platinum wire, oxygen vacancies move from the TiO_{2-x} layer to the TiO_2 layer. As a result the length of doped region increases. This situation is termed as "ON" state which is characterized by a decrease in the memristance. When the positive voltage is applied for a long time, at a certain point in time no change in memristance takes place. The device then enters a hysteresis region. If a voltage with opposite polarity is now applied across the memristor, the boundary separating the doped and undoped region moves in the opposite direction, that is towards TiO_{2-x} layer. Therefore, the length of doped layer now decreases and memristance increases. This situation corresponds to "OFF" state. The hysteresis curve of a memristor is shown in Fig. 2 below. The simulation has been performed using the parameters: R_{ON}=100 R_{OFF}=10K R_{INIT}=5K D=8N μ_v=1f p=10

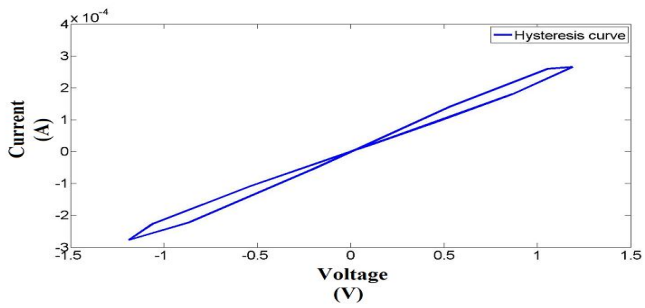


Fig. 2 Hysteresis curve of memristor

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

TCAM [5] is a specialized type of high-speed memory that searches its entire contents in a single clock cycle. Data stored on a TCAM can be accessed by comparing the input search data against a table of stored data and returning the address [6] of the matching data. TCAMs deal with three states (high, low and mask states). The third state i.e. the mask state can be used to search a portion of the word when necessary, instead of searching the whole word. Thus this includes greater flexibility in the search operation.

II. Proposed 8 Transistor 4 Memristor TCAM Cell

We have designed a TCAM cell with a reduced number of transistors compared with the conventional TCAM cell. We have designed a hybrid structure that contains 8 transistors and 4 memrisors whereas a conventional TCAM cell contains 16 transistors.

A. Basic Cell

The proposed cell is shown in Fig. 3. Each cell can store three states namely high, low and mask states. The high, low and mask states are denoted by DATA1 DATA2= 10, DATA1 DATA2= 01 and DATA1 DATA2= 11 respectively. The two Memristors MEM3 and MEM4 are connected in such a way that their memristance is high when current flows through them. For simulation purpose we have used R_{OFF}=100k for MEM3 and MEM4; R_{ON}=100 and R_{OFF}=100k for MEM1 and MEM2. Let us consider the case when DATA1 DATA2 =01, then DT1 DT2 =10. Since, DATA1 DATA2 = 01 is written, resistance of MEM1 goes high e.g. R_{OFF}=100k and resistance of MEM2 goes low e.g. R_{ON}=100.

During search cycle, R_{MEM3}=100k and R_{MEM4}=100k, R_{MEM1}=100k and R_{MEM2}=100

Voltage at DT1 will be,

$$V_{DT1} = \frac{R_{MEM1}}{R_{MEM1} + R_{MEM3}} V_{DD} \tag{4}$$

which is 1.65V, a high voltage.

Voltage at DT2 will be,

$$V_{DT2} = \frac{R_{MEM2}}{R_{MEM2} + R_{MEM4}} V_{DD} \tag{5}$$

which is 0.00329V, low voltage. Thus 2 bits are stored.

B. Precharge Unit

The proposed precharge unit (Fig. 4) charges the ML line to VDD prior to search operation. Hence, any mismatch will discharge the matchline to ground voltage and any match will keep the matchline charged to VDD.

C. Sense Amplifier Unit

The proposed Sense Amplifier unit contains a NOT gate and a pass transistor. The sense amplifier detects the state of ML line during search. The pass transistor holds its output value when search operation is not performed. A single row of our proposed TCAM array is shown in Fig. 4. Each row consists of 64 TCAM cells, a precharge unit and a sense amplifier unit.

III. Circuit Operation

The circuit operates in two phases, write and search. These phases are described as follows.

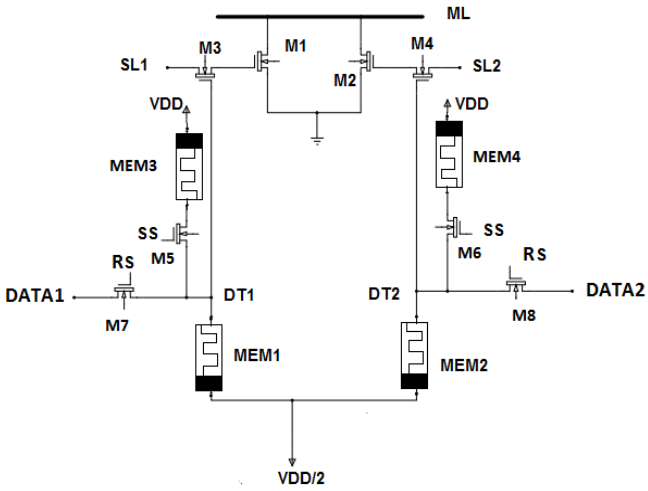


Fig. 3 Eight transistor-four memristor proposed TCAM cell.

A. Write

We have designed a 64 bit TCAM array. The Row Select (RS) signal is used to select a particular row where we want to write a 64 bit data. During write operation, Search Select (SS) signal is set to low state while RS is set to high state (Fig. 3). When a high voltage is applied to data line, the memristance connected to that line goes low and voltage across it is also low. However, a low voltage in the data line causes the memristance and hence the voltage drop across it to go high. The circuit that results during write is shown in Fig. 5(a).

When for example, DATA1 DATA2=10, the memristance 1 and memristance 2 will be low and high respectively. Hence DT1=0 and DT2=1. When DATA1 DATA2=01, DT1=1 and DT2=0.

B. Search

During search the Search Select (SS) signal (Fig. 3) is set to HI state and RS is set to low. Hence the write signals DATA1 and DATA2 are cut off from the rest of the circuit and only search occurs. Then, we get the circuit configuration as in Fig. 5(b).

The search data are placed on search lines SL1 and SL2. These values are compared with the stored data in TCAM cell. A match is denoted by the condition DATA1 DATA2= SL1 SL2. This will keep the match line ML charged to V_{DD}. This is sensed at the sense amplifier output as a low voltage. When a mismatch occurs i.e. DATA1 DATA2 ≠ SL1 SL2, the ML line is discharged to ground voltage. This is detected as a high voltage at the sense amplifier output.

IV. Simulation Results For 64 Bit TCAM Array

The simulation is performed for 64 bit array. The applied Write Select (WS) signal is of 3n duration. SS is of 2n duration. VDD/2 is 1.65 volt during write operation. As soon as write is finished, VDD/2 becomes zero.

For the purpose of analysis the graphs for cell 22 is plotted. A mismatch occurs during the second cycle of search signal in cell 22.

A. Match

During the first cycle of write, DATA1 DATA2=01 is written to cell 22. This is shown in Fig. 6. As a result, R_{MEM1} goes high and R_{MEM2} goes low. During the search cycle SL1 SL2=01. Hence a match occurs and the matchline remains at high voltage. This is indicated in the plot of ML in Fig. 7.

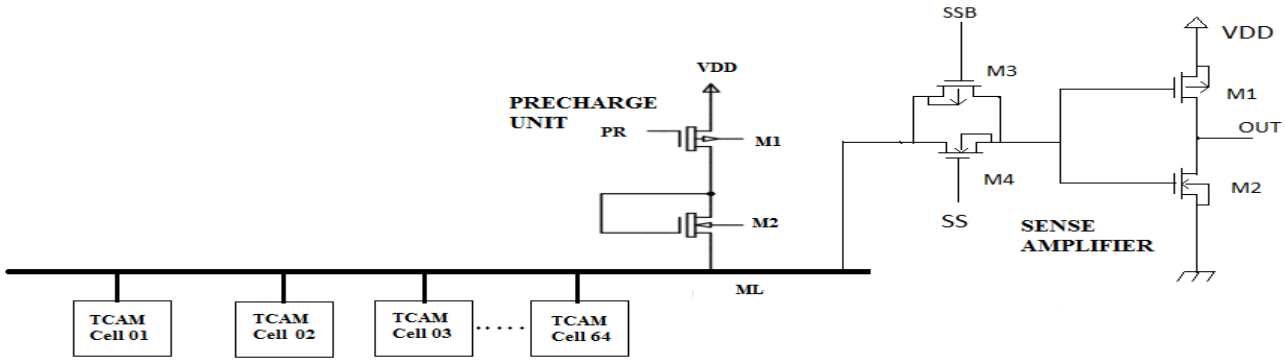


Fig. 4 Block Diagram of the proposed 64 bit TCAM array.

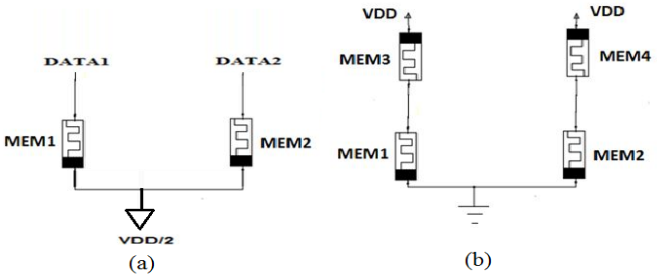


Fig. 5 (a) Circuit when SS=0, RS=1 (b) Circuit when SS=1, RS=0.

Since ML is high, the sense amplifier output OUT goes low.

B. Mismatch

During the second cycle of write, DATA1 DATA2 =10 is written to the TCAM cell 22 (Fig. 6). Hence, R_{MEM1} goes low and R_{MEM2} goes high. But during the search cycle SL1 SL2=01. Hence a mismatch occurs and ML is discharged to ground voltage. At the same time the sense amplifier output goes high (Fig. 7).

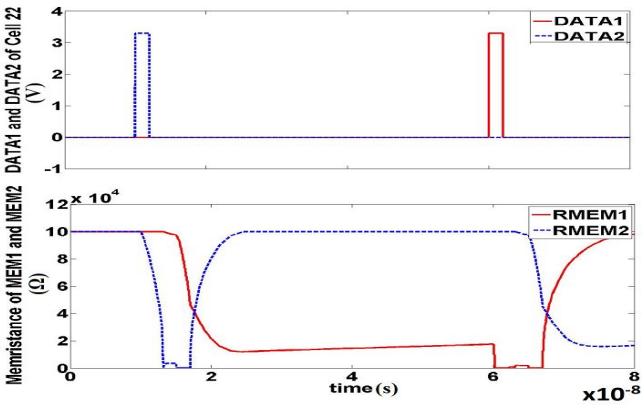


Fig. 6 Write operation in a 64 bit TCAM array

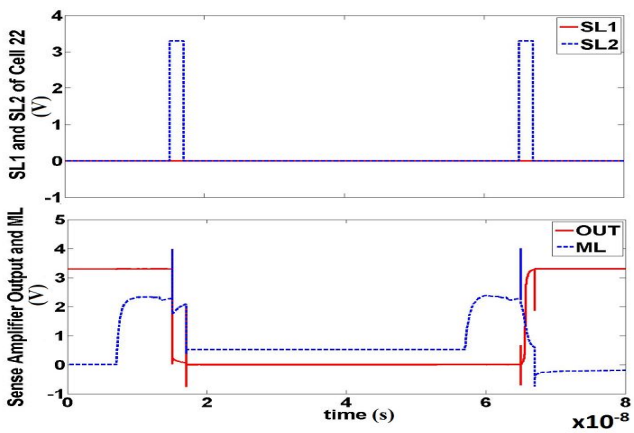


Fig. 7 Search operation in a 64 bit TCAM array

V. Three Simulation Corners

Our proposed design is compared with the conventional design with respect to speed, energy per bit and voltage margin. All these criteria are tested for three design corners i.e. typical, fast and slow. The process parameters are given below.

TABLE I.
Design Parameters for Three Corner Simulations

		Typical	Fast	Slow
VTH0	**NMOS**	0.366247V	0.329622V	0.402872V
	PMOS	-0.390601V	-0.351541V	-0.429661V
Temperature		27 ºC	0ºC	100 ºC
Supply voltage		3.3V,1.65V	3.63V,1.815V	2.97V,1.485V

VI. Comparison of Performance Parameters

A. Speed and Energy Per Bit

Search time during full-match is calculated by taking the time difference between 50% rise time of Search Enable (SE) signal and 50% fall time of the sense amplifier output (OUT).

TABLE II.
Comparison of Performance Characteristics
(For Search Cycles)

		16T NOR Type (Conventional design)	8T 4M NOR Type (Proposed design)
TYPICAL MODE			
Charging (full match)	Time	0.626 ns	0.026 ns
	Energy per bit	374.8 fJ	7.7 fJ
Discharging (1-bit mismatch)	Time	1.06 ns	0.94 ns
	Energy per bit	573.66 fJ	268 fJ
FAST MODE			
Charging (full match)	Time	0.495 ns	0.024 ns
	Energy per bit	417 fJ	165 fJ
Discharging (1-bit mismatch)	Time	0.892 ns	0.65 ns
	Energy per bit	615.92 fJ	1.1 pJ
SLOW MODE			
Charging (full match)	Time	0.776 ns	0.031 ns
	Energy per bit	325.75 fJ	120fJ
Discharging (1-bit mismatch)	Time	1.13 ns	1.3 ns
	Energy per bit	426.32 fJ	1.1 pJ

The time needed for changing the voltage level during write cycle is very low. However for this voltage to be stored even after the power supply is turned off, it is necessary to change the states of the memristors according to the data. So we have kept the write signal (RS) enabled for a longer time (irrespective of data). As we have defined the write time to be the entire write enable period, the write time for different data combination is considered to be the same. Though, this fact causes our design to require more time and energy for write operation compared to the conventional design, it is expected that these criteria can be improved by using memristors with better model parameters.

TABLE III.
COMPARISON OF PERFORMANCE CHARACTERISTICS (FOR WRITE CYCLES)

Data		16T NOR Type (Conventional design)	8T 4M NOR Type (Proposed design)
TYPICAL MODE			
0 1	Time	0.145 ns	1.75 ns
	Energy per bit	0.873 pJ	1.2 fJ
1 0	Time	0.143 ns	1.75 ns
	Energy per bit	0.901 pJ	1.1 fJ
1 1	Time	0.159 ns	1.75 ns
	Energy per bit	0.657 pJ	0.81 fJ
FAST MODE			
0 1	Time	0.135 ns	2 ns
	Energy per bit	1.178 pJ	1.77 fJ
1 0	Time	0.134 ns	2 ns
	Energy per bit	1.181 pJ	1.75 fJ
1 1	Time	0.14 ns	2 ns
	Energy per bit	0.855 pJ	1.56 fJ
SLOW MODE			
0 1	Time	0.158 ns	4 ns
	Energy per bit	0.662 pJ	1.54 fJ
1 0	Time	0.161 ns	4 ns
	Energy per bit	0.65 pJ	1.54 fJ
1 1	Time	0.164 ns	4 ns
	Energy per bit	0.438 pJ	1.42 fJ

B. *Voltage Margin*

Voltage margin is determined by taking the difference of the ML voltage level for full-match and for worst case (1 bit) mismatch. The match line is more prone to be detected erroneously in case of 1-bit mismatch. So, the higher the voltage margin, the better the design.

TABLE IV.
COMPARISON OF VOLTAGE MARGIN

	16T NOR Type (Conventional design)	8T 4M NOR Type (Proposed design)
TYPICAL MODE		
Voltage Margin	0.314 V	0.75 V
FAST MODE		
Voltage Margin	0.276 V	0.86 V
SLOW MODE		
Voltage Margin	0.326 V	0.72V

VII. PHYSICAL LAYOUT AND AREA ESTIMATION

The layout of our proposed design is drawn in Microwind 2.0 layout design and simulation software using lambda (λ) based rules at TSMC 180nm technology.

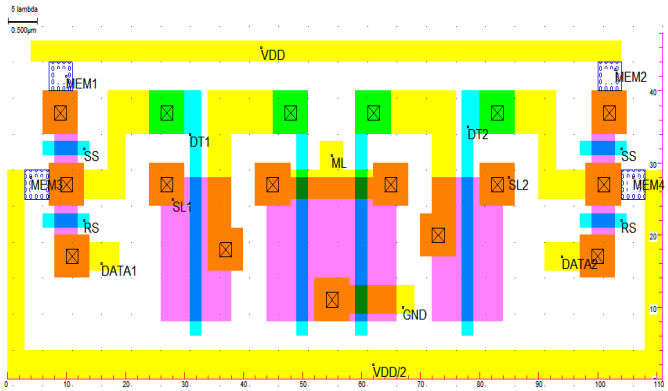


Fig. 8 Layout implementation of proposed TCAM cell.

Total area of our proposed design is found to be 48.4 μm^2. Each memristor requires an area of 0.002 X 0.002 μm^2 = 0.000004 μm^2, which is negligible with respect to the total area of the design. However, for clarifying the position of the memristors we have not drawn the memristors up to scale. Memristors are drawn somewhat larger to make them noticeable in the layout.

TABLE V.
AREA OF BASIC CELL

Parameter	**16T NOR Type (Transistor based)**	**12T NOR Type (Memristor based)**
Area	134.7675um^2	48.4μm^2

VIII. CONCLUSION

In this paper we have proposed a hybrid architecture of Ternary Content Addressable Memory (TCAM) with 8 transistors and 4 memristors. After extensive simulation it is shown that our proposed design offers greater speed, lower energy consumption, higher voltage margin and smaller area requirement than conventional TCAM design. From this observation it can be delineated that, our proposed design has the perspective of being a promising alternative for future memory based applications.

REFERENCES

[1] D.B. Strukov, G.S. Snider, D.R. Stewart, and R.S. Williams, *"The missing memristor found,"* Nature (London) 453, 80-83 (2008); J.M. Tour and T. He, "The fourth element," ibid 42-43 (2008).

[2] O. Kavehei, Y. S. Kim, A. Iqbal, K. Eshraghian, S. F. Al-Sarawi, and D. Abbott, *"The fourth element: Insights into the memristor,"* In the IEEE International Conference on Communications, Circuits and Systems, ICCCAS, pages 921–927, July 2009.

[3] Y. N. Joglekar and S. J. Wolf, *"The elusive memristor: Properties of basic electrical circuits,"* European Journal of Physics, 30(4):661, 2009.

[4] L.O. Chua, *"Memristor - the missing circuit element,"* IEEE Trans. Circuit Theory 18, 507-519 (1971).

[5] N. Mohan, *"Low-power high-performance ternary content addressable memory circuits,"* Ph.D. dissertation, Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario, Canada, 2006.

[6] Syed Iftekhar Ali and M. S. Islam, *"An energy efficient design of high-speed ternary CAM using match-line segmentation and resistive feedback in sense amplifier,"* Journal of Computers (ISSN 1796-203X) Special Issue: Selected Papers of the 13th International Conference on Computer and Information Technology (ICCIT 2010), vol. 7, no. 3, pp. 567-577, March 2012.

Achieving Tetra Band Performance in Simple Dielectric and in Metamaterial filled Patch Antennas

Abdullah Al Noman Ovi[1*], M.R.C. Mahdy[2], Md. Rashedul Alam Zuboraj[3], Ayed Al Sayem[4], Md. Ababil Hossain[4], and Md.Abdul Matin[4]

[1]Department of Electrical and Computer Engineering, *Purdue University*, USA
[2]Department of Electrical and Computer Engineering, National University of Singapore, Singapore
[3] Department of Electrical and Computer Engineering, Ohio State University, USA
[4]Department of Electrical and Electronics, Bangladesh University of Engineering & Technology, Bangladesh
[*]ovi_1000@yahoo.com

Abstract— **In this contribution, we have proposed a robust technique to achieve tetra-band performance in both conventional dielectric (elliptical shape) and also metamaterial loaded (rectangular shape) patch antennas. Incorporation of symmetric slots over the patch in both type of antennas have been shown here by using the concept of additional modified modes and perturbation approach. At first, we have shown that if and only if, elliptical shape patch antenna is properly optimized along with symmetrical slots over metallic patch, it is possible to achieve satisfactory four- band performance without the aid of any complex procedure currently present in literature. Later, we have reported a compact metamaterial loaded tetra band rectangular patch antenna to achieve extremely high gain and directivity performances for all the four bands, which may not be possible with conventional dielectric antennas.**

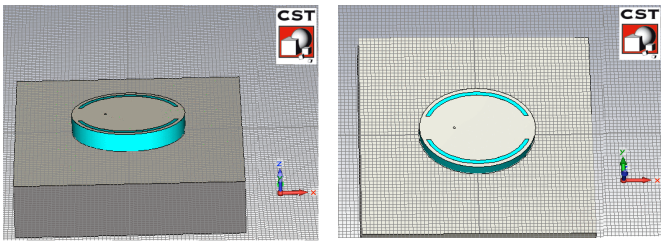
Index Terms— **Elliptical and rectangular patch antenna, Broadside radiation, metamaterial, tetra band antenna.**

I. Introduction

Multi-band performance, high directivity and size reduction are desirable for radar, satellite and in many other important communication systems such as mobile cellular and Wi-fi communications [1-5]. It would be interesting if all these qualities were achieved in a unique radiator. As huge size reduction of patch antennas is not possible with conventional ways, Alu et al. [6-9] has proposed design method to obtain electrically small patch antennas using DPS (Double Positive)-MM (metamaterial) juxtaposed-layer. But we have shown [1-5] such patches give broadside null radiation pattern at sub wavelength regime. Finally, it was predicted that all these electrically small rectangular antennas loaded with metamaterial can only be good resonators but may not be good radiators [1-5], [10]. Later, in [1] we proposed the idea of additional modified modes to design successful dual band radiators according to designers' will. Although theoretically maximum 35% size reduction is possible [2,3], electrically small patch antenna may not be possible with any shape using juxtaposed layer of MM [10]. So, the question arises: is there any real practical benefit with these metamaterial based small antennas [11]? A probable solution of this problem is using multi-band (or multi-functional) antennas instead of reduced size antennas. In literature, there are several designs of multi-band patch antennas with conventional size. However, a notable fact is that many of them [12, 13] do not fulfil the criteria of -10 dB S-parameter performance to consider it a real frequency band among some of their multi bands [12]. Moreover, any cavity that does not radiate at broadside direction with gain performance more than '0' dB and also with not maximum directivity at broadside may not be considered as a successful microstrip patch antenna (i.e. see [13]). As a result, only few 'successful' designs of tetra band PIFA antennas with complex designs have been demonstrated in literature [14], [15] so far. But as per we know, for such four-band performance, conventional dielectric filled microstrip patch antennas have not been demonstrated in literature. In this article, at first we have demonstrated a tetra band elliptical microstrip patch antenna with quite satisfactory gain & directivity, broadside radiation and S-parameter performance. Notable fact is that our proposed design is extremely simple and probably will be very useful for commercial designs. As per we know, due to the use of conventional material, the gain and directivity performance are not extremely high of the antennas present in current literature. As a result, here we have also demonstrated a rectangular patch antenna with ENG MM to achieve extremely high broadside radiation performance (directivity and gain). Moreover, in metamaterial loaded 2nd antenna, designers can choose the resonant frequency according to their own will [1-5]. Such degrees of freedom for getting desired four-band performance 'according to the designers will' may not be available in literature. Here we have also analysed the robustness of such antennas by varying some important antenna parameters.

II. Tetra Band Elliptical Patch Antenna Filled With Conventional Dielectric Material

This antenna is a simple tetra band antenna that needs no other materials than the substrate itself which is of single layer and composed of normal dielectric material found in nature. It needs only symmetrical slots to obtain higher order modes over the metallic patch. It shows tetra band characteristics with quite acceptable radiation performance. The geometry and parameters of that antenna are given below:



(a) (b)

Fig 1: Geometry of the elliptical microstrip patch antenna (a) side view (b) top view.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

TABLE I
ELLIPTICAL MICROSTRIP PATCH ANTENNA MEASUREMENTS FOR DESIGNING TETRA BAND

Ground plane	Elliptical patch & substrate	Substrate height	Substrate's relative permittivity, ε_R	Substrate's relative permeability, μ_R	Feed position	Probe radius
60mm×60mm	Major axis= 15mm Minor axis= 12 mm	5 mm	9.8	1	x=-6, y=0	0.3 mm

From $|S_{11}|$ Parameter analysis (Fig 2) it has been observed that microstrip antenna with elliptical patch giving tetra band resonance frequency at 2.2 GHz, 5.22 GHz, 6.22 GHz & 6.77 GHz.

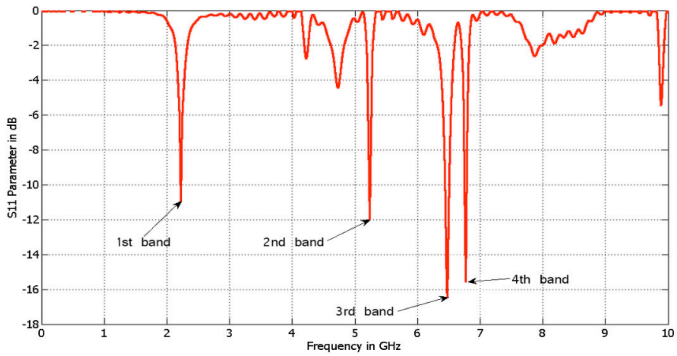


Fig 2: S-parameter performance of proposed elliptical tetra band microstrip patch antenna.

The radiation patterns found from the full wave simulation (Fig 3-6) is describing that at broadside the antenna shows maximum radiation. From both the 3D and polar plot, they provide satisfactory radiation performance at all bands without adding any complexities to the design. For a simple dielectric elliptical antenna, the expressions for the TE^z and TM^z modes can be written as (before the incorporation of the slots):

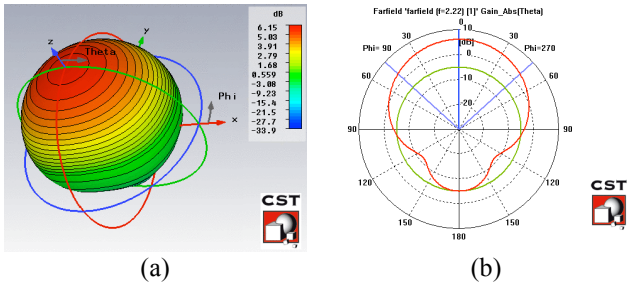
TE^z modes:

$$H_z(\xi,\eta,z)=\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\sum_{p=0}^{\infty}\{A^{TE}_{e_{nmp}}R^{TE}_{e_{nm}}(q^{TE}_{mn},\xi)S^{TE}_{e_{nm}}(q^{TE}_{mn},\eta)+A^{TE}_{o_{nmp}}R^{TE}_{o_{nm}}(q^{TE}_{mn},\xi)S^{TE}_{o_{nm}}(q^{TE}_{mn},\eta)\}\sin(k_{z_p}z) \quad (1)$$

TM^z modes:

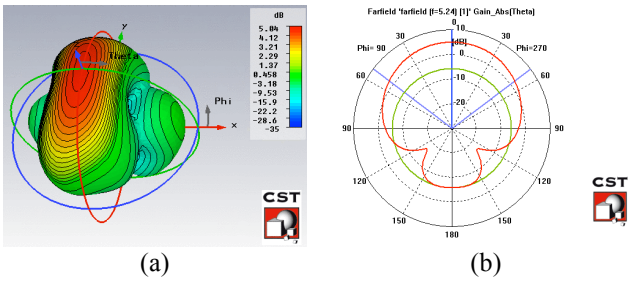
$$E_z(\xi,\eta,z)=\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\sum_{p=0}^{\infty}\{A^{TM}_{e_{nmp}}R^{TM}_{e_{nm}}(q^{TM}_{mn},\xi)S^{TM}_{e_{nm}}(q^{TM}_{mn},\eta)+A^{TM}_{o_{nmp}}R^{TM}_{o_{nm}}(q^{TM}_{mn},\xi)S^{TM}_{o_{nm}}(q^{TM}_{mn},\eta)\}\sin(k_{z_p}z) \quad (2)$$

where, A^{TE}_{onmp}, A^{TE}_{enmp}, A^{TM}_{onmp} and, A^{TM}_{enmp} are unknown expansion coefficients and the subscript e and o are used to refer even and odd mode respectively [16]. Except the first harmonic (n=0), all modes can be even or odd. The elliptical shape (θ and φ being different) is one of the reasons, which gives rise to different even and odd mode. Even mode and odd mode come from the horizontal and vertical polarisation of the antenna which are not same in case of elliptical shape. When they are seperated by a sufficient distance, they act like individual mode.



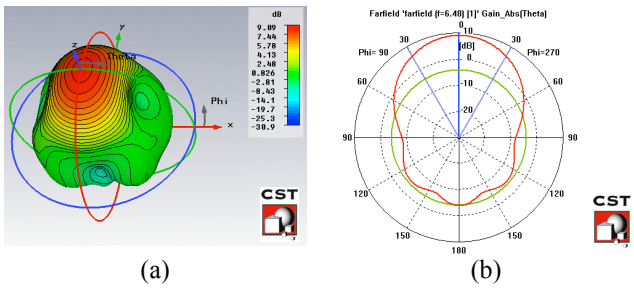
(a) (b)

Fig 3: (a) 3-D radiation pattern (b) Polar plot at 2.22 GHz



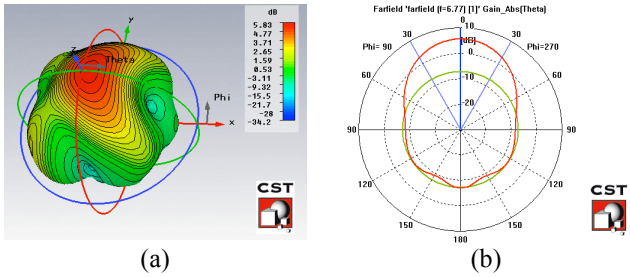
(a) (b)

Fig 4: (a) 3-D radiation pattern (b) Polar plot at 5.24 GHz



(a) (b)

Fig 5: (a) 3-D radiation pattern (b) Polar plot at 6.48 GHz



(a) (b)

Fig 6: (a) 3-D radiation pattern (b) Polar plot at 6.77 GHz

In addition, we have found that the impedance matching performance of elleptical shaped antennas is better than other shaped [1-5] patch antennas if the feed position is properly optimized. Thus dual band performance is obtained just due to the elliptical shape. Using boundary conditions in [16], the resonant frequencies can be written before adding slots:

$$f_{e_{nmp}/o_{nmp}}^{\binom{TE}{TM}} = \frac{c}{2\pi\varepsilon_r}\sqrt{k_{c_{enm/onm}}^{\binom{TE}{TM}2} + k_{zp}^{\ 2}} \qquad (3)$$

which can also be expressed in the following form:

$$f_{e_{nmp}/o_{nmp}}^{\binom{TE}{TM}} = \frac{c}{2\pi\varepsilon_r}\sqrt{\frac{4q_{mn}^{\binom{TE}{TM}}}{f_o^{\ 2}} + k_{zp}^{\ 2}} \qquad (4)$$

The additional arc shaped parallel slits are responsible for further additional modified modes [2-4]. Maci et al. [17] showed that two parallel slots close to the radiating edges can cause mode modification and by effective current path elongation, it is possible to get multiband performance. In elliptical patch antennas, similar concept to get mode modification will be found in our previous work [18]. Here, we have used that idea to get further mode modification to obtain tetra band performance. The idea of splitted modes proposed by Alu et al. [8,9] is observed in our proposed simple elleptical antenna for 3rd and 4th band due to the additional slots. The generalized perturbation equation can be applied to model the slight frequency shifts for the 1st and 2nd band due to those slots:

$$\frac{\Delta k}{k_0} \approx \frac{\Delta W_e - \Delta W_h}{\Delta W_e + \Delta W_h} = \frac{(\frac{1}{2}\int_{\Delta v}\varepsilon|E|^2 dv) - (\frac{1}{2}\int_{\Delta v}\mu|H|^2 dv)}{(\frac{1}{2}\int_{\Delta v}\varepsilon|E|^2 dv) + (\frac{1}{2}\int_{\Delta v}\mu|H|^2 dv)}$$

The left most term indicates the frequency shifts criteria [19]. *Kzp* written in Eq (1) - (4) can be recalculated using above equation after the incorporation of the symmetric slots.

III. TETRA BAND RECTANGULAR PATCH ANTENNA PARTIALLY LOADED WITH ENG MM FOR HIGH GAIN AND DIRECTIVITY

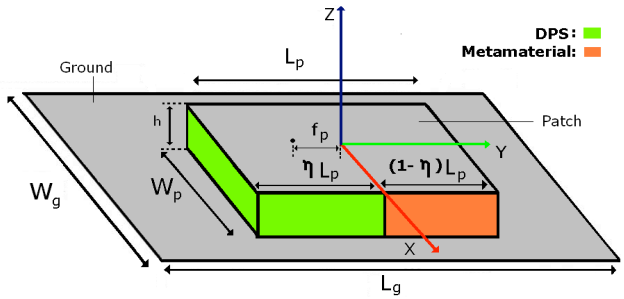


Fig. 7: Geometry of a un-slotted rectangular microstrip patch antenna partially loaded with metamaterial (ENG, MNG or DNG) [1]. f_p= feed position from DPS and metamaterial interface. The filling ratio of the DPS is η (i.e. the length of DPS block under the patch is ηL). The underneath substrate (which is $L_p \times W_p \times h$ in volume) is inhomogeneous, filled with two isotropic and homogeneous materials with permittivity and permeability ε_1, μ_1 and ε_2, μ_2. Substrate height is h.

According to J.Xiong et al. [20], any etched slot in the metallic patch may deteriorate the performance of a rectangular patch antenna loaded with metamaterial. But here we have used two symmetrical etched slots like [17] to introduce the concept of additional higher modified $TM_{0\delta0}$ (3<δ<4) mode(s). On the other hand, if we use such symmetric slots for MNG MM loaded dual band antennas [1]; we fail to get Tetra band performance from them. However, to get tetra band performance with ENG Metamaterial loaded antennas, we follow similar procedure followed by us in [5].

A very important fact is that the parameters of this tetra band design should be optimized very carefully. If the positions of f_0 and f_r are changed, it may behave like a triple band antenna proposed in [5]. Actually if f_r and f_0 are not far away from TM_{030} mode; bulk resonance & effective high value of permittivity of ENG MM may destroy the modified TM_{030} mode radiation performance and the antenna behaves like a triple band antenna proposed by us in [5]. Here, the detailed design procedure is given of this highly directive antenna:

(i) Firstly, we have adopted a conventional 40mm × 35mm patch to design such an antenna (not MM loaded). Using proposed idea of Maci et al. [17] and taking ε_r= 2.2, we find dual band performance with conventional TM_{010} & TM_{030} mode (modified like TM_{010} mode) at 2.37 GHz, 3.54 GHz and respectively. We get TM_{040} mode at 5.64 GHz (having broadside null rad. pattern). (ii) Now, to introduce an additional modified mode we have to use metamaterial. As we are targeting to get our desired mode over modified TM_{030} mode (behaves like conventional TM_{010} mode [1]), we have to use ENG MM. This is because in [1], we have shown that if we want to get any additional mode higher than conventional TM_{010} mode, we have to use ENG MM (Fig. 2 of [1]). So, we have chosen f_r= 4.5 GHz here (far from TM_{030} mode found in 3.54 GHz). Then using our proposed idea [1] & algorithm [21], we have derived necessary design parameters (given in the caption of Fig. 8). Lorentz dispersive lossy model [2],[4] has been used for ENG metamaterial loading with parameter values: ε_∞ = 1.0, ε_s= 1.1355, ω_0= 2670 rad/s, δ = 50MHz; when: relative permittivity, $\varepsilon_{2r}(\omega) = \varepsilon_\infty + \frac{(\varepsilon_s - \varepsilon_\infty)\omega_0^{\ 2}}{\omega_0^{\ 2} - \omega^2 + j\omega\delta}$ (iii)

Then, by using full wave analysis [22]; we have found two 'interface resonance' modes at 4.5 GHz and 4.75 GHz.

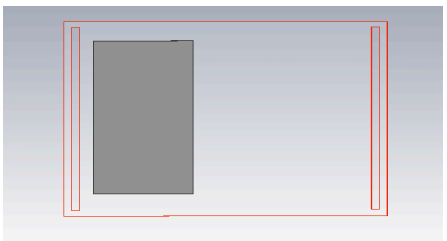


Fig. 8: proposed 40mm × 35mm ENG metamaterial loaded microstrip patch antenna with symmetrical slots and parasitic patch (Top view).Parasitic patch is underneath the DPS substrate. Parasitic patch dimension: 28.66 mm (parallel to slots or x-axis) × 14.32 mm × 0.1 mm . Distance between main and parasitic patch: 0.2 mm. Dimension of both slots: 33mm × 1 mm . Design parameters: Lp = 40 mm, Wp = 35 mm, Lg = 75 mm, Wg = 75 mm , fp = 15.272 mm, h = 1 mm, η = 0.5818 , ε_{1r} = 2.2, μ_{1r} = 1, μ_{2r} = 1 and ε_{2r} (at fr) = -0.1182 in case of ENG loading, fr = chosen resonant frequency = 4.5 GHz.

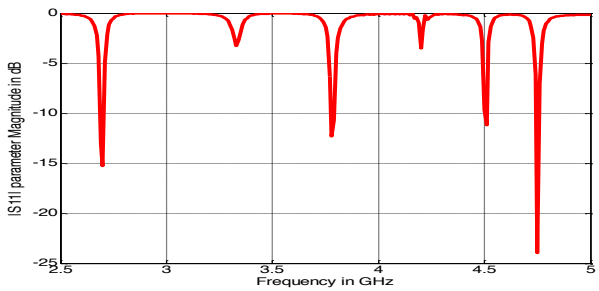


Fig. 9: $|S_{11}|$ performance of proposed antenna with parasitic patch

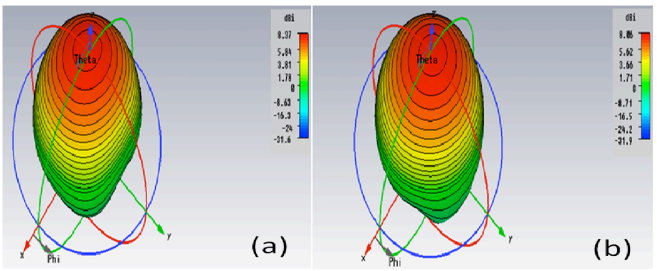


Fig. 10: 3-D radiation pattern radiation of proposed antenna (Fig. 2) (a) Conventional TM_{010} mode (at 2.7GHz); rad. efficiency 97.89% and (b) TM_{030} mode (at 3.78 GHz) ; rad. efficiency 77.678%.

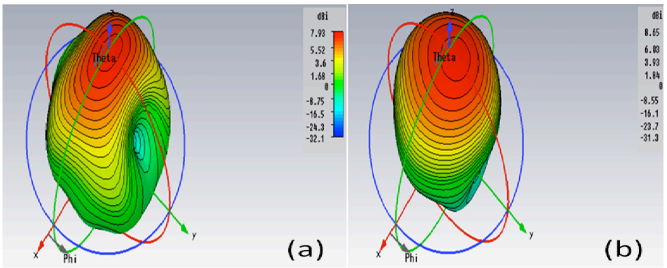


Fig. 11: 3-D pattern of proposed antenna (Fig. 2) (a) Unconventional $TM_{0\delta 0}$-^{1}mode (at 4.51GHz); rad. efficiency 67.67% and (b) $TM_{0\delta 0}{}^{2}$mode (at 4.75GHz); rad. efficiency 73.60%.

To solve the problem of bandwidth enhancement and impedance matching, we have applied the concept of parasitic patch [5] in this contribution to minimize the inductive effect caused by probe, ENG medium and mutual coupling between probe and patch. In this case, the parasitic patch must have larger dimension parallel to symmetrical slots (Fig. 8). Satisfactory $|S_{11}|$ Parameter performance of this antenna with this additional parasitic patch is shown in Fig. 9.

It is a noticeable fact of ENG loaded tetra band antenna that TM_{030} mode was in 3.54 GHz without metamaterial loading (using normal DPS with symmetric slots). After the loading of ENG metamaterial, the two modified modes are found at 4.51 GHz and 4.75 GHz. Both modes are quite far from 3.54 GHz. We are tagging these modes as $TM_{0\delta 0}{}^{1}$ mode and $TM_{0\delta 0}{}^{2}$ mode. The directivities at 2.7 GHz, 3.78 GHz, 4.51 GHz and 4.75 GHz are 8.374 dBi, 8.057 dBi, 7.925 dBi and 8.655 dBi respectively at broadside (Fig. 10 and 11). These extremely high directivities at broadside for all the four bands are the most important features of this compact sized antenna. Last but not the least, the practical implementation of such multi-band designs ([5] & this article) can be investigated with the MM inclusions proposed in [23], [24].

IV. CONCLUSION

In this article, we have shown one conventional and one metamaterial loaded design of 'successful' tetra band patch antenna. Although we have made only theoretical investigation; full wave analysis based simulations [22] and robustness analysis strongly bear the possibility of practical implementation of such antennas. In contrast, our proposed elliptical shaped conventional dielectric filled patch antenna can be considered very convenient for commercial designs because of its extremely simple design along with satisfactory four-band performance.

REFERENCES

[1] M. R. C. Mahdy, M. R. A. Zuboraj, A. A. N. Ovi, and M. A. Matin,"An Idea of Additional Modified Modes in Rectangular Patch Antennas Loaded With Metamaterial", *IEEE Antennas and Wireless Propagation Letters*, vol. 10, pp. 869-872. Aug. 2011.

[2] Saimoom Ferdous, Md.Ababil Hossain, Shah Md Mahmud Hasan Chowdhury, Mahdy Rahman Chowdhury Mahdy and Md. Abdul Matin, "Reduced And Conventional Size Multi-Band Circular Patch Antennas Loaded With Metamaterials", *IET Microwaves, Antennas & Propagation,* Vol. 7, issue 9, pp. 768 – 776,18 June 2013.

[3] Shah Md Mahmud Hasan Chowdhury, Ababil Hossain, Saimoom Ferdous, Mahdy Rahman Chowdhury Mahdy, Md. Abdul Matin, "Conceptual and Practical Realization of Reduced size Multi-band Circular Microstrip Patch Antenna Loaded with MNG Metamaterial", *7th International Conference on Electrical and Computer Engineering (ICECE),* pp 834-837, Dec. 2012.

[4] M. H. Ababil, M. F. Saimoom, S. M. H. Chowdhury, M. R. C. Mahdy, M. A. Matin "Highly Directive Epsilon Negative Metamaterial-Loaded Circular Patch Antenna for Triple-Band Performance*", Ultra-Wideband, Short-Pulse Electromagnetics 10,* pp 219-228, 2014.

[5] M. R. C. Mahdy, M. R. A. Zuboraj, A. A. N. Ovi and M. A. Matin, "Novel Design of Triple Band Rectangular Patch Antenna Loaded with Metamaterial," *Progress In Electromagnetics Research Letters*, Vol. 21, 99-107, 2011.

[6] Alù, A.; Bilotti, F.; Engheta, N.; and Vegni, L.; "Subwavelength, Compact, Resonant Patch Antennas Loaded with Metamaterials," *IEEE Trans. Antennas Propag.*, vol. 55, no. 1, pp. 13–25, Jan. 2007.

[7] Bilotti, F.; Engheta, N.; and Vegni, L.; "Design of Miniaturized Metamaterial Patch Antennas with μ-negative Loading," *IEEE Trans. Antennas Propag.*, vol. 56, no. 6, pp. 1640–1647, Jun. 2008.

[8] Chen, P. Y. and Alù, A.; "Sub-wavelength Elliptical Patch Antenna Loaded with μ-negative Metamaterials," *IEEE Trans. Antennas Propag.*, vol. 58, no. 9, pp. 2909–2919, Sep. 2010.

[9] Chen, P. Y. and Alù, A.; "Dual-Mode Miniaturized Elliptical Patch Antenna With μ -Negative Metamaterials," *IEEE Antennas and Wireless Propagation Letters*, vol. 9, pp. 351 - 354, 2010.

[10] Jahani, S., Rashed-Mohassel, J., Shahabadi, M.: 'Miniaturization of Circular Patch Antennas using MNG Metamaterials', *IEEE Antennas Wirel. Propag. Lett.*, 2010, **9**, pp. 1194–1196.

[11] Ikonen, P.; "Electrically Small Metamaterial-based Antennas — have we seen any real practical benefits?"; *3rd European Conference on Antennas and Propagation (EuCAP),2009,* pp 866-869.

[12] Abdel Fattah Sheta, Ashraf Mohra and Samir F. Mahmoud, "Multi-band Operation of a Compact H-shaped Microstrip Antenna", *Microwave and Optical Tech. Lett.,*Vol 35, Issue 5, pp 363–367, 2002.

[13] J.S.Row, "A Triple-Band Microstrip Patch Antenna," *Microwave and Optical Tech. Lett.* , Vol. 38, No. 2, pp 120-123, 2003.

[14] Kearney, D; John, M. and Ammann, M. J., "Miniature Ceramic PIFA for UWB Band Groups 3 and 6," *IEEE Antennas and Wireless Propagation Letters,* Vol. 9, 2010

[15] Sanz-Izquierdo, Benito; Batchelor, John C.; Langley, Richard J. and Sobhy, Mohammed I.; "Single and Double Layer Planar Multiband PIFAs," *IEEE Transactions on Antennas and Propagation*, Vol. 54, No. 5, May 2006.

[16] Tadjalli, A. and Sebak, A. "Resonance Frequencies And Far Field Patterns Of Elliptical Dielectric Resonator Antenna: Analytical Approach", *Progress In Electromagnetics Research journal, JPIER* 64, 81–98, 2006

[17] Maci, S; Gentili, G. Biffi; Avitabile, G.; "Single-Layer Dual-Frequency Patch Antenna," *Electronics Letters*, vol. 29, no. 16, 1993.

[18] Ovi, A. A. Noman; Saha, N; Dey S. and Alam, N. N.; "Symmetrical Slot Loaded Dual Band Elliptical Microstrip Patch Antenna," *PIERS Proceedings,* 538-541, August 19-23, Moscow, Russia, 2012.

[19] R. F. Harrington, "Time Harmonic Electromagnetic Fields", *John Wiley & Sons, Inc*, 2001.

[20] Xiong, J.; Li, H.; Jin, Y. and He, S.; "Modified TM_{020} Mode of a Rectangular Patch Antenna Partially Loaded with Metamaterial for Dual-Band Applications," *IEEE Antennas and Wireless Propagation letters,* vol. 8, pp. 1006–1009, 2009.

[21] M. R. C. Mahdy, M. R. A. Zuboraj, A. A. N. Ovi and M. A. Matin.,"A Novel Design Algorithm' And Practical Realization of Rectangular Patch Antenna Loaded With SNG Metamaterial," *Progress In Electromagnetics Research M*, Vol. 17, 13-27, 2011.

[22] CST Microwave Studio 2009, CST of America [Online]. Available: http://www.cst.com.

[23] Imhof, Pascal D.; Ziolkowski, Richard W. and Mosig, Juan R.; "Highly Subwavelength Unit Cells to Achieve Epsilon Negative (ENG) Metamaterial Properties," *APS international symp.*,2006,IEEE

[24] Anne F. de Baas, *Nanostructured Metamaterials*. Luxembourg: Publications Office of the European Union, 2010.

Design and Evaluation of a Novel Octagonal UWB Wearable Textile Antenna under Different Bending Conditions

Shahriar Hasan Shehab[1*], Shahreen Hassan[2] and Farhadur Arifin[3]

[1,2,3]Department of Electrical and Electronic Engineering, American International University-Bangladesh (AIUB)
Banani, Dhaka-1213, Bangladesh.
[*]Shahriar.hasan.shehab@gmail.com

Abstract— **In this paper a novel Ultra Wide Band textile antenna is designed which operates from 2.6–14 GHz, covering entire Ultra Wide Band frequency range of 3.1–10.6 GHz, as approved by FCC. The proposed antenna consists of a octagonal patch and a partial ground plane with a square notch. The substrate of the designed antenna is made of Dacron fabric (ε_r=3) while patch and ground is made of copper tape. The dimension of the substrate is 43×34 mm^2. Simulated results of different antenna parameters such as return loss (S11), gain and radiation patterns are presented. This antenna performance are also analysed at four different bending angles. The characteristics, low profile and compact size of the proposed antenna depicts as a potential candidate for UWB applications. The antenna design details and simulated results are presented by CST microwave studio.**

Index Terms—**Ultra Wide Band, Textile Antenna, Octagonal patch, Partial ground.**

I. Introduction

In recent years, the rapid development of wearable intelligent textile systems has increased the necessity for WBAN (Wireless Body area Network) system to provide standalone outfit. WBAN is a network where its nodes are placed close to the body which enables wearable computer devices to exchange digital information with each other using the electric conductivity of human body as a data network [1].

The antenna plays a very basic part of Wireless Body Area Network. In this modern era, body worn antenna is one of the most popular and cutting edge research areas. Body worn antennas may be made from textile materials [2] and fixed on body, into clothing or may be worn as a button antenna [3]. It can even be implanted inside a human body [4]. Textile antenna system has a wide variety of applications. It is used for detecting motion on the body during exercise, monitoring heart rate and blood pressure, emergency services such as fire fighters, detective and police [5]. It also helps to communicate among the soldiers of the modern battle field including unmanned vehicles [6]. Many researchers have reported single frequency band wearable antennas [7] and later dual frequency band wearable antennas [8]. In the recent researches Ultra Wide Band technology is one of the most favourable choices for WBAN applications. The Ultra Wide Band (UWB) system covers the frequency range from 3.1 – 10.6 GHz, which is based on narrow pulses, low power consumption and looks similar to random noise to conventional radio systems. The transmitter and receiver have simple structure because there is no carrier frequency instead UWB emits timed pulses of electromagnetic energy which is necessary for portable devices.

However, in case of wearable antennas it is quite difficult to obtain large bandwidth with smaller size and a very low thickness. As the antenna is mounted on textile material it is also difficult to operate the antenna when it is bent frequently.

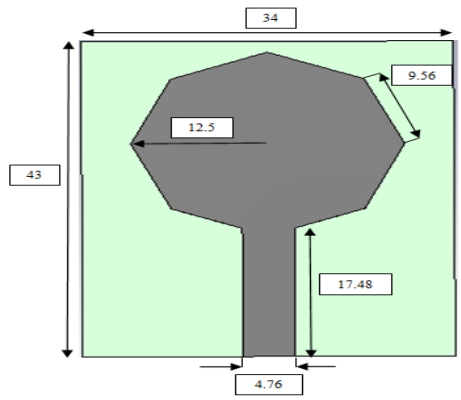
In this paper a novel design of compact octagonal UWB textile antenna is proposed. Also bending of the antenna for different angles is focused. The outline of the paper is structured as followings. In section 2 describes the antenna geometry. Bending of the proposed is shown in section 3. The antenna performances are evaluated in section 4 and compared the performances under several angles of bending in section 5. A conclusion is given is section 6. The antenna structures are designed and analysed using CST Microwave Studio package.

II. Antenna Design

The proposed octagonal UWB textile antenna operates from the frequency of 2.6 GHz to 14 GHz. The antenna is designed on hypalon coated Dacron fabric substrate with thickness of 1.68 mm. Selected textile material has high tensile strength, high resistance to stretching, durability, outstanding electric property. The specification of the substrate for dielectric permittivity, ε_r is 3 and tangent loss is 0.002. The octagonal patch of the antenna is fed by a 50 ohm microstrip feed line with a width of 4.76 mm. The radius of the patch is calculated using equation (1),

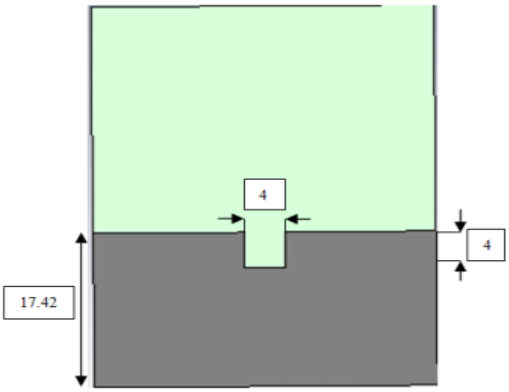
$$a = {87.94}/{f_r\sqrt{r}}\ , \quad \text{in mm } (f_r \text{ in GHz}) \qquad (1)\ [9]$$

Fig. 1 shows the front view (a) and back view (b) of the proposed antenna.



(a) Front View.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE



(b) Back View

Fig. 1: Geometry of the proposed UWB antenna.

The conductive octagonal plane and partial ground plane are made of copper tape. A partial ground plane is placed at the back of the substrate and the dimension is 18×34 mm^2. In addition, the partial ground has a square notch and the dimension of the slot is 4×4 mm^2. The area of the substrate is 43×34 mm^2. In Table I the antenna measurements are shown.

TABLE I
MEASUREMENTS OF OCTAGONAL UWB TEXTILE ANTENNA

Parameters	Size (mm)
Substrate Length	43
Substrate Width	34
Substrate Thickness	1.68
Patch Thickness	0.7
Ground Thickness	0.52
Feed Length	17.48
Feed Width	4.76

III. Bending of Octagonal UWB Textile Antenna

In case of textile antenna system, bending of the antenna occurs at different angles. It is mandatory for the textile antennas to keep their performances with respect to different angles of bending. In this paper the bending effect at four different angles –namely 20, 40, 60 and 80 degrees for designed octagonal UWB antenna is analysed. From the equation (2), the value of the radius (r) for the intended angle of the cylinder was calculated.

$$S = r\theta \quad \text{, in mm} \qquad (2)$$

Where, S is the arc length, r is the radius of the circle and θ is the measurement of the central angle in radians. Width of the antenna was used as the arc length. All the antenna dimensions were maintained the same in all bending angles. CST microwave studio is used to design and simulations purposes. Fig. 2, antenna measurements for bending purposes are given.

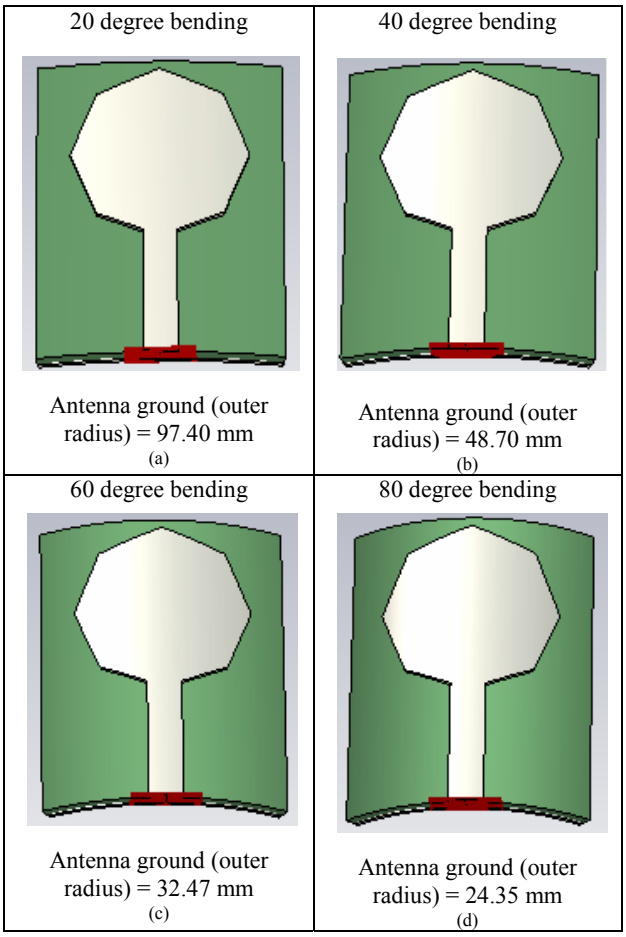


Fig. 2: Octagonal UWB Textile Antenna Measurements for Bending Purpose; 20 degrees bending (a), 40 degrees bending (b), 60 degrees bending (c), 80 degrees bending (d).

IV. Performance of the Proposed UWB Textile Antenna

A. S-parameter Analysis of UWB Textile Antenna

The simulated return loss of the proposed antenna is shown is Fig. 3.

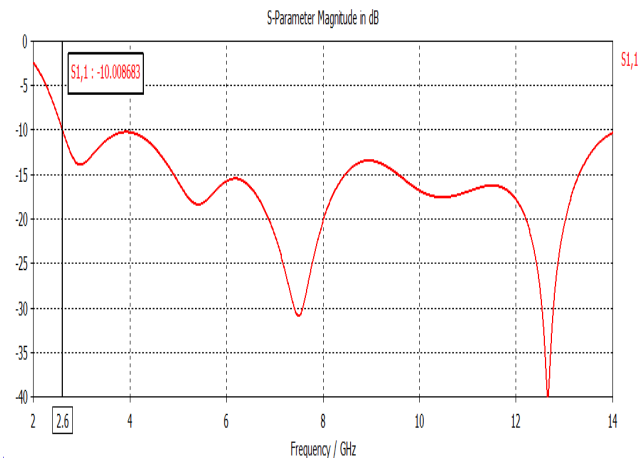


Fig. 3: Simulated return loss (S11) of the proposed antenna.

From the simulated return loss graph, it can be seen that the UWB antenna occupies a bandwidth of 11.4 GHz (2.6 GHz – 14 GHz) for a return loss (S11) less than -10 dB (VSWR < 2).

B. *Gain of The Proposed UWB Textile Antenna*

Fig. 4 shows the gain of the UWB textile antenna.

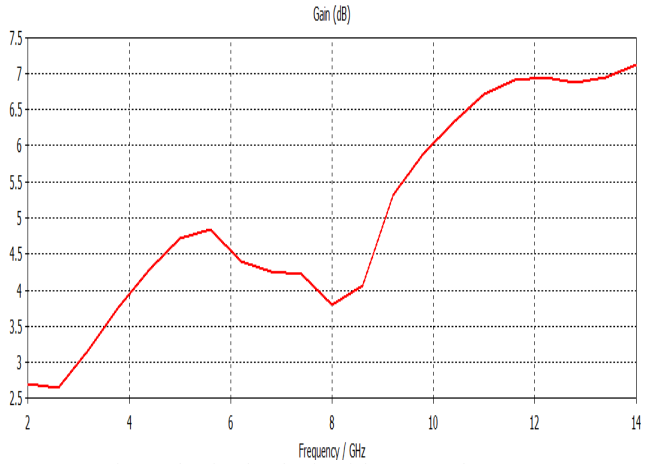


Fig. 4: The simulated gain of the proposed antenna.

In Fig. 4, it can be seen that the peak gain of the proposed antenna in a frequency range of 2 GHz to 14 GHz and gain variations are less than about 7.11 dB.

C. *Radiation Pattern for Different frequencies of UWB Textile Antenna*

The azimuth and elevation radiation pattern of the UWB antenna at 5 and 7 GHz have been shown in fig. 5. For attaining the azimuth pattern, theta (θ) has been set to 90 degree for all values of phi (φ). Similarly for elevation pattern, phi (φ) has been set to 90 for all values of theta (θ).

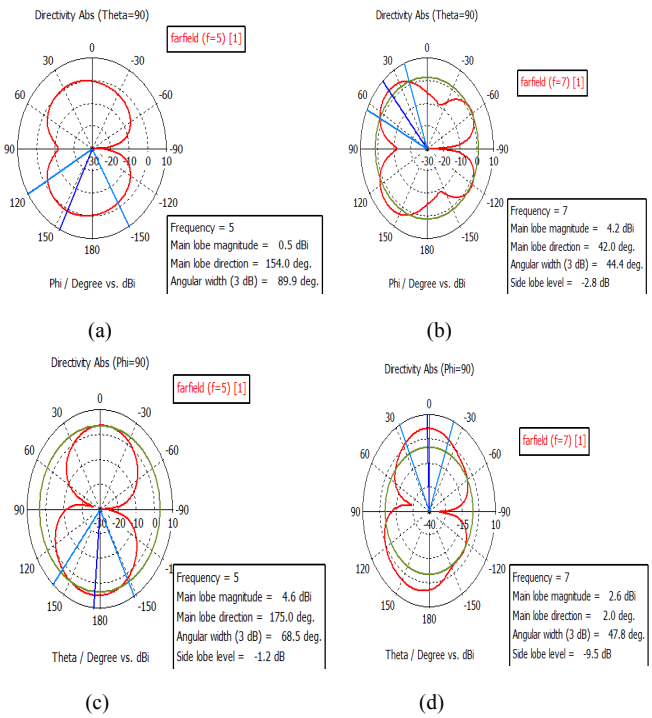


Fig. 5: Azimuth radiation pattern at 5GHz (a) and 7 GHz (b) and Elevation radiation pattern at 5GHz (c) and 7 GHz (c).

In both 5 GHz and 7 GHz the radiation characteristics shows an Omni-directional pattern but at 7 GHz a little deformation from 2 GHz pattern is observed.

V. PERFORMANCE ANALYSIS OF UWB TEXTILE ANTENNA UNDER BENDING CONSEQUENCES

A. *S-parameter Analysis of Bended Proposed UWB Textile Antenna*

In fig. 6, the simulated return loss (S11) under different angle of bending of the proposed UWB antenna is shown.

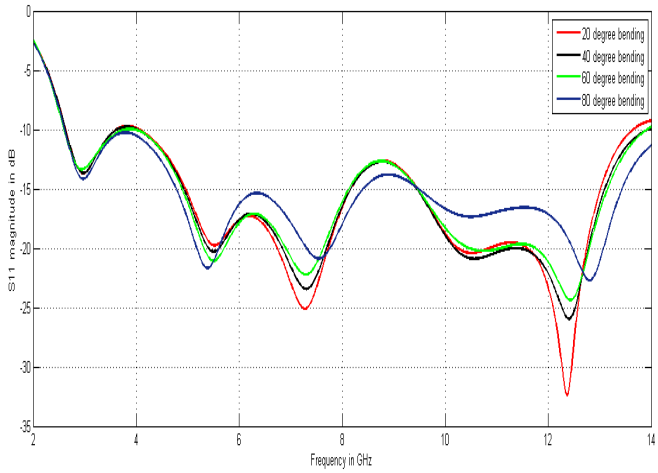


Fig. 6: The comparison of simulated return loss (S11) of the bended UWB antenna under four different angles.

Figure 6 shows the comparison of UWB textile antenna for different angle of bending. The proposed antenna shows better impedance matching with return loss below -10 dB for higher bending angles. At 20 degree bending highest return loss of -32.41 dB at frequency 12.37 GHz is achieved. However, a small portion of bandwidth lies above -10 dB at frequency 3.8 GHz where return loss is -9.7 dB. At higher angle of bending the return loss is below -10 dB and their pattern is quite similar. In 40, 60 and 80 degrees angles of bending maximum return loss of -25.97 dB, -24.39 dB and -22.73 dB are achieved respectively.

B. *Radiation Pattern of UWB Antenna at Four Different Angles of Bending*

Figure 7.1 and 7.2 illustrates the radiation pattern at 5 GHz and 7 GHz for different angles of bending.

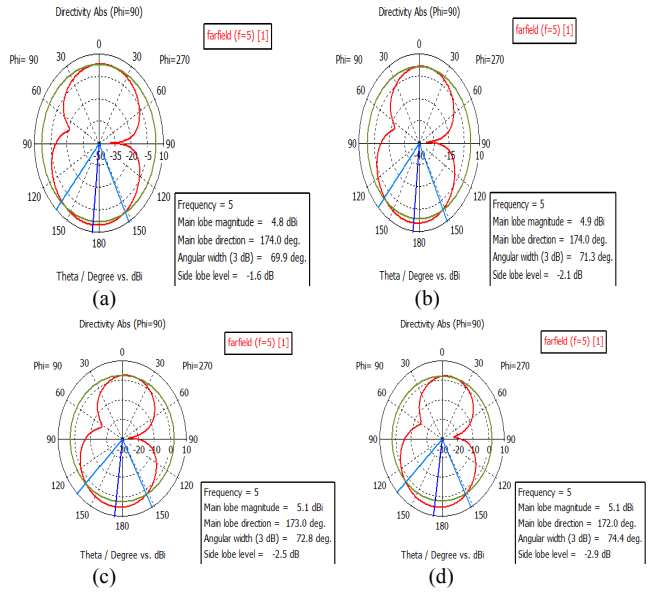


Fig. 7.1: Comparison of elevation pattern at 5 GHz; for 20 degree of bending (a), for 40 degree of bending (b), for 60 degree of bending (c), for 80 degree of bending (d).

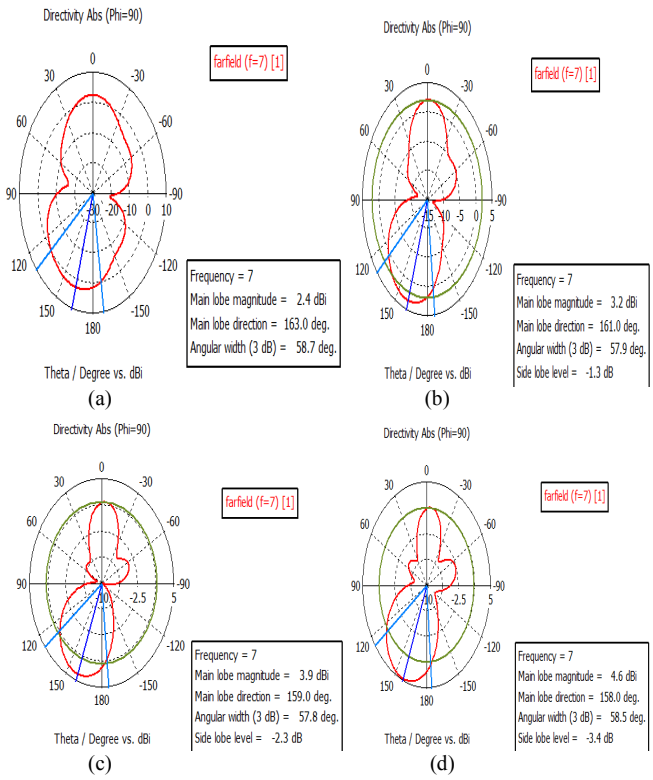


Fig. 7.2: Comparison of elevation pattern at 7 GHz; for 20 degree of bending (a), for 40 degree of bending (b), for 60 degree of bending (c), for 80 degree of bending (d).

In figure 7.1, for different angles of bending Omni-directional radiation pattern is observed at 2 GHz. At 7 GHz, radiation pattern got slightly distorted for different angles of bending. However, the increasing bending angles have very little impact on the radiation pattern of the antenna.

VI. Conclusion

In this paper a design for a compact UWB textile antenna has been proposed. The characteristics of the designed antenna having operating frequency from 2.6 GHz to 14 GHz were analyzed. The bending of the textile antenna at different angles was also focused. At different level of frequencies and bending angles the return loss, radiation characteristics and gain of the proposed textile antenna had very small fluctuations but maintained good efficiency. This performance of the proposed antenna indicates that it is suitable for UWB applications. Increasing the bandwidth as well as decreasing the size could be focus towards the further step of this design.

References

[1] M. Klemm and G. Troester, "Textile UWB antennas for wireless body area networks," *IEEE Trans. Antennas Propag.*, Vol. 54, no. 11, pp. 3192-3197, Nov. 2006.

[2] Shuvashis Dey, Nandita Saha and Subrata Biswas, "Design and Performance Analysis of UWB Circular Disc Monopole Textile Antenna and Bending Consequences," *IEEE European Conference for Antennas and Propagation (EuCAP 2011),* Rome, Italy, April 2011.

[3] B. Sanz-Izquierdo, F. Huang, and J. C. Batchelor, "Convert dual band wearable button antenna," *IET Electron. Lett.*, vol. 42, no. 12, pp. 668–670, Jun. 2006.

[4] T. Karacolak, A. Z. Hood, and E. Topsakal, "Design of a Dual Band Implantable Antenna and Development of Skin Mimicking Gels for Continuous Glucose Monitoring*," IEEE Transactions on Microwave Theory and Techniques*, vol. 56, no. 4, pp. 1001-1008, April 2008.

[5] C. Hertleer, A. Tronquo, H. Rogier, and L. Van Langenhove, "A Textile Antenna for Off-Body Communication Integrated Into Protective Clothing for Firefighters" *IEEE Antennas Wireless Propagation. Letter,* vol. 57, pp. 919-925, 2009.

[6] C. Hertleer, A. Tronquo, H. Rogier, and L. Van Langenhove, "An aperture-coupled patch antenna for integration into wearable textile systems*," IEEE Antennas Wireless Propagation Letter*, vol. 6, pp. 392–395, 2007.

[7] G. A. Conway and William G. Scanlon, "Antennas for over-body-surface communication at 2.45GHz," *IEEE Transactions on Antennas and Propagation*, vol. 57, no. 4, pp. 844-855, April 2009.

[8] S. Zhu and R. J. Langley, "Dual-Band Wearable Textile Antennas on An EBG Substrate," *IEEE Transactions on Antennas and Propagation,* Vol. 57, No.4, Apr 2009.

[9] Garg. R., Bhartia P., Bahl. I., and Ittipiboon. A., *Microstip Antenna Design Handbook*. London: Artech House, 2001.

A Miniaturized Spiral Antenna for Energy Harvesting of Implantable Medical Devices

Qingyun Ma, Laxmi Ray, and Mohammad Rafiqul Haider

Department of Electrical and Computer Engineering, The University of Alabama at Birmingham, Birmingham, AL 35294, USA

E-mail: {maq, laxmiray, mrhaider}@uab.edu

***Abstract*— In this paper, a wireless power transfer system for implantable medical devices using a miniaturized 3 mm diameter single arm circular spiral antenna is presented. The transmission characteristics, S_{21} between the transmitting and receiving antennas with air and tissue as a medium between antennas are analyzed and measured. The receiver antenna receives almost 2 mW average power with power transfer efficiency of 53%. The power transfer efficiency could still reach to 4% with 274 µW average received power, with placement of 1.5 cm tissue and 1 cm air in between two antennas. With one PN diode rectifier, the total DC power from the receiver side is 489 nW.**

***Index terms*— Wireless power transfer, implantable, spiral antenna, biomedical tissue, power transfer efficiency.**

I. Introduction

The implantable medical devices market is expected to exceed $50 billion in 2015 in the US alone due to its immense applications in cardiovascular, neurological, orthopaedic and ophthalmological disorders. For continuous operation, the implanted units could be powered up either from an external source or implantable batteries. The tethered cables for power and data transmission have potential risks of skin infections and irritation. The implantable batteries pose potential battery fluid leakage and biohazard. Wireless power transfer to the implanted medical units has become increasingly important because of the problems seen with less feasible battery power sources. The wireless power transfer shows the great promise by eliminating the risk from the skin irritation and biohazard from battery fluid leakage.

Two most common forms of wireless power transmission are the inductive power link with magnetically coupled resonating coils and the antenna based power transmission through electromagnetic (EM) radiation. Inductive power link consists of two loosely coupled resonating coils separated by a certain distance. The coils in the inductive link must be in perfect alignment, and the distance should be less than one half of the coil diameter. The position of the receiving and transmitting coil could vary due to the body movement of the patients. Furthermore, to satisfy FCC standards the link frequency needs to be in between 5 to 10 MHz which results the inductor diameter over several centimetres [1].

The antenna based power transmission through electromagnetic field has high tolerance on the mismatch of the positions of the transmitting and the receiving antennas. The optimal EM frequency for biomedical tissue of 1.5 cm thickness with high link efficiency could be below 3 GHz [2]. Because of the higher frequency, the antenna diameter can be decreased into millimetre level. To accommodate the antenna for medical applications the antenna must be low profile and small size in order to make the system compact.

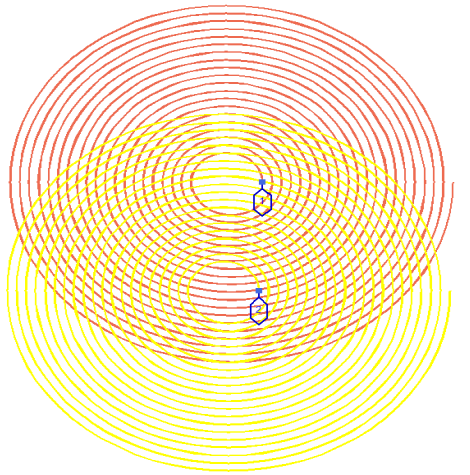
Spiral antenna has widespread usage because of its low-profile planar structure, radiation of circular polarization and flexibility of changing resonant and radiation characteristics for desired application by changing circular radius *(r)*, number of turns *(N)*, spacing between turns *(s)* and width of the spiral arm *(w)*. The spiral antenna was investigated previously [3] and feeding of the antenna can be done either from the center or side end of the spiral antenna. The concept of side-fed spiral antenna [4, 5] and center-fed spiral antenna [6, 8] are well known. Several different possible configurations of spiral antenna have been presented lately like flexible double arm side fed circular spiral antenna with microstrip infinite balun for wearable applications [7]. In addition, a single arm circular spiral antenna has advantage on adaptability of changing polarization and beam characteristics using changes in the center and side feed [8].

In this paper we propose a 3 mm diameter compact single arm center-fed circular spiral antenna operating at 524.8 MHz. The communication between transmitting and receiving antenna is studied to analyse the power transfer efficiency. The general characteristics of antenna like return loss, transmission coefficient, and its transient response analysis, power transfer efficiency with air and tissue as a medium between the spiral antennas are shown.

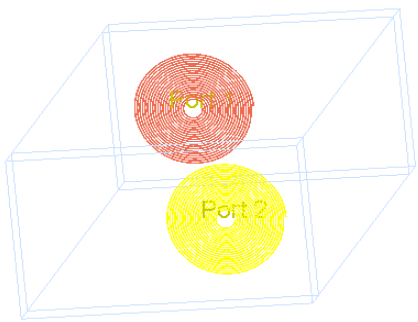
II. Geometry of Spiral Antenna and System of Wireless Power Transfer

A center-fed 3 mm diameter circular spiral antenna operating at 524.8 MHz was designed and simulated. The substrate chosen for realizing the antenna has dielectric constant ε_r=4.6, loss tangent tanδ=0.01 and thickness of 1mm with 0.01 mm thick aluminium metal layer. The Agilent Advanced Design System (ADS) software was used to simulate and analyze the performance and functionality of the proposed spiral antenna. The configuration of the antenna is shown in Fig. 1. The radius*(r),* the spacing *(s),* the number of turns*(N)* and the width of the spiral arm(*w*) are important design parameters of this single arm spiral antenna. For tuning the desired resonant frequency a radius of 1.5 mm with 20 turns and the arm width of 0.015 mm and 0.05 mm spacing between each turn are

978-1-4799-4166-7/14/$31.00 ©2014 IEEE



(a)



(b)

Fig . 1. (a) Geometry of center-fed Transmitting (port1) and Receiving antennas (port2) (b) 3D simulation setup of wireless power transfer.

maintained. The optimized design parameters of the proposed center-fed spiral antenna are summarized in Table I.

III. Simulation Results

The simulated return loss (S_{11}) plot of the transmitting spiral antenna is shown in Fig. 2. The proposed antenna satisfies -10 dB return loss characteristics from 520 MHz to 529 MHz for 524.8 MHz application. The transmitting antenna when excited at the center gives low return loss of about 30.931 dB at the operating frequency. The simulated gain and directivity of the transmitting spiral antenna at 524.8 MHz is shown in Fig.3. The average gain of the transmitting spiral antenna is about -39dBi and the directivity is 2 dBi at 524.8 GHz operating frequency. The simulated far-field radiation pattern for circular polarization of the transmitting spiral antenna at 524.8 MHz for θ = 0° and θ = 90° are depicted in Figs. 4 (a) and (b), respectively. For the study of wireless power transfer between transmitting and receiving antenna two identical spiral antennas are used. Fig. 5 shows transmission characteristics S_{21} between the transmitting and receiving antennas.

TABLE I: OPTIMIZED PARAMETER OF THE SPIRAL ANTENNA

Parameters	Values
Radius	r = 1.5 mm
Number of turns	N=20
Spacing between turns	s = 0.05 mm
Width of spiral arm	w = 0.015 mm

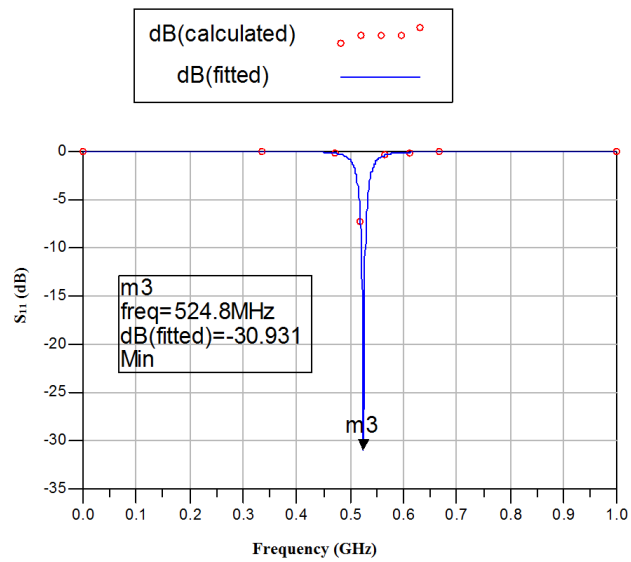


Fig. 2. Return Loss plot of transmitting spiral antenna.

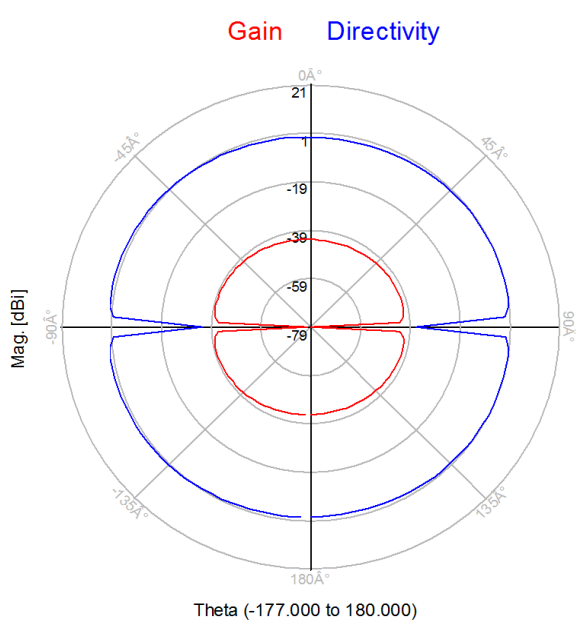
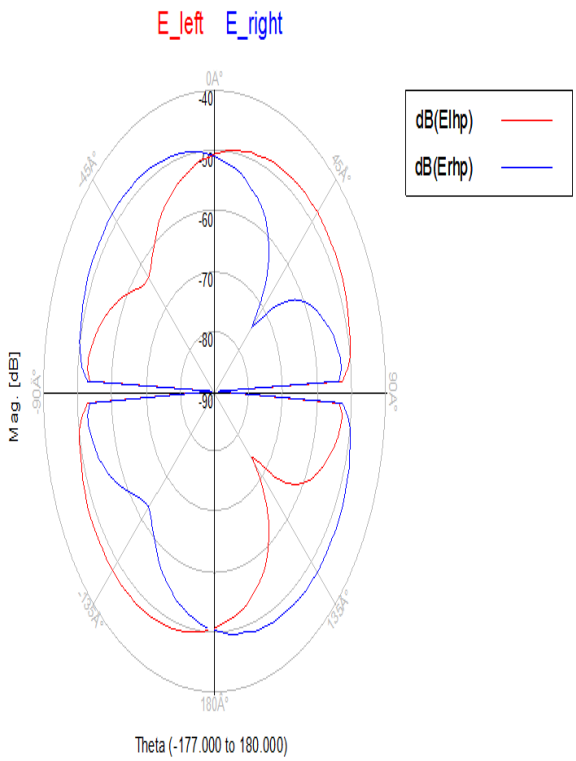
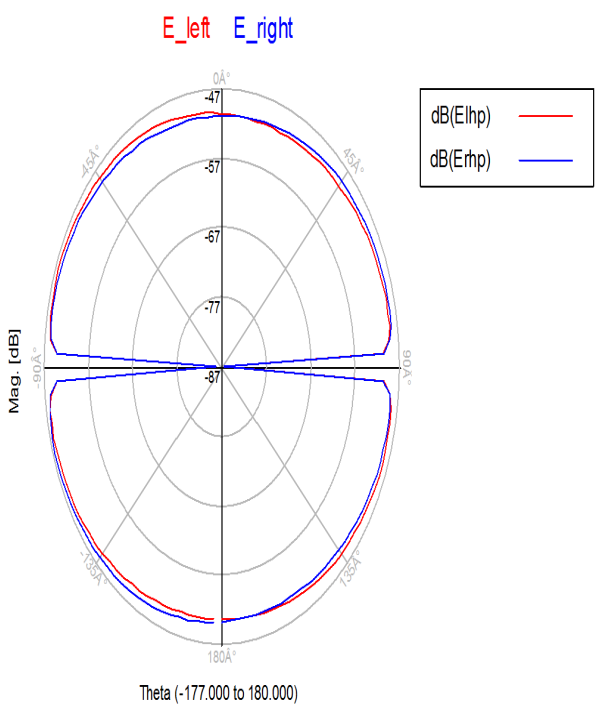


Fig. 3. Gain and directivity of the spiral antenna at 524.8 MHz.



(a)



(b)

Fig. 4. Circular polarization of the spiral antenna at (a) θ = 0° (b) θ = 90°.

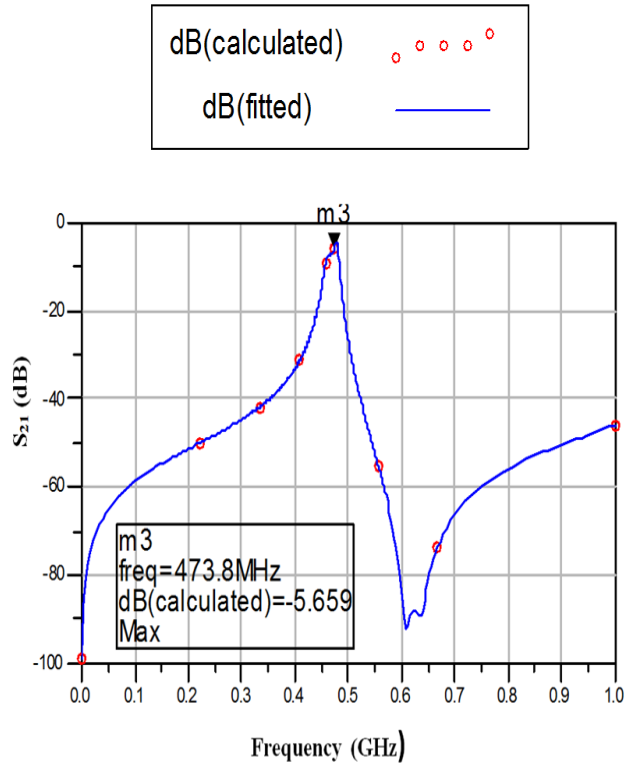


Fig. 5.Transmission characteristics S_{21} between spiral antennas

Due to the coupling of the mutual impedance of two antennas, the frequency is shifted to 473.8 MHz with maximum transmission characteristics of -5.659 dB. This strong coupling shows the influence of nearby placed receiving spiral antenna on radiation characteristics of the antenna.

IV. TRANSIENT SIMULATION RESULTS

The wireless power transfer system structure is converted to a symbol and the performance of this system is analysed in the ADS circuit simulator by using the circuit schematic as shown in Fig. 6. In the transient simulation, a sinusoid source is directly connected to the transmitting antenna through a matching network. The result from this transient simulation is shown in Fig. 7. The received voltage signal is about 1.8 V_{p-p} with 2 V_{p-p} input signal. The transmitting antenna and the receiving antenna are on the two different substrates with 1.5 cm separation distance. With this configuration, the receiver side could receive almost 2 mW average power with power transfer efficiency of 53%. The power transfer efficiency reduces to 4% with 274 µW average received power, when we put 1.5 cm tissue and 1 cm air in between two antennas. The tissue layer acts as a lossy medium which absorbs most of the transmitted power and reduces power transfer efficiency. With one PN diode rectifier, the total DC power from the receiver side is 489 nW.

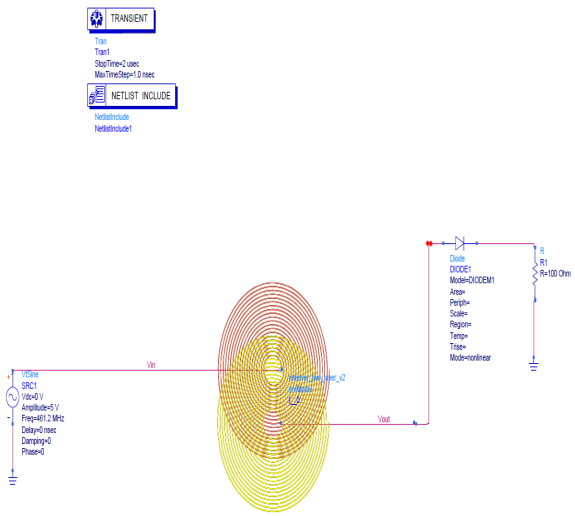


Fig. 6. Circuit simulation setup of the wireless power transfer link system.

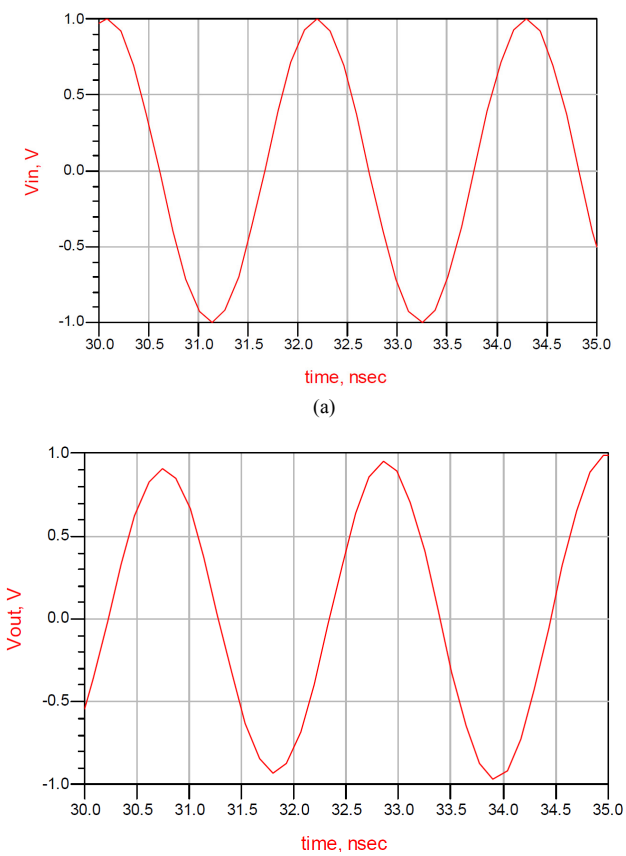


Fig. 7. (a)The input voltage signal (v_{in}) from the transmitting antenna (b) The received voltage signal (v_{out}) from the receiving antenna.

V. Conclusion

This paper presents a wireless power transfer link for medical applications. The proposed single arm center-fed spiral antenna covers 524.8 MHz frequency band. By properly selecting the design parameters of the spiral antenna tunable resonant and radiation characteristics for desired band can be achieved. The transmission characteristic between the transmitting and receiving spiral antennas gives maximum coupling. The transient simulation results shows that the power transfer efficiency is directly proportional to the distance between the center points of two antennas rather than the alignment condition of the two spiral antennas. This is well explained by simulation results with medium as air and tissue between transmitting and receiving antennas. The reduced power transfer efficiency when tissue is used as a medium shows that tissue acts as a lossy medium. From the above discussed result we confirmed that the antenna based wireless power transmission shows the greater promise on tolerance of the position mismatch and size than the inductive power link.

References

[1] M. Kiani, Uei-Ming Jow; M. Ghovanloo, "Design and optimization of a 3-coil inductive link for efficient wireless power transmission," *IEEE Transactions on Biomedical Circuits and Systems*, vol.5, no. 6, pp. 579-591, Dec. 2011.

[2] S. D. O'Driscol,"A mm-sized implantable power receiver with adaptive matching," *2010 IEEE Sensors*, pp. 83-88, Nov. 1-4, 2010.

[3] R. Bawer;J.J Wolfe, "The spiral Antenna," IRE International Convention Record, Pt. T, pp. 84-95, May, 1960.

[4] Eubanks,T.W.;Ming-Yi li;Kai Chang, " Modified equiangular spiral antennas side-fed by multi-stage wideband impedance transforming transitions," *Antennas and Propagation Society International Symposium ,2009.APSURSI '09. IEEE,* pp.1-4, 1-5 June 2009.

[5] Eubanks, T.W.; Kai Chang, "A Compact Parallel-Plane Perpendicular-Current Feed for a Modified Equiangular Spiral Antenna," *IEEE Trans. Antennas and Propag.*, vol.58, no.7, pp.2193-2202, July 2010.

[6] Karlsson,M.;Shaofang Gong, " An integrated spiral antenna system for UWB," *Radar Conference,2005. EURAD 2005.European*, pp.283-286, 6-7 October 2005.

[7] Hoseon Lee;Tentzeris,M.M;Geiger,J.,"Flexible Spiral Antenna with Microstrip Tapered Infinite Balun for Wearable Application," *Antennas and Propagation Society International Symposium ,2012.APSURSI '12. IEEE,* pp.1-2, 8-14 July 2012.

[8] C.W.Jung; Cetiner, B.A.; De Flaviis,F., " A single-arm Circular Spiral Antenna with Inner/Outer Feed Circuitry for Changing Polarization and Beam Characteristics," *Antennas and Propagation Society International Symposium,,2003APSURSI '03. IEEE,* vol.4, no., pp.474-477, 22-27 June 2003.

Design of an X-Band Microstrip Array Antenna For Circular Polarization

Muhammad Asad Rahman, Quazi Delwar Hossain, Md. Azad Hossain
Chittagong University of Engineering and Technology, Chittagong-4349, Bangladesh
asad31@cuet.ac.bd, quazi@cuet.ac.bd, azad@cuet.ac.bd

Eisuke Nishiyama, Ichihiko Toyoda
Saga University,
1 Honjo-machi, Saga-shi, Saga 840-8502, Japan
nisiyama@ceng.ec.saga-u.ac.jp, toyoda@cc.saga-u.ac.jp

Abstract – **In this paper, the design of a circularly polarized array antenna with 9.56 dBi of gain at 10 GHz frequency is presented. The proposed antenna consists of a 2×2 array antenna with feed network where all microstrip patches are linearly polarized. "Both-Sided MIC Technology" is used to design the proposed array antenna. The simulated return loss of the proposed array antenna is less than -35dB. The 3 dB axial ratio bandwidth of 2.21% is obtained. The unequal feed line is used in order to realize 90° phase difference between the linearly polarized patches. Therefore, the circular polarization is realized by the combination of linearly polarized patches and unequal feed line. The structure and the basic behavior along with the simulation results of the proposed linearly polarized array antenna are demonstrated in this paper.**

Keywords-- circular polarization; patch antenna; array antenna; axial ratio; Both-Sided MIC technology

I. INTRODUCTION

Microstrip array antennas are used in a large number of microwave communication applications due to the attractive features of low profile, light weight, easy fabrication, and conformability to mounting hosts [1-3]. Linear polarization is very simple to generate but circular polarization (CP) is widely used in many microwave application as it exhibits multiple benefits over linear polarization. As circular polarization avoids polarization losses due to misalignment, there is no need to keep the transmitting and receiving antenna in the same alignment. This property is useful for RFID systems [4]. It has the ability to decrease interference between direct and reflected signal due to multipath propagation [5] and this is of huge benefit for navigation systems, especially satellite based. Circular polarization is also immune to Faraday rotation. This is the main reason for using CP in almost all Earth–satellite communication systems [6].

Two orthogonal field components with equal amplitudes and 90° phase difference are needed to radiate circular polarized waves [7]. Many CP antenna design techniques have been developed. The method of producing single-feed CP operation using a square microstrip antenna by truncating a pair of patch corners [8]or a single-fed equilateral-triangular microstrip antenna with a truncated tip [9] have been reported. Circularly polarized antenna with a cross slot of unequal arm lengths using single feed has been demonstrated in [10], [11]. The excited patch surface current path is lengthened due to slot perturbation that splits the dominant resonant mode of the microstrip patch into two orthogonal near degenerated modes. The two near degenerate orthogonal resonant modes can have equal amplitudes and a 90° phase differenceby selecting the proper slot length and the patch at a suitable position; as a result CP radiation can be achieved. Applying same concept, Y-shaped slot in antenna can also be used for CP radiation [12]. By inserting slits or spur lines to the boundary of a microstrip patch, CP radiation can be achieved [13], [14]. A circularly polarized equilateral-triangular-ring microstrip antenna design has been reported with a truncated tip and a Y-shaped conducting strip [15]. Tuning stub or a bent tuning stub can be an effective way to excite circular polarization due to the frequency tuning capability of the tuning stub. This design concept has been successfully applied to a circular patch [16], [17], a square-ring patch [18], and a triangular patch [19].

Use of an external power divider such as the quadrature hybrid, the ring hybrid, the Wilkinson power divider and the T-junction [20] to excite single square or circular patch can be another way to radiate circularly polarized waves. In this configuration the dual-orthogonal feeds excite two orthogonal modes with equal amplitude but in-phase quadrature.

A circularly polarized array can be configured using circularly polarized microstrip patches with appropriate feeding [20]. Arrays of linearly polarized patches can be used for circular polarized radiation [21], [22]. Array of two linearly polarized antennas will radiate circularly if the patches are placed orthogonally with one of the patches being fed 90° out of phase. A method has been presented in [22] for generating circularly polarized radiation from an array composed of linearly polarized elements having unique angular and phase arrangements.

In this paper a circularly polarized array antenna is proposed using Both-Sided MIC technology [23] where circular polarization is achieved having a 2×2 array composed of single feed linearly polarized patches. The patches are arranged in such a way that the resultant array radiation is circularly polarized. The feed network is designed using both

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

mirostrip lines and slot line instead of using only micostrip lines that reduces the problem of step by step impedance matching and provides a simple circuit configuration. The proposed antenna provides a gain of 9.56 dBi at design frequency.

II. ARRAY ANTENNA DEVELOPMENTS

A. Configuration of Array

Fig. 1 shows the complete layout of proposed array antenna. Its configuration consists of radiating square patches with a special feed network on one side of the dielectric substrate which has a ground plane on the reverse side.

The feed network shown in Fig. 2 is composed of microstrip lines on the obverse side and a slot line on the reverse side of the substrate. Two types of branch circuits are used in this feed network and they are the microstrip-slot branch circuit and the slot-microstrip branch circuit. The microstrip-slot branch circuit is connected in parallel and slot line impedance has to be double than microstrip line for proper impedance matching. In this case two output signals at the equal distance point from the branch point on the slot line are of same amplitude and in phase [24], [25].

Moreover, there are two slot-microstrip branches at the two opposite end of slot line. In this case slot-microstrip branch circuit is series coupled and the condition for impedance matching is impedance of microstrip line has to be half than that of slot line. Due to the series coupling, two output signals at the equal distance point from the branch point on the microstrip line are of same amplitude and 180°out of phase. The impedance of the slot line in this design is to be chosen as 100 ohm and the microstrip line impedance is 50 ohm.

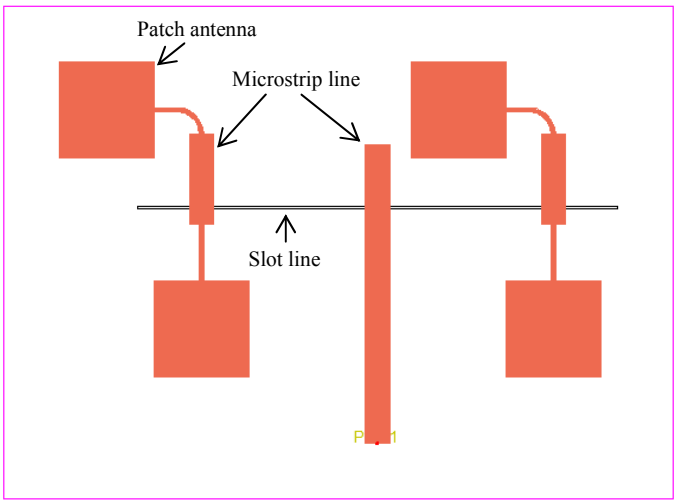


Fig. 1. Layout of proposed array antenna

The design frequency of the each patch antenna used in array is 10 GHz. Each side of square patches is 9.46 mm. Upper two patch elements are placed at 90° angle with respect to other two lower patch elements. Positional distance of upper patch elements from slot line is determined in such a way that the feed line length is quarter wave length (λ\4) greater than that of lower patch elements. This strategy creates 90° phase shift between upper patches and lower patches to excite circular polarized waves. For proper matching between the each patch and slot-microstrip branch circuit, 136 ohm quarter-wavelength transmission lines are used as impedance transformers.

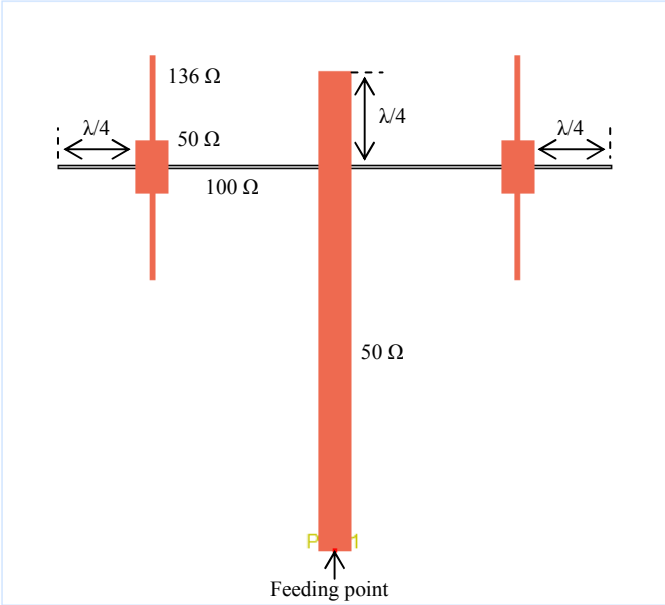


Fig. 2. Basic configuration of feed line network

Teflon glass fiber substrate with a thickness of 0.8 mm and dielectric constant of 2.15 is used to design the proposed array. Copper with a thickness of 0.018 mm is used as radiating plane and ground plane conductor.

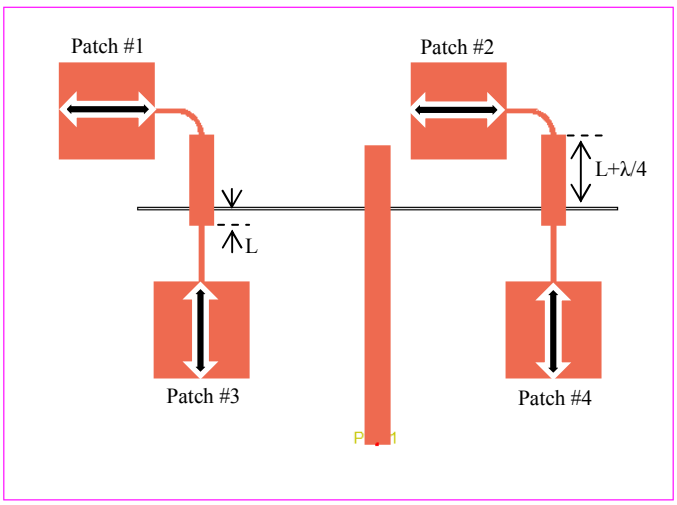


Fig. 3. Basic behavior of the array antenna

B. Basic Behavior of the Array

To generate circular polarized waves, two orthogonal signals of equal amplitude with 90° phase shift between them are needed. The basic behavior of the proposed circularly polarized array antenna can be explained using Fig. 3. When the patch #1 and the patch #2 are 0° linearly polarized, the

patch#3 and the patch #4 are 90° linearly polarized. Combining the signals of patch #1, #3 and patch #2, #4, the modes are orthogonal.

Moreover the length of feed lines is adjusted such a way that there is a 90º phase shift between patch #1 and #3, and patch #2 and #4. Therefore, in the proposed array structure, two orthogonal signals are formed with 90° phase shift and as a result circular polarized signal is excited.

III. SIMULATION RESULTS OF ARRAY ANTENNA

Advanced Design System (ADS) of Agilent Corporation has been used to simulate and optimize the proposed circularly polarized array antenna.

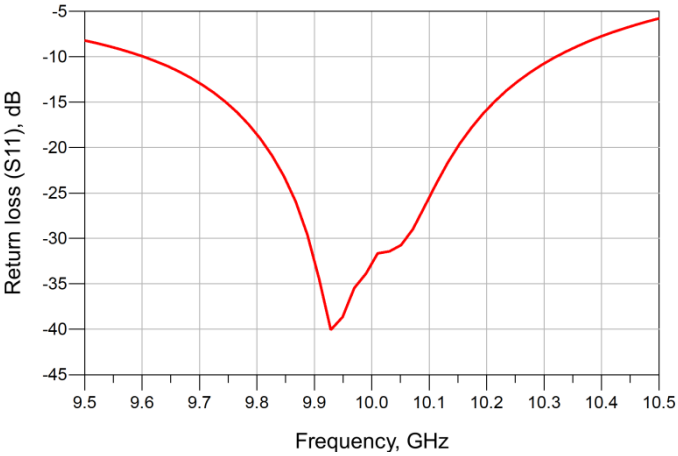


Fig. 4.Input return loss of the array

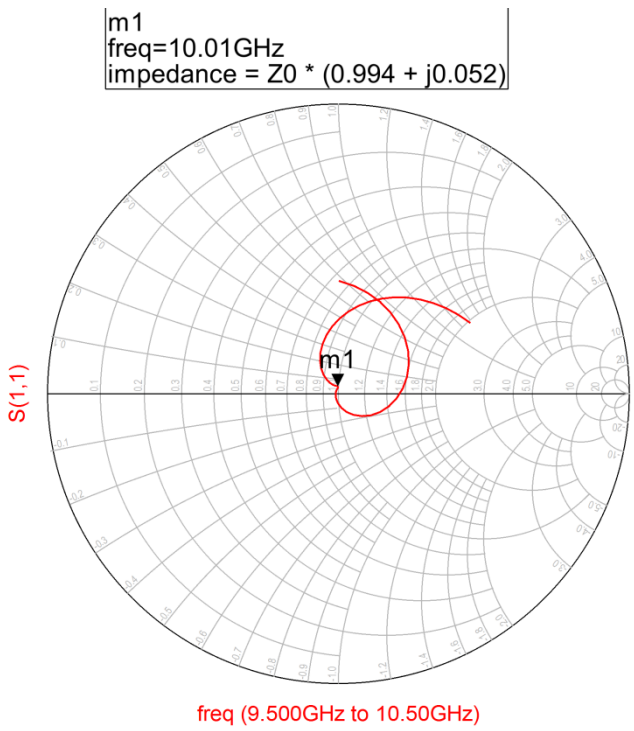


Fig. 5. Input impedance for the array antenna

The simulated return loss of the array antenna is shown in Fig. 4. The result indicates that the return loss is better than -10dB from 9.6 to 10.32 GHz (7.23%). It is the lowest at near of the design frequency of 10 GHz and excellent impedance matching is obtained.

The Input impedance for the array is shown in Fig. 5. The figure shows a dip in the impedance locus near 10 GHz, which indicates that two resonant modes are excited at very close frequencies. This suggests that the fundamental mode in the present design is split into two near-degenerate resonant modes [26]. It can be found that, when upper patch elements are ahead by quarter wave length from lower patches, these two resonant modes can be excited with equal amplitudes and a 90° phase difference, resulting in circular polarized radiation.

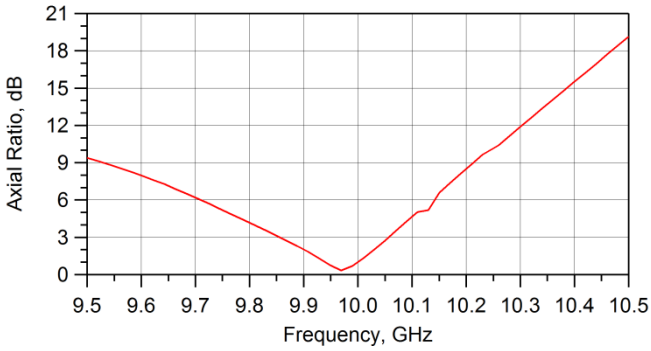


Fig. 6. Axial ratio of the antenna array

Simulated result for the axial ratio is shown in Fig. 6. This result provides a good axial ratio smaller than 3 dB at the design frequency and it indicates that the proper circular polarization is achieved. The 3-dB axial ratio bandwidth of the designed array antenna is 0.22GHz or about 2.21% respect to the center frequency of 9.95 GHz and the minimum value of axial ratio is 0.344 dB at 9.97 GHz.

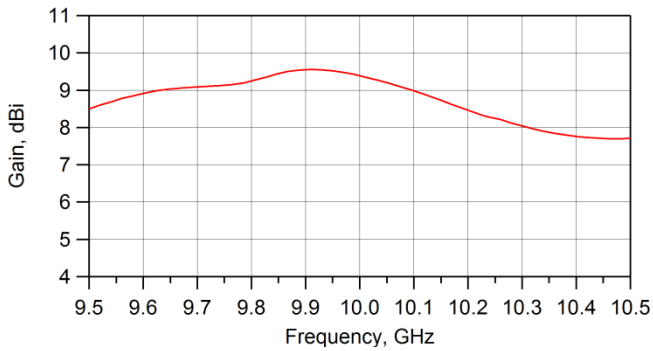


Fig. 7. Array antenna gain

Fig. 7 shows the gain of the array. It is found that from Fig. 7, the peak gain is 9.56 dBi at 9.92 GHz which is much higher than the average 2×2 circularly polarized array antenna where each patch element is linearly polarized. The average gain from 9.6 to 10.3 GHz is above 8 dBi. Fig. 8 shows the radiation pattern of the proposed array antenna. From fig. 8, it is confirmed that the antenna radiation is right hand circularly polarized (RHCP). The cross-polar level is better than -30 dB that indicates good radiation performance of the antenna.

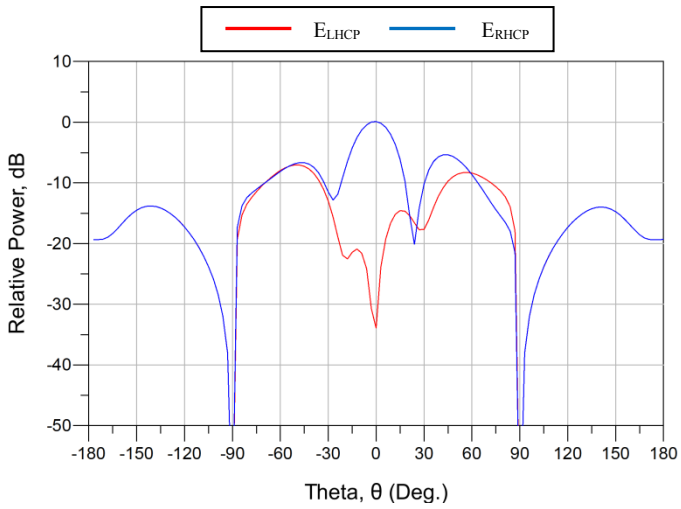


Fig. 8. Radiation pattern of the proposed array antenna at 9.97 GHz

IV. CONCLUSION

In this paper, a circularly polarized array antenna has been presented and the characteristics of the proposed array antenna are discussed.The simulation results show good axial ratio and 9.56 dBi gain at design frequency. All results indicate that the proposed array antenna radiation is circularly polarized with good return loss performance and proper impedance matching. The most remarkable feature of this array antenna is that CP radiation is excited using only linear polarized patches with special feed line network that consists of microstrip lines and slot line. This feed line structure minimizes the impedance matching problem that is faced in conventional microstrip feeding circuits. Moreover, each patch element is excited at a single point. From this study, it is also clear that, the newly designed array antenna has simple structure and besides, the both-sided MIC technology is effectively used at the design process. The experimental investigation of the proposed array antenna will be conducted very soon.

REFERENCES

[1] F. Yang and Y. Rahmat-sami, "Patch Antennas with Switchable Slots (PASS) in wireless communications: Concepts, Design and Application," IEEE Trans. on Antennas and Propagation, Vol. 47, No. 2, pp. 13-29, 2005.

[2] M.K. Fries, R.Vahldieck, "Uniplaner circularly polarized slot-ring antenna architectures", Radio Sci., vol. 38, no. 2, Nov. 2002.

[3] J. Y. Park, C. Caloz, Y. Qian and T. Itoh, "A Compact Circularly Polarized Subdivided Microstrip Patch Antenna," IEEE Microwave and Wireless Components Letters, Vol. 12, No. 1, pp. 18-19, 2002.

[4] J. Garcia, A. Arriola, F. Casado, X. Chen, J. I. Sancho, and D. Valderas, "Coverage and read range comparison of linearly and circularly polarized radio frequency identification ultra-high frequency tag antennas," IET Microwaves, Antennas and Propagation, vol. 6, no. 9, pp.1070–1078, 2012.

[5] Y. Qin, S. Gao, and A. Sambell, "Broadband high-efficiency circularly polarized active antenna and array for RF front-end application," IEEE Trans. Microw. Theory Tech., vol. 54, no. 7, pp. 2910–2916, July 2006.

[6] K. Davies and E. K. Smith, "Ionospheric effects on satellite land mobile systems," IEEE Antennas and Propagation Magazine, vol. 44, no. 6, pp. 24–31, Dec 2002.

[7] R. Garg, P. Bhartia, I. Bhal, and A. Ittipiboon, Microstrip Antenna Design Handbook, Artech House, London, pp. 493, 2001.

[8] W. S. Chen, C. K. Wu, and K. L. Wong, "Single-feed square-ring microstrip antenna with truncated corners for compact circular polarization operation," Electron. Lett. 34, 1045–1047, May 28, 1998.

[9] C. L. Tang, J. H. Lu, and K. L. Wong, "Circularly polarized equilateral-triangular microstrip antenna with truncated tip," Electron. Lett. 34, 1277–1278, June 25, 1998.

[10] H. Iwasaki, "A circularly polarized small-size microstrip antenna with a cross slot," IEEE Trans. on Antennas and Propagation, vol.44, pp.1399–1401, Oct. 1996.

[11] J. H. Lu, C. L. Tang, and K. L. Wong, "Single-feed slotted equilateral-triangular microstrip antenna for circular polarization," IEEE Trans. on Antennas and Propagation, vol.47, pp.1174–1178, July1999.

[12] K. P. Yang, K. L. Wong, and J. H. Lu, "Compact circularly-polarized equilateral-triangular microstrip antenna with Y-shaped slot," Microwave Opt. Technol. Lett. 20,pp.31–34, Jan. 5, 1999.

[13] J. H. Lu, C. L. Tang, and K. L. Wong, "Circular polarization design of a single-feed equilateral-triangular microstrip antenna," Electron. Lett.34, pp. 319–321, Feb. 19, 1998.

[14] H. M. Chen and K. L. Wong, "On circular polarization design of annular-ring microstrip antennas," IEEE Trans. on Antennas and Propagation, vol. 47, pp. 1289–1292, Aug. 1999.

[15] C. L. Tang and K. L. Wong, "A modified equilateral-triangular-ring microstrip antenna for circular polarization," Microwave Opt. Technol. Lett. 23, 123–126, Oct. 20, 1999.

[16] K. L. Wong and Y. F. Lin, "Circularly polarized microstrip antenna with a tuning stub," Electron. Lett. 34, 831–832, April 30, 1998.

[17] K. L. Wong and M. H. Chen, "Slot-coupled small circularly polarized microstrip antenna with modified cross-slot and bent tuning-stub," Electron. Lett. 34, 1542–1543, Aug. 6, 1998.

[18] W. S. Chen, C. K. Wu, and K. L. Wong, "Square-ring microstrip antenna with a cross strip for compact circular polarization operation," IEEE Trans. Antennas Propagat. 47, 1566–1568, Oct. 1999.

[19] J. H. Lu and K. L. Wong, "Single-feed circularly-polarized equilateral-triangular microstrip antenna with a tuning stub," IEEE Trans. Antennas Propagat. 48, 1869–1872, Dec. 2000.

[20] J. Huang, "A Ka-band circularly polarized high-gain microstrip array antenna," IEEE Trans. on Antennas and Propagation, vol. 43, no.1, pp. 113-116, Jan. 1995.

[21] R. Garg, P. Bhartia, I. Bhal, and A. Ittipiboon, Microstrip Antenna Design Handbook, Artech House, London, pp. 494-500, 2001.

[22] J. Huang, "A Technique for an Array to Generate CircularPolarization with Linearly Polarized Elements," IEEE Trans. on Antennas and Propagation, vol. AP-34, no.9, pp. 1113-1124, Sep. 1986.

[23] M. Aikawa, and H. Ogawa, "Double-Sided MIC□s and Their Applications," IEEE Trans. Microwave Theory Tech., vol.37, no.2, pp.406-413, Feb. 1989.

[24] K. Kodama, E. Nishiyama and M. Aikawa, "Slot array antenna using Both-sided MIC technology," IEEE International Symposium on Antennas and Propagation , Vol. 3, pp. 2715-2718, 2004.

[25] K. Egashira, E. Nishiyama and M. Aikawa, "Microstrip array antenna using Both-sided MIC feed circuits," Asia-Pacific Microwave Conference, APMC, 2002.

[26] Kin-Lu Wong, Compact and Broadband Microstrip Antennas, John Wiley & Sons, Inc., New York, pp. 165, 2002.

Design and Development of a Patch Antenna for ISM Band Passive RFID Tag

Avijit Hira*, Nayan Sarker and Md. Osman Goni
Department of Electronics & Communication Engineering, Khulna University of Engineering & Technology
Khulna 9203, Bangladesh.
*avi_ece910@yahoo.com

Abstract— **This work deals with the design and simulation of a passive radio frequency identification (RFID) tag antenna operating in ISM band. The antenna has an impedance bandwidth of 60 MHz (2419MHz - 2479 MHz) resonant at 2449.5 MHz. The microstrip antenna has a planar geometry and consists of a ground plane, a substrate, a modified rectangular patch as a radiator. The proposed antenna provides excellent SWR of 1.0435 and a gain of 3.351 dB. The possible adjustment over the entire ISM band is demonstrated by adjusting a single branch. The basic theory and design are analyzed, and the simulation results are obtained using CST Microwave Studio commercial software is employed to optimize the antenna's properties.**

Index Terms—**Passive RFID Antenna, Sensitivity, CST microwave studio, ISM Band**.

I. Introduction

Radio frequency identification (RFID) is an automatic non-contact bidirectional communication identification technology taking advantage of electromagnetic induction and radiation. A generalized RFID system comprises a scanner unit, called a reader, and a set of remote transponders, attached to the objects to be identified denoted as tags [1]. It also includes software (middleware) and a server processor (generally a PC) to communicate with remote tags. RFID is one of the promising technologies employed as an alternative for overcoming some issues associated with current technology used for identifications, such as bar code and optical systems. It employs radio frequency waves to transfer data between a reader and a movable item to identify, categorize and track. Now a days RFID is dominating in sensing technology because of its improved read range, no line of sight stipulation, low cost, capability of handling multiple objects within a single scan and so on.

The RFID tag element consists of an antenna integrated with a microchip. A single antenna both transmits and receives radio waves to or from readers. According to the presences of battery in the transponder and their radiation nature RFID tag may be classified in three major categories, i) Active tag ii) Battery assisted passive or semi active tag and iii) Passive tag [2]. Active tag bears a battery and emits radio wave continuously. But semi active tag has a small on board battery and emits radio wave after receiving reader's command only. Passive tag contains no on board power supply. The reader antenna transmits electromagnetic wave to tag. This signal is received by the tag via antenna. This received energy excites the passive tag and provides the necessary power to the tag that allows the microchip to operate. Reader sends carrier wave to activate the tag as well as commands. When the tag receives the command signal, the microchip enables the tag to scatter a modulated signal back to the reader in response to reader command [3]. This process is referred to as passive backscatter. The reader then demodulates and decodes the signal and delivers the data stored in the tag to host PC.

The two most sensitive factors of RFID tag performance and measurement are i) Tag sensitivity and ii) Read range [4]. Tag sensitivity is defined as the minimum signal strength either field or power at the tag location needed to power up the tag [5]. Tag field sensitivity E_{tag} (V/m) and power sensitivity P_{tag} (W) are directly related. The power collected by a perfectly matched antenna load (chip) can be expressed either via incident power and tag antenna gain or via incident power density and effective tag antenna area [6]:

$$P_{tag}\, G = S\, A_e \tag{1}$$

Here, G is the gain of the tag antenna; S is the power density of an electromagnetic wave with the electric field E incident on the RFID tag, given in free space separation d. by:

$$S = \frac{E^2}{120\pi} = \frac{P_t\, G_t}{4\pi d^2} \tag{2}$$

Here, P_t is the transmitted power and G_t is the gain of the reader antenna. A_e is the effective area of the antenna inversely related to the wavelength.

$$A_e = \frac{\lambda^2}{4\pi}\, G \tag{3}$$

Hence, Tag field sensitivity;

$$E_{tag} = \frac{4\pi}{\lambda}\sqrt{30 P_{tag}} \tag{4}$$

Tag power sensitivity can also be readily expressed as:

$$P_{tag} = G\, p\tau P_{th} \tag{5}$$

Where, p is the polarization efficiency, τ is the impedance matching or the power transmission coefficient between the tag antenna and the microchip and P_{th} is the chip power threshold sensitivity. Equation (5) describes the necessity of a high gain antenna to collect more power from incident wave and to involve maximum reader tag separation to make sure that the sufficient amount of it gets absorbed in the RFID chip and activates it.

The most important tag performance characteristic is read range, the maximum distance at which RFID reader can detect the backscattered signal from the tag. Though the reader receives modulated backscatter from tag due to its high sensitivity in comparison with tag, the read range is defined by the tag response threshold. Read range is also sensitive to the tag orientation, the material on which tag is attached, and to the propagation environment [7]. The read range can be reckoned using Friis free-space formula:

$$r = \frac{\lambda}{4\pi}\sqrt{\frac{P_t G_t G \tau}{P_{th}}} \tag{6}$$

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

In tag antenna designing convenient impedance matching between tag and microchip must be confirmed. So, the antenna should be conjugate matched with the highly capacitive chip impedance i.e. Za = Zc, (where Za = (Ra+jXa) Antenna Impedance and Zc = (Rc+ jXc) Microchip Impedance), which allow the tag to receive the maximum power from the reader's antenna [8]. To achieve highest radiation efficiency at the resonant frequency of the tag is at where Xa = - Xc and Ra close to Rc, which ensures the conjugate matching. Again the power transmission coefficient which indicates the impedance mismatch between the tag antenna and the microchip can be expressed by,

$$\tau = \frac{4R_aR_c}{|Z_a + Z_c|} \quad (7)$$

Based on the above discussion and mathematical expression, it is very convenient to declare that all other factors (P_t, G_t, λ, temperature, environment and so on) remaining constant, the performance and accuracy of RFID depends on tag antenna specifications only such as gain, perfect impedance matching with micro-chip.

In this paper, modified planer microstrip antenna is designed for RFID application in ISM band. RFID transponders operating in ISM have the advantage of being smaller than UHF tags, but generally have shorter ranges [7]. Again passive RFID tags are extensively used due to its several features such as low cost, absence of on board power supply, while its lifetime is theoretically infinite and many more. The antenna is designed by targeting the mid frequency of ISM band, and then possible adjustment over the entire ISM band is investigated.

II. Tag Antenna Design

In this letter, a microstrip patch antenna is designed. The proposed antenna is constructed using several rectangular arms of equal width of 5 mm and oriented irregularly. The proposed antenna shape is shown in Fig. 1. FR4 substrate of thickness 1.6 mm is used on which the total structure is mounted. Dielectric constant (□) is chosen 4.3 while loss tangent is 0.025. The opposite side of the substrate contains the copper plate that acts as a ground plane. This enhanced the directivity of the antenna to a desired level. Copper metal is also used as the radiating element. The thicknesses of both the radiator and ground plane are 0.1 mm.

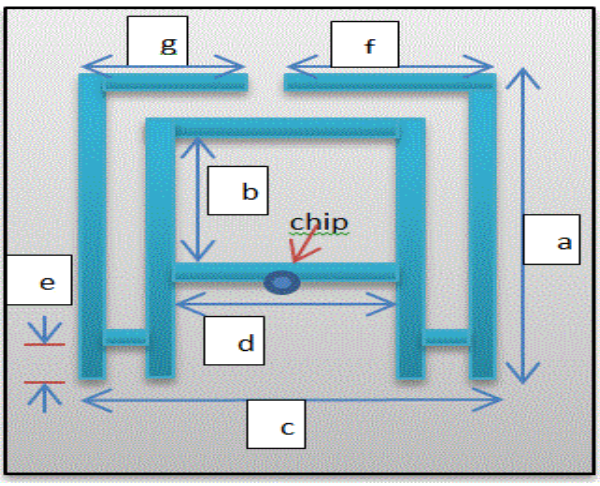


Fig. 1 Structure of proposed tag antenna.

Hence the approximate arm lengths are given in Table 1. The microchip position is allotted in arm "d". All the arms have an equal width of 5 mm and separated from the neighbor arm by 5 mm also.

TALBE I
GEOMETRICAL PARAMETERS OF ANTENNA STRUCTURE

Arm	Length (mm)
A	60
B	25
C	60
D	25
E	5
F	28
G	22

The arms of the proposed antenna along a meandered path, a configuration is produced by distributed capacitive and inductive reactance's which globally affect the antenna input impedance. In this proposed structure, horizontal lines mainly control the radiation resistance; adjacent vertical lines give the storage of electric energy and loss; the overall conductor length affects the inductance.

III. Simulated Results

The proposed antenna is designed and simulated in the CST Microwave Studio in order to extract its electromagnetic parameters. It provides excellent characteristics which demonstrate that proposed antenna can be used in the ISM band RFID applications. The proposed is resonant at 2449.5 MHz with accuracy of 0.02% from the mid of ISM band. Simulated return loss is -33.44 dB with standing wave ratio (SWR) 1.0435. The frequency range having return loss less than -10 dB is found from 2419 MHz to 2479 MHz. So impedance bandwidth is 60 MHz, which is 2.45% of resonant frequency. These are depicted in Fig. 2 and in Fig. 3.

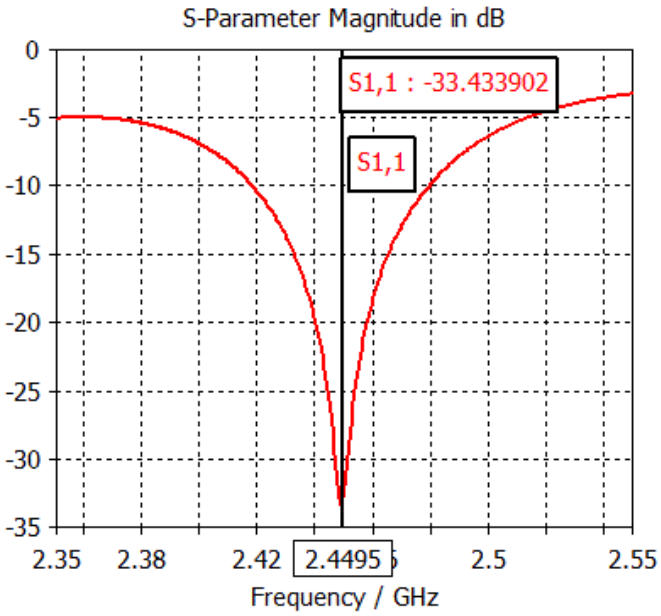


Fig. 2 Simulated return loss of proposed antenna.

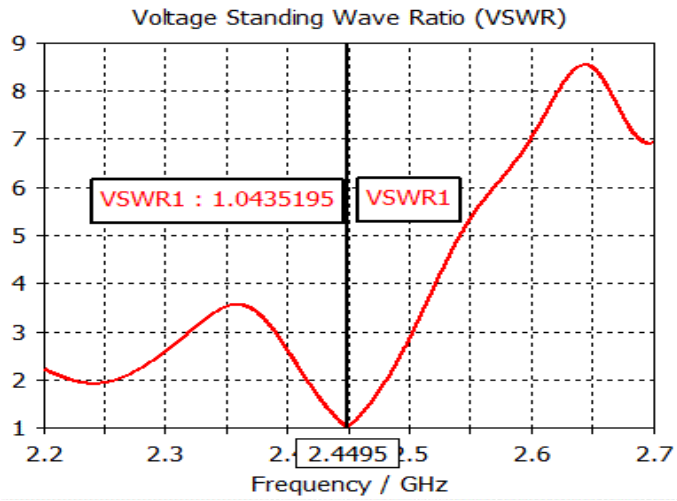


Fig. 3 Computed VSWR plots of the proposed antenna.

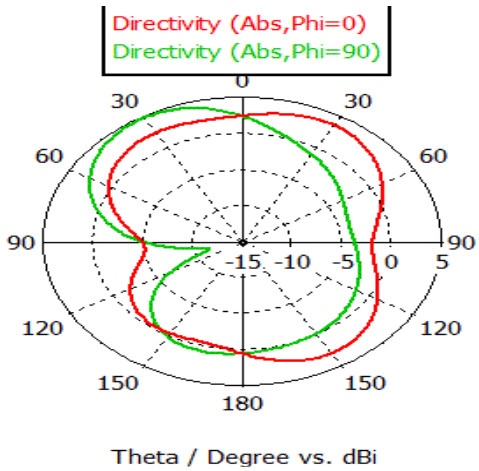


Fig. 4 Computed far field directivity of the proposed antenna.

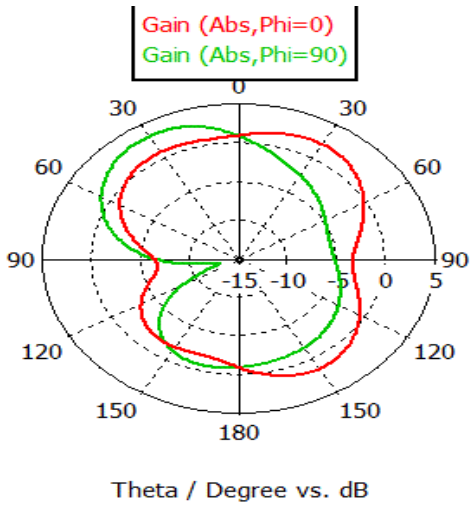


Fig. 5 Computed far field gain of the proposed antenna.

From Fig. 4 and Fig. 5, the pictorial representation of simulated far field radiation pattern, it is very convenient to state that the radiation pattern of the proposed antenna is fairly unidirectional; which is very important criteria of designing tag antenna. So, it could provide necessary power to microchip irrespective to the reader angular position. Proposed antenna bears a directivity of 4.84 dBi and gain 3.351 dB at 2450 MHz.

Passive RFID tag has no on board power supply. The tag antenna should have power collecting capability from the RF signal emitted by reader. Antenna supplies this collected power to the microchip to activate it in order to answer in response to the readers command via backscatter. Amount of collected power would be transferred to microchip depends on the conjugate matching between tag and microchip. The antenna input impedance at the allotted chip location is extracted as 10.43 + j148.5 Ω using inductive coupled feeding instead of direct connection with the excitation source. So it can be perfectly matched with microchip having complex impedance of 10 – j150 Ω. Hence, the power transmission coefficient (τ) would be 0.994.

$P_t G_t$ is termed as effective isotropic radiated power, EIRP. An RFID standard of EIRP is 4 W [6]. The available commercial chip power threshold sensitivity may vary between -20 dBm to -10 dBm. The maximum read range of this proposed antenna would be 3.60 meter while the tag sensitivity is limited to -12 dBm which is close to the worst sensitivity as compared to others.

IV. Adjustment over entire ISM band

Mostly used electromagnetic application in ISM band covers the frequency range of 2400 MHz to 2500 MHz. RFID application in this band is adopted in different frequencies in different country. Designed antenna covers most of the band (60 MHz) with return loss below -10 dB and VSWR less than 2. Possible adjustment of resonant frequency entire the band is investigated with a simple adjustment of length of arm "f" hence satisfactory results are obtained. Comparative simulated results of length adjustment are summarized in Table 2.

TALBE II
Effect of Length Adjustment of Arm " f "

Length of arm " f " (mm)	Resonant frequency (MHz)	Return loss (dB)	VSWR
33	2415	-29.32	1.07
30	2434	-35.63	1.03
28.5	2444	-35.06	1.03
25	2460	-30.79	1.06
22	2477	-32.94	1.04
18	2491	-18.33	1.27

The simulated results obtained with adjustment of length of arm "f" in the CST microwave studio. Corresponding return losses are depicted in Fig. 6 for further clarification. Satisfactory results are obtained for each adjustment. Now it may conclude that by selecting appropriate length of arm "f" any resonant frequency can be achieved within the ISM band. So proposed antenna can be used in other applications such as WiMAX, Wi-Fi, Bluetooth and WLAN at 2.4 GHz to 2.5 GHz band.

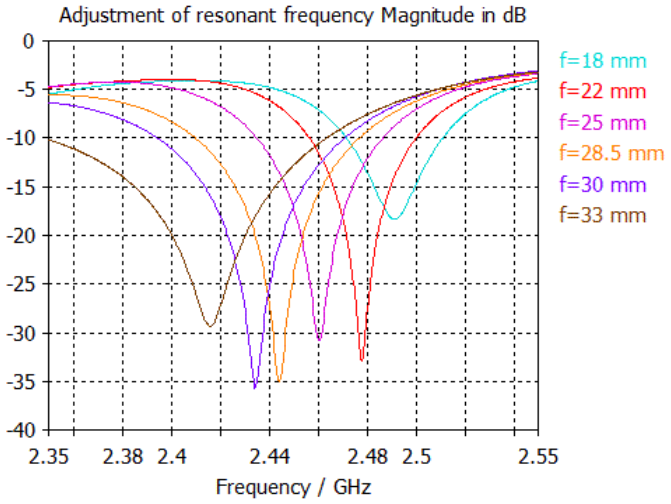


Fig. 6 Adjustment of resonant frequency.

V. Conclusions

Modified rectangular shaped microstrip antenna is designed and its electromagnetic features are investigated. The results are found satisfactory and can be frequently used in RFID application at ISM band. The simulation results show that a high gain, a good impedance matching with the microchip and a fairly unidirectional far field radiation pattern would be obtained. The adjustment of resonant frequency over the entire ISM band is investigated simply by varying length of a single arm; hence satisfactory results are obtained in almost every trial. So, proposed antenna can be used in others commercial communication application. Antenna is designed with 60 mm X 60 mm FR4 substrate with metallic radiators of much less surface area than substrate results in small radiation loss.

References

[1] Ning Zhang, Xiuping Li, "RFID Tag Antenna Design and its EM Simulation Based Measurement Method."

[2] H. W. Son, and G. Y. Choi, "Orthogonally proximity coupled patch antenna for a passive RFID tag on metallic surfaces," Microwave and Optical Technology Letters, vol.49, no. 3, 2007.

[3] K. H. Anabi1, J. S. Mandeep, M. Islam, J. J. Tiangl, "A quarter-wave Y-shaped patch antenna with two unequal arms for wideband Ultra High Frequency."

[4] R. Islam, A. A. Yussuf, A. Z. Alam, A. F. Ismail, J. Chebil, S. Khan, "Design of a Passive RFID Tag Antenna at 2.45 GHz for Mounting on Various Platforms" IEEE International RF and Microwave Conference 2011.

[5] Curty JP, Declercq M, Dehollain C, Joehl N (2007). Design and Optimization of Passive UHF RFID Systems, Springer.

[6] Pavel V. Nikitin, Senior Member, IEEE, and K. V. S. Rao, Senior Member, IEEE, "Antennas and Propagation in UHF RFID Systems".

[7] K. V. Seshagiri Rao, Senior Member, IEEE, Pavel V. Nikitin, Member, IEEE, and Sander F. Lam; "Antenna Design for UHF RFID Tags: A Review and a Practical Application" IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, VOL. 53, NO. 12, DECEMBER 2005.

[8] K. Finkenzeller, "RFID Handbook", ISBN 0-470-84402-7, John Wiley & Sons 2003.

[9] Smart Border Alliance "RFID Feasibility Study Final Report", Application note.

[10] Balanis, C. A., "Antenna Theory: Analysis and Design" 3rd ed. John Wiley & Sons, 2005.

[11] C. Occhiuzzi, S. Cippitelli, G. Marrocco, "Modeling, Design and Experimentation of Wearable RFID Sensor Tag" IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, VOL.58, NO. 8, AUGUST 2010.

[12] S. Nambi, S. Nyalamadugu, S. M. Wentworth, and B. A. Chin, "Radio frequency identification sensors," in Proc. 7th World Multiconf. Systemics,Cybernetics and Informatics (SCI 2003), 2003, pp. 386–390.

Soil Moisture Measurement Using Smart Antennas

Muhsiul Hassan,[1,*] and Nemai C. Karmakar[2]

[1,2] Department of Electrical & Computer Systems Engineering, Monash University
Bldg 72, Clayton Campus, Melbourne, VIC 3800, Australia
[*]muhsiul.hassan@monash.edu

Abstract—**This review article presents an overview of the soil moisture measuring technique using microwave radiometry. Remote sensing of soil moisture is a promising field of research if accuracy can be achieved. The optimal frequency to estimate at this state using microwave radiometry is at L-band (~1.4GHz). In this article analysis of basics of soil microwave sensing, parameters responsible for radiation variation, soil dielectric properties and retrieval algorithm is discussed mainly. In the end discussion on synchronization of soil microwave radiation receiving by polarimetric radiometer, conclusions and future perspective has been presented.**

Index Terms—**Soil moisture retrieval, Patch antenna, Polarimetric radiometer.**

I. Introduction

Soil moisture monitoring is important because it reflects land degradation and drought of a land. It is a critical component of climate, hydrology, ecology and agriculture systems too. It is also related to the precipitation variability in an area. The soil can be considered as consisted of water, air and solid particles and with different percentage of those components result different types of soil. An eloquent model formulation would account for the observed effects of various soil components on the complex dielectric behavior of the soil water-air system as a function of frequency [1]. Dry soil and moist soil has different dielectric constants [2], so it can be concluded that it will radiate different patterns for different soils. Using radiation from microwaves makes detection possible in almost all-weather conditions, which is a definite advantage that is not possible using visible or infrared remote sensing.

Earth receives electromagnetic radiation from sun and to maintain thermal equilibrium radiates and absorbs it. Radiometry is the field of study that deals with radiated electromagnetic power. There are two types of microwave remote sensors: Active sensors and Passive sensors. The active type receives the back-scattering which is reflected from the transmitted microwave which is incident on the ground surface. The microwave radiometer is one of the passive microwave sensors [3]. Radiometers measures and detect the radiated energy by radiating body in the microwave frequency range (300 MHz-300 GHz). Passive sensors are sensitive to the emissivity and brightness temperature of the soil. Both are function of dielectric properties, surface roughness, physical temperature, frequency, angle of observation and polarization etc. [4].

The frequency most suitable for this remote sensing is L-band (1 GHz-2 GHz). To design antenna for this frequency range, microstrip patch antennas are natural choice because of their light weight, conformability and low cost [5]. The radiation properties of microstrip structures have been known since the mid 1950's [5]. Rectangular and circular micro strip resonant patches have been used extensively in a variety of array configurations.

Soil moisture measuring technique starting with basics of microwave remote sensing will be discussed next. Microwave radiometer system and future aspects of this research will be discussed in the end.

II. Soil Moisture Measuring Technique Using Microwave Remote Sensing

A. Basics of Remote Microwave Sensing

The power collected by an antenna which is surrounded by a distribution of incident power $B(\theta,\varphi)$ [6] can be calculated using the equation

$$P = \frac{1}{2} A_r \int_{f}^{f+\Delta f} \iint_{4\pi} B_f(\theta,\varphi) |F_n(\theta,\varphi)|^2 \, d\Omega df \qquad (1)$$

Where, A_r is the effective area of the antenna, solid angle of the transmitting antenna Ω_t is $\Omega_t = A_t/R^2$, and R is the distance between the antenna and the radiating target. Δf is the bandwidth and F_n is the normalized antenna pattern. If the brightness is dependent on frequency then $B_f(\theta,\varphi)$ is called spectral brightness density. The equation is illustrated with Fig. 1.

The power absorbed by an antenna with normalized radiation pattern $F_n(\theta,\varphi)$ surrounded by a black body at a constant physical temperature T_0 is given by modifying expression (1) and can be expressed as

$$\begin{aligned} P_{bb} &= k \cdot T_0 \cdot \Delta f \cdot \frac{A_r}{\lambda^2} \cdot \iint_{4\pi} |F_n(\theta,\varphi)|^2 \, d\Omega \\ &= k \cdot T_0 \cdot \Delta f \end{aligned} \qquad (2)$$

Where the antenna solid angle Ω has been expressed as a function of its effective area and T_0 is the absolute temperature.

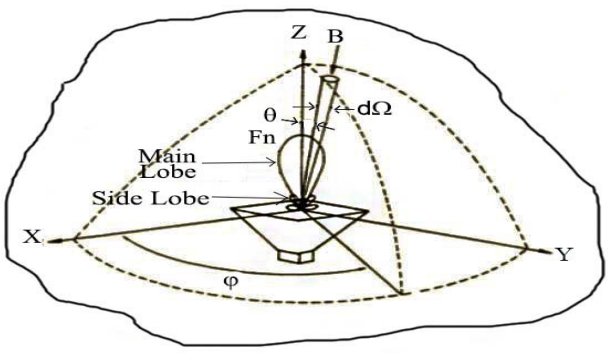


Fig. 1 Geometry of the radiation incident over antenna

Black-bodies are ideal elements that absorb all incident energy and radiate them in omnidirectional way. But in reality, materials do not radiate energy such as black-body and thus called grey-body materials. To express grey-body radiation, two new concepts are introduced, the brightness temperature, $T_B(\theta,\varphi)$ and emissivity, $e(\theta,\varphi)$.

The apparent temperature T_{AP} is an equivalent temperature related to the total brightness incident over the antenna $B_i(\theta,\varphi)$ and the relation of them is

$$B_i(\theta,\varphi)=\frac{2k}{\lambda^2}\cdot T_{AP}(\theta,\varphi)\cdot\Delta f$$

where (3)

$$T_{AP}=T_{UP}+\frac{1}{L_a}(T_B+T_{SC})$$

Where, T_{UP} is the atmospheric upward radiation

T_{SC} is the atmospheric downward radiation scattered reflected by the surface and L_a is the atmospheric radiation.

Brightness temperature can also be calculated by Stokes parameters. The four Stokes parameters describe the polarization and total energy transported by a wave [7], and if the energy acquired at horizontal and vertical polarization are separate issues, then in terms of brightness temperatures expression can be written as-

$$T_B=\begin{pmatrix}T_{Bh}\\T_{Bv}\\U\\V\end{pmatrix}=\frac{\lambda^2}{k\Delta f\eta}\begin{pmatrix}\langle|E_h|^2\rangle\\\langle|E_v|^2\rangle\\2\Re\langle E_vE_h^*\rangle\\2\Im\langle E_vE_h^*\rangle\end{pmatrix} \quad (4)$$

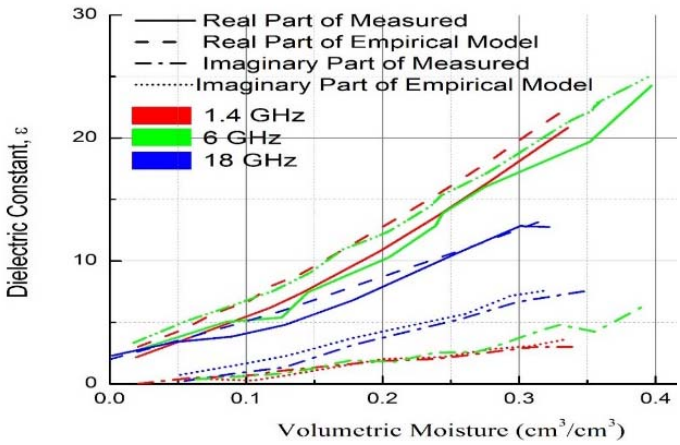


Fig. 2 Comparison of measured ε and that calculated by semi-empirical model for sandy loam in 1.4, 6 and 18 GHz [1]

B. Soil Moisture Measuring

The most important property of soil for measuring moisture using microwave antenna is its dielectric constant. Soil dielectric constant ε, which has no unit and can be expressed as, $\varepsilon=\varepsilon'+j\varepsilon''$ $\varepsilon=\varepsilon'+j\varepsilon''$ $\varepsilon=\varepsilon'+j\varepsilon''$. The real part is responsible for velocity of wave and the imaginary part is for energy losses. The dielectric constant of dry soils is almost independent of temperature [8] and frequency. A series of dielectric constant measurements of five soils with different texture composition at frequencies between 1.4 and 18 GHz was performed [9] and found out that texture has a strong effect on the dielectric behaviour. Models that use a semi-empirical was followed by two widely used models presented in [1, 10]. In [10] a means to estimate a soil's dielectric properties from a knowledge of its texture was proposed. Methods related with theoretical four-component dielectric mixing model to proposed empirical model in [1] were discussed and result is shown in Fig. 2. These relationships can be used for estimating the brightness temperature. Using complex dielectric constant and the volume fraction of each of the soil components as the function the relative gravimetric moisture m in [9] is determined as-

$$\varepsilon^\alpha=1+\frac{\rho_b}{\rho_s}(\varepsilon_s^\alpha-1)+V_a\varepsilon_a^\alpha+V_{fw}\varepsilon_{fw}^\alpha+V_{bw}\varepsilon_{bw}^\alpha-m_v \quad (5)$$

Where, ρ_b , ρ_s s and m_v are bulk density, particle density and volumetric moisture respectively. The relative gravimetric moisture m is determined as, $m=\frac{w_w-w_d}{w_d}$ and the volumetric moisture, $m_v=m\frac{\rho_b}{\rho_w}$. The dielectric constant modelling plays a crucial role because the constant is one of the most also changes and the radiometer output so changes. If the bound water and free water are considered as the same [1] the equation of dielectric constant becomes

$$\epsilon=\left(1+\frac{\rho_b}{\rho_s}(\epsilon_s^\alpha-1)+m_v^\beta\epsilon_{fw}^\alpha-m_v\right)^{1/\alpha} \quad (6)$$

In [4], the variation of Brightness temperature at different angles of incidence with different moisture content for both horizontal vertical polarization was studied and it was observed that the moisture content increases, the value of brightness temperature reduces. The value of brightness temperature increases as the received power increases (see Fig. 3).

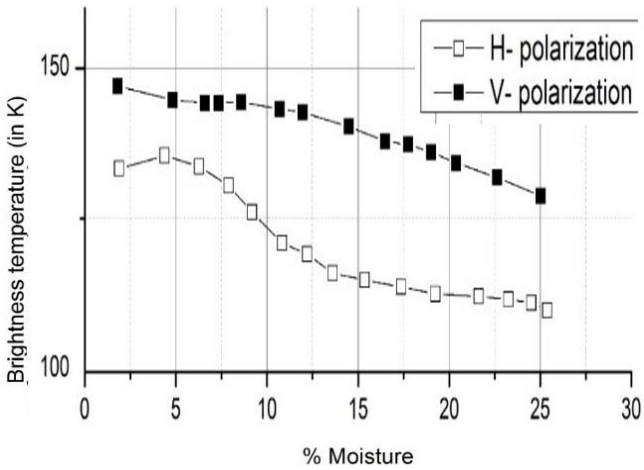


Fig. 3 Variation in brightness temperature with percent of Moisture

Several attempts have been made to explain the effect of surface roughness on the observed brightness temperature. When a radar wave strikes on a smooth surface at a specific incident angle, some energy is scattered in all directions, some of which returns to the sensor as backscatter, and the rest is reflected in the specular direction [11-16]. Physical models [17] are generally driven by surface characteristics derived from measurements of surface height profiles [18]. Using only

the coherent term of the scattered field, [11] proposed a simple model for the reflectivity

$$\Gamma_{sp} = \Gamma_{op} \exp(-4k^2{}_{\lambda}\,\sigma^2{}_s \cos^2(\theta))$$

Where,

$$\Gamma_{oh} = \left| \frac{\cos\theta - \sqrt{\varepsilon_s - \sin^2\theta}}{\cos\theta + \sqrt{\varepsilon_s - \sin^2\theta}} \right|^2, \quad (7)$$

$$\Gamma_{ov} = \left| \frac{\varepsilon_s \cos\theta - \sqrt{\varepsilon_s - \sin^2\theta}}{\varepsilon_s \cos\theta + \sqrt{\varepsilon_s - \sin^2\theta}} \right|^2$$

Where Γ_{sp} is the reflectivity at p-polarisation ($p = v$ or h) of a smooth surface, Where ε_s is the dielectric constant of soils, $k = 2\pi/\lambda$ is the electromagnetic wave number, σ_s is the standard deviation of the height of the surface and θ is the incidence angle.

The soil moisture models are used as forward models, and the geophysical variables are retrieved by minimization of a cost function of the type [19]

$$F = \sum_{\theta} \frac{(T_{Bh} - T_{Bh}^{meas})^2 + (T_{Bv} - T_{Bv}^{meas})^2}{\sigma_{TB}^{\ 2}} + \sum_{n} \frac{(P_n^{ini} - P_n)^2}{\sigma_{Pn}^{\ 2}} \quad (8)$$

Where, the simulated brightness temperature T_{Bp} is computed using a land emission model,T_{Bp}^{meas} is the measured brightness temperature, and P_n is any of the parameters on which T_{Bp} depends. Estimates can be constrained to be close to the initial value by choosing low σ_{Pn} values, or they can be left as a free parameter by selecting $\sigma_{Pn} \ll 1$.

III. Microstrip Patch Antenna and PMR System

Many techniques have been proposed to determine microstrip antenna characteristics. Of them, transmission line model, generalized transmission line model, cavity model, and multiport network model are the popular methods. In details of antenna structure and analysis in not scope of this article, but some basic feeding techniques and polarimetric radiometer (PMR) have been discussed.

1) Microstrip Patch Antenna: Most basic structure of a patch antenna consists of a radiating patch on one side of a dielectric substrate which has a ground plane on the other side as shown. The patch is generally made of conducting material such as copper or gold and can take any possible shape [5]. The radiating patch and the feed lines are usually photo etched on the dielectric substrate [20]. Early microstrip antennas used either a microstrip feed line or a coaxial feed [21-23] offering one degree of freedom by adjusting feed point to adjust input impedance level. But having simplicity it shows some major problems. Increasing substrate thickness can lead to increase in spurious feed radiation, surface wave power and feed inductance [24]. For the feed point is selected on the non-radiating edge of the patch the cross-polarized radiation is high but can be minimized by optimizing the aspect ratio [25]. A number of non-contacting feeds have been developed for microstrip antennas. Of them, proximity feed or electromagnetic coupling uses a two-layer substrate with a microstrip line on the lower substrate, terminating in an open stub below the patch which is printed on the upper substrate [26, 27] thus having two degrees of freedom: the length of the feeding stub and the patch width to line width ratio. This design reduces spurious radiation and coupling in return of complex fabrication process. Aperture coupled microstrip antenna is another type of non-contacting feed method [28, 29]. This configuration uses two parallel substrates separated by a ground plane. A microstrip feed line on the lower substrate is coupled through a small aperture in the ground plane to a microstrip patch on the top substrate. A thin, high dielectric constant substrate for the feed and a thick, low dielectric constant substrate for the antenna decrease spurious radiation and increases bandwidth. In addition, the ground plane eliminates spurious radiation from the feed from interfering with the antenna pattern or polarization purity. This geometry has at least four degrees of freedom: the slot size, its position, the feed substrate parameters, and the feed line width. Also along with different feeding methods, perforating patch, stubs and some innovative features such as dual polarization etc. are used for different requirements.

Polarimetric Radiometer: Polarimetric radiometer receives radiation from soil/ocean, digitize the received data by embedded computer program. The final outcome is the brightness temperature which is calibrated using special set and number of measurements and finally result is verified using stability tests. As discussed in the previous section, radiation fields can be characterized using Stokes parameters. It is convenient to use the modified Stokes parameters [30]. In brightness temperature: the first and second parameter describe the vertically and horizontally polarized brightness temperatures respectively and the third and fourth parameter describe linearly and circularly polarized components respectively. The first successful operation of analogue direct cross-correlation technique for polarimetric radiometry is demonstrated by FPoR (Helsinki University of Technology Fully Polarimetric Radiometer) [31]. The first three Stokes parameters has confirmed the feasibility of these parameters in measuring surface wind vectors in sea areas. The geophysical noise of these parameters was found to be extremely small, showing very homogeneous wave conditions in the test area. The results found in [31] suggest that the relationship between the wind vector and the Stokes parameters may vary locally. Though the centre frequency of [31] was 36.5GHz, these result can be used to form a frequency range of 6-94 GHz. Therefore multiple band operation can be a topic of further research. The available commercial radiometer for L-band is PLMR by Prosensing Inc. [32] which calculates the brightness temperature by a series of calculations. It is an airborne system for remote sensing of environmental parameters such as soil moisture or ocean salinity. This system has two modules which can be divided as receiver module and communication module. This system is capable of measuring 12 measurements in 2 polarization at 6 antenna beam positions. By calibrating initial temperature and setting several gains brightness temperature T_B is measured with accuracy and precision using and its accuracy/stability can be checked. This [32] polarimeter is for only L-band but the principle can be extended for multiband which can be a good future work. If L-band detection is incorporated with some higher frequency bands it is possible to increase the precision and sample number from a certain area under observation. In general the total PMR system for total phased array network for soil moisture measurement can be proposed as Fig. 4 The beamforming network is essential because of the phased array system.

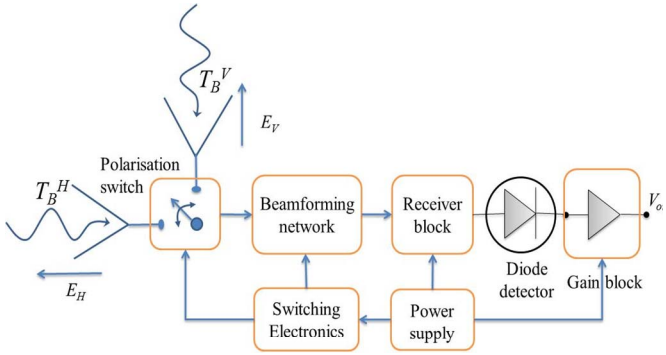


Fig. 4 Block diagram of PMR system

IV. Conclusion and Future Perspective

This article presents a review on the soil moisture measurement using smart antennas. Water, being the most important element of nature, controls the existence and propagation of life in the planet. Soil moisture also plays vital role in agriculture and animal diversity and so water management requires detailed temporal and spatial soil moisture monitoring system. Soil contains several component whose composition determines the dielectric constant of the soil, thus different types of soils have different dielectric constants. The dielectric constant of soil is primarily varied with the soil composition, roughness of soil and temperature of soil. The radiated power from soil also changes with changes in this dielectric constant. The radiation by the soil can be remotely sensed by PMR (Polarimetric Microwave Radiometer). The bands that are used for this PMR are P-, L-, C-, X-, Ku-, Ka- and W-bands. Among them L-band has been proved by research as the optimal band. To provide accurate but high spatial resolution higher frequency band data can be used to downscale the low resolution L-band data. To combine this system with precision and making the system cheap, low profile it is necessary to design multiple band antenna in a compact design. Microstrip antenna becomes the optimum choice for such requirements because of its planar and low profile attributes. Comprehensive studies have been done in this interest and this field has huge possibility for further research. To combine the experimental data regarding soil moisture with brightness temperature calculation and decipher to match the actual will be a big challenge.

References

[1] M. C. Dobson, F. T. Ulaby, M. T. Hallikainen, and M. A. El-Rayes, "Microwave Dielectric Behavior of Wet Soil-Part II: Dielectric Mixing Models," *IEEE Transactions on Geoscience and Remote Sensing,* vol. GE-23, 1985.

[2] T. J. Jackson and T. J. Schmugge, "Passive microwave remote sensing system for soil moisture: some supporting research," *Geoscience and Remote Sensing, IEEE Transactions on,* vol. 27, pp. 225--235, 1989.

[3] S. Murai, *Remote Sensing Note Chapter 3*: Japan Association on Remote Sensing, 1993.

[4] O. P. N. Calla, D. Bohra, R. Vyas, B. S. Purohit, R. Prasher, A. Loomba, *et al.*, "Measurement of soil moisture using microwave radiometer," in *Recent Advances in Microwave Theory and Applications, 2008. MICROWAVE 2008. International Conference on*, 2008, pp. 621-624.

[5] C. A. Balanis, *Antenna Theory: Analysis and Design*: Wiley, 2005.

[6] F. T. Ulaby, R. K. Moore, and A. K. Fung, *Microwave remote sensing: active and passive. Radar remote sensing and surface scattering and emission theory. Vol. II*: Addison-Wesley Publishing Company, Advanced Book Program/World Science Division, 1981.

[7] J. P. Hamaker and J. D. Bregman, "Understanding radio polarimetry. III. Interpreting the IAU/IEEE definitions of the Stokes parameters," *Astronomy and Astrophysics Supplement,* 1996.

[8] G. C. Topp, J. L. Davis, and A. P. Annan, "Electromagnetic determination of soil water content: Measurements in coaxial transmission lines," *Water Resources Research,* vol. 16, pp. 574-582, 1980.

[9] M. T. Hallikainen, F. T. Ulaby, M. C. Dobson, M. A. El-Rayes, and W. Lil-Kun, "Microwave Dielectric Behavior of Wet Soil-Part 1: Empirical Models and Experimental Observations," *Geoscience and Remote Sensing, IEEE Transactions on,* vol. GE-23, pp. 25-34, 1985.

[10] J. R. Wang and T. J. Schmugge, "An Empirical Model for the Complex Dielectric Permittivity of Soils as a Function of Water Content," *IEEE Transactions on Geoscience and Remote Sensing,* vol. GE-18, 1980.

[11] B. J. Choudhury, T. J. Schmugge, R. W. Newton, and A. T. C. Chang, "Effect of surface roughness on the microwave emission from soils," *Journal of Geophysical Research,* vol. 84, pp. 5699--5706, 1978.

[12] J. R. Wang and B. J. Choudhury, "Remote sensing of soil moisture content, over bare field at 1.4 GHz frequency," *Journal of Geophysical Research: Oceans,* vol. 86, pp. 5277-5282, 1981.

[13] J. R. Wang, "Passive microwave sensing of soil moisture content: The effects of soil bulk density and surface roughness," *Remote Sensing of Environment,* vol. 13, pp. 329-344, 9// 1983.

[14] T. Mo and T. J. Schmugge, "A parameterization of the effect of surface roughness on microwave emission," *IEEE Transactions on Geoscience and Remote Sensing,* vol. GE-25, pp. 481--486, 1987.

[15] U. Wegmuller and C. Matzler, "Rough bare soil reflectivity model," *Geoscience and Remote Sensing, IEEE Transactions on,* vol. 37, pp. 1391-1395, 1999.

[16] J. P. Wigneron, L. Laguerre, and Y. H. Kerr, "A simple parameterization of the L-band microwave emission from rough agricultural soils," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 39, pp. 1697--1707, 2001.

[17] A. K. Fung, *Microwave Scattering and Emission Models and Their Applications*: Artech House, Incorporated, 1994.

[18] A. A. Chukhlantsev, *Microwave radiometry of vegetation canopies*, 2006.

[19] M. Parde, J. P. Wigneron, P. Waldteufel, Y. H. Kerr, A. Chanzy, S. S. S??bj??rg, *et al.*, "N-parameter retrievals from L-band microwave observations acquired over a variety of crop fields," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 42, pp. 1168--1178, 2004.

[20] D. M. Pozar, "Microstrip antennas," *Proceedings of the IEEE,* vol. 80, pp. 79-91, 1992.

[21] K. R. Carver and J. Mink, "Microstrip antenna technology," *Antennas and Propagation, IEEE Transactions on,* vol. 29, pp. 2-24, 1981.

[22] I. J. Bahl, *Microstrip antennas*. Dedham, Mass.: Dedham, Mass. : Artech House, 1980.

[23] J. R. James, P. S. Hall, and C. Wood, *Microstrip Antenna Theory and Design*: Institution of Engineering & Technology, 1981.

[24] D. H. Schaubert, D. M. Pozar, and A. Adrian, "Effect of microstrip antenna substrate thickness and permittivity: comparison of theories with experiment," *Antennas and Propagation, IEEE Transactions on,* vol. 37, pp. 677-682, 1989.

[25] M. L. Oberhart, Y. T. Lo, and R. Q. H. Lee, "New simple feed network for an array module of four microstrip elements," *Electronics Letters,* vol. 23, pp. 436-437, 1987.

[26] H. Oltman and D. Huebner, "Electromagnetically coupled microstrip dipoles," *Antennas and Propagation, IEEE Transactions on,* vol. 29, pp. 151-157, 1981.

[27] P. B. Katehi and N. Alexopoulos, "On the modeling of electromagnetically coupled microstrip antennas--The printed strip dipole," *Antennas and Propagation, IEEE Transactions on,* vol. 32, pp. 1179-1186, 1984.

[28] D. M. Pozar, "Microstrip antenna aperture-coupled to a microstripline," *Electronics Letters,* vol. 21, pp. 49-50, 1985.

[29] G. Gronau and I. Wolff, "Aperture-coupling of a rectangular microstrip resonator," *Electronics Letters,* vol. 22, pp. 554-556, 1986.

[30] L. Tsang, J. A. Kong, and R. T. Shin, *Theory of microwave remote sensing*: Wiley, 1985.

[31] J. Lahtinen, J. Pihlflyckt, I. Mononen, S. J. Tauriainen, M. Kemppinen, and M. T. Hallikainen, "Fully polarimetric microwave radiometer for remote sensing," *Geoscience and Remote Sensing, IEEE Transactions on,* vol. 41, pp. 1869-1878, 2003.

[32] www.prosensing.com/, "Polarimetric L-band Multibeam Radiometer (PLMR)," ed.

Quantum Mechanical Equivalent of Electromagnetic Band gap and Perfect Tunneling in Multilayer Semiconductor Hetero-structures

Md. Jubayer Shawon,[1] Arif Shahriar,[2**] M.R.C. Mahdy,[3*] Ayed Al Sayem,[2] Golam Dastegir Al-Quaderi,[4] and Md. Saifur Rahman[2]

[1]Department of Electrical Engineering, University of Malaya, 50603 Kuala Lumpur, Malaysia
[2]Dept. of EEE, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh
[3]Department of Electrical and Computer Engineering, National University of Singapore, 4 Engineering Drive 3, Singapore
[4]Dept. of Physics, University of Dhaka, Dhaka, Bangladesh
*elemrc@nus.edu.sg
**arif.eee0@gmail.com

Abstract—**In this article, we have applied Maxwell-Schrodinger analogy to introduce the idea of electromagnetic metamaterial based band gap in semiconductor regime. We have found a new type of band gap in semiconductor hetero-structures, which is analogous to the spatial average single negative (SASN) band gap in multi-layer photonic crystal. Interestingly, width of such band-gaps is controllable according to designers' will. Perfect tunneling of the propagating stationary states and strong spatial delocalization of bounded electronic states have also been observed, which is electronic analogue of the enhancement of evanescent waves in Pendry's lens. Finally, a multilayer structure has been proposed, which supports band-gap at some energy bands and perfect tunneling at target energy value. As the on-off ratio (step transition from low transmission to high transmission) is very high even if the applied energy is very low, this kind of multi-layer structures can be very useful as future low power electronic device and logic transistor.**

Index Terms—**electronic metamaterial, semiconductor hetero-structures, perfect tunneling, probability wave, Kane model.**

I. Introduction

Metamaterials are artificially engineered materials which has unusual physical properties. In photonic crystal domain, several metamaterial structures have already been investigated and they have found several interesting applications in cloaking, Optical fibers, perfect lenses, antennas [1-5]. Recently, SNG (Single Negative) type metamaterials has been investigated thoroughly. SNG type metamaterials are such type of material which has only one material parameter negative [3, 5]. It can be epsilon negative ($\varepsilon < 0$) but mu positive ($\mu > 0$) or mu negative ($\mu < 0$) while epsilon is positive ($\varepsilon > 0$). The Epsilon negative SNG materials are usually called ENG (Epsilon Negative) and mu negative materials are called MNG (Mu Negative). Using ENG-MNG multilayer periodic structure [6, 7], we can obtain a photonic band gap which is called Spatially Averaged Single Negative (SASN) band gap [8]. Unlike Bragg band gap [9], SASN band gap originates from the interaction of evanescent waves in the MNG and ENG media [8]. Due to the localization of the field at the interfaces, zero effective phase gap [7] is robust against disorder, independent of scaling and incident angle and has same PBG for both TE and TM polarization [6-11]. SASN band gap [8] is the sub-wavelength version of the zero effective phase band gap [7]. Because SNG PC thickness does not need to be quarter wave thick at operating wavelength, it enables us to reduce the device size whereas for devices with other PBGs this is not possible. Besides, by applying the concept of SASN band gap, it is possible to control the band gap and center frequency according to designers will [8, 11]. Such degrees of freedom are not found in conventional Bragg bang gap based optical devices. At a particular frequency, if the spatially averaged permeability ($\sum_i \mu_i d_i$) or permittivity ($\sum_i \varepsilon_i d_i$) of a layered photonic structure becomes zero, there emerges a band edge of SASN band gap [8, 11]. If both of the conditions are met at two different frequencies then there will be a band gap in between those two frequencies [8, 11]. Using that analogy, we have demonstrated quantum mechanical equivalent of SASN band-gap in semiconductor array structure. Interestingly, width of such equivalent band-gaps is also controllable according to designers' will. In addition, it is possible to get perfect tunneling in semiconductor hetero-structure [12], which has been shown recently. But in multi-layer hetero-structures (array of hetero-structure), such perfect tunneling is counterintuitive. In this article, we have shown such a perfect tunneling is also possible from the analogy of zero width electromagnetic band-gap at target perfect tunneling energy. Finally, the multi-layer structure, which has been proposed in this article, can support band gap (but this band-gap is not equivalent SASN band-gap) and perfect tunneling at different energies. This structure can play a key role in future semiconductor device technology obtaining low power logic transistors with high on-off ratio.

II. Schrödinger - Maxwell analogy and equivalent SASN band-gap in semiconductor

It is somewhat interesting to notice that Schrödinger and Maxwell equations are analogous [12] for some situations. The Schrödinger and Maxwell equations are as follows-

$$\frac{\delta}{\delta z}\begin{bmatrix} E_x \\ H_y \end{bmatrix} = \begin{bmatrix} 0 & i\omega\mu \\ i\omega\varepsilon & 0 \end{bmatrix}\begin{bmatrix} E_x \\ H_y \end{bmatrix}, \tag{1}$$

and

$$\frac{\delta}{\delta z}\begin{bmatrix} \psi \\ \frac{-i\hbar}{m}\frac{\delta\psi}{\delta z} \end{bmatrix} = \begin{bmatrix} 0 & i\frac{m}{\hbar} \\ 2i\frac{(E-V)}{\hbar} & 0 \end{bmatrix}\begin{bmatrix} \psi \\ \frac{-i\hbar}{m}\frac{\delta\psi}{\delta z} \end{bmatrix}. \tag{2}$$

Clearly, we can see an analogy can be drawn from the equations. The analogies are as follows-

$$E_x \to \psi \qquad \mu \to m \qquad \varepsilon \to 2(E-V) \qquad \omega \to 1/\hbar\,. \tag{3}$$

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

From this analogy, we can assume that, may be, we can find an analogous condition of electromagnetic SASN band gap [8,11] in quantum mechanical domain. The conditions for the quantum mechanical equivalent of the SASN band gap would be -

$$\sum_i m_i d_i = 0 \quad \text{and} \quad \sum_i 2(E - V_i)\cdot d_i = 0, \tag{4}$$

at two different target energy values. We have also derived the condition for band edge analytically. For an infinite periodic quantum mechanical structure ($N \to \infty$) consisting of 3 layer unit cells (Fig. 1), and as per [15] the dispersion relation of the quantum modes based on the Bloch condition would be

$$\begin{aligned}\cos(Kd) &= \cos(k_1 d_1)\cos(k_2 d_2)\cos(k_3 d_3)\\ &-\frac{1}{2}\left(\frac{k_2 m_3}{k_3 m_2}+\frac{k_3 m_2}{k_2 m_3}\right)\cos(k_1 d_1)\sin(k_2 d_2)\sin(k_3 d_3)\\ &-\frac{1}{2}\left(\frac{k_1 m_3}{k_3 m_1}+\frac{k_3 m_1}{k_1 m_3}\right)\sin(k_1 d_1)\cos(k_2 d_2)\sin(k_3 d_3)\\ &-\frac{1}{2}\left(\frac{k_1 m_2}{k_2 m_1}+\frac{k_2 m_1}{k_1 m_2}\right)\sin(k_1 d_1)\sin(k_2 d_2)\cos(k_3 d_3),\end{aligned} \tag{5}$$

where $K_i = \sqrt{\frac{2m_i(E - V_i) - K_{||}^2}{\hbar^2}}$, m_i is the effective mass of the electron within a particular media, d_i is the thickness of a particular quantum layer, $d = d_1+d_2+d_3$, and V_i is the potential energy of the electron in a particular quantum layer, E is the total energy of the electron, and $K_{||}$ is the parallel component of the electron wave vector. In the sub-wavelength limit i.e., $k_i d_i \ll 1$, one can expand $\cos(k_i d_i)$ and $\sin(k_i d_i)$ in Taylor series,

$$\begin{aligned}\cos(k_i d_i) &= 1 - \frac{(k_i d_i)^2}{2} + O((k_i d_i)^2),\\ \sin(k_i d_i) &= k_i d_i + O(k_i d_i).\end{aligned} \tag{6}$$

Substituting (6) into (5) and neglecting the higher order terms, the dispersion relationship can be simplified as-

$$\begin{aligned}\cos(Kd) &= 1 - \frac{(k_1 d_1)^2}{2} - \frac{(k_2 d_2)^2}{2} - \frac{(k_3 d_3)^2}{2}\\ &-\frac{1}{2}\left(\frac{k_2 m_3}{k_3 m_2}+\frac{k_3 m_2}{k_2 m_3}\right)k_2 d_2 k_3 d_3 - \frac{1}{2}\left(\frac{k_1 m_3}{k_3 m_1}+\frac{k_3 m_1}{k_1 m_3}\right)k_1 d_1 k_3 d_3\\ &-\frac{1}{2}\left(\frac{k_1 m_2}{k_2 m_1}+\frac{k_2 m_1}{k_1 m_2}\right)k_1 d_1 k_2 d_2.\end{aligned} \tag{7}$$

It is obvious that the band gap occurs at the position where $\cos(Kd) = 1$. The gap edges appear as (also see [16]) :

$$\begin{aligned}&\frac{(k_1 d_1)^2}{2} + \frac{(k_2 d_2)^2}{2} + \frac{(k_3 d_3)^2}{2} + \frac{1}{2}\left(\frac{k_2 m_3}{k_3 m_2}+\frac{k_3 m_2}{k_2 m_3}\right)k_2 d_2 k_3 d_3\\ &+\frac{1}{2}\left(\frac{k_1 m_3}{k_3 m_1}+\frac{k_3 m_1}{k_1 m_3}\right)k_1 d_1 k_3 d_3 + \frac{1}{2}\left(\frac{k_1 m_2}{k_2 m_1}+\frac{k_2 m_1}{k_1 m_2}\right)k_1 d_1 k_2 d_2 = 0.\end{aligned} \tag{8}$$

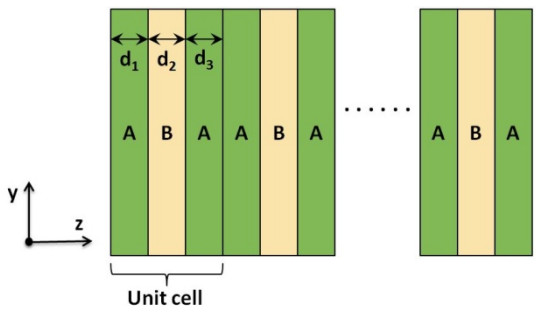


Fig. 1 Schematic drawing of 1D photonic crystal structure Here, $Hg_{0.85}Cd_{0.15}Te$ is the background material (first and last medium).

Solving this equation assuming no parallel component of the electron flux yields exactly the same condition that we drew from the analogy (also see [16] : Type-A structures for analogous electromagnetic conditions of Eq (4) given in this article).

Now we shall use this analogy to predict the band edge energy at which our semiconductor super lattice will show no electron transmission. Of course, just like SASN band gap [8, 11], our semiconductor equivalent band gap only works in sub wavelength regime. This means, the wavelength of the electron wave within the super lattice should be greater than the thickness of our unit cell. In this work, we have used $Hg_{1-x}Cd_xTe$ ternary alloy as our material which has zinc blend structure [13]. At different layers, we have used different mole fractions. The reason of using this II-VI semiconductor hetero-structure is that it is a narrow band gap material. Hence, its mole fractions can be changed to yield inverted band structure which yields negative effective mass within the material [13]. We have calculated the transmission coefficient of the envelop function using Kane model. This material has energy dependent effective mass. The effective mass is related to the energy by the following relation [14,15]-

$$\frac{1}{m} = \frac{2P^2}{3}\left[\frac{2}{E - E_{\Gamma_8}} + \frac{1}{E - E_{\Gamma_7}}\right], \tag{9}$$

Here, E_{Γ_7} and E_{Γ_8} are respectively the band edge energy of the Γ_7 and Γ_8 bands. The quantity,

$$\begin{aligned}P &= \frac{-i}{m_0}\langle S|\frac{\hbar}{i}\frac{\partial}{\partial x}|X\rangle = \frac{-i}{m_0}\langle S|\frac{\hbar}{i}\frac{\partial}{\partial y}|Y\rangle\\ &= \frac{-i}{m_0}\langle S|\frac{\hbar}{i}\frac{\partial}{\partial z}|Z\rangle,\end{aligned} \tag{10}$$

Which is the element of the Kane matrix, where $|S\rangle$, $|X\rangle$, $|Y\rangle$ and $|Z\rangle$ are respectively S-like and P-like Eigen functions for $K = 0$. For semiconductors with a zinc blende structure, the effective potential V can be assumed energy independent, and is equal to $V = E_c$, where $E_c = E_{\Gamma_7}$ is the conduction band (S-type symmetry) edge energy.

For simulation, we have chosen $Hg_{0.73}Cd_{0.27}Te$| $Hg_{0.92}Cd_{0.08}Te$| $Hg_{0.73}Cd_{0.27}Te$ unit cell. The parameters needed for calculating the envelope function can be obtained from reliable measurements or from first principle calculations which suggest $2m_0P^2 \approx 18.5$ eV, $E_{\Gamma_6} = (1.47x + 0.008)$ eV, $E_{\Gamma_7} = (-0.36x + 0.36)$ eV, and $E_{\Gamma_8} = (-0.36x - 0.59)$ eV where x represents the $Hg_{1-x}Cd_xTe$ mole fraction [12].

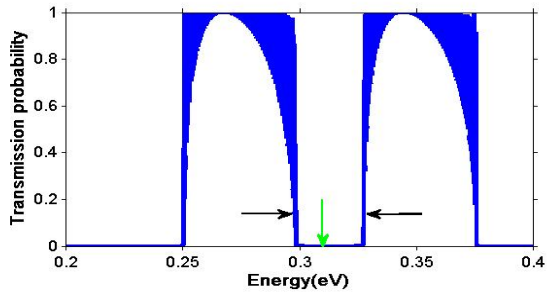


Fig. 2 Transmission Probability vs. energy graph for $Hg_{0.73}Cd_{0.27}Te|$ $Hg_{0.92}Cd_{0.08}Te|$ $Hg_{0.73}Cd_{0.27}Te$ system. The $\sum_i m_i d_i = 0$, condition is met at 0.29eV (left arrow) and the $\sum_i 2(E - V_i)d_i = 0$, condition is met at 0.33eV (right arrow). Here 100 unit cells have been cascaded and dimensions are - $d_1 = 5nm, d_2 = 10nm, d_3 = 5nm$.

We have chosen this material combination to meet the SASN condition and have found band gap at our predicted energy range (Fig. 2). Fig. 2 shows that electronic states cannot exist in between 0.29eV and 0.33eV. So the electrons having the energy in between this energy range cannot propagate through our semiconductor super lattice. From Fig. 3, we can see that the probability wave associated with the energy of 0.31eV (vertical arrow, Fig. 2), which lies in the forbidden region decays exponentially when it penetrates the super lattice.

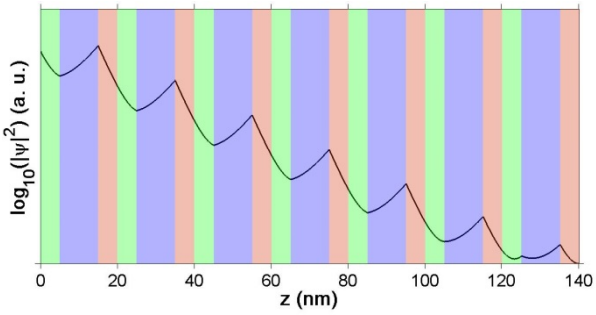


Fig. 3 Amplitude of the envelope function (arbitrary units) within the designed semiconductor heterostructure (seven unit cells has been used).

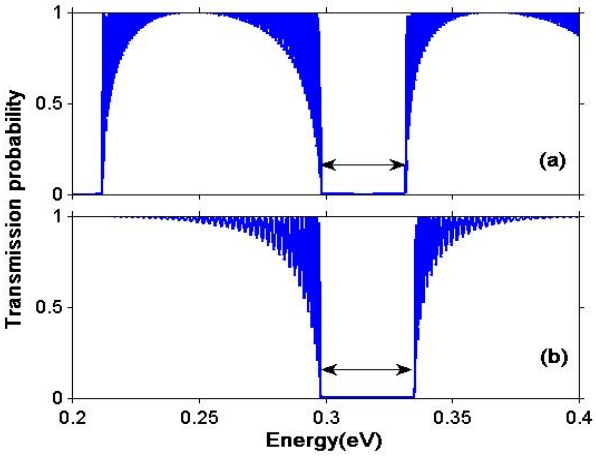


Fig. 4 Lattice constant is scaled by 0.7 (a) and 0.4 (b) for the $Hg_{0.73}Cd_{0.27}Te|$ $Hg_{0.92}Cd_{0.08}Te|$ $Hg_{0.73}Cd_{0.27}Te$ system. The results clearly demonstrate the scaling independent nature of SASN band gap.

Another interesting property of this band gap is- unlike Bragg band gap, it is independent of scaling. Different scaling factors have been employed on our previous design. Still the band gap remained almost same (Fig. 4). Besides it is also possible to vary the zero effective phase gap by varying the ratio of the layer thicknesses in a very controlled and desired way (see Fig. 5)

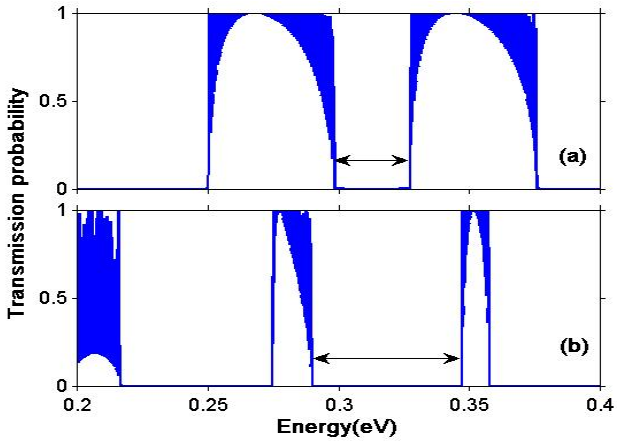


Fig. 5 Layer thickness d_1/d_2 has been changed from 0.5 (a) to 1.0 (b) for the $Hg_{0.73}Cd_{0.27}Te|$ $Hg_{0.92}Cd_{0.08}Te|$ $Hg_{0.73}Cd_{0.27}Te$ system. Variation in the band gap has been obtained by changing the thickness ratio.

III. BANDGAP AND PERFECT TUNNELING IN THE SAME STRUCTURE

It is possible to get perfect tunneling in semiconductor hetero-structure [12], which has been shown recently. But in multi-layer hetero-structures (array of hetero-structures), such perfect tunneling is counterintuitive. Now we shall demonstrate that such a perfect tunneling is also possible at perfect tunneling energy for our array structure, which means transmission probability of electron through the crystal is 100%. Our designed structure has CdTe| HgTe| CdTe sub-wavelength unit cell. When the unit cell is repeated 100 times, we still found perfect tunneling at 0.34eV for which two different conditions in Eq (4) of this article has been applied at single target energy instead of two different energy values (Fig. 6) (also see the condition of perfect tunneling of electromagnetic wave for Type-A unit cells in [16]). Then simply a question arises: how such a perfect tunneling is possible in a multi-layer structure. The answer lies with the idea of Type-A electromagnetic unit cell reported in [16]. Recently, we have shown that such unit cells support perfect tunneling just at the target frequency even if they are used to make an array structure [16] (specially for normal incidence of light) . In addition, it has been observed that electromagnetic wave does not gain any phase shift at the target perfect tunneling frequency for normal incidence in absence of loss [16]. The detail quantum mechanical behavior and physical insights of similar kind of unit cells (not multi-layers) have been rigorously investigated in semiconductor regime in [17]. But the most probable explanation of such perfect tunneling in multi-layer semiconductor hetero-structure (array of unit cells) for ballistic electron is that: at perfect tunneling target energy, the electron wave behaves like photon (canonical or wave nature of photon proposed by Barnett et al. [18]) instead of particle nature and it shows perfect tunneling at normal incidence due to the two matching conditions of Eq (4) satisfied at a single energy. Instead of opening a band gap, at that single target energy, the probability wave will pass through our structure as if the multi-layer super-lattice wasn't even there. According to the band-gap theory of array structures, a more rigorous physical explanation can be drawn [19] in favor of our proposal. The statement reported in [19] : " ….. which

results $K_{BZ} = 0$. The latter quantity assures that the wave is restored after passing through the crystal" is true only for the target frequency for Type-A electromagnetic array structures. K_{BZ} is the Bloch wave vector reported in [19]. That idea of zero width band gap [19] most probably remains valid also for semiconductor hetero-structures at the target energy and hence for ballistic electrons, such perfect tunneling may occur at that energy (see Left side of Eq (7) in this article). We have also plotted the envelope function at the energy of 0.34eV (Fig. 7). The envelope function shows that the input and output wave function has the same amplitude which proves our claim. We can also see from the graph that there is a strong delocalization of electronic states at the interface of equivalent MNG-ENG.

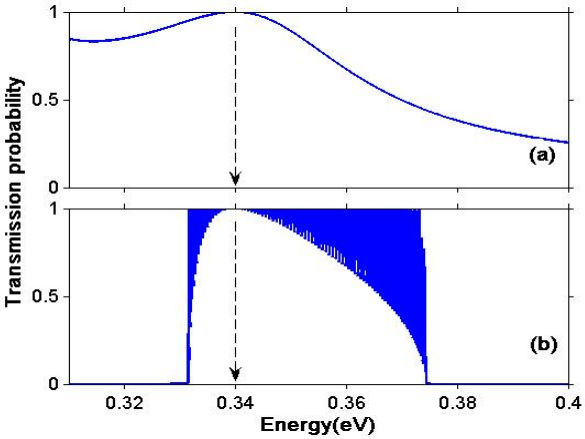


Fig. 6 Transmission probability of electron through CdTe| HgTe| CdTe sub-wavelength unit cell (a) and 100 cascaded unit cells (b). Perfect tunneling is achieved at 0.34eV when two conditions of Eq (4) are applied at that specific energy. Here 100 unit cells have been cascaded and dimensions are - $d_1 = 1.26nm, d_2 = 10nm, d_3 = 1.26nm$. It is important to note that the band gap found in this structure (Fig. 6b) is not the equivalent SASN band-gap that we have proposed in this article. The band-gaps of this multi-layer structure have opened naturally due to repeated array of unit cell hetero-structures.

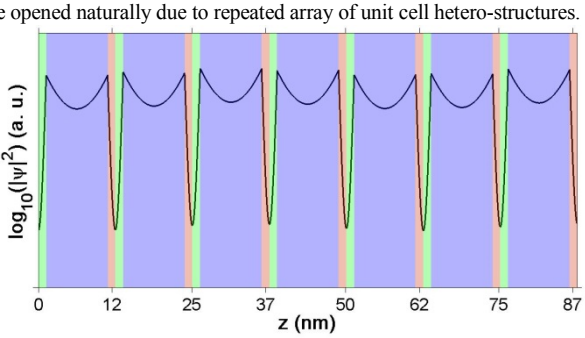


Fig. 7 Envelope function (arbitrary units) for 10 unit cells at perfect tunneling energy (0.34eV)

This perfect tunneling phenomenon (with band-gap) can have significant impact on device technology. High on-off current ratio is always wanted in CMOS technology. Graphene transistors are not considered as logic transistor because of low on-off ratio [20] although having better transport capabilities. Our proposed device shows a steep transition in transmission probability (Fig. 6) which may play a key role in obtaining high on-off ratio even if the applied energy is very low.

IV. Conclusion

We have shown the Maxwell-Schrodinger analogy to import the idea of electromagnetic SASN band gap to quantum mechanical domain as the major contribution of this work. We have successfully demonstrated that the quantum mechanical band gap can be engineered according to designers will and this band gap is independent of scaling given that the sub-wavelength condition remains valid. Moreover, our proposed multi-layer structure is capable of showing perfect tunneling of probability wave (ballistic electron has been considered) through the hundreds of layers even though the wave function is decaying in individual layers. We have also shown strong localization of electrons near interfaces. As the on-off ratio (step transition from low transmission to high transmission) is very high even if the applied energy is very low, this kind of tunneling structures can be very useful as future low power electronic device and logic transistor.

References

[1] U. Leonhardt, "Science 312, Optical Conformal Mapping and Dielectric Invisibility Devices," *Science*, vol. 312, no. 5781, pp. 1777-1780, June 2006.

[2] J. B. Pendry, D. Schurig, D. R. Smith, "Controlling Electromagnetic Fields," Science, vol. 312, no. 5781, pp. 1780-1782, June 2006.

[3] M. R. C. Mahdy, M. R. A. Zuboraj, A. A.N. Ovi, and M. A. Matin,"An Idea of Additional Modified Modes in Rectangular Patch Antennas Loaded With Metamaterial", *IEEE Antennas and Wireless Propagation Letters*, vol. 10, pp. 869-872. Aug. 2011.

[4] J. B. Pendry, "Negative Refraction Makes a Perfect Lens," *Phys. Rev. Lett.*, vol. 85, pp. 3966, 2000.

[5] M. M. Hasan, D. S. Kumar, M. R. C. Mahdy, D. N. Hasan, M. A. Matin, "Robust Optical Fiber Using Single Negative Metamaterial Cladding" ,*IEEE Photonics Technology Letters*, Vol. 25, pp. 1043-1046 , June 2013.

[6] Y. J. Xiang, X. Y. Dai, S. C. Wen, Z. X. Tang, D. Y. Fan, "Zero-effective-phase bandgaps in photonic multilayers: analytic expressions for band-edge frequencies and broadband omnidirectional reflection," *J. Opt. Soc. Am. B*, vol. 28, pp. 1187-1193, 2011.

[7] Haitao Jiang, Hong Chen, Hongqiang Li, Yewen Zhang, Jian Zi, and Shiyao Zhu, "Properties of one-dimensional photonic crystals containing single-negative materials," *Phys. Rev. E*, vol. 69, pp. 066607, June 2004.

[8] Yi Weng, Zhi-Guo Wang, and Hong Chen, "Band structures of one-dimensional subwavelength photonic crystals containing metamaterials," *Phys. Rev. E*, vol. 75, pp. 046601, April 2007.

[9] T.B. Wang, J.W. Dong, C.P. Yin, H.Z. Wang, "Complete evanescent tunneling gaps in one-dimensional photonic crystals," *Phys. Lett. A*, vol. 373, pp. 169–172, 2008.

[10] Li-Gang Wang, Hong Chen, and Shi-Yao Zhu, "Omnidirectional gap and defect mode of one-dimensional photonic crystals with single-negative materials," *Phys. Rev. B*, vol. 70, pp. 245102, December 2004.

[11] Xiang, Y.; Dai, X.; Wen, S.; Tang, Z.; Fan, D., "Extending the zero-effective-phase photonic bandgap by one-dimensional ternary photonic crystals," *Appl. Phys. B*, vol. 103, pp. 897–906, 2011.

[12] L Jelinek, J D Baena, J Voves, R Marques, "Metamaterial-inspired perfect tunnelling in semiconductor heterostructures," *New Journal of Physics*, vol. 13, pp. 083011, 2011.

[13] J. Chu, A. Sher, "Physics and properties of narrow band gap semiconductor," Springer, 2008.

[14] G. Bastard, "Wave Mechanics Applied to Semiconductor Heterostructures," New York: InterscienceWiley-Interscience, 1988.

[15] G. Bastard, "Theoretical investigations of superlattice band structure in the envelope-function approximation," *Phys. Rev. B*, vol. 25, pp. 7584, June 1982.

[16] Arif Shahriar, M.R.C. Mahdy, Jubayer Shawon, Golam Dastegir Al-Quaderi, M. A. Matin, "On the Relation between Perfect Tunneling and Band Gaps for SNG Metamaterial Structures," arXiv:1308.32.

[17] Mario G. Silveirinha, Nader Engheta, "Spatial Delocalization and Perfect Tunneling of Matter Waves: Electron Perfect Lens," *Phys. Rev. Lett.*, vol. 110, pp.213902, 2013.

[18] Stephen M. Barnett, "Resolution of the Abraham-Minkowski Dilemma," *Phys. Rev. Lett.*, vol. 104, pp.070401, 2010.

[19] S. Zouhdi, A. V. Dorofeenko, A. M. Merzlikin and A. P. Vinogradov, "Theory of zero-width band gap effect in photonic crystals made of metamaterials," *Phys. Rev. B*, vol. 75, pp. 035125, 2007.

[20] Fengnian Xia , Damon B. Farmer , Yu-ming Lin and Phaedon Avouris ,"Graphene Field-Effect Transistors with High On/Off Current Ratio and Large Transport Band Gap at Room Temperature," *Nano Lett.*, 10 (2), pp 715–718, 2010.

Effect of Surface Wettability Controlled by Ultraviolet-Ozone treatment on Spin-Coated ZnO Seed Layer and ZnO Nanowires Grown by Hydrothermal Method

Imtiaz Mahmud[1], Ji-Sub Park[1], Han Jae Shin[1,2], Min-Kyu Park[1], Do Kyung Lee[3] and Hak-Rin Kim[1,*]

[1]School of Electronics Engineering, Kyungpook National University,
Daegu, 702-701, South Korea

[2]Gumi Electronics and Information Technology Research Institute,
Gumi-si, Gyeongsangbuk-do, 730-853, South Korea

[3]Department of Advanced Energy Material Science and Engineering, Catholic University of Daegu,
Gyeongsan-si, Gyeongsangbuk-do, 712-702, South Korea

[*]rineey@ee.knu.ac.kr

Abstract—**We presented improved density and aspect ratio of ZnO nanowires which were hydrothermally grown on spin coated ZnO seed layer. After UVO treatment on silicon surface, spin coated ZnO seed layer showed a high coverage density due to increased surface wettability without forming island like growth in conventional method. ZnO active nucleation sites showed a positive increasing relation with increase in coverage density. Whereas, increased spin coating time had risen surface roughness with rise in coverage area, UVO treated surface wettability change had showed lower surface roughness. Moreover, seed layer thickness remained unaffected for surface wettability changing effect. This let ZnO nanowires grow on the increased ZnO active nucleation sites without hindrance of other nanowires. Hence, higher density and aspect ratio ZnO nanowires were achieved by controlling surface wettability via UVO treatment.**

Index Terms—**ZnO seed layer, ZnO growth, UVO treatment, Surface treatment, Surface wettability.**

I. Introduction

Zinc oxide (ZnO) has attracted much attention during the past decade owing to its applicability as building blocks in numerous electronic, optic and photonic devices [1]-[3]. Being an ample, nontoxic and II-VI semiconductor material, it provides high electron mobility, wide band gap energy (3.3 eV), high exciton binding energy (60 meV) and large piezoelectric coefficient [4], [5]. Especially, in nanoscale extent, ZnO nanowires or nanostructures provide several important chemical and physical properties due to its high aspect ratio and growth density [1]. Therefore, it has showed significant performance in advanced applications like gas sensors, chemical sensors, light emitting diodes, photodiodes, solar cells, piezoelectric devices and so on [6]-[12]. Several growth methods for ZnO nanowires have been developed like pulsed-laser deposition (PLD), chemical vapour deposition (CVD), vapour-phase transport (VPT), thermal evaporation, electrode deposition and so on [6]-[17]. However, costly apparatus and complex processes with high working temperature have made these techniques tough to implement for industrial production. Instead, hydrothermal method offers facile operating technique in low temperature with low cost simplest apparatus, thus it is regarded as much compatible method for large scale production [18].

However, ZnO seed layer plays a vital role for ZnO nanowires growth in hydrothermal method. ZnO nanowires grow on ZnO active nucleation sites presented in seed layer in accordance with their crystallinity [19]. Uniformity, coverage area and thickness of seed layers need to be optimized for better growth of ZnO nanowires. In conventional method, with increased time spin coating to increase seed layer coverage area and ZnO active nucleation sites, surface roughness and thickness is also increased which ultimately results in lower density and aspect ratio ZnO nanowires growth [20]. In this study, we introduced controlled surface wettability via UVO treatment to enhance coverage area and density of ZnO active nucleation sites without affecting seed layer surface roughness and thickness.

Although, surface wettability change has been applied to enhance various applications of ZnO, but it has not been implemented to control ZnO seed layer growth in hydrothermal method. Furthermore, ultraviolet ozone (UVO) treatment has already verified itself as a well-established surface treatment method to change surface wettability in low cost approach besides cleaning the surface via vacuum cleaning process [21]. Previously, works like ECR plasma treatment effect on ZnO nucleation by atomic layer epitaxy method, effect of UVO treated seed layer on ZnO nanowire growth, UVO and self-assembled monolayers for patterned growth of ZnO nanowires had been conducted, but no work had investigated surface wettability changing effect on ZnO seed layer and nanowires growth using UVO treatment in hydrothermal method [22]-[24].

In this work, UVO treated surface wettability changing effect on spin coated ZnO seed layer growth was explored. Furthermore, its consequence on hydrothermally grown ZnO nanowires was examined as well.

II. Experimental Procedure

ZnO nanowires were grown on silicon substrate in hydrothermal method. Silicon substrate was first cleaned with acetone, isopropyl alcohol (IPA) and deionized (DI) water successively in sonic cleaner. Then it was etched in buffered

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

oxide etch (BOE) to remove native oxide layer, afterwards it was UVO treated to change surface wettability and surface energy. For seed growth, the seed solution was prepared using zinc acetate dehydrate ($Zn(CH_3COO)_2.2H_2O$, 30 mM) dissolved in ethanol (50 mL, 99.9%). The seed solution was dropped onto the silicon substrate surface and spin-coated at 850 rpm for 30 sec. It was followed by a soft baking at 80°C. To observe UVO treatment effect on various times spin coating, this spin coating and baking process was done for 1, 3 and 5 times. Afterwards, the spin coated substrates were annealed at 80˚C for 1 hour. After annealing process, ZnO nanowires were formed over the seed layer using aqueous solution consisting of zinc nitrate hexahydrate ($Zn(NO_3)_2.6H_2O$, 25 mM), hexamethylenetetramine (25 mM) and DI water. The ZnO nanowires were grown using sample dipping method. The spin coated samples were suspended vertically in the solution for 3 hours at 92˚C for the final growth of the ZnO nanowires.

Contact angle measurement equipment (KRUSS DSA100) was used to check the contact angle as well as surface wettability for different UVO treatment time. Energy dispersive x-ray spectroscopy (EDX) was used to characterize ZnO active nucleation sites in the seed layer depending on spin coating time and UVO treatment. Field emission scanning electron microscope (FE-SEM) (Hitachi S4800) was used to characterize the growth density and aspect ratio of ZnO nanowires.

III. Result and Discussion

A. *Effect of UVO Treatment and Spin Coating Conditions on Seed Layer*

The silicon substrate surface was UVO treated to control surface wettability. Before UVO treatment, silicon substrates were etched with BOE which removed the native oxide layer over silicon substrate surface and grew SiH or SiH_2 groups over it. The ozone presented in UVO treatment had reacted with hydrogen sited over the silicon substrate surface and converted it into hydroxyl group [21], [25]. This hydroxyl group attracted water molecules and increased surface wettability as well as surface energy [21], [25], [26].

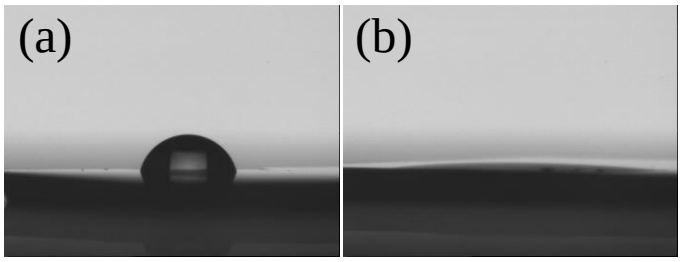


Figure 1. Effect of UVO treatment on silicon substrate surface wettability: (a) etched sample with no UVO treatment (water contact angle: 78˚), (b) etched sample with 540 second UVO treatment (water contact angle: <1˚).

Initially, Surface wettability changed rapidly with UVO treatment, but slowed down gradually with increased surface wettability by more UVO treatment time as shown in Fig. 1.

ZnO seed layer growth was enormously enhanced with surface wettability change via UVO treatment. The presence of extra hydroxyl group over the silicon substrate surface attracted zinc ions of the seed solution and bind them on the surface. Thus more ZnO nucleation sites were created. This would be comprehended with EDX images of Fig. 2, as dense active nucleation sites were observed in the seed layer of UVO treated substrate surface in each case of different time spin coating. The inside of this phenomena could be explained with the following seed solution reaction mechanism [28]:

$$Zn(CH_3COO)_2 \,.\, 2H_2O + 2CH_3\text{-}CH_2\text{-}OH \longrightarrow ZnO + 2CH_3\text{-}CH_3\text{-}CH_2\text{-}CO + 3H_2O \quad (1)$$

In this substitution reaction, hydroxyl group from ethanol replaced acetate group of zinc acetate and formed zinc oxide. In our case, extra hydroxyl group sited over the silicon substrate via UVO treatment reacted with the seed solution. This correspondingly forwarded the reaction and produced more ZnO active nucleation sites in the seed layer for each case of spin coating time. Additionally, with increased surface wettability, seed solution was moderately spread over the surface during spin coating, that ultimately increased interaction with hydroxyl group presented in the surface and produced moderate seed layer all over the surface. That resulted in continuous thin film growth for each case of spin coating time.

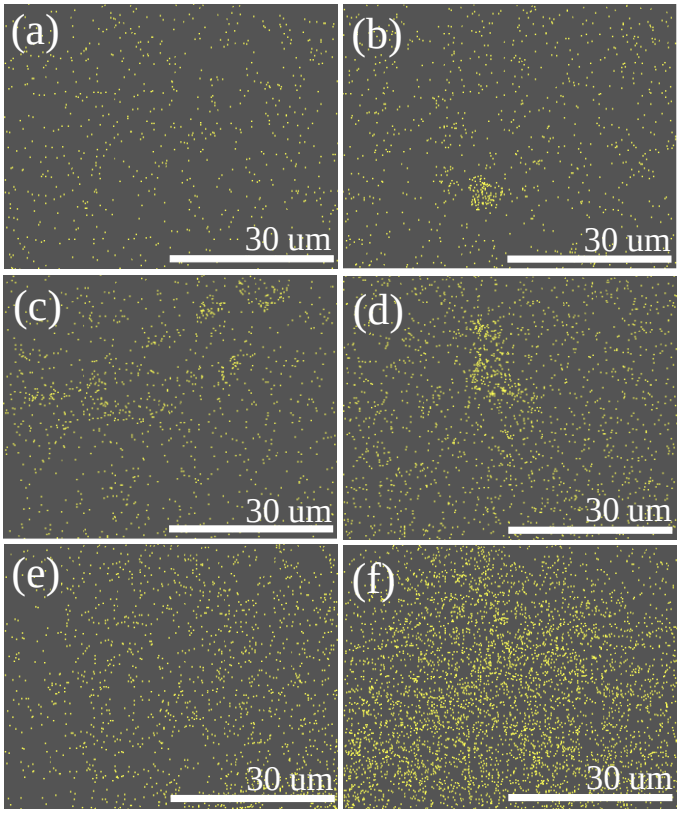


Figure 2. EDX images showing relative position of ZnO seed layer active nucleation sites with different UVO treatment and different times spin coating: (a), (c), (e) are 1, 3 and 5 time spin coated samples with no UVO treatment respectively and (b), (d), (f) are 1, 3 and 5 times spin coated samples with 540 second UVO treatment.

However, in seed growth without UVO treatment, seed layer was not uniformly distributed throughout the surface, rather island like growth was observed. With increased time spin coating, this island area and thickness were increased. However, ZnO active nucleation sites were also increased as shown in Fig.2. Therefore, with increased time spin coating, seed layer growth covering whole surface area would be achieved but would be accompanied by higher seed layer surface roughness and thickness [20].

In accordance with this phenomenon, it was confirmed by the AFM images that, with increased time spin coating,

surface roughness was increased, but UVO treated surface wettability change had lowered surface roughness. Furthermore, thickness of the seed layer was unchanged with change in surface wettability. This would remove the influence of other nanowires in growth process of each nanowires in its neighborhood, which would ultimately result in higher density and aspect ratio ZnO nanowires growth [20].

B. ZnO Nanowires Growth Dependence on Seed Layer

Change in the properties of ZnO seed layer actively affected ZnO nanowires growth in hydrothermal method. As shown in Fig. 3, uniformly distributed high density growth of ZnO nanowires all over the substrate surface were achieved for each case of spin coating by changing surface energy using UVO treatment. This was due to uniformly distributed growth of ZnO seed layer and active nucleation sites for change in surface wettability. In contrast, without UVO treatment, island like growth of ZnO nanowires were observed according to seed layer growth for various time spin coating.

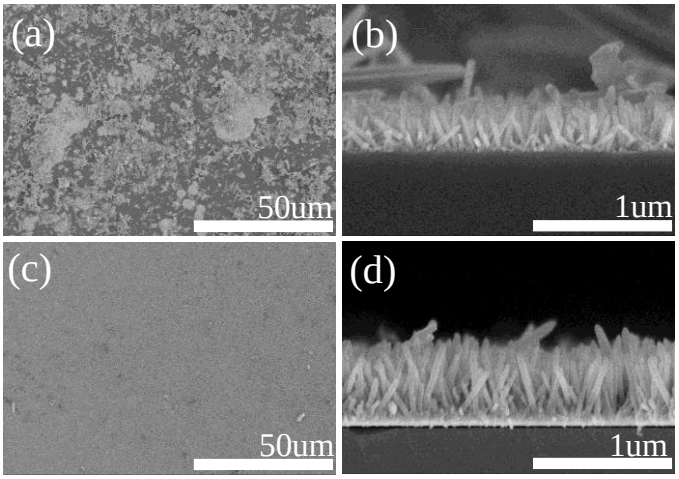


Figure 3. SEM images of ZnO nanowires growth showing effect of UVO treatment, (a),(b) surface and cross section of ZnO nanowires growth at 3 time spin coating with no UVO treatment and (c),(d) surface and cross section of ZnO nanowires growth at 3 time spin coating with 540 second UVO treatment.

However, Fig. 4(a) and Fig. 4(b) respectively shows change in density and aspect ratio of as grown ZnO nanowires at different conditions. Density was uniformly increased with change in surface wettability for 1 time and 3 times spin coating, but drastic negative change was observed in case of 5 times spin coating. This was because, increased surface roughness had produced more thick ZnO nanowires which was directly affected by enriched ZnO active nucleation sites of increased time spin coated thick seed layer. Each nanowire growth was affected by nearby nanowires that resulted comparatively bad crystalline, high diameter nanowires. Additionally, 5 times spin coating with and without UVO treatment results were almost same, because higher time spin coating had produced seed layer covering almost whole surface area of the silicon substrate as achieved in surface wettability controlled growth. But in case 1 time or 3 times spin coating, the uniformly distributed seed layer with moderate roughness and distributed ZnO active nucleation sites with surface wettability change, enabled nanowires to grow without hindrance from other nanowires in the vicinity. Thus higher density nanowires growth was achieved with precise surface wettability.

Moreover, aspect ratio of ZnO nanowires was always increased with surface wettability change via UVO treatment. As shown in Fig. 4(b), a highest 30% increase of aspect ratio was observed for surface wettability change in case of 3 times spin coating. Well crystalline uniform growth of seed layer over the surface resulted this phenomenon. Additionally, 5 times spin coated with and without UVO treated substrates showed almost same density, but aspect ratio was greatly enhanced. It was because, with increased time spin coating, although seed layer covered whole surface area, but its surface roughness was drastically increased, on the other hand, UVO treated sample had lower surface roughness that resulted in higher aspect ratio growth. Moreover, the uniformly distributed seed layer active nucleation sites by lower surface roughness removed interference of other nanowires during growth process. This enhanced the ZnO nanowires growth process and higher aspect ratio nanowires were formed.

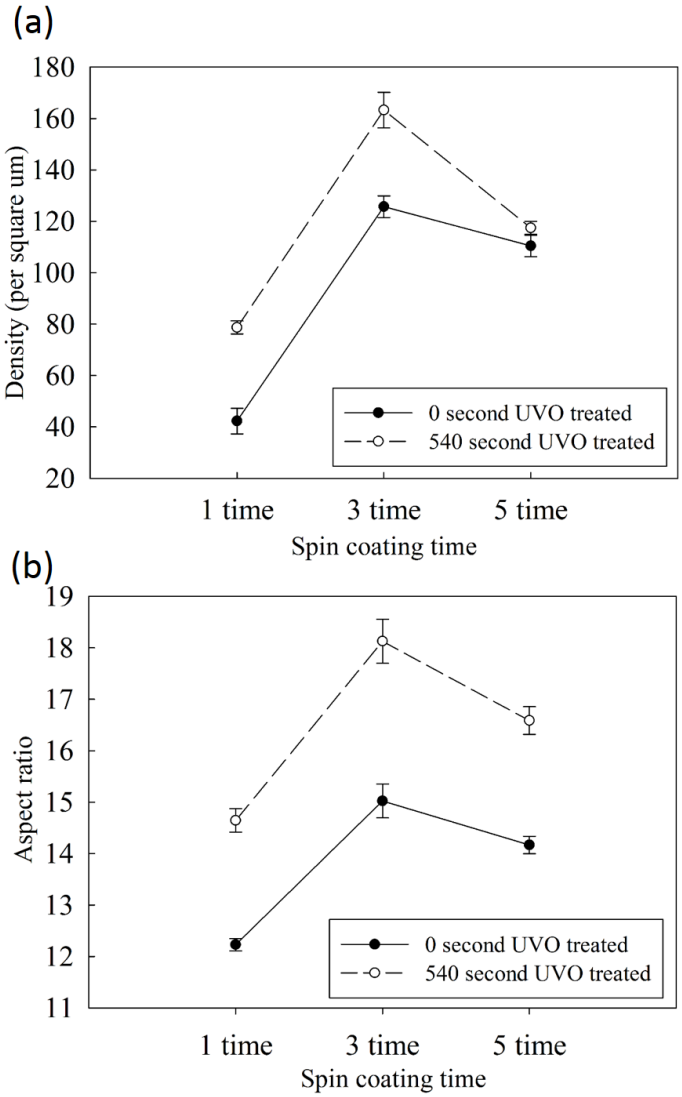


Figure 4. UVO treatment surface wettability changing effect on ZnO nanowire growth at different spin coating time; (a) effect on growth density, (b) effect on growth aspect ratio.

IV. CONCLUSIONS

UVO treated controlled change in surface wettability positively and significantly affected ZnO nanowires growth. Interaction between the extra hydroxyl group formed over the silicon substrate during UVO treatment and ZnO seed solution had produced uniformly distributed seed layer all over the

surface in each case of spin coating time. It had produced moderately distributed extra ZnO active nucleation sites in seed layer. Most importantly, seed layer surface roughness was enhanced that ultimately resulted in uniformly distributed higher density and aspect ratio ZnO nanowires growth covering the whole surface of the silicon substrate. Furthermore, we have been investigating its applicability for patterned growth in low cost hydrothermal method. Therefore, optimized spin coating time with UVO treatment would be very useful for large scale practical implementation in elementary and industrial applications.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (2008-0062617) and was financially supported by the "Over regional linked 3D convergence industry promotion program" through the Ministry of Trade, Industry & Energy(MOTIE) and Korea Institute for Advancement of Technology (KIAT) (N0.R0001519).

REFERENCES

[1] M. Law, J. Goldberger and P. Yang, "Semiconductor nanowire and nanotubes," *Annu. Rev. Mater. Res.*, vol. 34, pp. 83-122, Aug. 2004.

[2] C. M. Lieber and Z. L. Wang, "Functional nanowires," *MRS Bull.*, vol. 32, pp. 99-108, Feb. 2007.

[3] E. C. Garnett, M. L. Brongersma, Y. Cui and M. D. McGehee, "Nanowire solar cells," *Annu. Rev. Mater. Res.*, vol. 41, pp. 269-295, Aug. 2011.

[4] Z. L. Wang, "Zinc oxide nanostructures: growth, properties and applications," *J. Phys.: Condens. Matter.*, vol. 16, pp. R829-R858, Jun. 2004.

[5] G. C. Yi, C. Wang and W. I. Park, "ZnO nanorods: systhesis, characterization and applications," *Semicond. Sci. Technol.*, vol. 20, pp. S22-S34, Mar. 2005.

[6] P. Mitra, A. Chatterjee and H. Maiti, "ZnO thin film sensor," *Materials Letters*, vol. 35, pp. 33-38, Apr. 1998.

[7] G. Shen, P.-C. Chen, K. Ryu and C. Zhou, "Devices and chemical sensing applications of metal oxide nanowires," *J. Mater. Chem.*, vol. 19, pp. 828-839, 2009.

[8] S. Liang, H. Sheng, Y. Liu, Z. Huo, Y. Lu and H. Shen, "ZnO Schottky ultraviolet photodetectors," *Journal of Crystal Growth*, vol. 225, pp. 110-113, May. 2001.

[9] N. Saito, H. Haneda, T. Sekiguchi, N. Ohashi, I. Sakaguchi and K. Koumoto, "Low-Temperature Fabrication of Light-Emitting Zinc Oxide Micropatterns Using Self-Assembled Monolayers," *Adv. Mater.*, vol. 14, no. 6, pp. 418-421, Mar. 2002.

[10] J. Lee, Y. Choi, J. Kim, M. Park and S. Im, "Optimizing n-ZnO/p-Si heterojunctions for photodiode applications," *Thin Solid Films*, Vols. 403-404, pp. 553-557, Feb. 2002.

[11] J. B. Baxter, A. M. Walker, K. v. Ommering and E. S. Aydil, "Synthesis and characterization of ZnO nanowires and their integration into dye-sensitized solar cells," *Nanotechnology*, vol. 17, pp. S304-S312, May. 2006.

[12] X. Wang, J. Zhou, J. Song, J. Liu, N. Xu and Z. L. Wang, "Piezoelectric field effect transistor and nanoforce sensor based on a single ZnO nanowire," *Nano Letters*, vol. 6, pp. 2768-2772, Oct. 2006.

[13] Y. Sun, G. M. Fuge and M. N. R. Ashfold, "Growth of aligned ZnO nanorod arrays by catalyst-free pulsed laser deposition methods," *Chem. Phys. Lett.*, vol. 396, pp. 21-26, Sep. 2004.

[14] H. Wan and H. E. Ruda, "A study of the growth mechanism of CVD-growth ZnO nanowires," *J. Mater. Sci.: Mater. Electron.*, vol. 21, pp. 1014-1019, Oct. 2010.

[15] S. C. Lyu, Y. Zhang, C. J. Lee, H. Ruh and H. J. Lee, "Low-temperature growth of ZnO nanowire array by a simple physical vapor-deposition method," *Chem. Mater.*, vol. 15, pp. 3294-3299, Jul. 2003.

[16] Z. W. Pan, Z. R. Dai and Z. L. Wang, "Nanobelts of semiconducting oxides," *Science*, vol. 291, pp. 1947-1949, Mar. 2001.

[17] M. J. Zheng, L. D. Zhang, G. H. Li and W. Z. Shen, "Fabrication and optical properties of large-scale uniform zinc oxide nanowire array by one-step electrochemical deposition technique," *Chem. Phys. Lett*, vol. 363, pp. 123-128, Sep. 2002.

[18] M. Guo, P. Diao and S. Cai, "Hydrothermal growth of well-aligned ZnO nanorod arrays: Depensence of morphology and alignment ordering upon preparing conditions," *J. Solid State Chem.*, vol. 178, pp. 1864-1873, Jun. 2005.

[19] K. A. Wahid, W. Y. Lee, H. W. Lee, A. S. Teh, D. C. S. Bien and I. A. Azid, "Effect of seed annealing temperature and growth duration on hydrothermal ZnO nanorod structures and their electrical characteristics," *Appl. Surf. Sci.*, vol. 283, pp. 629-635, Jul. 2013.

[20] T. A. N. Peiris, H. Alessa, J. S. Sagu, I. A. Bhatti, P. Isherwood and K. G. U. Wijayantha, "Effect of ZnO seed layer thickness on hierarchical ZnO nanorod growth on flexible substrates for application in dye-sensitised solar cells," *J. Nanopart. Res.*, vol. 15:2115, pp. 1-10, Nov. 2013.

[21] M. Liehr and P. Thiry, "High resolution electron energy loss spectroscopic characterization of insulators for Si technology," *J. Electron Spectroscopy and Related Phenomena*, vol. 54-55, pp. 1013-1032, Dec. 1990.

[22] Y. Masuda, N. Kinoshita, F. Sato and K. Koumoto, "Site-Selective Deposition and Morphology Control of UV- and Visible-Light-Emitting ZnO Crystals," *Crystal Growth & Design*, vol. 6, pp. 75-78, 2006.

[23] J. Lim, K. Shin and C. Lee, "ECR Plasma Pretreatment for ZnO Epitaxy on Sapphire and Silicon Substrates," *Journal of the Korean Physical Society*, vol. 44, pp. 1088-1091, 2004.

[24] S.-Y. Chang, N.-H. Yang and Y.-C. Huang, "Hydrothermal Growth and Interface Correlation of Highly Aligned ZnO Nanorod Arrays on UV-Activated Sol–Gel Transparent Conducting Films," *Journal of The Electrochemical Society*, vol. 156, pp. K200-K204, Sep. 2009.

[25] W. Kern, *Handbook of Semiconductor Wafer Cleaning Technology-Science, Technology and Applications*, William Andrew Publishing/Noyes, 1993.

[26] E. F. Aziz, N. Ottosson, M. Faubel, I. V. Hertel and B. Winter, "Interaction between liquid water and hydroxide revealed by core-hole de-excitation," *Nature Letters*, vol. 455, pp. 89-91, Sep. 2008.

[27] D. Janssen, R. D. Palma, S. Verlaak, P. Heremans and W. Dehaen, "Static solvent contact angle measurements, surface free energy and wettability determination of various self-assembled monolayers on silicon dioxide," *Thin Solid Films*, vol. 515, pp. 1433-1438, Dec. 2006.

[28] H. Du, F. Yuan, S. Huang, J. Li and Y. Zhu, "A new reaction to ZnO nanoparticles," *Chemical Letters*, vol. 33, pp. 770-771, Mar. 2004.

Effect of Addition of Ni Particles on the Microstructural, Mechanical and Electrical Properties of Zn-based High Temperature Solders

M. M. Hasan, *A. Sharif and S. Mallick

Department of Materials and Metallurgical Engineering
Bangladesh University of Engineering and Technology (BUET)
Dhaka-1000, Bangladesh.
*asharif@mme.buet.ac.bd

***Abstract*—Electronic solders are widely used for making interconnections between the electronic chips with the circuit board. The substitution of traditional Lead bearing solders is one of the key challenges to the electronic industries in the current drive for eco-friendly manufacturing. In the present study, Ni powder is added with Zn to investigate the effect of Ni on the microstructural, mechanical and electrical properties of the newly developed solder composites. The microstructures of the Ni/Zn solder composites were composed of pure Zn and Ni-Zn δ phase. Addition of Ni causes to increase the hardness of Zn and keeps the electrical resistivity unchanged.**

***Index Terms*—Pb-free solder, Ni/Zn composite, microstructure, micro-hardness, electrical resistivity etc.**

I. Introduction

In these days, the development of Pb-free solders has become an important issue for electronic interconnection materials because of the health and environmental safety concerning lead usage [1-3]. Pb-free solder systems that have been disclosed in the literature include Sn-based [4-5], Au-based [6-7], Bi-based [8-9] and Zn-based [10-17] alloys. Moreover, the tendency for miniaturization of the electronic devices causes to higher heat flux dissipation and thus requires the solder materials which have a melting temperature higher than 300^0C [3].

Among the several systems Zn-based solders have a lot of prospects due to its competitive price and mechanical properties. The effect of the addition of Al, Cu, Sn, Ca, Mg, Cr and Ti is reported so far [10-17]. Zn-Sn alloys show a much improved ductility compared to other Zn-based alloys, excellent electrical properties and oxidation resistance at high temperature and humidity [13]. Though it has an eutectic composition of 85wt.% Sn at 199˚C, but it can withstand multiple reflows of up to 260 ˚C. In these days, Zn-Al alloys are used to join Al structures [10]. Zn-Al phase diagram shows only one eutectic point with a composition of 6wt.% Al at 381˚C. The high oxygen affinity of these alloys with low wettability is the only drawback [11]. Addition of Cu, Mg, Ga and In as third element in Zn-Al binary system has been investigated [15-17]. Zn–Al–Cu solder alloys showed a reasonable electrical resistivity and spreadability making them suitable for electronic applications [15]. Zn–4Al–3Mg–3Ga shows solidus and liquidus at 309˚C and 347˚C respectively [16]. Good die-attaching with a small amount of voids was achieved at 320˚C or higher in the die-attaching test.

The lifetime of electronic devices depends on the reliability of a solder. It is important for a solder to match the coefficient of thermal expansion, have a high elastic modulus, hardness, yield strength and shear strength together with resistance to corrosion, fatigue and creep [2]. Studies have shown that forming a composite by adding an appropriate second phase particles, of ceramic, metallic or intermetallic is a potentially viable and economically affordable approach to improve the mechanical properties of a solder matrix. In previous research, the control of solderability by the formation, presence and growth of the second phase has been proposed by Guo et al. [18].

So far, the composites of Sn-based [29-32] and Bi-based [33] Pb-free solders are reported. Gain et al. found that the shear strengths and hardness of solder joints containing higher percentage of Ni nano-particles added Sn–9Zn and Sn-8Zn-3Bi exhibited consistently higher value than those of plain solder joint due to control of the fine microstructure as well as homogeneous distribution of Ni nano-particles [19]. Tama et al. [20] and Ahmed et al. [21] found the similar result for $Al_2O_3$and Ag nano-particles added eutectic Sn–9Zn composite solders. Shen et al. [22] successfully prepared a ZrO_2 reinforced composite solder by mechanically dispersing ZrO_2 particles into a eutectic Sn–3.5Ag solder paste, and the composite solders had a refined microstructure with an improved microhardness.

In the present study, different weight percentages of Ni particles were mixed into Zn to prepare Zn/Ni composite solders. The influence of a small amount of reinforcing Ni particles on the microstructural evolution, especially, the morphology and distribution of intermetallic compounds (IMCs) were investigated using metallographic analysis. In addition, the thermal, mechanical and electrical properties of these Zn/Ni composites were measured depending on the content of the reinforcing Ag micro-particles.

II. Experimental

This project aimed to develop high temperature lead-free solder composite. Key materials were commercially available pure Zn (99.99%) and Ni Powder (size < 10 μm, M = 58.71

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

g/mol, Purity > 99.5%, E Merck, Germany). Three different amounts of Ni powders were added to Zn, i.e. 0.20, 0.30 and 0.40 wt.% Ni. In this paper pure Zn, 0.20Ni/Zn, 0.30Ni/Zn and 0.40Ni/Zn (wt.%) composites are referred as 2NZ, 3NZ and 4NZ respectively.

A batch weighing 1.5 kg was cast in each time. Depending upon the composition the amount of raw materials for mixing varied as that of Table I. The constituent elements were melted in a gas fired pit furnace and homogenized at ~530°C. The melts were then poured in a steel mold to prepare the chill cast ingot. The schematic diagram of the cast product is shown in Fig. 1. Our aim was to keep the mould preheat temperature to 210 ± 10°C and the superheat temperature of the molten metal at 120 ± 10°C. The heating parameters that were controlled for casting is listed in Table IIIII.

TABLE IV

AMOUNT OF RAW MATERIALS USED (IN GRAMS)

Sample	Zn	Ni
Z	1496	-
2NZ	1490	2.98
3NZ	1560	4.76
4NZ	1590	6.42

The as-cast solders were sectioned and polished according to non-ferrous metallography with 0.5 µm Al_2O_3 particles in order to obtain the microstructure. After cleaning with acetone, the samples were investigated by an optical microscope with digital camera (LEICA-MZFLIII, Italy).

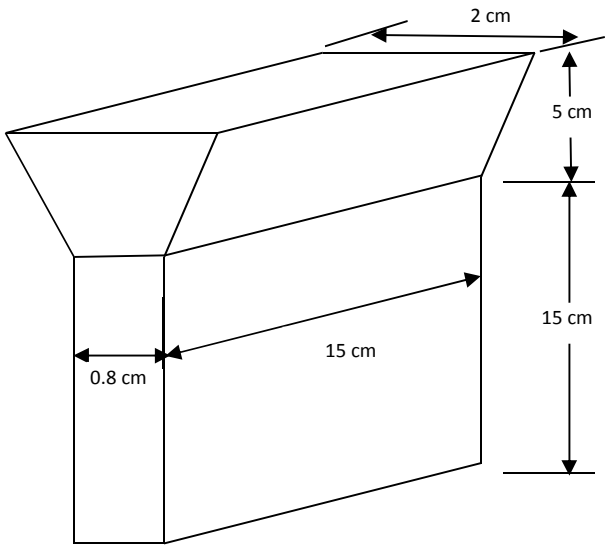


Figure 1: Showing the schematic diagram of the casting with dimensions.

Grinding and polishing were necessary to obtain polished, smooth and flat parallel surface before indentation testing. Thus, the polished samples were placed in a Vickers micro-hardness tester (FV-800, FUTURE-TECH, Japan) to measure the micro-hardness. The applied load was 10 kg for 10 s and at least 10 readings of different indentation were taken at room temperature to obtain the mean value.

TABLE VVI

HEATING PARAMETERS USED FOR CASTING

Samples	Z	2NZ	3NZ	4NZ
Mould Preheat Temperature (°C)	220	219	213	211
Pouring Temperature (°C)	536	534	530	540
Solidus Temperature (°C)	419	419	418	419
Superheat Temperature (°C)	117	115	112	121

In order to reveal the electric resistivity of the composites, solid blocks of 30 mm x 30 mm x 6 mm was prepared. The surfaces of the blocks were then grinded and polished. The electrical resistivity of the samples was measured by electric conductivity meter (with digital display), type 979, Technofour, India.

III. RESULTS AND DISCUSSIONS

A. *Microstructural Analysis*

Fig.2 shows the optical micrographs of NZ samples at magnification of 500X. Fig. 2 (a) shows the presence of pure Zn grains with some pores. Fig. 2 (b) proeutectic Zn phase in a matrix of eutectic (pure Zn + Ni-Zn δ) in 2NZ sample. Fig. 2 (c) shows the evidence of incomplete dissolution of Ni particles in Zn and positioned preferentially at the grain boundaries in the form of a second phase of Ni-Zn δ. There is a certain possibility that the grain boundary particles are affected by compositional variation along the diameter, the indication of coring. Fig. 2 (d) shows the existence of a second phase which is also the evidence of incomplete dissolution of Ni in Zn. It is observable that the complete dissolution of Ni causes to reduce the melting temperature and therefore increases the grain size of Zn. This is due to the longer time of solidification of the proeutectic Zn grains, seen in Fig. 2 (b). Whereas the incomplete dissolution of the Ni particles is responsible for the pinning of the Zn grains at the boundaries and therefore results to the smaller grain sizes in 3NZ sample in Fig. 2 (c). A mixed effect is observed in the case of 4NZ sample in Fig. 2 (d).

B. *Vicker Microhardness Testing*

In Fig. 3 the hardness plot of Vicker's microhardness of the representative samples of Z, 2NZ, 3NZ and 4NZ is shown for general comparison. The evolution of δ-phase is the reason for the higher hardness in the NZ samples as compared to the Z sample. The pinning of the particles at the grain boundaries of Zn is seen in the microstructure 3NZ sample. Whereas lamellar structure of Zn and δ-phase is seen in 2NZ sample and spheroidal δ-phase is seen in the 4NZ sample. As the order of increase of hardness of the samples are Z<3NZ<2NZ<4NZ, we can say that the order of effects are uniphase hardening < grain boundary pinning < eutectic hardening < second phase hardening.

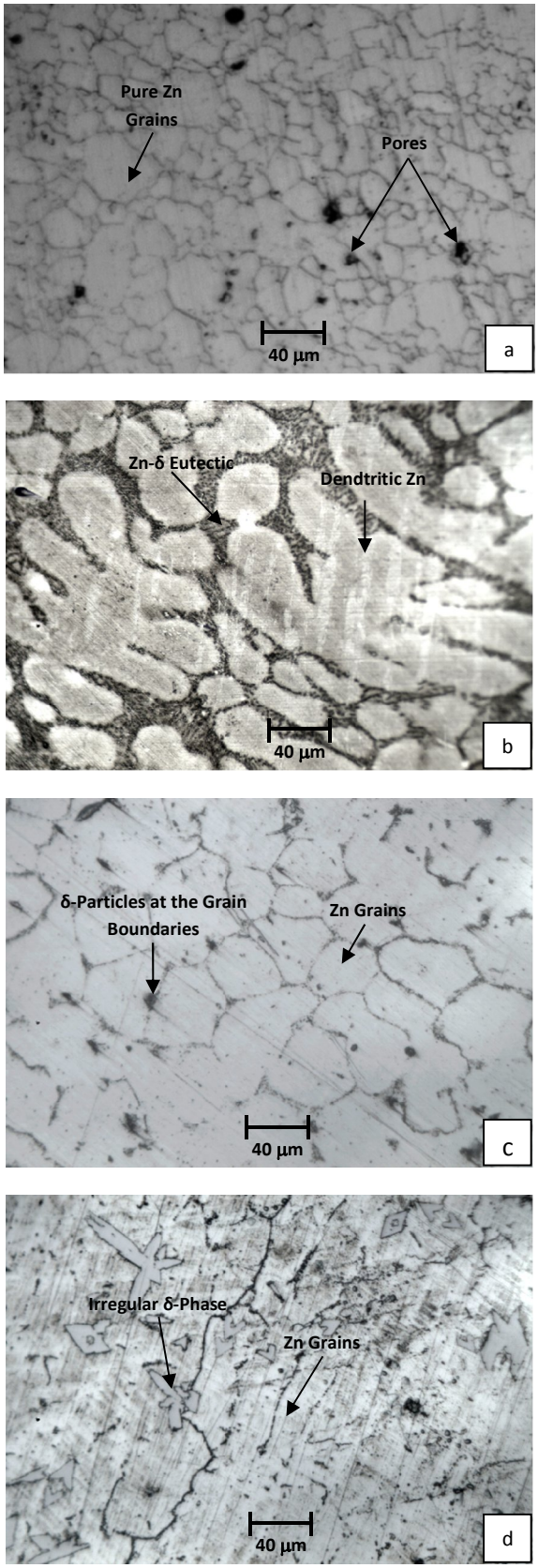


Figure 2: Showing optical micrographs of (a) Z, (b) 2NZ, (c) 3NZ and (d) 4NZ at 500X magnification.

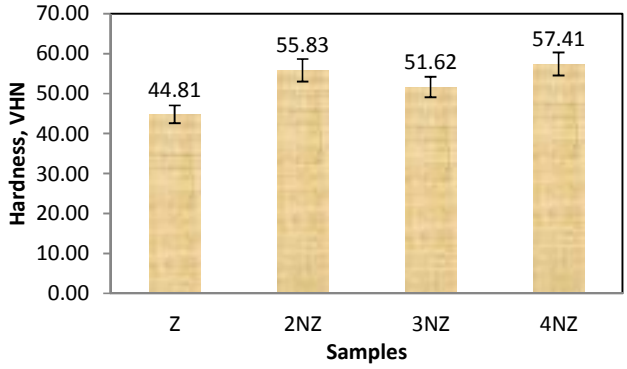


Figure 3: Showing Vicker's microhardness of the NZ samples.

C. *Electrical Resistivity Analysis*

Electrical resistivity of the Z and NZ samples are measured and listed in below (Fig. 4). It is observed that the electrical resistivity of NZ samples is slightly higher compared to the Z sample. Between the NZ samples, 4NZ has the maximum resistivity value of 6.53 µΩ.cm. 2NZ and 3NZ samples has almost equal resistivity value. The higher resistivity of the NZ samples is due the presence of Ni in Zn as δ-phase. As the amount of δ-phase is increased in the order of 3NZ < 2NZ < 4NZ, the order of increasing the electrical resistivity is also 3NZ < 2NZ < 4NZ.

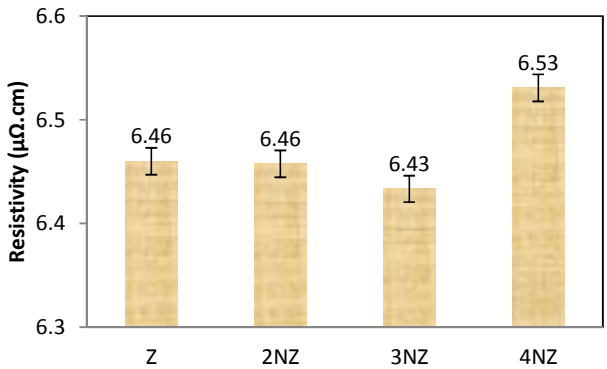


Figure 4: Showing the electrical resistivity of NZ samples.

IV. Conclusions

The results found in this research are concluded as follows:

- Presence of proeutectic Zn phase in a matrix of eutectic (pure Zn + Ni-Zn δ) is identified in the microstructure of 2NZ sample. 3NZ sample shows the evidence of incomplete dissolution of Ni particles in Zn and positioned preferentially at the grain boundaries in the form of second phase particles. Microstructure of 4NZ sample shows the existence of a second phase. As the addition of Ni on Zn causes to reduce the melting temperature, it was expected that the NZ-composites will show the larger grains compared to pure Zn. Pinning of the δ particles at the grain boundaries of Zn causes to the smaller grain size in the 3NZ sample.

- Vicker's Hardness of the NZ samples is higher than the Z sample. Among the NZ samples 3NZ sample has the lowest hardness of 51.62 VHN and 4NZ sample has the highest of 57.41 VHN.
- Between the NZ samples, 4NZ has the maximum resistivity value of 6.53 μΩ.cm. 2NZ and 3NZ samples has almost equal resistivity value 6.4 μΩ.cm. It can be concluded that the effect of addition of Ni on NZ composites are completely dependent on the microstructure.

ACKNOWLEDGMENT

The authors would like to thank the Bangladesh University of Engineering and Technology (BUET) in funding the present research. The authors are highly grateful to the Pilot Plant and Process Development Centre (PP and PDC) of Bangladesh Council for Scientific and Industrial Research (BCSIR) for providing the necessary facilities to carry out the research successfully.

REFERENCES

[1] M. McCormack, S. Jin, H.S. Chen, "New lead-free, Sn-Zn-In solder alloys", J Electron Mater, Vol. 23, pp. 687-690, 1994.

[2] K. Suganuma, "Advances in lead-free electronics soldering", Curr Opin Solid State Mater Sci, Vol. 5, pp. 55-64, 2001.

[3] A. Kroupa, D. Andersson, N. Hoo, J. Pearce, A. Watson, A. Dinsdale, S. Mucklejohn, "Current problems and possible solutions in high-temperature lead-free soldering", J Mater Engg Perform, Vol. 21, pp. 629-637, 2012.

[4] S. Yagi, T. Ichitsubo, E. Matsubara, M. Yamaguchi, H. Kimura, K. Sasamori, "Interfacial reaction of gas-atomized Sn–Zn solder containing Ni and Cu additives", J Alloys Compd, Vol. 484, pp. 185-189, 2009.

[5] W.R. Osório, L.C. Peixoto, L.R. Garcia, N.M. Noël, A.Garcia, "Microstructure and mechanical properties of Sn–Bi, Sn–Ag and Sn–Zn lead-free solder",J Alloys Compd, Vol. 572, pp.97-106, 2013.

[6] A. Debski, W. Gasior, Z. Moser, R. Major, "Enthalpy of formation of intermetallic phases from the Au–Sn system", J Alloy Compd, Vol. 491, pp. 173–177 2010.

[7] V. Chidambaram, J. Hald, J. Hattel,"Development of Au–Ge based candidate alloys as an alternative to high-lead content solders" J Alloy Compd, Vol. 490, pp. 170-179, 2010.

[8] J. M. Song, H. Y.Chuang, Z. M. Wu,"Interfacial reactions between Bi–Ag hightemperature solders and metallic substrates" J Electron Mater, Vol. 35, pp. 1041-1049, 2006.

[9] J. M. Song, H. Y. Chuang, Z. M. Wu,"Thermal and tensile properties of Bi–Ag alloys", Metall Mater Trans A, Vol. 38, pp. 1371-1375, 2007.

[10] T. Nagaoka, Y. Morisada, M. Fukusumi and T. Takemoto, "Selection of soldering temperature for ultrasonic-assisted soldering of 5056 aluminum alloy using Zn–Al system solders", J Mater Proc Tech, Vol. 211, pp. 1534–1539, 2011.

[11] S. F. Gueijman, C. E. Schvezov and A. E. Ares, "Tracking interphases in directionally solidified Zn-Al binary alloys", Mater Perform Character, Vol. 1, Issue 1, DOI: 10.1520/MPC20120017, 2012.

[12] Y. M. Kim, C. Y. Oh, H. R. Roh and Y. H. Kim, "A new Cu-Zn solder wetting layer for improved impact reliability", Electron Comp Techn Conf, IEEE 2009, pp. 1008-1013

[13] S. Kim, K. S. Kim, S. S. Kim and K. Suganuma, "Interfacial reaction and die attach properties of Zn-Sn high-temperature solders", J Electron Mater, Vol. 38, pp. 266-272, 2009.

[14] S.W. Park, T. Sugahara, K.S. Kim and K. Suganuma, "Enhanced ductility and oxidation resistance of Zn through the addition of minor elements for use in wide-gap semiconductor die-bonding materials", J Alloys Compd, Vol. 542, pp. 236–240, 2012.

[15] N. Kang, H. S. Na, S. J. Kim and C. Y. Kang, "Alloy design of Zn–Al–Cu solder for ultra-high temperatures", J Alloys Compd, Vol. 467, pp. 246–250, 2009.

[16] A. Haque, B. H. Lim, A. S. M. A. Haseeb and H. H. Masjuki, "Die attach properties of Zn–Al–Mg–Ga based high-temperature lead-free solder on Cu lead-frame", J Mater Sci: Mater Electron, Vol. 23, pp. 115–123, 2012.

[17] T. Gancarz, J. Pstrus, P. Fima, and S. Mosinska, "Thermal properties and wetting behavior of high temperature Zn-Al-In solders", J Mater Engg Perform, Vol. 21, pp.599-605, 2012.

[18] F. Guo, J. Lee, S. Choi, J. P. Lucas, T. R. Bieler andK. N. Subramanian,. "Sn–3.5Ag solder reinforced with mechanically incorporated Ni particles", J Electron Mater, Vol. 30, pp. 1073–82,2001.

[19] A. K. Gain, Y.C. Chan,W. K.C. Yung, "Effect of nano Ni additions on the structure and properties of Sn–9Zn and Sn–Zn–3Bi solders in Au/Ni/Cu ball grid array packages", Mater. Sci Engg B, Vol. 162, pp. 92–98, 2009.

[20] T. Fouzder, A. K. Gain, Y.C. Chan, A. Sharif, W. K.C. Yung, "Effect of nano Al2O3 additions on the microstructure, hardness and shear strength of eutectic Sn–9Zn solder on Au/Ni metallized Cu pads", Microelectron Reliab, Vol. 50, pp. 2051–2058, 2010.

[21] M. Ahmed, T. Fouzder, A. Sharif, A. K. Gain, Y.C. Chan, "Influence of Ag micro-particle additions on the microstructure, hardness and tensile properties of Sn–9Zn binary eutectic solder alloy", Microelectron Reliab, Vol. 50, pp. 1134–1141, 2010.

[22] J. Shen, Y. C. Liu, Y. J. Han, Y. M. Tian, H. X. Gao, "Strengthening effects of ZrO2 nanoparticles on the microstructure and microhardness of Sn–3.5Ag lead- free solder", Mater Sci Engg A, Vol. 44, pp.135–41, 2006.

Sensitivity in Polycrystalline HgI_2 X-ray Imagers for Mammography

Farzana Kabir,[1,*] Dewan Fahim Noor,[2] Zubayer Ibne Ferdous[3] and Shaikh Asif Mahmood[4]
[1]Department of Electrical and Electronic Engineering , Bangladesh University of Engineering and Technology
Dhaka-1205, Bangladesh
*orpia.1337@gmail.com

***Abstract*—The sensitivity of polycrystalline HgI_2 (poly-HgI_2) X-ray detector is determined. A generalized expression for charge carrier transport and absorption-limited sensitivity of X-ray photoconductors is derived by analytically solving the continuity equation for both holes and electrons considering the drift of electrons and holes in the presence of deep traps. Sensitivity is studied as a function of varying applied electric field across the photoconductor for both positive and negative bias under mammographic condition. The theoretical model is fitted to the published results for physical vapor deposited (PVD) poly-HgI_2 mammographic detector. The fitted values of the electron and hole ranges are $1.6896 \times 10^{(-5)}cm^2V^{-1}$ and $5.916 \times 10^{(-6)}cm^2V^{-1}$ respectively. Concentration of deep trap for electrons and holes were found out to be $2.6082 \times 10^{13}/cm^3$ and $3.3806 \times 10^{12}/cm^3$ respectively . Effect of decreasing electron and hole lifetime on sensitivity is studied for varying detector thickness for positive bias under mammmographic condition. Under these conditions, deteriorating hole lifetime has much more impact on sensitivity than deteriorating electron lifetime. Presence of an optimum detector thickness has been found at low hole lifetime**

***Index Terms*—X-Ray Sensitivity, Photoconductors, Monoenergetic Beam,Polycrystalline Mercuric Iodide .**

I. Introduction

Polycrystalline mercuric iodide (poly-HgI_2) is a promising photoconductor to be used in direct conversion flat-panel X-ray detectors. One of the main advantages of polycrystalline mercuric iodide is that, it has very low effective work function (W±∼5 eV) compared to other available photoconductors [1][2]. Low effective work function is a desirable property for a photoconductor since it results in higher sensitivity. Another big advantage is that, the effective work function does not depend on applied electric field. Low effective work function, high atomic number and the possibility of large area deposition makes poly-HgI_2 an attractive candidate for use in direct conversion flat-panel X-ray image detectors[3].The HgI_2 layer is deposited onto conductive glass plates by either physical vapor deposition (PVD) or screen printing methods to fabricate a poly-HgI_2 detector.

In this paper, we have investigated sensitivity of poly-HgI_2 X-ray imagers under mammographic condition. The carrier ranges of poly-HgI_2 deposited by screen printing method have been determined for chest radiology conditions in a previous work [4]. In this paper, carrier ranges of poly-HgI_2 deposited by PVD method have been determined for mammographic condition.

The X-ray sensitivity of a photoconductor is defined as the collected charge per unit area per unit exposure of radiation. The continuity equation for both holes and electrons is analytically solved by considering the drift of electrons and holes in the presence of deep traps under the situation of exponentially decaying distribution of electron hole pair generation across the thickness of the photoconductor. An expression is derived for determining sensitivity in terms of electron hole pair creation energy, W± , monoenergetic X-ray photon energy E, linear attenuation coefficient α, energy absorption coefficient α_{en} of the photoconductor, transport properties of the photoconductor i.e. carrier mobility μ and carrier lifetime τ, the operating conditions(electric field applied across the detector, F and detector thickness L).

II. Theoretical Model

Consider a photoconductor sandwiched between two large area parallel plate electrodes and biased with a voltage V to establish an electric field F. A monoenergetic beam of X-ray is applied. The X-ray photons are absorbed exponentially as exp $(-\alpha x)$ along the photoconductor thickness and generate an electron hole pair concentration that follows the X-ray photon absorption profile. The electrons and holes are drifted in opposite directions by the applied field and give rise to a transient X-ray photocurrent. The radiation receiving electrode can be biased positively or negatively.

To deduce the mathematical equation of the total collected charge some necessary assumptions are made- (1) The thermal equilibrium concentration of charge carriers is negligibly small (2) The diffusion of carriers is negligible compared to their drift because of high applied field across the photoconductor (3) A constant drift mobility μ and a single deep trapping time (lifetime) τ are assigned to each type of carriers (holes and electrons) since the interrupted field time-of-flight measurements indicate a single deep trapping time for both types of carriers [2][5]. (4)The field remains relatively uniform. (5) The loss of carriers by deep trapping is more significant than bulk recombination. (6) The trapped charge concentrations are very small compared to the trap center concentrations and thus trap saturation effect is negligible. (7) The photoconductor is exposed to a monoenergetic pulse of X-ray radiation that has a very short duration compared to the charge carrier transit times across the sample thickness.

The continuity equation for holes for these assumptions with positive bias applied on the radiation receiving electrode is-

$$\frac{\partial \rho(x,t)}{\partial t} + \mu_h F \frac{\partial \rho(x,t)}{\partial x} = -\frac{\rho(x,t)}{\tau_h} \text{.....(1)}$$

where $\rho(x,t)$ is the concentration of holes at location x at time t, and F is the applied electric field. According to the features of X-ray absorption and electron hole pair generation, the initial hole or electron distribution at time t=0 across the photoconductor is

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

$$\rho(x,0) = n(x,0) = \frac{\alpha_{en}\varphi_0 E}{W_{\pm}e^{(\alpha x)}} = Be^{(-\alpha x)} .....(2)$$

Where E is X-ray photon energy assuming a monoenergetic beam has been applied and

$$B = \frac{\alpha_{en}\varphi_0 E}{W_{\pm}}$$

The incident X-ray photons per cm^2 is related to X-ray exposure X by [6]

$$\varphi_0 = \frac{5.45 \times 10^{13}}{\frac{\alpha_{air}}{\rho_{air}}E} X .....(3)$$

The solution of Equation (1) is a drifting hole distribution that is given by-

$$\rho(x,t) = Be^{(\alpha\mu_h F - \frac{1}{\tau_h})t} e^{(-\alpha x)}; \mu_h F t < x < L$$

$$\rho\,(x,t) = 0; (x < \mu_h F t) or (x > L) .....(4)$$

The collected positive charge per unit area per unit exposure of radiation would be

$$Q_h = \frac{\mu_h F e}{L} \int_0^{t_h} \int_{\mu_h F t}^{L} \rho\,(x,t)\, dx dt .....(5)$$

This gives

$$Q_h = \frac{B\tau_h \mu_h F e}{\alpha L} \left\{ \left(1 - e^{-\alpha L}\right) + \frac{\left(e^{-\alpha L} - e^{-\frac{t_h}{\tau_h}}\right)}{1 - \frac{1}{\tau_h \alpha \mu_h F}} \right\} ..(6)$$

where $t_h = L/(\mu_h F)$=transit time of hole across the semiconductor and e is the charge of an electron or hole. Similarly, we can calculate the collected charge per unit area per unit exposure of radiation,Q_e due to electron drift.

Sensitivity is determined by S=Q_h+Q_e. Sensitivity can be written as-

Sensitivity,S=

$$\frac{e\varphi_0 E}{W_{\pm}} \frac{\alpha_{en}}{\alpha} \frac{\mu_h \tau_h F}{L} \left[\left(1 - e^{-\alpha L}\right) + \frac{\left(e^{-\alpha L} - e^{-\frac{t_h}{\tau_h}}\right)}{1 - \frac{1}{\tau_h \alpha \mu_h F}} \right]$$

$$+ \frac{e\varphi_0 E}{W_{\pm}} \frac{\alpha_{en}}{\alpha} \frac{\mu_e \tau_e F}{L} \left[\left(1 - e^{-\alpha L}\right) - \frac{\left(1 - e^{-\alpha L - \frac{t_e}{\tau_e}}\right)}{1 + \frac{1}{\alpha \mu_e F \tau_e}} \right] ...(7)$$

When negative bias is applied on the radiation receiving electrode, sensitivity can be written as-

Sensitivity,S=

$$\frac{e\varphi_0 E}{W_{\pm}} \frac{\alpha_{en}}{\alpha} \frac{\mu_h \tau_h F}{L} \left[\left(1 - e^{-\alpha L}\right) - \frac{\left(1 - e^{-\alpha L - \frac{t_h}{\tau_h}}\right)}{1 + \frac{1}{\alpha \mu_h F \tau_h}} \right]$$

$$+ \frac{e\varphi_0 E}{W_{\pm}} \frac{\alpha_{en}}{\alpha} \frac{\mu_e \tau_e F}{L} \left[\left(1 - e^{-\alpha L}\right) + \frac{\left(e^{-\alpha L} - e^{-\frac{t_e}{\tau_e}}\right)}{1 - \frac{1}{\tau_e \alpha \mu_e F}} \right] ...(8)$$

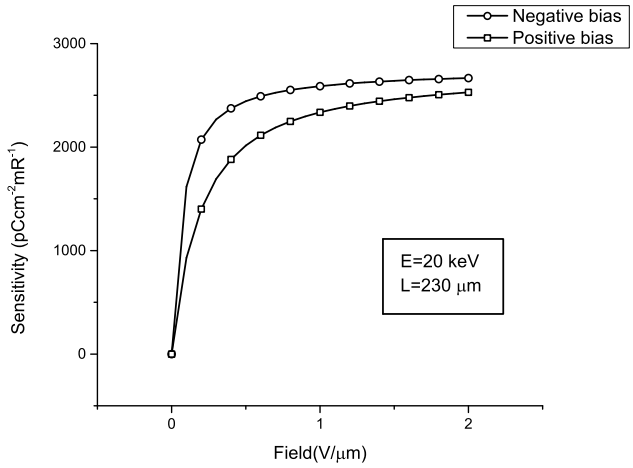


Fig. 1. Sensitivity versus applied electric field with both positive and negative bias applied on the radiation receiving electrode for a detector of thickness 230 μm

III. Result and Discussion

The sensitivity model is applied to PVD type polycrystalline HgI_2 to evaluate its performance as X-ray photoconductor. The sensitivity calculations were carried out using equation (7)and equation (8) for a monoenergetic beam of X-rays at 20 keV to very approximately represent mammographic condition.

In Fig.1, Sensitivity is examined as a function of varying applied electric field F for a monoenergetic beam of X-rays having photon energy, E=20 keV for a typical detector thickness L of 230 μm.This thickness corresponds to about 7δ (δ = 1/α =absorption depth) to absorb nearly all the incident radiation to minimize the patient exposure. It is observed that, the sensitivity increases with increasing electric field before reaching an asymptotic level of maximum collection efficiency. After reaching the maximum value, sensitivity becomes constant as the electron hole pair creation energy of polycrystalline HgI_2 is a constant and thus does not depend of the applied field. It is also observed that the sensitivity is higher when negative bias is applied. This can be explained by the following facts. Sensitivity is mainly controlled by charges having the same polarity as the bias applied at the radiation receiving electrode as they get collected at the bottom electrode. When negative bias is applied on the radiation receiving electrode, electrons control sensitivity and electron range is much higher in polycrystalline HgI_2 than hole range.

In Fig.2, The sensitivity model is fitted with experimental data for sensitivity of PVD polycrystalline HgI_2 when a monoenergetic beam of 20 KeV is applied with negative bias applied on the radiation receiving electrode. This is approximately the mammographic condition. The data have been extracted from Fig.7 of [1]. As apparent from Fig.2, there is a very good fit of the experimental data to Equation 8. From the curve fitting in Fig.2, in case of the PVD sample, the electron lifetime was found to be, $\tau_e = 1.917 \times 10^{(-7)} s$ and hole lifetime was found to be $\tau_h = 1.479 \times 10^{(-6)} s$. From the lifetime data, electron range and hole ranges for PVD poly HgI_2 are determined. Electron range has been found to be $\mu_e \tau_e = 1.6896 \times 10^{(-5)} cm^2 V^{-1}$ and hole range has been found to be $\mu_h \tau_h = 5.916 \times 10^{(-6)} cm^2 V^{-1}$.The electron and hole range of PVD poly HgI_2 have been reported to

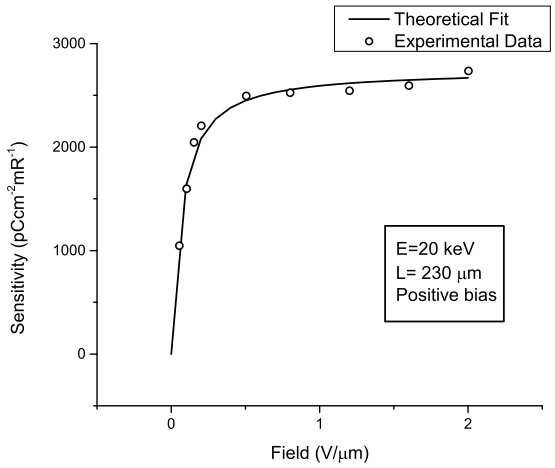


Fig. 2. Sensitivity versus applied electric field with negative bias applied on the radiation receiving electrode for a detector of thickness 230 μm. The sensitivity model is fitted with the experimental data

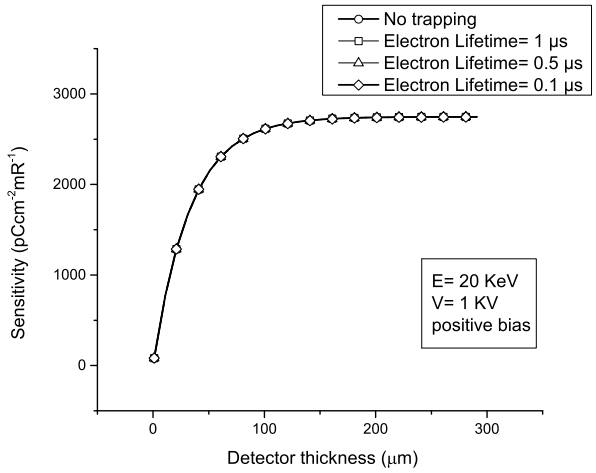


Fig. 3. Sensitivity versus detector thickness operating at an applied voltage V = 1 kV for various levels of electron trapping. Hole trapping neglected

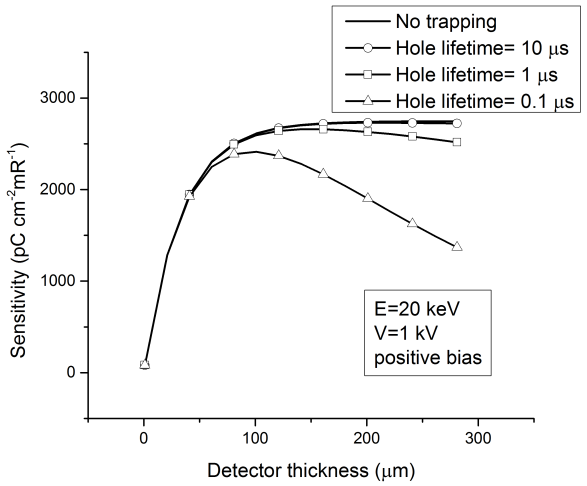


Fig. 4. Sensitivity versus detector thickness operating at an applied voltage V = 1 kV for various levels of hole trapping. Electron trapping neglected

be $\mu_e\tau_e \approx 1 \times 10^{(-5)} - 1 \times 10^{(-4)} cm^2V^{-1}$ and $\mu_h\tau_h \approx 1\times10^{(-6)} cm^2V^{-1}$ respectively [7][8]. Therefore, the electron and hole ranges are found to be within acceptable range. It is further observed that, electron range is far greater than hole range in polycrystalline HgI_2.

Carrier lifetime depends on concentration of deep trap centers. In the sensitivity model, neutral deep trap centers are assumed so that there is no bulk recombination. Deep trapping levels at room temperature is related to carrier lifetime, τ by the following equation [9] -

Carrier lifetime,

$$\tau = \frac{1}{vC_tN_t} .....(9)$$

Here v is the carrier velocity(1×10^7 cm/s). C_t is the Charge capture cross section of the trapping centers of HgI_2 ($2 \times 10^{(-14)} cm^2$) and and N_t is the concentration of deep trap centers for carriers. Concentration of deep trap centers for both electrons and holes are determined from equation (9).

Concentration of deep trap centers for electrons,

$$(N_t)_{electron} = 2.5963 \times 10^{13}/cm^3$$

Concentration of deep trap centers for holes,

$$(N_t)_{hole} = 3.3652 \times 10^{12}/cm^3$$

It is instructive to observe detector thickness dependence of sensitivity for an incident monoenergetic X-ray beam of 20 keV (mammographic condition) operating at an applied voltage of 1kV with positive bias applied on the radiation receiving electrode. Fig.3 and Fig.4 illustrate the effects of deteriorating electron and hole trapping times respectively while varying detector thickness.

It is observed that, in case of deteriorating electron lifetime, sensitivity increases with increasing detector thickness as more charge carriers are generated and collected as detector thickness increases. It is also noticed that, deteriorating electron lifetime does not have much impact on sensitivity. On the other hand, it is observed that, as discussed above, effect of deteriorating hole lifetime is more pronounced. This is also explained by the fact that holes control sensitivity when positive bias is applied on the radiation receiving electrode. The interesting case is when the hole lifetime is so poor that the sensitivity can only be optimized by reducing the detector thickness L (thereby reducing the amount of absorbed radiation) just to enable the generated holes to be collected. It is also noticed that, S versus L behaviour has a maximum that moves to smaller thicknesses as the hole lifetime gets shorter.

If negative bias is applied on the top electrodes, the shape of the Fig.3 and Fig.4 are swapped as sensitivity is controlled by electrons when negative bias is applied.

IV. Conclusions

A model for determining sensitivity of photoconductors has been developed. The model has been applied to polycrystalline mercuric iodide detectors. Sensitivity has been observed under varying applied electric field. The proposed physics based theoretical model shows a very good agreement with the experimental results. The carrier ranges of PVD polycrystalline mercuric iodide have been determined by curve fitting under mammographic condition. Electron range has been found to be $\mu_e\tau_e = 1.6896 \times 10^{-5} cm^2V^{-1}$ and hole range has been found to be $\mu_h\tau_h = 5.916 \times 10^{-6} cm^2V^{-1}$. The electron

and hole range of PVD poly HgI_2 have been reported to be $\mu_e\tau_e \approx 1 \times 10^{(-5)} - 10^{(-4)} cm^2V^{-1}$ and $\mu_h\tau_h \approx 1 \times 10^{(-6)} cm^2V^{-1}$ respectively previously [7][8]. So, the carrier ranges determined from the sensitivity model are in very good agreement with the reported carrier ranges for poly HgI_2. The trap center concentrations for both electron and holes have been determined. Concentration of deep trap centers for electrons and holes were found out to be $2.5963 \times 10^{13}/cm^3$ and $3.3652 \times 10^{12}/cm^3$ respectively. Effect of deteriorating electron and hole lifetime on sensitivity is observed while varying detector thickness. Deteriorating hole lifetime had much more impact on sensitivity than deteriorating electron lifetime for positive bias. Presence of an optimum detector thickness for maximizing sensitivity has been observed at low hole lifetime.

ACKNOWLEDGMENT

We are grateful to the Robert Noise Simulation Lab of Department of Electrical and Electronic Engineering , Bangladesh University of Engineering and Technology for the contribution to our effort.

REFERENCES

[1] Z. Su. et al., "Systematic investigation of the signal properties of polycrystalline HgI_2 detectors under mammographic, radiographic, fluoroscopic and radiotherapy irradiation conditions,"*Phys. Med. Biol.*, vol. 50, pp. 2907-28, 2005.

[2] R. A. Street et al., "Comparison of PbI_2 and HgI_2 for direct detection active matrix X-ray image sensors," *Jounrnal of Applied Physics*, vol. 91, pp. 3345-55, 2002.

[3] H. Du. et al., "Investigation of the signal behavior at diagnostic energies of prototype, direct detection,active matrix, flat-panel imagers incorporating polycrystalline HgI_2," *Physics in Medicine and Biology*, vol. 53, no. 5, p. 1325-1351, March 2008.

[4] M. Zahangir Kabir and S. O. Kasap, "Charge collection and absorption-limited sensitivity of X-ray photoconductors: Applications to a-Se and HgI_2," *photoconductors: Applications to a-Se and HgI_2*, vol. 80, no. 9, p. 1966, March 2002.

[5] S. O. Kasap, "X-ray sensitivity of photoconductors: application to stabilized a-Se," *J. Phys. D: Applied Physics*, vol. 33, p. 2854, 2000.

[6] J. M. Boone, *Handbook of Medical Imaging*, J Beutel, H L Kundel, and R L Van Metter, Eds.

[7] A. Zuck, M. Schieber, . Khakhan, and Z Burshtein, "Near single-crystal electrical properties of polycrystalline HgI_2 produced by physical vapor deposition," *IEEE Trans. Nucl. Sci.*, vol. 50, pp. 991-997, 2003.

[8] R. C. Whited and L. van den Berg, "Native defect compensation in HgI_2 crystals," *IEEE trans. Nuc. Sci.*, vol. NS-24, p. 165, 1977.

[9] G. Xu, W .Jie, G. Li, and G. zha, "Study of trapping levels in a-HgI_2 crystals by UV and isothermal current measurements," *Nuclear Instruments and Methods in Physics Research A*, p. 278, 2010.

Barrier Height Control of Metal/n-$In_{0.2}Ga_{0.8}N$ with an Ultrathin Interfacial Layer

Sudipta Paul,[1,*] Md. Soyaeb Hasan,[1] and Md. Rafiqul Islam[1]

[1]Electrical and Electronic Engineering, Khulna University of Engineering and Technology
Khulna-9203, Bangladesh
[*]sudipto.paul011@gmail.com

Abstract—**In this paper the barrier height control of n-type InGaN has been demonstrated by a simple physics based unified model. Metal induced gap states (MIGS) model for metal-semiconductor contact has been extended to metal-insulator-semiconductor (MIS) contact for unpinning of Fermi level and low contact resistivity formation. Metal/n-$In_{0.2}Ga_{0.8}N$ contact is strongly pinned with a pinning factor value of 0.28. This work has improved the pinning factor from 0.28 to almost closer to 1 by inserting an ultrathin GeO_2 (2 nm) insulator layer and by varying this layer thickness, density of interface electronic states has been controlled which modulated the barrier height. Finally an excellent ohmic contact has been formed by using Cu as the metallic contact with contact resistivity as low as 1.4×10^{-8} Ω-cm^2 at 600K.**

Index Terms—**Pinning factor, MIGS Model, MIS contact, Barrier Height Control, Ohmic contact.**

I. Introduction

InGaN has got a great attention by the researcher for its structural properties. It has already involved in some applications such as high temperature and high power electronics [1], [2] light emitting diodes and detectors operating in the blue and ultra-violet wavelength range [3], [4]. It has a band gap ranging from 0.7 eV to 3.4 eV which covers the entire visible spectrum. That's why they are of concern for the future optoelectronics applications. To make a device stable and efficient, some outlines such as contact issues, surface preparation are needed to be considered carefully. Schottky Barrier-FET (SB-FET) implementation using III-V semiconductors including both binary and turnery materials such as GaAs, InAs, GaN, AlGaN, InGaAs can be challenging because Fermi level at metal/III-V interfaces are pinned [5]-[9]. Wave-function of electron in the metal penetrates the semiconductor band gap. This penetration forms electron states. This states in the band gap results in Fermi level (FL) pinning [7]. So, for implementing III-V materials in SB-FET, barrier height modulation is needed. S. Hara *et al.* first proposed a systematical method for controlling Barrier Height (BH) at metal/semiconductor interface by controlling surface electronic states [10]. Controlling of interface electronic state means to control of FL pinning factor. For the modulation of barrier height at the metal-semiconductor interfaces, two methods have been proposed [5]. One is dopant segregation which is the introduction of appropriate dopants at the semiconductor surface and it has been investigated both theoretically and experimentally [11]. Another is the insertion of an ultrathin insulator between metal and semiconductor [5], [7]-[9].

Monch through his Metal-Induced-Gap-States (MIGS) model described the FL pinning factor for Metal-Semiconductor (M-S) contact [12]. In our work this model has been extended to Metal-Insulator-Semiconductor (MIS) contact by a simple physics based approach. By insertion of a thin insulator layer between metal and semiconductor and varying its thickness, pinning factor can be varied. A thin insulator is capable of diminishing FL pinning caused by MIGS by subduing wave function of metal that penetrates into semiconductor band gap [5]. MIS technology to mitigate FL pinning by inserting interfacial layer (IL) has been demonstrated with Si [6], Ge [6], [13], GaAs [7], [9] and InGaAs [5], [6]. Also for some III-N materials such as GaN and AlGaN, theoretical band offsets using the MIGS Schottky pinning factor has been determined [9], [14] and there are a very few experimental works too [15], [16]. In this paper we first demonstrated the Barrier Height (BH) control mathematically for the InGaN using MIGS model. GeO_2 has been introduced as the insulator layer for the MIS contact. Simulations have been performed using MATLAB software. The pinning factor of the InGaN is 0.28. We have improved this value to closer to 1 and by varying this pinning factor the variation of the barrier height for the same metallic contact has been demonstrated. This work has been concluded with formation of an excellent ohmic contact. An ultralow barrier height of 0.07 eV has been found for Cu as the metallic contact. The contact resistivity in the 10^{-8} ohm-cm^2 range at various temperature (300-600K) has been investigated which is the evidence of an outstanding ohmic contact. The total working procedure of BH control and ohmic contact formation demonstrated in this paper has brought out in a flow diagram as shown in Fig. 1.

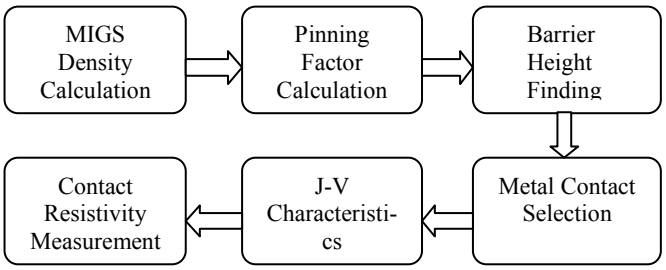


Fig. 1 Block Diagram of Barrier Height Control and ohmic contact formation

II. MIGS Model For Barrier Height Control

A. MIGS Density Calculation

Physical explanation of the FL pinning factor has been nicely described by Monch through his MIGS model for M-S contact [12]. In this paper this model has been improved in case of MIS contact. We have proposed that metal-interfacial layer (M-IL) has MIGS and its density is expressed as

$$D_{M0} = \frac{2}{\pi a^2 E_g} \quad (1)$$

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

Here E_g and a are bandgap and lattice constant of the insulator respectively. The reduction of this MIGS density into the insulator is exponential in nature and is expressed by the equation [13]

$$D_M = D_{M0} e^{-\frac{t}{\delta_i}} \quad (2)$$

In that equation t is the width of the insulator and δ_i is the decay length of the MIGS density through the IL. This decay length is given by [13]

$$\delta_i = \frac{h^2}{2\pi m_0 a E_g} \quad (3)$$

Its decay in the semiconductor(δ_S) can be found from equation (3) by replacing the value of lattice constant (a) and band gap (E_g) for semiconductor.

B. Pinning Factor Calculation

When metal is deposited on the semiconductor, there will be a change of Fermi level. But in some semiconductors the surface states pin the FL. Surface states compel the FL to be there. Surface states are due to electron wave function of metal penetrating the semiconductor bandgap and causes FL pinning [5], [7]. The BH alteration is very little although by varying the metal work function to a wide range. The value of the FL pinning factor, S lies between 0 and 1. S=0 means FL is completely pinned and S=1 means FL is completely unpinned.

Monch found an empirical formula to the experimentally observed value of pinning factor for the M-S contact for MIGS model [9]:

$$S = \left[1 + 0.1(\varepsilon_\infty - 1)\right]^{-1} \quad (4)$$

But the pinning factor equation for a MIS contact considering intrinsic pinning factor of insulator layer is [13]

$$S = \left[1 + \frac{\varepsilon_i (1 - S_i) e^{-\frac{t}{\delta_i}} \left(\delta_s \varepsilon_i + t\varepsilon_s\right)}{S_i \delta_i (\varepsilon_i + \varepsilon_s)}\right]^{-1} \quad (5)$$

In this equation S_i is the intrinsic pinning factor of the insulator GeO_2 itself. It is given by [13]

$$S_i = \left(1 + q^2 \frac{D_{M0}\delta_i}{\varepsilon_0 \varepsilon_i}\right)^{-1} \quad (6)$$

C. Barrier Height Measurement

Due to difference in charge neutrality level (CNL) value of insulator and semiconductor, a dipole is risen which must not be ignored in case of calculating barrier height. If ϕ_S, χ_S and ψ_S are the work-function, electron affinity and total equilibrium band bending of the semiconductor respectively then the BH of an n type semiconductor can be written as [6]

$$\phi_{Bn} = \phi_S - \chi_S - \psi_S \quad (7)$$

The equilibrium band bending of semiconductor ψ_S depends upon insulator-semiconductor (I-S) dipole as [13]

$$\psi_S = -\frac{C_I}{C_I + C_S}\phi_{MS} - \frac{C_{il}}{C_{il} + C_{is}}\Delta_{IS} \quad (8)$$

Again, I-S dipole voltage Δ_{IS} which depends upon the CNL difference between semiconductor and insulator and pinning factor of I-S interfaces is expressed as [6].

$$\Delta_{IS} = (1 - S_{IS})(\phi_{CNLS} - \phi_{CNLI}) \quad (9)$$

Here S_{IS} is the pinning factor of I-S interface. ϕ_{MS} is the work function difference between metal and semiconductor. C_I and C_S are the insulator and semiconductor surface capacitance densities respectively. C_{il} and C_{iS} are the IL-semiconductor dipole capacitance densities for the IL and the semiconductor, respectively [6].

$$C_I = \frac{\varepsilon_I}{t}, \quad C_{il} = \frac{\varepsilon_{\infty l}}{d_{il}}, \quad C_{is} = \frac{\varepsilon_{\infty s}}{d_{is}} \quad (10)$$

Here d_{il} (d_{is})is the dipole layer thickness and $\varepsilon_{\infty l}(\varepsilon_{\infty s})$ is the high frequency dielectric constant. Now for an n-type semiconductor surface capacitance depends on band bending as [6]

$$C_s(\psi_s) = q\sqrt{\frac{\varepsilon_s N_D}{2K_B T}} \left| \frac{e^{\frac{q\psi_S}{K_B T}} - 1 - \left(\frac{n_i}{N_D}\right)^2 e^{\left(-\frac{q\psi_S}{K_B T}\right)}}{\sqrt{e^{\frac{q\psi_S}{K_B T}} - \frac{q\psi_S}{K_B T} - 1 + \left(\frac{n_i}{N_D}\right)^2 \left[e^{\frac{-q\psi_S}{K_B T}} - 1\right]}} \right| \quad (11)$$

Now combining equation (7) & (8) and then differentiating ϕ_{Bn} with respect to ϕ_M, the mathematical relation of pinning factor and semiconductor surface capacitance densities can be obtained.

$$S = \frac{d\phi_{Bn}}{d\phi_M} = \frac{C_I}{C_I + C_S(\psi_S)} \quad (12)$$

III. Ohmic Contact Formation

A contact is whether ohmic or rectifying determined by the contact resistivity value. An ohmic contact has a very low contact resistivity. The contact resistivity is determined from the current density equation. It has linear current-voltage characteristics. This type of contact formation is achievable by keeping the barrier height as low as possible and doping the semiconductor heavily. Current density equation for a MIS contact is [17]

$$J = A^* T^2 \exp(-\sqrt{\zeta}t) \exp\left(-\frac{q\phi_{Bn}}{\eta KT}\right)\left[\exp\left(\frac{qV}{\eta KT}\right) - 1\right] \quad (13)$$

The term $\exp(-\sqrt{\zeta}t)$ is known as the tunneling probability where ζ(in eV) is the effective barrier height and t (in A) is the thickness of the insulator layer. A^* in this equation is the Richardson constant and η is the ideality factor. Contact resistivity is related to the current density as

$$\rho_C = \left(\partial J / \partial V\right)^{-1} \quad (14)$$

IV. Result and Discussions

Unpinning of FL is a complex effect. It combines not only the effect of fixed surface charge but also interface dipoles

TABLE 1
PARAMETERS FOR $In_{0.2}Ga_{0.8}N$ AND GeO_2

Sl. No	Parameters	Values	
		$In_{0.2}Ga_{0.8}N$	GeO_2
01	Bandgap (eV)	2.63 [18]	4.3[13]
02	Work function (eV)	5 [19]	--
03	Electron affinity (eV)	4.56[20]	--
04	Lattice constant (A)	3.26 [18]	4.402[13]
05	Dielectric constant	11.146[18]	5.9[13]
06	High frequency dielectric constant	6.056[21]	1.8[13]
07	Charge neutrality level (eV)	2.48[22]	3.2[13]

[5], [7]. Pinning factor of $In_{0.2}Ga_{0.8}N$ calculated from equation (4) without the presence of GeO_2 is 0.28. So, FL is strongly pinned between metal and semiconductor. In this work the pinning factor of $In_{0.2}Ga_{0.8}N$ has been investigated by insertion of GeO_2 as the IL between metal and semiconductor.

Parameters used for $In_{0.2}Ga_{0.8}N$ and GeO_2 have been listed in Table. 1. Most of the parameters of $In_{0.2}Ga_{0.8}N$ have been calculated by using Vagard law. Strong FL pinning is due to the interface state density. This surface state forms by the penetration of electron wave function of metal into the semiconductor bandgap. This pins the Fermi level. Surface state density has reduced by inserting GeO_2 insulator layer between metal and semiconductor with suitable thickness. Presence of GeO_2 has mitigated electron wave-function, thus avoiding FL pinning.

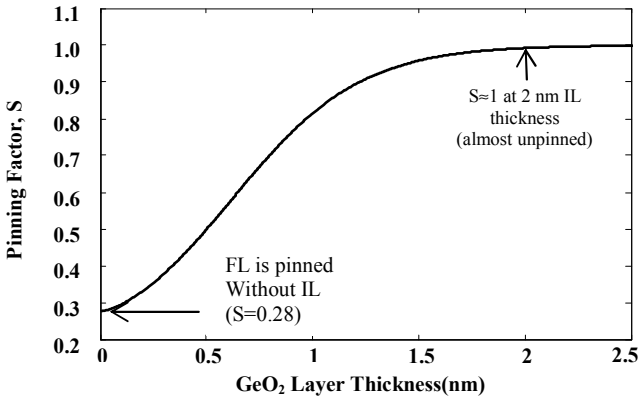


Fig. 2 FL pinning factor of $In_{0.2}Ga_{0.8}N$ as a function of GeO_2 thickness

Fermi level pinning factor variation of $In_{0.2}Ga_{0.8}N$ with the variation of GeO_2 layer has been plotted in Fig. 2. With the presence of IL, electron wave function of metal becomes attenuated. So, pinning factor is rising. For complete unpinning of the FL(S=1), the thickness required is almost 3 nm. But the noticing thing is that the more the thickness of the IL, the more the tunneling resistance. So, here a tradeoff is needed between pinning factor and thickness of insulator. Fig. 2 indicates that a 2 nm thickness of GeO_2 can make the value of S almost closer to 1. So, FL is almost unpinned (S≈1) with 2 nm GeO_2 layer.

Pinning factor shift also shifts the BH for the same metallic contact. Pinning factor varies with the variation of the insulator layer thickness. It results a barrier height variation for the same metal contact. We have investigated this variation with some metals having work functions ranging from 4.5 eV to 6 eV. Fig. 3 shows the barrier height modulation for the same metallic contact with various S values.

Cu (4.65) [23], Ag (4.73) [24], Au (5.1) [23], Ni (5.15) [8] and Pt (5.65) [23] have been used to test the BH variation. From Table.2 we can see that $In_{0.2}Ga_{0.8}N$ has a barrier height of 0.33 for Cu without having any IL between them. But with the presence of an ultrathin GeO_2 of 2 nm, barrier height shifts from 0.33 to 0.07 eV. Similarly, barrier height shifting happens for rest of the metals. Barrier height with Cu and Ag decreases with the increase of pinning factor because their work-function is less than that of $In_{0.2}Ga_{0.8}N$. On the other hand Ni, Au and Pt contacts show larger BH because of having work-function more than $In_{0.2}Ga_{0.8}N$. So, with the increase of pinning factor barrier height shows more dependence on metal work-function.

Insulator (C_I) and semiconductor (C_S) surface capacitance densities vary with the IL thickness.C_I is inversely proportional

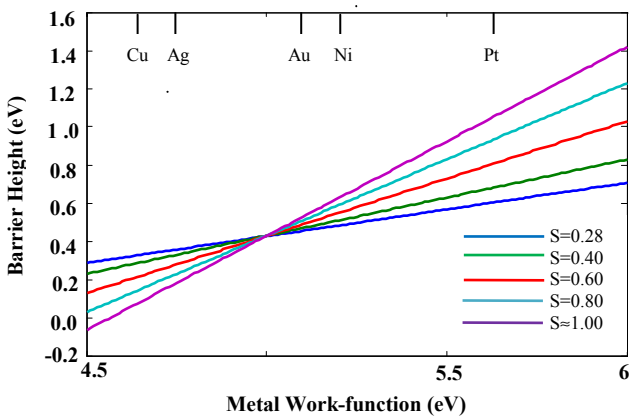


Fig. 3 Barrier Height variation with metal work-function for different pinning factor

to GeO_2 thickness. On the other hand, the initial FL position of the InGaN was near the conduction band and it is shifted from near the conduction band to near the mid-gap. BH shift is due to dipole formation between GeO_2 and InGaN (Δ_{IS}) and the insulator (C_I) and semiconductor (C_S) surface capacitance densities. Because of this dipole, the atomic structure of the InGaN changes. GeO_2 creates a polar I-S interface, thus inducing a dipole which leads to a barrier shift. This magnitude of the dipole is independent of the metal. It depends on the charge neutrality level (CNL) misalignment between insulator and semiconductor.

TABLE 2
CALCULATED BARRIER HEIGHT FOR DIFFERENT METALS

Pinnig factor	ϕ_{Bn} for Cu(4.65 ev)	ϕ_{Bn} for Ag(4.73 ev)	ϕ_{Bn} for Au(5.1 ev)	ϕ_{Bn} for Ni(5.15 ev)	ϕ_{Bn} for Pt(5.65 ev)
S=0.28	0.33	0.36	0.44	0.47	0.60
S=0.40	0.28	0.32	0.47	0.50	0.69
S=0.60	0.21	0.26	0.49	0.53	0.82
S=0.80	0.15	0.21	0.51	0.56	0.95
S≈1.00	0.07	0.16	0.54	0.60	1.08

Semiconductor surface capacitance changes with the variation of the pinning factor as shown by equation (12). The pinning factor is closely related to IL thickness and it increases with the increase of the thickness. So that's why the surface capacitance densities also vary with the variation of the GeO_2 layer thickness.

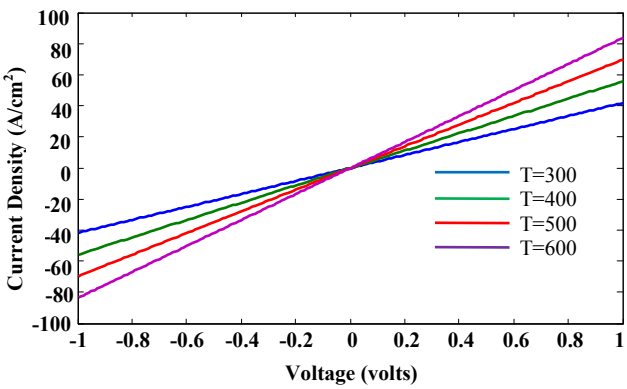


Fig. 4 Temperature dependent J-V characteristics for Cu/GeO_2 /$In_{0.2}Ga_{0.8}N$

In our investigation Cu formed an ultralow barrier height of 0.07 eV with 2 nm GeO_2 between Cu and InGaN. So, an excellent ohmic contact can be formed using Cu. Fig. 4 represents the Current density-voltage characteristics of

Cu/GeO_2 (2nm)/$In_{0.2}Ga_{0.8}N$ for different temperature. Doping concentration (N_d) has been taken as 10^{19} cm^{-3}.

Current density-Voltage characteristics formed straight lines which is a clear indication of ohmic contact. So, tunnelling resistance offered by IL didn't create strong impact on contact resistivity which has been plotted as a function of temperature for Cu/GeO_2 (2nm)/$In_{0.2}Ga_{0.8}N$ at Fig. 5. So, considering the effect of tunneling resistance, the contact resistivity has been found as low as 1.4×10^{-8} Ω-cm^2 at 600K which is an excellent ohmic contact.

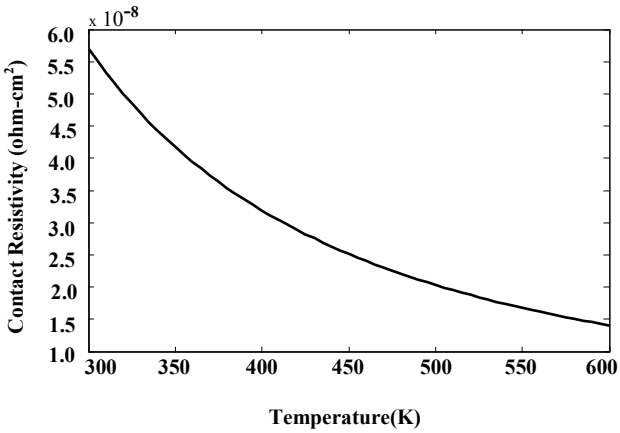


Fig. 5 Contact resistivity of Cu/GeO_2(2nm)/$In_{0.2}Ga_{0.8}N$ at various temperatures

V. CONCLUSION

In summary, MIGS model to change the FL pinning position and BH modulation at metal/InGaN interfaces has examined by insertion of ultrathin GeO_2(2 nm) interlayer. This work has set out an excellent framework for MIGS-based FL pinning in MIS contacts in case of InGaN. FL depinning is a complex effect which combines both the effect of fixed surface charge and interface dipole. Insertion of IL attenuates the electron wave-function of metal which prevents FL pinning and improves the pinning factor. BH modulation also activates due to the creation of an electric dipole at the GeO_2/ $In_{0.2}Ga_{0.8}N$ interface which introduces the barrier shift. Its alteration is due to the variation of insulator and semiconductor surface capacitance densities with IL thickness. This works concluded with forming a suitable ohmic contact for n-InGaN. Cu formed an ultralow barrier height of 0.07 eV with pinning factor closer to 1. With a moderate doping concentration and considering the tunnelling resistance offered by IL, Cu/GeO_2 (2nm)/$In_{0.2}Ga_{0.8}N$ establishes an excellent ohmic contact with contact resistivity 1.4×10^{-8} ohm-cm^2. However, this ultralow contact resistivity value put the MIS contact in compatibility with incorporation with conventional CMOS for source/drain applications.

REFERENCES

[1] T. S. Kim, J. Y. Park, T. V. Cuong, H. G. Kim, H. J. Lee, E-K. Suh and C-H. Hong, "Effect of Annealing on InGaN/GaN Multiple Quantum Wells," *Journal of the Korean Physical Society*, Vol. 47, pp. 871-875, Nov. 2005.

[2] T. Makimoto, Y. Yamauchi and K. Kumakura, "High-power characteristics of GaN/InGaN double heterojunction bipolar transistors," *Applied Physics Letters*, vol. 84, pp. 1964-1966, Mar 2004.

[3] K. Johnson, V. Bousquet, S.E. Hooper, M. Kauer, C. Zellweger, J. Heffernan, "High-power InGaN light emitting diodes grown by molecular beam epitaxy," *Electronics Letters*, vol. 40, pp. 1299-1300, Sept. 2004.

[4] Jun Ohsawa, Takahiro Kozawa, Osamu Ishiguro and Hiroshi Itoh, "Selective Detection of Blue and Ultraviolet Light by an InGaN/GaN Schottky Barrier Photodiode" *Japanese Journal of Applied Physics,* vol. 45, pp. 24-28, May. 2006.

[5] Runsheng Wang, Xu Min, Peide D. Ye, Ru Huang, "Schottky-barrier height modulation of metal/In0.53Ga0.47As interfaces by insertion of atomic-layer deposited ultrathin Al2O3," *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures*, vol. 29, pp. 041206, 041206-4, Jul. 2011.

[6] A. Agrawal, N. Shukla, K. Ahmed, and S. Datta, "A unified model for insulator selection to form ultra-low resistivity metal-insulator-semiconductor contacts to n-Si, n-Ge, and n-InGaAs," *Applied Physics Letters,* vol. 101, pp. 042108, 042108-4, Jul. 2012.

[7] Crystal Kenny, "Exploiting non-linear Arrhenius dependence of diode IV curves to determine schottky barrier band diagrams," PhD thesis, Stanford University, USA, Nov. 2012.

[8] M. H. Kim, S. N. Lee, C. Huh, S. Y. Park, J. Y. Han, J. M. Seo, and S. J. Park, "Interfacial reaction and Fermi level movement induced by sequentially deposited metal on GaN: Au/Ni/GaN", *Phys. Rev. B*, vol. 61, pp. 10966–10971, Apr. 2000.

[9] J. Robertson, B. Falabretti, "Band offsets of high K gate oxides on III-V semiconductors," *Journal of Applied Physics*, vol. 100, pp. 014111, 014111-8, Jul. 2006.

[10] S. Hara, T. Teraji, H. Okushi, and K. Kajimura, "Control of Schottky and Ohmic interfaces by unpinning Fermi level," *Applied Surface science,* vol. 117-118, pp. 394-399, Jun. 1997.

[11] M. Zhang, J. Knoch, Q. T. Zhao, S. Lenk, U. Breuer and S. Mantl, "Schottky barrier height modulation using dopant segregation in Schottky-barrier SOI-MOSFETs," *Solid-State Device Research Conference ESSDERC 2005.* pp. 457-460, 12-16 Sept. 2005.

[12] W Monch, "On the physics of metal-semiconductor interface" *Reports on Progress in Physics*, vol. 53, pp. 221-278, Mar. 1990.

[13] Shashank Gupta, Prashanth Paramahans Manik, Ravi Kesh Mishra, Aneesh Nainani, Mathew C. Abraham, Saurabh Lodha, "Contact resistivity reduction through interfacial layer doping in metal-interfacial layer-semiconductor contacts," *Journal of Applied Physics* , vol.113, pp. 234505,234505-7, Jun. 2013.

[14] G. Heidelberger, "Investigation of 'high-K' materials as alternative dielectrics for AlGaN/GaN based metal-insulator-semiconductor heterostructure field effect transistors (MISHFET)," PhD Dissertation, Rheinisch-Westf€alischen Technischen Hochschule Aachen, Apr. 2009.

[15] Ya-Lan Chiou; Lee, Ching-Ting, "Band Alignment and Performance Improvement Mechanisms of Chlorine-Treated ZnO-Gate AlGaN/GaN Metal–Oxide–Semiconductor High-Electron Mobility Transistors," *Electron Devices, IEEE Transactions on*, vol.58 pp. 3869- 3875, Nov. 2011.

[16] Y. L. Chiou, C. S. Lee and C. T. Lee, "AlGaN/GaN metal-oxide-semiconductor high electron mobility transistor with ZnO gate layer and $(NH_4)_2S_x$ surface treatment" *Applied Physics Letters*, vol.97, pp. 032107-3, Jul. 2010.

[17] S. M. Sze, Kwok K. Ng, *Physics of Semiconductor Devices*, 3rd Ed, John Wiley & Sons, Inc., Hoboken, New Jersey, 2007, pp. 169,170.

[18] W. J. Aziz, K. Ibrahim, "Study of design the (In0.04Ga0.96N/SiC) solar cells," *Optoelectronics and advanced materials – Rapid Communication,* vol. 5, pp. 12-15, Jan. 2011.

[19] P. Mahala, S. K. Behura, A. Ray, C. Dhanavantri, O. Jani, "Effect of indium concentration on metal/$In_XGa_{1-X}N$ Schottky junction solar cell under low level injection," *International Conference on Advances in Energy Research, ICAER 2011,* 9-11 Dec. 2011.

[20] M. Mehta, O. Jani, C. Honsberg, B. Jampana, I. Ferguson, A. Doolittle, "Modifying PC1D to model spontaneous and piezoelectric polarization in III-V nitride solar cells," *Proceedings of the 2zro European Photovoltaic Science and Engineering Conference, Milan, Italy*, pp. 409-413, September 3-7, 2007.

[21] Hadis Morkoc, *Nitride Semi-conductors and Devices,* R. M. Osgood Jr. Ed. Berlin, Germany: Springer-Verlag, 1999. pp. 246.

[22] V.N. Brudnyi, A.V. Kosobutsky, N.G. Kolin, "The charge neutrality level and the Fermi level pinning in A^3N (BN, AlN, GaN, InN) nitrides," *Russian Physics Journal,* vol. 51, pp. 1270-1278, Dec. 2008.

[23] A. A. Fursina, R. G. S. Sofin, I. V. Shvets and D. Natelson, "Interfacial transport properties between a strongly correlated transition metal oxide and a metal: Contact resistance in Fe_3O_4/M (M=Cu/Au/Pt) nanostructures,"*Phys.Rev.B*, vol. 82, pp. 245112,245112-5, Dec. 2010.

[24] A. L. Falk, F. H. L. Koppens, C. L. Yu, K. Kang, N. d. L. Snapp, A. V. Akimov, M.-H. Jo, M. D. Lukin, and H. Park, "Near field electrical detection of optical plasmons and single-plasmon sources," *Nature Physics,* vol. 5, pp. 475-479, May. 2009.

Low Activation Energy of Mg-doped $In_xGa_{1-x}N$ (x~0.4) and Self-compensation Modelling

Md. Nazmul Hasan, [1,*] Md. Soyaeb Hasan[1] and Md. Rafiqul Islam[1]

[1]Dept. of Electrical and Electronic Engineering, Khulna University of Engineering & Technology (KUET)
Khulna-9203, Bangladesh
[*]nh.shawon@gmail.com

Abstract— **p-InGaN epitaxial layer is crucially important for advanced electronic and optoelectronic devices. The activation energy (E_A) of Mg-acceptor, universally accepted p-type dopant, and hole concentration (p) of Mg-doped $In_xGa_{1-x}N$ alloys (x~0.4) have been investigated herein. The E_A has been calculated using an equation almost fitted with experimental data available in literatures in which E_A decreases with the increase in In content (x) in $In_xGa_{1-x}N$. The observed E_A in Mg doped $In_{0.4}Ga_{0.6}N$ alloys is about 41 meV which is few times smaller than Mg doped GaN, widely used active p-layer from III-nitride. The increased carrier concentration (p) due to increasing In content is near about 9.7×10^{18} cm^{-3} for x~0.4. The hole concentration starts to decrease at around the Mg concentration of 4.5×10^{19} cm^{-3}. These results indicate that self-compensation occurs in Mg-doped InGaN at higher-doping levels.**

Index Terms—**Mg-doped $In_xGa_{1-x}N$, Activation energy, Hole concentration, Self-compensation model.**

I. Introduction

III-nitride semiconductors and their alloys are extensively being explored for the assessment of advanced practical devices. Amongst them, InGaN is a direct bandgap semiconductor alloy having bandgap tunability from NIR (E_g^{InN}=0.65 eV) to UV region (E_g^{GaN}=3.45 eV) [1]. This attributes make this alloy apposite for advanced single junction and multi-junction solar cells, LEDs and LASER [2]. It is a highly potential alloy for High electron mobility transistor (HEMTs), Heterojunction Bipolar Transistors (HBTs) and some other advanced electronic devices [3], [4]. Inspite of this, still now InGaN is not available in device level due to the complexity of fabricating p-InGaN [5]. Among III-V semiconductors, generally Mg-doped GaN epilayers are used as p-type active layers for device purposes. The factors limiting p-GaN is the higher activation energy, E_A of Mg. Due to higher E_A of Mg in p-GaN, the self-compensation (SC) of Mg atoms increases which results in low conductivity in p-GaN [6]. Recently, it has been reported that p-InGaN can replace the p-GaN layer [7]. The high hole concentration (p) in Mg-doped InGaN/GaN super-lattice with optical and electrical properties of p-InGaN layer has been reported by hall measurement [8].

With decrease in bandgap, E_A decreases [9]. Hence Mg-doped $In_xGa_{1-x}N$ (x~0.4) is expected to have higher hole concentration than Mg-doped GaN. Besides low etching damage and low temperature requirement to synthesis p-InGaN than p-GaN [10] are other superior characteristics for device application. After the synthesis of p-InGaN in 1995 [11], a number of literatures have been available on the properties of p-InGaN [12-16]. This paper addresses the investigation of hole concentration (p) as function of In content (x), in Mg doped $In_xGa_{1-x}N$ alloys, for x up to 0.4. The E_A of Mg has been reported mathematically with a fitting equation and compared with practical aspects. It also includes the limited conductivity in Mg-doped $In_xGa_{1-x}N$ is due to donor impurity background compensation, namely self-compensation by intrinsic lattice defects such as vacancies or interstitial host atoms.

II. Investigation Of Low Activation Energy

A. Activation Energy of Mg in p-InGaN

E_A of the dopants in an epitaxial layer is a key parameter to get a doped layer. Fig. 1 shows the E_A of Mg-acceptor in Mg doped p-$In_xGa_{1-x}N$ alloys as a function of In content (x). It has been found that E_A decreases with the increase in In content (x). Lower value of E_A is the main reason for higher value of p of Mg doped $In_xGa_{1-x}N$ alloys with higher In content. For $In_xGa_{1-x}N$ epitaxial layer B.N Pantha et al. [17] reported experimentally that E_A is 43 meV when x is up to 0.35 which is roughly four times smaller than that in Mg doped GaN. The minimum E_A, (~41 meV) for x=0.4, is calculated by following fitting equation which coincides with the experimental results [10], [17], [15], [12]. The fitted parameters for E_A are chosen arbitrarily considering physical phenomena. This equation provides such a close agreement with the experimental findings and able to calculate E_A of Mg acceptor with different In composition, x in p- $In_xGa_{1-x}N$ epitaxial layer.

$$E_A = \frac{117}{1+(5x^{4/5})^3} + 33 \quad (1)$$

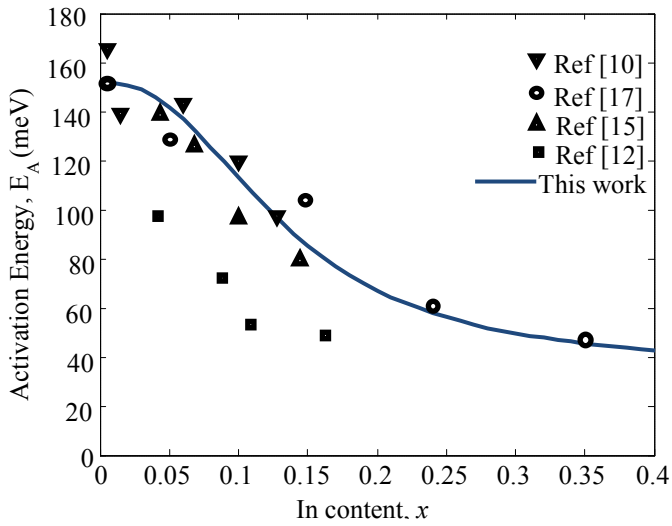


Fig. 1 Decrease of Activation Energy, (E_A) of Mg acceptors in Mg doped p-$In_xGa_{1-x}N$ alloys with increase in In content, x.

B. Hole Concentration in p-InGaN

The hole concentration (p) in Mg doped $In_xGa_{1-x}N$ which depends on the activation energy, has been shown in Fig. 2 as a function of In content, x at room temperature. The increase in hole concentration is caused by the decrease in Mg acceptor activation energy [10], [12]. By using equation (2) [18] and from Fig. 2, it is found that the hole concentration continuously increases from 4.9×10^{17} cm^{-3} for x=0 (GaN) to 9.7×10^{18} cm^{-3} for x=0.4. Such values suggest that it can be able to achieve good ohmic contacts between metal-contact and p-InGaN which is very congenial to reduce the contact resistance. Due to the presence of high background electron concentration in p-$In_xGa_{1-x}N$, the value of p reduces for x>0.40, has been discussed latter.

$$p = -\frac{N_D + K}{2} + \sqrt{\left(\frac{N_D + K}{2}\right)^2 + K(N_A - N_D)} \qquad (2)$$

When

$$K = \frac{1}{\beta} N_{val} \exp(-\frac{E_A}{k_B T}) \qquad (3)$$

Where N_D is the donor impurity density, N_A the Mg doping concentration, β is the valance band (vb) degeneracy factor taken as 3.6, N_{val} is the tem perature dependent vb density of states (DOS), E_A is the thermal Mg-acceptor ionization energy obtained from equation (1), k_B is the Boltzmann constant and T is the room temperature. N_{val} is obtained 3.2×10^{19} cm^{-3} for 40% In composition in the InGaN epitaxial layer [19].

$$N_{val}\,(In_xGa_{1-x}N) = N_{val}(InN)\times(x) + N_{val}(GaN)\times(1-x)$$
$$= \{5.3\times x + 1.8\times(1-x)\}\times10^{19} \qquad (4)$$

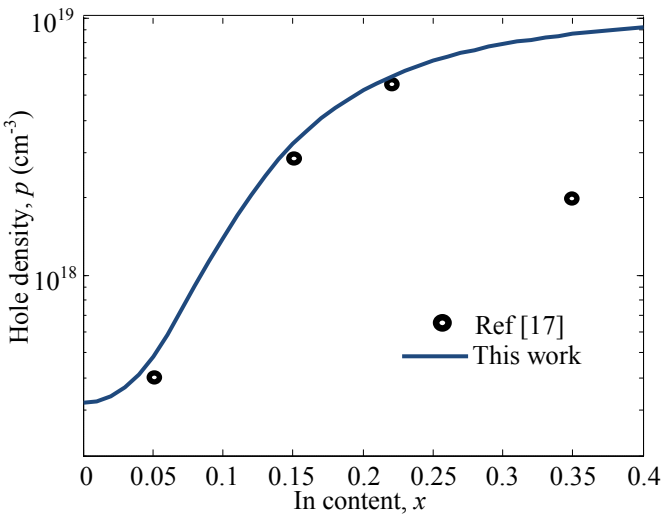


Fig-2: Room temperature hole concentration (p) as a function of In content x in Mg doped $In_xGa_{1-x}N$

III. Modelling of Self-Compensation

Fig. 3 shows the hole density (p) with the variation of Mg-concentration (N_A). The three curves corresponds to three residual donor concentrations, N_D in an impurity compensation model which indicates that, in moderate doping range , 5×10^{18} cm^{-3} ≤ N_A ≤ 3.0×10^{19} cm^{-3} , hydrogen de-passivation is very effective and impurity compensation as well as self-compensation (SC) having less effect. They are calculated at room temperature with the expression of p

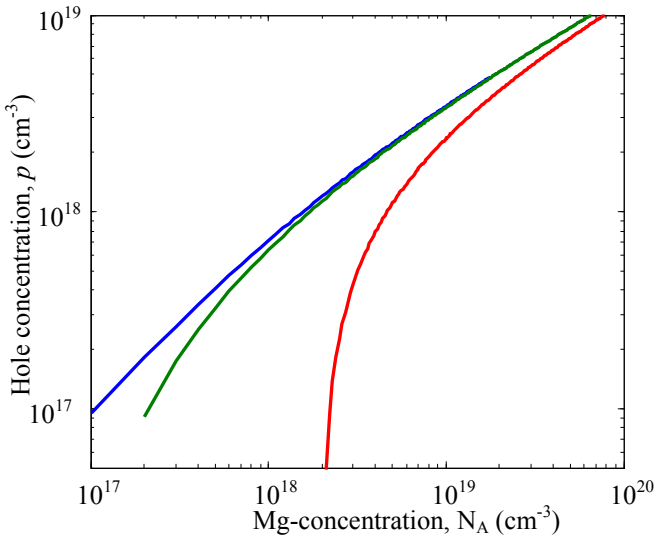


Fig. 3 Hole concentration (p) with respect to Mg-concentration (N_A) in p-$In_xGa_{1-x}N$ for three residual donor densities as shown in the legend

shown in equation (2). The values of donor concentration N_D are chosen 0, 1×10^{17} cm^{-3} and 2×10^{18} cm^{-3} and K at room temperature is obtained as K=1.8×10^{18} cm^{-3}. For this situation, E_A=41 meV has been kept fixed. Actually, this figure indicates that when N_A becomes larger than 3.0×10^{19} cm^{-3}, the measured hole concentration falls progressively due to the increase of the degree of compensation whatever the donor concentration N_D is.

Considering two idealized case of SC, one due to isolated nitrogen vacancy V_N which has a triple donor (multiplicity, m=3) and other due to the nearest neighbor complex MgV_N which has double donor (multiplicity, m=2) [20]. The former case is considered to be less realistic. But it has been included for the comparison purposes with the latter case. The latter case is considered important since the large amounts of dopants and vacancies are ionized and became mobile. Thus they tend to form near neighbor pair (MgV_N) when the doping concentration is high (upper 10^{19} cm^{-3}) and the mean separations are in the order of a few nanometers only. Self-compensation coefficient (c) can be defined by c=N_D/N_A where N_D and N_A are the donor and the Mg doping concentration, l indicates corresponds to case 1 (l=1) and case 2 (l=2) and N_{CC} introduces a fit parameter. Equation (1) can be rewritten for impurity compensation where all symbols have been defined above [21].

$$p = -\frac{mcN_A + K}{2} + \sqrt{\frac{(mcN_A + K)^2}{4} + KN_A(1-3c)} \qquad (5)$$

$$c = \frac{1}{3}\left(\frac{N_A}{N_{cc}}\right)^{l+1}\left(\frac{1+\sqrt{1+\alpha N_{cc}/N_{val}}}{1+\sqrt{1+\alpha N_A/N_{val}}}\right)^3 \qquad (6)$$

It should be noted that the parameter α and N_{val} in equation (6) are not independent, rather, α/ N_{val} ≡ 4/K where K has already been given. In Fig. 4, p vs N_A are plotted for case 1 and case 2. The fitting parameter N_{CC} for two curves were obtained N_{CC} ~1×10^{21} cm^{-3} and 1.5×10^{20} cm^{-3} for case 1 and case 2 respectively. Now, the desired $p(N_A)$ at room temperature can

be calculated from equation (5) by using the described input parameters of Mg acceptor E_A, the vb (β, N_{val}), as well as the fit parameter N_{CC}. If N_A is small compared to N_{CC}, the SC co-efficient c obtained from equation (6) are very small. This requirement is satisfied when $N_A<2\times10^{19}$ cm^{-3}.This indicates that SC in Mg doped InGaN is negligible for $N_A<2\times10^{19}$ cm^{-3}, but the SC becomes more significant when N_A approaches N_{CC}. The hole concentration (p) reaches at maximum near $N_A=4.5\times10^{19}$ cm^{-3} and decreases at higher Mg concentrations. Thus provides strong evidence that SC is the cause for the reduction in p at very high doping levels. If a better control of back-ground electron concentration could be achieved, it would be possible to improve p-type conductivity in Mg doped InGaN.

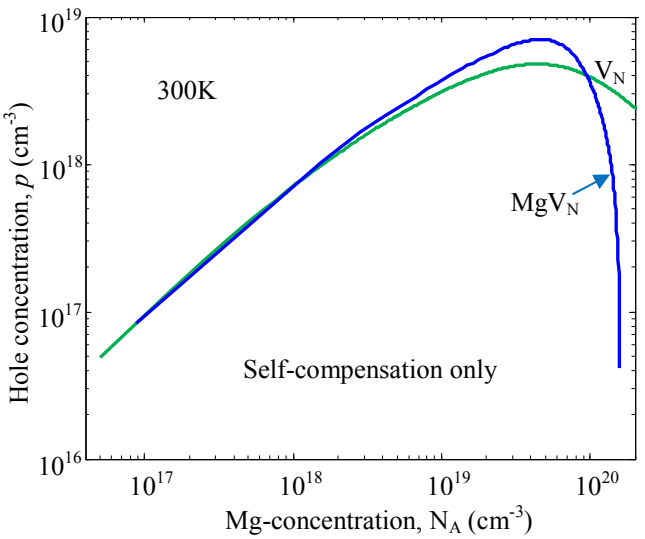


Fig. 4 Hole concentration (p) vs Mg concentration (N_A) based on self-Compensation

IV. Results & Discussion

Investigation of low activation energy (E_A) of Mg as well as higher hole concentration (p) in Mg doped p-$In_xGa_{1-x}N$ alloys is the promising issues for device application. The lower the E_A, the higher the p, will be possible if the better control of high background electron concentration in p-$In_xGa_{1-x}N$ is achieved. A comparison table of different E_A has been demonstrated below where calculated E_A from equation (1) has been compared with the practically obtained E_A. The calculated value of E_A is near about 41 meV and p was approximately 9.7×10^{18} cm^{-3} when In composition, x is 40% in Mg doped $In_xGa_{1-x}N$. At higher doping level when $N_A=4.5\times10^{19}$ cm^{-3}, the hole concentration has been investigated as maximum and then decreased for further increase of N_A due to the effect of self-compensation.

TABLE 1
COMPARISON OF E_A (CALCULATED AND PRACTICAL) AT DIFFERENT N_D AND N_A

Sample $In_xGa_{1-x}N$:Mg	$N_D\times10^{18}$ cm^{-3}	$N_A\times10^{19}$ cm^{-3}	E_A (meV)	E_A (meV) (This work)
x=0.04	1.0	0.53	141 [15]	144
x=0.10	2.0	3.0	120 [10]	111
x=0.35	2.0	3.0	43 [17]	43
x=0.40	2.0	3.0	….	41

V. Conclusion

In summary, the activation energies of Mg-acceptor in p-$In_xGa_{1-x}N$ alloys have been demonstrated mathematically by fitting equation for In composition, x up to 0.40. It has been found that the activation energy of Mg in p-$In_xGa_{1-x}N$ ($0<x<0.4$) decreased with increasing In content, x. The value, calculated using this equation is about 41 meV which is in close agreement with the experimental findings. The maximum hole concentration (p) of $In_{0.4}Ga_{0.6}N$ was 9.7×10^{18} cm^{-3}. Two semi quantitative Self-Compensation models with either isolated V_N or MgV_N pair predicts that p reaches maximum near $N_A=4.5\times10^{19}$ cm^{-3} and fall progressively beyond this N_A, that is why obtaining high hole concentration in p-$In_xGa_{1-x}N$ is difficult.

References

[1] J. Wu, W. Walukiewicz, K. M. Yu, J. W. Ager, III, E. E. Haller, H. Lu, W. J. Schaff, Y. Saito, and Y. Nanishi, "Unusual properties of the fundamental band gap of InN", *Appl. Phys. Lett.*, vol. 80, pp. 3967–3969, Apr. 2002.

[2] S. F. Fang, K. Adomi, S. Iyer, H. Morkoc, H. Zabel, C. Choi, and N. Otsuka, "Gallium Arsenide and Other Compound Semiconductors on Silicon", *Journal of Applied Physics* , vol.68, pp.31-58, Oct 1990.

[3] S. Nakamura, "Present status of InGaN-based laser diodes," *phys. stat. sol.*, vol. 176, July 1999.

[4] O. Jani, I. Ferguson, C. Honsberg, S. Kurtz, "Design and characterization of GaN/InGaN solar cells," *Appl. Phys. Lett.* vol. 91, pp. 13217-13220, Sep. 2007.

[5] Elaissa Trybus, Gon Namkoong, Walter Henderson, Shawn Burnham, W. Alan Doolittle, Maurice Cheung, Alexander Cartwright, "InN: A material with photovoltaic promise and challenges" *Journal of Crystal Growth,* Vol. 288, pp. 218-224, March 2006.

[6] K. Jagannadham, E. A. Berkman, N. Elmasry, "Thermal conductivity of semi-insulating, p-type, and n-type GaN films on sapphire," *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films*, vol.26, pp.375-379, May 2008.

[7] K. Kumakura, T. Makimoto, N. Kobayashi, "Low-resistance nonalloyed ohmic contact to p-type GaN using strained InGaN contact layer," *App. Phys. Lett*, vol.79, pp.2588-2590, Oct 2001.

[8] K. Kumakura, T. Makimoto, N. Kobayashi," Efficient Hole Generation above 10^{19} cm^{-3} in Mg-Doped InGaN/GaN Super lattices at Room Temperature" *Jpn. J. Appl. Phys*. vol. 39, L195, Jan. 2000.

[9] J. Li, T. N. Oder, M. L. Nakarmi, J. Y. Lin, and H. X. Jiang, "Optical and electrical properties of Mg-doped p-type $Al_xGa_{1-x}N$" *Appl. Phys. Lett*. vol. 80, pp. 1210-1214, Dec. 2001.

[10] Kumakura, K.; Makimoto, T.; Kobayashi, N., "Mg-acceptor activation mechanism and transport characteristics in p-type InGaN grown by metalorganic vapor phase epitaxy," *Journal of Applied Physics*, vol.93, pp.3370-3375, Mar 2003.

[11] S. Yamasaki, S. Asami, N. Shibata, M. Koike, K. Manabe, T. Tanaka, H. Amano, and I. Akasaki, "p-type conduction in Mg-doped Ga0.91In0.09N grown by metalorganic vapor-phase epitaxy," *Applied Physics Letters* , vol.66, pp.1112-1113, Feb 1995

[12] K. Kumakura, T. Makimoto, and N. Kobayashi," Activation Energy and Electrical Activity of Mg in Mg-Doped $In_xGa_{1-x}N$ (x<0.2)," *Jpn. J. Appl. Phys.*, vol. 39, L 337,March 2000.

[13] K. Kumakura, T. Makimoto and N. Kobayashi," High hole concentrations in Mg-doped InGaN grown by MOVPE," *J. Crystal Growth,* vol.221, pp. 267-270, June 2000.

[14] T. C. Wen, W. I. Lee, J. K. Sheu, G. C. Chi," Characterization of p-type In x Ga 1− x N grown by metalorganic chemical vapor deposition" *Solid-State Electron*. vol. 45, pp.427-3, Jan. 2001

[15] D. Iida, M. Iwaya, S. Kamiyama, H. Amano, and I. Akasaki," High hole concentration in Mg-doped *a*-plane $Ga_{1-x}In_xN$ ($0<x<0.30$) grown on *r*-plane sapphire substrate by metalorganic vapor phase epitaxy" *Appl. Phys. Lett.* Vol.93, pp.182108-182112, Nov.2008.

[16] C. A. Chang, T. Y. Tang, P. H. Chang, N. C. Chen, and C. T. Liang," Magnesium Doping of In-rich InGaN" *Jpn.J. Appl. Phys.,* Part 1 vol.46, pp.2840-2844, Jan. 2007.

[17] B. N. Pantha, A. Sedhain, J. Li, J. Y. Lin, and H. X. Jiang, "Electrical and optical properties of p-type InGaN," *Applied Physics Letters* , vol.95, pp.261904-261907, Dec 2009.

[18] J. S. Blakemore," Semiconductor Statistics" *Dover publication,* New York, 1987.

[19] Levinshtein, M.E., Rumyantsev, S.L. and M.S., "Properties of advanced semiconductor materials,"*Willey student edition*, Chichester, UK, pp.1-90, 2001.

[20] S. Nakamura, T. Mukai, M. Senoh, and N. Iwasa," Thermal Annealing Effects on P-Type Mg-Doped GaN Films," *Jpn. J. Appl. Phys.* vol. 31, Part 231, L139, Feb. 1992.

[21] U. Kaufmann, P. Schlotter, H. Obloh, K. Ko¨hler, and M. Maier," Hole conductivity and compensation in epitaxial GaN:Mg layers," *Phys. Rev. B*, vol. 62, pp. 10867-10872,Oct. 2000.

Analytical Modeling of Accumulation Layer Thickness in SOI G^4FET: Introducing Geometric Factor

Mehedy Hasan Bappy[1*] and M. Ziaur Rahman Khan [1**]

[1] Department of Electrical and Electronic Engineering,
Bangladesh University of Engineering and Technology,
Dhaka-1205, Bangladesh

[*]bappy.eee.buet.07@gmail.com [**]zrkhan@eee.buet.ac.bd

***Abstract*— An analytical model has been developed to determine the accumulation layer thickness of an inverted back interface silicon-on-insulator (SOI) four-gate transistor (G^4 - FET) by solving two dimensional Poisson's equation. To do so a geometric factor is introduced that combines the neglected inversion effects along with the existing device physics. Moreover, this factor considers the charge redistribution due to change in bias condition in different electrodes of the device. Physically, it indicates the non-availability of space for both depleted and inverted charges. The proposed model, named as modulated model for accumulation thickness considering inversion effects is verified numerically. The model commenced here successfully correlates the effect of all four gate bias.**

***Index Terms*— G^4-FET, Accumulation layer thickness, Inversion operation, Geometric Factor.**

I. Introduction

The four gate transistors (G^4-FET), formerly known as MOS-JFET, first introduced in 2002, features the maximum number of gates that can be accommodated in a single transistor and is a genuine four-gate transistor whose four gates can be biased independently [1]. Preliminary investigation confirms the potential of the G^4-FET for innovative analog circuits that can be conceived within a standard partially depleted SOI technology. Due to its multiple independent gates, the use of the G^4-FET enhances circuit design flexibility while reducing transistor count as compared to standard CMOS implementation [2]. This device can be manufactured using conventional SOI processes [3], without necessitating additional fabrication steps, being adaptable to more aggressive technology nodes [4].

Charge sheet model is vastly applicable to long channel devices. It leads to a very simple algebraic formula for the current of long-channel devices [5]. The same formula applies in all regimes from sub-threshold to saturation, neither splicing nor parameter changes are needed, no discontinuities occur in either the current or the small-signal parameters, or in the derivatives of the small-signal parameters. Also it can be easily extended to two to three dimensional modeling.

When the device is scaled to nano-range, this accumulation layer or channel can no longer be approximated as a sheet of charge of zero thickness, since the thickness becomes comparable to the silicon film thickness [6]. So mathematical modeling of accumulation layer thickness is necessary for different bias condition to determine the range of operation where charge sheet model can be applied. The analysis to determine the dependence of accumulation layer on the structural parameters of the device is also needed.

II. Device Structure

G^4-FET is a majority carrier, buried channel, accumulation mode device where the source, drain, and body / channel regions are all same type. It has two vertical MOS gates: a conventional polysilicon top gate, as well as the substrate, which can act as a bottom gate. In addition, it has two lateral JFET gates that form PN junctions with the channel region [7]. Figure 1 shows basic structure of G^4-FET illustrating different region.

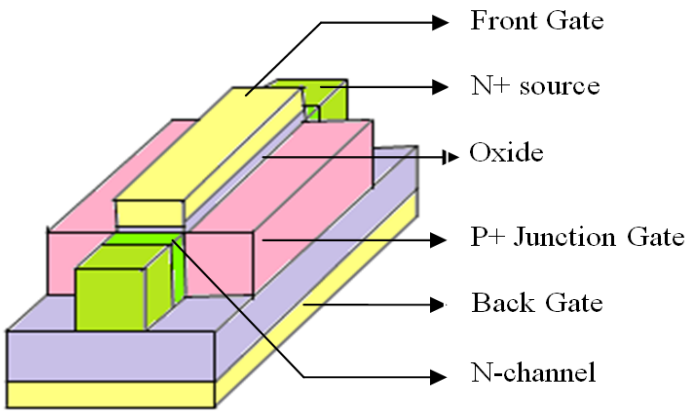


Figure 1 : G^4-FET Basic Structure (n-channel G^4-FET)

Application of a negative voltage to the top gate of an n-channel G^4-FET depletes the channel region. Negative voltages at JFET gates cause the PN-junction depletion region to broaden, further narrowing the channel. So both MOSFET and JFET like control is present in G^4-FET.

In partially depleted G^4-FETs, two distinct mode of operation are identifiable. In MOSFET mode, operation of G^4-FET is based on the front surface accumulation current controlled by V_{G1} . In this case V_{G2} and V_{JG} are parameters of the device [8]. Meanwhile in JFET mode, operation of the G^4-FET is based on the variation of the cross section of the channel (volume neutral current) controlled by V_{JG} while V_{G1} and V_{G2} are parameters inducing depletion or inversion at the front and back interfaces [9].

III. Modeling Accumulation Layer Thickness

The accumulation surface potential is given by the drops across the accumulation layer formed under the front oxide-silicon interface. Under fully depleted approximation, total charge can be approximated as

$$\rho(x,y) = qN_D.$$

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

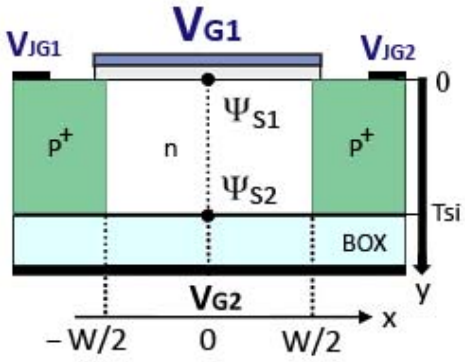


Figure 2 : n-channel G^4-FET cross section

Then the 2D Poisson equation is given by,

$$\frac{d^2\Psi}{dx^2} + \frac{d^2\Psi}{dy^2} = -\frac{qN_D}{\epsilon_{si}}$$

Assuming,

$$\Psi(x,y) = ax^2 + bx + c \text{ ... (1)}$$

Potential at the centre of bulk Silicon in G^4FET,

$$\Psi(-W/2,y) = \Psi(W/2,y) = V_{JG} - \varphi_B \text{ ... (2)}$$

So from (1) and (2),

$$\frac{d^2\Psi}{dx^2} = -\frac{8}{W^2}\Psi + \frac{8}{W^2}\left(V_{JG} - \varphi_B\right) \text{ ...(3)}$$

After some mathematical calculation,

$$dy = \frac{d\Psi}{\sqrt{\left\{\frac{8}{W^2}[\Psi^2 - 2(V_{JG} - V_P)] + C\right\}}} \text{ ... (4)}$$

Where, $V_P = \varphi_B - \frac{W^2 qN_D}{8\epsilon_{si}}$

Integrating from surface of bulk Silicon to accumulation layer thickness (0 to y_{acc}) at the front gate centre,

$$y_{acc} = \frac{W}{2\sqrt{2}}\left[\sinh^{-1}\left\{\frac{\Psi_{s1} - V_{JG} + V_P}{\zeta}\right\} - \sinh^{-1}\left\{\frac{V_P - V_{JG}}{\zeta}\right\}\right] \text{ ... (5)}$$

Here, $\zeta^2 = \frac{W^2}{8}C - \left(V_{JG} - V_P\right)^2$

IV. Introducing Geometric Factor (G)

A charge sheet model was proposed [10] to analyze the behavior of fully depleted SOI four gate field effect transistors (G4-FET) above the threshold condition. That model was derived assuming parabolic potential variation between junction gates and by solving 2-D Poisson's equation. Using the charge sheet model current-voltage characteristics were analyzed. Different back surface conditions were considered for each analysis. The models were compared with 3-D Silvaco/Atlas simulation results. They showed pretty good agreement most often. But observing each characteristic for different back gate conditions, it was found that models barely follows the simulated results for back gate inversion condition even in depletion. This section is going to investigate the reason behind this matter and search for a remedy to overcome the discrepancy. Here the device physics is been rethought from different angles as well as reconsider some neglected phenomena which are no longer negotiable for an inverted back gate bias.

A. *Problem with Usual Assumption on Charge Distribution in Bulk of Device*

The accumulation layer thickness is a function of front gate biases and junction gate biases for different back surface charge conditions. While full depletion approximation is used model, that is derived and stated in previous section (Equation 5) shows good agreement with simulation results, from onset of accumulation to the point where strong accumulation is achieved. The layer thickness becomes constant at strong accumulation.

With classical physics the foremost wrong was to assume the Silicon body is still depleted fully when back gate is in moderate or strong inversion mode. Obviously this assumption makes the analysis comparatively simpler as analytical model is derived through 2D Poisson equation, that has charge quantity on the right hand side. When fully depleted body is considered the charge density associated is simply given by $\rho(x,y) = qN_D$. But with a moderate inversion condition, this quantity is no longer valid. Moreover in case of a strong inversion at back surface, the sum of charge, encircled by the body may be the opposite (from +ve to –ve in n-channel / from –ve to +ve in p-channel) to that of in depleted back gate case. Thus fully depleted body assumption is no longer valid for inverted back surface potential.

The uniformly distributed charge assumption lessens the complexity of device electrostatics as well as electromagnetic. That assumption is pretty much valid with small front and junction gate bias (near zero) and sub-threshold condition (very small drain bias, nearly zero) while back gate is depleted. But an inverted back surface ends the validity of this assumption. Uniform charge distribution is hampered by each and every electrodes of the device on prior of bias voltage value.

B. *Geometric Factor (G)*

With an inverted back gate potential neither the fully depleted idea nor the uniform charge distribution is appropriate or acceptable. So the charge distribution phenomena need to be modified for using 2D Poisson equation. But it is very difficult to consider all the electrostatic and electromagnetic phenomena, involving in a device structure and then transform them into mathematical representation. Thus different assumptions have been made usually while analysis (as was the fully depleted body assumption), until the assumption results significant scale of error.

So it is required to rethink and reconsider the charge distribution fact. One thing is to keep in mind that there is not enough scope to consider one electrode and find its effect on non-uniformity as it is a nano range device and all electrodes' effects are interconnected.

One way to reconsider the charge density is to modify the doping density (N_D). When back gate inverted, no longer uniform charge distribution, charge density $\rho(x,y) = qN_D$ is violated. As charge is quantized, it can be called as "effective doping density" (N_D^{eff}) for inversion operation. The modified doping density modifies the charge density that will ultimately compensate the non-uniform charge and the error of fully depleted body assumption as well.

Now, consider a cross section of G4FET. Cross section of this plane has four electrodes - two junction gates, front gate and back gate .This is an n-channel device. The back gate has a high negative bias thus back surface is in strong

inversion. Junction gates have small negative bias. The charge profile is shown in figure 3 approximately. Clearly charge is non-uniformly distributed across the plane. Let the doping density of the body is N_D.

For a given V_{JG}, the charge condition at the back interface is governed by V_{G2}. Back gate bias can be represented by the following equation.

$$V_{G2} = V_{FB2} + \left(1 + \alpha\frac{C_{jg}}{C_{ox2}}\right)\Psi_{s2} - \gamma\frac{C_{jg}}{C_{ox2}}\Psi_{s1} + (\gamma - \alpha)\left(\frac{C_{jg}}{C_{ox2}}\right)(V_{JG} - V_P)$$

The back surface point, which is at mid-width between the junction gates, reaches accumulation for $V_{G2} \geq V_{G2}^{acc2}$ and is in inversion for $V_{G2} \leq V_{G2}^{inv2}$ [12].

$$V_{G2}^{acc2} = V_{FB2} + (\gamma - \alpha)\left(\frac{C_{jg}}{C_{ox2}}\right)(V_{JG} - V_P)$$

$$V_{G2}^{inv2} = V_{FB2} + \left(1 + \alpha\frac{C_{jg}}{C_{ox2}}\right)2\phi_F - (\gamma - \alpha)\left(\frac{C_{jg}}{C_{ox2}}\right)V_P + (1 + \gamma\frac{C_{jg}}{C_{ox2}})V_{JG}$$

That is the minimum back gate voltage that, make the back surface onset of inversion. Now with the increase in back gate voltage more charges gradually pile up at the back gate surface.

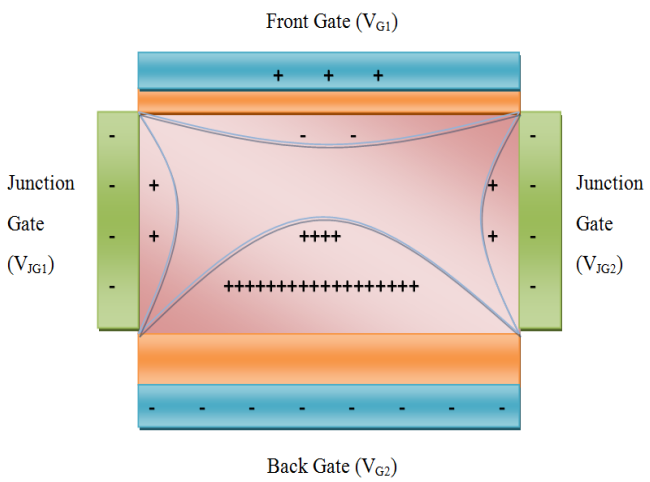


Figure 3 : Charge Profile in n-channel G4FET Cross Section.

Near the back surface in inversion, average charge per unit area can be given as

$$Q_{Bulk} = -(V_{G2} - V_{G2}^{inv}) \cdot C_{OX2}$$

where ,

$$C_{OX2} = \square_{OX2}/t_{OX2}$$

So, average charge density throughout the bulk Silicon can be given as

$$Q_n = Q_{Bulk} / t_{Si}$$

It is important to note that while finding the charge density the charge per unit area has been divided by t_{Si}. This implies that Q_{Bulk} is considered to be the charge per unit area along the whole Si thickness i.e. throughout the bulk. But it is not true, as earlier the non-uniformity of charge has been emphasized. To compensate for this assumption, geometric factor (G) has been introduced. Geometric factor (G) can be given by the following equation.

$$G = \frac{(1 + aV_{G1})(1 + c|V_{G2}|^r)}{1 + \frac{1}{(1 + |V_{JG}|)^b}}$$

Without considering this geometric factor, effective doping density

$$N_D^{eff} = Q_n / q$$

Where, q = charge of an electron =1.6 x 10^{-19} C

Taking the geometric factor into consideration,

$$N_{inv} = Q_n / qG$$

The name "geometric factor" represents that G is a mathematical way to indicate geometric area for both depleted and inverted charges. An increase in the value of G, due to the increase in bias voltage, results in more non-availability of space for both depleted and inverted charges.

Empirical constants (a,b,c,d) are dependent on the device design parameters and geometric structure as well. With proper modification of these constants or finding the dependency of these constants on the design parameters, proposed geometric factor can be used in wide range of device operations.

V. Numerical Verification

Fully depleted body approximation is applicable when the device is operating in depletion mode, so model obtained from this assumption considering classical physics cannot be applied for all bias conditions and modes of operation. Though until the onset of inversion this model is in good agreement with simulation result, when back gate is in inversion mode, the device shows significant deviation from assumed fully depleted Model [11].

In this section, the proposed model (with geometric factor) to include inversion effects, will be numerically verified. For all the simulations used device parameters are: Channel width, W=120nm; front oxide thickness, t_{ox1} = 5nm; back oxide thickness, t_{ox2} = 50nm; silicon body thickness, t_{si} =40nm; channel doping concentration, $N_D=10^{17}cm^{-3}$; silicon body doping concentration, $N_A=10^{20}$ cm^{-3}.

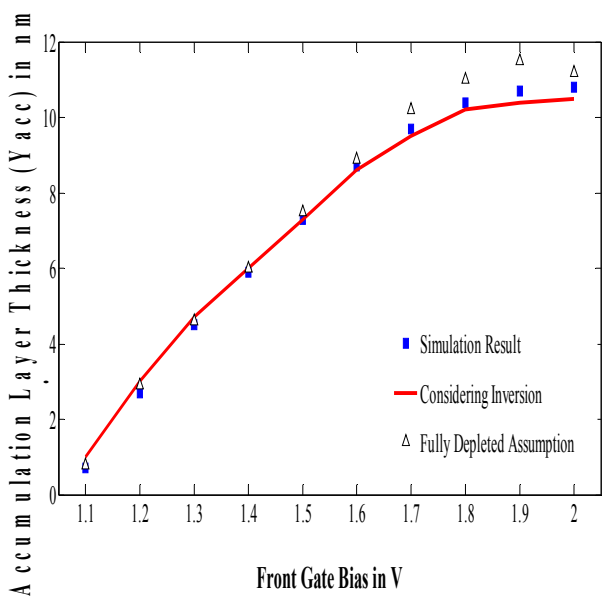


Figure 4 : Comparison between Inversion Model, Depletion Model and Simulation Result (Back gate bias -10V)

Figure 4 illustrates the simulation result along with the model results (both depletion assumption and considering inversion effects) for accumulation thickness for varying front gate bias while a back gate bias of -10V is applied. It is apparent that with lower front gate bias, depletion Model is fairly accurate. It can be deduced that, at -10 V device is not even at the onset of inversion and with a low voltage at front gate, not significant level of inversion occurs. So, fully depleted approximation is valid enough. But at higher front gate bias full depletion assumption wear out.

With a back gate bias much high (V_{G2} =-20V), device is in moderately strong inversion region. So, uniform charge distribution and fully depleted condition are no longer valid. In Figure 5 depicts the compatibility of Model considering inversion with simulation result much better than Model considering full depletion.

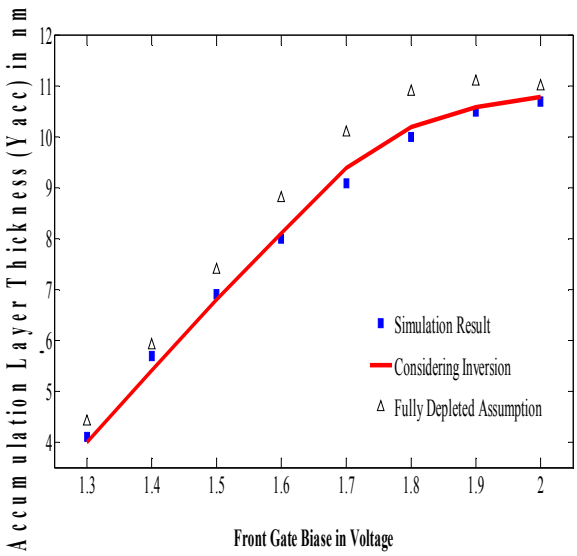


Figure 5 : Comparison between Depletion Model, Inversion Model and Simulation Result (Back Gate Bias -20V)

Figure 6 plots Inversion Model at -20 V back gate bias varying junction gate bias. It is apparent that simulation result is fairly in accord with the deduced model. Percentage of error is significantly less for model considering inversion.

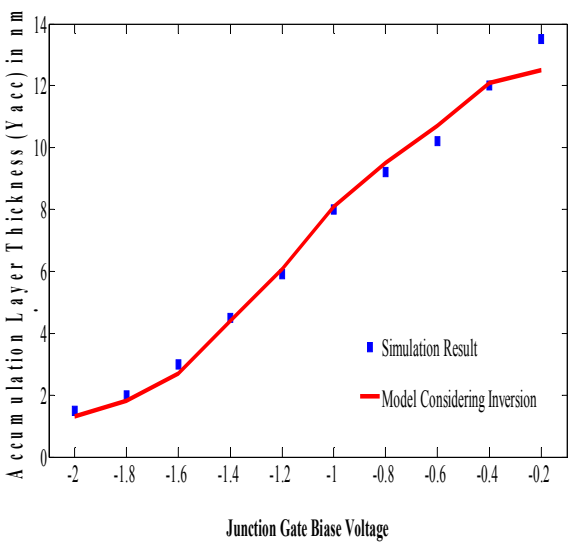


Figure 6 : Accumulation Layer Thickness Vs Junction Gate Bias (comparison between different models and simulation result)

VI. Conclusions

The proposed model that includes geometric factor provides enough satisfactory results in case of inverted back surface and depleted back surface as well. The Stronger the inverted back surface, the more compatible the analytical model with the numerically simulated ones. Most importantly the bias voltage dependent geometric factor gives us the scope for considering dynamic charge profile with the change of bias voltage in different electrodes of the device. With proper modification and improvisation of this factor, it is possible that, analysis of any device with any structural parameter will be much error free in any operating conditions.

References

[1] K. Akarvardar, S. Chen, J. Vandersand, B. Blalock, R. Schrimpf, B. Prothro, C. Britton, S. Cristoloveanue, P. Gentil, and M. M. Mojarradi, "Four Gate transistors Voltage- Controlled Negative Differential Resistance Device and Related Circuit Applications," in *International SOI Conference,2006 IEEE,* pp.71-72,2006.

[2] K. Akarvardar, B.J. Blalock, S. Cristoloveanue, P. Gentil, M. M. Mojarradi, "A Novel Four-Quadrant Analog Multiplier Using SOI Four-Gate Transistors (G^4-FETs)," in *31st ESSCIRC, 2005, pp.499-502*

[3] Kerem Akarvardar, Sorin Cristoloveanu and Pierre Gentil- "Analytical Modeling of the Two-Dimensional Potential Distribution and Threshold Voltage of the SOI Four-Gate Transistor"- *IEEE Transactions on Electron Devices*, VOL. 53, NO. 10, pp.2569-2577, 2006.

[4] B.Dufrene, K.Akarvardar, S.Cristoloveanu, B. J. Blalock, P. Gentil, E. Kolawa, and M. M. Mojarradi -"Investigation of the Four-Gate Action in G^4–FETs"- *IEEE Transactions on Electron devices*, Vol. 51, No. 11, pp 854-857, 2004.

[5] J.R. Brews "A charge-sheet model of the MOSFET," *Solid-State Electronics* vol. 21, no.2, pp. 345–355, 1978.

[6] S. Wind, D. Frank, and H. Wong, "Scaling MOS device to their limits," Microelectronics Engg.,vol.32, pp. 271-282, 1996.

[7] B.Dufrene, K.Akarvardar, S.Cristolovenue, B.J.Blalock, R.Gentil, E.Kolawa, and M.M.Mojarradi, "*Investigation of the four-gate action G^4-FETs,"Electron Devices, IEEE Transactions on,* vol.51, pp.1931-1935,2004.

[8] R.H.Yan, A. Ourmazd, and K. F. Lee, "Scaling the Si MOSFET; from bulk to SOI to bulk, " *Electronics Devices, IEEE Transactions on* vol.39, pp. 1704-1710, 1992.

[9] B.J. Blalock, S.Cristoloveanu, B. M. Dufrene, F.Allibert, and M.M. Mojarradi, "The Multiple-Gate MOS-JFET Transistor," *International Journal of High Speed Electronics Systems,* vol.12, pp.511-520, 2002.

[10] Shehrin Sayed, Md. Iftekhar Hossain, and M. Ziaur Rahman Khan-" A Subthreshold Swing Model for Thin-Film Fully Depleted SOI Four-Gate Transistors"- *IEEE Transactions on Electron Devices*, vol. 59, no. 3, pp. 854-857, 2012.

[11] Shehrin Sayed and M. Ziaur Rahman Khan, "Analysis on Surface Accumulation Behavior of Fully Depleted G^4-FETs," *2012 7th International Conference on Electrical and Computer Engineering,* 2012, pp. 795-798.

[12] Kerem Akarvardar, Sorin Cristoloveanu, Pierre Gentil, Ronald D. Schrimpf and Benjamin J. Blalock- "Depletion-All-Around Operation of the SOI Four-Gate Transistor"- IEEE TRANSACTIONS ON ELECTRON DEVICES, VOL. 54, NO. 2, pp. 323-331, 2007

Imaging Beyond Human Vision

Achyut K. Dutta
Banpil Photonics, Inc., 4800 Patrick Henry Drive, Suite # 120, Santa Clara, CA 95054, USA
E-mail: akdutta@banpil.com

***Abstract*— Traditional imaging has been limited to the visible spectrum that humans can see. Imaging technologies are transcending into the invisible spectrums in the infrared (IR) region to enable imaging beyond human vision. We highlight the capabilities and advantages and therefore the value of IR imaging focused primarily on shortwave IR (SWIR) with some elements of long wave IR (LWIR) over the visible spectrum (VIS). We present an overview of SWIR and LWIR device structures based on InP and HgCdTe devices for the respective bands. We also highlight our development of a high-performance multispectral camera system for VIS-SWIR imaging for the most demanding high-sensitivity and high-speed commercial and industrial applications. The 640x512 pixel VIS-SWIR camera system is designed to provide a compact, small-form factor to within a cubic inch, high sensitivity needing less than 100 electrons, high dynamic range exceeding 190 dB, and high-frame rates (>1000 FPS) at full resolution.**

***Index Terms*—** Infrared, camera, multispectral, VIS-SWIR, InP, InGaAs, HgCdTe, MCT, FPA, high-performance

I. INTRODUCTION

Traditional imaging has been closely tied to human vision so that whatever our eyes could see, process and interpret naturally were our limits. Science and technology has over the decades developed other forms of non-visual imaging including radio frequency and thermal signatures, but the most prominent unnatural mode of imaging to emerge has been infrared (IR) imaging. IR has blossomed on a slow trajectory in technical fields and especially in military applications, and in recent years due to various strides in materials, devices and technologies, IR imaging is steadily finding applications and adoption in commercial and industrial applications.

The extension of imaging system sensitivity from the visible (VIS) spectrum into the IR region has the potential to bring tremendous improvements and new capabilities to dual-use commercial and industrial imaging systems that are already standard in defense imaging applications. However, even as IR imaging systems are becoming commercially available from a number of companies with applications as diverse as security, biological imaging, and agricultural processing, the benefits, advantages and ultimate value over visible imaging is not well understood by target users. Furthermore, existing infrared imaging technologies are still very expensive. At costs in the tens of thousands of dollars per camera [1], it is difficult to make a strong case for its necessity and value as a precursor for penetrating high-volume commercial markets. Last, but no least, to make imaging beyond human vision practical, IR imaging needs to achieve high-performance, small form-factor or reduced size, weight, power and cost effective imaging units suitable for broader adoption.

This paper provides an overview of IR imaging sensors that will extend human vision beyond the limited capabilities provided by solid-state silicon-based image sensors (CCD/CMOS) already in wide use for the visible spectrum. Specifically, we discuss short wave IR (SWIR) sensors using InP devices and long wave IR (LWIR) sensors implemented with mercury cadmium telluride (HgCdTe or MCT) devices. We provide an assessment of the value that IR imaging in the SWIR and LWIR bands provides over the visible (VIS) spectrum by taking a close comparative look at capabilities and performance in a variety of environments including, haze, fog, smoke, and clouds. In addition to presenting device structures for IR sensors extending imaging to broadband or multispectral coverage, we also discuss high performance requirements for IR imaging, emerging applications and trends, strategies for their mass adoption as well as the challenge and opportunities with bearing to our activities in developing a high-performance multispectral camera system for dual-use applications.

II. TRADITIONAL HUMAN VISION – STANDARD VISIBLE IMAGING

Solid-state image sensors have become commonplace in everyday commercial applications including digital cameras, smartphones, automotive sensors, security and surveillance systems, and novel medical applications such as a pill-sized camera that could replace colonoscopy for cancer detection, to name a few. Such imaging sensors typically comprise of CCD (charge coupled device) or CMOS (complementary metal oxide semiconductor) image sensors with associated switching elements, and address (scan) and read out (data) lines [2]. Both CCD and CMOS image sensors are fabricated from silicon (Si) technology, which is so matured today that several millions of pixels and surrounding circuitry can be fabricated in small devices quite cheaply. The number of sensors and the variety of functions they perform is also increasing largely due to the proliferation of the CMOS sensor, which has become the de facto image sensor displacing the CCD sensor, primarily due to its low cost.

CMOS has also been great in increasing imaging resolution into multiple megapixels while also becoming more compact into smaller form factor and achieving pixel sizes as small as few micron, the smallest to date. The advantage has been the ability to achieve high definition imaging at the consumer level with devices that are as small as the tip of your finger and at nearly disposable cost structures.

The CMOS sensor has closely emulated human vision with its ability to image in the visible spectrum generally with good performance between 0.4 to 0.9 μm with the availability of good lighting. However, as CCD and CMOS image sensors technology are based on Si-technology, the detectable spectral ranges are limited to wavelengths below 1 μm where Si exhibits absorption. Besides that, CCD and CMOS image sensors also have other shortcomings such as lower responsivity combined with lower quantum efficiency over broad spectral range. Therefore when the light level begins to deteriorate in low visibility situations such as fog, haze, smog, and especially night time, CMOS performance drops

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

drastically. This is because silicon's absorption only extends to 0.9 μm, its cutoff wavelength and thereafter becomes blind.

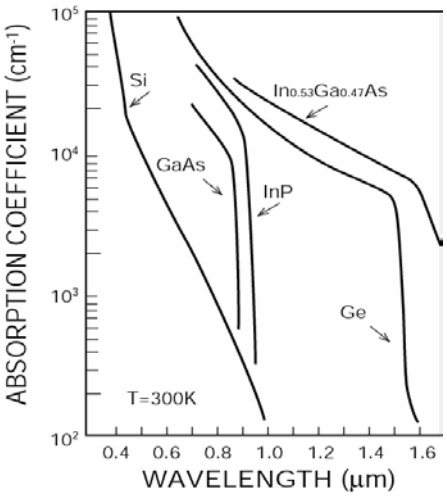


Figure 1: Dependence of the absorption coefficient on the wavelength for different semiconductors used in image sensors.

It is at this point where alternative imaging technologies come into play using other materials with capabilities beyond silicon's absorption wavelength and therefore beyond human vision.

Figure 1 shows the dependence of the absorption coefficient on the wavelength for different semiconductors used in photodetectors. Absorption has asymptotic behaviour near bandgap wavelength λ_g where the material is transparent, and there is a strong dependence of the absorption on the wavelength shorter than λ_g.

III. EXTENDED HUMAN VISION – FROM VISIBLE TO SWIR IMAGING

As we shift from the visible spectrum into increasing wavelengths, we get into the IR spectrum, which is itself made up of several bands from near IR (NIR) into SWIR, mid wave IR (MWIR) and LWIR. Through a selection of materials with absorption at appropriate wavelengths, we are able to fabricate image sensors with capabilities within the various bands. VIS is the 0.4 to 0.7 μm band, NIR is the 0.7 to 1.0 μm band, and SWIR is 1.0 to 1.7 μm band.

Below we discuss imaging capabilities across various environmental conditions which primarily compares and contrasts the visible spectrum with the SWIR spectrum and therefore highlighting the benefits and advantages offered by the ability to see beyond normal human vision.

III.1 Atmospheric Transmission: Atmospheric transmission, which includes light absorption and scattering through various types of atmospheric conditions, provides a good reason to move imaging towards longer wavelengths in the reflective bands. Atmospheric transmission ranges from the less challenging types such as the U.S. and Arctic, weighted more heavily on light scattering, to the more challenging atmospheres that are more visually degraded such as tropical, desert, and desert dust storms, which have more absorption. At the longest ranges, where range measures the distance between sensor and target, the SWIR band can have 10 times (or 10x) the atmospheric transmission of the VIS band. This is even more pronounced in better atmospheres as scattering is more important and less absorption occurs. Long range and low visibility makes for SWIR transmission that can be 100x and even as high as 1000x the transmission of VIS [3]. Atmospheric transmission or penetration clearly favors SWIR and then NIR over VIS imaging.

In studies [4] comparing the performance of VIS-NIR systems against SWIR, performance has been shown to be similar under good atmospheric conditions meaning with visibility up to 23 km. However, as atmospheric conditions are degraded for example with visibility at 6 km, VIS-NIR system exhibit severely poor performance compared to SWIR systems, which is consistent with the conclusions that SWIR imaging is the best band for long range and therefore also for degraded atmospheric spectral discrimination.

III.2 Imaging Through Haze: Haze is a commonly occurring visually degraded environment due to a variety of small particles that cannot be seen or felt individually, but on aggregate diminishes horizontal visibility. Haze is differentiated from fog by its optical density. It has an optical density of less than 2 per km at a wavelength of 0.55 μm while fog has a higher optical density. Haze is made up of water and dust particles that are small (diameter ~0.1 μm) that in effect reduces scattering at longer wavelengths. The optical density of Haze reduces at longer wavelength. As such there is a marked advantage to imaging through haze with SWIR sensors, which have longer wavelengths over VIS sensors whose shorter wavelengths result in much higher optical density and therefore difficulty penetrating though haze as illustrated in Figure 2.

Figure 2: Imaging through Haze with VIS sensor (left) and SWIR sensor (right).

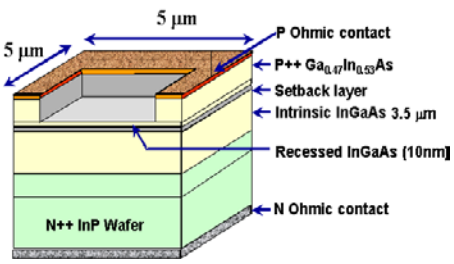
III.3 Imaging Through Fog: Fog is comprised of water particles that are larger than those of haze. In addition, these particles are constantly larger than VIS, NIR and SWIR wavelengths. As such there is little difference by wavelength and therefore not much advantage between VIS and SWIR sensors in penetrating fog since the larger particles scatter light in a manner that results in the same optical density for both bands. However, moving to longer wavelengths, which exceed the large particle diameter, i.e. LWIR sensors are capable of penetrating fog and overcoming the VIS/SWIR limitations.

III.4 Imaging through Clouds: Understanding sensor performance through cloud cover is important especially for sensors operating in high altitude. Cloud has essentially larger water particles, behaves in much the same way as fog except at higher altitudes with the same impact on VIS, NIR

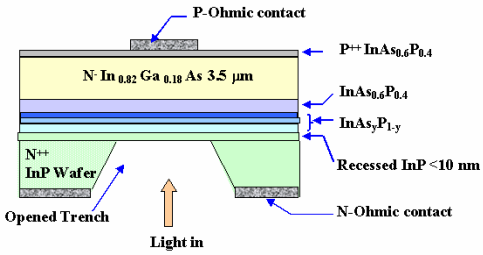
Environmental Conditions	VIS	SWIR	LWIR
Haze			
Fog			
Atmospheric Transmission			
Clouds			
Smoke			
Laser Designators			
Camouflage			
Long Range			

Figure 3: Comparison of VIS, SWIR & LWIR Imaging Capabilities.

and SWIR. However, NIR and SWIR exhibit an advantage over VIS in that they can image around clouds meaning demonstrating the ability to image through the gaps between clouds. Such gaps behave in the same way as haze with smaller particles than the clouds and therefore a marked benefit to SWIR imaging.



(a) Top-illuminated type photodetector



(b) Bottom-illuminated type photodetector

Figure 4: The schematics of a pin photodiode with recessed (a) top contact layer and (b) bottom contact layer for allowing penetration of incoming photons

III.5 Imaging Through Smoke: Smoke is a recognized obscurant used as a battlefield cloaking technique. It is also essential to be able to penetrate smoke as it is a common hazard in forest, industrial and residential fires. The optical densities of smoke are driven by smoke size particles with many approximately or less than 0.1 µm in diameter. The type of smoke also matters; smoke that appears white is mostly scattered light, for example in forest fires burning wood, whereas smoke that appears black is caused by both scattering and absorption of light, for example in burning oil. Due to the particle sizes, the advantage for imaging through smoke both white and black is for longer wavelengths. Black smoke has a higher optical density due to both scattering and absorption, while white smoke has a smaller optical density. Both white and black smoke optical densities reduces significantly with wavelength and a marked advantage and benefit is provided by SWIR sensors over NIR and VIS.

III.6 Imaging Camouflage, Background, and Target Contrast: Another area of interest and differentiation is background contrast, which is important for driving, navigation, pilotage, situational awareness and search. NIR and VIS reflectivity are quite similar; however material reflectivity is very different for SWIR and even further differentiated by environment whether urban or rural. In urban environments, while there are contrast differences between VIS and SWIR, they are overall very similar. On the other hand, in rural environments, vegetation has a much higher reflectivity in SWIR and most images look very bright with very few low reflectivity objects, suggesting contrast not as varied in SWIR as it is in VIS. Figure 3 summarizes the imaging capability at various wavelength bands.

IV. DEVICE STRUCTURES- SWIR

IV.1 Broadband Photodetector Structure: The SWIR imaging device structure consists of *InGaAs* (lattice matched to *InP*) as a absorption layer, and an ohmic contact layer which does not absorb significant number of photons, the photodetector can be used in wide spectral regions covering from near *UV to 1.7 µm*, and can be extended its detection capability to *2.5 µm* by varying the *InGaAs* composition. Figures 4(a) and 4(b) are the cross-sectional views of the top-illuminated and bottom-illuminated type devices, respectively. The devices can offer quantum efficiency (QE) of more than *90%* at *1.6 µm* wavelength. The diode can be designed in circular or square in shape. The buffer layers with graded composition of $InAs_yP_{1-y}$ are placed between absorption layer $In_{0.8}Ga_{0.2}As$ and the substrate *InP*. This dislocations bend at step interface of each graded layer and are disturbed to enter into the next layer. So it is very effective to reduce the dislocation density in the lattice mismatch materials [5], [6]. The epi-layer is grown on a highly doped *n-InP* wafer and the n-ohmic contact is placed at the bottom of *n+ InP* wafer. With varying the InGaAs composition, the proposed detector's sensitivity can be extended to UV to 2.5 µm. If we consider absorption of $In_{0.8}Ga_{0.2}As$ material is 5000 /cm at cut-off of 0.47 eV, this device is estimated to be > 80% at 2.5 µm of illumination for 3.5 µm thick absorption layer. This device can also have lower wavelength absorption into down to 0.4 µm. In effect the sensor device becomes a VIS-SWIR device with multispectral capabilities able to image in all lighting and weather conditions in a single monolithic device.

V. IMAGING DEVICES- MWIR/LWIR

Another alternative to the CCD/CMOS sensor is the capability to image beyond the capabilities afforded by a SWIR image by using materials that absorb longer wavelengths MWIR/LWIR. One such viable material is Mercury cadmium telluride (MCT), which is capable between 1 µm and 12 µm.

V.1 HgCdTe Device Structure: Planar or mesa schemes

are the most frequently used photodiode architectures in HgCdTe detectors for LWIR imaging devices (Figure 5). HgCdTe detectors are typically broadband, and by varying compound composition the spectral band can be selected. Further, if backside substrate is removed, potentially the detector could response to visible light. Generally however, HgCdTe detectors are designed as broad band sensors responding in NIR to VLWIR; 1.7 to 2.5 µm, 3 to 5 µm, 8 to 12 µm and 12 to 17 µm spectral bands. The base active layer generally doped n-type with a concentration ~ 1 x 10^{15} cm^{-3}. The cap layer is doped p-type ~ 5 x 10^{17} cm^{-3} and has a wider bandgap than the base layer (Δx ~ 0.06). Growth of a single type (n-type or p-type) HgCdTe layer permits the fabrication of planar diodes. If the base layer is grown n-type, a group V element such as Arsenic is then used to form the p-type region. Arsenic is implanted in selected regions and annealed to form a planar p+ on n diode architecture [7], [8]. To activate the As-pocket implants regions, annealing has to be done under Hg-overpressure [8]. When both n and p type layers are grown prior to diode delineation, the only diode architecture permitted is a mesa scheme. Speaking briefly to n-on-p cases, when the base layer is grown p-type, boron implantation is typically employed in selected regions to form n-on-p planar diodes [9]-[11]. In both cases, the device structure is known as double layer heterojunction structure (DLHJ).

The choice of detector architecture determines the type

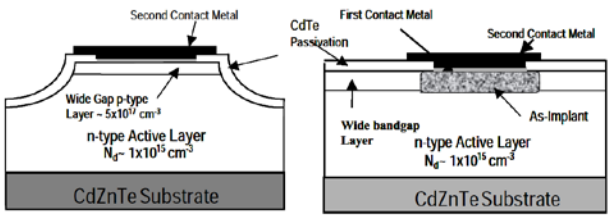


Figure 5: (a) (left) p+ on n mesa diode and (b) (right) p+ on n planar diode.

of dark currents that impact detector performance. The dominant current mechanisms for this specific case are diffusion current from the field-free bulk active layer side, generation-recombination current in the depletion region along with band-to-band or trap-assisted tunneling currents. The double layer heterostructure architecture coupled with recent improvements in passivation techniques have reduced the impact of surface tunneling currents and with the state-of-the-art quality MBE-grown HgCdTe have reduced G-R, tunneling currents [12], [13]. The R_oA product and QE(η) were taken as the figure of merit in order to maximize the detector parameters. Normally, detectors are AR-coated to reduce the reflection losses due to CZT back interface. The reflection loss from the back CZT interface is 21% while from the back of a Si substrate reflection loss is 30%.

VI. HIGH PEFORMANCE MULTISPECTRAL IMAGING SYSTEMS

We at Banpil established a long history of innovation in the optical imaging space having pioneered multispectral imaging through innovations in novel devices that combine visible and infrared imaging in the same monolithic sensor device. Our imaging device innovations and development progression towards next generation image sensors have previously been discussed [14]–[31]. We demonstrated linear arrays and 2-dimensional (2-D) array sensors including devices with micro-nano pillars (nano rods or wires) to increase sensitivity over standard image sensors as well as increased the absorption spectra. We have developed several generations of Focal Plane Array (FPA) with a wide spectral range from 0.35 µm to 1.7 µm, and as already discussed with a target to extend to 2.5 µm.

VI.1 Banpil Camera System Specifications: Banpil's high-performance camera [32] is an extended VGA 640x512 pixels imaging system. It consists of several innovations with foundations in our multispectral imaging technologies comprising of: (1) a high sensitivity FPA with low dark current; (2) a power-efficient multipurpose ROIC that is dynamically reconfigurable in supporting advanced high-sensitivity FPAs; and (3) an advanced reconfigurable multispectral camera with optics, electronics and embedded software**,** which by utilizing the innovations of the Banpil FPA and ROIC components could support applications with capabilities in both day and night without illumination, in all weather and visibility conditions, and all at optimally minimum SWaP-C. These characteristics and its attractive low-cost scheme position the Banpil camera system for dual-use applications in commercial and industrial sectors.

At the core of Banpil's low-cost VIS-SWIR camera is a fourth-generation Banpil developed high-performance InGaAs image sensor chip hybridized with a CMOS-based ROIC component chip with additional image processing functions integrated into the camera system specifically for use in high-performance applications. The Banpil camera provides features with emphasis on high sensitivity, high-speed (>1000 FPS), and high dynamic range (190 dB) that differentiate it from current SWIR cameras. It boasts a high-sensitivity over a broad spectral range (0.35 - 1.7 µm optionally extendable to 2.5 µm) with the ability to support poor visibility below 0.0001 lux and night vision (moonless night) as a VIS to SWIR camera. SWIR cameras today exclusively cover the SWIR region with the visible spectrum covered separately by conventional cameras. The Banpil VIS-SWIR camera has an exceptionally high dynamic range exceeding 190 dB, allowing unencumbered image processing capabilities in rapidly changing scenes e.g. from high brightness to darkness, and high quantum efficiency (>80%) consistently maintained well beyond the peak.

VII. EMERGING APPLICATIONS

The low-cost high-performance camera with its unique features and capabilities will enhance some existing applications as well as enable some emerging applications in commercial sectors. Multispectral imaging such as VIS-SWIR has significant advantages over conventional low light level imaging approaches, providing increased sensitivity and capturing unique target characteristics associated with the SWIR spectral band and thereby enhancing human vision. For instance, a SWIR camera exploits night glow to illuminate objects with great clarity on moonless nights. In addition to sensitivity advantages, SWIR imaging adds the capability to convert illumination outside the spectral range of conventional low light level devices; the capability to detect hidden and camouflaged objects as highlighted earlier.

A low-cost high performance Banpil camera that is optimized for small form factor, weight and power (SWaP-C), will open up a new class of sensors to the intelligence, surveillance, and reconnaissance (ISR) capabilities of the warfighter in defense applications. Due to a procurement-

friendly price-point, virtually every warfighter in the field will be able to exploit enhanced imaging without the burden and exclusivity that limits them today.

In commercial and industrial applications, it is already understood that security, automotive, and medical imaging are beneficiaries of multispectral imaging. The key to adoption remains cost. The autonomous vehicle of the future and even the passive and active driver safety and assistance features of the present will become more robust with VIS-SWIR sensors that overcome the limitations of CMOS-based imaging with a low-cost high performance camera. Finally, there are numerous emerging applications such as aerial photography for agricultural crop management, low earth orbit remote sensing for environmental monitoring, less destructive, cost effective, and safer natural resource exploration methodologies, and many others that are ideal applications, but would not be traditional users of the current high-priced multispectral imaging technologies. These applications collectively will support a symbiotic relationship with the low-cost high performance camera as they will be growth sectors in years ahead which will en masse help drive down costs to be competitive with CCD/CMOS.

VIII. Challenges and Opportunities

In our analysis of the cost structure of the camera and the trends we see emerging with rising applications and increased volume, image sensor hybridization stands out as an obstacle to cost reduction that is yet to be addressed fully. The hybridization cost component remains high even as we achieve significant cost advantages on key component costs. In fact, we estimated that successful achievement of transitioning to high density devices, such as a 5 µm pixel pitch VGA device, will reduce components cost so much so that the hybridization cost will exceed component costs on a percentage basis. Our own R&D into improving the hybridization process in conjunction with our hybridization partners reveals that it is ripe for more automation and additional innovation to bring down costs.

There is an opportunity also in smaller pixel and larger format arrays. Banpil is already conducting R&D into 5 µm devices [18]. Even with larger formats such as 1280x1024 pixels to provide higher resolution imaging over VGA, there are still an increased number of die per wafer which provides a cost advantage. Additional innovation is required to balance the tradeoffs to preserve functionality at the pixel level with such a tight pixel area to maintain high performance and SWaP-C.

IX. CONCLUSION

We explained various imaging devices which could extend human vision, mainly limited to the visible spectrum, which is where standard solid state image sensors (CCD/CMOS) operate. By capturing wavelengths such as NIR, SWIR, MWIR, and LWIR (beyond human vision), these imaging devices based on InGaAs and HgCdTe can provide visibility through different spectral signatures which could not be possible by human vision. The advantages and therefore the value of these imaging capabilities based on several environmental conditions have been shown. We at Banpil Photonics have been developing VIS-SWIR and LWIR image sensors and their imaging systems through innovating novel imaging devices to achieve high performance. By adopting the innovations, techniques and strategies presented, we have successfully developed a SWaP-C camera system with functionalities that will provide a path to broader commercial adoption, more emerging applications, and ultimately enable significant strides in enhancing imaging beyond human vision.

X. Acknowledgements

The authors and Banpil Photonics, Inc., would like to acknowledge the significant contribution and support of DARPA under DARPA/MTO Contract No. D10PC20063, also the Office of Army Research Laboratories / FiberTek under letter subcontractor numbers 224064-51611 and 224080-070975, and also in part supported by the NSF under Grant No. OII-0637454.

References

[1] E. Mounier, "Microbolometers: a market perspective," Proceedings of SPIE Vol. 8252, 825210 (2012).

[2] B. Ackland and A. Dickinson, “Camera on a chip,” in ISSCC Dig. Tech. Papers, pp. 22-25, 1996.

[3] R. G. Driggers, V. Hodgkin, and R. Vollmerhausen, "What good is SWIR? Passive day comparison of VIS, NIR, and SWIR", Proc. SPIE 8706, 87060L (2013)

[4] P. Villeneuve, A. Stocker, and A. Schaum, “VNIR/SWIR Trade Study: Hyperspectral Detection at Long Slant Range (U)” Military Sensing Symposium (2001).

[5] P. Poulin et al. “InGaAs photodiodes for the 1.0 to 2.4 micron spectral region prepared by low pressure MOCVD” in Proc. Int. symp Gallium Arsenide and Related Compounds {Inst. Phys.Conf. Ser No. 74} 1984 ch. 6.

[6] K. Makita et. al., “InGaAs/InAsP pin photodiodes for long wavelength regions grown by hybrid vapor phase epitaxy,” Electron Letters., **vol. 24**, p. 379, 1988.

[7] S. H. Shin, J. M. Arias, M. Zandian, J. G. Pasko, L. O. Bubulac and R. E. DeWames, 1993, J. Electron. Mater., **22**, 1039.

[8] J. M. Arias, J. G. Pasko, M. Zandian, L. J. Kozlowski and R. E. DeWames, 1994, Opt. Eng. Vol. **33**, 1422.

[9] Y. Nemirovsky and D. Rosenfeld, 1990, J. Vac. Sci. Tech. A, **V. 8**, No.2, 1159.

[10] H. G. Robinson, D. H. Mao, B. L. Williams, S. Holander-Gleixner, J.E. Yu, C.R. Helms, 1996, J. Elect. Mat. **V 25**, 1336.

[11] J. P. Zanatta, P. Ferret, R. Loyer, G. Petroz, S. Cremer, J. P. Chamonal, P. Bouchut, A. Million and G. Destefanis, 2000, Proc. SPIE **4130**, 441.

[12] P. S. Wijewarnasuriya, M. Zandian, J. Phillips, D. Edwall, R. E. DeWames, G. Hildebrandt, J. Bajaj, J. M. Arias, A. I. D’Souza, and F. Moore; 2002, J. of Electronic Materials, **31**, 726.

[13] J. Bajaj, 2000, Proc. SPIE, Vol. **3948,** 42.

[14] A. K. Dutta, R. Sengupta, A. Krishnan, S. Islam, P. S. Wijewarnasuriya, N. K. Dhar, "Broadband image sensors for Biomedical, Security, and Automotive Applications," Proceedings of SPIE Vol. 6769, pp.67690C-67690C-12 (2007).

[15] A. K. Dutta, N. K. Dhar, P. S. Wijewarnasuriya et al., "Novel Multi-Color Image Sensors for Bio-Chemical, Bio-Medical, and Security Applications," Proceedings of SPIE Vol. 6008, pp.600812 (2005).

[16] A. Lange, R. Olah, A. K. Dutta, N. K. Dhar, "Optimization of broadband detector for optical communication: design-simulation and fabrication", Proceedings of SPIE Vol. 8283, 82830B (2011).

[17] A. K. Dutta, R. Olah, G. Mizuno, and N. K. Dhar, "Novel Multicolor Photodetectors for Short- and Long-Distance Optical Communication," Proceedings of SPIE Vol. 7959, pp. 795906 (2011).

[18] T. Lin, R. Olah, A. K. Dutta, N. K. Dhar, “Effect of dense planer focal plane array on device performances,” Proceedings of SPIE Vol. 9100, 91000K (2014).

[19] A. K. Dutta, R. Sengupta, A. Krishnan, S. Islam, N. K. Dhar, "Broadband Detector for Optical Communication," Proceedings of SPIE Vol. 6775, pp. 67750G-67750G-12 (2007).

[20] A. K. Dutta, and M. S. Islam, "Novel Broadband Photodetector for Optical Communication," Proceedings of SPIE Vol. 6014, pp. 60140C-60140C-12 (2005).

[21] US Patent # 7,115,910, *Multicolored Photodiode Array and Method of Manufacturing Thereof,* -by A. K. Dutta, Assignee Banpil Photonics, Issued Oct. 3, 2006.

[22] US Patent # 7,535,033. *Multicolored Photodiode Array and Method of Manufacturing,* -by A. K. Dutta, Assignee Banpil Photonics, Inc. Issued May 19, 2009.

[23] US Patent # 7,972,885, *Broadband Imaging Device and manufacturing thereof,* -by A. K. Dutta and R. Olah, Assignee Banpil Photonics, Issued July 5, 2011.
[24] US Patent # 8,035,184, *Broadband Imaging Device*s, -by A. K. Dutta and R. Olah, Assignee Banpil Photonics, Inc. Issued Oct. 11, 2011.
[25] US Patent # 8,101,971, *Multicolored Photodiode Array and Method of Manufacturing Thereof,* -by A. K. Dutta, Assignee Banpil Photonics, Issued Jan. 24, 2012.
[26] US Patent # 8,304,759, *Integrated Image Sensor System on Common Substrate,* -by A. K. Dutta, Assignee Banpil Photonics, Issued Nov. 6, 2012.
[27] US Patent # 8,357,960, *Broadband Image Sensor and Manufacturing Thereof,* -by A. K. Dutta, Assignee Banpil Photonics, Issued Jan. 22, 2013.
[28] US Patent # 7,662,659, *Methods of forming arrays of nanoscale building blocks,* -by A. K. Dutta, Assignee Banpil Photonics, Issued Feb 16, 2010
[29] US Patent # 7,922,976, *High sensitivity sensor device and manufacturing thereof,* -by A. K. Dutta, Assignee Banpil Photonics, Issued Apr 12, 2011
[30] US Patent # 8,174,059, *Multicolor photodiode array and method of manufacturing,* -by A. K. Dutta, Assignee Banpil Photonics, Issued May 8, 2012
[31] US Patent # 8,641,975, *High sensitivity sensor device and manufacturing thereof,* -by A. K. Dutta, Assignee Banpil Photonics, Issued Feb 4, 2014
[32] P. Oduor, G. Mizuno, R. Olah, A. K. Dutta, *"Development of low-cost high-performance multispectral camera system at Banpil*", Proceedings of SPIE Vol. 9100, 910006 (2014).

Tunable Slow Light with Graphene based Hyperbolic Metamaterial

Ayed Al Sayem, [1*] M.R.C. Mahdy, [2**] D.N.Hasan, [2] and Md. Abdul Matin[1]

[1]Dept. of EEE, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh

[2] Department of Electrical and Computer Engineering, National University of Singapore, 4 Engineering Drive 3, Singapore

*ayedalsayem143@gmail.com

**elemrc@nus.edu.sg

Abstract— **Slow light has always been a topic of extreme interest for researchers and scientists in order to utilize the full potential of light in switching, memory devices and in quantum optics. However, slow light phenomena and its potential applications in switching and memory devices by tunable graphene based hyperbolic metamaterial have not still been introduced in literature. Here we theoretically propose and numerically analyze a graphene based tunable slow light device from the concept of controlling the group velocity of light in hyperbolic metamaterials. In THz range, graphene-dielectric stack can behave as hyperbolic metamaterials and hyperbolic metamaterials can demonstrate slow light phenomena while being cladded by air. By utilizing the tunability option of graphene, proposed devices can be made tunable, which is the most important criteria to manage the full advantage of slow light. It ultimately paves a fully new way to find novel applications in photonic switch, optical buffers and memory devices. Interestingly, in such graphene based devices, voltage will play a vital role to control the group velocity of light from slow to fast and the carriers are photons instead of electrons.**

***Index Terms*—Graphene, Slow light, Hyperbolic metamaterial**

I. Introduction

Slow light has always been a topic of extreme interest for researchers and scientist in order to utilize the full potential of light in switching, memory and quantum optics, where the full control of the speed of light is necessary [1]. In such devices, the term 'slow light' means light with very small group velocity [1]. On the other hand, in last ten years, hyperbolic metamaterials (HMMs) [2] have shown their extreme aptness and applications in plasma physics[3], exhibiting negative refraction [4], negative index [5], sub-wavelength imaging [6], nano-scale wave guiding [7], light confinement [8, 9], nonlinearity enhancement [9], biosensing [10] etc. And recently HMM has been introduced in solar application by using the slow light concept [11, 12]. Graphene which is a single atom thick two dimensional plane of carbon atoms forming a hexagonal lattice layer with a lattice constant of 0.264 nm [13] has manifested some fascinating electrical properties and devices[14, 15]. Now graphene is getting seriously engaged in enthralling applications of photonics and plasmonics, last of them being inaugurated in constructing HMM replacing metal [16, 17]. Due to tunability option of graphene, all the fascinating properties supported by HMMs are supposed to be boosted by a marginal range, which in recent literature is already being reflected [16, 17, 18]. Surprisingly slow light phenomena and its potential applications in switching and memory devices by graphene based hyperbolic metamaterial have not still been introduced in literature to the best ofour knowledge. Here in this paper, we describe such an enthralling physics and working principle of a novel tunable slow light device, where the device is controlled by voltage and the carrier is photon instead of electron. Such a device may find exciting application in photonic switch and memory devices, where voltage will play the role to control the speed of light from slow to fast.

II. Theory

There are two ways to engineer hyperbolic metamaterial structure: a) Multilayer realization consisting of alternating sub-wavelength metal and dielectric layers. b) Metal nanowires in a dielectric host. Uniaxial anisotropic metamaterial can be identified by the permittivity tensor.

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_{||} & 0 & 0 \\ 0 & \varepsilon_{||} & 0 \\ 0 & 0 & \varepsilon_{\perp} \end{bmatrix}$$

These uniaxial materials can show hyperbolic dispersion if they satisfy the condition $(\varepsilon_{||}\varepsilon_{\perp}) < 0$

For metal dielectric alternating metamaterial structure, as long as the length of one unit cell consisting of one metal and one dielectric layer is less than the wavelength of light used, we can employ effective medium theory.

By effective medium theory we get,

$\varepsilon_{\parallel} = f\,\varepsilon_m + (1-f)\varepsilon_d$ where $\varepsilon_{\parallel}$ is the effective permittivity in transverse palne and $\varepsilon_{\perp} = \left(\frac{f}{\varepsilon_m} + \frac{(1-f)}{\varepsilon_d}\right)^{-1}$ is the effeticve permittivity in propagation direction and $f = \frac{t_m}{t_d + t_m}$ is the filling ratio where t_m is the thickness of metal layer and t_d is the thickness of dielectric layer.

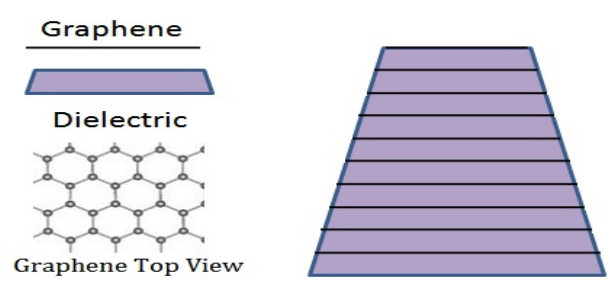


Fig. 1 Graphene-dielectric multilayer hyperbolic metamaterial considered in this work.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

In case of graphene, at THz frequency range intra-band transition dominates. The effective permittivity of graphene can be obtained by two different ways. Frequency and chemical potential dependent conductivity of a monolayer graphene can be given by [16],

$$\sigma(\omega,\mu_c) = -\frac{j4\pi e^2 k_B T}{h^2(\omega - j2\mathcal{T})}\left(\frac{\mu_c}{k_B T} + 2\ln\left(e^{\frac{-\mu_c}{k_B T}} + 1\right)\right) - \frac{j4\pi e^2(\omega - j2\mathcal{T})}{h^2}\int_0^\infty \frac{f_D(-\xi) - f_D(\xi)}{(\omega - j2\mathcal{T})^2 - 16\left(\frac{\pi\xi}{h}\right)^2} d\xi \quad (1)$$

Where $f_D(\xi)$ is the Fermi Dirac function which can be given by, $f_D(\xi) = \frac{1}{e^{\frac{\xi-\mu_c}{k_B T}}+1}$

The permittivity of graphene dielectric metamaterial can be given by,

$$\varepsilon_\parallel = \varepsilon_\parallel' - j\varepsilon_\parallel'' = \varepsilon_d - j\frac{\sigma(\omega,\mu_c)}{\omega\varepsilon_o t_g} \quad (2)$$

$$\varepsilon_\perp = \varepsilon_d \quad (3)$$

Again the effective permittivity can be derived in following form effective medium approach,

$$\varepsilon_g = 1 - j\frac{\sigma(\omega,\mu_c)}{\omega\varepsilon_o t_g} \quad (4)$$

$$\varepsilon_\parallel = \frac{t_g\varepsilon_g + t_d\varepsilon_d}{t_g + t_d} = \varepsilon_d - j\left(\frac{\sigma(\omega,\mu_c)}{\omega\varepsilon_o t_d}\right);\ t_g \ll t_d \quad (5)$$

$$\varepsilon_\perp = \frac{\varepsilon_g\varepsilon_d(t_g + t_d)}{t_g\varepsilon_d + t_d\varepsilon_g} \cong \varepsilon_d;\ t_g \ll t_d \quad (6)$$

Where k_B is Boltzmann constant, T is temperature, μ_c is chemical potential of graphene , h is plank's constant, , ω is angular frequency, $\mathcal{T}$ is phenomenological scattering rate (.1meV) , t_g is the thickness of graphene layer, t_d is the thickness of dielectric layer and ε_d is the permittivity of dielectric. As we can see, in graphene based HMM, $\varepsilon_\perp$ is always positive, so to satisfy $(\varepsilon_{||}\varepsilon_\perp) < 0$ condition, $\varepsilon_\parallel$ has to be negative which can be obtained at THz frequency for graphene based HMM. So in the THz frequency range, we can expect graphene-dielectric stack to act as a hyperbolic metamaterial.

As conductivity of single layer graphene is dependent on frequency, chemical potential and temperature, permittivity of a single layer graphene and consequently the parallel permittivity of graphene-dielectric anisotropic metamaterial possess the same dependence.

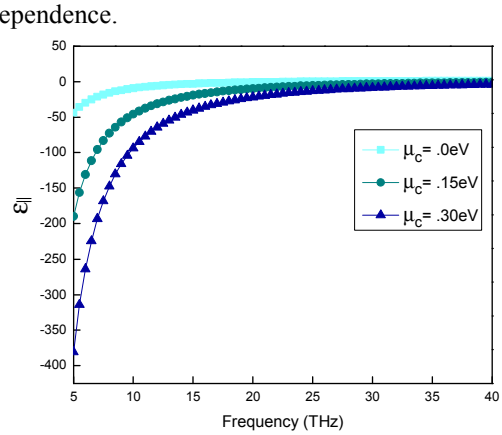


Fig. 2 Parallel permittivity (real part) of graphene-dielectric HMM for different chemical potential at different frequency.

Fig.2 shows the parallel permittivity of the anisotropic structure at different chemical potential vs different frequencies. As longitudinal permittivity of the structure is always positive, to obtain hyperbolic dispersion and also for securing slow light condition, parallel permittivity of the structure has to be negative. Fig.3 shows the permittivity of single layer graphene.

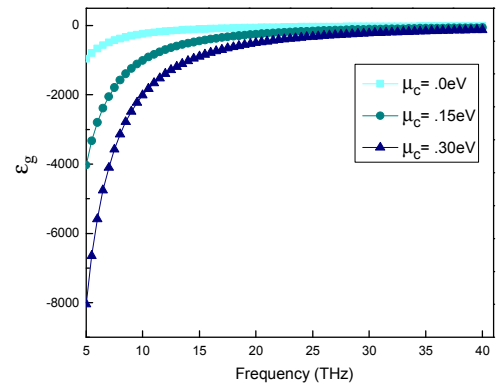


Fig. 3 Permittivity of graphene (real part) for different chemical potential at different frequency.

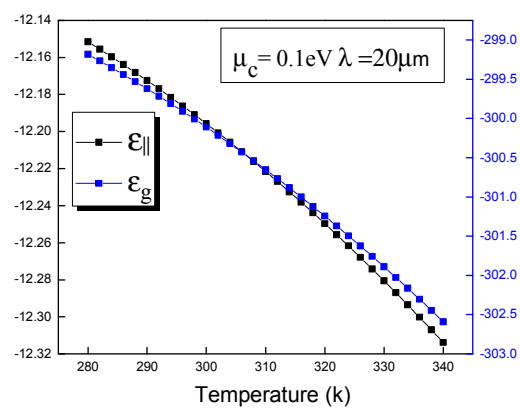


Fig. 4 Temperature dependent permittivity (real part) and parallel permittivity (real part) of graphene and graphene based HMM respectively.

Fig.4 shows the temperature dependence of parallel permittivity of the structure. To control slow light for switching effect, temperature might be used as a tuning mechanism but because of low temperature dependence of parallel permittivity, temperature controlled switch might not be possible. But for high temperature dependent signal generation, this structure might be used.

The most important parameter to change the parallel permittivity of the structure is chemical potential which can be changed by external dc field or gate voltage [18]. In this paper we will discuss the condition of slow light with respect to chemical potential; at last adopting the relation from [18] with gate voltage and chemical potential will show a relation of absorption and transmission with respect to gate voltage.

A waveguide filled with hyperbolic anisotropic metamaterial cladded by air [11, 12] or an air waveguide cladded by anisotropic metamaterial [19] can both exhibit slow light phenomena. It has been revealed that light can come to complete

standstill in hyperbolic metamaterial wave guide if the width of the waveguide is tuned to a certain value for a specific frequency or wavelength. This slow light phenomenon has been used to absorb ultra-broadband solar spectrum [11]. Necessary condition for complete standstill of light is identified by the condition when two propagating constants (forward and backward) come to a degeneracy point. For different wavelengths, the waveguide width at which light will be stopped will be modified. Solving a uniaxial anisotropic waveguide one can obtain the following equation for supporting even and odd TM modes respectively,

$$\frac{\varepsilon_{2z}k_1}{\varepsilon_1 k_2} = \tan\left(\frac{k_2 W}{2}\right) \; for\ even\ modes \qquad (7)$$

$$\frac{\varepsilon_{2z}k_1}{\varepsilon_1 k_2} = -\cot\left(\frac{k_2 W}{2}\right) \; for\ odd\ modes \qquad (8)$$

Where

$$k_1 = (\beta^2 - k_\omega^2 \varepsilon_{air})^{\frac{1}{2}}$$

$$k_2 = \left(k_\omega^2 \varepsilon_{2z} - \frac{\varepsilon_{2z}}{\varepsilon_{2\parallel}}\beta^2\right)^{\frac{1}{2}}$$

$$k_\omega = \frac{\omega}{c}$$

$\varepsilon_{air} = relative\ permeticity\ of\ air$, $\varepsilon_{2z} = \varepsilon_d\ from$ (6), $\varepsilon_{2\parallel} = \varepsilon_\parallel\ \ from$ (5), W is width of the waveguide.

In this work, we clearly show that slow light is not only dependent on the width of the waveguide but also the parallel and longitudinal permittivity. But why they were not mentioned so strictly in pervious works of slow light using hyperbolic metamaterial? Simply because using metal and dialectic, those values come naturally; there is no way to change to permittivity of the metal or dielectric at a certain frequency because every metal or dielectric has a specific permittivity at a specific wavelength. Though there is one option to tune the structure, which is the filling ratio of metal and dielectric but once the structure is fabricated there is no way to change the permittivity of metal or dielectric or width of the structure or filling ratio. So a hyperbolic structure built to stop light at a specific frequency or wavelength at a specific width will always stop light at that specific width of that specific wavelength or frequency. Light of that specific wavelength cannot simply pass through structure. But using graphene we can tune the permittivity of the structure after fabrication, why? Because permittivity of graphene is simply dependent on conductivity and that conductivity is dependent on chemical potential which can be tuned by external electric field or dc voltage. Though tuning the chemical potential might be difficult in this complicated structure (but by no means impossible [18]). In addition, we don't need a huge variation in chemical potential to disobey the condition for slow light. It means that a graphene-dielectric structure, which is tuned to stop light of specific wavelength, we can let that light to be a transmitting one by changing chemical potential. Thus we can make a photonic memory switch. So basically we can control slow light at our will.

III. Discussion and Results

Fig.5 demonstrates the temperature dependent slow light. It can be observed that the width of the waveguide which is required to achieve slow light is changed by temperature slightly. This is because the conductivity of graphene has weak temperature dependence which is indicated in fig.4 and consequently the parallel permittivity of the structure is also very slightly temperature dependent. So we conclude that temperature can't be a control mechanism for tuning slow light. So it is veritably not practicable to realize a photonic memory or switch with

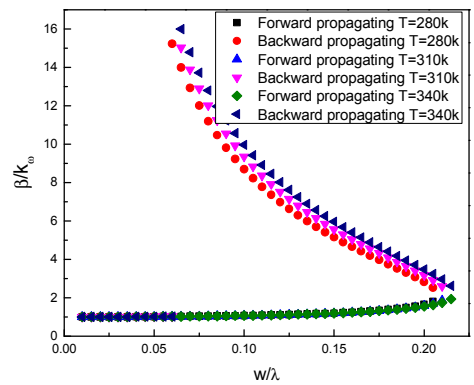


Fig. 5 Temperature dependent Slow Light , normalized propagation constant $\boldsymbol{\beta}$ as a function of waveguide width w for different temperature (T).

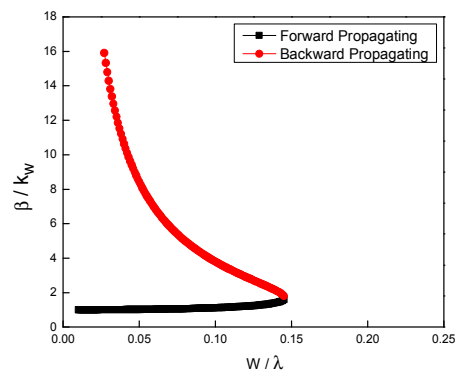


Fig. 6a Slow Light for Graphene based HMM, normalized propagation constant β as a function of waveguide width w. Light is stopped at a certain width where forward and backward propagating wave coincides, at zero chemical potential the permittivity of the structure is such that light stops at width $w = \sim.15\lambda$ where $\lambda =$ wavelength of the incident light.

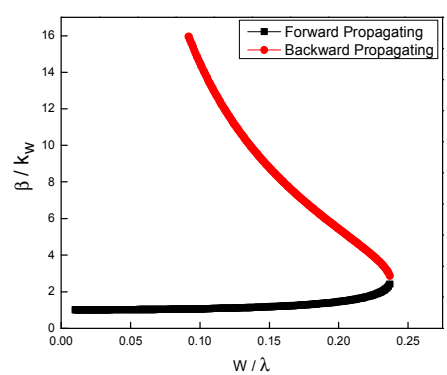


Fig. 6b Slow Light for Graphene based HMM, normalized propagation constant β as a function of waveguide width w. Light is stopped at a certain width where forward and backward propagating wave coincides, at .2eV chemical potential the permittivity of the structure is such that light stops at width $w = \sim.235\lambda$ where $\lambda =$ wavelength of the incident light ,thus violating the condition for slow light for zero chemical potential.

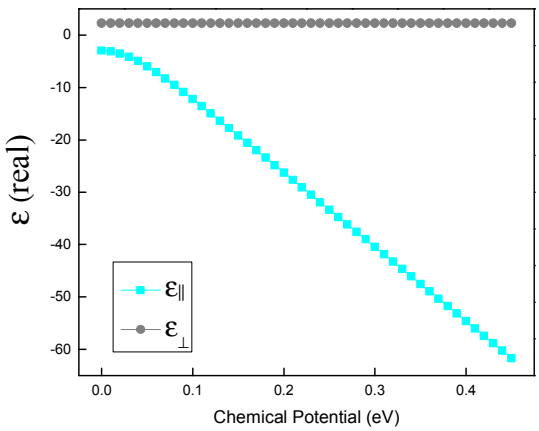


Fig. 7 Parallel and longitudinal permittivity (real part) of Graphene-dielectric metamaterial structure at different chemical potential.

temperature control but it may find application where temperature sensitive light device is obligatory.

Fig.6a and 6b clearly shows the chemical potential dependent slow light in graphene based HMM. At first we want to discuss the slow light mechanism for unbiased graphene or with zero chemical potential. In fig.5a the black and red line represents the forward and backward propagating waves respectively. When $\frac{\text{waveguide width}}{\text{wavelength of light}} = \sim.14$ light is stopped at that width. In the above example, we are using light of wavelength of $20\mu m$, then the width of the waveguide must be at least $20\mu m \times .14 = 2.8\mu m$ to achieve slow light condition. So now if we can change the permittivity of the structure then the width at which light would be trapped will be changed. Changing chemical potential of graphene will change the permittivity (parallel) (fig. 7) of the graphene-dielectric HMM structure. So the change of chemical potential will shift the trapped position to higher wave guide width. From fig.6b for .2eV chemical potential, for light to be trapped or stopped, the necessary condition is $\frac{\text{waveguide width}}{\text{wavelength of light}} = \sim.235$. So using previous wavelength of light of $20\mu m$, the width of the waveguide must be at least $20\mu m \times .235 = 4.7\mu m$

Now let us consider a rectangular waveguide which has the necessary width to trap light of a specific wavelength at 0 chemical potential. If we change the chemical potential from 0 to any higher value, then the waveguide doesn't have the required width to trap light. So the condition for light to be trapped is not satisfied and so light will be transmitted. Thus the device can act as a switch or photonic memory device.

IV. Conclusions

In this work, we have briefly pointed out the possibility of tunable slow light by using a graphene based hyperbolic medium. As graphene is actively tunable, the device is also tunable. Slow light is itself a fascinating phenomenon. Active tunability in order to achieve complete stopping of light and letting it transmit again via gate voltage may pave the way to future voltage controlled phonotic memory devices. The control mechanism will be gate voltage and the carriers will be photons. Group velocity of the photons will be controlled by gate voltage. In future, such devices may ultimately pave a fully new way to find novel applications in photonic switch, optical buffers and memory devices replacing current technologies.

References

[1] Thomas F. Krauss, "Why do we need slow light? ," *Nature Photonics* , Vol 2, (2008)

[2] Jacob Z, Alekseyev L V, Narimanov E 2006 Optics Express 14 8247.D. R. Smith and D. Schurig, "Electromagnetic wave propagation in media with indefinite permittivity and permeability tensors," *Phys. Rev. Lett.* 90(7), 077405 (2003).

[3] L. Felsen. "Focusing by an anisotropic plasma interface. Antennas and Propagation," *IEEE Transactions on*, 12(5):624-635, 1964

[4] D.R. Smith, P. Kolinko, and D. Schurig. "Negative refraction in indefinite media," *JOSA B*, 21(5):1032-1043, 2004.

[5] V.A. Podolskiy and E.E. Narimanov, "Strongly anisotropic waveguide as a nonmagnetic left-handed system," *Physical Review B*, 71(20):201101, 2005.

[6] Z. Jacob, L.V. Alekseyev, and E. Narimanov, "Optical hyperlens: Far-field imaging beyond the diffraction limit," *Optics Express*, 14(18):8247{8256, 2006.

[7] A.A. Govyadinov and V.A. Podolskiy. Metamaterial photonic funnels for sub-diffraction light compression and propagation. *Physical Review B*, 73(15):155108, 2006.

[8] J. Yao, X. Yang, X. Yin, G. Bartal, and X. Zhang. "Three-dimensional nanometer-scale optical cavities of indefinite medium," *Proceedings of the National Academy of Sciences*, 108(28):11327, 2011.

[9] GA Wurtz, R. Pollard, W. Hendren, GP Wiederrecht, DJ Gosztola, VA Podolskiy, and AV Za-yats. "Designed ultrafast optical nonlinearity in a plasmonic nanorod metamaterial enhanced by nonlocality," *Nature nanotechnology*, 6(2):107-111, 2011

[10] AV Kabashin, P. Evans, S. Pastkovsky, W. Hendren, GA Wurtz, R. Atkinson, R. Pollard, VA Podol-skiy, and AV Zayats, "Plasmonic nanorod metamaterials for biosensing. *Nature materials*," 8(11):867-871, 2009.

[11] Haifeng Hu, Dengxin Ji, Xie Zeng, Kai Liu Qiaoqiang Gan, "Rainbow Trapping in Hyperbolic Metamaterial Waveguide," *Scientific Reports* 3, Article number: 1249.

[12] Yanxia Cui,Kin Hung Fung,Jun Xu,Hyungjin Ma,Yi Jin,Sailing He and Nicholas X. Fang, "Ultrabroadband Light Absorption by a Sawtooth Anisotropic Metamaterial Slab," Nano Lett., 2012, 12 (3), pp 1443–1447

[13] A. H. Castro Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, and A. K. Geim, "The electronic properties of graphene," *Rev. Mod. Phys.* 81, 109 (2009)

[14] Frank Schwierz,"Graphene transistors," *Nature Nanotechnology* 5, 487–496 (2010)

[15] Thanasis Georgiou, Rashid Jalil, Branson D. Belle, Liam Britnell, Roman V. Gorbachev, SergeyV.Morozov, Yong-Jin Kim, Ali Gholinia, Sarah J. Haigh, Oleg Makarovsky, Laurence Eaves, Leonid A. Ponomarenko, Andre K. Geim, Kostya S. Novoselov & Artem Mishchenk, "Vertical field-effect transistor based on graphene–WS_2 heterostructures for flexible and transparent electronics," *Nature Nanotechnology* 8, 100–103, (2013).

[16] Mohamed A. K. Othman, Caner Guclu, and Filippo Capolino, "Graphene-based tunable hyperbolic metamaterials and enhanced near-field absorption", *Optics Express*, Vol. 21, No. 6.

[17] K. V. Sreekanth, A. De Luca, and G. Strangi, "Negative refraction in graphene-based hyperbolic metamaterials," *Applied Physics Letters* 103, 023107 (2013)

[18] Bofeng Zhu, Guobin Ren, Siwen Zheng, Zhen Lin and Shuisheng Jian,"Nanoscale dielectric-graphene-dielectric tunable infrared waveguide with ultrahigh refractive indices," *Optics Express*, Vol. 21, No. 14 (2013).

[19] Tian Jiang, Junming Zhao and Yijun Feng, "Stopping light by an air waveguide with anisotropic metamaterial cladding," *Optics Express*, Vol. 17, Issue 1, pp. 170-177 (2009).

Quantum Mechanical Modelling of Steady State Photoconductive Behaviour of AlGaAs/GaAs QWIP at Varying Bias and Optical Density

Soumitra Roy Joy[1]* and Farseem Mannan Mohammedy[1]
[1]Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
*joy.soumitra@gmail.com

***Abstract*—A quantum mechanical approach of first order born approximation is taken to numerically model AlGaAs/GaAs Quantum Well Infrared detector that operates at long wavelength region (8-9μm). Our numerical analysis of the photocurrent spectra as well as the broadening of the same at 8.31μm peak wavelength is found to be in close resemblance with the experimental result with 8.55μm peak wavelength as reported by Alves et. al. We have gone one step further by subsequently exploring the photo-response behavior of QWIP at the limit of strong electron-photon interaction and found that, besides the modification in photo-current spectrum in terms of relative magnitude and spectral shift, the strong non-linear photo-response enforced by light-matter coupling possesses the potential of accommodating multi-color detection mechanism under moderate bias and high optical density in active region.**

***Index Terms*—Quantum Well Infrared Detector, photo-current spectrum, electron-photon interaction, NEGF model**

I. Introduction

From thermal imaging to medical diagnostics, and in commercial and military application, infrared detection technology has been a vibrant topic for perpetual research. Since late '90s or so, Quantum Well infrared detector has successfully emerged to be one of the strongest competitors of the widely used HgCdTe (MCT) photodetector in this infrared detection regime [1]. An active research is going on to bring forth the high sensitivity, high selectivity and multispectral capability that the Quantum Well interband and intersubband detector promises for detection and identification of specific emission signatures. Among the different types of quantum well infrared photodetectors (QWIPs), the technology of GaAs/AlGaAs multiple-quantum-well (MQW) detectors is the most mature [2].

In this paper, we have formulated a numerical model to investigate the photocurrent spectrum of AlGaAs/GaAs intersubband Quantum Well photoconductor subjected to a widely varied degree of light-matter mutual coupling. The photoconductive response, which is considered to maintain a linear relation with the strength of incident light in some other previous work as in [3], may not remain linear when it experiences a continually increasing exposure to optical intensity. In this work, we have also peered into the non-linearity in electrical response and broadening in current spectra invoked by strong light-matter interaction.

The motivation of exploring the possibility of non-linear response of photoconductor primarily emanated from the fact that, researchers today has been showing an increasing interest in using plasmonic/ other mechanism of field enhancement for photodetectors in order to provide strong near-fields to increase light absorption and improve the detector performance in terms of NEDT, dark current and other figure of merits [4]. The behavior of the sensor exposed to condensed light can be significantly different, such is our conviction.

II. Device Description

In this paper, we have analyzed the photoconductive behavior of $Al_{0.26}Ga_{0.74}As/GaAs$ Quantum Well Infrared photodetector, which functions through intersubband transition for long wavelength light (8-9μm). The physical structure of the device is reported in [5], which comprises of 20 periods of active regions sandwiched between thick GaAs contact and buffer layer, as in Fig. 1. We have taken a cross section of the device along the vertical (growth) direction and perform our simulation over a single period of active region, as shown in Fig. 2.

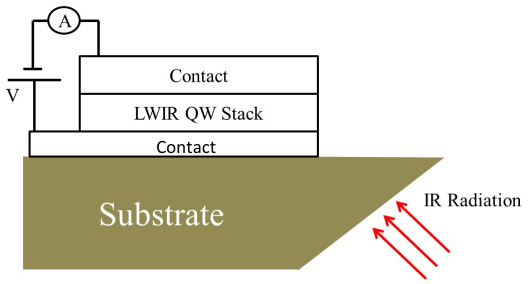


Fig. 1. Schematic diagram of Quantum Well Photo-detector with readout

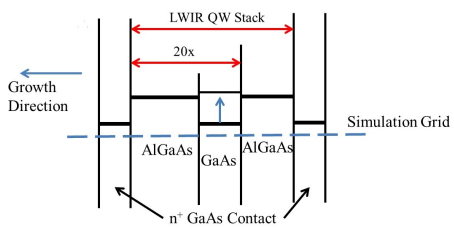


Fig. 2. Active region and the cross-section along which the simulation is performed. Thick lines indicate conduction band offset

III. Quantum Mechanical Modeling

A. Numerical Formulation

The fundamental unit of the periodic layer of active region of QWIP can be described by the Hamiltonian:

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

$$H = H_0 + H_{ep} \tag{1}$$

H_0 gives the description of ballistic transport; it includes the kinetic energy, the band structure and electric potential. H_{ep} contains scattering of electron by photon/ optical phonon. In our model, we adopted a single band tight binding effective mass approximation model [6] where we have kept aside all other kinds of possible scattering, i.e- scattering by acoustic phonon, ionized impurities, interface roughness etc since we would like to keep our interest more to the maxima of the photocurrent spectra and less to its broadening. The operator for electron photon interaction reads:

$$H_{ep} = \frac{e}{m_0} A \cdot p \tag{2}$$

Where the quantized photon field for monochromatic light under standard dipole approximation is given by [7]

$$A = \sqrt{\frac{\hbar\sqrt{\mu\varepsilon}\phi_\omega}{2N_\omega\omega\varepsilon_0}}\widehat{a}(be^{-i\omega t} + b^\dagger e^{i\omega t}) \tag{3}$$

$$\phi_\omega = \frac{N_\omega c}{V\sqrt{\mu\epsilon}} \tag{4}$$

Here $\widehat{a}$ is the polarization unit vector and A_0 is the incident photon flux, N_ω is the number of photon per unit absorbing volume. In the Plane Orbital Basis for a layered system, the Hamiltonian for electron-photon interaction is of the following form:

$$H_{ep} = \sum_{L,L'} M_{L,L'}(be^{-i\omega t} + b^\dagger e^{i\omega t}) \tag{5}$$

$$M_{L,L'} = \frac{e}{m_0} A_0 \langle L|p|L'\rangle \tag{6}$$

$$A_0 = \sqrt{\frac{\hbar}{2\epsilon_0 V\omega}}\widehat{a} \tag{7}$$

L indicates semiconductor layers, and A_0 is the solution of Maxwell's electromagnetic equation, where we take the electromagnetic field of a photon to be a single mode monochromatic plane wave [9].

$$\nabla^2 A_0(x) + \frac{\omega^2}{c^2} A_0(x) = 0 \tag{8}$$

Within the NEGF formalism, the steady state equation of motion for the Green's functions is given by the Dyson's equation [8], [11]:

$$G_R(E) = [(G_0^R)^{-1} - \Sigma_{ep}^R(E) - \Sigma^{RB}(E)]^{-1} \tag{9}$$

$$G_0^R(E) = [E + i\eta - H_0]^{-1} \tag{10}$$

$$G^<(E) = G^R(E)\Sigma^{<B}(E)G^\dagger(E) \tag{11}$$

In this paper, we have employed the numerical model proposed in [12] to accommodate a single inelastic scattering of electron inside the device due to photon absorption which is otherwise subjected to ballistic conduction.

$$(E - H_0 - \Sigma^{RB} - D \otimes G_1^R(E+\omega))G_0^R(E) = 1 \tag{12}$$

$$((E+\omega) - H_0 - \Sigma^{E+\omega})G_1^R(E+\omega) = 1 \tag{13}$$

The self-energy, $D \otimes G_1^R$ in the equation of motion for G_0^R accounts for the broadening and spectrum shift resulting from the out-scattering of state ψ_0 to ψ_1. The electron correlation functions are:

$$G_0^<(E) = G_0^R(E)\Sigma^{<B}(E)G_0^\dagger(E) \tag{14}$$

$$G_1^<(E+\omega) = G_1^R(E+\omega)[D \otimes G_0^<(E)]G_1^\dagger(E+\omega) \tag{15}$$

$$D^<(E) = 2\pi M_{L,L'}[N_\omega\delta(E+\omega)]M_{L,L'} \tag{16}$$

The physical explanation of the equations is as following: A plane wave,ψ_0 , injected from the contact at energy E propagates into the device and due to subsequent photon absorption, may scatter into another quantum state ψ_1 at some higher energy $(E+\omega)$ which have no phase coherence with ψ_0. This secondary wave ψ_1 then propagates ballistically until it exits through the barrier [12].

B. Model Manoeuvring

Determination of photo-induced current using the exact conduction band offset of $AlGaAs/GaAs$ hetero-structure poses a major computational problem. Photo-excitation process involves inelastic scattering of electron to higher electronic state; however, the ground state must be populated for proper functioning of the mechanism. Since the contact acts as the primary source of carrier to the device, no electron can get into any of the energy states that lie below the continuum level of the contact, unless tunnelling/ optical phonon or other types of elastic relaxation process is taken into account to populate ground state.

In this paper, we have made a careful maneuvering over the original hetero-structure conduction band profile so as to circumvent the inclusion of phonon relaxation process and still mimic the photoconductive behavior of the QWELL up to a satisfactory extent without introducing any possible computational artifact. We have lowered the extreme left and right corner of the conduction band effectively leaving a barrier between the QWELL and the extreme left (or right) lead. The width of the barrier is kept marginally enough so as to inject carrier into the ground state of the QWELL. Too wide barrier would completely seal the injection process whereas too thin barrier would result in an overbroadening of ground states and washing out of spectral peaks.

Once injected from contact, the carrier is assumed to be ballistically transported apart from the single inelastic scattering it undergoes for the appropriate wavelength of incident light. No consecutive/multiple scattering is taken into account.

IV. Result and Discussion

Fig. 3 shows the calculated local density of states of $Al_{0.26}Ga_{0.74}As$ Quantum Well structure simulated for unit cell period. It reveals two sharp peaks of energy states at 0.0654eV and 0.2159eV. The latter energy states are posited at the edge of the Quantum Well height ($Al_{0.26}Ga_{0.74}As$ has 0.2282eV band offset over GaAs QW) and hence are called quasi-bound state. The Local Density of States(LDOS) implicates that a sharp photo-absorption spectrum is likely to be found around 8.3 μm incident light wavelength, which is equivalent to the energy difference between two strong density of states (0.22eV-0.065eV=0.15 eV). It may be noted that, the Density of States does not appreciably alter within the applied low bias region.

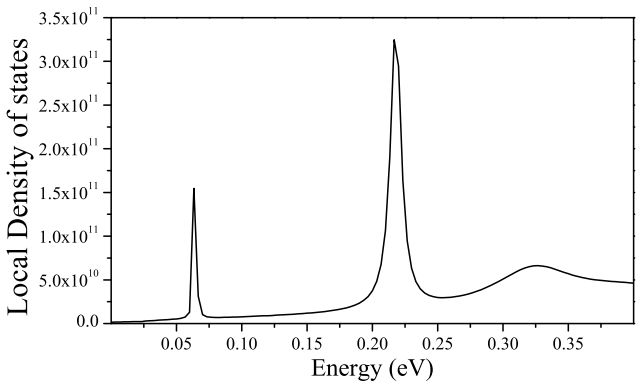


Fig. 3. Local density of states of the AlGaAs/GaAs QWELL structure

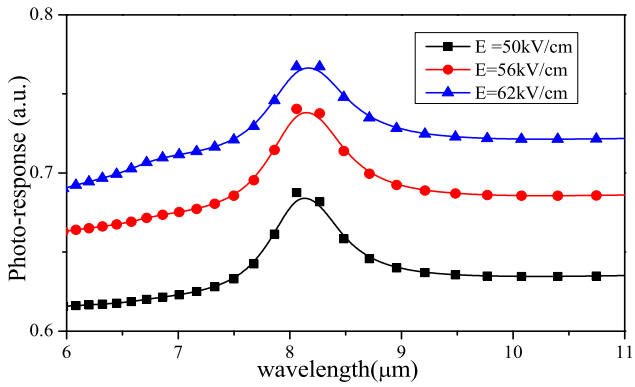


Fig. 5. Photo-current spectra of QWIP structure at various electric field and moderate photon density

A. Comparison between Calculation and Experimental Result

Fig. 4 is the comparison between photocurrent spectrum found by our calculation and that found by experiments in [5](and also in [10]). The close resemblance between the two both in terms of peak wavelength and the FWHM substantiates the validity of our numerical approach. The peak photocurrent from our calculation is found at 8.31μm for moderate photon density, which agrees with the experimentally found peak at 8.55μm within less than 3% deviation. It may be interesting to notice that, the experimental data shows a minor peak at 7.929μm, which is also echoed in our simulation at 7.793μm, with a bit exaggeration, although.

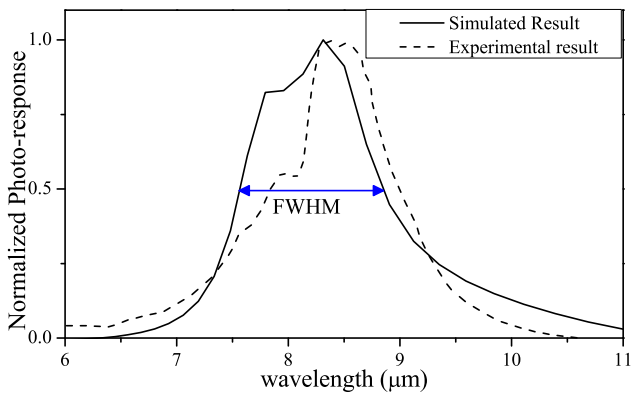


Fig. 4. comparison between our simulation and experimental result from [5]

Fig. 5 shows the photocurrent response of AlGaAs/GaAs QW detector resolved in spectrum under various bias. For moderate photon density, we observe that the position of peak wavelength at 8.3 μm does not get modulated appreciably by the bias voltage; apart from stepwise addition of an offset to the spectrum. Since AlGaAs/GaAs behaves as a photoconductor, the bias voltage is not kept too high, lest the peak of the spectrum should get flattened out.

B. Extrapolation of Model towards strong electron photon interaction limit

Fig. 6 shows the modification in photocurrent spectrum due to increasing amount of photon existing per unit absorbing volume of active region. Since availability of more photon per volume is a consequence of more incident light, the photocurrent peak gets enhanced accordingly. The point to be noted is, along with the rise is photocurrent peak response, the broadening, i.e. the FWHM of the photocurrent spectrum keeps increasing as well. This is certainly as expected, since the broadening of the spectrum indicates reduced lifetime of the carrier in the ground state.

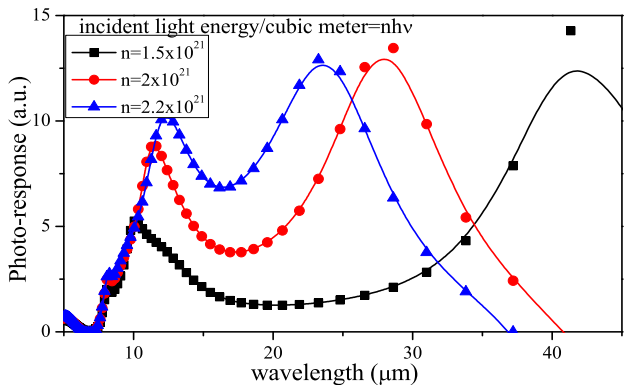


Fig. 6. Change in photocurrent spectra with the increasing number of photon available per unit absorbing volume in high photon density regime

The photo-response, for sufficiently increased amount of photon incidence, does no longer get increased at a proportionate rate, it rather undergoes saturation. What we find more intriguing is the spitting or bifurcation of photocurrent spectra into a major peak (in far-infrared region) and a minor peak (in long infrared region) and their tendency to come closer to each other (the minor peak undergoes red-shift while the major peak sees blueshift) with the increasing number of photon per absorbing volume. This phenomenon of spectral splitting is otherwise inexplicable if the dynamic re-structuring of energy bands inside the active region ensued by the change in optical density inside active device detector is not taken into account. If not operated under strong electron-photon interaction, this multi-color spectra could have never been obtainable otherwise in such a simple intersubband Quantum Well detector, since a two (or more) color detection scheme would have necessitated at least 3 bound/quasi-bound states.

Our simulation asserts that, even at moderate photon density, the spectra contains a secondary peak in far-deep infrared region. However, this far-infrared peak is not supposed to appear in experimentally found FTIR spectra until the corresponding photon energy of the secondary peak rises several order higher over thermal energy (kT). Interestingly enough, the device, if subjected to strong electron-photon interaction,

should re-orient its band-structure dynamically with the change in incident optical intensity so as to cause blue-shifting of the far-infrared peak (as well as red-shift of infrared peak) eventually resulting in two-color spectrum while still retaining a simple photo-sensing system.

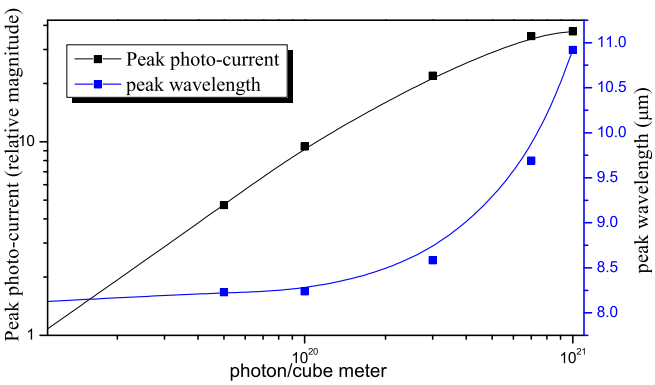


Fig. 7. Change in peak photocurrent and spectral maxima with increasing number of photon per unit volume (blue-shifting of far-infrared is not shown)

Fig. 7 shows that the photo-response keeps increasing almost linearly upto 10^{20} photon per m^3 of active region. Photon density exceeding this limit will cause the photo-response to get saturated. Moreover, the volume density of photon does not appreciably change the spectral maxima upto 10^{20} photon count per unit volume. Beyond that point, the spectrum sees a red shifting in peak wavelength for considerably large amount of optical intensity. At this limit of strong carrier-photon coupling, the band structure is not solely determined by device engineering, but also by the nature of light the device is responding to.

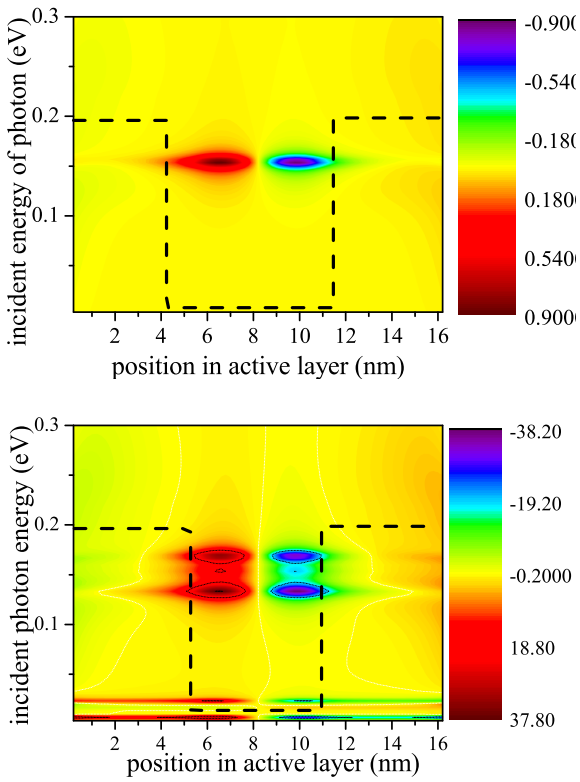


Fig. 8. Spatially resolved photocurrent spectrum at (top) low electron photon coupling, and (bottom) stronger coupling at thermal equillibrium

Fig. 8 contains spatially resolved photocurrent spectrum along the growth direction at thermal equillibrium. One can observe how the spectrum is gettting not only broadened but also splitted into a major and a minor peak around 15eV energy of photon incidence as the strength of electron photon coupling goes from moderate to stronger degree. An almost evanescent spectral line also appears near 0.035eV (corresponding to far-infrared line). Our simplified assumption of spatially invariant photon density along the structure gave rise to non-conservation of current; however, it has an added benefit of vividly depicting the QW region where photo-absorption does actually take place.

V. Conclusions

From the numerical analysis illustrated here, it can be asserted that the quantum mechanical description of a Quantum Well infrared Photo-detector we employed here with a careful computational manoeuvring with some straightforward assumption circumvents a considerable amount of computational burden and yet yields a satisfactory result which is in good agreement with that found experimentally by others within less than 3% error. Our simulation predicts the interesting possibility of multiple color detection mechanism in a two level system with the strong non-linearity in responsivity invoked by the intense coupling between electron-photon. Although we have mimicked the impact of strong electron-photon coupling by increasing photon volume density resulting from light incidence through the device aperture, it can effectually take place by few other and more plausible ways, i.e- plasmon assisted condensation of light. The model presented here is assumed to be excited by monochromatic beam; however, and modification in spectrum due to gaussian excitation beam is under our current investigation and will be published elsewhere.

References

[1] S. D. Gunapala and K. M. S. V. Bandara, in Thin Films, edited by M. H. Francombe and J. L. Vossen,. *Academic Press, New York*, 1995, Vol. 21, p. 113.

[2] Levine, B. F., "Quantum well infrared photodetectors", *Journal of applied physics* 74, no. 8 (1993): R1-R81.

[3] Henrickson, Lindor E. "Nonequilibrium photocurrent modeling in resonant tunneling photodetectors." *Journal of applied physics* 91.10 (2002): 6273-6281.

[4] Schuller, Jon A., Edward S. Barnard, Wenshan Cai, Young Chul Jun, Justin S. White, and Mark L. Brongersma. "Plasmonics for extreme light concentration and manipulation", *Nature materials* 9, no. 3 (2010): 193-204.

[5] Alves, Fabio Durante P., G. Karunasiri, N. Hanson, M. Byloos, H. C. Liu, Andrew Bezinger, and Margaret Buchanan,"NIR, MWIR and LWIR quantum well infrared photodetector using interband and intersubband transitions",Infrared physics & technology 50, no. 2

[6] Datta, S., "Electronic Transport in Mesoscopic Systems", *Cambridge University Press*, 1995

[7] Aeberhard, U., and R. H. Morf,"Microscopic nonequilibrium theory of quantum well solar cells." *Physical Review B*, 77.12: 125343,2008

[8] Datta, Supriyo. "Nanoscale device modeling: the Green?s function method.", *Superlattices and Microstructures* 28, no. 4 (2000): 253-278.

[9] Gasiorowicz, S. "Quantum Physics", Wiley, New York, 1984

[10] Gunapala, Sarath D., Sumith V. Bandara, A. Singh, John K. Liu, Sir B. Rafol, E. M. Luong, Jason M. Mumolo et al. "640 486 long-wavelength two-color GaAs/AlGaAs quantum well infrared photodetector (QWIP) focal plane array camera", *Electron Devices, IEEE Transactions* on 47, no. 5 (2000): 963-971.

[11] Keldysh, L. V.,"Diagram technique for nonequilibrium processes", *Sov. Phys. JETP* vol. 20.4, pp. 1018-1026, 1965

[12] Lake, Roger, Gerhard Klimeck, R. Chris Bowen, and Dejan Jovanovic. "Single and multiband modeling of quantum electron transport through layered semiconductor devices." *Journal of Applied Physics* 81, no. 12 (1997): 7845-7869.

Extraction of Lorentz Model parameters for Dielectrics and Their Application in Nanoplasmonics

KH. Arif Shahriar,* Ikramul Hasan Sohel, Md. Mahamudun Arif Joy, Rakibul Hasan Sagor, and Md. Ghulam Saber
Department of Electrical and Electronic Engineering, Islamic University of Technology (IUT)
Gazipur, Bangladesh
*arif.sjn@gmail.com

***Abstract*—The Lorentz model parameters for alumina, aluminum arsenide, zinc oxide, potassium bromide, graphene and gallium arsenide are optimized using a nonlinear optimization algorithm based on the Hooke and Jeeves method. The complex relative permittivity of each of the mentioned materials is calculated over a long wavelength range using the optimized parameters. The obtained permittivity curves exhibit very good match with those obtained from the experimental results. The symmetric SPP mode propagation through MDM waveguides containing alumina and aluminum arsenide as dielectric layers sandwiched between layers of silver has been investigated using their optimized modeling parameters.**

***Index Terms*—Lorentz model; Optical properties; Material optics; Material modeling; Finite-difference time-domain; Hooke and Jeeves algorithm; Surface Plasmon Polariton; Plasmonic waveguide.**

I. Introduction

The electromagnetic wave phenomena can be represented using Maxwell's modern theory of electromagnetism. For the purpose of computer simulation, Maxwell's equations can be efficiently discretized using different methods like the Finite Volume Time-Domain (FVTD) method, Finite-Difference Frequency-Domain (FDFD) method, Finite-Difference Time-Domain (FDTD) method, Finite-Element Method (FEM), Discontinuous Galerkin methods, Spectral method etc [1] . Among them, the Finite-Difference Time-Domain (FDTD) method, first proposed by Yee [2] in 1966, is one of the most commonly used methods.

Some significant advantages of FDTD are – it is a fully explicit computational technique which uses nonlinear algebra and treats the impulse response of an electromagnetic system naturally. Thus it can provide wideband frequency response in a single simulation. Besides, modeling of a new structure can be simplified to a problem of mesh generation avoiding complex reformulation of an integral equation [1], [3–5].

The original formulations provided by Yee require that the constitutive parameters i.e. conductivity, permeability and permittivity of the materials must be independent of frequency [6]. Though perfect conductors, ideal dielectrics and free space can satisfy this condition, it is rarely true for real materials. For some materials, this approximation can be applied for a narrow range of frequencies only [4]. So, the wavelength (or frequency) range of interest can be divided into several narrow subranges such that the approximation can be applied to each subrange. But a significant drawback of this approach is that the FDTD simulation has to be performed separately for each subrange and then the results have to be merged together to get a complete solution. As a result, the simulation becomes computationally inefficient and cumbersome [7].

As FDTD does not accept data in tabular form [7], so for accurate and efficient simulation of real materials using FDTD method, it is very important to mathematically model the material in such a way that the model parameters are applicable for a wide frequency range. Mathematical models like the Debye model, Lorentz model, Drude Model, Lorentz Drude model etc. are available for this purpose [3]. Therefore, researchers have paid a great deal of attention towards finding new ways to optimize the optical parameters of different materials based on these models.

The modeling parameters of several materials have been reported till date to our knowledge. Deinega et al. [7] have reported the modified Lorentz model parameters for silicon applicable in the wavelength range of 300-1000 nm. Rakic et al. [8] have reported the parameters for 11 materials using Brendel-Bormann and Lorentz-Drude model. Sagor et al. [9] have reported the Modified Debye Model parameters for Nickel, Palladium, Iridium, Titanium and Hexagonal Cobalt. Saber et al. [10] determined the Modified Debye and Lorentz Model parameters for Cuprous Oxide and Silicon-Germanium alloy. Alsunaidi et al. [11] have reported the parameters for AlGaAs. M.A. Ordal et al. [12] have reported the modeling parameters for 14 metals applicable for the infrared and far-infrared region.

In this paper we present the Lorentz model parameters for 6 materials optimized using a nonlinear optimization algorithm based on the Hooke and Jeeves method [13]. The complex relative permittivity of each of these materials has been calculated using these parameters and the calculated permittivities are in very good agreement with the experimental data. The experimental data for Al_2O_3, AlAs, KBr, GaAs are taken from [14], ZnO from [15] and Graphene from [16].

The obtained parameters can be used to simulate electromagnetic phenomena involving these materials. Recent research in the field of electromagnetic phenomena has given rise to the study of Surface Plasmon Polaritons (SPPs). SPPs are electromagnetic waves which travel along metal-dielectric interface with the unique capability to propagate beyond diffraction limit [17]. This property can lead to miniaturization of integrated photonic devices to a great extent [18]. But due to high attenuation, plasmonic signals die out over a distance of a few micrometers in the visible frequency range [17]. So, it is important to find new materials which can guide SPP with

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

low attenuation.

Using the obtained parameters of AlAs and Al_2O_3 and the parameters of Ag reported in [8], we have simulated two Metal-Dielctric-Metal (MDM) waveguides for symmetric SPP mode propagation. We investigated the power propagation profile for both the waveguides. It was observed that Ag/Al_2O_3/Ag configuration gives better performance in terms of symmetric SPP mode power propagation.

II. MATERIAL MODEL

The frequency dependent complex permittivity function for Lorentz model with N number of pole pairs[3] is given by:

$$\varepsilon(\omega) = \varepsilon_\infty + \sum_{n=1}^{N} \frac{\Delta\varepsilon_n \omega_n^2}{\omega_n^2 + 2j\omega\delta_n - \omega^2}$$

where the infinite frequency relative permittivity is represented by ε_∞, the number of the Lorentz pole pair is represented by n, the change in relative permittivity due to the nth pole pair is represented by $\Delta\varepsilon_n$, the undamped frequency of the nth pole pair is represented by ω_n and δ_n is the damping factor for the nth pole pair.

III. OPTIMIZATION METHOD

A nonlinear optimization algorithm based on the Hooke and Jeeves [13] method has been developed in order to extract the parameters. To begin with, the experimental values of real (ε_{rex}) and imaginary (ε_{iex}) parts of complex permittivity of a material are taken from [14–16].

Then a group of guess values of the parameters which need to be optimized is provided. The complex relative permittivity is calculated by inserting the guess values into the mathematical model. An objective function is used to calculate the deviation of the calculated permittivity from the experimental one. This objective function is minimized by varying the parameters according to the Hooke and Jeeves method so that the deviation becomes smaller in successive iterations and ultimately comes down to a predefined tolerance level.

We defined the objective function as follows:

$$f(\varepsilon_\infty, \Delta\varepsilon_n, \delta_n, \omega_n) = \ x_1 + x_2$$

where

$$x_1 = \sum (\varepsilon_{rc} - \varepsilon_{rex})$$
$$x_2 = \sum (\varepsilon_{ic} - \varepsilon_{iex})$$

Where ε_{rc} is the real part of permittivity calculated using the obtained parameters, ε_{rex} is the experimentally obtained real part of permittivity, ε_{ic} is the calculated imaginary part of permittivity, ε_{iex} is the experimentally obtained imaginary part of permittivity.

IV. FORMULATION AND SIMULATION OF THE WAVEGUIDE

A two dimensional plasmonic waveguide simulator has been developed based on the FDTD method. The original FDTD algorithm ignores the frequency dependency of permittivity of materials. The simulator is designed to overcome this limitation by implementing the general auxiliary differential equation (ADE)-FDTD algorithm [19].

In order to prevent reflection from the waveguide boundaries we used the perfectly matched layer (PML) proposed by Berenger in 1994 [20]. The region for simulation is divided into small cells. The step size is taken as Δx = Δy = 2nm and the time step is chosen as

$$\Delta t = 0.95c^{-1}[\frac{1}{\Delta x^2} + \frac{1}{\Delta y^2}]^{-2}$$

where c=3×10^8 ms^{-1} is the speed of light. These values of Δx,Δy,Δt were found to be sufficient to ensure the stability of the algorithm as well as the Courant stability criterion [3]. The schematic diagram of the waveguide is shown in Fig. 2(a).

The symmetric SPP mode is modulated by a Gaussian pulse of width 3fs and 1550nm wavelength and is pumped into the waveguide. The normalized Ey profile pumped into the structure is shown in Fig.2(b).

V. RESULTS AND DISCUSSION

The parameters extracted using the nonlinear optimization algorithm are presented in TABLE I. and TABLE II. The curves for real and imaginary parts of the complex relative permittivity are shown for both optimized and experimental values in Fig. 1. From the tables we observe that the maximum RMS deviation is 0.0019 for Potassium Bromide. Besides, the figures also exhibit very good agreement between the experimental and optimized curves.

We have used double pole pairs for modeling graphene and gallium arsenide and all other materials have been modeled using single pole pair. So, the time required to run the FDTD algorithm using these parameters will be less compared to the FDTD algorithm that uses multiple pole pair models.

The distribution of the E_y field of the symmetric SPP mode propagating through the MDM waveguide is shown in Fig.2(d).

The normalized power curves for Al_2O_3 and AlAs are shown in Fig.2(c). From the curves we see that the attenuation of SPP is less if Al_2O_3 is used as a dielectric layer in the MDM waveguide compared to AlAs.

TABLE I: OPTIMIZED PARAMETERS FOR ALUMINA, ALUMINUM ARSENIDE, ZINC OXIDE AND POTASSIUM BROMIDE FOR SINGLE POLE PAIR LORENTZ MODEL

Parameters	**Materials**			
	AlAs	**Al_2O_3**	**KBr**	**ZnO**
ε_∞	2.642392	1.001	1.562775	2.259007
$\Delta\varepsilon_1$	5.5635	2.057793	0.800918	1.465165
δ_1 (rad/sec)	1.8850×10^{12}	2.7×10^{12}	2.2×10^{10}	3×10^{10}
ω_1 (rad/sec)	6.2430×10^{15}	1.9678×10^{16}	1.1368×10^{16}	7.168×10^{15}
Range of λ (nm)	$495 - 1770$	$200 - 1960$	$220 - 2000$	$400 - 2070$
RMS Deviation	0.0305	0.013	0.0019	0.0079

TABLE II: OPTIMIZED PARAMETERS FOR GRAPHENE AND GALLIUM ARSENIDE FOR TWO POLE PAIR LORENTZ MODEL

Parameters	**Materials**	
	Graphene	**GaAs**
ε_∞	1.001	6.230234
$\Delta\varepsilon_1$	6.574	0.51418
$\Delta\varepsilon_2$	74.5407	5.251958
δ_1 (rad/sec)	-3.61×10^{15}	-1.3006×10^{14}
δ_2 (rad/sec)	-2.86×10^{15}	-5.1648×10^{14}
ω_1 (rad/sec)	9.14×10^{15}	4.4466×10^{15}
ω_2 (rad/sec)	1.21×10^{15}	5.1195×10^{15}
Range of λ (nm)	$400 - 1000$	$410 - 830$
RMS Deviation	0.023	0.0618

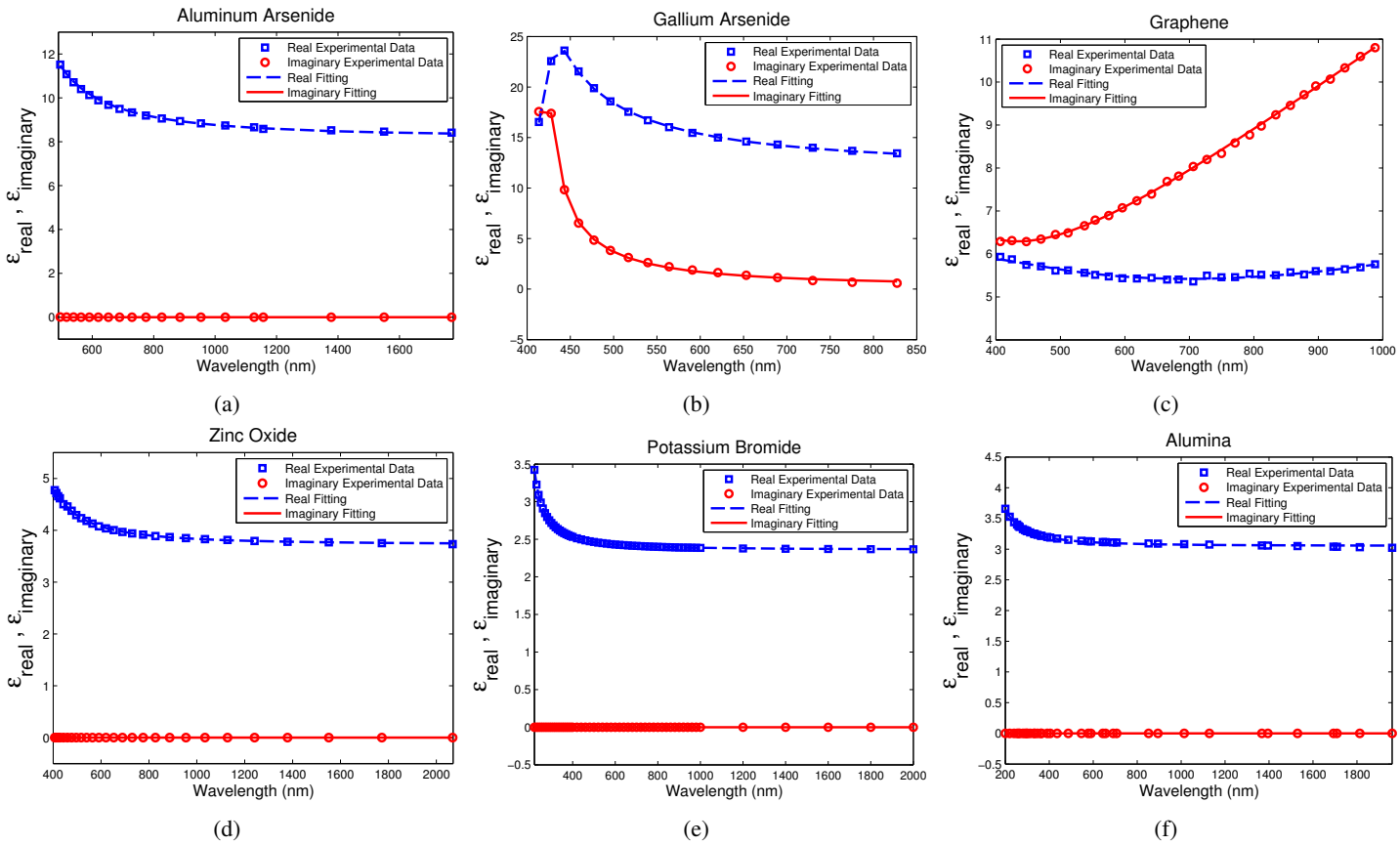


Fig. 1: Comparison between modeled and experimental complex permittivities of aluminum arsenide, gallium arsenide, graphene, zinc oxide, potassium bromide and alumina.

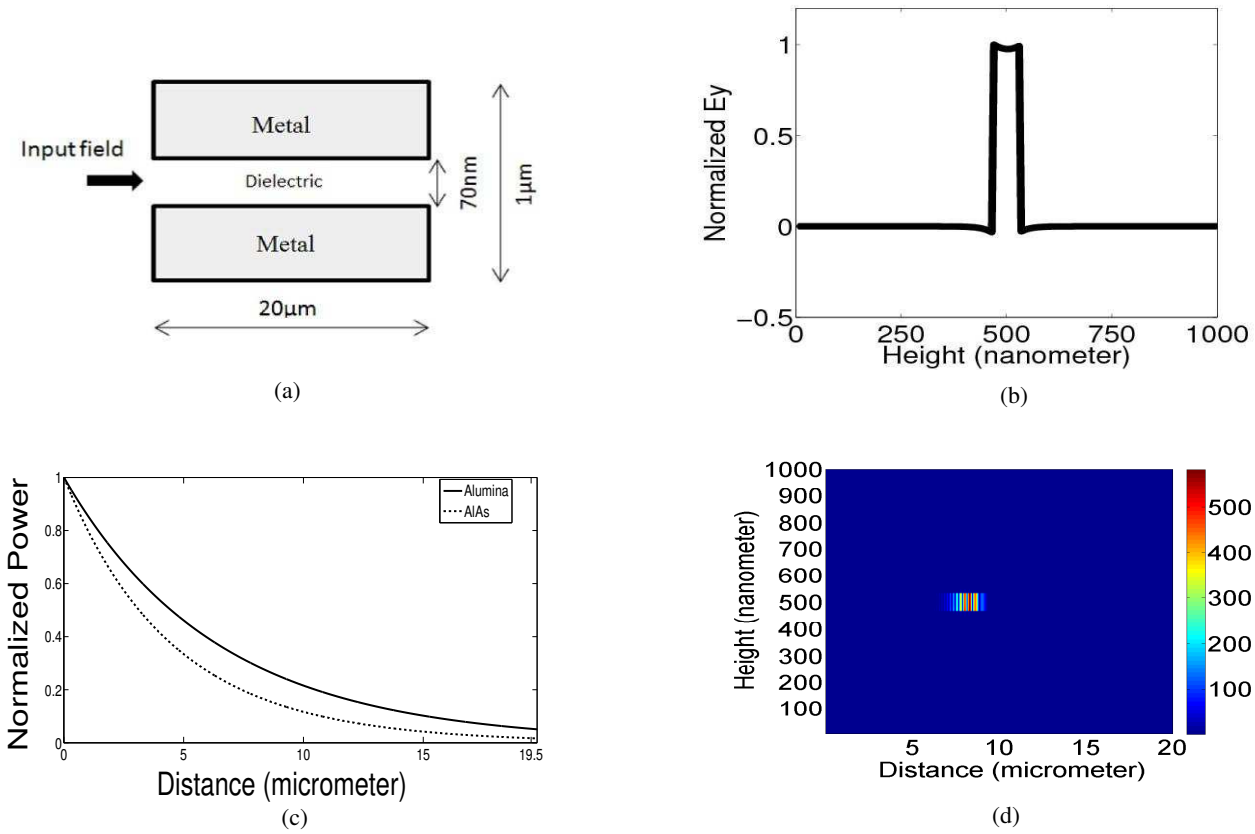


Fig. 2: (a) The schematic diagram of the waveguide. (b) Normalized E_y profile pumped into the waveguide. (c) Normalized power profile through the MDM waveguides. (d) Electric field distribution inside the waveguide.

VI. CONCLUSION

We have presented the Lorentz model parameters for alumina, aluminum arsenide, zinc oxide, potassium bromide, graphene and gallium arsenide. To the best of our knowledge, the Lorentz model parameters of these materials have not been presented in any work till date. The complex relative permittivities calculated using these parameters are in very good agreement with the experimental values. Since the obtained parameters are applicable for broad frequency ranges, broadband calculation can be performed in a single run of the FDTD algorithm. From the simulation of Ag/AlAs/Ag and Ag/Al_2O_3/Ag waveguides we can conclude that compared to AlAs, Al_2O_3 is a better choice as dielectric layer in MDM waveguide in terms of attenuation. In order to completely characterize the performance of AlAs and Al_2O_3 as dielectric layers in MDM waveguide we also need to consider the confinement factor which is left as future work . Using the same procedure, other materials and configurations can be simulated to compare their performance as SPP waveguides.

ACKNOWLEDGMENT

The authors would like to acknowledge the support of Islamic University of Technology.

REFERENCES

[1] Umran S Inan and Robert A Marshall. *Numerical Electromagnetics*. Cambridge University Press, 2011.

[2] Kane S Yee et al. Numerical solution of initial boundary value problems involving maxwells equations in isotropic media. *IEEE Trans. Antennas Propag*, 14(3):302–307, 1966.

[3] Allen Taflove and Susan C Hagness. Computational electrodynamics. 2000.

[4] Karl S Kunz and Raymond J Luebbers. *The finite difference time domain method for electromagnetics*. CRC press, 1993.

[5] Stephen D Gedney. Introduction to the finite-difference time-domain (fdtd) method for electromagnetics. *Synthesis Lectures on Computational Electromagnetics*, 6(1):1–250, 2011.

[6] Raymond Luebbers, Forrest P Hunsberger, Karl S Kunz, Ronald B Standler, and Michael Schneider. A frequency-dependent finite-difference time-domain formulation for dispersive materials. *Electromagnetic Compatibility, IEEE Transactions on*, 32(3):222–227, 1990.

[7] Alexei Deinega and Sajeev John. Effective optical response of silicon to sunlight in the finite-difference time-domain method. *Optics letters*, 37(1):112–114, 2012.

[8] Aleksandar D Rakic, Aleksandra B Djurišic, Jovan M Elazar, and Marian L Majewski. Optical properties of metallic films for vertical-cavity optoelectronic devices. *Applied optics*, 37(22):5271–5283, 1998.

[9] Rakibul Hasan Sagor, Md Ghulam Saber, Md Thesun Al-Amin, and Asif Al Noor. An optimization method for parameter extraction of metals using modified debye model. *SpringerPlus*, 2(1):426, 2013.

[10] Md Ghulam Saber and Rakibul Hasan Sagor. Optimization of the optical properties of cuprous oxide and silicon-germanium alloy using the lorentz and debye models. *Electronic Materials Letters*, 10(1):267–269, 2014.

[11] MA Alsunaidi and FS Al-Hajiri. Time-domain analysis of wideband optical pulse shg in layered dispersive material. *Session 5A1 Optics and Photonics*, 2:795.

[12] Mark A Ordal, Robert J Bell, RW Alexander Jr, LL Long, MR Querry, et al. Optical properties of fourteen metals in the infrared and far infrared: Al, co, cu, au, fe, pb, mo, ni, pd, pt, ag, ti, v, and w. *Applied optics*, 24(24):4493–4499, 1985.

[13] Robert Hooke and To A Jeeves. "direct search"solution of numerical and statistical problems. *Journal of the ACM (JACM)*, 8(2):212–229, 1961.

[14] Edward D Palik. *Handbook of optical constants of solids*, volume 3. Academic press, 1998.

[15] Sadao Adachi. *Optical constants of crystalline and amorphous semiconductors: numerical data and graphical information*. Springer, 1999.

[16] JW Weber, VE Calado, and MCM Van de Sanden. Optical constants of graphene measured by spectroscopic ellipsometry. *Applied Physics Letters*, 97(9):091904–091904, 2010.

[17] Stefan Alexander Maier. *Plasmonics: Fundamentals and Applications: Fundamentals and Applications*. Springer, 2007.

[18] Harry A Atwater. The promise of plasmonics. *Scientific American*, 296(4):56–62, 2007.

[19] Mohammad A Alsunaidi and Ahmad A Al-Jabr. A general ade-fdtd algorithm for the simulation of dispersive structures. *Photonics Technology Letters, IEEE*, 21(12):817–819, 2009.

[20] Jean-Pierre Berenger. A perfectly matched layer for the absorption of electromagnetic waves. *Journal of computational physics*, 114(2):185–200, 1994.

Semi-analytical Model for Enhanced Surface Plasmon Polaritons in a Corrugated Interface

M A Awal, Zabir Ahmed, and Muhammad Anisuzzaman Talukder*
Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology
Dhaka - 1205, Bangladesh
*anis@eee.buet.ac.bd

***Abstract*—We propose and present a semi-analytical model to study the interaction between surface plasmon polaritons generated by phase-matching structures in corrugated dielectric/metal interfaces. The proposed semi-analytical model calculates surface plasmon polariton dynamics that agree well with that found from the full-field vectorial finite difference time domain (FDTD) simulations. The proposed model takes almost no computational time in comparison with FDTD simulations. Therefore, the proposed model can be used instead of FDTD simulations to design and engineer sub-wavelength structures to enhance light confinement. In particular, the proposed model can save significant amount of time to efficiently predict the optimal structural parameters to achieve resonant absorption in metal/dielectric interface of a thin-film solar cell.**

***Index Terms*—Periodic corrugations, plasmonics, surface plasmon polariton, plasmonic solar cell, plasmonic light trapping.**

I. Introduction

With the ever increasing global population along with rapid technological development and increased demand of mechanizaton for comfort, global energy demand keeps on growing. It is predicted that with the current rate of growth of energy consumption, global oil reserve will be depleted within 50 to 100 years; natural gas reserve will be exhausted in 60 to 70 years; and coal reserve will be depleted in around 200 years [1]. Apart from the limited reserve of fossil fuels, the use of fossil fuels raises severe environmental concerns. At present, the burning of fossil fuels produces around 21.3 billion tonnes of carbon dioxide (CO_2) per year [2]. However, it is estimated that the natural processes can absorb only about half of that amount produced due to the burning of fossil fuels. Therefore, there is a net increase of 10.65 billion tonnes of atmospheric carbon dioxide per year [2]. Carbon dioxide is one of the greenhouse gases that enhances radiation trapping and contributes to global warming, which is strongly believed to cause major adverse effects on global climate in near future. This calls for urgent and rigorous exploration for alternative clean and environment-friendly energy sources.

Solar irradiation is the primary source of energy on earth. The amount of solar radiation striking earth is around 1.76×10^6 TW, which is orders of magnitude higher than the global energy demand [3]. Among all renewable energy sources, solar energy is the most promising, however, still has a higher cost relative to other sources as photovoltaic energy conversion is not very efficient. Reducing the cost per Watt by increasing the efficiency has been the driving force behind photovoltaic research till date.

Historically, crystalline Si (c-Si) has been the mostly used semiconductor for photovoltaics. In c-Si solar cells, approximately 50% of the total cost is due to the cost of the materials themselves, with the remainder due to cell production and module fabrication [4]. There are several advantages of using c-Si as a material such as its abundancy, low cost, and availablility of matured technology for fabrication. However, c-Si is a weak light absorber requiring a layer of several hundreds of microns to fully absorb the incident sunlight and thus increasing the cost. Alternative to bulk PV cells, i.e., thin-film PV cells have been developed, which have thickness of only a few microns. The thin-film PV cells have lower processing cost, but with lower efficiencies [5].

Incorporation of plasmonic design in thin-film cells increases effective absorption length that leads to increased light absorption which can significantly compensate for the lower active region thickness [6]–[8]. Plasmon is a quasi particle that represents macroscopic oscillation of free electron plasma. Free electrons in metals under proper excitation give rise to several forms of plasmons, i.e., volume plasmon, localized surface plasmon, and surface plasmon polariton (SPP). In this work, we have studied the dynamics of SPPs as a means to increase the efficiency enhancement of thin film PV cells. Recent studies have shown significant promise of corrugated dielectric/metal interfaces to support the excitation of SPPs for increased light confinement in sub-wavelength structures [9]. To optimize such a structure, thorough understanding of the underlying physics is very important. Several studies have been done to investigate the transmission and reflection properties of a groove to incident SPPs [10], [11]. The phase relation between SPPs excited by individual phase-matching structures depends on the detail geometry of the structure. Therefore, the dynamics of SPPs can be engineered by changing the geometry of the corrugated interface.

In this work, we present a semi-analytical model to characterize the interaction between SPPs coupled by finite or infinite number of periodic rectangular grooves in a dielectric/metal interface. We apply this model to calculate the field along a c-Si/Ag interface, which is often used in photovoltaic structures. We also calculate the field with full finite difference time domain (FDTD) simulation. We find good agreement between

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

the results found using the semi-analytical model with that found using the full FDTD simulation. The semi-analytical model is much faster than FDTD simulation, and therefore, can be used to design sub-wavelength structures such as thin film PV cells for increased light confinement.

II. Interaction of SPP with Unexcited Grooves

Periodic groove structure on the back contact of a PV cell is very efficient to confine light in SPP modes in the dielectric/metal interface, consequently enhancing absorption and photo-current [12], [13]. For efficient light-trapping, groove geometry, groove dimensions and pitch width of the periodic groove structure are important design parameters. Since resonances of plasmonic modes depend on all these parameters simulatenously, it poses a significant design challenge to find optimal values of these parameters. Parameter sweeps for different design parameters require considerable amount of time. The semi-analytical model that we propose in this work resolves this issue by calculating the response of a groove by its transmission properties.

We confine our study to rectangular groove geometry. To understand and predict the interaction of generated SPPs with adjacent grooves, it is important to study the response of an unexcited groove to an incident SPP mode. A proper insight of the reflection and transmission properties of an unexcited groove is helpful to develop understanding toward an improved structure design.

A. Simulation Setup

To simulate the response of an unexcited groove to incident SPP mode, we carry out FDTD simulation for a Silver/Silicon interface with only two grooves. Since our focus is to study the dynamics of grooves in confining light in plasmonic modes, we solve only for plasmonic modes and neglect other scattering and photonic modes. To cancel the photonic mode from the simulation, a layer of Silicon with semi-infinite thickness is placed on Silver substrate and the source is placed inside the Silicon layer. We model the materials, i.e., Silver and Silicon using optical constants obtained from Palik [14].

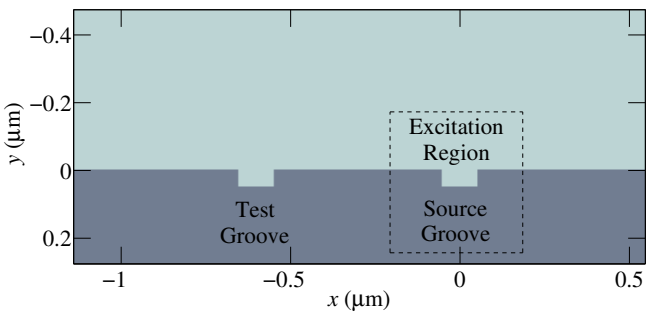


Fig. 1. Simulation setup for modelling response of unexcited groove to incident SPP wave.

In this setup, light is incident on only one of the grooves, which we identify as the "source groove." The incident light on the source groove excites SPP, which propagates and is incident on the sidewall of the second groove, which we identify as the "test groove."

Figure 1 shows the simulation setup where a rectangular source groove with width of 100 nm and depth of 50 nm was excited by normal incidence of a plane polarized monochromatic wave with free space wavelength of 1000 nm. In this particular case, test groove is separated from the source groove by 50 nm and has the same dimensions as the source groove.

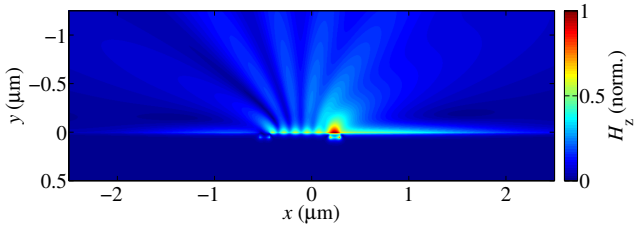


Fig. 2. Standing wave-like mode profile between the test and source groove. Both grooves have a width of 100 nm and a depth of 50 nm.

To model the response of an unexcited groove to incident SPP, we carry out another FDTD simulation to obtain the field amplitude profile for a system with only the source groove. Figure 2 shows the resulting normalized magnetic field amplitude for the simulation domain shown previously in Fig. 1 except when there is no test groove.

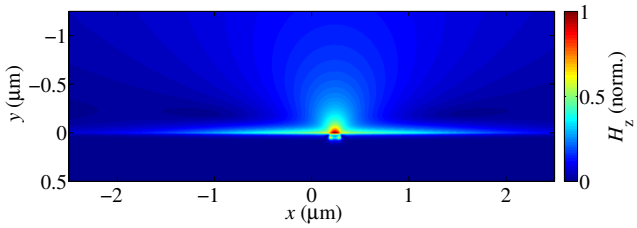


Fig. 3. Resulting normalized H_z field amplitude for the source groove in the absence of test groove

In Fig. 3, the magnetic field distribution shows the SPP mode profile excited by the source groove, which is essentially a reference quantity relative to which the transmission characteristics of the test groove will be determined. To calculate these transmission characteristics, we concentrate on the field distribution along the metal/dielectric interface. Figure 4 shows normalized H_z field along $y = 0$ for both of the output responses from Fig. 2 and Fig. 3. We determine three parameters, i.e., field transmission coefficient (T), field reflection coefficient (R), and transmission phase shift (ϕ_t) to quantify transmission characteristics of an unexcited groove of a specific dimension. These parameters will be employed to develop the semi-analytical model that we propose in this work.

B. Calculation of T, R, and ϕ_t

When SPP generated from the source groove is incident upon the test groove, part of it is transmitted through the test groove, part of it is reflected back toward the source groove, and the rest is decoupled from it to couple into the photonic mode. Field transmission coefficient T signifies the fraction of SPP mode that propagates through the test groove. To calculate T, first we determine the field amplitude at a point beyond the test groove along Ag/Si interface from the simulation with both grooves. Then we determine the field amplitude at the same point from the simulation with only the source groove. We define T as the ratio of these two field quantities. In the case of Fig. 4, we take the ratio of normalized magnetic field amplitudes at $x = -2$ μm to calculate T.

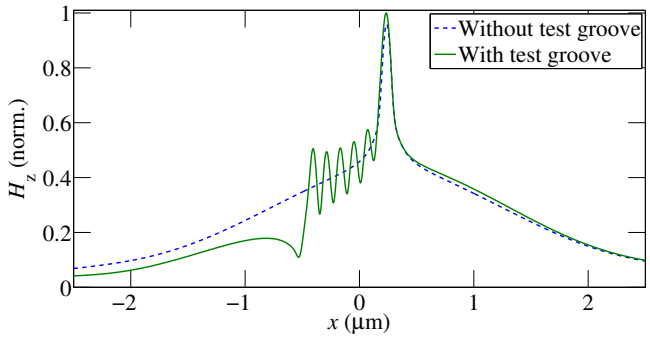


Fig. 4. Resulting normalized H_z field amplitude along $y = 0$ for the source groove with and without the test groove

Calculation of field reflection coefficient R is not straightforward. To calculate R, we use the observation that a standing wave-like mode is created in the region between two grooves as shown in Fig. 2. This mode shape is created due to superposition of the incident SPP from source groove and the reflected SPP from the sidewall of the test groove. If the exponential decay associated with the SPP mode is to be ignored over small propagation length, H_z at the maxima of the standing wave-like mode near to the test groove can be written as

$$\begin{aligned} H_{z,\text{maxima}} &= H_{z,\text{incident}} + H_{z,\text{reflected}} \\ &= H_{z,\text{incident}} + R H_{z,\text{incident}}. \end{aligned} \tag{1}$$

From Eq. (1), we can evaluate R at different groove dimensions. For the case shown in Fig. 4, local maxima at $y = -0.4075$ μm is considered for calculation of R.

Transmission phase shift, ϕ_t is calculated by comparing the relative phase difference between the SPP modes with and without the test groove at a point beyond the test groove. For this purpose, field profile at a fixed time step for both simulation domain is calculated.

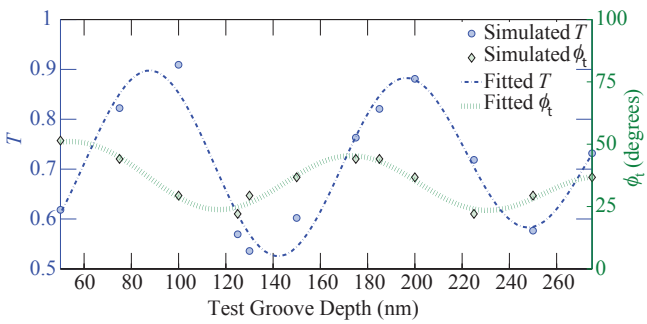


Fig. 5. Field transmission coefficient T and transmission phase shift ϕ_t versus the depth of test groove.

With the changes of groove dimensions, the parameters T, R, and ϕ_t show significant variations. Figure 5 shows the fluctuations of T and ϕ_t for a range of test groove depths while keeping groove width constant at 100 nm. These modeling parameters calculated from full-field simulations are pivotal for the semi-analytical model presented in the next section.

III. Semi-analytical Model

A. Formulation

The magnetic field of the SPP mode excited in a dielectric/metal interface is described as

$$H_z(x,t) = A e^{j(\beta_{\text{spp}} x - \omega t)}. \tag{2}$$

In Eq. (2), β_{spp} is the propagation constant that depends on the structure. In general, we can write $\beta_{\text{spp}} = -a + jk_{\text{spp}}$, where $a, k_{\text{spp}} > 0$ for $x > 0$ and $a, k_{\text{spp}} < 0$ for $x < 0$. The amplitude of the SPP mode in a region, which has N number of excited grooves either on the left or right side, is given by [15]

$$A = \sum_{m=0}^{N} A_m T^m e^{-amd} e^{jm(k_{\text{spp}} d - \phi_t)}, \tag{3}$$

where d is the pitch width between grooves. Now, if we have an interface with identical grooves, $A_0 = A_1 = ... = A_N$. In the presence of N grooves on both sides, superposition of two such modes gives us the SPP mode. When each of these two modes is incident on a groove, a part is reflected back. Therefore, the magnetic field between two grooves, $H_{z,\text{Left}}$ contributed by N grooves on the left side and $H_{z,\text{Right}}$ for N grooves on the right side can be described as

$$\begin{aligned} H_{z,\text{Left}} =& A\{\mathrm{e}^{-ax}\sin[k_{\text{spp}}x - \omega t] \\ &- R\,\mathrm{e}^{a(x-2d)}\sin[-k_{\text{spp}}(x-2d) - \omega t]\}, \end{aligned} \tag{4a}$$

$$\begin{aligned} H_{z,\text{Right}} =& A\{\mathrm{e}^{a(x-d)}\sin[-k_{\text{spp}}(x-d) - \omega t] \\ &- R\,\mathrm{e}^{-a(x+d)}\sin[k_{\text{spp}}(x+d) - \omega t]\}. \end{aligned} \tag{4b}$$

In Eq. (4), R is the field reflection coefficient, which is accounted only for the first stage. The SPP mode between two grooves is found as $H_z = H_{z,\text{Left}} + H_{z,\text{Right}}$. The values of T and ϕ_t presented in Fig. 5 are used to evaluate Eq. (4). The values of a and k_{spp} can be obtained from analytical calculation for a given metal/dielectric interface [16].

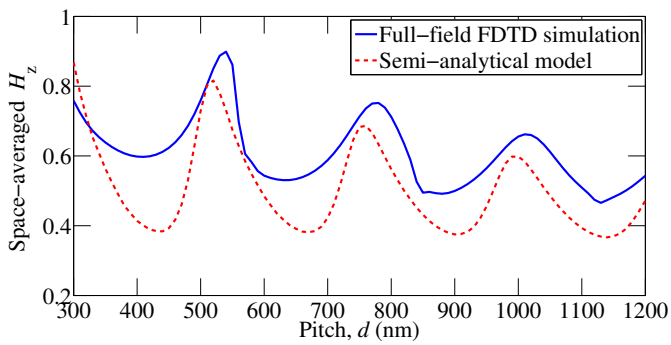


Fig. 6. Space-averaged H_z versus pitch of the corrugated interface from semi-analytical model and full FDTD simulation.

B. Results

We compare the results obtained from the semi-analytical model with that obtained from full FDTD simulation. In Fig. 5, we plot the field transmission coefficient and phase shift for varying test groove depth from FDTD simulation when the light is incident only on the source groove. We use the parameter values obtained from Fig. 5 in the semi-analytical model and calculate the space-averaged field between two adjacent grooves for a structure with $N = 1000$ grooves on each side of these two adjacent grooves of interest. The results obtained from semi-analytical model are given in Fig. 6 along with the results obtained from the full FDTD simulation with periodic boundary conditions. In Fig. 6, space-averaged H_z along the interface is plotted versus pitch of the grooves on the interface. We note that the semi-analytical results match well with full FDTD simulation results. The difference between the semi-analytical and full FDTD simulation results may occur due to the fact that the semi-analytical model does not take the field at the groove faces into account.

IV. Conclusion

The model presented here employs characteristic transmission properties of a single phase-matching structure to semi-analytically predict the field enhancement in periodic structures. The results closely match with that obtained from full FDTD simulations. Design and optimization of a sub-wavelength device requires numerous full FDTD simulations. By contrast, our proposed model presents a design rule, which could easily be used to calculate optimum parameter values, such as the groove depth and pitch, for a sub-wavelength geometry for enhancement and engineering of the confined field along a dielectric/metal interface.

V. Acknowledgement

This work was partly supported by the funding from Ministry of Education, People's Republic of Bangladesh, under the project titled, "Plasmonic Photovoltaics for Next Generation Solar Cells."

References

[1] Terra Symbiosis. Depletion of fossil resources. [Online]. Available: http://www.terra-symbiosis.org/EN_epuisement-fossiles.pdf

[2] (2007) US Department of Energy. What Are Greenhouse Gases. [Online]. Available: http://www.eia.gov/oiaf/1605/ggccebro/chapter1.html

[3] David Archer, *Global Warming: Understanding the Forecast*, Wiley, 2011.

[4] A. Luque and S. Hegedus, *Handbook of photovoltaic science and engineering*, John Wiley & Sons, 2011.

[5] M. A. Green, "Third generation photovoltaics: solar cells for 2020 and beyond," *Physica E: Low-dimensional Systems and Nanostructures*, vol. 14, no. 1, pp. 65–70, 2002.

[6] H. A. Atwater and A. Polman, "Plasmonics for improved photovoltaic devices," *Nature materials*, vol. 9, no. 3, pp. 205–213, 2010.

[7] K. Catchpole and A. Polman, "Plasmonic solar cells," *Optics express*, vol. 16, no. 26, pp. 21793–21800, 2008.

[8] S. Pillai, K. Catchpole, T. Trupke, and M. A. Green, "Surface plasmon enhanced silicon solar cells," *Journal of Applied Physics*, vol. 101, no. 9, pp. 093105, 2008.

[9] V. E. Ferry, L. A. Sweatlock, D. Pacifici, and H. A. Atwater, "Plasmonic nanostructure design for efficient light coupling into solar cells," *Nano letters*, vol. 8, no. 12, pp. 4391–4397, 2008.

[10] J. S. Liu, J. S. White, S. Fan, and M. L. Brongersma, "Side-coupled cavity model for surface plasmon-polariton transmission across a groove," *Optics express*, vol. 17, no. 20, pp. 17837–17848, 2009.

[11] M. Kuttge, F. J. García de Abajo, and A. Polman, "How grooves reflect and confine surface plasmon polaritons," *Optics express*, vol. 17, no. 12, pp. 10385–10392, 2009.

[12] Wang, Zhu, T. White, and K. Catchpole. "Plasmonic near-field enhancement for planar ultra-thin photovoltaics," *Photonics Journal*, vol. 5, no. 5, pp. 8400608, 2013.

[13] V. E. Ferry, "Light Trapping in Plasmonic Solar Cells," PhD. Dissertation, California Institute of Technology, 2011.

[14] E. D. Palik, *Handbook of optical constants of solids*. Academic press, 1998, vol. 3.

[15] H. Liao, Z. Li, J. Chen, X. Zhang, S. Yue, and Q. Gong, "A submicron broadband surface-plasmon-polariton unidirectional coupler," *Scientific reports*, vol. 3, 2013.

[16] S. A. Maier, *Plasmonics: Fundamentals and Applications*. Springer, 2007.

Miniaturized VCSEL Modules for Optical Manipulation of Microparticles

Niazul Islam Khan[1,*], Anna Bergmann[2], and Rainer Michalzik[2]

[1]Department of Electrical and Electronic Engineering, East West University, Aftabnagar, Dhaka-1212, Bangladesh
[2]Institute of Optoelectronics, Ulm University, 89069 Ulm, Germany
[*]niaz.eee04@gmail.com

Abstract— **In recent years, optical manipulation has been emerging as a powerful technique in a wide range of applications, especially in the field of biological sciences. It is a mechanism to handle micrometer sized particles without mechanical or thermal damage or contamination. Lately, the use of VCSELs (vertical-cavity surface-emitting lasers) as light sources in optical manipulation has got significant attention because of their various advantages compared to conventional lasers. Particularly, their small dimensions with typical active diameters of less than ten micrometers, the possibility of simple and inexpensive formation of two-dimensional VCSEL arrays and their integration capability have made them convenient for optical manipulation applications. Another field of interest in biological research is microfluidics, which studies the manipulation and analysis of fluids in microscale. In this paper, we combine these three basic elements – VCSELs, optical manipulation, and microfluidics – to form VCSEL-based miniaturized optical trap array modules. In our previous work, the fabrication of VCSEL arrays and their characterization results were presented. In this paper, the integration of the VCSEL arrays into the module will be presented. Optical deflection experiments with microparticles will also be shown.**

Index Terms— **VCSEL, miniaturization, integration, optical manipulation, microfluidics.**

I. Introduction

Optical manipulation is a technique of manipulating particles using light. It is based on the phenomenon that micrometer-sized particles can be moved by the radiation pressure of light. The concept of laser-based optical trapping was first suggested by Arthur Ashkin in 1970 [1]. He observed in an experiment with a weakly focused laser beam that particles were moving along the beam axis and towards the beam center. The working principle of optical trapping is based on the light-matter interaction. When light hits a dielectric particle, there is a momentum transfer between the particle and the light. This effect is illustrated in Fig. 1, where a transparent particle with a diameter of several wavelengths and a refractive index higher than that of the surrounding medium is located off-axis in an unfocused Gaussian beam. Two exemplary parallel rays A and B are considered. When these rays are incident on the particle, they are refracted as well as reflected at the particle surface. Considering only refractions, we see a change in the directions of the rays. This change causes a change in momentum of the incident rays and gives rise to two forces F_A and F_B oriented orthogonal to the respective ray. With ray A having a higher intensity than ray B, the force F_A is larger than the force F_B. This leads to a net force that pulls the particle towards the intensity maximum, or towards the beam center.

Today, optical manipulation has found numerous applications in different branches of science and technology. Particularly, it has gained much interest in the field of biological sciences and chemistry because it offers the advantage of manipulating biological material without contamination, mechanical or thermal damage [2].

Microfluidics, another emerging field in today's biological research has expanded very rapidly because it has the merits of miniaturization, portability, and high throughput with reduced sample volume [3]. VCSELs, another component of our module, have attained significant attention as laser sources in almost all areas where miniaturization is an important issue [4]. Because of their versatile properties, they are highly suited for optical manipulation in microfluidics. Their emission in the near-infrared region falls in the transparency window of biological material, i.e., biological material shows low absorption in this wavelength range, minimizing the risk of thermal damage [5]. Moreover, their surface emission eases the fabrication of two-dimensional laser arrays, which enables the formation of multiple optical lattices. This would otherwise require extensive beam splitting or steering setups [4]. Besides, their compact size with typical dimensions of some ten micrometers facilitates the realization of a drastically miniaturized module. Hence, by the integration of VCSEL arrays with microfluidic channels, a portable, inexpensive particle manipulation device is feasible.

II. Vertical-Cavity Surface-Emitting Lasers

Fig. 2(a) depicts the schematic of a surface-relief VCSEL structure emitting its light perpendicular to the surface. Epitaxially grown on a GaAs substrate using an AlGaAs-based molecular beam epitaxy system, the structure is targeted for an emission wavelength of 850 nm. The laser resonator is formed by sandwiching an active region by two mirrors, namely the top and the bottom DBR (distributed Bragg reflector). To enable back-side n-contacting, both the bottom DBR and the GaAs substrate are n-doped. The top DBR is p-doped and has a

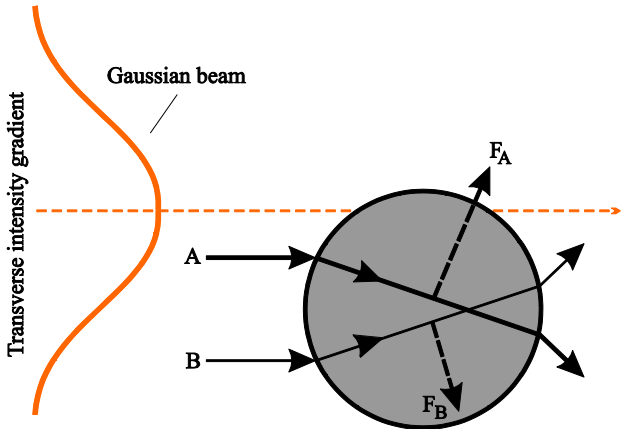


Fig. 1. Working principle of a two-dimensional optical trap.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

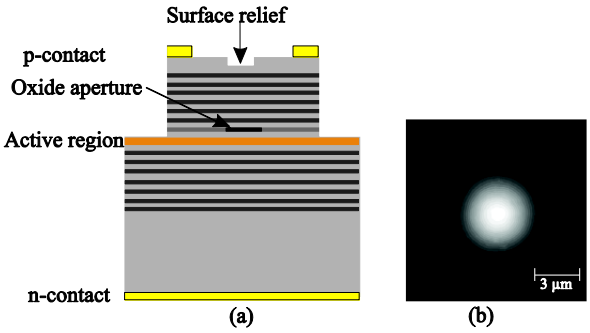


Fig. 2. Schematic of a surface-relief VCSEL structure (a) and its measured Gaussian-like near-field intensity pattern (b).

reduced number of mirror pairs compared to the bottom DBR to enhance light outcoupling from this side. In contrast to a conventional VCSEL structure, the p-doped DBR is followed by an additional antiphase layer on top, with a thickness of one quarter of the material wavelength. During fabrication process, the antiphase layer is selectively etched in the center of the laser facet to form a so-called surface-relief. This laser structure favors the emission of a Gaussian-like beam profile as in Fig. 2(b), in contrast to ring-like beam profiles from conventional VCSELs with the same aperture diameter. A ring-like emission is unfavourable for the targeted miniaturized trapping scheme because it leads to an offset between beam and trapping center [6]. Besides, lower beam quality increases the beam diameter at the focal point.

To enable light emission from the top side, the top p-contact is structured as a ring on top of the VCSEL mesa. Sandwiched between the bottom and the top DBR, the active region consists of three 8 nm thick GaAs quantum wells that are separated by 10 nm thick barriers. Undoped spacer layers are inserted on both sides of these quantum wells in order to make the active region one wavelength thick, which improves the longitudinal carrier confinement [7]. Above the active region is the oxidation layer with high aluminium content (typically 98–100%). During the fabrication process, this layer is selectively oxidized to form an oxide aperture. The aperture provides both carrier and current confinement in the transverse direction on the p-doped side.

III. Integration Approach of the Miniaturized Trap Array Module

The miniaturized trap array module targeted for our experiments requires densely packed surface-relief VCSEL arrays and the integration of the microfluidic chip directly above the VCSEL facets [8]. The integration is a critical consideration for the fabrication of the module. Electrical accessing of the p-contact in the conventional way by bond wires hinders this close integration of the microfluidic chip. Therefore, the p-contacting is done by bondpad metals placed in 30 to 40 µm deep bonding grooves. The integration

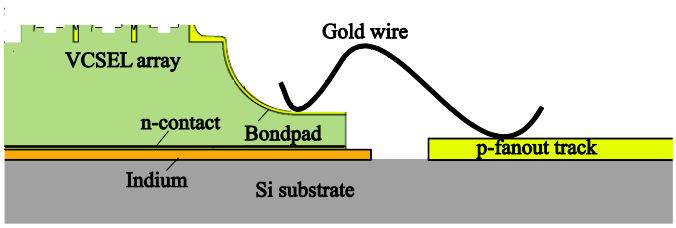


Fig. 3. Cross-sectional schematic view of a wire-bonded VCSEL chip on a chip carrier with fanout tracks.

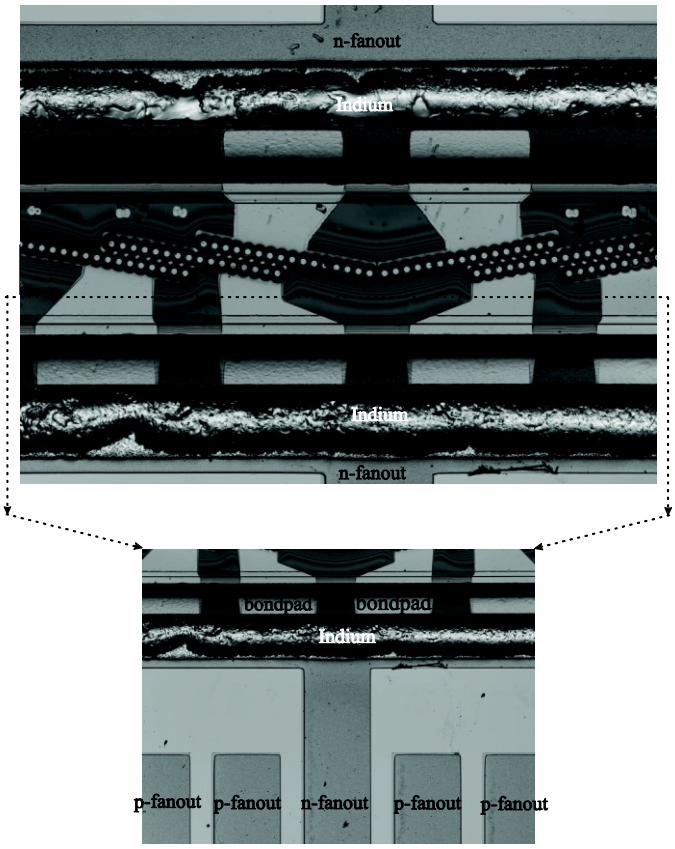


Fig. 4. Laser microsocopic image of the top side of the chip carrier with an integrated VCSEL chip (top) and enlarged view of the metal fanout tracks (bottom).

concept is illustrated in Fig. 3. A VCSEL chip soldered to a structured chip carrier can be seen. The heat sink contains several fanout tracks. For accessing the p-contact of the VCSEL arrays, the fanout tracks are connected to the bondpads by gold wires. It is clear from the figure that because of the deep bonding groove, the gold wire will not create an obstacle for the microfluidic chip to be placed close above the VCSEL facets.

IV. Fabrication of the Module

For the module, integration-capable top-emitting VCSEL arrays are required. The fabrication process of the arrays is presented in Ref. [8]. After the fabrication, the fabricated VCSEL sample is sawed into 1.8 mm × 0.6 mm chips. Then the chips are soldered on a silicon chip carrier with a size of 20.7 mm × 2.53 mm. To form the fanout tracks, first a seed structure of titanium and gold is deposited on the wafer, then thick gold is electroplated on this seed structure. Later indium is evaporated for soldering of the VCSEL chips. A top view of a chip carrier with a VCSEL is partially visible in Fig. 4. The heat sink has five fanout tracks on each side – one fanout track for the common n-contact and four tracks for four p-contacts. The p-fanout tracks are connected to the bondpads by wire bonding. The characterization results are presented in section V.

In the next step, the optoelectronic chip is integrated with the microfluidic chip. It is glued and then treated with ultraviolet curing [9]. The chip contains a microfluidic channel (Y-junction) with one inlet and two outlets. The channel is fabricated from a polymer called Topas® COC (Cyclic olefin copolymer) and has a width of 60 µm. During

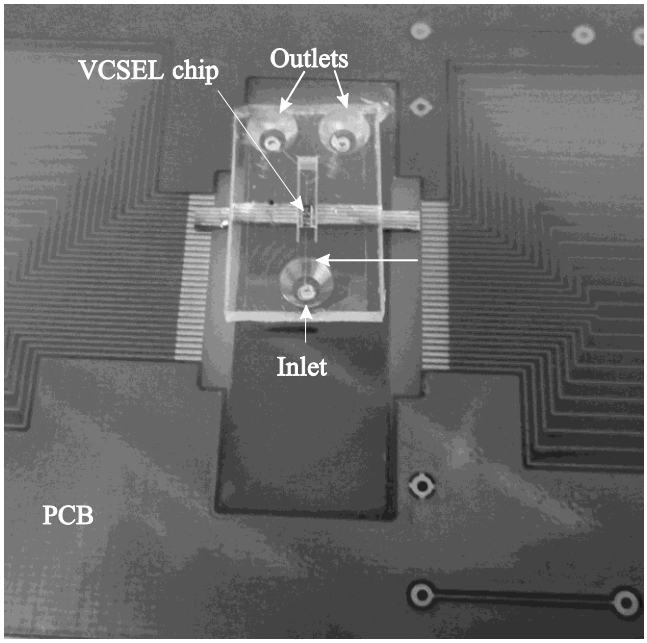


Fig. 5. Photograph showing the assembled VCSEL chip integrated with the microfluidic chip on a printed circuit board.

the integration, the channel is aligned to the VCSEL arrays. The chip is then assembled on a printed circuit board (PCB) using glue. Figure 5 shows the assembled chip on the PCB.The inlets and outlets of the channels are connected via tubes to a syringe containing a particle solution. Electrical

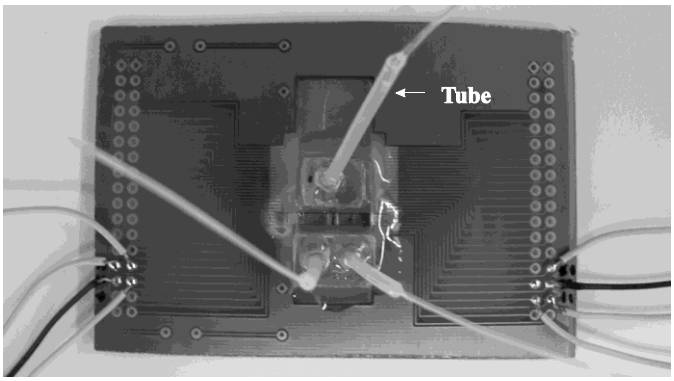


Fig. 6. Photograph of a fully functional VCSEL-based miniaturized trap array module.

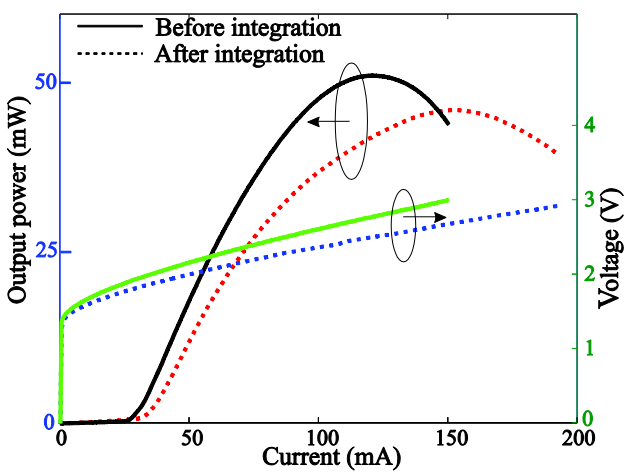


Fig. 7. LIV characteristics of a surface-relief laser array consisting of 13 emitting VCSELs with an individual active diameter of 10 µm.

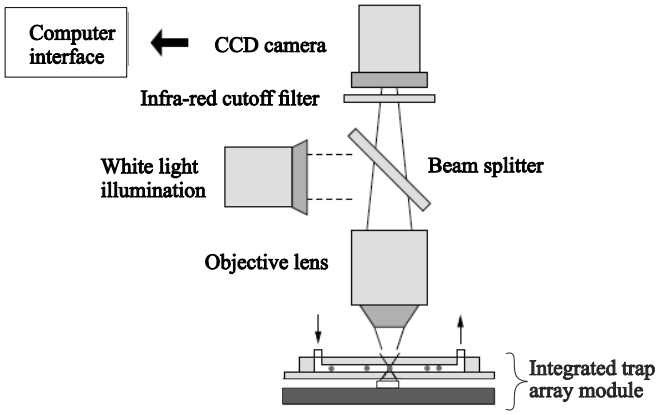


Fig. 8. Schematic of the experimental setup used for the observation of the optical deflection experiment.

wires are then attached to the PCB. The fully functional miniaturized VCSEL array module can be seen in Fig. 6.

V. Characterization After Heat Sink Integration

Figure 7 shows the LIV (light-current-voltage) curves of a fabricated VCSEL array before and after integration with the chip carrier. The array consists of 13 emitting VCSELs with an individual active diameter of 10 µm. A threshold current of around 26 mA, which increases slightly after integration, can be observed. The maximum power decreases from 52 mW to around 42 mW. This reduction of power might be caused by the increased leakage current after the soldering of the VCSEL chip to the carrier.

VI. Trapping Experiments

A. *Experimental Setup*

To observe the experiment, a CCD camera is used in combination with a beam splitter and white light illumination. An IR-cutoff filter is used to block the laser light so that the particle movement during laser operation is visible in the CCD camera. Therefore, the laser beams cannot be seen in the images. The schematic of the experimental setup is displayed in Fig. 8. The flow of the liquid is regulated by a computer-controlled syringe pump. Figure 9 shows the laser arrays just under the channel. The output power is 3 mW per laser.

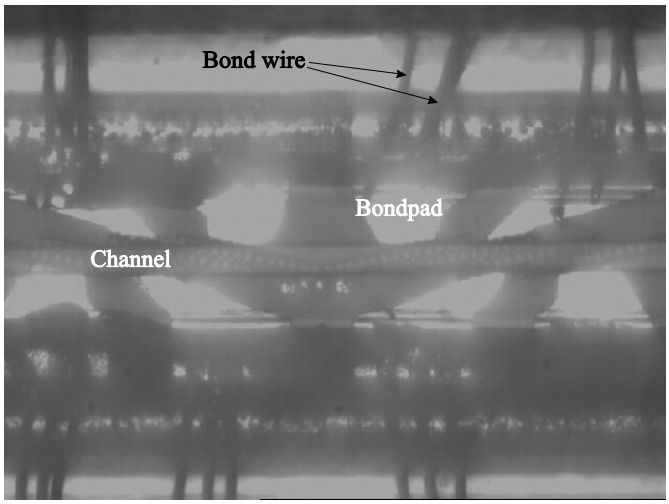


Fig. 9. Photograph of the laser arrays just under the microfluidic channel. The laser arrays are tilted by about 12° with respect to the channel edge.

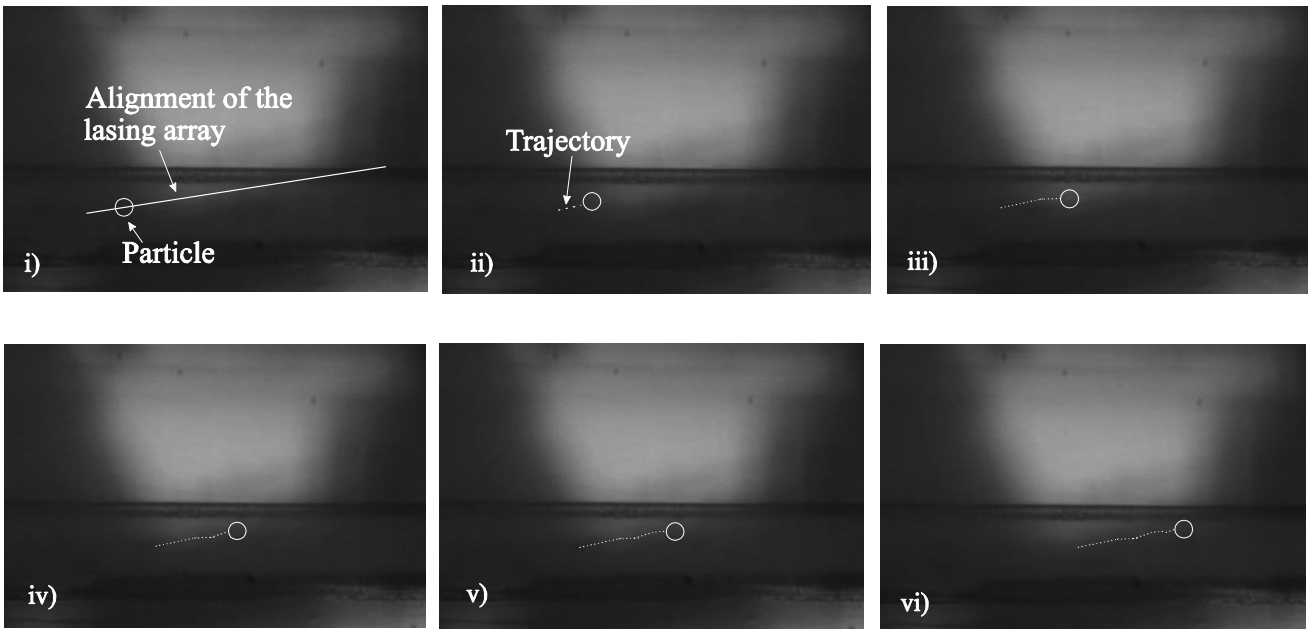


Fig. 10. Snapshots of a continuous deflection experiment with a 10 µm polystyrene particle. The particle is slowly redirected to the upper part of the channel while passing through the channel.

B. *Experimental Results*

The experiment is performed with 10 µm diameter polystyrene particles. Figure 10 depicts the trajectory of the particles when the laser array is in operation. The alignnment of the laser array is indicated. The particles are moving from left to right with a certain velocity. In Fig. 10 (i), a 10 µm particle approaches the laser array and it is attracted by the laser beams. As the particles flows forward, it is further deflected (Figs. 10 (ii)-(vi)) by the attraction of the tilted laser arrays. As a result, the particles are deviated from their initial path.

VII. Conclusion

In this paper, the fabrication of a VCSEL-based miniaturized module for microparticle manipulation has been presented. The module requires the fabrication of single-mode integration-capable VCSEL arrays. Single-mode emission is realized by the surface-relief technique and the integration capability is achieved by deep bonding grooves. Moreover, the VCSEL arrays are fabricated as densely packed arrays to enable continuous optical deflection experiments. After the fabrication of the VCSEL arrays, they are sawed to form VCSEL chips. Then the chips are soldered on silicon chip carriers containing heat sinks. Next, the p-fanout tracks of the heat sink are connected to the bondpads by wire bonding. A microfluidic chip is then integrated with the laser module by fixing it directly above the VCSEL facets. Next, the chip is assembled on a PCB, which is soldered with wires for electrical connection to an external current source. In this way, the miniaturized module is realized.

After the fabrication of the module, optical deflection experiments are performed with 10 µm polystyrene particles and particle deflection is successfully observed. The final goal is to use the module for biological cell sorting.

Acknowledgment

We would like to thank the Institute of Optoelectronics for providing the materials and equipment which allowed to do this research at Ulm University.

References

[1] A. Ashkin, “Acceleration and trapping of particles by radiation pressure,” Phys. Rev. Lett., Vol. 24, pp. 156–159, 1970.

[2] K. Dholakia, P. Reece, and M. Gu, “Optical manipulation,” Chem. Soc. Rev., Vol. 37, pp. 42–55, 2008.

[3] J.C. McDonald and G.M. Whitesides, “Poly(dimethylsiloxane) as a material for fabricating microfluidic devices”, Acc. Chem. Res., Vol. 35, pp. 491–499, 2002.

[4] A. Bergmann, N.I. Khan, J.A. Martos Calahorro, D. Wahl, and R. Michalzik, “Hybrid integration of VCSELs for miniaturized optical deflection of microparticles ”, in Semiconductor Lasers and Laser Dynamics V, K. P. Panajotov, M. Sciamanna, A. A. Valle, R. Michalzik (Eds.), Proc. SPIE 8432, pp. 843204-1–12, 2012.

[5] K. Svoboda and S. M. Block, “Biological applications for optical forces”, Annu. Rev. Biophys. Biomol. Struct., Vol. 23, pp. 247–285, 1994.

[6] A. Kroner, I. Kardosh, F. Rinaldi, and R. Michalzik, “Towards VCSEL-based integrated optical traps for biomedical applications”, Electron. Lett., Vol. 42, pp. 93–94, 2006.

[7] B. Weigl, M. Grabherr, C. Jung, R. Jäger, G. Reiner, R. Michalzik, D. Sowada, and K.J. Ebeling, “High-performance oxide-confined GaAs VCSELs”, IEEE J. Select. Topics Quantum Electron., Vol. 3, no. 2, 1997.

[8] N. I. Khan, A. Bergmann, and R. Michalzik, “Fabrication of integration-capable surface-relief VCSEL arrays for miniaturized optical manipulations of microparticles”, in Proc. 2nd Int. Conf. on Advances in Electr. Engin., ICAEE, pp. 368-373. Dhaka, Bangladesh, Dec. 2013.

[9] UV curing [Online]. Available at: http://www.fusionuv.com/uvlearningcenter.aspx?id=206 [Accessed on 26/06/2014]

Design and Simulation of Three Wavelength Terahertz GaN Quantum Cascade Laser

Md. Shahadat Hasan Sohel,* A. F. M. Saniul Haq,
and Muhammad Anisuzzaman Talukder
Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology, Dhaka 1000, Bangladesh
*shasan@eee.buet.ac.bd

***Abstract*—Quantum cascade laser (QCL) is a promising candidate for comparatively less developed terahertz spectral range. GaN based QCLs provide significant advantages over traditional GaAs based QCLs due to their higher LO phonon energy. Multiple wavelength hetero active region cavity offers various advantages like multiple emission wavelength for spectroscopy, nonlinear optical phenomena, and second harmonic generation. In this paper, we propose and design a multiple wavelength hetero cavity terahertz GaN based QCL emitting at 1.7 THz, 4.3 THz, and 7.2 THz. We present detail modeling of the system and final gain spectrum. We observe high gain (~ 250 cm^{-1}) with multiple wavelength output emission. We also observe the sensitivity of the emission on temperature and find that the gain decreases sharply with increase in the temperature.**

***Index Terms*—Quantum Cascade Laser, Hetero Cavity, Multiple Wavelength, Terahertz QCL**

I. Introduction

The terahertz frequency range, loosely defined as the frequencies from 300 GHz to 1 THz has historically been immature as it lacks suitable radiation sources, detectors, and transmission technology [1]. However, terahertz technology is important for its potential applications, including astrophysics and atmospheric science, biological and medical sciences, security screening and illicit material detection, nondestructive evaluation, communications technology, and ultrafast spectroscopy. Due to significant application potentials, many researchers have explored ways to find sources for Terahertz emission. Photonic approaches to direct terahertz generation are limited due to the lack of appropriate materials with sufficiently small bandgaps, e.g., the longest wavelength lead salt laser diodes do not extend below 15 THz [2]. Unlike the conventional bandgap structure of semiconductor lasers, in quantum cascade lasers (QCLs), quantum confinement splits the conduction band into a number of distinct subbands [2]. The energy spacing of the lasing subbands determines the radiation frequency, allowing in principle to produce light at arbitrarily long wavelengths and even extend into the terahertz frequency range [2]. Terahertz QCL was first demonstrated in 2002 and remarkable progress on terahertz QCLs has been made since then [3].

In QCLs, the emission wavelength can be chosen over a very wide wavelength range by choice of the quantum well thicknesses in the active regions. Typically 30–50 periods of alternated active and injector regions are cascaded to make the laser core. As electrons transport through the cascade, active regions are pumped sequentially. Thus, as the various active regions do not compete for pumping energy, the combination of wavelength agility and cascading principle can advantageously be used for multi-wavelength emission [4]. Heterogeneous active regions, i.e., periods emitting at different wavelengths have been employed to obtain broadband tunable lasers for spectroscopic applications [4]. Straub et al. proposed a two-wavelength QCL with heterogenous cascades emitting at 8 μm and 9.5 μm. They have used traditional InGaAs/AlInAs material system for operating in the mid-IR range. No work has been reported on heterogenous cascade in terahertz frequency range.

In this work, we have proposed a novel GaN based heterogeneous active region QCL design, which will emit in three distinct frequencies, namely 1.7 THz, 4.3 THz, and 7.2 THz. GaN based active region enables us to operate in the terahertz frequency range. With its comparatively higher LO-phonon resonance energy, QCL can operate at comparatively higher temperature. We have done detail modeling of GaN based QCL systems. We also have calculated detailed electronic transport for the system. We have calculated the gain for each of the emission wavelengths and also observed the effect of change of temperature on the output emission. The remaining of this paper is organized as follow: Section II describes the mathematical modeling required for the simulation of various parameters. Section III presents the proposed design and its various aspects. Section IV describes the characterization for the proposed design. Section V concludes the paper and discusses the scope for further extension of the work.

II. Theoretical Modeling

We want to calculate the gain spectrum of a GaN based Quantum cascade structure. To find the the output emission wavelength, first we must find the mode that will lase. The axial modes that are supported by a laser cavity of length L are given by [5]

$$f = m\frac{c}{2nL}, \tag{1}$$

where c is the speed of light, n is the index of refraction, and m is an integer that denotes mode number. In practice, a number of axial modes lie within the gain linewidth of QCLs. However, in QCLs, which have a homogenously broadened gain medium, the mode that first experiences gain greater than loss becomes the sustained mode in the cavity.

Now, we need to calculate the gain of the axial mode. The peak modal gain of a QCL for radiative transition from subband j to subband i is given by [6]

$$G_M = \frac{4\pi q^2}{\varepsilon_0 n\lambda}\frac{z_{ij}^2}{2\gamma_{ij}L_p}(n_j - n_i), \tag{2}$$

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

where q is the charge of an electron, ε_0 is the permittivity of free space, n is the index of refraction, λ is the transition wavelength, $2\gamma_{ij}$ is the full width at half maximum (FWHM) linewidth, z_{ij} is the dipole matrix element, L_p is the length of a period, Γ is the confinement factor, and n_i, n_j are the carrier densities in subbands i and j, respectively. The gain spectrum is found by multiplying the peak modal gain by the Lorentzian lineshape, L_{ij}, given by

$$G = G_M L_{ij}. \tag{3}$$

The calculation of the gain spectrum using Eq. (2) requires the determination of the dipole matrix element z_{ij} between the levels i and j, the FWHM linewidth γ_{ij} for the transition from level i to level j, and the carrier densities in the subbands i and j. The methods for calculations of z_{ij} and γ_{ij} can be found in Ref. [6] and Ref. [7], respectively. The carrier density in each of the subbands is calculated using the rate equation approach as discussed in ref. [8].

Next, we have considered electrical and optical properties of GaN based materials. The temperature dependent bandgap energy E_g can be found by methods described in [9]. The polarization, which is responsible for internal electric field of GaN based materials consists of spontaneous and piezoelectric components. The spontaneous polarization in C/m^2 for the $Al_xGa_{1-x}N$ alloy as function of the Al content x can be approximated by [10]

$$P_{spAlGaN}(x) = -0.090x - 0.034(1-x) + 0.021x(1-x) \tag{4}$$

Wurtzite group III nitrides exhibit an additional piezoelectric polarization P_{piezo} when stress is applied. The piezoelectric polarization can be written as

$$P_{piezo} = 2\frac{a-a_0}{a_0}\left(e_{31} - e_{33}\frac{C_{13}}{C_{33}}\right) \tag{5}$$

The parameters used for the calculation of polarization are taken from references [9] and [11].

The existence of both built-in polarization fields in wurtzite III Nitride semiconductors leads to energy band bending. The electrostatic fields in each layer j-th as a result of total polarization fields can be expressed as [12],

$$E_j = \frac{\sum_k (P_k - P_j)\frac{l_k}{\epsilon_k}}{\epsilon_j \sum_k \frac{l_k}{\epsilon_k}} \tag{6}$$

where P is the total macroscopic polarization, ϵ is the static dielectric constant, and l is the thickness of each layers (k-th, j-th). The subscripts k and j correspond to the k-th and j-th layers.

To ensure zero average electric field in the layers, note that the electric field expression in (6) needs to satisfy the periodic boundary conditions as follows [12]

$$\sum_j l_j E_j = 0. \tag{7}$$

III. Design

We propose to use a GaN-based system with large LO-phonon energy of 90 meV for THz QCLs. The advantages are threefold. First, the large LO-phonon energy in a GaN-based system can reduce the thermal population of the lower laser state. Second, ultrafast LO-phonon scattering in GaN/AlGaN QWs can be used for the rapid depopulation of the lower laser state. Third, the large LO-phonon energy can also increase the lifetime of the upper laser state by reducing the relaxation of electrons with higher in-plane kinetic energy via emission of a LO-phonon [13].

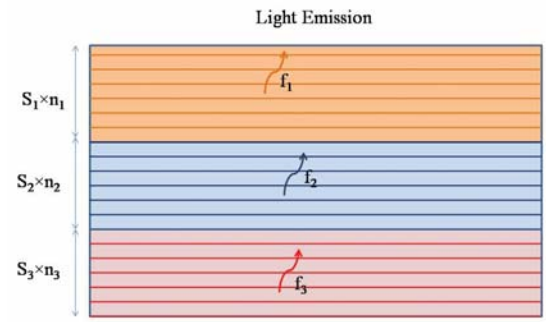


Fig. 1: Schematic of the overall design structure. Here, structures S_1, S_2, and S_3 periods are repeated n_1, n_2, and n_3 times respectively and emit f_1, f_2, and f_3 frequency respectively.

In our design, three structures S_1, S_2, and S_3 each emitting a different wavelength are stacked together as schematically shown in Fig. 1. The active waveguide core consists of 30 period of each structures. S_1 is designed to emit at 7.2 THz, S_2 is designed to emit at 4.3 THz, and S_3 is designed to emit at 1.7 THz. We use GaN for the quantum wells and ternary $Al_{0.2}Ga_{0.8}N$ alloys for the barriers. Each structure is so designed that they satisfy the periodic boundary conditions explained by Eq. (7).

All three structures S_1, S_2, and S_3 consist of four quantum wells. The layer sequence for one period of S_1 is (in nm) 5.7/**0.62**/6.18/**0.73**/2.85/**0.52**/3.6/**0.83**, beginning from an injection layer for a depopulation layer. The layer sequence for one period of the S_2 is (in nm) 5.7/**0.62**/6.18/**0.73**/2.8/**0.52**/4.2/**0.83**, beginning from an injection layer for a depopulation layer. The normal and bold scripts denote the GaN wells and the AlGaN barriers, respectively. In both S_1 and S_2, the 6.18 nm thick GaN wells are n-doped to 2×10^{11}cm^{-2}.

S_1 and S_2 are variation of a same structure, differing only in active region. The active region of S_1 consists of two quantum wells of 2.85 nm and 3.6 nm, whereas the two quantum wells in S_2 are of 2.8 nm and 4.2 nm. The wider active region results in a lesser energy difference between two energy levels and hence we get a smaller frequency output in S_2 than in S_1. The slight modification of the other active region layers results in better electron transition to create sufficient population inversion.

The layer sequence for one period of the structure S_3 is (in nm) 7.1/**0.8**/7.6/**0.8**/2.8/**0.8**/3.6/**0.9**, beginning from an injection layer for a depopulation layer. The 7.6 nm thick GaN well is n-doped to 5×10^{10}cm^{-2}. Here, S_3 has a different design for injector region than that of S_1 and S_2. We also note that the barrier width is changed. The injector wells are 7.1 nm and 7.6 nm in S_3, instead of 5.7 nm and 6.18 nm in S_1 and S_2. The active region barrier is also changed to 0.8 nm from 0.52 nm. These changes result in a much smaller emission frequency than that of the other two structures.

As S_2 and S_3 will have different structures and hence dif-

ferent current density, we need a special transitional structure for electron transition from last period of S_2 to first period of S_3. We have designed a special injector region as the basic structures' energy levels do not match. The injector level layer sequences are 2.85/**0.7**/3.8.

IV. Characterization

In this section, we will discuss the characteristics of the proposed structures. Figures 2, 3 and 4 show the simulated bandstructures and moduli squared of wavefunctions for an applied bias of 67.4 kV/cm for one and a half period of structures S_1, S_2, and S_3, respectively, at temperature 100 K.

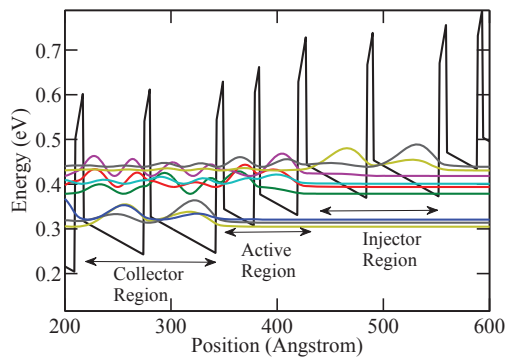


Fig. 2: Bandstructure and the moduli squared of the relevant wave functions for three periods of S_1 structure

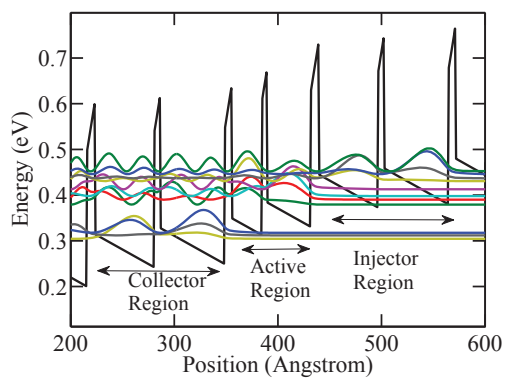


Fig. 3: Bandstructure and the moduli squared of the relevant wave functions for three periods of S_2 structure

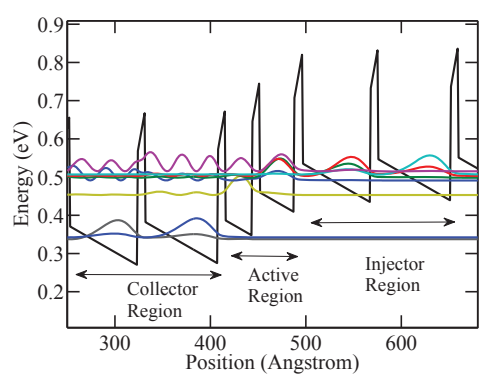


Fig. 4: Bandstructure and the moduli squared of the relevant wave functions for three periods of S_3 structure

The structures are grown on one substrate and on top of each other. This means all the structures are connected together like a series circuit. Therefore, same current runs through the structures and the current is limited by the smallest current produced by any structure [14]. To ensure current matching between the structures, we have calculated the current densities in individual structures for different biases, which is shown in Fig. 5. As S_1 and S_2 are similar structures varying only in active regions, currents in S_1 and S_2 are close to each other. Here, the collector levels and injector level energy levels closely match. So, injector level of S_2 can act as the collector level of S_1. Hence, carrier transition from S_1 to S_2 can be made directly without the need of any special injector region. We can see that current density in S_3 is significantly different from S_2. As S_2 and S_3 varies in the injector regions, injector region of S_3 can not work as the collector level of S_2. So, we have designed a special injector region for transition between S_2 and S_3, which consists of two quantum wells of 2.85 nm and 3.85 nm, separated by a barrier of 0.7 nm. This special injector region matches the energy levels of active region of S_2 to energy levels of injector level of S_3. As a result, we have a closely matched current levels in the combined structure, which is shown in Fig. 6.

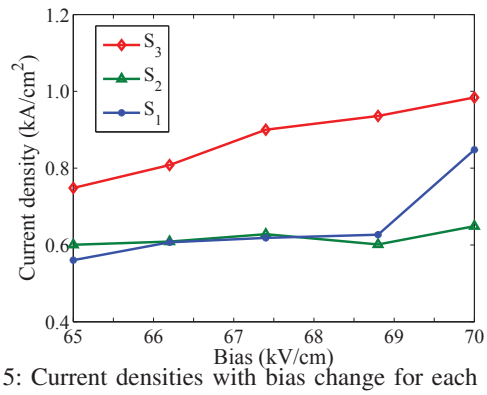


Fig. 5: Current densities with bias change for each structure

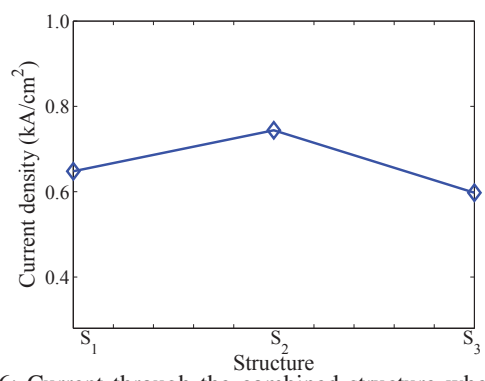


Fig. 6: Current through the combined structure when S_1, S_2, and S_3 are cascaded and biased at 67.4 kV/cm

We have calculated the population density in each of the structures. The population densities for S_2 is shown in Fig. 7. Here, we note that significant amount of population inversion was achieved in this structures. From Eq. (2) we can see that population inversion is directly proportional to the gain of the laser. As a result, we will be able to achieve high gain from the design. Again, Eq. (2) shows that modal gain is inversely proportional to wavelength. At terahertz range, the wavelength is significantly less than mid-IR range wavelength. So, this fact will also result in higher gain. Overall gain spectrum shown

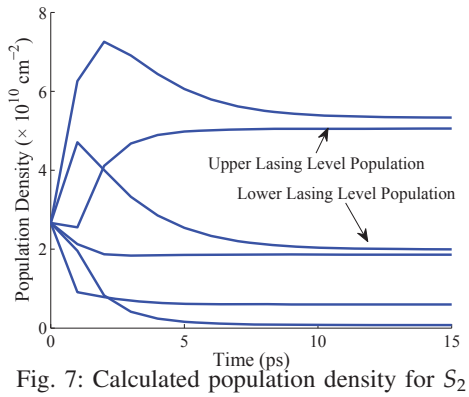


Fig. 7: Calculated population density for S_2

in Fig. 8 agrees with these result. This higher gain will enable the emission to overcome the loss in the extended cavity and will result in sustained high power emission.

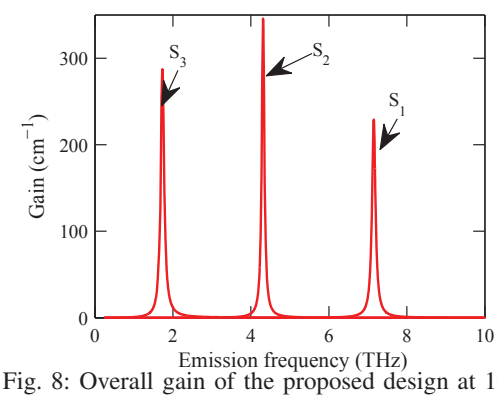


Fig. 8: Overall gain of the proposed design at 100 K

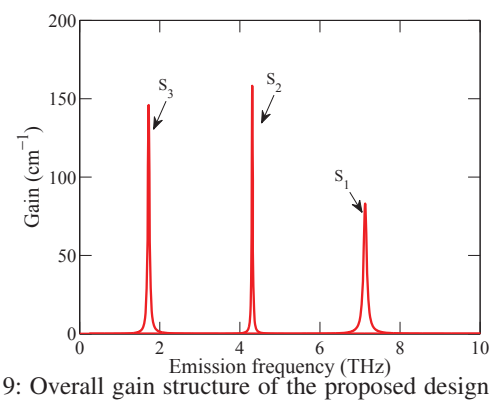


Fig. 9: Overall gain structure of the proposed design at 150 K

We have also performed the simulation at higher temperatures. Figure 9 shows the output emission of the laser at a higher temperature of 150 K. At this temperature, we note that the emission has significantly lower gain than the emission at 100 K temperature. This lower gain can be attributed to the thermal backfilling of electrons in the lower lasing level, which lowers the population inversion approximately by 40%. As the gain is directly proportional to the population inversion, gain also reduces by the same amount. If we further increase the temperature to 200 K the laser fails to lase, as the gain decreases more and becomes less than loss.

V. Summary

We have proposed a novel multiwavelength heterostructure GaN based Terahertz design. The proposed design simultaneout detail carrier transport calculation and found significant population inversion in each structure. The current between the structures is matched so that the gain of one structure will not suffer for the other. We have shown that output emission of this laser will have significantly high gain in the range of ~250 cm^{-1}. This high gain will result in higher output power. This can be very useful in terahertz imaging and spectroscopy applications. We also found that higher temperature results in lower gain due to thermal backfilling of electrons into the lower lasing level.

References

[1] B. S. Williams, "Terahertz quantum-cascade lasers," *Nature Photonics*, 2007.

[2] X. Yin, B. W.-H. Ng, and D. Abbott, *Terahertz Imaging for Biomedical Applications*. Springer, 2012.

[3] H. Yasuda, T. Kubis, I. Hosako, and K. Hirakawa, "Design of terahertz quantum cascade lasers for high temperature operations using non-equilibrium green's function method," *PIERS Proceedings, Malaysia*, 2012.

[4] A. Straub, C. Gmachl, and T. Mosely, "Two-wavelength quantum cascade lasers with heterogeneous cascades," in *Optoelectronic and Microelectronic Materials and Devices, IEEE conference*, 2002.

[5] A. Siegman, *Lasers*. University Science Books, 1986.

[6] H. C. Liu and F. Capasso, *Intersubband transitions in Quantum well*. Academic press, 2000.

[7] T. Unuma, M. Yoshita, T. Noda, and H. Sakaki, "Intersubband absorption linewidth in GaAs quantum wells due to scattering by interface roughness, phonons, alloy disorder, and impurities," *J. App. Phys.*, vol. 93, 2003.

[8] D. Indjin, P.Harrison, R. Kelsall, and Z. Ikonic, "Self-consistent scattering theory of transport and output characteristics of quantum cascade lasers," *Journal Of Applied Physics*, vol. 91, 2002.

[9] I. Vurgaftman and J. R. Meyer, "Band parameters for nitrogen-containing semiconductors," *J. Appl. Phys*, vol. 94, 2003.

[10] O. Ambacher, J. Majewski, C. Miskys, A. Link, and M. Hermann, "Pyroelectric properties of Al(In)GaN/GaN hetero- and quantum well structures," *J. Phys.: Condens. Matter*, vol. 14, 2002.

[11] E. Baumann, "Near infrared intersubband absorption and photovoltaic detection in GaN/AlN multi quantum well structures," Ph.D. dissertation, Univ. of Neuchatel, 2007.

[12] H. Zhao, R. A. Arif, Y.-K. Ee, and N. Tansu, "Self-Consistent Analysis of Strain-Compensated InGaN-AlGaN Quantum Wells for Lasers and Light-Emitting Diodes," *IEEE Journal of Quantum Electronics*, vol. 45, 2009.

[13] G. Suna, Richard, A. S. b, and J. B. Khurginc, "Active region design of a terahertz GaN/AlGaN quantum cascade laser," *Elsevier Superlattices and Microstructures*, vol. 37, 2005.

[14] G. Cherucheril, S. March, and A. Verma, "Multijunction solar cells," Ph.D. dissertation, Iowa State University, 2011.

Improved texture retrieval by combining different variants of Local Binary Patterns

Nabeel Mohammed,[1,*] Sohel Rana,[2]
[1,2]Department of CSE, University of Asia Pacific
Dhanmondi, Dhaka, Bangladesh
*nabeel.mohammed@uap-bd.edu

***Abstract*—This paper proposes the combination of different variants of local binary patterns for texture retrieval. Recent studies have shown that LBP features extracted at multiple resolutions give the best performance for a standard texture retrieval collection. Techniques have been proposed to create a multi-dimensional histogram of these features. In this study we use a simpler approach. We hypothesize that the different variants of LBP may actually extract slightly different but useful image information. If this were the case, then using them in combination will result in further improved performance, compared to using each one independently. We demonstrate this by using LBP, LBP^{ri}, and LBP^{u2} features in combination. We take ideas from the GNU Image Finding Tool (GIFT), and use separate normalisation to ensure the different feature vector lengths do not bias the system towards one of the features. We performed experiments on the two standard texture collections. Our results demonstrate that combining the different LBP features do indeed result in improved performance. In fact, for one of the collections (Outex_TR_00000) collections, using our method give better performance than the current state-of-the-art.**

***Index Terms*—Image retrieval, Local binary patterns, Separate normalisation**

I. Introduction

This work is inspired by the research reported in [1] and [2], where the authors extract LBP features at multiple resolutions, and apply them to an image retrieval task. Both these studies report their results on the Outex_TR_00000 texture collection [3]. [1] tried 16 variants of local binary patterns (LBP) along with Gabor filters, Tamura features, and two different GLCM features to find the most effective texture feature for the Outex_TR_00000 image collection. In reality the numbers are even more extensive, as they tried the LBP variants in different combination of resolutions, for a total of 38 LBP features. These were tested using three different distance measures: the Bhattacharya, chi-squared and L_1 distances. The non-LBP features were only tested using the chi-squared distance. There was in total 118 different attempts to find the best performing texture feature. After this thorough search, they found the best performing texture feature for this collection was a concatenation of $LBP_{1,8}$, $LBP_{3,8}$ and $LBP_{5,8}$, which gave an accuracy of 63.74% when using the L_1 distance.

In [2], the authors wanted to address the shortcomings of a scheme which concatenated the feature vectors, and proposed a multi-dimensional LBP descriptor, which they state performs better than the concatenated version. However, there is some confusion at this point. In the first paper [1], the authors report that the average retrieval rate (ARR) of $LBP_{1,8} + LBP_{3,8} + LBP_{5,8}$ is 63.74% . However in their latter paper [2], when reporting the results of the multi-dimensional LBP feature, they report the ARR of $LBP_{1,8}+LBP_{3,8}+LBP_{5,8}$ as 58.85% and the ARR of the multi-dimensional LBP feature as 63.04%. In our experiments we could found an ARR of 63.12% for $LBP_{1,8} + LBP_{3,8} + LBP_{5,8}$.

Whichever may be the correct number, what is noticeable in both the papers is that the authors concentrated on using the same type of LBP feature (whether it be LBP, LBP^{ri}, LBP^{u2} etc.) in different resolution combinations. Given that they found a combination of $r = 1, 3$ and 5 to perform best, we don't want to repeat their work. Instead the contribution of this study is to try to see whether a combination of the different types of LBP features, extracted at these resolutions, can give better performance.

The following section gives a brief description of LBP. We follow this with a description of our experiments. We then describe our results, which justify our claims.

II. Background

Local Binary Patterns (LBP) were first proposed in [4]. Some variants were then presented in [5]. Since then it has become a widely used texture features and has gained wide acceptance due to good performance in different image processing applications and also through simplicity of calculation. Here we will discuss the version presented in [5].

LBP features are calculated over a circular neighbourhood around a central pixel. The size of the number of pixels in the neighbourhood (P) and its radius (R) can be varied. For this discussion we will assume that the images are grey scale.

The authors begin by defining a texture T in a local neighbourhood of an image as the joint distribution of the gray scale pixel values of $P, (P > 1)$ pixels

$$T = t(g_c, g_0, ...g_{P-1}) \tag{1}$$

Where g_c is the grey level value of the central pixel in the neighbourhood and $g_p(p = 0, ..., P - 1)$ are the grey level values of P equally spaced pixels at a distance of radius R, forming a circular neighbourhood. The grey level values which do not correspond to an actual pixel can be found by interpolation. To achieve invariance to luminance the value of the center pixel g_c is subtracted from the neighbourhood pixels G_p, giving

$$T = t(g_c, g_0 - g_c, ...g_{P-1} - g_c) \tag{2}$$

Assuming that the differences $g_p - g_c$ is independent of g_c, Equation 2 can be refactored to

$$T \approx t(g_c)t(g_0 - g_c, ...g_{P-1} - g_c) \tag{3}$$

As $t(g_c)$ here describes the overall luminance in the image, and as LBP features want to gain invariance to luminance, the

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

g_c components is dropped and the texture of the neighbourhood is approximated by

$$T \approx t(g_0 - g_c, ...g_{P-1} - g_c) \tag{4}$$

This data can be used to produce a P dimensional histogram which can be used as a feature. But to achieve invariance to scaling of the grey scale values just the sign of the differences are considered.

$$T \approx t(s(g_0 - g_c), ...s(g_{P-1} - g_c)) \tag{5}$$

where

$$s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{6}$$

Incorporating a binomial factor operator 2^p for each sign, Equation 5 can be converted to a number. This number can be used as a representation of local texture. [5] and other works have referred to this in terms of P and R as $LBP_{R,P}$.

$$LBP_{R,P} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \tag{7}$$

This process is illustrated in Figure II, showing how the LBP pattern can be calculated for an image region with the following grey scale values:

30	25	10
85	50	120
70	60	100

The calculation is performed around the central pixel, with a grey scale value of 50, with $R = 1$ and $P = 8$. Three neighbours have a pixel value less than 50, resulting in a corresponding 0 in the pattern entry. The others pattern entries have a value of 1, as their pixel value is larger than that of the central pixel. The order of how the pattern is read is immaterial as long as it is read consistently. In this example, we can start reading from the top right in a clockwise manner to get 01111100 as the pattern with $LBP_{P,R} = 124$ using Equation 7.

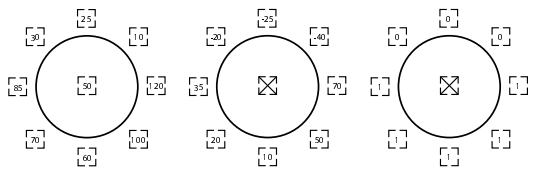


Fig. 1. Illustration of the LBP calculation.

An extension to this is presented by the authors based on the observation that certain binary patterns occur frequently in texture images. These "uniform" patterns have a small number of spatial transitions in their circular neighbourhood. These can be thought of as templates which can help identify microstructures in an image such as bright spots, dark spots and edges. Equation 8 is used to measure the uniformity of a region.

$$U(LBP_{R,P}) = |s(g_{P-1} - g_c)| + \sum_{p=1}^{P-1} |s(g_p - g_c) - s(g_{p-1} - g_c)| \tag{8}$$

In a scheme which is called LBP^{u2}_{PR}, patterns which have a U value of less than or equal to 2 are placed in a bin corresponding to their number. Other patterns are all placed in one bin. So, for a an $LBP_{R,P}$ value which has 8 bits, there are 58 uniform patterns, resulting in a histogram of 59 bins when using LBP^{u2}_{PR} .

$LBP_{R,P}$ is easily transformed into its rotation invariant version $LBP^{ri}_{R,P}$ by rotating the pattern to maximise the number of 0s in the most significant bits starting from g_{p-1}. So, the pattern 01111100 found from the illustration in Figure II would become 00011111. Using equation 9 each uniform $LBP^{riu2}_{R,P}$ pattern gets a unique label assigned to it, which depends on the number of 1s in the pattern. Every non-uniform pattern gets the label $P + 1$. This scheme produces $P + 2$ unique values. The image feature is calculated as a histogram of $LBP^{riu2}_{R,P}$ values, summed over a texture region.

$$LBP^{riu2}_{R,P} = \begin{cases} \sum_{p=0}^{P-1} s(g_p - g_c) & if\ U(LBP_{R,P}) \leq 2 \\ P + 1 & otherwise \end{cases} \tag{9}$$

Although LBP^{ri}, LBP^{u2} and LBP^{riu2} can all be calculated from LBP, upon reflection it is clear that each of them represent different image information, however slight the difference may be. The classic LBP histogram is simply a histogram of texture patterns. In achieving rotation invariance LBP^{ri} will also help identify similar texture patterns, even if they are rotated. The uniform patterns are interesting in that they do not claim to find similar patterns, but instead it can be thought of as a statistic of patterns which fulfil the uniform criteria. LBP^{riu2} would then be a the rotation invariant version of this statistic. Once we think of the different LBP features in this light, we hypothesize that using them in combination will give better overall performance, compared to using just one. Also, as they can all be extracted from LBP, the computational overhead is not significant. In this next section we describe the experiments we conducted to test our hypothesis.

III. Experiments

A. Texture collections

We present our results using two texture collections. The Outex_TR_00000 collection is used to facilitate comparison with [1] and [2].

We also present results using the VisTex collection, to show that the findings of this study can be repeated on other texture collections. The version of the VisTex collection is one used in [6], [7], [8], which has 1670 texture images in 124 relevance classes.

B. Features extracted

As mentioned previously, a large number of extensions and variants of LBP have been proposed over the last few years. Showing the performance of these variants in combination is a mammoth task and not required to establish our claim. We use the variants proposed in [5], as it is sufficient to make our case. Each of the variants were extracted at resolutions of

$1, 3$ and 5 pixels, as this resolution combination was found to be the best by [1] and [2]. Table I lists the features we have extracted. As writing the full details is space consuming, we have decided to label them for future reference in this paper.

Feature combinations	Feature Vector length	Label
$LBP_{1,8} + LBP_{3,8} + LBP_{5,8}$	768	LBP-Doshi
$LBP^{ri}_{1,8} + LBP^{ri}_{3,8} + LBP^{ri}_{5,8}$	108	LBP-Doshi-ri
$LBP^{u2}_{1,8} + LBP^{u2}_{3,8} + LBP^{u2}_{5,8}$	177	LBP-Doshi-u2
$LBP^{riu2}_{1,8} + LBP^{riu2}_{3,8} + LBP^{riu2}_{5,8}$	30	LBP-Doshi-riu2

TABLE I
THE DIFFERENT TYPES OF LBP FEATURES EXTRACTED TO BE USED IN COMBINATION. THE '+' INDICATES CONCATENATION.

C. Distance Measurement

[1] and [2] report that the multi-resolution LBP features performed best when used with the L_1 distance measure, as defined in equation 10, where V_p and V_q are two feature vectors of n dimensions, and j is the index into the vectors.

$$d_1(V_p, V_q) = \sum_{j=1}^{n} |V_p(j) - V_q(j)| \tag{10}$$

All results reported in this paper use this distance measure to judge the similarity between two images.

D. Combining features

Combining features by using simple concatenation is problematic. One reason for it has been discussed in [2], however that is not the main concern of this study. We have a different problem, where we are looking to combine features, whose feature vector lengths are quite different. Table I lists the length of the feature vectors of each feature extracted for this study.

If we were to use a simple concatenation scheme, the larger feature vectors would dominate in the distance measurements, giving results biased towards certain features.

To implement a more sensible system, we take ideas from the GNU Image Finding Tool (GIFT) system [9], which uses separate feature normalisation as a default setting. This has been cited as one of the reasons why it was shown to be the best of seven CBIR systems tested by [10]. Our normalisation scheme is much simpler than GIFT's, but the main theme is to ensure features are given their due weight. Our normalisation scheme shown in Equation 11, where the distance between images I_p and I_q are calculated using k features $V_{p,1}..V_{p,k}$ and $V_{q,1}..V_{q,k}$. The distance is calculated as the sum of the distances of the individual normalised features.

$$distance(I_p, I_q) = \sum_{j=1}^{k} d_1 \left(\frac{V_{p,j}}{sum(V_{p,k})}, \frac{V_{q,k}}{sum(V_{q,k})} \right) \tag{11}$$

The following features combinations have been used for this study.

- LBP-Doshi
- LBP-Doshi-ri
- LBP-Doshi-u2
- LBP-Doshi-riu2
- LBP-Doshi and LBP-Doshi-ri
- LBP-Doshi,LBP-Doshi-ri and LBP-Doshi-u2
- LBP-Doshi, LBP-Doshi-ri, LBP-Doshi-u2 and LBP-Doshi-riu2

E. Evaluation

[1] use a measure of the average percentage of relevant images retrieved in the top 20 retrieved images. In this case, all the images in the database are used as query images. This is no different from calculating the average retrieval rate (ARR) used by [11] and [12] to report results for a subset of the VisTex collection. The recall rate (RR) for a query image I_q calculated for the first N images retrieved is calculated as

$$RR_N(I_q) = \frac{number\ of\ relevant\ images\ retrieved\ by\ N}{N} * 100 \tag{12}$$

and ARR for a set of query images Q is then simply

$$ARR_N = \frac{\sum_{q \epsilon Q} RR_N(I_q)}{size(Q)} \tag{13}$$

We will report ARR_N results, using all the images in the database as queries, where N is the number of images per relevance class in the collection. Also, we present precision-recall (PR) graphs [13] where for each unique recall value, the precision values are averaged over all query images. Again, for these graphs every image in the collection is used as a query.

IV. RESULTS

Table II and III lists the ARR of the different combinations of features tried in this study. Using LBP-Doshi, LBP-Doshi-ri together in combination also improves performance, compared to using just LBP-Doshi. However adding the information gathered from LBP-Doshi-u2 consistently improves the performance, and the best results are obtained when using LBP-Doshi, LBP-Doshi-ri and LBP-Doshi-u2 together in combination. In fact for the Outex_TR_00000 collection, the results obtained using a combination of the three features IS 65.30% is better than the currently reported state-of-the-art results of 63.74% by [1].

Features	ARR
LBP-Doshi	63.12%
LBP-Doshi-ri	62.1%
LBP-Doshi-u2	62.28%
LBP-Doshi-riu2	58.816%
LBP-Doshi and LBP-Doshi-ri	**64.7%**
LBP-Doshi,LBP-Doshi-ri and LBP-Doshi-u2	**65.3%**
LBP-Doshi, LBP-Doshi-ri, LBP-Doshi-u2 and LBP-Doshi-riu2	62.78%

TABLE II
PERFORMANCE FOR THE OUTEX_TR_00000 COLLECTION. THE RESULTS IN BOLD SHOW THE BEST RESULTS.

In Figures 2 and 3 we see that the improved performance is also reflected in the precion-recall graphs. Again combining LBP-Doshi,LBP-Doshi-ri and LBP-Doshi-u2 gives the best performance.

Features	ARR
LBP-Doshi	77.10%
LBP-Doshi-ri	75.05%
LBP-Doshi-u2	76.52%
LBP-Doshi-riu2	71.73%
LBP-Doshi and LBP-Doshi-ri	**78.13%**
LBP-Doshi,LBP-Doshi-ri and LBP-Doshi-u2	**78.59%**
LBP-Doshi, LBP-Doshi-ri, LBP-Doshi-u2 and LBP-Doshi-riu2	76.18%

TABLE III
PERFORMANCE FOR THE VISTEX COLLECTION. THE RESULTS IN BOLD SHOW THE BEST RESULTS.

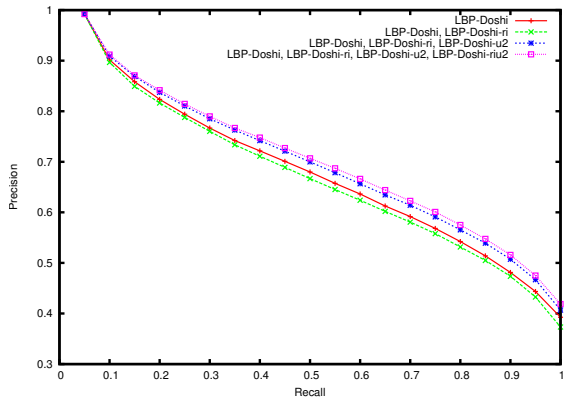


Fig. 2. Precision-Recall graphs comparing the performance of the different combinations of LBP feature types for the Outex_TR_00000 collection.

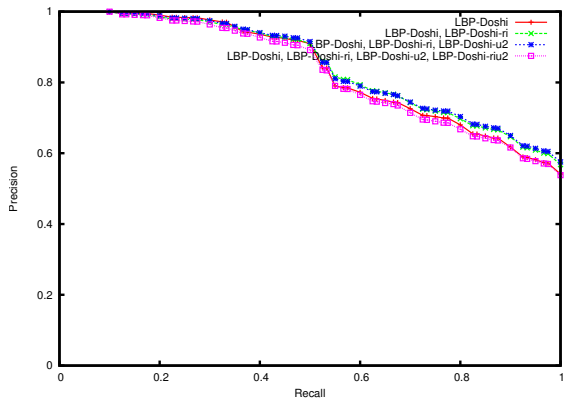


Fig. 3. Precision-Recall graphs comparing the performance of the different combinations of LBP feature types for the VisTex collection.

For both texture collections we see that adding LBP-Doshi-riu2 to the combination reduces performance. This may be understood by studying the LBP^{riu2} feature. LBP^{ri} rotates the LBP pattern so that there is a maximum number of zeros in the most significant bits, thus causing some texture patterns to be mapped to a single LBP. Applying the uniformity criteria (Equation 8) causes a further merging of texture patterns. This two step loss of image information can be a cause of degraded performance.

V. CONCLUSION

This paper explored the possibility of using different variants of LBP in combination, based on the hypothesis that if each variant extracts different but useful image information, then using them in combination will give better performance, compared to using a single variant. We have conducted our experiments on two standard texture collections. We extracted four variants of LBP features at multiple resolutions. We have shown that combining LBP, LBP-ri, and LBP-u2, through separate normalisation, gives improved performance compared to just using these features individually, for two standard texture collections. In fact for the Outex_TR_00000 collection, our method outperforms the current state-of-the-art.

In conclusion, we believe it is a flawed approach to claim one individual feature type outperforms another individual feature type, as done in [1] and [2]. It is possible, as we have demonstrated, that each individual feature extracts useful but different image information. It is through using such features together we can achieve better performance. We have demonstrated this using different variants of LBP, but it can be reasonably concluded that such combined approach can be beneficial when using other features.

REFERENCES

[1] Niraj P Doshi and Gerald Schaefer, "A comprehensive benchmark of local binary pattern algorithms for texture retrieval," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR), 2012*. IEEE, 2012, pp. 2760–2763.

[2] Gerald Schaefer and Niraj P Doshi, "Multi-dimensional local binary pattern descriptors for improved texture analysis," in *21st International Conference on Pattern Recognition (ICPR), 2012*. IEEE, 2012, pp. 2500–2503.

[3] Timo Ojala, Topi Maenpaa, Matti Pietikainen, Jaakko Viertola, Juha Kyllonen, and Sami Huovinen, "Outex-new framework for empirical evaluation of texture analysis algorithms," in *Proceedings of the 16th International Conference onPattern Recognition, 2002*. IEEE, 2002, vol. 1, pp. 701–706.

[4] Timo Ojala, Matti Pietikäinen, and David Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern recognition*, vol. 29, no. 1, pp. 51–59, 1996.

[5] Timo Ojala, Matti Pietikainen, and Topi Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[6] Nabeel Mohammed and David McG. Squire, "Improved texture features for CBIR using response scaling and locally normalised convolution," in *Proceedings of the 11th International Workshop on Content-Based Multimedia Indexing*, Veszprém, Hungary, June 17–19 2013.

[7] Nabeel Mohammed and David McG. Squire, "Efficient and accurate independent component filter-based features for texure similarity," in *Proceedings of the 20th IEEE International Conference on Image Processing (ICIP)*, Melbourne, Australia, September 15–18 2013.

[8] Nabeel Mohammed and David McG. Squire, "Effectiveness of ICF features for collection-specific CBIR," in *Proceedings of the 9th International Workshop on Adaptive Multimedia Retrieval, co-located with the 22nd International Joint Conference on Artificial Intelligence (IJCAI 2011)*, Barcelona, Spain, July 18–19 2011.

[9] David McG. Squire, Wolfgang Müller, Henning Müller, and Jilali Raki, "Content-based query of image databases, inspirations from text retrieval: inverted files, frequency-based weights and relevance feedback," in *Pattern Recognition Letters*, 1999, pp. 143–149.

[10] Harald Kosch and Paul Maier, "Content-based image retrieval systems–reviewing and benchmarking.," *JDIM*, vol. 8, no. 1, pp. 54–64, 2010.

[11] Cong Bai, Wenbin Zou, Kidiyo Kpalma, and Joseph Ronsin, "Efficient colour texture image retrieval by combination of colour and texture features in wavelet domain," *Electronics letters*, vol. 48, no. 23, pp. 1463–1465, 2012.

[12] Lionel Bombrun, Yannick Berthoumieu, N-E Lasmar, and Geert Verdoolaege, "Multivariate texture retrieval using the geodesic distance between elliptically distributed random variables," in *Proceedings of the 18th IEEE International Conference on Image Processing, 2011*. IEEE, 2011, pp. 3637–3640.

[13] Henning Müller, Wolfgang Müller, David McG. Squire, Stéphane Marchand-Maillet, and Thierry Pun, "Performance evaluation in Content-Based Image Retrieval: Overview and Proposals," *Pattern Recognition Letters*, vol. 22, no. 5, pp. 593–601, 2001, (special issue on Image/Video Indexing and Retrieval).

A Novel Adaptive Volterra Filter to Compensate for Speaker Non-linearity

Tonmoy Roy*, Parvez Ahmmed, Md. Shafiur Rahman, Abdullah Ash-Saki and A. B. M. Harun-Ur-Rashid
Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology, Dhaka–1205, Bangladesh
Email: *tonmoy_roy@live.com

***Abstract*—The implementation of an adaptive Volterra filter to compensate for speaker non-linearity using a pipelined recurrent neural network based architecture is demonstrated. The proposed architecture consists of two stages: nonlinear stage performing a nonlinear second order Volterra (SOV) mapping from the input space to an intermediate space and a linear combiner performing a linear mapping from the intermediate space to the output space. The filter design is tested by implementing it on a dataplane processing unit configured for audio processing. The implemented algorithm is adjusted for Xtensa Processor and uses HiFi-2 DSP standard. The collected data confirms smaller error of the output with the speaker and also a very low settling time.**

***Index Terms*—Volterra, JPPSOV, Xtensa, adaptive filter, non-linearity.**

I. Introduction

The Volterra Series filter is one of the most widely used non-linear systems. The ease with which existing linear adaptive filters can be extended to fit this model makes it a popular non-linear filter. The Volterra Series expansion can be interpreted as a Taylor Series expansion with memory [1]. A major disadvantage of the filter is the computational complexity arising when the complete series has to be implemented. One key reason is that the number of the Volterra coefficients increases geometrically as the delays (or memory) and orders increases.

There are a number of methods applied to reduce the computational complexity. These include iterative factorization technique [2], the fast Kalman algorithm [3], multi memory decomposition [4]. With these strategies, the computational complexity can be reduced significantly compared with the SOV filter, but the stable performance is not guaranteed. Again, in Volterra filter using the normalized least mean square (NLMS) the output of the system becomes a nonlinear function of the filter coefficients [5].

However, a novel architecture based on Pipelined Recurrent Neural Network (PRNN) [6], [7] provides a more computationally efficient way to approximate non-linear systems in a modular approach. In this Xtensa processor [8] based implementation, the novel adaptive Joint Process filter using Pipelined Second Order Volterra (JPPSOV) architecture is followed. The attractive attribute of such pipelined architecture is that the capability of such realizations to efficiently approximate nonlinear systems using less computational burdens. Moreover, the combination of the two stages, nonlinear and then linear, makes the output dependent on the filter coefficients linearly.

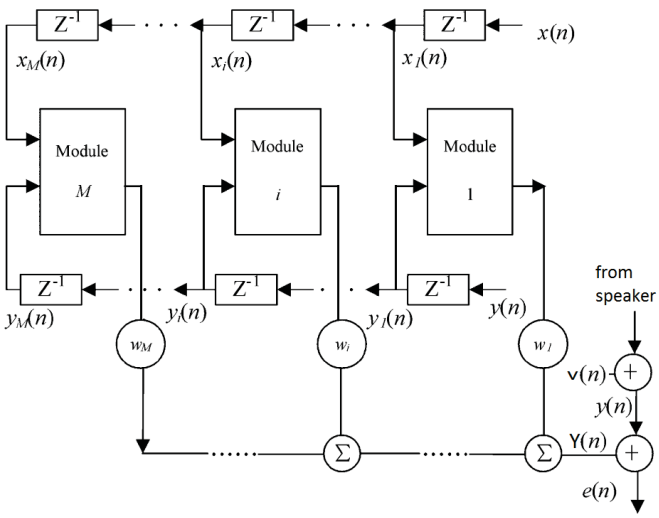


Fig. 1. Block diagram of modified adaptive Joint Process filter using Pipelined Second Order Volterra architecture.

II. Theory

The algorithm effectively splits the system into two subsystems, one is nonlinear and the other is linear. The nonlinear subsystem contains M number of SOV cores (Fig. 1). It performs nonlinear mapping from a input space to an intermediate space which linearizes the nonlinear signals. Each module takes one of the external inputs and the noisy output from the speaker as inputs. The benefit of this arrangement is in the uniformity of numbers of inputs and outputs and the synaptic weight matrix.

For the i-th module, the input can be written as:

$$X_i(n) = [X_{i,1}(n), X_{i,2}(n)] \tag{1}$$

where, the column $X_{i,1}(n)$ is vector of previous external input at the n-th time point and is described by:

$$X_{i,1}(n) = x(n-i)$$

and $X_{i,2}(n)$ contains the noisy output from the speaker,

$$X_{i,2}(n) = y(n-i)$$

At the n-th time point, input signal vector $X_i(n)$ is expanded to a row vector $XX_i(n)$ by the SOV series (2) where the number of columns of the matrix is defined by $L = 6$ which is much less compared to $(N+1)(N+2)/2$ of the ordinary Volterra theory.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

$$\begin{aligned} XX_i(n) &= [XX_{i,1}(n), XX_{i,2}(n), \ldots, XX_{i,L}(n)] \\ &= [1, X_{i,1}(n), X_{i,2}(n), X_{i,1}^2(n), \\ &\quad X_{i,1}(n)X_{i,2}(n), X_{i,2}^2(n)] \end{aligned} \tag{2}$$

Since the weight vectors are same for all the modules, the synaptic weight vectors of the SOV series of each module is denoted by the column vector $H(n)$, whose length is also L.

$$H(n) = [h_1(n), h_2(n), ..., h_L(n)]^T \tag{3}$$

Thus, the output $y_i(n)$ of the i-th modules is defined by:

$$y_i(n) = XX_i(n)H(n) \tag{4}$$

The linear subsystem simply performs a mapping from the intermediate space to the output. The outputs of the nonlinear subsystems are combined in a ratio provided by a weight vector to produce an estimation of the original signal. The weight vector is given by:

$$W(n) = [w_1(n), w_2(n), ..., w_M(n)] \tag{5}$$

If the outputs of the previous non-linear modules are

$$YY(n) = [y_1(n), y_2(n), ..., y_M(n)]^T \tag{6}$$

the outputs of the linear subsystem would simply be the inner products of these two vectors.

$$Y(n) = W(n)YY(n) \tag{7}$$

The error between the desired signal and estimated signal is given by

$$e(n) = y(n) - Y(n) \tag{8}$$

Under the assumption that the coefficient vector $H(n)$ of the nonlinear subsection is time variant, the adaptation of $H(n)$ is similar to the classical case. Therefore, the normalized least mean square (NLMS) equation driving the update of the nonlinear subsection finally becomes

$$H(n+1) = H(n) + \eta_2 e(n)\frac{U(n)}{||U(n)||_2^2} \tag{9}$$

where $U(n)$ is defined by,

$$U(n) = W(n)XX_i(n)$$

Similarly, the NLMS equation driving the update of the linear filter is

$$W(n+1) = W(n) + \eta_1 e(n)\frac{YY(n)}{||YY(n)||_2^2} \tag{10}$$

where $||YY(n)||_2^2$ is the L_2 norm.

Here, η_1 and η_2 are the two learning rates. A necessary condition for the mean convergence is [9],

$$0 < \eta_1 < 2 \qquad \text{and} \qquad 0 < \eta_2 < 2$$

The values of the learning rates are very crucial for steady response of the filter. In this work, $\eta_1 = \eta_1 = 1/8$ is used.

TABLE I
THE VALUE OF THE NON-LINEARITY COEFFICIENTS FOR SIMULATING DIFFERENT SPEAKERS

Co-efficient	b	c
Speaker-1	-0.1	0.1
Speaker-2	0.05	0.05

III. OPTIMIZATION TECHNIQUES

As illustrated in [10], the filter with only 4-5 modules ensures lowest mean square error (MSE). As the Xtensa processor gives an opportunity of two parallel multiplication operation, 4 modules are used in this implementation. Moreover, the inputs of each module is also modified in order to decrease the no. of memory operations.

According to (9), $U(n)$ has to be divided by $||U(n)||_2^2$ and then η_1 is multiplied to obtain the change of the vector H. The filter converges as long as $0 < \eta_1 < 2$. The performance of the filter does not change much within this range of η_1. If η_1 is kept constant on the other hand, the division result can be allowed to vary to mimic the change of η_1. Since doing complete division using any DSP processor like Xtensa is inefficient, the division is replaced with an efficient bit shift operation. This bit shift can be modeled by,

$$q = \left\lfloor \frac{2^{\lfloor \log_2 a \rfloor}}{b} \right\rfloor \tag{11}$$

where a is the dividend, b is the divisor and q is the quotient. It can be observed that the quotient can be off by only one fourth of its intended value. Since $\eta_1 = 1/8$, the division can be replaced with the bit shift operation, effectively making $\eta_1 = 1/32$ in the worst case scenario.

It can be seen from (10) that, the computation of the change of the vector W can be done in a similar manner to improve efficiency.

IV. TEST STRATEGY

The Xtensa instruction set simulator (ISS) has been used to test the performance of the algorithm. From the tests run in the simulator, various performance characteristics like the time required to process each sample and the residual error are obtained.

The audio for the speaker input and the audio output obtained from the microphone both are required during run-time for the filter. Both the audios are assumed to be mono channel with 16 kHz sampling rate and each sample was assumed to consist of 16 bits. The errors in output are obtained as 16-bit values as well.

For the actual testing of the adaptive filter algorithm, the microphone input signal is modeled from the speaker input signal. Furthermore the speaker input is considered to be non-linearly distorted by,

$$z = \frac{x + bx^2 + cx^3}{1 + |b| + |c|} \tag{12}$$

where the non-linear terms are assumed to be less than or equal 10% of the linear term. A random noise has also been added to the microphone input. The noise is within 1% of the feedback signal at most.

Using the speaker input and the modeled microphone input, the filter is tested for various different scenarios. Three different audio samples each 1 second in length are used to test the

TABLE II
THE REQUIRED CYCLES & MINIMUM PROCESSOR CLOCK SPEED FOR THE DIFFERENT SCENARIOS

Audio type	Speaker	Total cycles	Cycles per sample	Minimum clock speed (MHz)
Human voice	Speaker-1	17238241	1077	17.24
	Speaker-2	17238871	1077	17.24
Instrumental music	Speaker-1	17564397	1098	17.56
	Speaker-2	17642333	1103	17.64
Periodic signal	Speaker-1	17995837	1125	18.00
	Speaker-2	18049309	1128	18.05

TABLE III
THE SETTLING TIME AND THE RESIDUAL NOISE AT DIFFERENT SCENARIOS

Audio type	Speaker	Residual Noise (dB)	Settling Time (ms)
Human voice	Speaker-1	-41.0048	8.75
	Speaker-2	-39.0622	7.50
Instrumental music	Speaker-1	-28.7217	12.50
	Speaker-2	-27.9533	9.38
Periodic signal	Speaker-1	-45.1022	32.19
	Speaker-2	-44.0620	25.00

filter. These audios are: human voice, instrumental music and simple triple frequency audio. All of the audios are of 1 second in duration. Different values of the non-linearity coefficients in the microphone input model are tested to simulate different speaker types. The used coefficients, b and c, are summarized in Table I.

The processor has been configured in the built-in 'HiFi-2 standard' mode [11] for the particular implementation. In this configuration, the processor can handle audio signal processing like multiplication and addition of 24-bit numbers very efficiently. Hence for this implementation, the 16-bit integers from the input are shifted left in order to convert to 24-bit. Processor's parallel multiplication instruction is utilized whenever possible to reduce the number of cycles required per audio sample.

V. RESULTS AND DISCUSSION

The number of instructions required by the processor to operate on the audio samples of 1 second and minimum clock frequency required to do the processing in real-time is given in Table II. Since the processor can operate at frequencies well above the obtained minimum required clock frequency, it will be possible to use the processor for real-time operation with ease. As illustrated in Fig. 2, use of only four modules shows the lowest Mean Square Error (MSE).

The average error in decibels was calculated with respect to the input signal. As the settling time is negligible, the residual noise was calculated as the average of all the error values in decibel (dB). The time after which the error becomes less or equal to the mean error was considered to be the settling time. The mean local error vs. sample plots for the human voice sample, music sample and the simple periodic audio signals through different speakers are shown in Fig. 3 and Fig. 4 respectively. Again, the settling time and the residual noise is presented for the six scenarios in Table III.

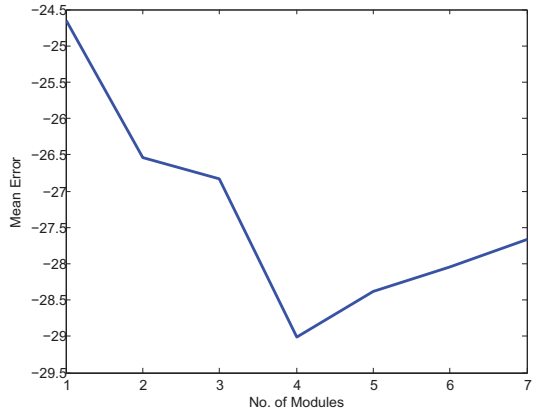


Fig. 2. relation between the MSE and the number of modules for speaker-1.

VI. CONCLUSION

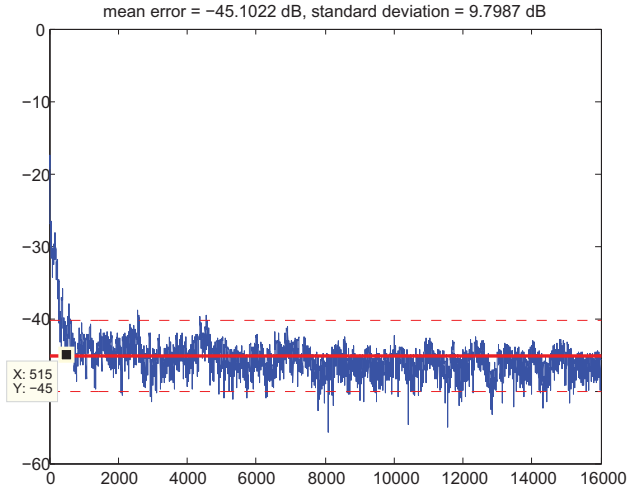
An implementation of a novel adaptive joint process filter employing the pipelined architecture is presented in this paper. It is seen that the algorithm can easily be implemented using the Tensilica Xtensa dataplane processing unit. The obtained residual error was very low and the settling time was very small. Furthermore, size of the code and the required processor clock speed was within desired levels for the implementation indicating the efficiency of the algorithm.

ACKNOWLEDGMENT

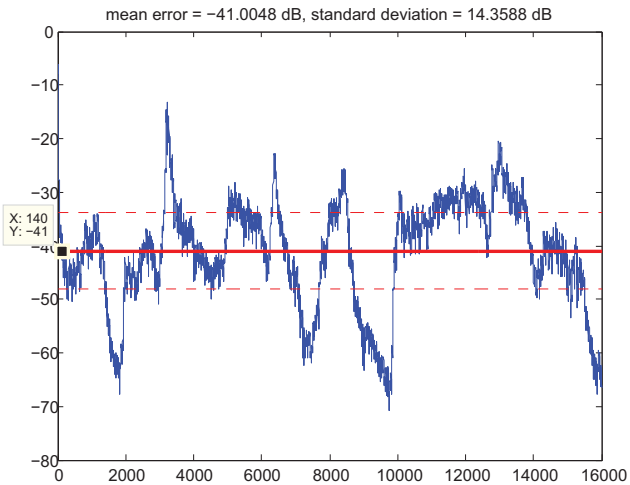
The authors would like to thank Cadence Design Systems, India for providing the software, license and samples required for the work.

REFERENCES

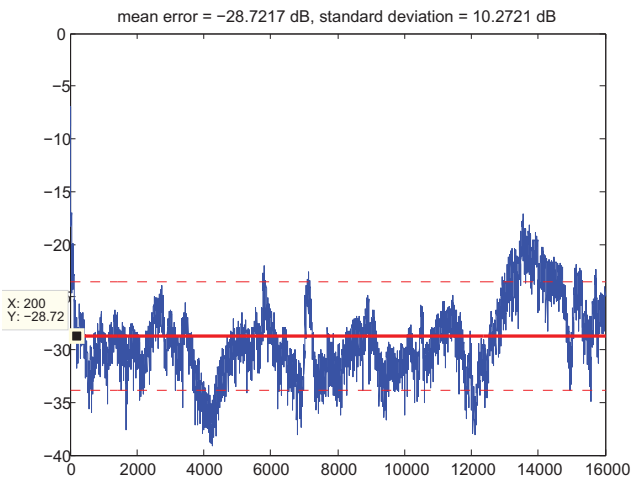
[1] V. Mathews, "Adaptive polynomial filters," *IEEE Signal Processing Magazine*, vol. 8, no. 3, pp. 10–26, July 1991.
[2] T. Koh and E. Powers, "Second-order volterra filtering and its application to nonlinear system identification," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 33, no. 6, pp. 1445–1455, Dec 1985.
[3] C. Davila, A. Welch, and I. Rylander, H., "A second-order adaptive volterra filter with rapid convergence," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 35, no. 9, pp. 1259–1263, Sep 1987.
[4] Y. Lou, C. Nikias, and A. Venetsanopoulos, "Efficient VLSI array processing structures for adaptive quadratic digital filters," *Circuits, Systems and Signal Processing*, vol. 7, no. 2, pp. 253–273, 1988. [Online]. Available: http://dx.doi.org/10.1007/BF01602100
[5] T. Harada, M. Muneyasu, and T. Hinamoto, "A pipeline architecture of quadratic adaptive volterra filters based on NLMS algorithm," in *The IEEE International Symposium on Circuits and Systems, ISCAS*, vol. 2, May 2001, pp. 785–788.
[6] S. Haykin and L. Li, "Nonlinear adaptive prediction of nonstationary signals," *IEEE Transactions on Signal Processing*, vol. 43, no. 2, pp. 526–535, Feb 1995.
[7] L. Li and S. Haykin, "A cascaded recurrent neural network for real-time nonlinear adaptive filtering," in *IEEE International Conference on Neural Networks*, vol. 2, 1993, pp. 857–862.
[8] (2014) Xtensa customizable processors - Cadence IP. [Online]. Available: http://ip.cadence.com/ipportfolio/tensilica-ip/xtensa-customizable
[9] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice Hall, 2002.
[10] H. Zhao and J. Zhang, "A novel adaptive nonlinear filter-based pipelined feedforward second-order volterra architecture," *Signal Processing, IEEE Transactions on*, vol. 57, no. 1, pp. 237–246, Jan 2009.
[11] (2014) Tensilica HiFi audio voice DSP IP. [Online]. Available: http://ip.cadence.com/ipportfolio/tensilica-ip/audio



(a) periodic signal

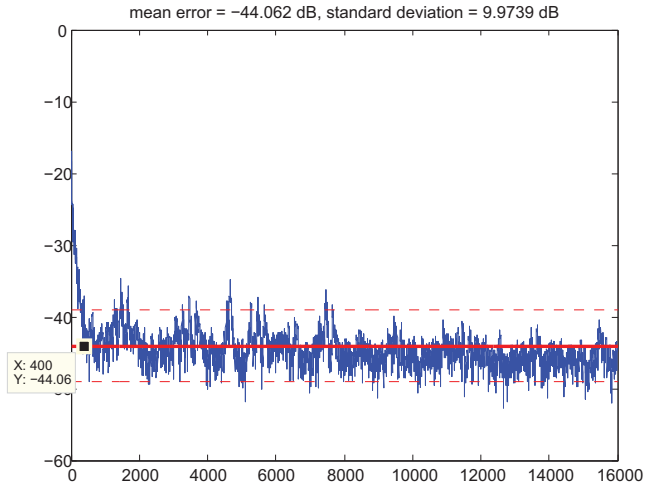


(b) human voice

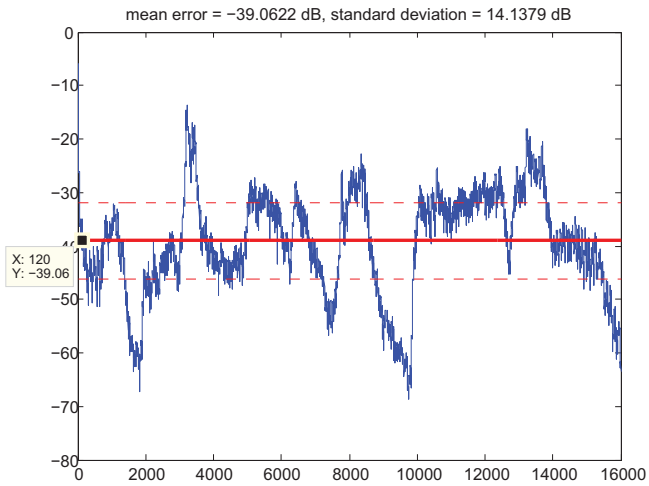


(c) instrumental music

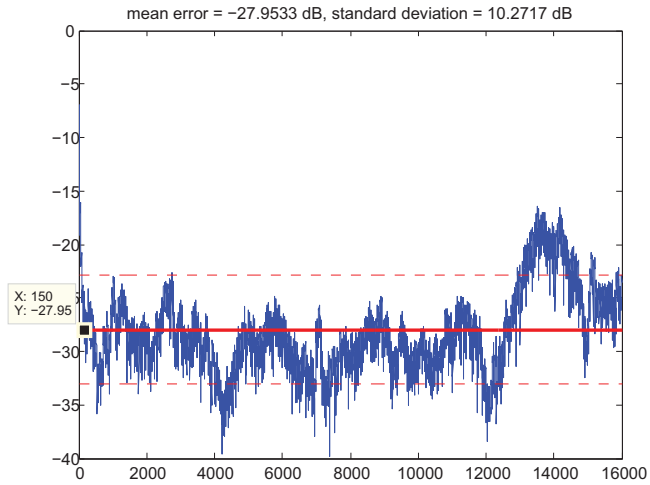
Fig. 3. Error in speaker-1 where $b = -0.1$ and $c = 0.1$



(a) periodic signal



(b) human voice



(c) instrumental music

Fig. 4. Error in speaker-2 where $b = 0.05$ and $c = 0.05$

Stereo facial image clustering using double spectral analysis

Georgios Orfanidis,[1,*] Nikos Nikolaidis, [1], Anastasios Tefas, [1] and Ioannis Pitas [1]
[1]Department of Informatics
Aristotle University of Thessaloniki, GREECE
*eypros@aiia.csd.auth.gr

***Abstract*—In this work we proposed a new variant of spectral clustering using double spectral analysis which proved to be able to achieve better clustering results. The present work focuses on the special case of 3D videos and the implication of their use. Various improvements are introduced including the use of stereo over mono videos, the use of double spectral clustering over spectral clustering and the use of multiple representative images per trajectory for more robust trajectory representation. Extended experiments have been conducted in three 3D full feature commercial films which revealed the power of stereo face clustering in comparison with single channel face clustering.**

***Index Terms*—3D, stereo facial image clustering, Spectral clustering, Ncut, HSV color space, facial image clustering**

I. Introduction

Clustering is one of the fundamental topics in computer science. It has been studied at a great degree but even at present day it is not considered solved. Clustering deals with dividing an existing set of objects $\mathcal{P}$ (in our case, images) into a number of subsets (clusters) $\mathcal{C} = \{C_i | C_i \subseteq \mathcal{P}\}$. Those subsets have to fulfill the following conditions: $\bigcup_{C_i \in \mathcal{C}} C_i = P$ and $\forall C_i, C_j, i \neq j \in \mathcal{C} : C_i \bigcap C_j = \oslash$. Essentially clustering assigns each sample of a data set to a cluster, although there are exceptions to this rule like in the fuzzy c-means algorithm.

This paper deals with facial image clustering applied on 3D commercial feature films, where the goal is to separate face images into groups, for which within-cluster similarity is high whereas between-cluster similarity is smaller. There is some work with single channel feature films like [1] but similar work on 3D films is quite limited. Also, in [1] these films there is limited variance in optical effect, illumination, or special features of the images in the used test films. The film used in [1] has just 1 shots with consistent clothing, the film focuses on character development rather than visual effect etc. There are also various works with facial image clustering not deriving from videos like [2], [3] but focusing on still images. in this case the requirements are different. On the other hand, even in mono videos, often the dataset derives from much shorter videos as in [4] or is confined as in [5]. Spectral clustering has also been used in previous works [6],[7].

The rest of this work is organized as follows: Section 2 contains a short presentation of Mutual Information and its normalized version used in this work as similarity measure among images. Section 3 provides the problem statement and prior work on the field while Section 4 introduces the double spectral clustering and the improvements proposed. Section 5 contains the experimental results and finally Section 6 concludes the paper.

II. Mutual Information

Mutual Information (MI) is used in this work as the similarity measure between facial images, thus we will briefly present some related important definitions. the interested reader may refer to [6] for details on MI.

Mutual Information is defined as the common information of two distributions. Entropy of a random variable X is defined as:

$$H(X) \triangleq -\sum (p(x) log(p(x))) \tag{1}$$

while joint entropy of two random variables X, Y is defined as:

$$H(X,Y) \triangleq -\sum (p(x,y) log(p(x,y))) \tag{2}$$

where $p(x)$ is the probability density function and $p(x,y)$ is the joint probability density function of random variables X and Y respectively. The normalized mutual information used is:

$$D(X,Y) = \frac{H(X) + H(Y)}{2H(X,Y)} \tag{3}$$

as in [8]. In this work the HSV color space of images was used for calculating similarity as it has been proved to be more robust in illumination changes compared to RGB color space. More specifically we will follow the approach of [6] where only Hue (H) and Saturation (S) are used. According to this approach, a 4D normalized MI is used:

$$D(X,Y) = \frac{H(H_1) + H(S_1) + H(H_2) + H(S_2)}{2H(H_1, S_1, H_2, S_2)} \tag{4}$$

where H_i is the Hue and S_i is the Saturation of each image respectively. For the 4D joint histogram an approach similar to [6] was used.

III. Problem statement and prior work

In order to use spectral clustering techniques we define our face clustering problem as an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ cut problem. We consider each facial image to be a graph vertex and then we construct a similarity matrix $\mathbf{W}$ in which each element w_{ij} represents the edge that connects the two images (vertices), which has a weight representing the similarity of the corresponding image pair and has an assigned value of $1/d_{ij}$.

We can then use different spectral approaches to solve the problem. Crucial to all approaches is the definition of the Laplacian matrix $\mathbf{L}$. The normalized Laplacian is defined as:

$$\mathbf{L} = \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2} \tag{5}$$

where $\mathbf{D}$ is the Degree matrix. The normalized Laplacian $\mathbf{L}$ has been used in our case since it has been proven to have greater merits than the unnormalized one [9].

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

Having defined the Laplacian matrix $\mathbf{L}$ different approaches for clustering can be used. One possible solution could be the use of Normalized Cuts (Ncut) [10] described in subsection III-A while another solution could be the approach introduced in [11] and summarized in subsection III-B below.

A. Normalized cut

Ncut is defined primarily for bipartition, namely for splitting the graph into two parts, and it attempts to find the cut that minimizes the edge weights (connections) between the two clusters while simultaneously maximizing the edge weights within the two clusters.

The problem the Ncut algorithm tries to solve is defined as $\text{argmin}_{\mathbf{y}} \frac{\mathbf{y}^T(\mathbf{D}-\mathbf{W})\mathbf{y}}{\mathbf{y}^T\mathbf{D}\mathbf{y}}$ subject to $\mathbf{y}^T\mathbf{D}\mathbf{1} = 0$ which uses the matrices defined in section III. The Ncut problem is usually solved by relaxing the solution using eigen-analysis [10].It is proven in [10] that the eigenvector corresponding to the second smallest eigenvalue is the optimal graph cut for the given criterion. More eigenvectors or an iterative eigen-analysis can be applied in order to derive more than two clusters.

B. Spectral clustering

While Ncut solves a binary problem, the spectral clustering approach proposed in [11] solves a multiclass clustering problem. This solution also uses the Laplacian matrix $\mathbf{L}$ defined in (5). The solution applies eigen-analysis in $\mathbf{L}$ and uses the first k eigenvectors $\mathbf{u}_2, ..., \mathbf{u}_{k+1}$ excluding the first one. A new matrix containing those eigenvectors is defined as $\mathbf{U} \in \mathbb{R}^{N\times k}$ (N being the number of images). New samples are defined as the rows of this matrix after being normalized to norm 1. The final step is to use k-means [12] to these new samples.

IV. Introducing double spectral analysis

The two approaches briefly mentioned in sections III-A and III-B although have proved quite efficient in various cases, have some drawbacks. In this Section, we list these drawbacks and also propose ways to overcome them while preserving their merits.

At first, Ncut is designed to solved two-class problems which can have some implications when applied to multi-class problems. One simple example is demonstrated in figure 1(a). In this example a simple similarity matrix $\mathbf{W}_{sample}$ is used, created by inverting the Euclidean distance of each pair of points $\mathbf{x}_{ij}$.

Assume that we cluster the samples using the Ncut algorithm and matrix $\mathbf{W}_{sample}$. We plot the elements of the second smallest eigenvector in Figure 1(b) which is the optimal solution for the Ncut problem according to [10] and also choose to use the simplest threshold, namely zero. While there are 4 easily distinguishable classes, shown in Figure 1(a), Ncut fails to separate correctly the classes as it is obvious in Figure 1(b). More specifically, class #1 is split in two by using this threshold. By observing the elements of the eigenvector corresponding to each sample we see that all elements corresponding to samples of the same class obtain similar values. That observation led us to use a more flexible threshold, the greatest gap threshold (GG threshold) as explained in section IV-B.

The main drawback of the spectral clustering approach [11] is the use of k-means as the main clustering algorithm. It is well known that k-means has various limitations; such as the assumption of spherical clusters, the effect of initialization on its performance and the possible convergence to a local minimum. Thus, we opted for a more efficient clustering algorithm. Since Ncut does not assume the above we opt to use it instead of k-means, as described in the next Section.

A. Double spectral clustering

The proposed approach is similar to the one in [11]. Assume that we have a dataset with N facial images. We then calculate the similarity matrix $\mathbf{W} \in \mathbb{R}^{N\times N}$ as described in Section II, III and the normalized Laplacian matrix $\mathbf{L} \in \mathbb{R}^{N\times N}$ in (5).

We apply eigen-analysis to $\mathbf{L}$ and create a new matrix $\mathbf{U}$ containing the k first eigenvectors of $\mathbf{L}$ as columns. The eigenvectors are ordered using their corresponding eigenvalue in ascending order. Furthermore, the first eigenvector corresponding to the trivial solution is not taken into consideration. Thus, $\mathbf{U} = \{\mathbf{u}_2, ..., \mathbf{u}_{k+1}\}$.

We create new samples $\mathbf{Y} = \{\mathbf{y}_1, ..., \mathbf{y}_N\}$ that correspond to the rows of matrix $\mathbf{U}$. Those new samples have dimension k which equals the number of eigenvectors being used. It must be noted that we do not normalize the samples $\mathbf{y}_i$ as in [11], in order to keep as much discriminality as possible. Normalization using L_2 norm would mean projecting all samples to the unitary circle which obviously reduces the distance among samples.

The general idea is to create new samples that benefit from the eigen-analysis of the Laplacian matrix $\mathbf{L}$ which tends to gather together samples with high similarity score in $\mathbf{W}$ and to distanciate samples with low similarity score in $\mathbf{W}$.

From this point we create a new similarity matrix $\mathbf{W}'$ considering as initial data set the $\mathbf{Y} \in \mathbb{R}^{N\times k}$ and using the inverse of the Euclidean distance of similarity score:

$$[\mathbf{W}']_{i,j} = \frac{1}{(\mathbf{y}_i - \mathbf{y}_j)^T(\mathbf{y}_i - \mathbf{y}_j)} \tag{6}$$

Then, we proceed by creating the correspoding Laplacian matrix $\mathbf{L}'_{sim}$ and apply the Ncut algorithm in order to split the dataset into 2 subclusters. In order to get the desired number of clusters we recursively apply the Ncut algorithm to those subclusters until this number has been reached.

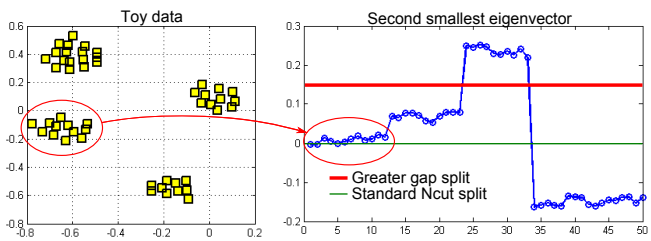
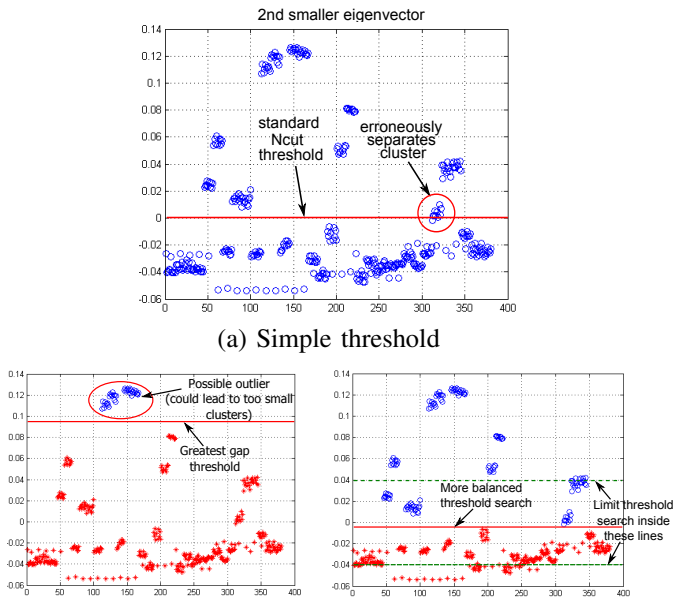


Fig. 1. Ncut unsuccessful clustering in toy example

B. Greater gap threshold

As it has been previously mentioned and also shown in Figure 1 the simple thresholding approach that uses a threshold equal to zero on the eigenvector elements often fails when used with multi-class problems. For this reason we propose a new approach which uses a more flexible threshold that has been proven to be more reliable. Figure 2 plots the elements of the eigenvector of a real problem that corresponds to the second smallest eigenvalue and their relative position with



(a) Simple threshold

(b) Overall greatest gap threshold

(c) Limited greatest gap threshold

Fig. 2. The effect of using different thresholds

respect to the zero axis. One possible problem with a fixed zero threshold has already been mentioned and concerns the separation of clusters which get values around zero, as the one shown in Figure 2(a) inside the red circle. Those clusters could have samples which correspond to very similar elements of the eigenvectors (and thus they should not be split) but the a fixed zero threshold does not take into account this similarity.

Thus, instead of using a fixed threshold, we first sort the elements of the second eigenvector and then search for the greatest difference between two consecutive elements. Then the threshold is set as the mean value of these two consecutive elements having the greatest gap. Thus we get a different threshold in each application of the Ncut algorithm.

One problem with this approach is that it tends to first separate the outliers from the other samples (Figure 2(b)) which may not be an optimal approach. For this reason we restrict the threshold search inside some limits as shown in Figure 2(c). The restriction limits are usually defined with respect to the greatest value v occurring in each eigenvector, that is as a proportion of this greatest value v: $Limits : \pm av,\ a \in (0, 1]$.

V. Experiments

A. Mono and stereo dataset creation

We applied our proposed method in three full length 3D feature films of different duration, size of cast and genre. For the creation of the facial image dataset, face detection and tracking was applied. Two different approaches were tested, in an effort to highlight the advantage of using stereo data. In the first approach, the face detection [13] is performed every n frames in one channel (video) of the stereoscopic video, followed by a face tracker [14], [15] at the same channel. In the second approach, face detection [13] was applied on both channels, mismatches between the two channels were rejected and a stereo tracking algorithm [15] was applied in both channels. By using the above approaches we end up with a number of facial trajectories, namely series of consecutive facial images.

As can be observed in Table I the trajectories generated by the stereo face detectors and trackers are fewer in number and longer (in number of frames). This has the advantage of fewer facial images to be clustered and better representation of the actor appearances. Another observation that can be made in Table I is that the number of clusters (that equals the cast size) is quite high. The cardinality of actor clusters varies significantly with some actors having many facial images while others have very few appearances.

Usually, each tracking trajectory is represented by one (usually detected) facial image. Another approach for dataset creation examined was the inclusion of multiple facial images per trajectory. We used all detected images as well as each middle image between two sequential detections belonging to the same trajectory. For 2 trajectories represented by 2 sequences of k and m images respectively, T_k and T_m we calculate the similarity between them as:

$$\text{Similarity Score} = \max_{ij} MI(x_i, x_j) \quad x_i \in T_k, x_j \in T_m$$

where x_i is an image belonging to the sequence T_k. The extreme values for the size of T_i is 1 (as the usual case) to the whole set of images belonging to each trajectory. By this way a robustness in variation in pose, illumination etc is applied. It can be seen that with multiple images the number of images is increased but the similarity matrix used in clustering is of the same size as before (having the size of the number of trajectories) because only one score per trajectory is finally used. Experiments have been conducted using this new approach and revealed an increase in performance.

B. Experimental results

Experiments were conducted in three 3D feature films with varying duration, size of cast, special effect etc. Some characteristics of each film are shown in Table I where it can be seen that stereo dataset are fewer in number and bigger (in number of frames). This has the advantage of fewer facial images to be clustered and also larger trajectories which signifies better representation of the films. Experimental results confirm this also in the performance. In Table I it is obvious that the number of clusters (that equals the cast size) is quite high while another fact is the unbalanced classes; with some persons having many images and also some persons having very few appearances.

Results have been accumulated in Table II where it can be observed that there is a significant improvement over Ncut by using both double spectral clustering and double spectral clustering with constrained greatest gap. Of the two latter methods best results were achieved using the double spectral clustering in combination with greatest gap in all 3 films. The use of stereo video also improves the performance and provides an extra way to further improve the results. There is a significant advantage in using both channels in comparison with a single channel.

For the evaluation of the clustering a variation of F-measure was used. Since the number of clusters is quite high the standard F-measure tends to punish the splitting of classes into more clusters. In our case, and taking into consideration the complexity of the facial image clustering task, a different clustering measure was used. This new measure evaluates more

TABLE I
VARIOUS CHARACTERISTICS OF THE THREE FEATURE FILMS

Size of cast	# of frames	# of trajectories	
		in stereo	in single channel
Feature film #1			
27	181,763	1246	1545
Feature film #2			
47	150,361	1222	1559
Feature film #3			
58	196,224	1726	2238

TABLE II
RESULTS ON THE 3 FEATURE FILMS FOR VARIANTS OF NCUT AND IMAGE/TRAJECTORY SIMILARITY

Feature film #1				
	Mono Left		Stereo Left	
	# of images		# of images	
	1	k	1	k
Ncut	0.3030	0.3984	0.3690	0.4096
Double spectral	0.4341	0.5234	0.5073	0.5234
GG Double spectral	0.4658	0.5419	0.5051	**0.5446**
Feature film #2				
	Mono Left		Stereo Left	
	# of images		# of images	
	1	k	1	k
Ncut	0.2272	0.3274	0.2235	0.3246
Double spectral	0.2989	0.3574	0.3379	0.3785
GG Double spectral	0.3804	0.4576	0.4111	**0.4626**
Feature film #3				
	Mono Left		Stereo Left	
	# of images		# of images	
	1	k	1	k
Ncut	0.1554	0.2183	0.1600	0.2231
Double spectral	0.2986	0.3490	0.3409	0.3305
GG Double spectral	0.3179	0.3607	0.3349	**0.3710**

the purity of clusters, which is one of the desired criteria, while on the same time validates the overall clustering as well. The new approach included oversplitting the set to more than the clusters needed and then merged them in a way a user would merge them; that is by merging all clusters having the majority of their elements derive from the same class.

VI. CONCLUSION

In this work we proposed a new variant of spectral clustering using double spectral analysis which proved to be able to achieve better clustering results. We also focus mainly on 3D films and the special issues regarding their use, like the possible improvement by the use of stereo video compared to mono ones, but also in other aspects of films like the use of multiple images per trajectory for further improving the results. Finally a more flexible threshold was introduced that also proved more efficient. Further work will include more extensive research on the use of 3D films for clustering purposes.

REFERENCES

[1] Vineet Gandhi and Remi Ronfard, "Detecting and naming actors in movies using generative appearance models.," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013.

[2] WT Chu, YL Lee, and JY Yu, "Visual language model for face clustering in consumer photos.," *Proceedings of the 17th ACM international conference on Multimedia*, 2009.

[3] Edoardo Ardizzone, Marco La Cascia, and Filippo Vella, "Mean shift clustering for personal photo album organization.," *IEEE International Conference on Image Processing*, 2008.

[4] S. Schwab, T. Chateau, C. Blanc, and L. Trassoudaine, "A multi-cue spatio-temporal framework for automatic frontal face clustering in video sequences.," *EURASIP Journal on Image and Video Processing*, vol. 1, pp. 1–12, 2013.

[5] Chen Guangliang and Gilad Lerman, "Spectral curvature clustering (scc).," *International Journal of Computer Vision*, vol. 81, pp. 317–330, 2009.

[6] N. Vretos, V. Solachildis, and I. Pitas, "A mutual information based face clustering algorithm for movie content analysis," *Image and Vision Computing*, vol. 29, pp. 693–705, 2011.

[7] Foucher Samuel and Langis Gagnon, "Automatic detection and clustering of actor faces based on spectral clustering techniques," *Fourth Canadian Conference on Computer and Robot Vision*, 2007.

[8] J. Pluim, J. Maintz, and M. Viergever, "Image registration by maximization of combined mutual information and gradient information," *IEEE Transactions on Medical Imaging*, vol. 19, pp. 809–814, August 2000.

[9] Ulrike Von Luxburg, "A tutorial on spectral clustering.," *Statistics and computing*, vol. 17, pp. 395–416, 2007.

[10] Shi Jianbo and Jitendra Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 888–905, August 2000.

[11] A. Ng, M. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm.," *Advances in neural information processing systems*, vol. 2, pp. 849–856, 2002.

[12] Stuart Lloyd, "Least squares quantization in pcm.," *IEEE Transactions on Information Theory*, vol. 28, pp. 129–137, 1982.

[13] GN Stamou, M Krinidis, N Nikolaidis, and I Pitas, "A monocular system for automatic face detection and tracking," *Visual Communications and Image Processing (VCIP)*, vol. 5960, pp. 794–802, 2005.

[14] Haris Baltzakis, Antonis Argyros, Manolis Lourakis, and Panos Trahanias, "Tracking of human hands and faces through probabilistic fusion of multiple visual cues," *Computer Vision Systems*, pp. 33–42, 2008.

[15] O. Zoidi, N. Nikolaidis, and I. Pitas, "Appearance based object tracking in stereo sequences.," *In 38th International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2013.

Video Characterization based on Activity Clustering

Nikolaos Kourous, Alexandros Iosifidis, Anastasios Tefas, Nikos Nikolaidis and Ioannis Pitas
Department of Informatics, Aristotle University of Thessaloniki, Greece
Email: {tefas,nikolaid,pitas}@aiia.csd.auth.gr

***Abstract*—In this paper, we propose a method for video characterization based on activity description information. We employ a state-of-the-art video representation in order to learn human activity concepts, i.e., video groups formed by videos depicting similar human activities. In order to exploit the enriched visual information that is available in multi-view settings, we propose the use of the circular shift invariance property of the coefficients of the Discrete Fourier Transform (DFT) that leads to a view-independent multi-view action representation. In the test phase, in order to assign a test video to one (or multiple) activity groups, we perform temporal video segmentation in order to determine shorter videos depicting simple actions. Experimental results on 2 multi-view action databases denote the effectiveness of the proposed approach.**

***Index Terms*—Video characterization, Activity clustering, Multi-camera setup, Temporal video segmentation.**

I. Introduction

Understanding human activities in video sequences is a challenging problem due to the large variability in temporal scale and the periodicity of human actions, the complexity of articulated motion, the prevalence of complex backgrounds, variations in observation angles, etc. It has been primarily approached by applying action/activity recognition techniques that are able to exploit visual information coming from one [1] or multiple [2], [3] cameras. The adoption of multiple cameras gives the advantage of exploiting information related to scene geometry and, thus, leads to higher classification performance in general. However, this performance gain is accompanied by a higher computational cost, since in multi-view methods multiple video streams should be processed and analyzed [2], [3].

One of the disadvantages of the above-described approach is that in an action recognition setting, the set of all possible actions should be a priori defined. In addition, an adequate number of (labelled) videos should be employed in order to train classifiers that will be subsequently used for action video characterization [4]. Given the complexity of such a task due to the aforementioned reasons, the cardinality of the training set required in order to achieve satisfactory performance in an unrestricted application scenario would be enormous. This in turn generates several problems, the most important being the underlying financial and computational costs of such an approach.

In several application scenarios, e.g., in movie production/post-production and content-based video retrieval, the objective is the determination of similar action/activity patterns, rather than to perform a strict characterization (i.e., recognition) of the performed actions. In this way, several treatments, e.g. filters, can be applied to similar video segments. This can, obviously, been addressed by applying action recognition techniques and determine video groups depicting the same action/activity. In addition, in an attempt to address the above-described issues relating to human action recognition, an alternative approach can be exploited that involves unsupervised learning. Unsupervised learning of human actions in this sense is a relatively new research topic [4] and it is expected to receive considerable research interest in the next years.

In this paper, we propose a method for video characterization that exploits activity description. We employ the Dense Trajectory-based video description [5], that has been shown to provide state-of-the-art performance on a related task, i.e. action recognition, and combine it with the Bag-of-Words (BoW) model [6] in order to represent videos depicting actions. Furthermore, it does not suffer from background subtraction problems [7], as is the case with silhouette-based action recognition methods [8]. Finally, there is no need to track particular body parts, e.g., arms, feet [9] for action recognition. In an offline process, we determine K action video groups by clustering training action videos depicting simple actions. During the test phase, a Dense Trajectory-based video representation is obtained in order to represent a new (unknown) video, which is subsequently assigned to the label of the group formed by the most similar to it videos. Since test videos are not guaranteed to depict simple actions, we also propose a method that exploits the same video description and is able to perform automatic temporal segmentation of a test video to shorter ones, each depicting a simpler action. Finally, we extend both methods in order to be able to exploit multi-view information that is available in the cases where actions are observed by using multi-camera setups. Experimental results denote that the adoption of multi-view information provides considerable performance gains, when compared to the single-view case, confirming the findings that have come from other relating studies [8], [10].

The rest of the paper is structured as follows. Section II describes the proposed methods. Experimental results conducted in order to evaluate its performance are described in Section III. Finally, conclusions are drawn in Section IV.

II. Proposed Method

Let us denote by $\mathcal{U}$ a video database containing videos depicting N_I videos, e.g., a walking instance. In the case where the action instances have been recorded by using one camera, these action instances are depicted in N_I action instances, while in the case where a multi-camera setup formed by N_C cameras has been used, each action instance is depicted in N_C videos and $\mathcal{U}$ is formed by $N_T = N_I N_C$ videos. We would like to determine K video groups $\mathcal{U}_k$, $k = 1, \ldots, K$, where $\bigcup_{k=1}^{K} \mathcal{U}_k = \mathcal{U}$, each of which will (ideally) be formed by videos depicting the same activity type.

We employ the Dense Trajectory-based video representation [5] in order to represent each video by five vectors, describing two video properties: motion and shape. The shape of various subjects appearing in the scene is described by the Histogram

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

of Oriented Gradient (HOG) descriptor, while motion is described by the remaining four descriptors, i.e., Histogram of Optical Flow (HOF), two channels of Motion Boundary Histogram (MBHx and MBHy) and the (normalized) interest point trajectory coordinates. Even though the four latter descriptors describe the depicted motion, each of them describes a different property of motion. This is important in order to distinguish different activity types [5], [11]. In addition, the adopted video description, by describing local properties of videos, is robust to occlusions [5]. By adopting such a video description, each video in $\mathcal{U}$ is represented by five vectors $x_i^v \in R^{D_v}, v = 1, \ldots, V$ $(V = 5)$, D_v being the dimensionality of each descriptor, obtained by using the BoW model on each descriptor type.

In order to determine the K video groups in $\mathcal{U}$, we apply clustering on the previously determined video representations $\mathbf{x}_i^v, i = 1, \ldots, N_I$. These will serve as train samples. Since BoW-based video representations are better combined with kernel methods, we employ kernel K-Means algorithm [12] to this end. In order to combine the different activity properties described in different BoW-based representations $\mathbf{x}_i^v$, we employ the RBF-χ^2 kernel, where different descriptor types are combined in a multi-channel approach [13]:

$$[\mathbf{K}]_{ij} = \exp\left(-\frac{1}{v}\sum_v \frac{1}{A^v} D(\mathbf{x}_i^v, \mathbf{x}_j^v)\right), \quad (1)$$

where $D(\mathbf{x}_i^v, \mathbf{x}_j^v)$ is the χ^2 distance between the BoW-based representation of videos i and j with respect to the v-th channel. A^v is set equal to the mean value of χ^2 distances between $\mathbf{x}_i^v$ for the v-th channel. After the determination of the cluster labels for the vectors $\mathbf{x}_i^v$, each video i is assigned to the corresponding cluster label.

In the test phase, when a new video depicting an unknown activity appears, we want to assign it to a video group formed by videos depicting the same activity. Thus,we want to characterize its content in terms of the performed activity. To this end, we employ the Dense Trajectory-based video representation in order to represent it by using five vectors $\mathbf{x}_t^v$. Subsequently, we assign the video to the group that provides the minimal χ^2 distance from the corresponding cluster center (in the feature space).

In the case where action instances are depicted in multiple (N_C) videos, each captured by a different viewpoint, we can exploit the available multi-view information in order to obtain a view-independent action representation. This can be achieved by exploiting the circular shift invariance property of the coefficients of the Discrete Fourier Transform (DFT) [10]. In order to avoid confusion, let us denote by $\mathbf{x}_{i,c}^v$, $i = 1, \ldots, N_I$, $c = 1, \ldots, N_C$, $v = 1, \ldots, V$ the N_C vectors representing action video i by exploiting the v-th descriptor type. By using the camera labeling information that is available for both the training and test videos, we define a multi-view action representation by concatenating $\mathbf{x}_{i,c}^v$, i.e., $\mathbf{x}_i^v = [\mathbf{x}_{i,l_i=1}^{v\,T}, \mathbf{x}_{i,l_i=2}^{v\,T}, \ldots, \mathbf{x}_{i,l_i=N_C}^{v\,T}]^T$. In order to obtain a view-independent action representation $\mathbf{y}_i^v \in \mathbb{R}^D$, where $D = N_C \sum_v D_v$, we calculate the coefficients of the DFT by using $\mathbf{x}_i^v$, i.e.,:

$$y_i^v(k) = |\sum_{n=0}^{D-1} x_i^v(n) e^{-i\frac{2\pi k}{D}n}|, \quad k = 1, ..., D-1. \quad (2)$$

After the calculation of the multi-view action representations $\mathbf{y}_i^v$, we can proceed by following the above-described unsupervised learning process for action video group determination. Accordingly, a test action instance can be represented by employing a view-independent action representation $\mathbf{y}_t^v$ and assigned to the corresponding cluster label.

A. Temporal Video Segmentation

As has been previously described, test videos are not guaranteed to depict simple actions. Thus, the assignment of such videos in video groups containing simple actions would be wrong. In order to address this issue, we propose a method for automatic temporal video segmentation that exploits the Dense Trajectory-based video description.

Let us assume that a test video appears. We employ the Dense Trajectory-based video description in order to calculate descriptors $\mathbf{d}_i^v, i = 1, \ldots, N_d, v = 1, \ldots, V$ on the trajectories of densely-sampled video frame interest points. N_d is the number of interest points detected in the test video. We concatenate the five descriptor vectors $\mathbf{d}_i^v$ in order to fuse the local shape and motion information appearing in each trajectory, i.e., $\mathbf{d}_i = [\mathbf{d}_i^{1\,T}, \ldots, \mathbf{d}_i^{5\,T}]^T$. Subsequently, we apply K-Means clustering [14] on $\mathbf{d}_i$ in order to calculate a set of descriptor prototypes (codebook). By using this codebook, which is exclusively derived from the video under consideration, and the video frame indices corresponding to each trajectory, we create BoW-based representations of (overlapping) video segments, each consisting of T video frames and succeeding its neighbor by one frame. In all our experiments we have used the value $T = 20$. Let us denote the number of the resulting video segments by M. Let us also denote by $\mathbf{v}_j, j = 1, \ldots, M$ the BoW-based representations of the resulting video segments. By exploiting the video segment temporal relationship, we create two sets of video segment representations $\mathcal{S}_i$, $i = 1, 2$, each consisting of M_i, $i = 1, 2$ vectors $(M_1 + M_2 = M)$. By employing $\mathbf{v}_j, j = 1, \ldots, M$ and the corresponding set labels c_j, the within-set and total variance can be measured by:

$$s_w = \sum_{i=1}^{2} \sum_{j, c_j = i} (\mathbf{v}_j - m_i)^T (\mathbf{v}_j - m_i), \quad (3)$$

$$s_T = \sum_{j=1}^{M} (\mathbf{v}_j - m)^T (\mathbf{v}_j - m), \quad (4)$$

where

$$m = \frac{1}{M}\sum_{j=1}^{M} \mathbf{v}_j = \frac{1}{M}\sum_{i=1}^{2} M_i m_i, \quad (5)$$

$$m_i = \frac{1}{M_i}\sum_{j, C_j = i} \mathbf{v}_j. \quad (6)$$

After calculating s_w, s_T, we can calculate the so-called *Fisher criterion* $J = s_w/s_T$. Since $\mathbf{v}_j$ represent the video segments in the test video, the minimization of J leads to the maximization of the compactness of the two video segment sets $\mathcal{S}_1, \mathcal{S}_2$. In order to determine an optimal temporal segmentation of the test video, in terms of J minimization, we follow a line search strategy for the determination of the best parameter value N_1. The above-described process is illustrated in Figure 1.

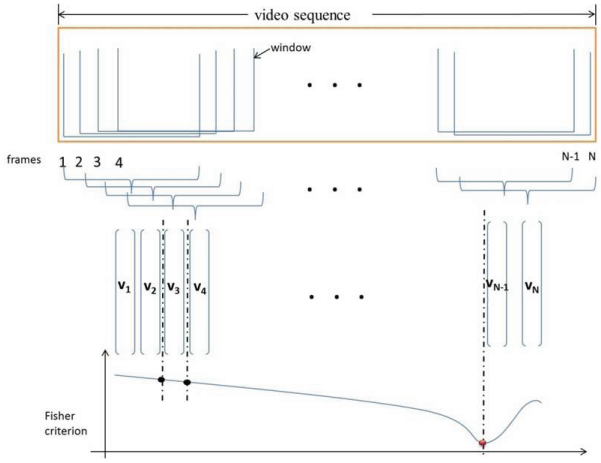


Fig. 1. Example of the temporal criterion.

An example of the above described temporal video segmentation process on a video sequence of the i3DPost database that depicts actions "walk" and "sit" is illustrated in Figure 2.

In the case where action instances are depicted in multiple (N_C) synchronized videos, each captured by a different viewpoint, we can exploit the available multi-view information in order to enhance temporal segmentation performance. Since all the N_C synchronized videos depict the same action instance, we expect that they should be temporally segmented at the same video frame index. We apply the above-described process on each of the N_C videos independently and determine the video frame cuts for each of them. Let us denote by t_i, $i = 1, \ldots, N_C$ the video frame indices determined for each of the N_C videos. We determine the action instance cut frame to be the mean video frame cut, i.e., $t = \frac{1}{N_C} \sum_{i=1}^{N_C} t_i$. In order to avoid taking into account outlier video frame cut values in the determination of the action instance cut, we can discard the α smaller and α larger t_i values before t calculation (α-trimming).

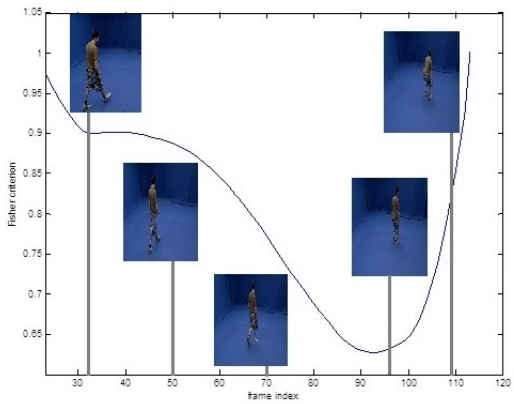


Fig. 2. Example of a video correspondence.

III. Experiments

In this Section we describe experiments conducted in order to evaluate the performance of the proposed methods. We have employed the i3DPost multi-view action database [15] and a new multi-view action database (that has been recorded for the needs of the FP7 EU project IMPART) [16] in order to evaluate both the temporal video segmentation and the video characterization methods. i3DPost action videos are captured in a constrained environment with blue background, while the IMPART database has complex background and periodical occlusions or interfering irrelevant movements. The i3DPost database contains 7 simple actions, specifically *run*, *walk*, *jump*, *jump-in-place*, *handwaving*, *bend* and *sit-standup*. It also contains video sequences containing complex actions, which are *run-fall* and *walk-sit* (2 actions) and *run-jump-walk* (3 actions). All actions are performed by 8 persons and are captured by 8 cameras, uniformly arranged inside the studio. We use 448 train samples (7 $actions \times 8\ persons \times 8\ cameras$) and 448 test samples, i.e., 256 2-action test samples (2 $instances \times 8\ pers. \times 8\ cam. \times 2\ segments$) and finally 192 3-action test samples (1 $instance \times 8\ pers. \times 8\ cam. \times 3\ segm.$). In multiview clustering, all N_C synchronized videos are combined, and thus the number of available samples is divided by the number of cameras. So, for this experiment we have 56 train and 56 test samples. The IMPART database is similar, with the difference that the actors are 7 and some of them perform the same action either in front of a scattered background or a blue one. Moreover, the cameras are 16 in this database. Here, train videos are 1024 and test ones are 848 in singleview setting. In the multiview setting, the training and test action sequences are equal to 64 and 53, respectively.

A. Temporal Segmentation experiments

The codebook cardinality was set equal to 50 for the i3DPost, and 70 for the IMPART datasets, since it was observed that for more complex environments a bigger codebook is preferable in encoding descriptors. Values larger than these gradually deteriorate the results. If S is the starting frame of the window presenting a minimum in the *Fisher criterion* J minimisation, the exact cut frame is set to $S + 0,4\,T$. Here we have to note that the diagram depicting obtained J values always presents a minimum. This is because the initial as well as the last values of the diagram are computed from degenerate clusters, one containing just a few histograms while the second containing all the others. This separation is not at all discriminative and gives high criterion values. As the cluster member number divergence gradually decreases, this problem is eliminated.

To decide whether a temporal segmentation is valid, we have manually determined action transition zones for each of the action sequences. A value of N_1 determined by the algorithm that falls inside the action transition zone was regarded to be a correct cut, while values of N_1 not belonging to the action transition zone were regarded to be incorrect. Transition zone is set by the ground truth cut frame allowing a tolerance of 10 frames, i.e. [t_i-10 t_i+10]. The obtained results are illustrated in Table I. As can be seen in this Table, by segmenting each of the available videos independently, classification rates equal to 75.78% and 76.71% have been obtained for the i3DPost and IMPART datasets, respectively. Multiview fusion enhances performance, since these classification rates are increased to 84.37% and 90.84%, respectively. Qualitative results on an action sequence of the i3DPost database can be seen in Figure 2.

TABLE I
2 & 3 action segmentation accuracy.

Database	Singleview (%)	Multiview(%) ($\alpha - trim$)
i3DPost	75.78	81.25 ($\alpha = 1$)
		84.37 ($\alpha = 2$)
IMPART	76.71	87.78 ($\alpha = 4$)
		90.84 ($\alpha = 6$)

B. Video Characterization Experiments

Clustering is performed either for each camera or in a multi-view setting. We have used a value of $D_v = 500, v = 1, \ldots, 4$, for video representation in all the experiments, employing all descriptors except *HOF*. We observed that it is better to use only the other 4, possibly due to their higher discriminative power in activity tasks, like this. In order to evaluate the clustering performance we need an appropriate metric. Let N be the number of labeled samples, ($N = N_I$ for the singleview case and $N = N_T$ for the multiview case), $C = \{c_1, c_2, \ldots, c_k\}$ be the clusters found by the clustering algorithm, $W = \{w_1, w_2, \ldots, w_c\}$ be the classes of the labeled objects, c_k be the objects composing cluster k and w_c the objects composing class c, $|C_k|$ be the cardinality of cluster k and $n_j^i = |w_i \cap C_j|$ be the number of objects of cluster i being in class j. We have chosen the F-measure as the evaluation metric, which combines the precision and recall measures [17]. The recall of cluster j with respect to cluster i, $r(i,j)$ is defined as $\frac{|c_j \cap w_i|}{|w_i|}$, where $||$ denotes the set cardinality. The precision of cluster j with respect to cluster i, $p(i,j)$, is defined as $\frac{|c_j \cap w_i|}{|w_j|}$. The F-measure combines both values as follows:

$$F_{i,j} = \frac{2}{\frac{1}{p(i,j)} + \frac{1}{r(i,j)}} = \frac{2p(i,j)r(i,j)}{p(i,j) + r(i,j)}. \quad (7)$$

The overall F-measure of clustering is given by:

$$F = \sum_{i=1}^{C} \frac{|w_i|}{N} max_{j=1,\ldots,K} F_(i,j). \quad (8)$$

The results of singleview-multiview segmentation and clustering are shown in Tables II and III for train and test videos, respectively. Clearly, multiview segmentation improves accuracy and also contributes to the enhancement of video characterization performance. We observed that in the singleview setting, the clustering algorithm finds it difficult to distinguish between similar actions, such as *walk-run* or *bend* and *sit-standup*. This is reasonable, since actions are quite different when observed from different viewing angles. However, by exploiting the enriched multiview visual information, discrimination is increased, achieving an F-measure value equal to 1 for the i3DPost train samples. Strangely enough, the same value for IMPART database is quite low and smaller than singleview, because samples in this case are much fewer and the problem becomes more difficult. However, test performance is much better in both databases. It is also remarkable that the method assigns more than 50% of test samples of an unknown class (*fall*) in a separate cluster, which indicates that the algorithm can recognize and group together new test samples of new concepts (action classes).

TABLE II
F-MEASURE (TRAIN VIDEOS)/ NUMBER OF CLUSTERS

Database	Singleview Segm. Singleview Clust.	Multiview Segm. Singleview Clust.	Multiview Segm. Multiview Clust.
i3DPost	0.666 / 9	0.6957 / 7	**1** / 7
IMPART	0.7264 / 8	**0.7383** / 8	0.5337 / 8

TABLE III
F-MEASURE (TEST VIDEOS)

Database	Singleview Segm. Singleview Clust.	Multiview Segm. Singleview Clust.	Multiview Segm. Multiview Clust.
i3DPost	0.5134	0.5627	**0.7438**
IMPART	0.5485	0.6271	**0.7770**

IV. CONCLUSION

In this paper, we proposed an efficient method for video characterization based on activity information. The method employs a state-of-the-art video representation in order to determine human activity concepts. A view-independent multi-view action representation is achieved by exploiting the circular shift invariance property of the coefficients of the DFT transform, while temporal video segmentation is performed as a pre-processing step in order to determine shorter videos containing simpler actions. Experimental results on two multi-view action databases denote the effectiveness of the proposed approach.

ACKNOWLEDGMENT

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement number 316564 (IMPART).

REFERENCES

[1] I. Laptev, M. Marszalek, C. Schmind and B. Rozenfeld, "Learning realistic human actions from movies," *Computer Vision and Pattern Recognition*, 2008.
[2] X. Ji and H. Liu, "Advances in view-invariant human motion analysis: a review", *IEEE Transactions on Systems, Man & Cybernetics*, Part-C, vol. 40, no. 1, pp. 13-24, 2010.
[3] A. Iosifidis, A. Tefas and I. Pitas, "Minimum Class Variance Extreme Learning Machine for Human Action Recognition", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 11, pp. 1968-1979, 2013.
[4] Y. Yang, I. aleemi and M. Shah, "Discovering motion primitives for unsupervised grouping and one-shot learning of human actions, gestures, and expressions", *IEEE Transactions on Pattern Anaysis and Machine Intelligence*, vol. 35, no. 7, pp. 1635-1648, 2013.
[5] H. Wang, A. Klaser, C. Schmid, and C. L. Liu, "Dense trajectories and motion boundary descriptors for action recognition", *International Journal of Computer Vision*, vol. 103, no. 60, pp. 1-20, 2013.
[6] Y. Huang, Z. Wu, L. Wang and T. Tan, "Feature coding in image classification: A comprehensive study". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, pp. 493-506, 2014.
[7] P. Spagnolo, T.D' Orazio, M. Leo, and A. Distante. "Moving object segmentation by background subtraction and temporal analysis", *Image and Vision Computing*, vol. 24, no. 5, pp. 411-423, 2006
[8] A. Iosifidis, A. Tefas and I. Pitas, "View-invariant action recognition based on Artificial Neural Networks", *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 3, pp. 412-424, 2012.
[9] M. Sigalas, H. Baltzakis, P. Trahanias, "Visual tracking of independently moving body and arms", *IEEE/ RSJ International Conference on Intelligent Robots and Systems*, 2009.
[10] A. Iosifidis, A. Tefas, N. Nikolaidis and I. Pitas, "Multi-view Human Movement Recognition based on Fuzzy Distances and Linear Discriminant Analysis", *Computer Vision & Image Understanding*, vol. 116, pp. 347-360, 2012.
[11] M. Jain, H. Jegou and P. Bouthemy, "Better exploiting motion for better action recognition", *Computer Vision and Pattern Recognition*, 2013.
[12] J. Tayor, and N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press, 2004.
[13] J. Zhang, M. Marszalek, M. Lazebnik and C. Schmid, "Local features and kernels for classification of texture and object categories: A comprehensive study", *International Journal of Computer Vision*, vol. 73, no. 2, pp. 213-238, 2007.
[14] R. Duda, P. Hart and D. Stork, *Pattern Classification*, 2nd ed., John Wiley &Sons, 2001.
[15] N. Gkalelis, H. Kim, A. Hilton, N. Nikolaidis and I. Pitas, "The i3DPost multiview and 3D human action/interaction database", *Conference on Visual Media Production*, 2009.
[16] [online]. Available: http://kahlan.eps.surrey.ac.uk/impart/
[17] N. Vretos, V. Solachidis and I. Pitas, "A Mutual Information Based Face Clustering Algorithm For Movie Content Analysis", *Image and Vision Computing*, vol. 29, no. 10, pp. 693-705, 2011.

Optimal Multichannel Equalization for Robust Speech Dereverberation

Abdullah-Al-Maruf* and Mohammed Ariful Haque

Electrical and Electronics Engineering, Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
*maruf3e@yahoo.com

Abstract—**Ideal acoustic multichannel equalization technique based on multiple-input/output inverse theorem (MINT) aims to completely invert the estimated room impulse responses (RIRs) between the source and microphone arrays. However, the method is known to be very sensitive to the estimation errors of the RIRs. In order to increase robustness, regularized MINT (R-MINT) has been proposed that overcomes the limitation by decreasing the energy of the equalization filters. In this paper, we analyze the solution of the R-MINT method and introduce a number of different ways to increase the robustness of the MINT method. Recently, regularized partial MINT (RP-MINT) method has been proposed that provides better perceptual speech quality compare to the R-MINT method. We also propose a new regularized optimal partial MINT (ROP-MINT) method which aims to optimize the equalized impulse response (EIR) between the source and output of the system. Simulation experiments conducted in various reverberant environments demonstrate that performance of our proposed ROP-MINT is better than that of the RP-MINT method.**

Index Terms—**speech dereverberation, multichannel equalization, regularization, optimization.**

I. Introduction

Speech signal recorded in an enclosed space by the microphones placed at a distance from the source is corrupted by its delayed and attenuated copies, which is known as reverberation. Reverberation decreases the speech intelligibility and degrades the performance of speech recognition systems. So, efficient dereverberation techniques can enhance the performance of many speech communication systems such as hands-free telephony, hearing aids and automatic speech recognition [1].

Dereverberation methods developed so far can be classified into single and multichannel techniques depending on the number of microphones used to record the speech signal [2]. Multichannel techniques are usually preferred than the single channel counterparts since spectro-temporal and spatial processing of the received microphone signals are possible. Acoustic multichannel equalization techniques, which aim to cancel out reverberation effect by inverse filtering, offer the most attractive approach to speech dereverberation. It is the only approach that offers perfect speech dereverberation in theoretical point of view.

The pioneer work in multichannel equalization approaches was done by Miyoshi and Kaneda [3] by proposing an elegant method commonly known as multiple-input/output inverse theorem (MINT). But the MINT method has a serious drawback in practice since it is extremely sensitive to the channel estimation errors and additive noise. In order to increase robustness of MINT, it has been proposed to reduce the energy of the equalizing filters by incorporating regularization [4]. The idea is to add equalizing filter energy by a regularization constant in the cost function which has to be minimized.

Besides regularized MINT (R-MINT) method, there is another way to increase the robustness of multichannel equalization techniques. It is reported in the literature that early reflections are not harmful for speech intelligibility [5], [6] and the late reverberation (defined as part of the room impulse response (RIR) after 50-80 ms) is the major reason of degrading speech quality. It has been experimentally shown that equalization of the late reflections is more robust than complete equalization in the presence of channel estimation errors [7]. Hence, more robust multichannel equalization techniques such as relaxed multichannel least-squares (RMCLS) [7] and channel shortening (CS) [8] are proposed that suppress the late reflections alone. Since any constraint is not imposed on the early reflections of the equalized impulse response (EIR), RMCLS and CS generally lead to undesired perceptual effects. Recently, a partial MINT (P-MINT) method has been proposed which aims to achieve both robustness and better perceptual speech quality [9]. It has been shown in [9] that by using the first part of one of the estimated RIRs as the early reflection part of the EIR, the regularized P-MINT (RP-MINT) can outperform all multichannel equalization techniques.

In this paper, we first analyze the solution of the R-MINT method and demonstrate how the regularization constant increases the robustness of the MINT method. We also present different ways to formulate robust MINT method other than the regularization approach. Secondly, we propose a new partial multichannel equalization technique that optimally determines the early reflection part of the EIR to increase perceptual quality while suppressing the reverberant tail of the RIRs. Our proposed regularized optimal partial MINT (ROP-MINT) is compared with the RP-MINT through simulation experiments in various reverberant conditions. Simulation results show that the dereverberated speech using the ROP-MINT method has better perceptual quality than that using the RP-MINT method.

II. Problem formulation

Consider a speech signal is recorded using M microphones in a reverberant environment. The ith microphone signal can be written as

$$x_i(n) = s(n) * h_i(n), \qquad i = 1, \dots, M \tag{1}$$

where $s(n)$ is clean speech signal, $h_i(n)$ is the RIR between the source and ith microphone and $*$ denotes convolution operation.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

The output signal of multichannel equalized system is given by the sum of the equalized microphone signals which can be written as

$$\hat{s}(n) = \sum_{i=1}^{M} x_i(n) * g_i(n) \tag{2}$$

$$= s(n) * \sum_{i=1}^{M} h_i(n) * g_i(n) \tag{3}$$

$$= s(n) * d(n) \tag{4}$$

where $g_i(n)$ is the equalization filter for the ith microphone, $d(n)$ is the EIR between the source and output of the systems. In matrix form, $d(n)$ can be written as

$$\mathbf{d} = \mathbf{Hg} \tag{5}$$

$$\mathbf{H} = [\,\mathbf{H_1}\ \mathbf{H_2} \dots \mathbf{H_M}\,] \tag{6}$$

$$\mathbf{g} = [\,\mathbf{g_1^T}\ \mathbf{g_2^T} \cdots \mathbf{g_M^T}\,]^T \tag{7}$$

where $\mathbf{H}$ is the multichannel convolution matrix and $\mathbf{g}$ is the stacked equalization filter vector.

The equation (5) is used to obtain the equalizing filters, $\mathbf{g}$. Since true RIRs are not available in reality, $\mathbf{g}$ is evaluated using the estimated multichannel convolution matrix, $\hat{\mathbf{H}}$ instead of using true convolution matrix, $\mathbf{H}$. Here, $\hat{\mathbf{H}}$ is constructed from estimated RIRs, $\hat{h}_i$, which can be written as

$$\hat{h}_i(n) = h_i(n) + e_i(n) \tag{8}$$

where $e_i(n)$ is the estimation error.

Our aim is to obtain the dereverberated speech signal $\hat{s}(n)$, which is perceptually close to the clean speech signal $s(n)$.

III. Improving the Robustness of the MINT Method

To make the MINT less sensitive to channel estimation errors, the energy of the equalizing filters needs to be reduced. So in the cost function for R-MINT, the energy of $\mathbf{g}$ is added by a regularization constant. Hence, the cost function that is to be minimized is given by

$$J(\mathbf{g}) = \left\|\hat{\mathbf{H}}\mathbf{g} - \mathbf{d}\right\|_2^2 + \delta\|\mathbf{g}\|_2^2 \tag{9}$$

where δ is a parameter known as regularization constant. Here, $\mathbf{d}$ is the expected EIR which can be written as

$$\mathbf{d} = [1\ 0\ 0 \dots 0] \tag{10}$$

The closed form solution for (9) can be written as

$$\mathbf{g} = \left(\hat{\mathbf{H}}^T\hat{\mathbf{H}} + \delta\mathbf{I}\right)^{-1}\hat{\mathbf{H}}^T\mathbf{d} \tag{11}$$

where $\mathbf{I}$ is an unity matrix. Applying eigenvalue decomposition, we get $\hat{\mathbf{H}}^T\hat{\mathbf{H}} = \mathbf{USU}^T$, where $\mathbf{U}$ is the eigenvector matrix and $\mathbf{S}$ is a diagonal matrix with eigenvalues s_i in the diagonal entries. Now, (11) can be rewritten as

$$\mathbf{g} = (\mathbf{USU}^T + \delta\mathbf{UU}^T)^{-1}\hat{\mathbf{H}}^T\mathbf{d} \tag{12}$$

$$= \mathbf{UTU}^T\hat{\mathbf{H}}^T\mathbf{d} \tag{13}$$

where $\mathbf{T}$ is a diagonal matrix with $\frac{1}{(s_i+\delta)}$ in the diagonal entries.

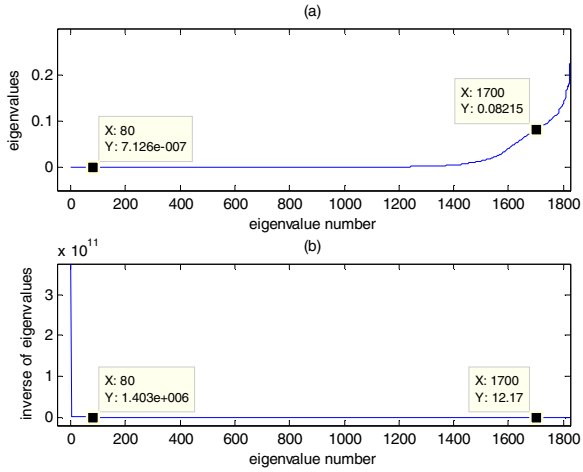


Fig. 1 (a) eigenvalus of $\hat{\mathbf{H}}^T\hat{\mathbf{H}}$, (b) inverse of the eigenvalues of $\hat{\mathbf{H}}^T\hat{\mathbf{H}}$

Initial few eigenvalues for any estimated channels are observed to be nearly zero which give extremely large values in inversion as it is shown in figure 1. In MINT, these extremely large inversed values get multiplied by entries of $\mathbf{H}$ instead of corresponding nearly zero valued entries of $\hat{\mathbf{H}}$. A small difference between $\mathbf{H}$ and $\hat{\mathbf{H}}$ creates undesired peaks in the EIR making MINT very sensitive to channel estimation errors. It is clear from (13) that R-MINT method actually adds regularization parameter δ with each eigenvalues s_i which prevents getting extremely large values in inversion. This lessens distortion in the EIR significantly making MINT more robust.

So instead of shifting eigenvalues with a constant amount δ, it is possible to incorporate regularization a number of different ways to make MINT less sensitive to channel estimation errors. Here we introduce some new ways to incorporate regularization.

A. Clipping the extremely large inverse of eigenvalues

Since extremely large inversed eigenvalues are the reason of distortion, it is not required to change all the inversed eigenvalues. Only the values over a threshold are needed to be changed while keeping intact the other inversed eigenvalues. Hence we propose the clipping method where only the inversed eigenvalues over a threshold will be changed and replaced by a moderate large value. Hence, for proposed clipping method the equation (13) becomes as

$$\mathbf{g} = \mathbf{UP}\,\mathbf{U}^T\hat{\mathbf{H}}^T\mathbf{d} \tag{14}$$

where $\mathbf{P}$ is a diagonal matrix. The diagonal element, p_i is computed as

$$p_i = \begin{cases} b, & if\ (1/s_i) \geq b \\ 1/s_i\,, & else \end{cases} \tag{15}$$

Here b is a threshold parameter which can be set to different values to get optimal performances for different reverberant environments.

B. Using variable regularization parameter

It is not required to use constant value of δ for each eigenvalue. Suitable value of the regularization parameter δ_i can be used for each eigenvalue s_i instead. Hence, we propose variable δ method for regularization which uses different value δ_i suitable for each eigenvalue s_i. So, for proposed variable δ method the equation (13) becomes as

$$\mathbf{g} = \mathbf{U}\mathbf{V}\,\mathbf{U}^T\hat{\mathbf{H}}^T\mathbf{d} \tag{16}$$

where $\mathbf{V}$ is a diagonal matrix with $\frac{1}{(s_i+\delta_i)}$ in the diagonal entries. So, the original R-MINT method is actually a special case of the variable δ method when all the δ_i become same.

C. Taking roots of the eigenvalues

In order to make MINT robust, we need to prevent getting extremely large values in the inversion of eigenvalues which can be done by increasing the small value of the eigenvalues. Since taking root increases the value of positive number less than 1, the eigenvalues can be replaced by their roots. Hence, we propose our root method for regularization where we replace each eigenvalues with their root in (13). So the equation (13) becomes as

$$\mathbf{g} = \mathbf{U}\mathbf{W}\,\mathbf{U}^T\hat{\mathbf{H}}^T\mathbf{d} \tag{17}$$

where $\mathbf{W}$ is a diagonal matrix with $\frac{1}{s_i^{1/\rho}}$ in the diagonal entries. The power constant ρ has to be selected for optimal performance like the regularization constant δ. Variable values can also be used for ρ which may improve performance but increases complexity.

IV. Proposed Regularized Optimal Partial MINT Method

The cost function of RP-MINT method [9] is given by

$$J(\mathbf{g}) = \left\|\hat{\mathbf{H}}\mathbf{g} - \hat{\mathbf{h}}_\mathbf{p}^\mathbf{d}\right\|_2^2 + \delta\|\mathbf{g}\|_2^2 \tag{18}$$

where $\hat{\mathbf{h}}_\mathbf{p}^\mathbf{d}$ is the desired EIR for RP-MINT. It contains one of the estimated RIRs in its early reflection part as long as the estimated RIRs are perceptually close to the true RIRs. So, $\hat{\mathbf{h}}_\mathbf{p}^\mathbf{d}$ can be written as

$$\hat{\mathbf{h}}_\mathbf{p}^\mathbf{d} = \left[\,\hat{h}_p(0)\;\; \hat{h}_p(1) \ldots \hat{h}_p(L_s-1)\; 0\; 0 \ldots 0\right] \tag{19}$$

where L_s is the length of the direct path and early path reflections in number of samples which is typically 50-80 ms.

When the environment is highly reverberant and estimated RIRs include large estimation errors, RP-MINT gives perceptually poor performance. It would be better to find an optimal equalization target rather than using a fixed $\hat{\mathbf{h}}_\mathbf{p}^\mathbf{d}$ as defined in (19). In order to find an optimal EIR we treat $\mathbf{d}$ as variable. So we propose a new least square cost function for our method as given below

$$J(\mathbf{g},\mathbf{d}) = \left\|\hat{\mathbf{H}}\mathbf{g} - \mathbf{d}\right\|_2^2 + \delta\|\mathbf{g}\|_2^2 + \gamma\|\mathbf{d} - \mathbf{d_0}\|_2^2 \tag{20}$$

where γ is a new constant, and $\mathbf{d_0}$ is an impulse function which is the ideal value for the EIR. Note that, in the cost function the last term $\|\mathbf{d} - \mathbf{d_0}\|_2^2$ tries to make EIR to be an impulse function and thus avoids EIR from taking any undesired value.

If we solve $\mathbf{g}$ and $\mathbf{d}$ after partially differentiating the cost function with respect to $\mathbf{g}$ and $\mathbf{d}$ and setting them to zero, we will get

$$\mathbf{g} = \left(\hat{\mathbf{H}}^T\hat{\mathbf{H}} + \delta\mathbf{I}\right)^{-1}\hat{\mathbf{H}}^T\mathbf{d} \tag{21}$$

and

$$\mathbf{d} = \frac{1}{1+\gamma}\left(\hat{\mathbf{H}}^T\mathbf{g} - \gamma\mathbf{d_0}\right) \tag{22}$$

respectively. These values of $\mathbf{g}$ and $\mathbf{d}$ will give minimum J. Iterative solution of above two equations gives the equalizing filter $\mathbf{g}$ for optimal EIR.

So the algorithm for ROP-MINT can be written as:

1) Set initial $\mathbf{d}$ equal to $\mathbf{d_0}$ (any initial value can be chosen, e.g., $\hat{\mathbf{h}}_\mathbf{p}^\mathbf{d}$.)
2) Calculate $\mathbf{g}$ from the equation (21).
3) Calculate $\mathbf{d}$ from the equation (22) using the value of $\mathbf{g}$ from step 2.
4) Use only first L_s number of coefficient of $\mathbf{d}$ obtained from step 3 normalized by $d(0)$ for the new value of $\mathbf{d}$.

$$\mathbf{d} \leftarrow \frac{1}{d(0)}\,[d(0)\; d(1) \ldots d(L_s-1)\;\; 0 \ldots 0]$$

5) Repeat from step 2 with the value of $\mathbf{d}$ obtained in step 4.

V. Simulation Results

In this section we present simulation results to evaluate the performance of our proposed techniques. Different objective measures of evaluating the quality of speech such as average segmental SNR (segSNR) and perceptual evaluation of speech quality (PESQ) [10] are used to evaluate performance. The higher the values of the segSNR and PESQ are, the better the speech quality is.

For simulation, the room size with dimension of $(5\text{m} \times 4\text{m} \times 3\text{m})$ and a linear array of 8 microphones starting from $(1.0, 1.6, 1.6)$ with 0.1m separation in y direction were used. The RIRs of length $L_s = 1600, 2400, 3200, 4000$ were generated using the well known image model [11] for the reverberation time $T_{60} = 0.2, 0.3, 0.4, 0.5$ ms respectively. To generate estimation errors in the estimated RIRs white zero mean Gaussian noise at $SNR = 20$ dB is added to each RIR.

A. Performance of proposed regularized MINT methods

Table I presents the PESQ values of the dereverberated speech using our proposed methods and original R-MINT method. The experiment was conducted for various reverberation times. In simulation $\delta = 0.001$ was chosen optimally for R-MINT to get the best performance. For clipping method threshold value b was chosen to be 1000. In variable δ parameter method, for the first 20% eigenvalues $\delta = 0.001$ was used while $\delta = .002, .005$ were used for the middle 60% and last 20% eigenvalues. In our root method $\rho = 1.6$ was used to get optimal performance. We see from the Table I that all these methods increase PESQ values of the obtained dereverberated signal. Note that, some PESQ values of dereverberated speech using our methods are even greater

than the dereverberated speech using original R-MINT method. Hence, it is clear from the PESQ values presented in the table that all the proposed methods are almost as effective as the original one.

TABLE I
PESQ OF VARIOUS METHOD OF INCORPORATING REGULARIZATION

T_{60} (s)	Rever-berated	Original R-MINT method	Proposed R-MINT methods		
			Clipping method	Variable δ method	Root method
0.2	2.263	2.851	2.823	2.822	2.856
0.3	2.073	2.486	2.449	2.487	2.483
0.4	1.947	2.301	2.278	2.294	2.326

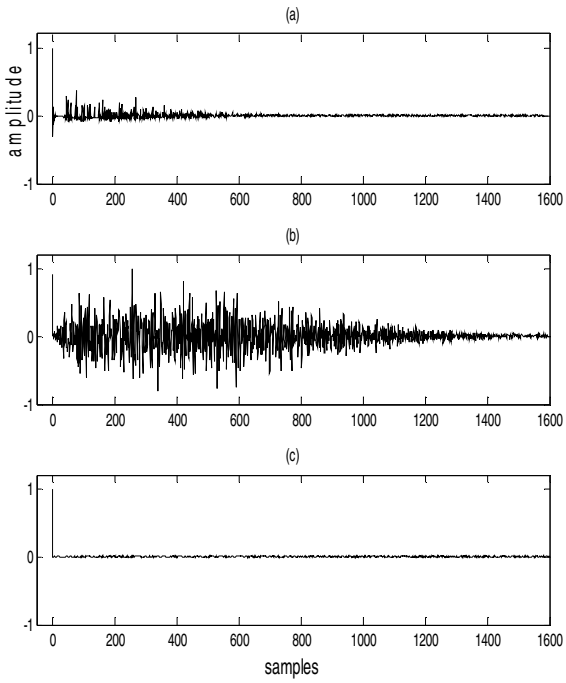


Fig. 2 (a) RIR for first microphone, (b) EIR obtained from MINT, (c) EIR obtained from root method for $\rho = 1.6$

Figure 2 shows the RIR for the first microphone and severely distorted EIR obtained from MINT method. EIR obtained from root method is also shown in the figure. It is clear from the figure that our proposed root method improves the EIR significantly making it almost like impulse function. All other proposed methods also show similar EIR pattern which make MINT more robust.

B. Performance of the proposed ROP-MINT method

Table II presents the PESQ and segSNR values of the dereverberated speech using our proposed ROP-MINT method and the RP-MINT method. The experiment was conducted for various reverberation times. In simulation, we used $\delta = 0.001$ and $L_s = 400$ for RP-MINT to get the best performance. Same values of δ and L_s were used for ROP-MINT while we used $\gamma = 5$ for the new constant. Initial $\mathbf{d}$ was set to $\mathbf{d}_0$. Only 5 iterations were taken for ROP-MINT method.

TABLE II
PERFORMANCE COMPARISON OF ROP-MINT

T_{60} (s)	Reverberated		RP-MINT		ROP-MINT	
	PESQ	seg-SNR	PESQ	seg-SNR	PESQ	seg-SNR
0.3	1.969	-2.181	2.486	-0.459	2.817	5.228
0.4	1.889	-2.688	2.395	-1.205	2.700	4.535
0.5	1.691	-3.029	2.326	-1.494	2.598	3.872

It is clear that in all different cases the PESQ and segSNR values of dereverberated speech using proposed ROP-MINT method are greater than the dereverberated speech using RP-MINT method. So our proposed ROP-MINT is more robust and perceptually advantageous method than the RP-MINT.

VI. CONCLUSION

In this paper, we have investigated the regularization technique and introduced various different ways of incorporating regularization in order to improve the robustness of the MINT method. Simulation results presented here shows that these new methods can dereverb speech signal almost as effectively as the original technique and sometimes even better for specific channels. We have also proposed a new equalization technique which tries to find an optimum EIR. Simulation results presented here shows that the perceptual quality of dereverberated speech obtained by our proposed ROP-MINT method is better than the RP-MINT method.

REFERENCES

[1] B.W. Gillespie and L. Atlas, "*Acoustic diversity for improved speech recognition in reverberant environments*," in ICASSP. IEEE, 2002, vol. 1, pp. 557–560

[2] P. A. Naylor and N. D. Gaubitch, *"Speech Dereverberation*," 1st ed. New York, NY, USA: Springer, 2010

[3] M. Miyoshi and Y.Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. Acoust., Speech, Signal Process*., vol. 36, no. 2, pp. 145–152, Feb. 1988.

[4] T. Hikichi, M. Delcroix, and M. Miyoshi, "Inverse filtering for speech dereverberation less sensitive to noise and room transfer function fluctuations," *EURASIP J. Adv. Signal Process.*, vol. 2007, 2007.

[5] I. Arweiler and J. M. Buchholz, "The influence of spectral characteristics of early reflections on speech intelligibility," *J. Acoust. Soc. Amer.*, vol. 130, no. 2, pp. 996–1005, Aug. 2011.

[6] Warzybok, J. Rennies, T. Brand, S. Doclo, and B. Kollmeier, "Effects of spatial and temporal integration of a single early reflection on speech intelligibility," *J. Acoust. Soc. Amer.*, vol. 133, no. 1, pp.269–282, Jan. 2013

[7] W. Zhang, E. A. P. Habets, and P. A. Naylor, "On the use of channel shortening in multichannel acoustic system equalization," in *Proc. Int. Workshop Acoust. Echo Noise Control (IWAENC)*, Tel Aviv, Israel, Sep. 2010

[8] M. Kallinger and A. Mertins, "Multi-channel room impulse response shaping – A study," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*

[9] I. Kodrasi, S. Goetze, and S. Doclo, "Regularization for partial multichannel equalization for speech dereverberation," *IEEE Trans. Audio, Speech, and Lang. Process,* vol. 21, no. 9, pp. 1879–1890, September 2013.

[10] . Yi Hu and P. C. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Trans. Audio, Speech, Language Process.*, vol. 16, no. 1, pp. 229–238, 2008

[11] J. B. Allen and D. A. Berkley, *"Image method for efficiently simulating small-room acoustics,"* vol. 65, no. 4, pp. 943–950, Apr. 1979

Utterance-based Speech Dereverberation using Blind Channel Estimation and Multichannel Equalization

Mohammad Ariful Haque
Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology
Dhaka,Bangladesh
arifulhoque@eee.buet.ac.bd

***Abstract*—Speech dereverberation considering noisy environment as well as speaker's movement is a challenging task. In this paper, we present an utterance-based noise-robust speech dereverberation technique that is suitable for non-stationary speaker. The acoustic impulse responses (AIRs) between the speaker and microphone array are estimated using the spectrally constrained frequency-domain least-mean-squares (LMS) algorithm. The AIRs are then equalized using the iterative multiple-input/output inverse theorem (MINT). It is assumed that the speaker stays still within an utterance, however, the speaker changes his/her position between the utterances. The simulation experiments conducted in various reverberant environment and speaker's position demonstrate that the proposed method can satisfactorily improve the perceptual quality of the noisy reverberated speech.**

***Index Terms*—speech dereverberation, blind multichannel estimation, iterative multichannel equalization, non-stationary speaker, additive noise**

I. Introduction

Dereverberation problem typically deals with extracting the original speech from the (reverberated) microphone signals. Efficient dereverberation can enhance the performance of many speech communication systems such as hands-free telephony, hearing aids and automatic speech recognition [1]. Dereverberation of speech considering practical scenario is a difficult task. First of all, it requires blind techniques in which no training pulses can be sent to estimate the long acoustic impulse responses (AIRs), typically consisting of thousands of coefficients. The degree of complexity of dereverberation problem increases many fold in a noisy environment as the channel vector does not remain collinear with the minimum mean squared error solution [2]. Moreover, the movement of the speaker yields a spatially-varying source-microphone geometry causing the AIRs to be changed from time-to-time. An effective dereverberation technique, therefore, should consider these frequent changes in the AIRs. Although a range of speech dereverberation techniques are available in the literature, the development of practical algorithms that can truly compensate the detrimental effect of reverberation is one of the holy grails for a long time [3].

The effect of reverberation can be reduced in many ways, such as, beamforming, speech enhancement, blind deconvolution, etc. However, the best approach from theoretical point of view would be to obtain the AIRs between the speaker and each microphone and then to cancel out the reverberation by inverse filtering. The pioneer work in this approach was done by Miyoshi and Kaneda in 1988 [4]. They proposed an elegant method commonly known as multiple-input/output inverse theorem (MINT) for multichannel inversion of the room acoustics exploiting the single-input multiple-output (SIMO) structure. But the MINT method has limited application in practice due to its extreme sensitivity to the channel estimation accuracy. Various methods have been proposed to improve the robust-ness of the MINT method. In [5], the authors proposed a so called system-identification-error-robust MINT method, however, it requires knowledge of the error term between true and estimated AIRs. Another approach for robust speech dereverberation, which does not require a priori information about the true AIRs, is to reduce the energy of the equalization filter [6]. A regularized partial MINT method [7] has been proposed capitalizing the same idea of reduced filter energy. The performance of the method is dependent on the appropriate setting of the regularization parameter, which is a computationally intensive task.

Most of the dereverberation techniques available in the literature cannot deal with non-stationary speaker. For example, harmonicity based dereverberation method proposed in [8] requires around 60 minutes of speech data to acquire the inverse filter under the assumption that the system is time-invariant. In fact, the blind identification of the time-varying AIRs are difficult because of the large number of channel coefficients. Therefore, some techniques have been developed to reduce the number of channel coefficients by sacrificing the dereverberation performance. In [9], the authors proposed a speech dereverberation technique considering a moving speaker but a simplified all-pole filter is used to model the time-varying AIRs. Similarly, probabilistic models are used in [10] to reduce the number of time-varying parameters for speech dereverberation.

In this paper, we consider speech dereverberation considering the full-length AIRs. The proposed utterance-based dereverberation technique consists of three stages, namely, the initial channel delay estimation, adaptive channel estimation and iterative channel equalization. The delay estimation stage performs two important tasks: (i) tracking the movement of the speaker in different utterances and (ii) estimating the initial zeros required for the initialization of the adaptive channel estimation algorithm. Then the AIRs are estimated for the given utterance using the noise-robust multichannel LMS algorithm proposed in our previous work [11]. It is to be mentioned here that the speech data from the same utterance is repeatedly feed to the channel estimation algorithm until the steady-state condition is reached. This is how we are able to estimate the AIRs using short-length utterances. Finally, the uttered speech data block is dereverberated using a multichannel equalization filter obtained using iterative MINT method. In order to reduce the noise level of the dereverberated speech, a denoising scheme is applied after the dereverberation

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

stage.

II. RELATION TO PRIOR WORK

The proposed technique incorporates some different ideas that make it suitable for practical reverberant environment as compared to other state-of-the-art methods. First, the channel estimation is performed utterance-by-utterance so that effect of speaker's movement (between the utterances) does not disturb the channel estimation process. In continuous adaptive mode, the channel estimation algorithms cannot track even a 10 cm movement of the speaker [12]. Second, the noise-robust equalization is performed using a multichannel filter obtained from iterative implementation of the MINT method. The proposed method does not require knowledge of the true AIRs (as in [5]) or tedious singular value decomposition of the channel convolution matrix for appropriate setting of the regularization parameter (as in [7]). The formulation of the iterative MINT is conventional [13], but the robustness of this method (without any constraint) to channel estimation error is not reported so far.

III. PROBLEM FORMULATION

Consider a speech signal recorded inside an echoic room using a linear array of microphones. The received signals at the microphones can be modeled as convolutional mixtures of the speech signal and the impulse responses of the acoustic paths between source and microphones. The channel outputs and observed signals are then given by

$$y_i(n) = s(n) * h_i(n) \quad (1)$$

$$x_i(n) = y_i(n) + v_i(n), \quad i = 1, 2, \cdots, M \quad (2)$$

where n represents the time-domain sample index, M is the number of microphones, $s(n)$, $y_i(n)$, $x_i(n)$, $v_i(n)$, and $h_i(n)$ denote, respectively, the clean speech, reverberant speech, the reverberant speech corrupted by background noise, observation noise, and impulse response from the source to the ith microphone. Using vector notation, (1) can be written as

$$y_i(n) = \mathbf{h}_i^T(n)\mathbf{s}(n) \quad (3)$$

where $\mathbf{h}_i(n) = [h_{i,0}(n)\ h_{i,1}(n) \cdots h_{i,L-1}(n)]^T$ represents the channel impulse response vector of length L, the superscript T denotes transpose and $\mathbf{s}(n) = [s(n)\ s(n-1) \cdots s(n-L+1)]^T$. Our objective is to obtain a dereverberated speech signal, $\hat{s}(n)$, from the received noisy signals, $x_i(n)$, $i = 1, 2, \cdots, M$.

IV. PROPOSED UTTERANCE-BASED SPEECH DEREVERBERATION TECHNIQUE

As in [3], we assume that the speakers stay still within an utterance. Therefore, the reverberated speech signal is first partitioned into successive utterances as

$$\mathbf{x}_i = [\underbrace{x_i(0)\ x_i(1)\ \cdots\ x_i(N_1-1)}_{\text{1st utterance}}\ \cdots \underbrace{x_i(N_{k-1})\ x_i(N_{k-1}+1)\ \cdots\ x_i(N_{k-1}+N_k-1)}_{\text{kth utterance}}\ \cdots\ ] \quad (4)$$

where i denote microphone number and N_k represent the length of kth utterance. Since each utterance is processed independently, we can simplify the representation of the utterances as follows

$$\mathbf{c}_i = [c_i(0)\ c_i(1)\ \cdots\ c_i(l)\ \cdots\ c_i(N_k-1)] \quad (5)$$

where, l indicates the sample index within an utterance $\mathbf{c}_i$. In what follows, we present the initial channel delay estimation, adaptive channel identification and equalization stages of our proposed technique.

A. Estimation of initial AIR delays

The delay estimation algorithm does not require the full utterance of speech data to estimate the initial zeros of the AIRs. Therefore, a short-length speech block is formed from $\mathbf{c}_i$ as

$$\mathbf{b}_i = [c_i(0)\ c_i(1)\ \cdots\ c_i(P-1)] \quad (6)$$

where P is the length of the shortened block. Now, the channel delays can be estimated using time-difference-of-arrival (TDOA) estimation algorithms. TDOA estimate, τ_{ij}, between ith and jth channels can be obtained from the cross-correlation function of $\mathbf{b}_i$ and $\mathbf{b}_j$. When the correlation function is evaluated in the frequency-domain, τ_{ij} can be estimated as

$$\tau_{ij} = \arg\max_{\beta} \int_{-\pi}^{\pi} B_i(\omega) B_j^*(\omega) e^{-j\omega\beta} d\omega \quad (7)$$

where $B_i(\omega)$ represents the discrete Fourier transform of $\mathbf{b}_i$ and * denotes complex conjugate. Eq. (7) does not perform well in reverberant environment. A simple technique that is robust to noise and dereverberation is mentioned in [14], which is

$$\tau_{ij} = \arg\max_{\beta} \int_{-\pi}^{\pi} \cos(\angle B_i(\omega) - \angle B_j(\omega) - \omega\beta) d\omega. \quad (8)$$

Eq. (8) is used in this paper to estimate all the relative delays τ_{ij}, where $i = 1, \cdots, M-1$ and $j = i+1, \cdots, M$.

B. Robust channel identification

The AIRs of a multi-microphone system can be blindly estimated by minimizing the cross-relation error [15]

$$e_{ij}(l) = c_i(l) * \widehat{h}_j(l) - c_j(l) * \widehat{h}_i(l) \quad (9)$$

where $\widehat{h}_i(l)$ is the estimate of the ith channel impulse response. Using (9), the multichannel LMS algorithm is formulated as [16]

$$\widehat{\mathbf{h}}(l+1) = \widehat{\mathbf{h}}(l) - \mu_l \nabla J(l) \quad (10)$$

where $\mathbf{h} = [\mathbf{h}_1^T, \cdots, \mathbf{h}_M^T]^T$, $\nabla J(l)$ is the gradient of the cost function comprising of the error terms, $e_{ij}(l)$, and μ_l represents the step size. In order to ensure noise-robust performance, a spectral flatness constraint $\nabla J_p(l)$ is attached to the update equation as [11]

$$\widehat{\mathbf{h}}(l+1) = \widehat{\mathbf{h}}(l) - \mu_l \nabla J(l) + \mu_l \alpha(l) \nabla J_p(l). \quad (11)$$

For implementation details of (11), the readers are referred to [12]. For adaptive estimation, the channel vectors are initialized utilizing the delay estimates τ_{ij} as

$$\widehat{\mathbf{h}}(0) = [\mathbf{0}_{\tau_i \times 1}\ 1\ \mathbf{0}_{(L-\tau_i-1)\times 1}] \quad (12)$$

where τ_i is the relative delay of the ith channel with respect to the minimum delay channel.

C. Multichannel equalization using iterative MINT

The MINT method aims to recover the original speech signal by designing a multichannel filter $\mathbf{g}$ such that

$$\widehat{\mathbf{H}}\mathbf{g} = \mathbf{d} \tag{13}$$

where $\widehat{\mathbf{H}}$ is the channel convolution matrix, and

$$\begin{aligned} \mathbf{g} &= [\mathbf{g}_1^T \ \mathbf{g}_2^T \ \cdots \ \mathbf{g}_m^T]^T \\ \mathbf{d} &= [1 \ 0 \ \cdots \ 0]^T. \end{aligned}$$

For iterative implementation, a cost function can be formulated as

$$J_M(m) = ||\widehat{\mathbf{H}}\mathbf{g}(m) - \mathbf{d}||^2. \tag{14}$$

The gradient of $J_M(m)$ is given by

$$\nabla J_M(m) = \frac{\partial J_M(m)}{\partial \mathbf{g}(m)} = -2\widehat{\mathbf{H}}^T\mathbf{d} + 2\widehat{\mathbf{H}}^T\widehat{\mathbf{H}}\mathbf{g}(m). \tag{15}$$

The update equation of the equalization filter is given by

$$\mathbf{g}(m+1) = \mathbf{g}(m) - \mu_m J_M(m) \tag{16}$$

where μ_m is the step size. Now, the dereverberated speech can be recovered as

$$\hat{s}(k) = \sum_{i=1}^{M} g_i(k) * c_i(k). \tag{17}$$

Finally, we apply a simple spectral subtraction based denoising technique [17] to reduce the noise level of the dereverberated speech. It is to be mentioned here that we need to estimate the AIRs and equalization filter only when the speaker changes his/her position. If the speaker remains stationary, as determined by the delay estimation stage, the equalization filter of the previous utterance segment can be used to obtain the dereverberated signal.

V. Simulation Results

In this section, we present simulation results to evaluate the performance of the proposed utterance-based speech dereverberation technique considering a non-stationary speaker. Different objective measures that are used to evaluate the quality of speech are the log-likelihood ratio (LLR), average segmental SNR (segSNR), weighted spectral slope (WSS) and perceptual evaluation of speech quality (PESQ) [18]. The higher the values of the segSNR and PESQ, and the lower the values of LLR and WSS, the better the speech quality.

For simulation, the dimension of the room size was taken to be $(5 \times 4 \times 3)$m and a linear array consisting of $M = 8$ microphones was used. The first microphone was positioned at $(1.0, 1.6, 1.6)$m and the location of the other microphones can be obtained by successively adding $d = 0.2$m to the y-coordinate of the first microphone. The experiments were conducted for different reverberation times $T_{60} = 0.25, 0.5, 0.7$s. The impulse responses were generated using the well-known image model [19] and the lengths of each AIR corresponding to the above mentioned reverberation times were set to $L = 1600, 3500,$ and 5000 samples, respectively. The additive noise to each microphone signal was white zero-mean Gaussian at SNR=30 dB. The sampling frequency was 8 kHz.

First, we present channel estimation results using (11) of Section IV-B. The accuracy of the channel estimation is indicated by the normalized projection misalignment (NPM) [20]. The lower the value of NPM, the better the channel estimation. Fig. 1 shows the performance of the channel estimation algorithm for various length of the speech segment used to estimate the AIRs. It is seen that a good NPM of the estimated channels can be achieved for speech segments of duration $2T_{60}$s only. Therefore, the minimum utterance lengths that can be considered for speech dereverberation using the proposed technique are $0.5, 1$ and 1.4s for the reverberation times $T_{60} = 0.25, 0.5,$ and 0.7s, respectively. Now, we present the channel equalization performance of the iterative-MINT method utilizing the estimated AIRs. Fig. 2 shows the impulse responses of the equalized channel as well as those of the actual acoustic channel considering different reverberation times. It is seen that both the early and late reflections are substantially mitigated in the equalized channels.

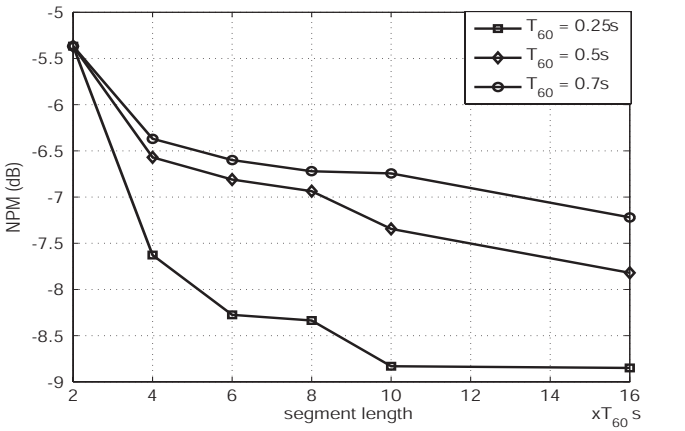


Fig. 1. Channel estimation accuracy in term of NPM for different length of speech segments.

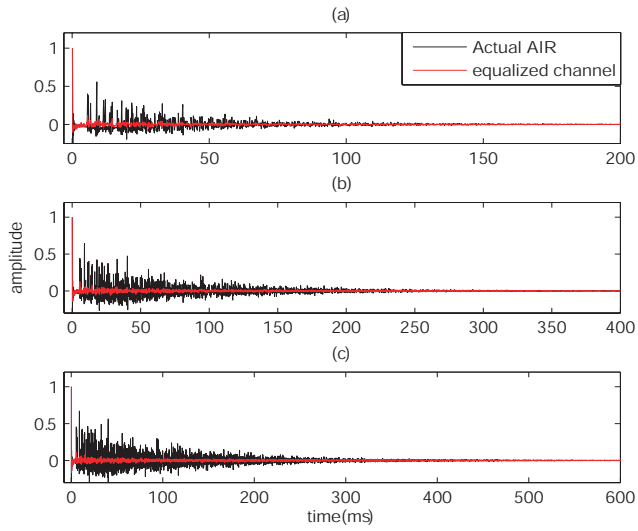


Fig. 2. Impulse response of the actual and equalized acoustic channel: (a) $T_{60} = 0.25$s, (b) $T_{60} = 0.5$s, and (c) $T_{60} = 0.7$s

The perceptual quality of the dereverberated speech using the proposed method in terms of different objective measures is presented in Table I. For comparison, the perceptual quality of the reverberated speech is also presented in the table. In both the cases, we perform spectral subtraction to reduce the noise level of the speech signal. The speaker was initially positioned at $(1.8, 2.0, 1.6)$ m. The speaker's position was successively changed in step-by-step manner at a step of 0.1m

TABLE I
PERFORMANCE OF THE PROPOSED ALGORITHM AT DIFFERENT REVERBERATION TIMES

Reverberation time (s)	NPM (dB)	LLR		segSNR		WSS		PESQ	
T_{60}	(estimated channel)	rev	derev	rev	derev	rev	derev	rev	derev
0.25	-8.028	0.603	0.459	-0.560	5.046	41.823	32.740	2.180	2.831
0.50	-6.618	0.752	0.532	-2.700	2.143	59.848	47.211	1.951	2.499
0.70	-6.269	0.893	0.618	-3.557	0.521	68.216	54.888	1.667	2.183

along x-coordinate. The duration of the uttered speech in each location was around $8T_{60}$s. The LLR, segSNR, WSS and PESQ measures presented in Table I shows the average value obtained for different locations of the speaker. The substantial improvement of the LLR, segSNR, WSS and PESQ measures with respect to the reverberated speech show that the proposed technique is effective for dereverberation for a wide range of reverberation time.

VI. CONCLUSION

We have demonstrated the effectiveness of the blind multichannel identification and equalization technique for utterance-based speech dereverberation. The proposed technique is suitable for noisy environment and non-stationary speaker. The time-varying AIRs are estimated using the spectrally-constrained frequency multichannel LMS algorithm for each utterance. It is shown that the AIRs can be effectively estimated from a speech segment of duration $2T_{60}$s. The AIRs are then equalized using the iterative-MINT method that exhibits robust performance against channel estimation error and additive noise. The quality of the dereverberated speech is substantially improved as compared to that of the received microphone signal which is clearly observed from the LLR, segSNR, WSS, PESQ measures.

REFERENCES

[1] B.W. Gillespie and L. Atlas, "Acoustic diversity for improved speech recognition in reverberant environments," in *ICASSP*. IEEE, 2002, vol. 1, pp. 557–560.

[2] M.A. Haque and M.K. Hasan, "Robust multichannel least mean square-type algorithms with fast decaying transient for blind identification of acoustic channels," *Signal Processing, IET*, vol. 2, pp. 431–441, December 2008.

[3] "Reverberant voice enhancement and recognition benchmark challenge," 2014.

[4] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 2, pp. 145–152, 1988.

[5] W. Zhang, E.A.P. Habets, and P.A. Naylor, "A system-identification-error-robust method for equalization of multichannel acoustic systems," in *ICASSP*. IEEE, 2010.

[6] T. Hikichi, M. Delcroix, and M. Miyoshi, "Inverse filtering for speech dereverberation less sensitive to noise and room transfer function fluctuations," *EURASIP J. Adv. Signal Process.*, vol. 2007, no. 1, pp. 62, 2007.

[7] I. Kodrasi, S. Goetze, and S. Doclo, "Regularization for partial multichannel equalization for speech dereverberation," *IEEE Trans. Audio, Speech, and Lang. Process*, vol. 21, no. 9, pp. 1879–1890, September 2013.

[8] T. Nakatani, T. Hikichi, K. Kinoshita, T. Yoshioka, M. Delcroix, M. Miyoshi, and B. Juang, "Robust blind dereverberation of speech signals based on characteristics of short-time speech segments," in *IEEE International Symposium on Circuits and Systems (ISCAS '07)*, 2007.

[9] J.R. Hopgood and C. Evers, "Block-based tvar models for single-channel blind dereverberation of speech from a moving speaker," in *14th Workshop on SSP*. IEEE/SP, 2007.

[10] M. Togami, Y. Kawaguchi, R. Takeda, Y. Obuchi, and N. Nukaga, "Optimized speech dereverberation from probabilistic perspective for time varying acoustic transfer function," *IEEE Trans. Audio, Speech, Language Process.*, vol. 21, no. 7, pp. 1369–1380, 2013.

[11] M. A. Haque and M. K. Hasan, "Noise robust multichannel frequency-domain LMS algorithms for blind channel identification," vol. 15, pp. 305–308, 2008.

[12] M.A. Haque, T. Islam, and T. Hasan, "Robust speech dereverberation based on blind adaptive estimation of acoustic channels," *IEEE Trans. Audio, Speech, Language Process.*, vol. 19, no. 4, pp. 775–787, 2011.

[13] S. Haykin, *Adaptive Filter Theory*, 3rd Edition, Prentice-Hall, Inc, Upper Saddle River, New Jersey, 1996.

[14] D. Halupka, A.S. Rabi, P. Aarabi, and A. Sheikholeslami, "Low-power dual-microphone speech enhancement using field programmable gate arrays," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3526–3535, 2007.

[15] L. Tong, G. Xu, and T. Kailatch, "A new approach to blind identification and equalization of multipath channels," in *25th Asilomar Conf. Signals, Syst., Comput.*, 1991, vol. 2, pp. 856–860.

[16] Y. Huang and J. Benesty, "A class of frequency-domain adaptive approaches to blind multichannel identification," *IEEE Trans. Signal Process.*, vol. 51, no. 1, pp. 11–24, Jan. 2003.

[17] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, and Signal Process.*, vol. 2, pp. 113–120, 1979.

[18] Yi Hu and P. C. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Trans. Audio, Speech, Language Process.*, vol. 16, no. 1, pp. 229–238, 2008.

[19] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," vol. 65, no. 4, pp. 943–950, Apr. 1979.

[20] D.R. Morgan, J. Benesty, and M.M. Sondhi, "On the evaluation of estimated impulse responses," *IEEE Signal Processing Lett.*, vol. 5, no. 7, pp. 174–176, July 1998.

A New Fast Inter Mode Decision Algorithm in H.264/AVC Video Coding

Md. Salah Uddin Yusuf[1*], Md. Imtiaz Hossain[2] and Mohiuddin Ahmad[3]

[1, 2, 3] Department of Electrical and Electronic Engineering
[1, 2, 3] Khulna University of Engineering & Technology
Khulna-9203, Bangladesh
[*]ymdsalahu2@gmail.com

Abstract—**In this paper, a new probability based fast inter mode decision algorithm is proposed for H.264/AVC video standard. The main idea is to classify the inter modes into large and small modes category and determine which one is suitable for a macroblock (MB). Depending on this, the probability of each large mode corresponding to the initial Hadamard SAD_{16x16} cost or the probability of each small mode corresponding to the initial Hadamard SAD_{s8x8} cost is determined. Then the best mode can be obtained from their probability characteristics. Thus, it is possible to determine the best inter mode only by calculating the Hadamard Transform (H-SAD) cost instead of rate distortion cost. Moreover, it avails the platform to determine the best mode without comparing the cost of each mode to other modes. As a result, this approach can avoid most of the complex computation processes like quantization, variable length coding, pixel reconstruction etc. The experimental results have shown that the proposed algorithm can reduce 60% to 65% of the total encoding time of H.264/ AVC with negligible distortion.**

Index Terms— **Inter Mode Decision, Macroblock, Hadamard Transform, H.264/ AVC.**

I. INTRODUCTION

H.264/AVC [1], [2] is the latest video standard which has outperformed previous video compression standard like MPEG- 1/2/3 ad H.261/263 in terms of video compression efficiency as well as video quality. It introduces many new techniques which include variable and hierarchical block transform, arithmetic entropy encoding including CABAC and CAVLC, directional spatial prediction for intra frame coding, multiple reference frame motions to outperform the previous standards. Among these features one of the key ones is it uses seven variable blocks sizes for motion estimation and compensation. Among these, large blocks (16×16, 16×8 and 8×16) are generally used for stationary image prediction with high coding efficiency and small blocks (8×8, 8×4, 4×8 and 4×4) are used for complex images for better prediction accuracy. The large blocks requires less computational time and has less accuracy than the small blocks while small blocks require more computational time and have more accuracy than the large blocks. Therefore, a trade of between the computational time and the prediction accuracy which plays a vital role for video quality is needed to decide which partition block size i.e. mode is the best one for motion estimation. This is done by computing rate distortion costs of all available blocks and selecting the best partition block size i.e. mode that can provide relatively the best motion compensation with minimum computational time. This cost calculation is also a highly complex that includes several processes like computation of integer transform, quantization and arithmetic entropy coding and pixel reconstruction process.

That's why this variable block technique involves lots of computational power. As a result, H.264/ AVC have limited applications in real time communication. In [3]-[11] different fast mode algorithms have been proposed applying which it is possible to reduce the computational complexity significantly enabling this video standard for real time application. For examples, in [3], Jing et al proposed mean absolute frame difference of a current frame and mean absolute difference of current MB to determine whether a MB is homogenous or not and decide the best mode among 16×16, 16×8 and 8×16. In [4], Kuo proposes the motion field distribution and correlation method for determining the best modes. In [5], Wei et al, proposes a mode decision algorithm where the mode of a MB is determined using the modes of adjacent MBs. Most of the algorithm proposes methods that skip some unnecessary modes, but need to calculate the rate distortion cost which needs high amount of computational time. In [7], a relative SAD (RSAD) based inter mode decision algorithm is proposed, where it is possible to determine the best mode without calculating rate distortion cost. In [9]-[11], probability based inter mode decision algorithms are proposed. In [9], Chiang proposes a H.264 encoding method based on statistical learning. Here, the best mode is determined using probability and statistic. In this paper, probability based fast new inter mode algorithm is proposed which can decide the best mode without calculating the rate distortion cost maintaining a high rate distortion performance of the H.264/ AVC codec.

Rest of the paper is organized with rate distortion cost functions for H.264/AVC standard in section II. The SKIP mode early termination algorithm is discussed in section III. Then the proposed probability based fast inter mode algorithm is described in section IV. Finally in section V, results and discussion are analysed comparing the standard and proposed algorithm.

II. RATE DISTORTION COST FUNCTION FOR H.264/AVC STANDARD

The objective of using variable block size mode in H.264/AVC standard is to ensure precise motion prediction. When a MB cannot be predicted well enough using a large block, it is suitable to divide it into small blocks and thus ensuring better motion compensation. But if it is possible to predict the MB with a large partition, there is no need for using smaller blocks because it introduces higher complexity calculating processing time. In JVT reference software JM 6.1.d [12], the rate distortion (RD) cost is define as

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

$$J_{SSD}(S, C, Mode|QP) = SSD(S, C, Mode|QP) + \lambda . R(S, C, Mode|QP \quad (1)$$

Where, $\lambda = 0.85 \times 2^{(QP-12)/3}$ is the Lagrangian multiplier and QP is the quantization parameter. The R in (1) can be written as $R = R_{header} + R_{motion} + R_{residual}$, $SSD\ (S.C)$ is the sum of the squared of the difference between the original blocks S and the reconstructed block C and it is defined as:

$$SSD(S, C) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (Sij - Cij)^2 \quad (2)$$

Here, S_{ij} is the (i, j) th element of the current block and C_{ij} is the (i, j) th element of the reconstructed block. N is the image block size and equal to 4 for H.264. In equation (1) best mode is determined from a set of modes mentioned as: $Mode \in \{SKIP, Inter\ 16 \times 16, Inter\ 16 \times 8, Inter\ 8 \times 16, Inter\ 8 \times 8, inter\ 4 \times 4, inter\ 16 \times 16\}$. Here, SKIP mode is an especial mode where no residual bit or motion is encoded. For J_{SAD} cost function, SAD value from the difference between the current MB and predicted MB is calculated. Here, J_{SAD} is defined as

$$J_{SAD} = \begin{cases} SAD(S, P) + \lambda_1 . 4K \\ SAD\ (S, P), \quad otherwise \end{cases} \quad (3)$$

where, $SAD(S, P)$ is the sum of absolute difference of the original block S and the predicted block P and defined as

$$SAD(S, P) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (Sij - Pij) \quad (4)$$

Here, S_{ij} is the (i, j) th element of the current block and P_{ij} is the (i, j) th element of the predicted block. K is equal to 0 for probable mode and 1 for other modes and λ_1 is almost the square of λ and almost the exponential function of quantization parameter QP. This SAD based can save a lot of computation than SSD cost function as it does not need reconstruction or variable length coding like the SSD cost function. But this results a degradation of coding efficiency. To compensate that degradation, sum of absolute transformed differences (SATD) is widely used which uses frequency transform, usually H-SAD of the difference between the original blocks S and the reconstructed block C. It is complex method but takes less computation than SSD based cost function method and also have a better coding efficiency than SAD based cost function. That's why H-SAD cost is considered for the proposed probability based fast inter mode algorithm.

III. PROPOSED PROBABILITY BASED FAST INTER MODE DECISION ALGORITHM

A. Determination of large and small modes

There are seven inter modes of different sizes and shapes as 16×16, 16×8, 8×16, 8×8, 8×4, 4×8 and 4×4 supported by H.264/AVC. There is also a SKIP mode which is employed to encode some static objects and background that remain almost same in adjacent frames. Modes 8×8, 8×4, 4×8 and 4×4 are considered as sub modes and are classified under Inter P 8×8 modes where $Inter\ p\ 8 \times 8 \in \{Inter\ 8 \times 8, Inter\ 8 \times 4, Inter\ 4 \times 8, Inter\ 4 \times 4\}$. Here we have classified these 7 modes into two categories; large modes and Small modes where,

$Large\ modes \in \{Inter\ 16 \times 16, Inter\ 16 \times 8, Inter\ 8 \times 16\}$
$Small\ modes \in \{Inter 8 \times 8, Inter\ 8 \times 4, Inter\ 4 \times 8, Inter\ 4 \times 4\}$.

To find the best mode for a MB, we first calculate the Hadamard $SAD_{16\times16}$ cost. Then, we divide the MB into four 8×8 sub blocks and calculate Hadamard $SAD_{s8\times8}$ cost for each block to find Hadamard $SAD_{P8\times8}$ which is defined as

$$SAD_{P8\times8} = \sum_{i=0}^{3} SAD_{s8\times8} \quad (5)$$

where, i is the ith sub block of the MB. If the cost Hadamard SAD_{16x16} is equal or less than SAD_{P8x8} then we choose the large modes, otherwise we choose the small modes.

B. Finding the best mode from large and small modes

To determine the best mode from large modes we first take the Hadamard $SAD_{16\times16}$ cost of a MB. Then, we find out which mode is the best for that cost for general cases. When a set of the Hadamard $SAD_{16\times16}$ costs are taken and their corresponding best mode is plotted in a curve an interesting characteristic of large modes is found. A characteristic curve for "Foreman" QCIF video sequence is shown in Fig.1.This is a normalized probability curve where it can be seen that a particular mode has a better probability of occurrence at a certain cost. To simplify this thing we divide the data set into small chucks and find out the dominating mode with best probability. The range at which a certain mode dominates is marked with a threshold value. For three large modes (16×16, 16×8 and 8×16) two thresholds are enough to distinguish among them.

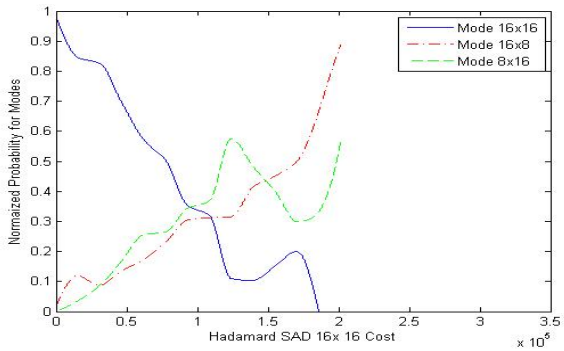


Fig.1 Probability Characteristic Curve of Large Modes

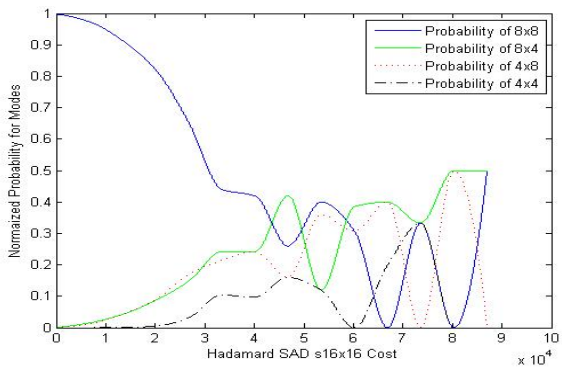
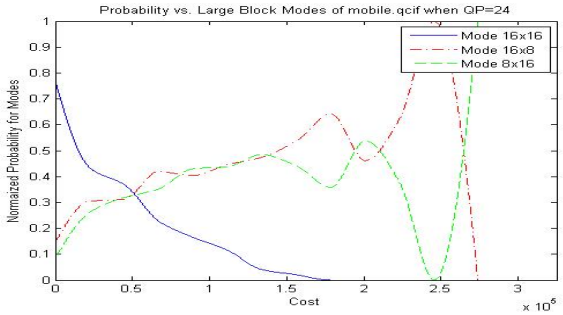


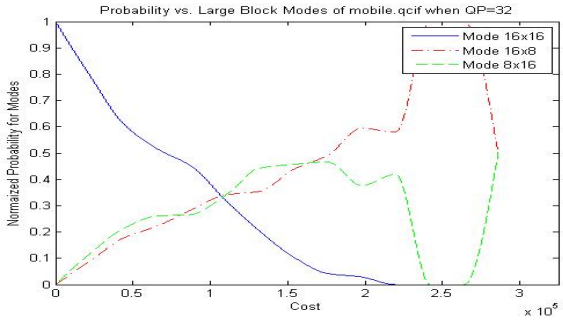
Fig. 2 Probability Characteristic Curve of Small Modes

To determine the best mode from the small ones Hadamard $SAD_{s8\times8}$ cost is calculated first for every sub- block and the probability of occurrence of each small mode to the value is found. Similar process can be applied to select the best mode. As there are four modes, three thresholds values are required to distinguish them. Characteristic curve for small modes of "Foreman" QCIF video sequence is shown in Fig. 2. Finally,

this algorithm is combined with SKIP mode early termination process to make it even faster.



(a)



(b)

Fig. 3 Probability characteristic curve of Large modes for News QCIF Sequence at QP = 24, 32

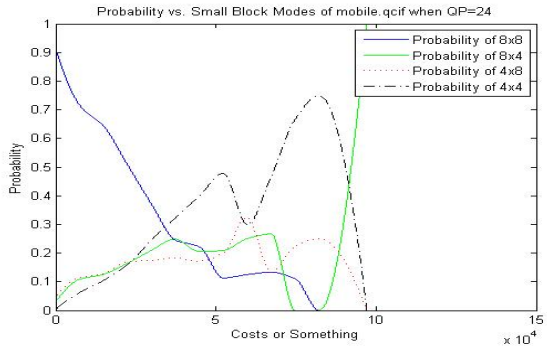
C. *Thresholds determination for each mode*

There are seven inter modes of different sizes and shapes including 16×16, 16×8, 8×16, 8×8, 8×4, 4×8 and 4×4 supported by H.264/AVC. In this paper, these seven block sizes are classified into two categories i.e. Large Blocks (16×16, 16×8, and 8×16) and Small Blocks (8×8, 8×4, 4×8 and 4×4). At first, a MB is taken and its Hadamard $SAD_{16\times16}$ Transform cost calculated. Then the characteristic curve of the probability of large modes vs. Hadamard $SAD_{16\times16}$ Transform cost of that MB is plotted. Finally, the characteristic curve of the probability of small modes vs.

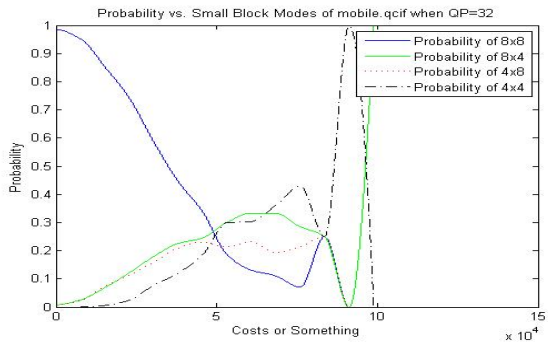
In this research, the probability of modes and vs. the initial MB cost characteristic curves of Mobile.qcif video sequence (100 frames) for different quantization parameter (QP=20, 24, 28, 32, 36, 40) are analyzed. Figure 3 and Figure 4 shows the characteristic changes between Large Modes and Small modes for different quantization parameters. From large modes probability characteristic curve, it has been observed that the range of domination of 16x16 mode increases as the quantization parameter increases. Similar scenarios are found from the small modes characteristic curves. Therefore, it is necessary to determine some thresholds for each mode to determine their range. But as the range for each mode varies with quantization parameter, the thresholds should be adaptive with quantization parameters.

D. *Parameters Selection and Proposed algorithm*

The proposed algorithm first checks whether the large or small modes are suitable for a MB. As there are 3 large blocks, therefore, two thresholds T2 and T3 have been introduced to distinguish them. 16×8 block is considered as the best mode if the initial cost is greater than T2. If the initial cost is greater than T3 then 8×16 mode is selected. Otherwise, 16x16 mode is considered as the best mode. Similarly, as there are four small modes, three thresholds T5, T6, T7 are introduced to distinguish among them. For a greater value than T5, T6, T7 8×4, 4×8 and 4×4 modes are selected as the best one respectively.



(a)



(b)

Fig. 4 Probability Characteristic Curve of Small Modes for News QCIF Sequence at QP = 24, 32

Otherwise, 8x8 mode is selected as the best sub-mode. The threshold for each mode is taken very carefully for different quantization parameters. Table 1 shows the thresholds for different video sequences for different QP obtained by observing and analyzing the mode vs. cost characteristic curves. To find optimum thresholds, polynomial curvature fitting theorem has been applied. Thus, optimized thresholds values of Foreman, News and Mobile video sequence for different QP's are obtained which are shown in Table 2. Therefore, it is possible to use the data from Table 2 to get the thresholds values for determining the best inter mode. Therefore, probability based fast inter mode algorithm can be represented with following steps:

Step 1: Initialize the thresholds T2, T3, T5, T6 and T7 for a specific QP value.

Step 2: Select a macroblock and determine its Hadamard $SAD_{16\times16}$ cost.

Step 3: Check if the MB is suitable for SKIP mode, if not go to Step 4, if yes select skip mode as the best mode.

Step 4: Divide the macroblock into 4 8×8 sub blocks and determine Hadamard $SAD_{s8\times8}$ cost for each block and calculate them to find Hadamard $SAD_{p8\times8}$ cost.

Step 5: If Hadamard $SAD_{16\times16}$ cost is less or equal to Hadamard $SAD_{p8\times8}$ cost choose large modes category, otherwise small modes category.

Step 6: If Large mode category is chosen, compare the Hadamard $SAD_{16\times16}$ cost with the T2 and T3 to determine the best Large mode (16×16, 16 ×8 and 8×16).

Step 7: If small mode category is chosen, compare the $SAD_{s8\times8}$ of a sub block with the T5, T6 and T7 to determine the best small mode (8×8, 8×4, 4×8 and 4×4).

Based on this algorithm it is possible to determine the best inter mode without calculating the rate distortion cost function.

TABLE I
THRESHOLDS T2, T3, T5, T6, T7 FOUND BY OBSERATION FROM THE SIMULATED PROBABILITY VERSUS 16x16 COSTS

Sequence		Large Blocks (Main Modes)		Small Block (Sub modes)		
News	**QP**	**T2**	**T3**	**T5**	**T6**	**T7**
	24	181500	52260	73690	21910	54710
	28	155900	62690	94940	26640	79920
	32	164100	112200	68650	39430	NAN
	36	187200	268300	114900	66410	NAN
Mobile	**QP**	**T2**	**T3**	**T5**	**T6**	**T7**
	24	194300	51060	91550	51320	33990
	28	212400	67450	39330	62330	41150
	32	164300	107200	48950	73110	83440
	36	201900	164900	67540	82680	87480

TABLE III
OPTIMIZED THRESHOLDS T2, T3, T5,T6,T7 FOUND USING POLYNOMIAL CURVATURE THEOREM ON THE DATA OF TABLE I

Sequence		Large Blocks (Main Modes)		Small Block (Sub modes)		
News	**QP**	**T2**	**T3**	**T5**	**T6**	**T7**
	24	181151	55550	81637	21776	54722
	28	154734	66891	107351	26384	79934
	32	161443	117202	87151	38989	NAN
	36	182092	273798	141460	65702	NAN
Mobile	**QP**	**T2**	**T3**	**T5**	**T6**	**T7**
	24	191714	50372	82839	50981	35040
	28	210707	66709	25576	61739	42468
	32	164256	106442	28391	72171	85050
	36	204472	164174	38072	81279	89405

IV. RESULTS AND DISCUSSION

Proposed algorithm was tested using the first 100 frames of different video sequences of QCIF format. Due to page limitation among them the results of "Foreman" that has medium motion changes with dominant luminance changes, "News" that has a low spatial details and changes in motion and "Mobile" that has a complex horizontal and vertical motion with slow zooming and panning are shown. The experiment is carried out in the JM 18.2 [12] encoder and the parameters selected as CABAC enabled; GOP structure is IPPPP; Maximum search range for motion estimation is 32; QP values are 20, 24, 28, 32, 36 and 40.

The thresholds determined by the experiment varied very slightly for different video sequences. Compared to the original H.264/ AVC video encoder, proposed algorithm has reduced the computational time significantly by around 60% to 65% for different QP values which are shown in Table 3. Though the bit rate increases and PSNR has reduced slightly, but the degradation of video quality for is very low and considered almost negligible. Therefore, the proposed algorithm is applicable for real time applications where reducing computational time is preferable compensating the video quality with negligible distortion.

REFERENCES

[1] I. E. G. Richardson, *"H.264 and MPEG-4 video compression; video coding for next generation multimedia"*. John Wiley and Sons, c2003.

[2] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/ AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.,* vol. 13, pp. 560-576, July 2003.

[3] X. Jing and L. P. Chau. "Fast approach for H.264 inter mode decision." *Electron, Lett*., vol. 40, no. 17, pp. 1051-1052, Aug. 2004.

[4] T. Y. Kuo and C. H. Chan, "Fast variable block size motion estimation for H.264 using likelihood and correlation of motion field". *IEEE Trans. Circuits Syst. Video Technol*., vol. 16, no. 10, pp. 1185-1195, Oct, 2006.

[5] Z. Y. Wei, K. N. Ngan, "A Fast Macroblock Mode Decision Algorithm for H.264". *Proc/ IEEE Signal Processing 2005*, vol. 1, pp. 243-246, Aug, 2005.

[6] Y. M. Lee, Y. T. Sun, and Y. Lin, "SATD- Based Intra Mode Decision for H.264/AVC Video Encoding", *IEEE, Trans. on Circuits and Systems for Video Technology*, vol 20, no. 3, pp. 463-469, Mar, 2010,.

[7] M. S. U. Yusuf and M. Ahmad, "Relative SAD- Based Inter Mode Decision Algorithm for H.264/ AVC Video Standard," *Journal of Multimedia,* vol. 6, No. 5, pp. 425-435, October 2011.

[8] I. Choi, J. Lee, B. Jeon, "Fast coding mode selection with rate-distortion optimization for MPEG-4 Part-10 AVC/ H.264". *IEEE Trans, Circuits Syst, Video Technol.*, vol. 16, pp, 1557-1561, July 2006.

[9] C.K Chiang and S.H Lai, "Fast H.264 Encoding Based on Statistical Learning," PCM 2010, Part II, LNCS 6298, pp. 179-189, 2010.

[10] Ma, W. , Yang, S. Gao L., Pei, C., Yan, S. , " Fast Mode Selection Scheme for H.264/ AVC Inter Prediction Bsed on Statistical Learning Method," IEEE, ICME (2009).

[11] Y. H. Kim, J. W. Yoo, S. W. Lee et al., "Adaptive mode Decision for H.264 encoder", Electron. Lett., Vol. 40, no. 19, pp. 1172-1173, September 2004.

[12] http://iphome.hhi.de/suehring/tml/

TABLE III
RATE DISTORTION AND ENCODING TIME REDUCTION OF PROPOSED ALGORITHM

Sequence	Performance Comparison						
Foreman	**QP**	**20**	**24**	**28**	**32**	**36**	**40**
	Bit Rate %	-8.7229	-9.57471	-8.50169	-5.77423	-2.4149	-0.31376
	PSNR -db	0.057	0.013	0.028	0.087	0.131	0.173
	Encoding Time (%)	63.70504	58.6266	11.28891	39.77372	40.02412	43.66658
News	**QP**	**20**	**24**	**28**	**32**	**36**	**40**
	Bit Rate %	-19.9206	-21.467	-22.3717	-6.05053	-4.81264	-3.58227
	PSNR -db	0.218	0.175	0.175	0.095	0.107	0.071
	Encoding Time (%)	51.71764	78.50289	82.57195	77.98118	74.19293	69.53723
Mobile	**QP**	**20**	**24**	**28**	**32**	**36**	**40**
	Bit Rate %	-2.60873	-3.85607	-5.71632	-6.72824	-5.76132	-3.69128
	PSNR -db	0.102	0.119	0.068	0.058	0.092	0.078
	Encoding Time (%)	58.34045	49.3016	48.91769	53.30804	54.98814	59.15103

Inclusion of Slack Bus in Newton Raphson Load Flow Study

S. M. Lutful Kabir[1]*, A. Hasib Chowdhury[2], Mosaddequr Rahman[3], Jahangir Alam[3]

[1]Institute of Info. and Comm. Tech., Bangladesh University of Engineering and Technology
[2]Department of EEE, Bangladesh University of Engineering and Technology
Dhaka-1205, Bangladesh
[3]Department of EEE, BRAC University
Dhaka-1212, Bangladesh
*lutfulkabir@iict.buet.ac.bd

Abstract - **Load flow solution is an inevitable tool for power system planning and operation. Among various methods of load flow solution techniques Newton Raphson (N-R) and its derivatives are the most popular and widely used ones. In Newton Raphson method one of the generator buses is treated as slack bus which is kept out of calculation during iterations. This is because of the fact that the slack bus is to carry entire loss of the system and total loss cannot be calculated before the end of iterations. The exclusion of slack bus from the load flow iteration restricts many analyses to make. Consideration of distributed slack bus, load flow for Microgrids, study on minimization of system losses etc. cannot be performed when the slack bus is excluded from the iterations. This paper reports the development of a method of including slack bus into the Jacobians for N-R load flow study. The method involves formulating the loss equation and then devising the formula for the terms to be included with the elements of the Jacobian matrix. The modified N-R method is applied to all of the buses and the losses are calculated in each iteration and included with the slack bus. The results obtained using the modified method are compared with those using conventional method. The agreement of the results confirms the accuracy of the developments.**

Index terms: **Inclusion of Slack bus, Jacobian modification, Microgrid load flow**

I. INTRODUCTION

The buses of a power system are divided into three categories in a load flow study. They are PV-, PQ- and Slack-bus [6]. Slack bus is also called infinite bus because theoretically the generation capacity at this bus is very large. It is treated as reference bus and with respect to it, all phase angle of the other bus voltages are calculated. In practice, in power flow solution, the slack bus is kept out of the iterative solution process and its power is calculated at the end of convergence of the iteration. Thus all losses of the network are burdened on this slack bus.

The new tools for planning and operating distribution systems have become essential with the increase of distributed generators (DGs) installed within power network called Microgrid network. Application of conventional power flow solution techniques in microgids using one slack bus is not possible [1]. This is because of the fact that the generating units are of small in capacity having comparable sizes and no one can be treated as slack generator of infinite capacity. So, the slack bus should be treated as normal PV bus. In other words slack bus should be included in the iterative process of load flow solution. Mohapatra [2] devised a method of eliminating burden of slack bus while maintaining the equal incremental loss criteria for optimum generation scheduling. Again, because of special characteristics of distribution systems, such as high R/X ratios and network imbalance, consideration of distributed slack bus is more realistic than a single slack bus [3, 4]. The previous work done for transmission system [5] cannot be applied directly in distribution system. Similar to the cases described, the consideration of keeping the slack bus outside the load flow study creates restrictions in different other studies.

This paper analyzes the system loss equation and it is found to be a function of bus voltage magnitudes and their angles. Since the elements of Jacobian in N-R method are the derivatives of real and reactive power with respect to voltage magnitudes and angles, hence with the inclusion of loss, it is necessary to modify the elements by adding terms involving the rate of change of loss with respect to the two variables. This paper devises a formulation for such inclusion. The method has been applied to solve the load flow considering slack bus to carry all losses and the results obtained from the new method are compared with those obtained using conventional method. The new method is also applied for the case of distributed losses.

II. MATHEMETICAL FORMULATION

For given loads in a system, the scheduled generations in the generator buses except the slack bus are known beforehand. The conventional load flow solution provides us the power to be maintained at the slack bus. In the developed method, at the beginning of iteration the generation power at the slack bus is assumed to be equal to total loads minus total generations in all other generating buses. It means that, at start, the losses are assumed to be zero and the generators are sufficient to supply the loads only. After each iteration, the total loss is calculated and included with the scheduled power of the slack bus. Simultaneously the rates of change of losses are included into the Jacobians in the iterative process.

A. The Power Equations

The apparent power injected at a bus *i*, is given by

$$S_i = P_i + jQ_i = V_i I_i^* = V_i \sum_{k=1}^{n} V_k^* Y_{bus_{ik}}^* \tag{1}$$

So, the real and reactive power is expressed as,

$$P_i = \mathrm{Re}\left[V_i^* \sum_{k=1}^{n} V_k Y_{bus_{ik}} \right] \tag{2}$$

$$Q_i = -\mathrm{Im}\left[V_i^* \sum_{k=1}^{n} V_k Y_{bus_{ik}} \right] \tag{3}$$

Since, we have considered that the slack bus will take the total loss of the system, so the loss calculated after each iteration is added with original scheduled power considered at

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

the beginning of the iteration. The modified scheduled real power should be equated with the injected real power, P_1 (assuming bus #1 as slack bus) obtained from equation (2). In other words, the original scheduled real power at slack bus can be equated with the injected power obtained from equation (2) minus the losses calculated after each iteration. The modified injected power may be termed as modified injected power at the slack bus. The modified injected power at bus #1 is given by,

$$P_1^{'} = P_1 - loss \tag{4}$$

By combining equations (2) and (4) we obtain,

$$P_1^{'} = \mathrm{Re}\left[V_1^{*} \sum_{ik=1}^{n} V_k Y_{bus_{ik}} \right] - loss \tag{5}$$

The total loss is summation of losses of all branches of the system. It is a function of voltages of the buses.

B. The Loss Equation

The total loss can be expressed as,

$$loss = \sum_{i=1}^{n-1} \sum_{k=i+1}^{n} \left| (V_i - V_k)^2 Y_{bus_{ik}}^2 \ \mathrm{Re}\left[\frac{-1}{Y_{bus_{ik}}} \right] \right|$$

$$= \sum_{i=1}^{n-1} \sum_{k=i+1}^{n} \left| (V_i - V_k)^2 C_{ik} \right| \tag{6}$$

$$where, C_{ik} = Y_{bus_{ik}}^2 \ \mathrm{Re}\left[\frac{-1}{Y_{bus_{ik}}} \right] \tag{7}$$

C. The Modification in Jacobian

Now in the Newton-Raphson method [2], the elements of the Jacobian corresponding to ΔP_1 is given by,

$$\frac{\partial P_1}{\partial \delta_1} = \frac{\partial}{\partial \delta_1} \left\{ \mathrm{Re}\left[V_1^{*} \sum_{k=1}^{n} V_k Y_{bus_{ik}} \right] \right\} - \frac{\partial loss}{\partial \delta_1} \tag{8}$$

$$\frac{\partial P_1}{\partial |V|_1} = \frac{\partial}{\partial |V|_1} \left\{ \mathrm{Re}\left[V_1^{*} \sum_{k=1}^{n} V_k Y_{bus_{ik}} \right] \right\} - \frac{\partial loss}{\partial |V|_1} \tag{9}$$

< Conventional term > < new term >

The derivation of the conventional terms can be found in ref. [7].

The new terms in equations (8) and (9) can be obtained by differentiating equation (6) with respect to δ_1 and $|V|_1$ respectively. They are,

$$\frac{\partial loss}{\partial \delta_1} = 2 \sum_{\substack{k=1 \\ k \neq i}}^{n} C_{1k} (V_1 - V_k) j V_1 \tag{10}$$

$$\frac{\partial loss}{\partial |V|_1} = 2 \sum_{\substack{k=1 \\ k \neq i}}^{n} C_{1k} (V_1 - V_k) e^{j\delta_1} \tag{11}$$

Therefore, the elements of the Jacobian [J] given by equation (8) and (9) are updated by including the new terms from equations (10) and (11).

D. Shifting of Reference

Since all busses are included in the iteration process, so the angles of all buses, including the slack bus, are subjected to change. Therefore after each iteration, all bus voltage is rotated in the opposite direction with the angle of bus #1 so that it always remains as reference. This is done by multiplying each voltage vector by $e^{-j\delta_1}$.

III. THE TEST SYSTEMS

A 6-bus system [8] is selected for the validation of the new method. Single line diagram of the system is shown in Figure 1. It consists of three generators and three loads. Apart from the 6-bus system, The developed method has also been applied to a 30-bus and a 57-bus IEEE test systems [9]. Salient information on the three systems is exhibited in Table I. The line parameters and other system data of the three systems are available in [8,9].

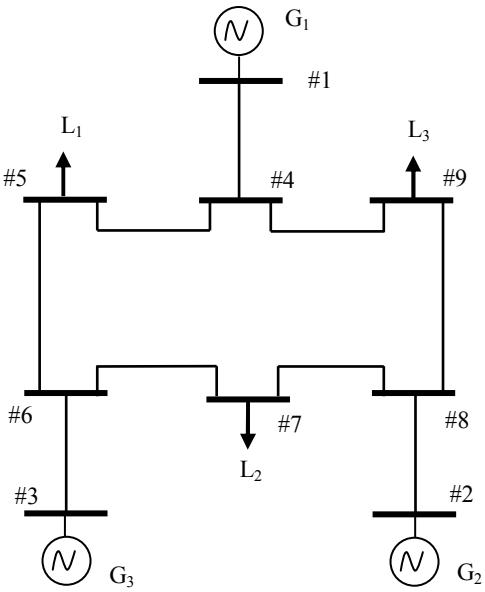


Fig. 1: 6-bus test system

TABLE I
DATA ON SYSTEMS STUDIED

Items	6-bus	30-bus	57-bus
Scheduled Gen. at Slack bus (at the beginning of iteration), MW	100	23.53	450.8
Scheduled generation at other gen. buses, MW	110	165.67	800.0
Total load, MW	210	189.3	1250.8
No. of genenrator bus	3	6	4
Total no. of bus	6	30	57

IV. RESULTS

Matpower [10], an open source MATLAB program for load flow studies, have been modified to implement the method developed and described in Section II. The original program implements conventional Newton-Raphson method for load flow solutions.

The validity of the formulation is established by applying the proposed method and the conventional N-R method for the load flow solutions of the test systems mentioned in section III and comparing results. While applying the proposed formulation two different case has been considered, (i) losses included in the slack bus generation only and (ii) losses distributed among all the generator buses.

A. Validation of the Developed Method

In the modified program the slack bus is included into the Jacobian and the program is run for the 6-bus system. It

converged after three iterations. The conventional N-R program is then run for the 6-bus system having the same slack bus. The conventional method also took three iterations to converge. The slack bus power after convergence is found to be 107.88 MW. It is found that the losses are exactly the same for both the methods. Also, voltage magnitudes and angles obtained by the two methods matched exactly.

In the modified program, the losses were calculated in each iteration and subtracted from the injected power to compare with the scheduled power of the slack bus. The losses obtained in different iterations are tabulated in Table II.

TABLE II
CONVERGENCE OF LOSS FOR 6-BUS SYSTEM

Iteration count	Losses, MW	Net Power at bus 1, MW
1	7.55	107.55
2	7.87	107.87
3	7.88	107.88

Similar studies have been carried out for the 30-bus and the 57-bus IEEE systems. The convergences of losses are shown in Table III and Table IV respectively. In both cases the numbers of iterations were also found to be the same for the conventional and the modified methods.

For all systems, it is observed that loss figures when calculated during iterative process converge earlier than the convergence of real and reactive powers. It indicates that the inclusion of loss and its effect on Jacobian is correct.

TABLE III
CONVERGENCE OF LOSS FOR 30-BUS SYSTEM

Iteration count	Losses, MW	Modified Scheduled Power at bus 1, MW
1	4.67	28.20
2	5.52	29.05
3	5.53	29.06
4	5.53	29.06

TABLE IV
CONVERGENCE OF LOSS FOR 57-BUS SYSTEM

Iteration count	Losses, MW	Modified Scheduled Power at bus 1, MW
1	27.84	478.64
2	27.86	478.66
3	27.86	478.66

B. Loss Distributed Over All Generator Buses

The capacities of the generators in microgrid are comparable and hence it may not be possible to impose all losses on a single generator. The developed method has been applied in 57-bus system and the total losses have been distributed over four generator buses in proportion to their capacities. The result is presented in Table V. The table compares losses for the two case, i.e. losses included in the slack bus generation only and losses distributed to all generator buses. It is seen from the result that the loss is reduced when it is distributed over different generation buses. This is a more accurate result. This type of study is now possible with the developed method.

TABLE V
COMPARISON BETWEEN NON-DISTRIBUTED AND DISTRIBUTED LOSS

Gen. Bus	Losses applied to slack bus		Losses distributed among gen. buses	
	Losses, MW	Gen. power, MW	Loss, MW	Gen. power, MW
1	27.86	478.36	9.64	459.64
3	0.0	40.0	0.86	40.86
8	0.0	450.0	9.63	459.63
12	0.0	310.0	6.63	316.63
Total	27.86	1278.36	26.76	1277.26

V. CONCLUSION

A method of including the slack bus during iterative process of Newton Raphson method has been presented. It is based on inclusion of line losses at every iteration. The effect of inclusion of loss has an impact on Jacobian elements. The mathematical formulation of this impact has been analyzed and the expressions of the variation of loss with voltage magnitudes and their angles have been presented in the paper. The validity of the formulation has been applied in three test systems. The agreement between the results of the new method with the conventional one proved the accuracy of the development. The new technique is also applied to a case of loss distributed among different generator buses. The ease of new method has opened up a new avenue to explore numerous studies which are not possible by the conventional method.

ACKNOWLEDGEMENT

Author[1] likes to express his thanks to Bangladesh University of Engineering and Technology for approving sabbatical leave during which the research was carried out. Thanks also to BRAC University for allowing the use of their facility in order to carry out the research.

REFERENCES

[1] R. M. Kamel, A. Chaouci, K. Nagasaka, : "Detailed Analysis of Micro-Grid Stability during Islanding Mode under Different Load Conditions", online journal www.SciRP.org, May 2011.
[2] A. Mohapatra,:"Distributed Slack Bus Algorithm for Economic Load Dispatch", An undergraduate thesis, NIT, Roukela, India, 2006.
[3] Shiqiong Tong, Miu K.N. : "Participation Factor Studies for Distributed Slack Bus Models in Three-Phase Distribution Power Flow Analysis", in Transmission and Distribution Conference and Exhibition, 2005/2006, IEEE PES, 21-24 May 2006, p. 92.
[4] P. H. Haley and M. Ayres : "Super Decoupled Load Flow With Distributed Slack Bus", IEEE transactions on Power Apparatus and Systems, Vol. PAS-104, pp 104-113, 1985.
[5] Chayakulkheeree, K. : "Application of Distributed Slack Bus Power Flow to Competitive Environments", in Australian Universities Power Engineering Conference, (AUPEC), Dec2007, p 1 .
[6] J. J. Grainger and W. D. Stevenson, : "Power System Analysis", Tata-McGraw Hill Education, 2003
[7] Hadi Saadat, : "Power System Analysis", McGraw-Hill Company, 2004.
[8] A. J. Wood and B. F. Woolenberg, "Power Generation, Operation and Control", John Wiley & Sons, Inc., 1996.
[9] Resources : "Power System Test Case Archive",[online] Available: www.ee.washington/-research/pstca/
[10] R. D. Zimmerman, C.E. Murillo-Sanchezand R. J. Thomas : "MATPOWER: Steady-State Operations, Planning, and Analysis Tools for Power Systems Research and Education", IEEE Transaction on Power Systems, vol. 26, Issue 1, pp 12-19, Feb. 2011.

Analytical Approach to Investigate the Loadability of a Long Transmission Line for Simultaneous AC-DC Power Flow

Mohammad Tawhidul Alam,[1,*] and Q. Ahsan[2]

[1]Department of Electrical and Electronic Engineering
Ahsanullah University of Science and Technology, Dhaka, Bangladesh
[2]Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology, Dhaka, Bangladesh
[*]*Email: tawhidul_bclt@yahoo.com*

***Abstract*—The simultaneous ac-dc power transmission system can increase the load carrying capability of a long transmission line. A mathematical model is developed in this paper for simultaneous ac-dc power transmission system to analyze the behavior of a composite system in comparison with the original ac system. The load carrying capability of an ac-dc composite system is compared with that of original ac system through numerical solution in this paper.**

***Index Terms*— AC-DC composite system, loadability, zigzag transformer, thermal limit.**

NOMENCLATURE

P_l	Steady state power flow through unconverted ac line in p.u
E_g	Internal generated voltage of the generator in p.u
E_r	Receiving end voltage in p.u
X	Transfer reactance of unconverted ac system in p.u
I_l	Current through unconverted ac line in p.u
I_{th}	Thermal limit of the conductor
I_L	Steady state current of unconverted line
δ	Load angle of the generator for unconverted system
V_{acon}	Per phase ac voltage of converted line
V_{dc}	DC voltage component in each phase.
P_{comb}	Combined steady state ac-dc power flow
P_{ac}	Steady state ac power in ac-dc composite system
P_{dc}	Steady state dc power in ac-dc composite system
I_{dc}	DC component of composite system line current
I_{AC}	AC component of composite system line current

I. INTRODUCTION

A long EHV ac transmission line has an upper limit of loading dictated by stability consideration and this line is not loaded to its thermal limit to keep sufficient margin for transient stability. Flexible AC Transmission System (FACTS) is used in existing ac transmission line to improve load carrying capability along with the overall system performance [1-2]. Series compensation with mechanical switch can be other option for increasing the loadability of existing transmission line [3]. Now a days, for power flow and stability improvement HVDC line is operating in parallel with HVAC line [4].

Simultaneous ac-dc transmission is a new approach for the increase of power flow through the existing line proposed by H.Rahman and B.H.Khan [5]. The simultaneous ac-dc power flow derives the benefit of parallel HVDC line without constructing it. In this system of power transmission existing ac line conductors are used to carry ac power along with dc power simultaneously. The proponents of this new approach conducted the numerical analysis considering only transmission line, disregarding the whole system from generator to infinite bus. A feasibility study is performed for small power tapping from simultaneous ac-dc transmission line in [6]. Simulation results of this paper clearly indicate that the tapping of a small amount of ac power from the composite ac–dc transmission line has a negligible impact on the normal functioning of the composite ac–dc power transmission system. To date, all the works those have been done on simultaneous ac-dc power transmission are based on numerical solution. From the numerical result it is difficult to formulate a generalized comment. However, to investigate the behavior of a system comprehensively an appropriate mathematical model is a necessity. Moreover, the change of which parameter of the system impact how much on its performance by using the mathematical model someone can evaluate easily.

In this paper, a comprehensive analysis is presented on simultaneous ac-dc system by developing an appropriate mathematical model. The model is used to investigate the load carrying capability of an existing long ac transmission line in case of simultaneous ac-dc power flow. A numerical analysis also performed for the validation of developed mathematical model.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

II. Simultaneous AC-DC Power Flow

In simultaneous ac-dc power transmission system same conductor is used to transmit ac power along with dc power, the schematic diagram of simultaneous ac-dc flow is shown in fig.1. The main concern of this paper is to convert existing long EHV ac line into ac-dc composite line. According to the figure, a remote generating unit is feeding to an infinite bus through a long transmission line. A portion of ac power, taking from the generator bus, converted into dc and feeding to the existing ac transmission line through the neutral point of zigzag transformer.

At the receiving end, dc portion of power is taking away from the neutral point of zigzag transformer and converted into ac before feeding to the infinite bus. Zigzag connection is used at both ends of the transmission line to avoid magnetic saturation in the transformer core due to dc current.

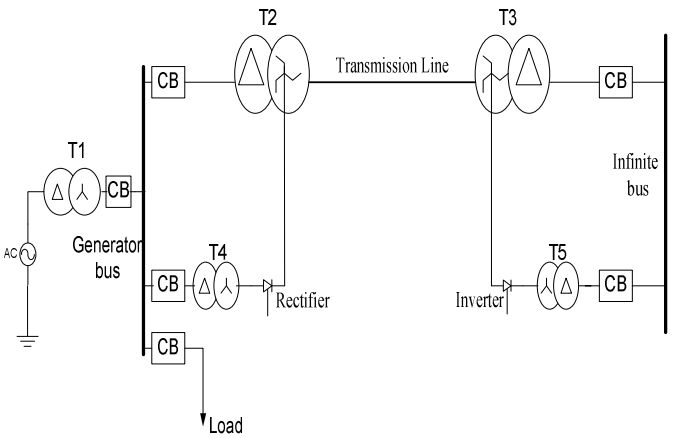


Fig. 1 AC-DC composite power transmission system

In converting ac line into ac-dc composite line the insulator remain unchanged and thus the maximum value of the phase voltages must be kept same. The output dc current of rectifier is equally divided among all the three phases and each conductor carries ac-dc modulated current wave. A communication link is needed between rectifier and inverter for better control of dc power flow.

III. Loadability Model

A simple ac transmission system is presented in fig. 2. In case of ac transmission line the power flow equation is given as follows.

$$P_l = \frac{E_g E_r}{X} sin\delta \quad (1)$$

$$I_l = \frac{E_g\angle\delta - E_r\angle 0}{jX} \quad (2)$$

$$X = X_{te} + X_{Tr} \quad (3)$$

Here,

X_{te} = Reactance that includes the reactance of generator and transformer in p.u.

X_{Tr} = Reactance of the transmission line in p.u.

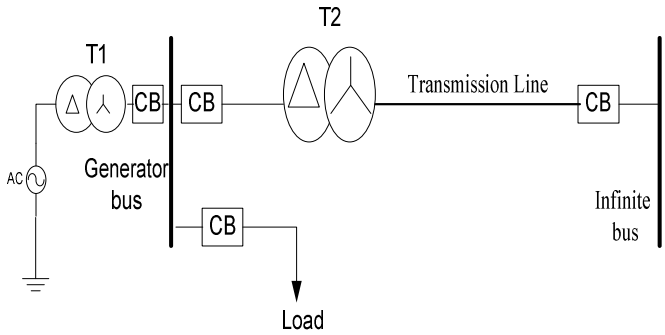


Fig. 2 A.C power transmission system

For a long ac transmission line minimum steady state stability margin is recommended as 30% and steady state power transfer angle would be 40^0 – 44^0 [7]. Although the thermal limit is qualified as line design and it depends on conductor selection but according to the loadability characteristics of a long transmission line the thermal limit is much higher than the steady state stability limit.

Let α is the ratio between steady state current and thermal limit of the conductor.

$$I_L = \alpha\, I_{Th} \quad (4)$$

As the mechanical infrastructure (i.e. insulating material, clearing distance and creepage distance) of composite line remain unchanged the maximum value of the phase voltage in converted line is to keep same as unconverted line. The sum of the dc voltage and per phase maximum ac voltage is equal to the maximum value of the unconverted line phase voltage but in this case maximum value of ac voltage must be greater than or at least equal to the dc voltage so that electric field changes its direction in each cycle, as the line insulators are not changing. Let k is a factor that quantifies the ac voltage level and 1-k quantifies dc voltage level. The value of k varies between 1 and 0.5 i.e.$0.5 \le k < 1$.

$$V_{\emptyset m} = \frac{\sqrt{2}}{\sqrt{3}} V_L \quad (5)$$

$$V_{acon} = \frac{1}{\sqrt{2}} k V_{\emptyset m} \quad (6)$$

$$V_{dc} = (1-k) V_{\emptyset m} \quad (7)$$

Here,

$V_{\emptyset m}$ = maximum phase voltage for unconverted ac system

V_L = Line voltage for unconverted ac system

V_{acon} = AC phase voltage for composite system

V_{dc} = DC voltage in each phase for composite system

Combined power flow is the summation of ac power and dc power in composite system.

$$P_{comb} = P_{ac} + P_{dc} \quad (8)$$

$$P_{ac} = P_{acmax} \sin\delta_{ac} = \frac{E_g E_r}{X'} \sin\delta_{ac} \quad (9)$$

δ_{ac} is the transmission angle and X' is the transfer reactance from generator to the infinite bus for ac power transmission in composite line.

Let β is the ratio between composite line ac current I_{AC} and unconverted line ac current I_L, where the expression of β can be found as

$$\beta = \frac{X}{kX'} \left| \frac{E_g\angle\delta_{ac} - E_r\angle 0}{E_g\angle\delta - E_r\angle 0} \right| \quad (10)$$

The expression of β can further be simplified by using the approximated current expression presented in [5].

$$\beta = \frac{X}{kX'} \left(\frac{\sin\delta_{ac}}{\sin\delta} \right) \quad (11)$$

Now the ac power flow equation (9) can be expressed in a simple form in terms of unconverted ac power.

$$P_{ac} = k\beta \frac{E_g E_r \sin\delta}{X}$$
$$= k\beta P_l \quad (12)$$

In case of ac-dc system the line can be loaded up to the thermal limit and it is assume that the line voltage at the sending and receiving end would be the same, so the dc current and power at this loading are given as ,

$$I_{dc} = \sqrt{I_{Th}^2 - I_{AC}^2} \quad (13)$$

$$I_{dc} = I_L \sqrt{\frac{1-\alpha^2\beta^2}{\alpha^2}}$$
$$= I_L\,\gamma \quad (14)$$

Here, $\gamma = \sqrt{\frac{1-\alpha^2\beta^2}{\alpha^2}}$

$$P_{dc} = 3V_{dc}I_{dc}$$
$$= \frac{3\sqrt{2}}{\sqrt{3}}(1-k)V_L\,I_L\,\gamma$$
$$= \frac{3\sqrt{2}}{\sqrt{3}\sqrt{3}}(1-k)\gamma\left(\frac{E_g E_r \sin\delta}{X}\right)$$
$$= \sqrt{2}\,(1-k)\gamma P_l \quad (15)$$

Now the combined power in terms of unconverted ac power is given by (16)

$$P_{comb} = [k\beta + \sqrt{2}(1-k)\gamma]P_l \quad (16)$$

The equation (16) clearly reveals that the combined ac-dc power depends on the values *k,* β and γ respectively. The value of *k* has huge impact on transfer reactance of converted ac system that is shown below.

$$X' = X_{te} + X'_{Tr}$$
$$= X_{te} + \frac{X_{Tr}}{k^2}$$
$$= X_{te} + mX_{Tr} \quad (17)$$

where, $m = \frac{1}{k^2}$

In case of long transmission line $X_{Tr} >> X_{te}$. Therefore, in case of ac-dc composite system with very much lower value of *k* the equivalent terminal system reactance would be negligibly small compare to the converted line equivalent reactance and considering the equation (3) and (17) the following approximation can be made.

$$X' \approx mX \quad (18)$$

The combined power flow equation (16) can be expressed in a simple form as follows

$$P_{comb} = AP_l \quad (19)$$

Here, $A = k\beta + \sqrt{2}(1-k)\gamma$

To get the combined power flow higher than the unconverted ac power the value of *A* must be greater than 1. i.e. $A > 1$.

For example, if the value of *k* is 0.5 that gives 50% ac voltage and 50% dc voltage mix and if the value of α is considered 0.5 the value of *A* would be 1.62 which means 62% improvement in combined power flow is obtained. In this case, same power transfer angle is considered for ac power transmission in composite ac-dc system and unconverted ac system which is 44^0 [7]. Considering some reasonable values of *k* and *α* the Loadability Improvement(LI) is presented in Table I.

TABLE I
LOADABILITY IMPROVEMENT FOR DIFFERENT AC AND DC VOLTAGE MIX

k	$\alpha = \frac{I_L}{I_{th}}$	P_{ac}	P_{dc}	P_{comb}	LI (%)
0.5	0.50	$0.25P_l$	$1.37P_l$	$1.62\ P_l$	62
	0.45	$0.25P_l$	$1.53P_l$	$1.78\ P_l$	78
	0.40	$0.25\ P_l$	$1.73\ P_l$	$1.98\ P_l$	98
0.52	0.50	$0.27P_l$	$1.31\ P_l$	$1.58\ P_l$	58
	0.45	$0.27P_l$	$1.47\ P_l$	$1.74\ P_l$	74
	0.40	$0.27\ P_l$	$1.66\ P_l$	$1.93\ P_l$	93
0.54	0.50	$0.29\ P_l$	$1.25\ P_l$	$1.54\ P_l$	54
	0.45	$0.29\ P_l$	$1.40\ P_l$	$1.69\ P_l$	69
	0.40	$0.29\ P_l$	$1.59\ P_l$	$1.88\ P_l$	88
0.56	0.50	$0.31P_l$	$1.19P_l$	$1.50\ P_l$	50
	0.45	$0.31P_l$	$1.39P_l$	$1.70P_l$	70
	0.40	$0.31\ P_l$	$1.51\ P_l$	$1.82\ P_l$	82
0.58	0.50	$0.33P_l$	$1.13\ P_l$	$1.46\ P_l$	46
	0.45	$0..33P_l$	$1.27\ P_l$	$1.60\ P_l$	60
	0.40	$0..33P_l$	$1.44\ P_l$	$1.77\ P_l$	77

Table I clearly shows that for a particular value of *α* if the value of *k* is increased the load carrying capability of the composite system will decrease. The increase in *k* represents increase in ac voltage mix and decrease in dc voltage mix. The value of α also has a significant impact on LI. For a particular value of *k* smaller decrease in α cause greater increase in LI.

IV. NUMERICAL ANALYSIS WITH SIMULATION RESULT

For the numerical simulation a power system is considered which presented in appendix. In case of unconverted ac line, considering 30% steady state stability margin [7], following results are found after simulation.

Steady state loading = 436MW
Steady state power transmission angle = 44.44^0
Line current = 0.7574kA
Value of α = 0.4208

Now, ac-dc composite system is analyzed through developed mathematical model and the results are compared with those obtained from the numerical simulation using MATLAB. The improvement in load carrying capability, shown in Table II, obtained using developed mathematical model is very much close to those observed from the numerical simulation. A small percentage of error is observed in loadability improvement with the developed model. This small percentage of error is due to some logical assumptions that are considered to make the model simple.

TABLE II
COMPARISON OF LOADABILITY IMPROVEMENT BETWEEN DEVELOPED MODEL AND NUMERICAL SIMULATION

$\alpha = \frac{I_L}{I_{th}}$	k	Loadability Improvement (LI)		% of error
		With developed model (%)	With numerical simulation(%)	
0.4208	0.50	95.37	101.49	6.03
	0.52	90.96	96.82	6.05
	0.54	86.62	92.21	6.06
	0.56	82.36	87.69	6.07
	0.58	78.17	83.25	6.10
	0.60	74.05	78.88	6.12

Again the ac-dc composite system is analysed, using developed mathematical model, to investigate the impact of different dc voltage mix on loadability improvement. The change of dc voltage in composite system has huge impact on steady state loading that shown in fig. 3

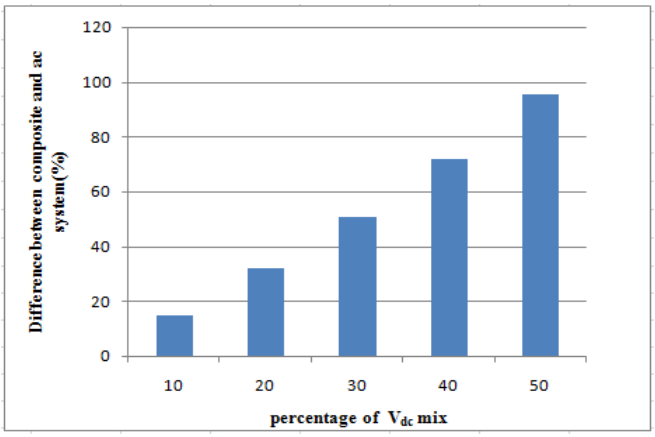


Fig. 3 Loadability improvement in ac-dc composite system

It is clearly observed from the figure that the loadability improvement increases with the increase in the value of dc voltage mix and obtained improvements for this system varies from 14.9% to 95.6% with the dc voltage mix variation from 10% to 50% respectively.

V. CONCLUSION

Conversion of a long EHV transmission line into simultaneous ac-dc composite system can give an improvement in load carrying capability of the line. Moreover, ac-dc composite power transfer system provides the benefit of ac-dc parallel transmission system without constructing separate line. The mathematical model developed in this paper for simultaneous ac-dc power flow through the existing line clearly shows, without numerical simulation, that the loadability of existing line increase with the increase of dc mix.

APPENDIX

A sample power system is considered for the simulation study with numerical data

A. *Unconverted AC line configuration*

Base power=550MVA

Line voltages are 24kV, 132kV, 345kV for the generator, generator bus and transmission line respectively.

Ratings of different components are as follows:

Two identical generators each has the following rating 550MVA, 24kV, 50Hz, H=3.5sec, 30% transient reactance.

Transformer T_1&T_2(Between generator and generator bus) each of 550MVA, 24/132 kV, 50Hz, 10% reactance.

Transformer T_3 : 1100MVA,132/345 kV,50Hz, 8% reactance.(At the sending end of the line)

Transformer T_4 : 1100MVA,345/132 kV,50Hz, 8% reactance.(At the receiving end of the line)

Transmission line: ACSR(Moose; commercial name) twin bundle conductor, 345kV, 50Hz, 400km single circuit line has following parameters

X_{Tr} = 0.33086 Ω/km/phase, R = 0.03252 Ω/km/phase

$Y = 3.33797*10^{-6}$ S/km/phase

current carrying capacity of each sub conductor is 0.9kA.

I_{th}(thermal limit)=1.8kA,

B. *Composite AC-DC configuration*

Transformer T_3 = 550MVA, 132/172.5kV, 50Hz, 8% reactance.

Transformer T_4 = 550MVA,172.5/132kV, 50Hz,8% reactance.

REFERENCES

[1] R. S. Wibowo, N. Yorino, M. Eghbal, Y. Zoka, and Y. Sasaki, "FACTS Devices Allocation With Control Coordination Considering Congestion Relief and Voltage Stability" *IEEE Transaction on Power System, Vol. 26 ,No. 4, November 2011.*

[2] F. A. R. Al Jowder and B.T. Ooi "Series Compensation of Radial Power System by a Combination of SSSC and Dielectric Capacitors" *IEEE Transaction on Power Delivery, Vol. 20 ,No. 14, January 2005*

[3] E. A. Leonidaki, D. P. Georgiadis, and N. D. Hatziargyriou, "Decision Trees for Determination of Optimal Location and Rate of Series Compensation to Increase Power System Loading Margin" *IEEE Transaction on Power System, Vol. 21 ,No. 3, August 2006.*

[4] A.E. Hammad "Stability and Control of HVDC and AC Transmissions in Parallel" *IEEE Transaction on Power Delivery, Vol. 14, No.4, October 1999.*

[5] H. Rahman and B. H. Khan "Power Upgrading of Transmission Line by combining AC-DC Transmission" *IEEE Transaction on Power System, Vol. 22 ,No. 1, February 2007.*

[6] H. Rahman and B. H. Khan, "Possibility of Power Tapping From Composite AC–DC Power Transmission Lines" *IEEE Transaction on Power Delivery, Vol. 23, No. 3, July 2008.*

[7] R.D. Dunlop, R. Gutman, P.P. Marchenko "Analytical development of Loadability Characteristics for EHV and UHV Transmission Line" *IEEE Transaction on Power Apparatus and Systems, Vol. PAS-98, No. 2 March/April 1979.*

A Study On Low Grid Voltage Problem Near Rooppur Nuclear Power Plant

A. Hasib Chowdhury, Md. Khurram Monir Rabby*
Dept. of EEE, Bangladesh University of Engineering and Technology
Dhaka-1205, Bangladesh
*khurram.rabby@gmail.com

Abstract— **Faced with the need of achieving the highest possible efficiency in the utilization of primary resources, nuclear power plants (NPPs) are considered as a viable alternative or addition. In case of transmission system voltage interruption, especially low voltage, NPPs have historically been susceptible to stations trip. Because of low voltage problem, the resulting loss in real and reactive power support can embitter transmission events. In this paper, the possibility of low bus voltage in the Western grid of the Bangladesh power system network, where the planned Rooppur NPP will be sited, is calculated and reported. It has been observed that even with small load factor, buses around the Rooppur NPP experience low voltage. A detailed investigation to study the problem and possible solutions is thus recommended.**

Index Terms—**Nuclear power plant, low voltage, loading factor, Bangladesh power system network**

I. Introduction

There are two basic characteristics of power system for identifying quality of supply and having great influence on operability of safety related electrical equipment. Those two characteristics are system voltage and system frequency [1].Problems of operating a NPP within an electric grid of limited capacity or of low voltage have long been a serious concern because the grid has direct bearing on the starting and running of the NPP [1]. To avoid the low voltage problem the design of the power system should be according to the NRC [2]. Similar requirements can be found in IAEA [3]. Moreover NRC's design criteria ensure that any mode of plant operation shall not degrade the performance of any safety system load below an acceptable level [4]. According to [5], emergency power supply (EPS) has the ability to detect and disconnect from its power source in event of degradation (over- and under-voltage, over- and under-frequency), and have possibility to connect directly to the alternative source. Limitation regarding degradation of voltage in power system depending issues are described in [6-10]. In [11], voltage stability, power transfer limits, reactive reserves, methods of improving voltage stability, load conservation procedures and effects of deregulation on nuclear power generating stations have been described. For the voltage case consideration, the phenomenon of voltage sag on the safety bus due to various transients as electrical operating modes is simulated and analyzed in [12-15].

The permitted deviations of NPP in standardized design may differ from one supplier to another and must be carefully verified by the prospective owner of the NPP with regard to the currently prevailing conditions of the grid. Typical IEEE approved supplier-guaranteed limits for correct operation at start-up conditions may include: high-voltage motors: down to 75% of nominal voltage, low-voltage motors: down to 70% of nominal voltage and valve drives: down to 80% of nominal voltage. Short-time voltage drops can be tolerated if high-capacity motors are started. For bus-bars of diesel back-up generators, a drop of 15-20% is generally IEEE acceptable. However, continuous operation of large-capacity fully loaded motors at such levels is not recommended. Off-nominal voltage conditions of prolonged duration can be compensated by on-load tap changers; usually, these compensate for the following deviations from nominal values:± 10% for normal transformers and ±15% for stand by grid transformers.

This paper presents a continuous power flow based approach to predict the most probable low voltage buses close to the Rooppur nuclear power plant buses in Western grid of the Bangladesh power system network.

II. Cases of Low Voltage

Off-nominal voltage conditions at various points of the grid are mainly due to the inability of the grid protection system to quickly clear transmission line faults, long transmission lines without intermediate power stations, lack of voltage control equipment, such as synchronous condenser and static capacitors, absence of on-load tap-changing transformers and insufficient capacity of the transmission lines to carry the peak load. The other types of low voltage reasons are described below.

A. Sharp Drop In Voltage

A rapid drop in voltage (voltage dip) is mainly due to an electric fault on a transmission line. The extent of the dip depends upon the proximity and nature of the fault, and upon the sensitivity of the automatic voltage regulator (AVR) equipment of the generators connected in the grid. The duration of the dip depends on the speed of the protective relays and their co-ordination as well as on the circuit breaker characteristics and their speed of operation. During sharp voltage dip conditions, all connected motors will be retarded. The magnitude of the retardation is determined by the voltage dip and its duration, the characteristics of the load, and the mass moment of inertia of the motor-pump assembly.

B. Inadequacy Of Voltage Control Equipment

When the voltage control equipment is inadequate, it may not be possible to maintain the same voltage at different points of the grid at all times, and poor voltage conditions may develop and persist over extended periods. Prolonged low voltage on the grid may create problems in starting and running large-capacity motors of the NPP. Some motors are provided with flywheels and have a relatively long start-up time.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

C. Voltage Constraints

It is normal for grid voltage to sway, as grid voltage is dependent of total system production and consumption of reactive power, and in the case of absence of the feeding transformer tap changer voltage present at the safety busbars will also change. There are two distinctive voltage events that can occur with different causes and effects, and are therefore separately considered.

D. Undervoltage/Degraded Voltage

Undervoltage state at safety busbars can be result of following events:

Degradation of high voltage (grid voltage) - Either imbalance of production/consumption of reactive power within system or loss of one or more phases can cause voltage to decrease significantly and have negative impact on the electrical safety related equipment of a NPP.

Consequences of degraded voltage condition are loss of multiple electrical components, trip of motors due to thermal protection actuation (Thermal overload) and inability of electrical (such as auxiliary relays) components to actuate or stay actuated, thus limiting the ability of the safety (and non-safety) related equipment to operate and perform its intended function.

III. System Description and Modelling

Bangladesh power system network (BPSN) is mainly divided into five operational regions. They are Dhaka, Central, Southern, Northern and Western region. Dhaka region has the largest bulk of generation. It exports power to the western part. The Eastern Grid is formed by Dhaka, Central and Southern regions, and Northern (Rajshahi-Rangpur) and Western (Khulna- Barisal) region form the Western Grid.

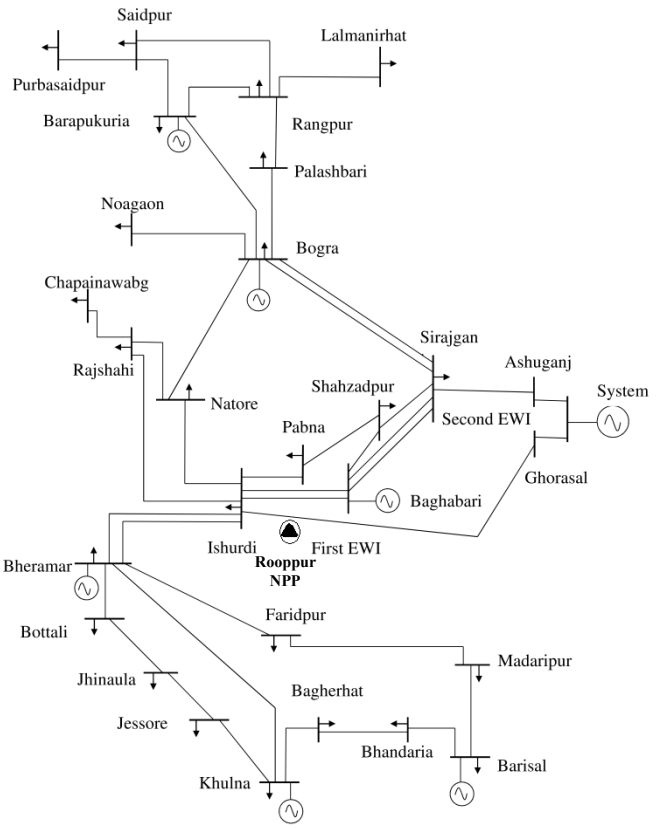


Fig. 1 Reduced EWIS of BPSN with proposed Rooppur NPP

The network model used in this study is fully based on the detailed transmission system of the BPSN, as available from Power Grid Company of Bangladesh (PGCB) website. Figure 1 shows the east-west interconnected system (EWIS) that has been created for representing BPSN in this work. Two 230 kV tie-lines interconnect the greater Eastern and Western Grid, which are considered as the first EWI and second EWI respectively, that are geographically divided by the combined flow of rivers Jamuna, Padma and Meghna [16].

Bulk power is generated in the eastern grid of Bangladesh power system network (BPSN). This power is transmitted to the western grid by means of two tie lines that are Ghorasal to Ishurdi (1st EWI) and Ashuganj to Sirajganj (2nd EWI). The newly proposed Rooppur NPP is indicated in the above figure. This topology has been derived from the actual network condition of BPSN.

Taking into account widely varied characteristics of the constituent load types, such as, resistive loads and mixed resistive-inductive loads, rotating inductive loads, thermostatic loads, extinction lighting loads etc., modeling of power system loads is a complex task [16]. Typical loads are represented as static voltage dependent load models where active and reactive powers are expressed as a function of bus voltages [16]. There is an exponential representation that represents the voltage dependency of the loads by varying the value of exponent, as shown in equation (1a) and (1b):

$$P = P_0 \left(\frac{V}{V_0}\right)^a \tag{1a}$$

$$Q = Q_0 \left(\frac{V}{V_0}\right)^b \tag{1b}$$

Where P_0 and Q_0 are the active and reactive power consumed at a bus at reference voltage V_0. If a=b=2, the resulting load model is constant impedance type where power varies with square of voltage magnitude. If a=b=1, it represents constant current load model, where power varies directly with voltage magnitude. Assuming a=b=0, i.e., ignoring voltage dependency, results in constant active and reactive power load model, which exerts maximum stress on the system [16].Constant P-Q static load model has been considered in this study due to unavailability of any reported study on load modelling for BPSN. System load data has been obtained from operational records of PGCB.

IV. Analytical Technique

Low voltage profile can be determined by bifurcation analysis. Bifurcation analysis is an effective tool for analysing nonlinear system stability. Only the saddle node and Hopf bifurcation are of interest. All type of bifurcation analysis is carried out by changing parameters of the system until bifurcation patterns immerge in system behaviour. In case of power system, the change in eigen value of the system due to changes in load bus power as a bifurcation parameter is studied. To locate the bifurcation, system parameter is presented with suitable mathematical equation [17]. The dynamic behaviour of the system can be presented by the following mathematical formulation [17,18],

$$\dot{x} = f(x, y, \lambda, p) \tag{2a}$$

$$o = g(x, y, \lambda, p) \tag{2b}$$

Where, $x \in R^n$ represents the system state variables, dynamic state of generators, loads and other time varying elements in the system, such as static VAR compensator controllers, exciter governor of the machine

etc.; $y \epsilon R^n$ corresponds to the algebraic variables, usually related to the transmission system and steady-state element models such as some generating sources and loads in the network; $\lambda \epsilon R^m$ corresponds to set of variables that are not controllable, varies continuously, typically loading factor is one of the parameters which varies as system load changes, and leads the system into bifurcation; $p \epsilon R^k$ stands for some directly controllable parameters such as degree of series shunt compensation; $f(.)$ and $g(.)$ represents vector function of algebraic equation and system constrains function, respectively [17]. From the bifurcation analysis PV diagram can be obtained. Using this PV diagram, low voltage profile can be predicted with LF analysis.

Another method of determining system stability is continuation power flow (CPF) technique. CPF technique consists of three steps: predictor, corrector and parameterization [17].

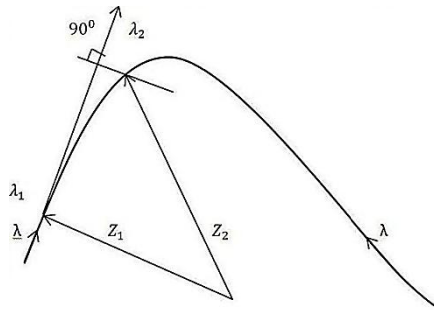


Figure 2 Continuation geometry analyses in state space and load parameter space

Predictor step involves arbitrary steps for λ and state variable to predict a new equilibrium point from the previously known equilibrium point. Corrector step is used to find intersection between the perpendicular plane to the tangent vector at the preceding equilibrium and branch as shown in Fig. 2. Step size is reduced if the process does not converge. As the system closes to the solution of power flow equation by varying loading parameter, system Jacobian becomes ill-conditioned. Then parameterization involves switching loading parameter to the new variables of the system that presents highest changes and guarantees the Jacobian of the power flow solution is non-singular at the saddle node bifurcation point [17].

V. RESULTS

The proposed Rooppur NPP is connected to the Ishurdi bus. So if Ishurdi and the buses directly connected to it in the Western grid (Bheramara, Rajshahi, Pabna, Baghabari and Natore) have low voltage effect, the NPP must have to bear the low system voltage impact.

The bus voltage and associated line length of Rooppur NPP are presented in Table I.

TABLE I
LINE VOLTAGE ANDLINE LENGTH CLOSE TO ROOPPUR NPP

Lines between buses	Voltage (kV)	Length (km)	
		Route	Ckt
Ghorasal-Ishurdi	230	178	356
Ishurdi-Baghabari	230	55	110
Ishurdi-Baghabari-Shahjadpur	132	57	57
Ishurdi - Pabna - Shahjadpur	132	56	56
Rajshahi - Natore	132	40	40

In Fig. 3, the PV diagram shows that for the system without compensation EWIS experiences low voltage for increasing loading factor (LF).

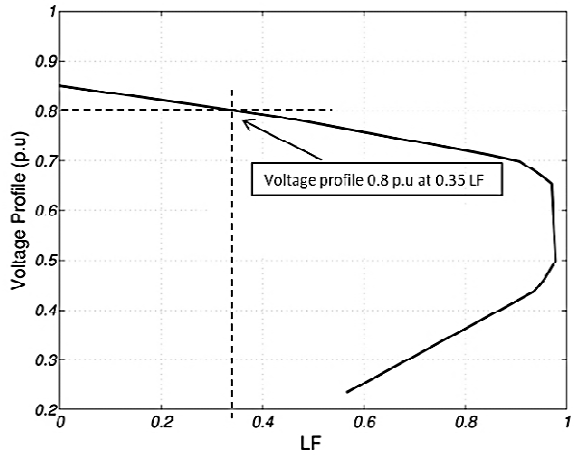


Fig. 3 Low voltage PV diagram

From the steady state analysis, it can be seen that with increasing LF, the voltage profile (p.u) of the above mentioned buses decreases. According to IEEE approval, around 0.8 voltage profile is considered as low voltage profile bus. From the Fig. 3, it is observed that at 0.35 LF voltage reaches 0.8p.u. After crossing the 0.35 LF, if the loading factor is further increased, the voltage profile (p.u) of the above buses falls so drastically that it does not able to meet the system voltage requirement.

The continuation power flow analysis using UWPFLOW has been performed for the overall system low voltage analysis. The result is presented in Table II, where the low voltage threshold is set at 0.80 p.u.

TABLE II
EWIS LOW VOLTAGE ANALYSIS

Bus Name	Corresponding Near Buses	LF (0.8 voltage p.u)	Actual minimum voltage (pu) (March-June, 14)
Barisal	Bhandaria, Madaripur	0.07	0.82
Bhandaria	Barisal, Bagherhat*	0.32	0.79
Faridpur	Madaripur, Bheramara*	0.07	0.86
Madaripur	Faridpur, Barisal	0.40	0.83
Saidpur	Purbasaidpur, Rangpur*	0.30	0.82
Purbasaidpur	Saidpur, Barapukuria Region*	0.20	0.83
ChapaiN'ganj	Rajshahi*	0.20	0.78

** means less affected buses*

If any buses of the power system experience low voltage then corresponding nearby buses also experience adverse effect due to it. As a result, not only the load buses but also the corresponding adjacent buses experience low voltage.

From above table it has been observed that actual minimum voltage (p.u) of Barisal, Faridpur, Madaripur, Saidpur and Purbasaidpur is more than 0.8p.u. Because loading of those buses is less. So voltage is comparatively high. But with the additional LF makes those buses voltage down to 0.8p.u. On the other hand, actual minimum voltage (p.u) of Bhandaria and ChapaiN'ganj is less than 0.8p.u. Because loading of

those buses is high. That's why voltage profile is low. So the above table shows their actual LF at 0.8 voltage profile (p.u).

From the above analysis, it has been observed that the buses around Rooppur NPP have the possibility to face the low voltage problem. In the system analysis for safety purpose in NPP criteria it is decided for 0.8 p.u voltage profile. Normally, after the 0.8 p.u voltage, with increasing LF the bus voltage decreases so drastically that system doesn't get enough time to recover its safe state.

VI. Conclusions

The deployment of NPPs to supply a portion of the electricity on a power grid brings with it a number of requirements on the grid design that is unique to it. The requirements must be considered by power grid planners and system operators. Unlike conventional power sources NPPs have long term shutdown cooling requirements that consume power and have stringent voltage and frequency limitations (imposed to assure the operability of critical emergency cooling systems). The operation process of electric grids must recognize the performance characteristics which are assumed by an NPP. All components and equipment within NPP electrical distribution system have their own protection and it is considered that the under voltage event can be only caused by events initiated outside the safety related electrical system. Degraded voltage condition is considered to be special type of under voltage event where voltage has lowered enough to cause equipment failure and/or damage if this state is present long enough. Under voltage event is the most likely event to occur among considered events.

In this paper a continuous power flow based approach to predict the most probable low voltage buses close to the Rooppur nuclear power plant buses in Western grid of the Bangladesh power system network has been presented. It has been observed that even with small load factor, buses around the Rooppur NPP experience low voltage which does tally with the actual situation. A detailed investigation to study the problem and possible solutions is thus recommended.

References

[1] International Atomic Energy Agency, "Interaction of Grid Characteristics With Design and Performance of Nuclear Power Plants," Technical Report Series No. 224, Vienna, Austria, 1983.

[2] International Atomic Energy Agency, "Electric Grid Reliability and Interface With Nuclear Power Plants," IAEA Nuclear Energy Series No. NG-T-3.8, IAEA, Vienna, Austria, 2012.

[3] United States Nuclear Regulatory Commission, "General Design Criteria for Nuclear Power Plants — Criterion 17 —Electric Power Systems,"10 CFR Appendix A, USNRC, Washington DC, USA, Jan. 2011.

[4] S. Zhang, C. Zhu, J. K. O. Sin, and P. K. T. Mok, "A novel ultrathin elevated channel low-temperature poly-Si TFT," *IEEE Electron Device Lett.*, vol. 20, pp. 569–571, Nov. 1999.

[5] Institute of Electrical and Electronic Engineers, Standard Criteria for Class 1E Power Systems for Nuclear Power Generating Stations, Nuclear Power Engineering Committee, IEEE Std 308-2001, IEEE New York, USA, 2001.

[6] International Atomic Energy Agency Report, Design of Emergency Power Systems for Nuclear Power Plants, IAEA Safety Guide No. NS-G-1.8, IAEA, Vienna, Austria, 2004.

[7] Institute of Electrical and Electronic Engineers, Recommended Practice for Protection and Coordination of Industrial and Commercial Power Systems, Industry Applications Society, IEEE Std 242-2001, IEEE New York, USA, 2001.

[8] Institute of Electrical and Electronic Engineers, Recommended Practice for Electric Power Distribution for Industrial Plants, Industry Applications Society, IEEE Std 141-1993, IEEE New York, USA, 1994.

[9] International Electromechanical Commission, Rotating electrical machines – Part 1: Rating and performance, IEC 60034-1 ed.12, IEC Geneva, Switzerland, 2010.

[10] National Electrical Manufacturers Association, Safety Standard and Guide for Selection, Installation, and Use of Electric Motors and Generators, NEMA Standards Publication MG 2- 2001 Revision 1, NEMA Rosslyn, USA, 2007.

[11] United States Nuclear Regulatory Commission, Branch Technical Positions (PSB) BTP PSB-1 - Adequacy of Station Electric Distribution System Voltages, USNRC, Washington DC, USA, 1981.

[12] N.K. Trehan, "Effect of de-regulation on operation of electrical grid and nuclear power generating stations," *in Proc. of Seventh International Conference on AC-DC Power Transmission*, pp. 1-6, 28-30 Nov. 2001.

[13] M. Y. Kim, C. H. Jeong, "Analysis for the effect on voltage sag of safety bus as electrical operating modes in nuclear power plants," *in Proc. of IEEE International Conference on Power System Technology*, pp.1-5, Oct. 30-Nov. 2, 2012.

[14] Steven Warren Blume, *Electric power system basics: for the nonelectrical professional*, New York, USA: John Wiley & Sons, 2007.

[15] Alan Wyatt, *Electric Power Challenges and Choices*, Toronto, Canada: The Book Press Limited, 1986.

[16] Md. Jane Alam, A. Hasib chowdhury, "Voltage collapse constrained loadability analysis of Bangladesh Power System Network," *in Proc. of Fourth IASTED Asian Conference on Power and Energy Systems*, Phuket, Thailand, Nov. 24 – 26, 2010.

[17] Md. Khurram Monir Rabby, Abdul Hasib Chowdhury, Md. Ali Azam, and Md. Asaduzzaman Towfiq, "Bifurcation Analysis to Identify Voltage Collapse in Bangladesh Power System," *in Proc. of International Conference on Informatics, Electronics & Vision*, pp.1-5, 17-18 May 2013.

[18] C. A. Canizares, Z. T. Faur, "Analysis of SVC and TCSC Controller in Voltage Collapse," *IEEE Transactions on Power Systems*, vol. 14, no. 1, pp. 158-165, Feb. 1999.

Determination of Affordable Production Cost And Nuclear Based Electricity Generation in Bangladesh

Sk. Samiul Reza*, Tasin Nawaz, Abid Yamin, Debabrata Das Tutul and T. A. Chowdhury
Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology, Dhaka–1205, Bangladesh
Email: *samiul.reza.tazin@gmail.com

***Abstract*—Among many great scientific achievements, discovery of electricity stands out in an unparalleled manner. The twentieth century has made use of electricity so extensively that it has almost changed the face of the earth. With this exponential rise in demand, power shortage has become a global phenomenon. South Asian countries alone are facing an average shortage of 15%-30% in electricity. In order to cope up with the increasing demand, short-term measure like oil based generation system has been introduced which has hiked up the cost of electricity. In a developing country like Bangladesh, cost of energy must be kept at bay to ensure minimum standard of life. That's where the term "affordable cost" comes into action. From the point of view of Bangladesh, where natural resource is scarce, nuclear based electricity generation is a suitable way to sustain increasing demand in an affordable manner. In this paper, fundamental concentration is given on both affordability and nuclear energy.**

***Index Terms*—Nuclear energy, affordable cost, reserve, tariff, consumption rate.**

I. Introduction

South Asia is an under-developed and energy resource deficient region that houses 1.65 billion people at present, which makes one-fifth [1] of the total population of the world. This region contains the highest number of people without access to electricity (658 Million) [2]. And most of the countries, including Bangladesh have a very low energy consumption rate. Key reasons behind this are low per capita income, high cost of energy, insufficient natural resources etc.

Affordable cost is an amount that a consumer feels comfortable to pay for energy and it doesn't lower his living standard by any means. It can be measured by considering net income, present tariff and consumption rate of energy. Scenario for the last 15 years in Bangladesh [3] is shown in Fig. 1.

It shows that, consumption rate and tariff have increased quite rapidly. In addition to that, Fig. 2 shows that electricity production cost is increasing day by day. However, in case of Bangladesh, maximum generation is much less than installed generation capacity which is shown in Fig. 3. To meet both affordable cost and high demand; we have to design a generation system versatile enough in both ends. Nuclear based generation system is an worthy candidate for that.

II. Affordable Cost Analysis

Affordable cost is a term dependent on numerous factors, most of them related to economics. Electricity has become one of the fundamental rights of the people. So, ensuring this right in a tranquil way is an imperative aspect for the country.

To determine affordable cost in the perspective of Bangladesh, a comparison is studied in Table I amongst the countries having the same economic standards as Bangladesh,

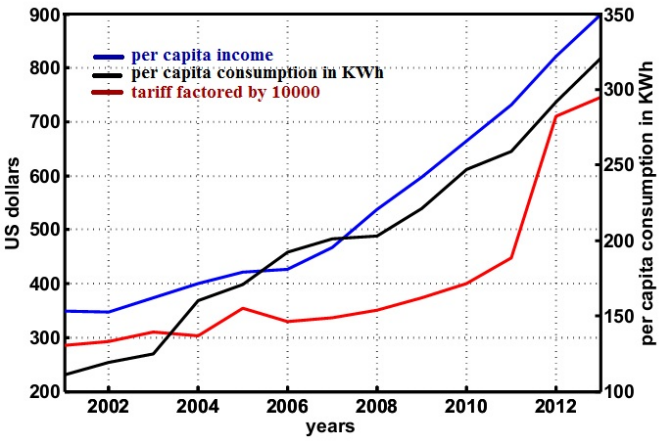


Fig. 1. Year-Wise Per Capita Income, Consumption Rate and Tariff in Bangladesh

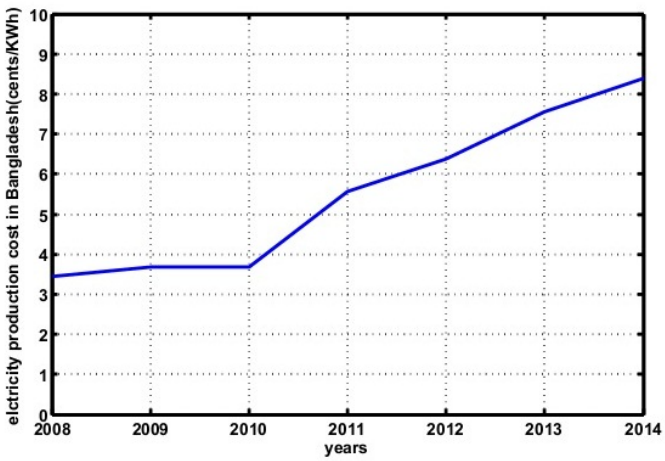


Fig. 2. Year-Wise Electricity Production Cost in Bangladesh

like Pakistan, India, Nepal, Sri lanka on the topics of per capita annual income, present tariff, per capita consumption, energy expense as a percentage of income etc. [3].

Here it is seen that, for most of the countries, energy expense as a percentage of income remains almost in the same range. This is the sign of a same level economic zone.

By observing year-wise energy expense as a percentage of income in these countries in Fig. 4, it can be deduced that except for Bangladesh, all other countries almost have a steady rate. In case of Bangladesh [4], it's evident that expense rate is getting steeper day by day. So conclusion can be made that, cost of energy in Bangladesh isn't affordable in general.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

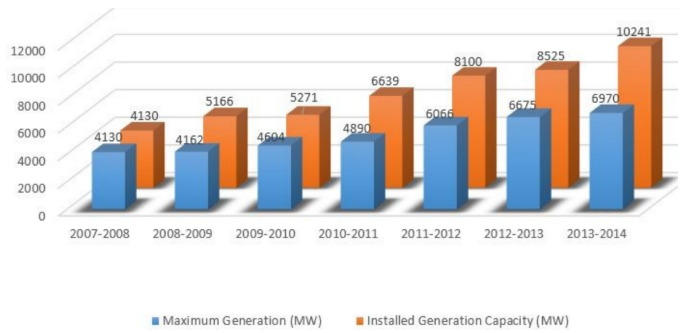


Fig. 3. Installed Generation Capacity vs Maximum Generation

TABLE I
COMPARISON OF ECONOMIC PARAMETERS AMONG SOUTH-ASIAN COUNTRIES IN 2013

Country name	Per Capita Annual Income ($)	Electric-ity Tariff (cents)	Per Capita Consumption (kWh/year)	Energy Expense as % Income
Bangladesh	903	5.833	321	2.0735
India	1504	9.48	787.3	4.9625
Pakistan	1307	10.92	431	3.601
Sri Lanka	3161	13	572	2.3524
Nepal	692	7.56	112	1.2236

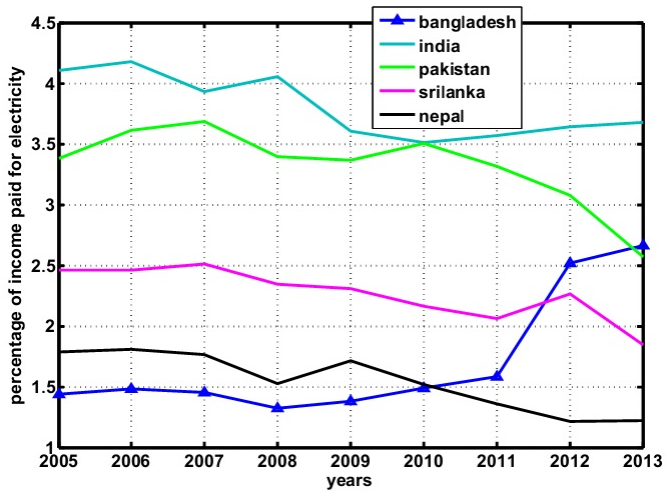


Fig. 4. Energy Expense As a Percentage of Income in Different South-Asian Countries

In order to obtain an affordable cost of energy, a seven step method is devised here which is shown in Fig. 5. Per capita annual income of Bangladesh in 2014 is around $ 1190. As it is seen that, in previous years, when the cost of energy was affordable, energy expense as a percentage of income was around 1.5% [4]. A highest limit of affordability can also be set and assumed that it is 2.0% of the income. From this point of view, a safe range of affordable cost can be ascertained which is, energy expense must remain within 1.5%-2.0% of the income. Comparing with the annual per capita income, annual cost of energy per capita should ranges from $ 17.85 to $ 23.8. Annual consumption rate per capita in Bangladesh is 321 kWh in 2014. So, affordable tariff of electricity should ranges from BDT 4.32 to BDT 5.76 per kWh.

Tariff consists of production cost alongside transmission cost, maintenance cost etc. So, affordable production would be a bit less than that derived tariff and can be assumed as

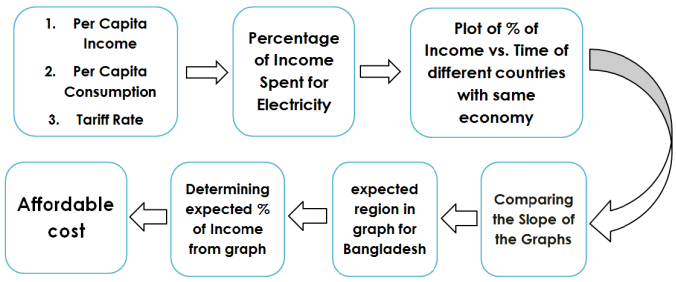


Fig. 5. Block Diagram of The Methodology of Determination of Affordable Cost

TABLE II
GAS AND COAL RESERVE OF BANGLADESH IN 2013

Gas Reserve		Coal Reserve	
Description	status	Location (exploration year)	Actual reserve (MMT)
Total no. of gas fields	23	Boropukuria (1985)	390
No. of gas fields in production	17	Khalashpir (1995)	143 (GSB), 685 (Hosaf)
Total reserve of extractable gas (proven & probable)	20.605 TCF	Fulbaria (1997)	572
Total reserve remaining after usage	11.174 TCF	Jamalganj (1965)	1050
Daily extraction	2035 MCF	Dighipara (1995)	200 (partial evaluation)
Daily demand	2500+ MCF		
Daily supply shortage	500+ MCF		

BDT 4.00 to BDT 5.00 per kWh. Bangladesh Government provides a huge subsidy in energy sector. In this production cost range, Government will not need to provide subsidy that much.

The production cost or tariff derived here is an average one and reflects the overall affordability of the whole population. No discrimination is done on the aspect of residential, industrial or commercial loads. Moreover, this range is for the whole consumption as no slab system is introduced here.

III. APPROACH TO AFFORDABILITY

There are some options regarding choice of fuel for electricity generation in our country. Natural gas, being the fundamental fuel for some time, has now reached its limit, whereas coal has a low production rate in Bangladesh. Feasibility of these fuels for future production is studied below:

A. Natural Gas

At present, around 80% of the produced electricity comes from natural gas. Being the fundamental fuel for electricity generation, natural gas is losing its reserve. At current rate of use in Bangladesh (1000 mmcfd), the estimated proven reserves would last 45 years [5]. If present rate of use increases at 10% per year, these reserves would last about 17 years. Another problem regarding this fuel is its fluctuation in supply, often the power plants do not get the necessary gas which halts the smooth generation system. Natural Gas reserve along with coal is shown in Table II.

TABLE III
RATE OF TAX ON COAL IN 2013 AND 2014

Solid Fuel	Rate of Tax at 1 May 2013 (US dollar per ton)	Rate of Tax at 1 May 2014 (US dollar per ton)
Coal	26.33	52.67
Peat:		
Peat briquettes	18.33	36.67
Milled peat	8.99	17.99
Other peat	13.62	27.25

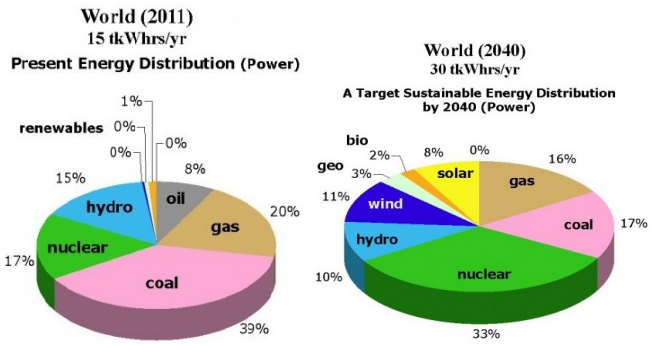


Fig. 6. Present and Proposed Future Energy Distribution of the World

B. Coal

Coal is another source which is getting low in reserve also and adding to that, coal of Bangladesh has low heat producing capacity. Coal reserves of five fields in Bangladesh are estimated at 3.0 billion tons equivalent to 67 tcf of gas, which can conveniently serve the energy needs of Bangladesh for 50 years [6]. Moreover, to tackle the environmental hazard, tax on coal has almost been doubled which is mandatory to be paid by the developing countries also. Short history on coal Tax is shown in Table III.

C. Hydro

Around 4% of total electricity generation comes from hydro power plant in Bangladesh. Though it has the advantage of almost zero fuel cost and longevity, hydro energy falls behind for its dependency on weather, extra-long setup time and extremely high capital cost.

D. Diesel and furnace oil

It has the advantage as short term solution but can't be used in the long term generation because of its extreme expense. Mainly to supplement the emergency crisis, oil based generation system can be used.

E. LNG

Liquefied natural gas (LNG) is natural gas (predominantly methane, CH_4) that has been converted to liquid form for ease of storage or transport. It might be used for very small scale generation but has a high risk of pressure related accident.

F. Nuclear

Nuclear fuel is increasing in percentage of the total fuel mix in the whole world. Because of low reserve, coal and gas sectors are shrinking and giving space to nuclear energy. Per kWh cost for overall generation in nuclear energy is quite low comparing to other sources.

TABLE IV
PRESENT AND TARGET SCENARIO OF ENERGY DISTRIBUION IN BANGLADESH

Fuel type	Installed Capacity in April 2012		Derated Capacity in April 2012		Target in 2021	
	MW	%	MW	%	MW	%
Gas	5086	76	4651	76.74	5651	30
Furnace Oil	110	1.64	85	1.4	65	0.3
Diesel	1017	15.2	905	14.9	500	2.97
Coal	250	3.74	200	3.3	9984	53
Hydro	230	3.44	220	3.63	188	1
Nuclear	0	0	0	0	1884	10
Renewable	0	0	0	0	565	3

Present and future fuel mix for electricity is shown in Fig. 6. In the perspective of Bangladesh, she doesn't have a single resource rich enough to produce majority share of the electricity. So, with the gas gone, she needs another long term generation system alongside coal. Nuclear based system fits right into the category. Proposed target for 2021 supports this fact as well [7] which is shown in Table IV.

IV. HISTORY OF NUCLEAR ENERGY AND FUTURE PLAN OF BANGLADESH

In 2011 nuclear power provided 10% of the world's electricity. In 2007, the International Atomic Energy Agency (IAEA) reported that there were 439 nuclear power reactors in operation in the world, operating in 31 countries. However, many have now ceased operation in the wake of the Fukushima nuclear disaster while they are assessed for safety. In 2011, worldwide nuclear output fell by 4.3%, the largest decline on record, on the back of sharp declines in Japan (-44.3%) and Germany (-23.2%).

The United States uses the highest nuclear energy, with nuclear power providing 19% of the total electricity it consumes, while France produces the highest percentage of its electrical energy from nuclear reactors that is 80% as of 2006. In the European Union as a whole, nuclear energy provides 30% of the electricity. Nuclear energy policy differs among European Union countries, and some countries such as Austria, Estonia, Ireland and Italy, have no active nuclear power station. In comparison, France has a large number of these plants, with 16 multi-unit stations in current use. In United States, while the coal and gas electricity industry is projected to be worth $85 billion by 2013, nuclear power generators are forecast to be worth $18 billion.

Bangladesh first considered building a nuclear power plant in 1961. Since then, several feasibility studies have been carried out, affirming the feasibility of the project. In 1963, Rooppur site was selected. In 2001, Bangladesh adopted a national Nuclear Power Action Plan. On 24 June, 2007 government of Bangladesh announced plans to build a nuclear power plant to quench electricity shortages. In May, 2010 Bangladesh signed a civilian nuclear agreement with the Russian Federation [8]. Bangladesh also has framework agreement for peaceful nuclear energy applications with the US, France and China. Although the Ministry of Science and ICT, the Ministry of Energy and Mineral Resources and Bangladesh Power Development Board (BPDB) are working on the electric grid system integration, the demographics and circumstances

TABLE V
EVALUATION OF ECONOMIC COEFFICIENT OF DETERMINATION FOR BANGLADESH VS NEIGHBORING COUNTRIES

Bangladesh vs	Per Capita Income	GDP	Tariff	Energy Consumption	Country Average
India	0.86	0.11	0.01	0.98	0.49
Pakistan	0.94	0.09	0.28	0.32	0.41

present do raise the question if indeed Bangladesh can produce energy from nuclear sources [9].

In February 2011, Bangladesh reached an agreement with Russia to build the 2,000 megawatt (MW) Rooppur Nuclear Power Plant with two reactors, each of which will generate 1,200 MW of power. The Perspective Plan of the Planning Commission of the government of Bangladesh for the period 2010-2021 has recommended an energy mix to achieve the generation of 20,000 MW by 2021. Targets of electricity production by 2013 and 2015 are 7,000 MW and 8000 MW, respectively [10].

V. SCRUTINIZING THE AFFORDABILITY OF NUCLEAR ENERGY IN BANGLADESH

As generation is yet to start in Bangladesh, there is no sample data to compare with the affordable rate which has been derived. So, study will be done with the data of the countries having almost same level of economy. India and Pakistan have been using nuclear energy for some time. First, economic co-relation will be calculated between India, Pakistan and Bangladesh, then respective co-relation factors will be multiplied with nuclear energy cost of these countries and that will be checked if it falls in the affordable cost region. If not, an alternative way will have to be fashioned to generate nuclear energy within desired affordable cost.

The co-relation [11] between x and y is expressed as,

$$R_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}} \quad (1)$$

In equation 1, $(x_i - \overline{x})$ is each x-value minus the mean of x and $(y_i - \overline{y})$ is each y-value minus the mean of y.

In statistics, the coefficient of determination, denoted R^2 and pronounced R squared, indicates how well data points fit a statistical model, sometimes simply a line or curve. It is a statistic used in the context of statistical models whose main purpose is either the prediction of future outcomes or the testing of hypotheses, on the basis of other related information. It provides a measure of how well observed outcomes are replicated by the model, as the proportion of total variation of outcomes explained by the model.

Obtained data by using this formula is shown in Table V. Average production cost of nuclear energy in Pakistan is 9.59 pakistani rupee/kWh [12], which is equivalent to BDT 7.58/kWh and average production cost of nuclear energy in India is 4.00 indian rupee/kWh [13], which is equivalent to BDT 5.16/kWh. These data are multiplied with the respective economic coefficients of determination between Bangladesh and both Pakistan, India which are 0.41 and 0.49. So, in comparison with Pakistan, average production cost of nuclear energy in Bangladesh should be around BDT 3.11/kWh and in comparison with India, average production cost of nuclear energy in Bangladesh should be around BDT 2.53/kWh. So, it can be said that, the range of average production cost of nuclear based electricity generation system in Bangladesh would be around BDT 2.53 to 3.11/kWh.

This range of cost is of course experimental and in comparison with the data received from neighboring countries. So, actual data may deviate from the findings.

VI. CONCLUSION

Asia is the main global region where electricity generating capacity and specifically nuclear power is growing significantly. In East and South Asia there are 119 operable nuclear power reactors, 49 under construction and there are plans to build a further 100 [14]. Many more are proposed. There may be some concerns regarding the safety of a nuclear power plant, but with the passage of time, safe modern technology like generation 3 and generation 4 reactors are ushering nuclear energy into a new era of possibility. Countries like France and Russia are acting as pioneers in nuclear energy sector. Bangladesh should tag along in this booming operation and ensure affordable electricity for her citizen.

REFERENCES

[1] S. Sunny, *Green Buildings, Clean Transport and the Low Carbon Economy: Towards Bangladesh's vision of a Greener Tomorrow*. Germeny: LAP Publishers, 2011.

[2] (2014) Energy poverty - wikipedia. [Online]. Available: http://en.wikipedia.org/wiki/Energy_poverty

[3] (2014) South Asia home. [Online]. Available: http://www.worldbank.org/en/region/sar

[4] (2014) Bangladesh energy situation - energypedia. [Online]. Available: https://energypedia.info/wiki/Bangladesh_Energy_Situation

[5] (2008) Natural gas reserve estimate of Bangladesh. [Online]. Available: http://www.hcu.org.bd/

[6] (2012) Coal - banglapedia. [Online]. Available: http://www.banglapedia.org/HT/C_0350.htm

[7] (2013) FDI in the energy and power sector in Bangladesh. [Online]. Available: http://www.cpd.org.bd/downloads/Session2_FK.pdf

[8] R. Akhter and M. I. A. Zahed, "Comparison of nuclear energy of Bangladesh with India & Pakistan," *IOSR Journal of Electrical and Electronics Engineering*, vol. 8, no. 3, pp. 64–68, 2013.

[9] S. S. Zaman. (2014) Bangladesh to pursue more nuclear energy - Dhaka Tribune. [Online]. Available: http://www.dhakatribune.com/development/2014/jun/05/bangladesh-pursue-more-nuclear-energy

[10] (2010) Power system master plan 2010. [Online]. Available: http://www.powerdivision.gov.bd/user/brec/40/55

[11] R. E. Walpole, R. H. Myers, S. L. Myers, and K. E. Ye, *Probability and Statistics for Engineers and Scientists*. Upper Saddle River, NI: Pearson Prentice Hall, 2011.

[12] I. Khan. (2014) Low cost power generation - Pakistan Observer. [Online]. Available: http://pakobserver.net/detailnews.asp?id=234815

[13] (2013) Cost of nuclear power proving high, DAE in a fix - The Times of India. [Online]. Available: http://timesofindia.indiatimes.com/india/Cost-of-nuclear-power-proving-high-DAE-in-a-fix/articleshow/27920490.cms

[14] S. Y. Raza. (2012) Regional cooperation on the way to sustainable nuclear energy systems. [Online]. Available: http://www.iaea.org/INPRO/4th_Dialogue_Forum/

Optimizing Economical Dispatch of a Microgrid in Islanding Mode

Amirhossein Ghods[1,*], Fahmid Al Farid[2], Shohag Barman[2], Ebrahim Sheikhi [3], Umme Salma[2]
[1]Electrical and Computer Engineering Department, University of Ulsan
Ulsan, South Korea
[2]Department of Computer Science & Engineering, University of Chittagong
Chittagong - 4331, Bangladesh
[3]Electrical and Computer Engineering Department, Power and Water University of Technology
Tehran, Iran
[*]amirhossein @mail.ulsan.ac.kr

***Abstract*—In this paper, the focus is on the dispatching strategy of microgrids with minimization in costs and emissions. The novelty of this paper is in proposing a new approach toward combined heat and power generation in a grid with renewable energy resources, such as wind turbines, photovoltaic cells which reduces the overall costs considerably. In this research, for simulating the optimal dispatch of DG units and other power generation resources, GAMS software is applied to the problem, which results in a lower calculation time.**

***Keywords*— Microgrid; Power Systems Operation; Distributed Generation; Optimization; Cost Minimization**

I. INTRODUCTION

Nowadays, local control and minimized cost are the most important factors in the field of generation and saving the energy at distribution level. To improve system performance, the Distributed Generation (DG) resources are widely used. They are applicable in ranges of a few kilowatts to tens of megawatts such as micro-turbines, fuel cells, photovoltaic systems, wind turbine systems, diesel motors, and gas turbines are examples of DG units [1]. Few works have analyzed the technical/economic benefits and potential issues of aggregated microgrid islands on the network comprehensively. There exist several algorithms to address this issue and find a method for an islanded-network online monitoring. In [2], the authors have studied the effects of switching from power factor to voltage control mode for distributed resources in case of transferring from grid-connected to islanded mode. Six main parameters are defined based on generators' terminal measurements of voltage and current to control the system in these two scenarios (grid-connected and islanded). Permutations of these parameters will provide 64 operating modes; the objective is to find the most appropriate mode for each scenario. Other similar research works, like those mentioned in [3-6] mostly discuss the reliability aspects of aggregation of several DGs either in grid-connected or islanded mode.

Also, due to the fact that the existence of distributed units decreases efficiency of conventional overcurrent protection methods which use same rate of current relays to the whole network [3], it is believed that new methods should be investigated. References [7-9] have proposed several novel algorithms regarding this matter; e.g. applying knowledge-based methods such as genetic algorithm to redesign the relay schemes with higher current limits.

The superior of CHP units compared to other distributed resources can be justified by the obtained efficiency, where the power loss and service reliability in the network increase because of usage of output heat in several applications [10-11]. A project has been conducted for supplying a limited area with electrical and thermal demands in New York City, and it has been shown that the efficiency of generation has roughly increased to 60% [12]. The process of demand response including CHP units is also performed in [13], and has been shown by adding CHP units to the network, total costs for households has been reduced by 1–4%.

In this paper, a novel approach for economic dispatch of a grid including distributed units (either in grid-connected or islanded mode) has been proposed. Solar power, photovoltaic cells, wind turbines, and CHPs are implemented in the system along with diesel generators, micro turbines, and fuel cells to be able to study the associated electrical and environmental issues comprehensively.

II. MODELLING OPTIMIZED OPERATION OF MICROGRIDS

In this part, distributed generation units have been modelled.

A. Diesel Generator

Diesel generator's fuel cost is mathematically formulated in polynomial form and depends on output power.

$$F_{DG,i} = \sum_{i=1}^{N} a_i + b_i P_{DG,i} + c_i P_{DG,i}^2 \quad (1)$$

N = Number of generators.

$P_{DG,i}$ = Output power and generator i

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

a_i, b_i, c_i = Generator's coefficients

From information of diesel generator's manufacturer, Caterpillar, Table I for parameters of (1) for a 6 kW, 50 Hz generator is resulted [14]:

TABLE I. PARAMETERS OF DIESEL GEMERATOR

Parameter	a_i	b_i	c_i
Value	0.4086	0.2419	0.0063

B. Micro Turbine

Efficiency of micro turbine is in a linear relationship with input power. In other words, the overall and electrical efficiencies are given by the following equations:

$$\eta_{tot} = \frac{P_J + Q_{dh}}{Q_{fuel}} \tag{2}$$

$$\eta_J = \frac{P_J}{Q_{fuel}} \tag{3}$$

where P_J is the electrical generated power, Q_{dh} is the output heat, and Q_{fuel} is fuel produced energy. With respect to (2), the cost for micro turbine's fuel is calculated as following:

$$C_{MT} = C_{nl} \sum_J \frac{P_J}{\eta_J} \tag{4}$$

In this equation, P_J is the output power generated in time period J, and C_{nl} is the price per unit of natural gas for supplying to micro turbine.

C. Wind Turbine

In design of wind turbines, there are two important factors that should be considered; one is accessibility to wind sources, and next is the output power characteristic of the turbine. Here, it is assumed that accessibility to wind sources is equal for all locations in the system; so for modelling of the performance of a wind turbine, the output power is analysed.

Based on information from wind turbine manufacturer-model AIR403, in wind speeds more than 18 m/s, the output of turbine is 130 W [15].

D. Photovoltaic Cell

Photovoltaic cells convert sun's heat energy to electricity. In this modelling, the information of manufacturer for model Solarex MSX-83 in standard situation (1000 W/m^2, and 25 □C) has been used [16]. The output power P_{PV} based on environment temperature is calculated as follow:

$$P_{PV} = P_{STC} \frac{G_{ING}}{G_{STC}} (1 + k(T_C - T_r)) \tag{7}$$

In (7), P_{STC} is the maximum module in standard situation, G_{ING} is the object radiation, G_{STC} is radiation in standard situation, k is heating coefficient of power, T_C is the cell temperature, T_r is the reference temperature.

The output characteristics for photovoltaic cell Solarex MSX-83 are the following: peak power equals to 83 W, voltage in peak power is 17.1 V, current in peak power 4.84 A, short circuit current 5.27 A, and open circuit voltage is 21.2 V [17].

E. Main Objective

The main objective includes cost function and air pollution released from each resource and is written as one polynomial. This objective should satisfy the following constraints:

1) Minimizing Operation Costs

The total operational costs, $CF(P)$, of the microgrid is defined as:

$$CF(P) = \sum_{i=1}^{N} (C_i \times F_i(P_i) + OM_i(P_i)) \tag{8}$$

$F_i(P_i)$ is the cost of fuel, $OM_i(P_i)$ is the cost for service of units.

The service cost is in relationship with generated power, and is shown as bellow. In (9), K_{OM_i} is coefficient of service cost.

$$OM_i = \sum_{i=1}^{N} K_{OM_i} P_i \tag{9}$$

In Table II, K_{OM_i} for each unit is presented:

TABLE II. K_{OM} VALUES FOR EACH UNIT

Name of Generator	Diesel Generator	Micro Turbine	Wind Turbine
Value of Coef.	0.01258	0.00587	0.00684

2) Minimizing Pollution

The air pollution usually includes gases such as SO_2, CO_2, etc. It is possible to model each of these polluting factors separately. In this problem, only CO_2 is modelled and considered in the objective problem. The final emission level of this polluter is given by:

$$E_P = \sum_{i=1}^{N} 10^{-2}(\alpha_i + \beta_i P_i + \gamma_i P_i) \quad (10)$$

In the above equation, α_i , β , γ_i are non-negative coefficients derived from micro turbine's and diesel generator's output characteristics. The values of these variables are presented in the Table III:

TABLE III. MICRO TURBINE AND DIESEL GENERATOR COEFFICIENTS IN EQUATION (12)

	α_i	β_i	γ_i
Micro Turbine	0.06490	0.05554	0.06490
Diesel Generator	-0.06131	0.05555	0.05151

3) Demand Response Constraints

For stability analysis of electric power system, the following constraints should be satisfied:

$$\sum_{i=1}^{N} P_i = P_L - P_{PV} - P_{WT} \quad (11)$$

As the first constraint suggests, in the above equation, $\sum_{i=1}^{N} P_i$ is total power generated by micro turbine and diesel generator, P_L is the total demand, P_{PV} is the output power of photovoltaic cell, and P_{WT} is the output of wind turbine in terms of kW.

The other constraint is supplying the heat demand. In this constraint, the output heat should exceed the input heat.

$$P_{L^{th}} - P_{MT^{th}} - P_{Heater} \geq 0 \quad (12)$$

In the above equation, $P_{MT^{th}}$ is the micro turbine's output heat, $P_{L^{th}}$ is the input heat, and P_{Heater} is the generated power by heater.

In order to prevent excess heat generation, penalty is considered dependent on the above variables.

$$Penalty = \left(\frac{P_{MT^{th}} + P_{Heater} - P_{L^{th}}}{40}\right) \quad (13)$$

In the (13), 40 is an arbitrary coefficient that will give an average reasonable penalty and is derived based on public surveys presented in [18].

For achieving a stable operation, the maximum and minimum output generated power for each unit should be kept in a range of minimum and maximum value.

$$P_i^{\min} \leq P_i \leq P_i^{\max}, \forall_i = 1, \ldots, N \quad (14)$$

F. Combined and Non-Combined Generation of Power and Heat Scenarios

The objective function is studied in two scenarios of combined and non-combined generation of heat and power in a network that is either grid-connected or islanded.

In the CHP mode, in addition to using heaters, the output heat of micro turbines is utilized. Conversely, in the non-CHP generation scenario, for responding to electrical demands, wind turbines, photovoltaic cells, diesel generators, and micro turbines are considered. Regarding the constraint for supplying thermal load, $P_{MT^{th}}$ is replaced with zero in this scenario.

III. SIMULATION RESULTS

Due to the fact that the problem has multi objectives to be satisfied and a nonlinear nature, it is not possible to use the conventional gradient solution for microgrid operation optimization. Therefore, a new mathematical approach has been proposed which is formulated in the previous section and is modelled in this section. This model follows a non-linear mathematical (polynomial) representation that includes several environmental and electrical constraints. Also, heat generation is included based on information derived from manufacturers' catalogues, and the effects of microgrids' islanding in CHP mode is studied as well.

In this paper, the calculations are done using GAMS software package. In this software, several algorithms are applied and examined to find the optimum solution for linear, non-linear, mixed integer, mixed integer non-linear, mixed integer complementarity problems.

A. Simulation Results

In this problem, operation costs, demand, and pollution level are considered for 24 hours. Figure 1 and 2 show the procedure of demand response between microgrid's generating units, in case that micro turbine is being operated in CHP mode, and $\omega = 0.7$. According to this information, with increase in demand at peak times, micro turbine and diesel generator's output will increase. Heater is turned on in the initial four hours that micro turbine's heat output is not adequate, and will be turned off in other hours. Photovoltaic cells and wind turbines provide the base demand and do not possess a large share of total generation.

Figure 1. Diesel Generator Daily Output in Different Weighted Coefficient

Figure 2. Micro Turbine Daily Output in Different Weighted Coefficient

Table IV shows total cost and pollution emmision level in three scenarios: in the first and second scenarios, optimization is done based on one single objective, and the third one will be based on a multi-objective problem. As it can be seen, in the first scenario, the level of emission is minimized, while in second one, operation cost is minimized, and finally in third scenario, both the emmision and operation costs are considered for optimization with $\omega = 0.7$. It can be seen that minimum cost for generation considering the level of pollution is about 22.21×10^3 USD for one day.

TABLE IV. TOTAL COST AND POLLUTION IN SCENARIOS

Scenario	Total Cost (USD×10^3 /day)	Total Pollution(Kg /day)
Single Objective (Pollution)	28.63	10.348
Single Objective (Cost)	21.22	14.307
Multi-Objective ($\omega = 0.7$)	22.21	13.171

In the Table V, the difference between these the cases could be clearly observed. In the case of CHP availability with $\omega = 0.7$ as in our case study, the overall cost is lower 12×10^3 USD averagely.

TABLE V. TOTAL COST IN CHP AND NON-CHP IN VARIOUS WEIGHTED COEFFICIENTS

Weighted Coefficient ω	1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2
Total Cost (CHP) (10^3×USD)	21.22	21.226	21.226	21.226	21.226	22.21	23.78	25.27	26.67
Total Cost (non-CHP) (10^3×USD)	33.47	33.47	33.47	33.47	33.47	34.52	35.99	37.56	39.09

IV. CONCLUSION

This paper proposes an optimal operation scheme for microgrids. For implementing such a network, the whole area should be investigated for all availabilities; in areas with adequate sun power, it is possible to use photovoltaic cells, and similarly wind turbines, diesel generators, and so forth. The important issue in such a network is the way of dispatching the sources, so the two main objectives, pollution level and operational costs, would be minimized. The main purpose of this research is to propose a new nonlinear analytical representation for optimal dispatch of these sources in a microgrid.

It was seen that because renewable sources that mostly provide base load, rely on environmental factors, their output is constant in all values of ω. For non-renewable sources such as diesel generators and micro turbines, for ω close to 1, micro turbine had the main responsibility for providing output power because of inexpensive generation cost, and vice-versa in ω near to 0 the major generation is from diesel generator because of low polluting factors. In addition, it was observed because the output heat of micro generation units can be recycled, considering CHP mode in micro turbines, total costs decrease dramatically.

REFERENCES

[1] F. A. Mohamed, "Microgrid modeling and online management." Ph.D. dissertation, Dept. Elect. Eng., Helsinki Univ., Espoo, Finland, 2005.

[2] S. Sishuba and M. A. Redfern, "Adaptive control system for continuity of supply using dispersed generators," *IEE Proc. Generation, Transmission and Distribution*, vol. 152, no. 1, pp. 23-30, Jan. 2005

[3] H. E. Brown et al, "Improving reliability of islanded distribution systems with distributed renewable energy resources," *IEEE Trans. Smart Grid*, vol. 3, no. 4, pp. 2028-2038, Dec. 2012

[4] C. Cho et al, "Active synchronizing control of a microgrid," *IEEE Trans. Power Electron.*, vol. 25, no.12, pp.3707-3719, Dec. 2011

[5] S. Conti et al, "Optimal dispatching of distributed generators and storage systems for MV islanded microgrids," *IEEE Trans. Power Del.*, vol. 27, no. 3, pp. 1243-1251, Jul. 2012

[6] T. L. Vandoorn et al, Transition from islanded to grid-connected mode of microgrids with voltage-based droop control," *IEEE Trans. Power Sys.*, vol. 28, no. 3, pp. 2545-2553, Mar. 2013.

[7] W. K. A. Najy et al, "Optimal protection coordination for microgrids with grid-connected and islanded capability," *IEEE Trans. Ind. Electron.*, vol. 60, no. 4, pp. 1668- 1677, Apr. 2013

[8] M. Dewadasa et al, "Fold back current control and admittance protection scheme for a distribution network containing distributed generators," *IET Generation, Transmission & Distribution*, vol. 4, no. 8, pp. 952-962, Aug. 2010

[9] M. Dewadasa et al, "Fault isolation in distributed generation connected distributed networks," *IET Generation, Transmission & Distribution*, vol. 5, no. 10, pp. 1053-1061, Oct. 2011

[10] A. Sheikhi et al, "A novel method to determine the best size of CHP for an energy hub system," in *Electric Power and Energy Conversion Systems Conf.*, Sharjah, UAE, Nov. 2011, pp. 1-7

[11] R. J. R. Budden, "CHP and the energy act," *IEE Rev.*, vol. 34, no. 9, pp. 341-343, Oct. 1988

[12] J. Kelley and G. Rouse, "The right combination," *IEEE Power & Energy Mag.*, vol. 6, no. 6, pp. 60-70, Oct. 2008

[13] M. Houwing et al, "Demand response with micro-CHP systems," *Proc. IEEE*, vol. 99, no. 1, pp. 200-213, Aug. 2010

[14] Q. Jakhrani et al. "Estimation of carbon dioxide footprints from diesel generator emmisions," in *Green and Ubiquitous Technology Int. Conf.*, Jakarta, Indonesia, July.2012, pp. 78-81.

[15] M. E. Amoli and K. Choma, " Electrical performance characteristic of a new micro-turbine generator," in *IEEE Power Engineering Society Winter Meeting*, Columbus, OH, Feb.2001, pp. 736-740.

[16] S. O. Ani et al. "Energy yield of two generator systems from small wind turbine application", in *IEEE Electric Machines & Drives Conf.*, Niagara Fall, ON, May.2011, pp. 735-740.

[17] J. Kaffashi. (2011, Dec.). *Photovoltaic Cells characteristics* [Online]. Available: http://193.190.56.244/~jelsen/msx83.pdf.

[18] S. Jadid and M. Rezaeian, "Investigations of CHP mode in decreasing microgrid's operating costs and pollution using multi-objective genetic algorithm," in *Int. Power Systems Conf.*, Tehran, 2010, pp. 25-32.

Extending Methods to Determine the Life Remaining of Transformer Paper Insulation

[1,*] Daniel Martin, [1] Tapan Saha, [1] David Allan, and [2] Kerry Williams

[1]School of Information Technology and Electrical Engineering, University of Queensland
St. Lucia, QLD 4072, Australia

[2]Western Power
Perth, WA 6000, Australia

[*]d.martin6@uq.edu.au

***Abstract*—Efforts have been underway to estimate the life remaining of the paper insulation used in power transformers, and thus estimate the remaining life of the whole unit. However, one problem encountered when estimating remaining life is that the current state of the paper is not often exactly known. One method to determine the condition of paper is to measure the furan content of the oil to estimate the degree of polymerization of the paper. However, one problem with this technique is that paper does not age uniformly within a transformer. In particular, paper usually degrades fastest at the top of the winding where the temperature is highest.**

We therefore set out to determine whether a correlation exists between the condition of the paper at the top of the winding and that inferred by the furan content of the oil. Samples of paper from sixteen decommissioned transformers were used to investigate the rates at which paper degrades at the top of the winding. These results imply that the rate of ageing of paper in this location did not exceed double that calculated from using oil furan content.

***Index Terms*—Power transformers; Paper products; Dielectric breakdown; Oil insulation**

I. Introduction

The life of a power transformer is ultimately dependent on the condition of its solid electrical insulation, which is usually of cellulosic paper and pressboard. This insulation paper is wrapped around the, usually copper, windings of the transformer. Over time chemical reactions occurring with the paper insulation cause it to become brittle [1], where it may eventually detach from the windings. These chemical reactions involve atmospheric oxygen, water and mineral acids [2 – 6]. These acids are generated from oxidation of the mineral oil [6]. Water can be generated from the oxidation of either paper or oil, or ingress from the atmosphere.

Models are being investigated to estimate the life remaining of paper insulation based on the water content of this insulation and the oxygen concentration of the oil in addition to temperature [5]. The problem with applying these models to an operating transformer is that the condition of its paper is not always known especially if no monitoring has been used. One technique frequently used to determine the state of paper is the furan measurement, in particular 2FAL [1]. This furan molecule is generated from the degradation of paper, and dissolves into the oil where its concentration can be measured by High performance Liquid Chromatography (HPLC). The condition of the paper has been found to be related by an empirical function of this concentration.

A problem with this technique is that paper does not degrade uniformly within a transformer. Paper often degrades fastest towards the top of the winding where the temperature is highest [7].

Given that insulation will fail on its weakest link, it is important to be able to estimate the condition of the brittlest paper within the transformer from a 2FAL measurement. In this study insulation paper extracted from decommissioned power transformers was studied, to investigate the difference between the condition of the paper at the top of the winding, and that inferred from the concentration of furans dissolved in the oil. Our strategy was to determine the average rate of ageing of paper derived from 2FAL and then investigate how it varies for paper located around a transformer winding.

II. Structure of Cellulosic Paper, its Ageing, and Furan Generation

A cellulose macromolecule is a very long chain formed of repeating glucose molecules. For the Kraft paper inside a newly manufactured transformer the average number of glucose rings per molecule, its degree of polymerization (DP), is about 1,000 to 1,200 [8]. Water, oxygen and oil acids steadily react with the cellulose breaking down the number of rings per molecule [9]. As the degree of polymerization falls the paper becomes more brittle. A DP of 200-250 is usually taken as being the end of life because the tensile strength of the paper has fallen to around a quarter of its original value [10, 11]. The key reaction modes affecting cellulose are [3]:

- Oxidation, involving atmospheric oxygen
- Hydrolysis, involving water which has either been sourced from the atmosphere or generated from chemical reactions
- Pyrolysis, involving only temperature

Oxidation changes the chemical composition of oil creating a large number of different compounds [12 - 13]. These compounds can then continue to react with the cellulose. For instance, mineral acids generated by the oxidation of the oil will react in combination with water promoting hydrolysis [6, 14 - 17].

In a previous paper, [5], two populations of transformers were analysed. One population was sealed while the other was free breathing. Despite these transformers having very similar levels of water dissolved in their insulation oil, furan levels indicated that the condition of the paper of the free breathing transformer population was worst.

As paper degrades furans are produced [1]. Furans are not generated by insulating oil, and so are indicative of the ageing of the paper [18]. The concentration of furan 2-

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

furfural (2FAL) dissolved in the oil has been shown to correlate with the DP of the paper by various researchers, including Chendong, Hill et al, De Pablo, Burton and Vuarchex [18].

However, problems have been noted such as the partitioning of the furan between the oil and paper is related to temperature, i.e. the concentration of furan can spike during periods of increased loading of the transformer [18]. Furans degrade over time, reacting chemically with the water and oxygen dissolved in the oil [18]. Consequently, determining a definitive mathematical relationship between the furan concentration and paper DP is challenging, as many factors must be taken into account. A higher precision in the relationship between DP and 2FAL levels would be possible with a large database by segregating the transformer classes of generation, transmission and distribution [18].

III. Modelling Degradation of Paper [17]

The reduction in DP of paper can be modelled using (1), where k is the rate of ageing, DP_0 and DP_t are the degree of polymerisation of the paper at the start (≈1,000) and after time t. Rate k is related to the temperature and conditions of ageing (i.e. water and oxygen content of system) given in (2), where A is a function of the water and oxygen content, E is the activation energy of the reaction (≈111 kJ/mol), R is the ideal gas constant (8.314 $JK^{-1}mol^{-1}$) and T is the temperature in Kelvin. Within normal transformer operating temperatures the A value is considered to be independent of the temperature, i.e. it does not change with temperature.

For approximately every 6-8 °C increase in temperature rate k is doubled [1], which is also evident in (2). However, in a transformer as the temperature increases there is a general tendency for cellulose to desorb water, slowing down rate k. This desorbed water can be adsorbed by cooler areas of the transformer's cellulosic materials.

$$\frac{1}{DP_t} - \frac{1}{DP_0} = kt \tag{1}$$

$$k = Ae^{-E/RT} \tag{2}$$

Online condition monitoring equipment is available which can measure the water content and temperature within a transformer, allowing these equations to be applied. For a new transformer the DP of the paper is known, being around 1000-1200 [1]. However, a challenge encountered when attempting to apply this model to an aged transformer is that the DP of the paper is not accurately known. It is also not a single value because the paper does not degrade uniformly; there is both a temperature and water concentration gradient along the winding.

As paper degrades furans are released. The De Pablo model suggests that each time three cellulose chains are broken, one 2FAL molecule is generated [18].

Chendong et al. found a linear relationship between the furfural concentration in logarithmic scale and the degree of polymerisation [19]. This was represented by (3), where F is the 2FAL concentration in mg/L. Stebbins et al [20] modified (3) to (4) where F is the 2FAL concentration in ppb.

$$Log\ F = 1.51 - 0.0035 \times DP \tag{3}$$

$$DP = \frac{\log(0.88 \times F) - 4.51}{-0.0035} \tag{4}$$

De Pablo published a comprehensive report on furanic compound analysis [21], reporting a similar log-linear relationship between the furans and DP. He has reported five different relationships based on experiments conducted at different conditions.

In [22] De Pablo reported a relation between furfural and DP as shown in (5) below, where 2FAL is the furan concentration expressed as mg/kg of oil.

$$DP = 7100/(8.88 + 2FAL) \tag{5}$$

Pahlavanpour [23] modified the formula based on transformer conditions, and an assumption that 20% of winding paper and the inner paper layer degrades twice as fast as the rest of the paper. The revised formula is (6).

$$DP = 800/[(0.186 \times 2FAL) + 1] \tag{6}$$

IV. Tested Transformers

Five Australian utilities provided data on power transformers which had been scrapped, and subsequently had had samples of paper removed for analysis and comparison with a furan measurement. To be of use for this study the paper had to have been noted as being extracted from the top of one of the windings, often paper had been taken yet the utility had not recorded where it had been extracted from. Data from 16 power transformers was sufficiently complete for this study. The DP estimated using the Chendong equation was compared with the DP of the paper from the top of the winding. The year when these transformers were manufactured was known, and therefore two average rates of ageing were calculated for both DP measures, and then they were compared. This data is given in Table I.

A. Transformer #1

This was a 110/33 kV 48/60/80 MVA transformer manufactured in 1978 and decommissioned in 2011. Paper was removed from the top, middle and bottom of different windings. The DP of these samples of paper was mostly less than that estimated using 2FAL (Fig 1). This is a concern since there was less life left in the paper than that estimated using furan.

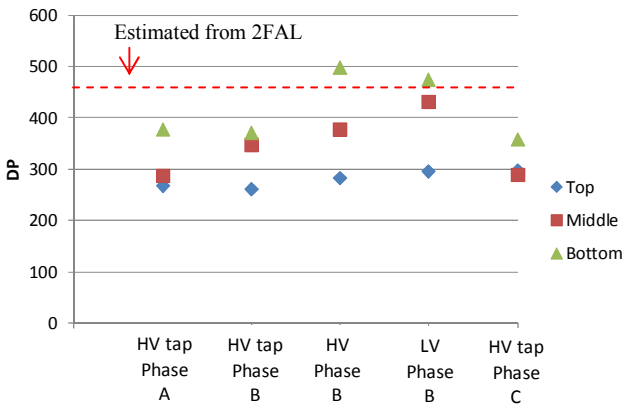


Fig. 1. Degree of polymerisation of paper from different sections of transformer #1.

B. Transformer #2

A scrapped 132/33/11 kV 45 MVA had several samples of paper removed from one winding. This unit was manufactured in 1963 and decommissioned in 2013. The DP

along one winding is shown in Fig 2. The 2FAL content was 987 ppb, providing a calculation for DP of 449. The values shown in the figure below are very close to calculated DP. As this transformer was used in the transmission network, it probably ran fairly cool with a minimal temperature gradient.

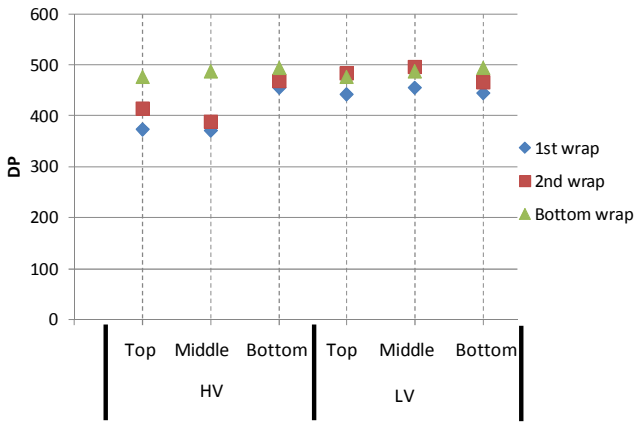


Fig. 2. Degree of polymerisation of paper from different sections of transformer #2.

C. Transformer #3

Samples of paper were removed from a failed 47 year old 20/27 MVA transformer [7], Fig 3. The DP of the paper from the top of the windings was nearly always lower than at the bottom of the same winding (only the paper on the HV tap of phase A did not show this relationship). The 2FAL concentration was used to estimate DP using (4). A value of 350 was calculated for this transformer from a 2FAL concentration of 2.2 ppm, which is around the average value of the measurements. However, the minimal DP shows that the paper at the top of the winding had very nearly reached the end-of-life criterion set by IEC 60076/7.

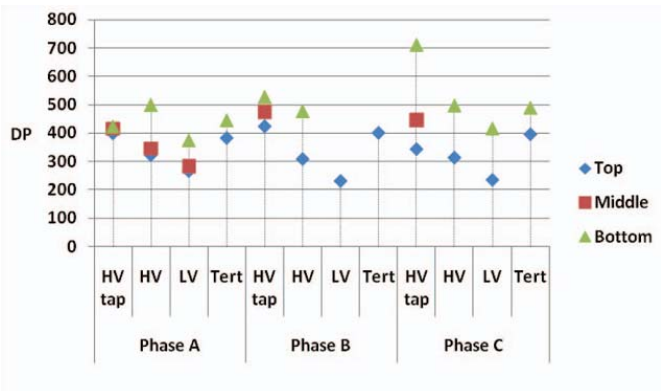


Fig. 3. Degree of polymerisation of paper from different sections of transformer #3.

D. Transformers #4 and #5

These transformers were operated by the same utility. Their DP measurements are shown in Fig 4. Transformer 4 (132/69 kV 30/45 MVA) was installed in 1981 and was 32 years old when decommissioned. Transformer 5 (unknown rating) was installed in 1961 and scrapped in 2012. Both their paper is more degraded than suggested by furan level, which is 600 and 440 respectively for the two units.

E. Transformers #6 to #16

A utility habitually collects at least one paper sample from each decommissioned transformer, however, does not have a policy on where the paper should be taken from. They supplied a database of results, which was filtered to only contain DP of paper knowingly taken from the top of the winding. These results are given in Table I.

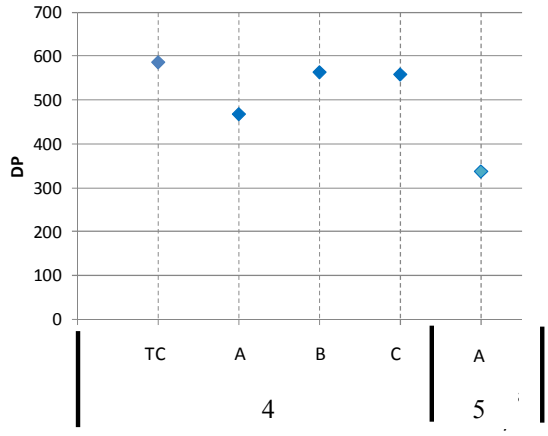


Fig. 4. Degree of polymerisation of paper from different sections of transformers #4 and #5.

V. Analysis

The rates of ageing obtained from using both the DP calculated from furan level, k_{DP-F} , and from the DP measurement on the paper removed from the top of the winding, k_{DP-TW}, are shown in Table I. The ratio of k_{DP-TW} to k_{DP-F} is also given in the table as k_R. For most of these transformers k_R was between 1.4 and 2.7, with an average of 1.9. The precise location of the tested paper was not always known, in reality there are likely to be temperature gradients between LV, HV and tertiary windings, which was not explored in this article.

The 2FAL level clearly did not match the DP expected for the paper from five transformers, numbers 8, 11, 13, 15 and 16 do not agree very well. For the first four transformers the furan concentration is less than expected this may infer that furans were removed from the oil, for instance if the oil has been dried during the life of the transformer [7, 18]. During an online dryout of a transformer furans were removed by the sieves used to extract water. For the fifth transformer the 2FAL level is higher than that expected. One possibility is that this transformer was running at a higher load than normal, which forced more 2FAL into the oil.

There does seem to be general trend that the paper at the top of a winding ages at a rate up to 3 times faster than the rate obtained from furan. However, one 2FAL measurement may not be sufficient for this calculation because in about a third of these cases it did not adequately reflect the condition of the paper. Furthermore, there does not appear to be an effective technique to indicate that there is a problem with the 2FAL level. The utility should therefore make regular, periodic, measurements and track trends.

For the transformer investigated in [7] the furan levels were adjusted after they had been removed by the online dryout system. The furan concentration was measured directly after the oil was processed, and then any increase was added to the original level. However, most of the furans stay absorbed in the paper, rather than dissolving into the oil [18]. It is likely that some of the increase in oil furan concentration would have been caused by furans moving from the paper into the oil.

TABLE I
AGEING RATES OF PAPER

	YOM	Age on Test (years)	2 Furfural (ppm)	Paper DP Determined from (4)	Rate of Ageing $\frac{1}{DP_t} - \frac{1}{DP_0}$ (h^{-1})	Paper DP value	Rate of Ageing	Ratio k_R
1	1978	33	0.460	460	4.1E-9	260	9.8E-9	2.4
2	1963	50	0.449	449	2.8E-9	371	3.9E-9	1.4
3	1963	47	0.350	350	4.5E-9	220	8.6E-9	1.9
4	1981	32	0.600	600	2.4E-9	473	4.0E-9	1.7
5	1961	51	0.440	440	2.8E-9	334	4.5E-9	1.6
6	1976	28	0.895	461	4.8E-09	300	9.5E-9	2.0
7	1971	37	0.193	651	1.6E-09	498	3.1E-9	1.9
8	1974	35	0.011	1007	-	526	2.9E-9	-
9	1986	22	0.055	807	1.2E-09	616	3.2E-9	2.7
10	1966	44	0.396	562	2.0E-09	478	2.8E-9	1.4
11	1965	41	0.263	613	1.8E-09	232	9.2E-9	5.1
12	1969	37	0.911	459	3.6E-09	352	5.7E-9	1.6
13	1963	43	0.147	685	1.2E-09	188	11.5E-9	9.6
14	1969	35	1.81	374	5.5E-09	196	13.4E-9	2.4
15	1952	53	0.261	614	1.4E-09	223	7.5E-9	5.4
16	1987	20	1.069	439	7.3E-09	615	3.6E-9	-

VI. CONCLUSIONS

Furan analysis is an effective method to estimate the condition of paper. However, it has several drawbacks such as furans degrade over time, furans exist in partition between the oil and paper where temperature affects this partition, and furans are removed during oil conditioning. A utility should therefore track furan level over time so they can detect these drawbacks. The paper within a transformer ages at different rates, thus one value of DP does not adequately describe the condition of the weakest paper. The remaining life of the overall transformer is dependent on this weakest paper.

In these sixteen analysed power transformers, the insulation paper at the top of the windings clearly degraded at a faster rate than that implied using the furan content of the oil. In most cases the rate of ageing for this paper was no more than twice the rate of ageing determined from furans. This is useful because when determining the life remaining of paper insulation this method can help indicate the condition of the most degraded paper within the transformer.

When a utility samples paper from a decommissioned transformer the location of the paper must be noted, and which layer. The sample should be taken from the top of a winding where the paper is likely to degrade fastest. As per [18], there may be better correlation between DP and furans when the class of transformer is known.

REFERENCES

[1] CIGRE Task Force D1.01.10, "Ageing of cellulose in mineral-oil insulated transformers", CIGRE Brochure No. 323, 2007.

[2] L. E. Lundgaard, W. Hansen, D. Linhjell, and T. J. Painter, "Aging of oil-impregnated paper in power transformers", IEEE Trans. Pow. Del., Vol. 19, pp. 230-239, 2004.

[3] N. Lelekakis, D. Martin, and J. Wijaya, "Ageing rate of paper insulation used in power transformers Part 1: Oil/paper system with low oxygen concentration", IEEE Trans. Dielectr. Electr. Insul., Vol 19, Iss. 6, pp. 1999 – 2008, 2012.

[4] N. Lelekakis, D. Martin, and J. Wijaya, "Ageing rate of paper insulation used in power transformers Part 2: Oil/paper system with medium and high oxygen concentration", IEEE Trans. Dielectr. Electr. Insul., Vol. 19, Iss. 6, pp. 2009 – 2018, 2012.

[5] D. Martin, Y. Cui, T. Saha, N. Lelekakis, and J. Wijaya, "Life Estimation Techniques for Transformer Insulation", Conf. AUPEC, session S21.7, Australia, 2013.

[6] N. Lelekakis, J. Wijaya, D. Martin, and D. Susa, "The Effect of Acid Accumulation in Power Transformer Oil on the Ageing Rate of Paper Insulation", IEEE Elect. Insul. Mag., Vol. 30, Iss. 3, pp. 19 – 26, 2014.

[7] N. Lelekakis, W. Guo, D. Martin, J. Wijaya, and D. Susa, "A Field Study of Ageing in Paper-Oil Insulation Systems," IEEE Electr. Ins. Mag., Vol. 28, no. 1, pp. 12–19, Jan.-Feb., 2012.

[8] T.V. Oommen and T. Prevost, "Cellulose Insulation in Oil-Filled Power Transformers: Part I – History and Development", IEEE Elect. Insul. Mag., Vol. 22, Iss. 1, pp. 28 - 35, Jan.-Feb., 2006.

[9] A. M. Emsley and G. C. Stevens, "Kinetics and mechanisms of the low-temperature degradation of cellulose", Cellulose, Vol. 1, pp. 26-56, 1994.

[10] "IEEE Guide for Loading Mineral-Oil-Immersed Transformers and Step Voltage Regulators", IEEE Std C57.91-2011, IEEE Power & Energy Society, USA, 2012.

[11] IEC 60076-7, "Power transformers - Part 7: Loading guide for oil immersed power transformers", Switzerland, 2005.

[12] J. Sabau, "On-line Reclamation of Aged Transformer Oils", IEEE Conference on Electrical Insulation, pp. 555-567, USA, 2001.

[13] P. Hodges, *Hydraulic Fluids*, Elsevier, Holland, 1996.

[14] D. Hill, T. T. Le, M. Darveniza, T. K. Saha, "A Study of Degradation in a Power Transformer-Part 3: Degradation Products of Cellulose Paper Insulation", Polymer Degradation and Stability, Vol. 51, pp. 211-218, 1996.

[15] D. Hill, T. T. Le, M. Darveniza, T. K. Saha, "A Study of Degradation of Cellulosic Insulation Materials in a Power Transformer-Part 1: Molecular Weight Study of Cellulose Insulation Paper", Polymer Degradation and Stability, Vol. 48, No. 1, pp. 79-87, 1995.

[16] D. Hill, T. T. Le, M. Darveniza, T. K. Saha, "A Study of Degradation of Cellulosic Insulation Materials in a Power Transformer-Part 2: Tensile strength of Cellulose Insulation Paper", Polymer Degradation and Stability, Vol. 49, No. 3, pp. 429-435, 1995.

[17] T. K. Saha, "Review of modern diagnostic techniques for assessing insulation condition in aged transformers", IEEE Trans. Dielectr. Electr. Insul., Vol. 10, No. 5, pp. 903-917, Oct. 2003.

[18] CIGRE Task Force D1.01.10, "Furanic Compounds for Diagnosis", CIGRE Brochure No. 494, 2012.

[19] X. Chendong, F. Qiming, and X. Shiheng, "To Estimate the Ageing Status of Transformers by Furfural Concentration in the Oil", CIGRE Study Committee 33 Colloquium, Russia, 1991.

[20] R. D. Stebbins, D. S. Myers, and A. B. Shkolnik, "Furanic compounds in dielectric liquid samples", Int. Conf. Properties and Applications of Dielectric Materials, pp. 921 – 969, Japan, 2003.

[21] A. De Pablo, "Furanic Compound Analysis: a Tool for Predictive Maintenance of Oil-filled Electrical Equipment", CIGRE Task force report, 2003.

[22] A. De Pablo, "Furfural and Ageing: How are they Related", IEE Power Division Colloquium Insulating Liquids, National Grid Leatherhead, UK, 1999.

[23] B. Pahlavanpour, "Power Transformer Insulation Ageing", CIGRE SC 15 Symposium, Sydney, Australia, 1995.

Mitigation of Rapid Voltage Variations Caused by Passing Clouds in Distribution Networks with Solar PV Using Energy Storage

M. J. E. Alam,[1,*] K. M. Muttaqi,[1,$] and Danny Sutanto[1,#]

[1]School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, NSW 2522, Australia
[*]mjea982@uowmail.edu.au, [$]kashem@uow.edu.au, [#]soetanto@uow.edu.au

Abstract—**Passing clouds can create unacceptable rapid voltage variations in weak distribution networks with a significant solar PV penetration. This paper investigates the benefit of a PV inverter ramp-rate control strategy using battery energy storage for the mitigation of rapid voltage variations. By controlling the discharging/charging operation of the energy storage appropriately, not only the inverter ramp-rate can be maintained at a desired level, but also the voltage variations can be controlled within acceptable limits. Results, from simulation using a real Australian distribution feeder with real load demand and PV output profile, show that with a high penetration of PV, the voltage variations can increase beyond the allowable limit stipulated by the IEC 61000-2-2 standards. However, with the proposed ramp-rate control, the variation can be maintained well below the limits.**

***Index Terms*—Solar PV, Rapid voltage variation, Ramp-rate control, Energy storage.**

I. Introduction

The impact of the fluctuation of the solar photovoltaic (PV) output due to cloud passing on electric utility operation has been reported in previous literature [1]. Voltage fluctuation caused by steep variations in PV output is particularly significant for low voltage (LV) distribution systems as the LV feeders are normally located far away from the main sub-station and often feature low strength or fault level [2]. Further, the LV distribution feeders with a high R/X ratio are highly vulnerable to the threat of unacceptable rapid voltage variations due to the active power variations of PV inverters. IEC standard 61000-2-2 [3] suggests that a rapid voltage variation over 3% of the nominal value is not acceptable for LV distribution systems. In a weak distribution feeder with a high penetration of PV systems, there is a possibility that this limit will be violated. Rapid voltage changes in distribution networks can cause unacceptable flickers in illumination devices, spurious tripping of relays and sensitive electronic equipment, stalling of induction motors [4], and may also create hunting of voltage regulating devices in the upstream networks [5]. Therefore, a solution for smoothing the sudden variations in PV output is necessary.

Previously reported research works have addressed the PV output smoothing problem using various methods including the use of electric double-layer capacitor [6], battery energy storage [7], or a combination of both [8, 9]. Although moving average control of energy storage was performed in some earlier works e.g. [6, 10] for PV inverter ramp-rate control, the authors of this paper pointed out a drawback of the traditional moving average method to achieve an accurate control of the inverter ramp-rate and to achieve this, the authors proposed a new strategy for ramp-rate control in [11]. While this approach mitigates the effect of the PV output fluctuation, it can also smooth out the rapid variations of voltage by injecting or consuming the appropriate amount of power using the energy storage device integrated with the PV system. The benefit of ramp-rate control in voltage fluctuation mitigation was not clearly identified in [11], as the paper mainly proposed the new approach and investigated its general application as a ramp-rate control algorithm. However, such an approach could have the value added benefit of mitigating the rapid voltage variations in distribution networks, especially in the LV feeders with sensitive installations where the limits of rapid voltage variation should be strictly adhered to.

This paper investigates the rapid voltage variations caused by cloud passing induced PV output fluctuations and explores the potential of the inverter ramp-rate control using energy storage for smoothing out the voltage fluctuations. A real PV output profile is imposed on the PV inverters connected to an Australian LV distribution feeder featuring the characteristics of a weak grid to analyze the true nature of the rapid voltage variations. Finally, energy storage devices integrated with rooftop PV systems are intelligently controlled using the proposed ramp-rate control strategy and its effect on the voltage smoothing, particularly its ability to maintain the allowable rapid variation limit, is assessed. Results presented in this paper are obtained from simulations using steady-state and dynamic analysis.

II. Analysis of Voltage Fluctuations Caused by Cloud Induced Fluctuations in PV Output

A PV system connected to a distribution network, as shown in Fig. 1, is considered for analyzing the voltage variation caused by steep changes in PV inverter output. For the discussions in this section, it is assumed that the energy storage is not yet integrated with the PV system.

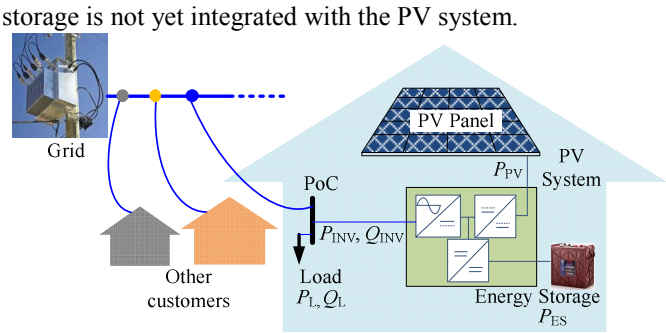


Fig. 1 A PV system connected with the grid.

The change in voltage magnitude ($\Delta|V|$) at the point of connection (PoC) of a PV inverter system is a function of the change in net change of active (ΔP) and reactive power (ΔQ) at the PoC as given in (1).

$$\Delta|V| = \frac{\Delta P\sigma_{Q\delta} - \Delta Q\sigma_{P\delta}}{\sigma_{P|V|}\sigma_{Q\delta} - \sigma_{P\delta}\sigma_{Q|V|}} \quad (1)$$

where, $\sigma_{P\delta}, \sigma_{Q\delta}, \sigma_{P|V|}, \sigma_{Q|V|}$ are the sensitivities of active and reactive power with the magnitude and angle of voltage, respectively, obtained from the Jacobian of the classical Newton-Raphson approach.

For a given equilibrium condition of the network, the variation in PoC voltage will mainly depend of the variation in load demand (ΔP_L, ΔQ_L) and PV inverter output (ΔP_{INV}, ΔQ_{INV}) as given below.

$$\Delta|V| = \frac{(\Delta P_L - \Delta P_{INV})\sigma_{Q\delta} - (\Delta Q_L - \Delta Q_{INV})\sigma_{P\delta}}{\sigma_{P|V|}\sigma_{Q\delta} - \sigma_{P\delta}\sigma_{Q|V|}} \quad (2)$$

As the PV inverters at LV feeders mainly operate at unity power factor and the load demand does not change as abruptly as the change in PV output caused by passing clouds, the rapid change in PoC voltage caused by cloud passing can be attributed mainly to ΔP_{INV} as given below.

$$\Delta|V| = \frac{-\Delta P_{INV}\sigma_{Q\delta}}{\sigma_{P|V|}\sigma_{Q\delta} - \sigma_{P\delta}\sigma_{Q|V|}} \quad (3)$$

To quantify the rapid changes in voltage in a PV integrated network caused by passing clouds, a performance index termed as Rapid Voltage Change Index (RVCI) is defined as given below.

$$\text{RVCI} = \frac{\Delta|V(k)|}{V_{Nom}} = \frac{|V(k)| - |V(k-1)|}{V_{Nom}} \quad (4)$$

where, $\Delta V(k)$ is the voltage change at PoC at any given k-th instant, determined from the change in voltage between two consecutive instants and V_{nom} is the nominal voltage. RVCI will be expressed in percentage to compare with applicable standards; a value of RVCI beyond 3% will result in violation of the IEC standard 61000-2-2 [3].

III. Ramp-rate Control of PV Inverter for Voltage Fluctuation Mitigation

This section develops an approach for smoothing out the sudden changes in PV output by controlling an energy storage unit integrated with the PV system. The smoothing out of the PV output will result in the mitigation of the rapid changes in PV output.

A. The Ramp-rate Control Algorithm

The ramp-rate control of the PV output is essentially the action of limiting the value of ΔP_{INV} in (3) to a desired value. This can be realized by integrating an energy storage device with the PV system, as shown in Fig. 1. The energy storage device will need to be controlled in such a manner that it delivers (discharges) or consumes (charges) an appropriate amount of power, P_{ES} so that ΔP_{INV} can be controlled at a desired value. The necessary change in storage power ΔP_{ES} at a given k-th instant for controlling the ramp-rate of P_{INV} to a desired value ($\Delta P_{INV\text{-}des}$) can be determined using the expression below [11].

$$\Delta P_{ES}(k) = \frac{\Delta P_{INV\text{-}des}(k)}{\eta_{INV}} - \Delta P_{PV}(k) \quad (5)$$

where, η_{INV} is the inverter efficiency. This change in P_{ES} will, in effect, contribute to mitigate the sudden voltage fluctuation at the k-th instant as evidenced in (6).

$$\Delta|V(k)| = \frac{-[\Delta P_{PV}(k) \pm \Delta P_{ES}(k)] \times \sigma_{Q\delta}}{\sigma_{P|V|}\sigma_{Q\delta} - \sigma_{P\delta}\sigma_{Q|V|}} \quad (6)$$

The ± sign in (6) indicates that either charging or discharging operation of the storage device needs to be performed for voltage fluctuation mitigation. If PV output is suddenly reduced due to cloud shadow, the storage device will discharge ΔP_{ES} amount of power which will be added to the PV power for minimizing a sharp decrease in voltage. If the PV output increases suddenly due to the removal of cloud shadow, then the storage device will charge the storage device with ΔP_{ES} amount of power which will be subtracted from the PV output to mitigate a sharp increase in voltage. The ramp-rate control algorithm given in (4) is illustrated in Fig. 2 using a fictitious PV output P_{PV} continuing for 4 consecutive time instants from (k-1)th to (k+2)th.

During the first time instant denoted by (k-1)th, P_{INV} is equal to P_{PV} (assuming η_{INV} is 100% for simplicity of the description). At the k-th instant, P_{PV} decreases sharply due to cloud shadow by such an amount that exceeds the desired ramp-rate, $\Delta P_{INV\text{-}des}$. To control the ramp-rate at the desired value, the storage power P_{ES} is changed by $+\Delta P_{ES}(k)$, determined using (5). At the (k+1)-th instant, P_{PV} increases sharply as the cloud passing period ends. To limit the change of P_{INV} at the desired value, P_{ES} is changed by $-\Delta P_{ES}(k+1)$, determined using (4). At the (k+2)-th instant, P_{PV} does not change; however, P_{ES} is changed by $\Delta P_{ES}(k+2)$ to increase P_{INV} for attaining the level of $P_{PV}(k+2)$. This change is performed in such a way that the ramp-rate of P_{INV} does not exceed $\Delta P_{INV\text{-}des}$. This brings an important consideration for developing a ramp-rate control strategy. The value of $\Delta P_{INV\text{-}des}$ during the in-ramp and post ramp events of a fluctuation needs to be selected in such a way that controlling of PES itself do not create any fluctuation. Interested readers are referred to [11] where a detailed methodology for the selection of appropriate values of $\Delta P_{INV\text{-}des}$ for PV inverter ramp-rate control has been proposed.

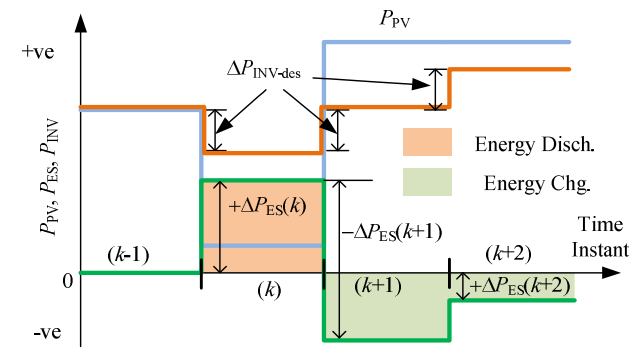


Fig. 2 An illustration of the ramp-rate control algorithm.

B. Control Flowchart

A flowchart of the ramp-rate control approach is given in Fig. 3. The PV panel output PPV is continuously monitored to detect any step change higher than the $\Delta P_{INV\text{-}des}$ for the given instant. If such a step change is detected, the controller determines the appropriate $\Delta P_{INV\text{-}des}$ depending on the event and then the required change in P_{ES} is obtained using (5).If the determined value of P_{ES} at the given k-th instant is positive, then SoC is checked to determine if it is higher than SoC_{min}. If "Yes", then discharge operation is performed; if "No", then the controller transits to the next time instant. On the other

hand, if P_{ES} is negative, the SoC is checked to determine if it is lower than SoC_{max}. Charging operation is performed if the result of the SoC status-check is "Yes", or transition to the next instant is initiated if the result is "No".

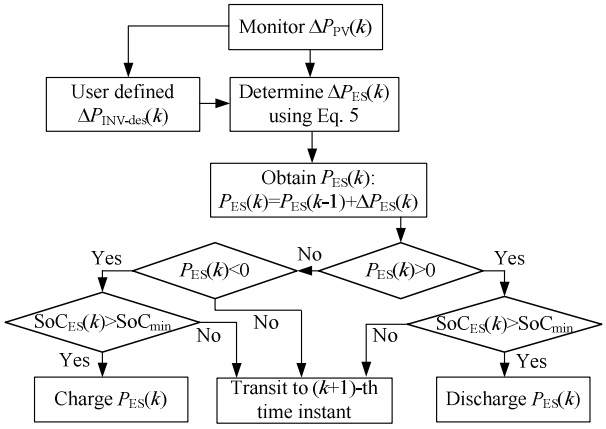


Fig. 3 Ramp-rate control flowchart.

C. Dynamic Model of the PV-Storage System

Power electronic systems and battery energy storage devices have time lags as any other physical devices. The smoothing operation of the ramp-rate control algorithm will be influenced by the associated time delays and controller characteristics. A dynamic model of the PV-storage integrated system is therefore necessary for investigating the transient response of voltage as a result of performing the ramp-rate control of the PV inverter. A dynamic model incorporating the time delays associated with the PV inverter and the energy storage device is developed for this purpose, as shown in Fig. 4 [11]. Details of the dynamic model including the state-equations and the method used for tuning the PI controller can be found in [11].

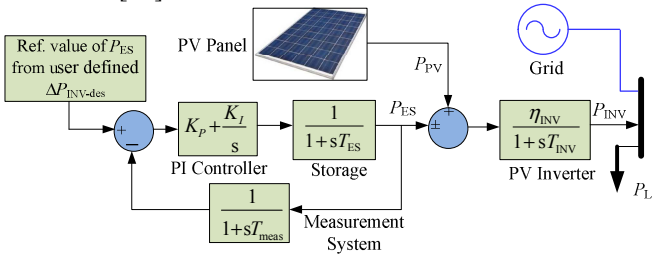


Fig. 4 A dynamic model of the PV-storage integrated system.

IV. Example of Voltage Fluctuation Mitigation in a Real Distribution Network

The benefit of PV inverter ramp-rate control in mitigating voltage fluctuation is investigated in this section using an Australian LV distribution feeder shown in Fig. 5. This LV feeder characterizes the features of a weak radial system located far from the main substation, as typically observed in rural Australian regions.

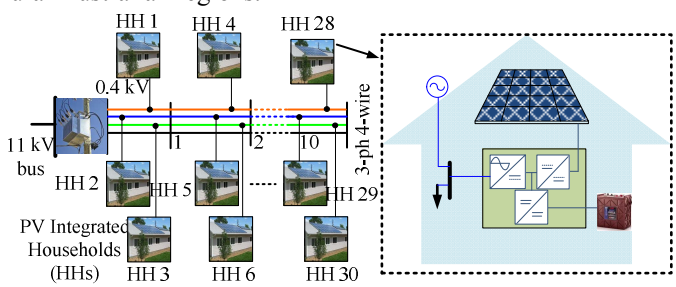


Fig. 5 Test distribution feeder.

All the households connected to the feeder are assumed to have rooftop PV systems with integrated battery storage devices. Valve Regulated Lead Acid batteries rated at 12 kWh is used for the investigations in this paper; SoC_{min} and SoC_{max} values are set to 40% and 100% respectively. The modelling of the storage devices incorporate the non-linear voltage versus SoC characteristics and the effect of internal resistance, as described in [12].

The results shown in this paper correspond to the household no. 28 (HH 28) connected to phase *a,* bus 10 of the feeder. The load demand profile is derived from the data captured by an Australian distribution utility and the PV output data correspond to the irradiance profile of 20th of March, 2010, captured by National Renewable Energy Laboratory's irradiance monitoring facility in the Oahu Island, Hawaii [13]. The normalized values of the PV output profile (P_{PV}) logged in 1-second resolution is shown in Fig. 6(a) that indicates severe power fluctuations occurred between 09:00 hours to 14:00 hours. The ramp-rate of PPV during this interval is shown in Fig. 6(b) where the highest negative ramp-rate caused by passing cloud is found to be -30%/second. A 5-minute interval around the highest negative ramp-rate event is considered for analyzing the effect of the ramp-rate control on voltage fluctuation.

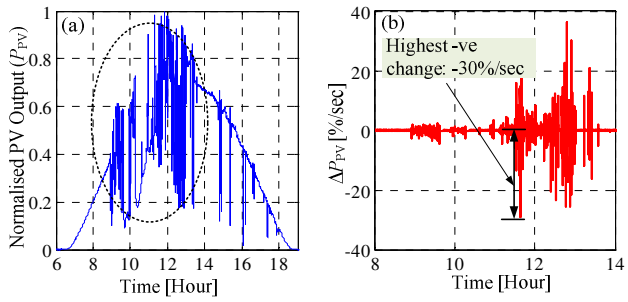


Fig. 6 Fluctuations in a real PV output profile. (a) Normalized P_{PV}. (b) Ramp-rate of PV output.

For illustration, the desired ramp-rate of the PV inverter ($\Delta P_{INV\text{-}des}$) is controlled at two values: 50 W/sec and 10 W/sec. The effects of the ramp-rate control are observable from Fig. 7. The P_{INV} profiles with and without ramp-rate control is shown in Fig. 7(a). Smoothing of the PV output fluctuation is clearly visible in Fig. 7(a), where the degree of smoothness increases with the decrease of the set value of $\Delta P_{INV\text{-}des}$ (from 50 W/sec to 10 W/sec).

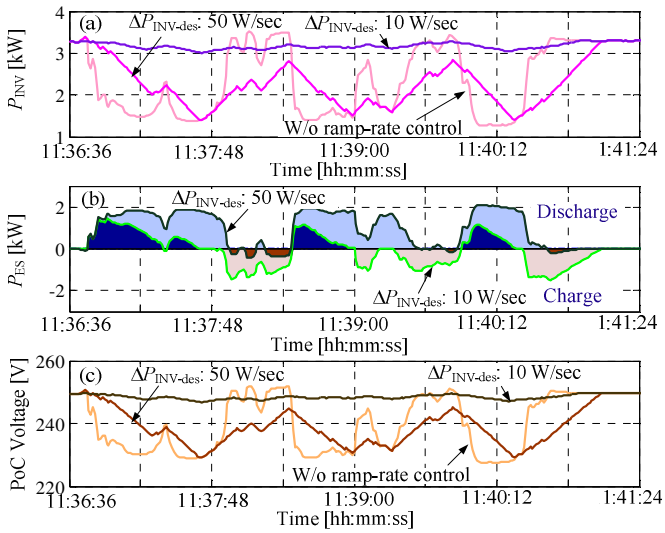


Fig. 7 Effect of the ramp-rate control on PV inverter output and PoC voltage. (a) Inverter power P_{INV}. (b) Storage power P_{ES}. (c) PoC voltage.

The storage power (P_{ES}) profiles that provide this smoothing effect are shown in Fig. 7(b) where the discharging

(+ve P_{ES}) and charging (-ve P_{ES}) operations required for ramp-rate control are identified. Finally, the mitigation of voltage fluctuation is shown in Fig. 7(c). Without the ramp-rate control, the voltage follows the similar trend of fluctuations as the PV output, and likewise, the voltage profile is smoothened with the application of the ramp-rate control.

As this paper focuses on the benefit of PV inverter ramp-rate control on voltage fluctuation mitigation, further investigations are performed on the results in Fig. 7(c). The rapid changes in the voltage profiles in Fig. 7(c) are plotted in Fig. 8(a) in terms of RCVI defined in (4). It is observed that without any ramp-rate control, RCVI can exceed the allowable limit of 3%. With the ramp-rate control, the value of RCVI is significantly reduced. A closer view of the highest negative ramp-rate event is provided in Fig. 8(b) where the decrease of RCVI from -4.7% to -0.2 % (for 50 W/sec) and to -0.05% (for 10 W/sec) is shown.

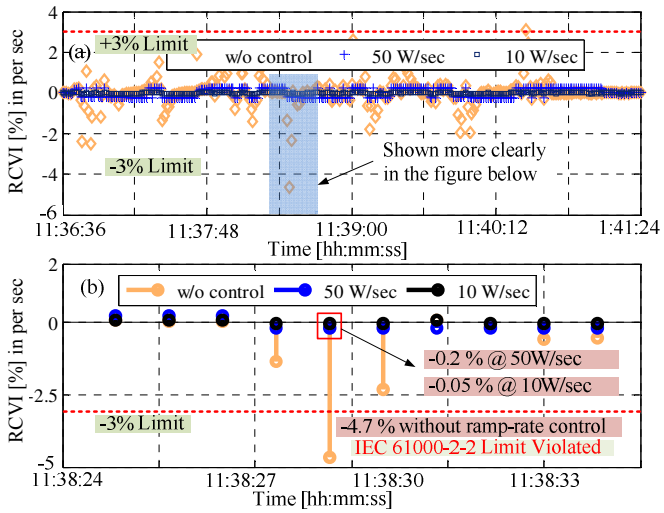


Fig. 8 Quantification of rapid voltage change with and without ramp-rate control. (a) RCVI over 5-min interval. (b) RCVI at the highest negative ramp-rate.

The dynamic performance of the ramp-rate control strategy in voltage fluctuation mitigation is investigated in Fig. 9. A 500 millisecond snapshot around the highest negative ramp-rate event of the PV output profile is used for this investigation. The dynamic model parameters (corresponding to Fig. 4) are given in Fig. 9(a).

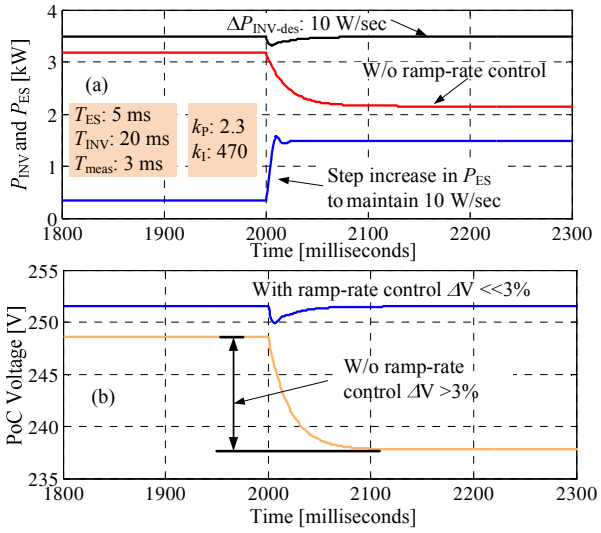


Fig. 9 Dynamic performance of the ramp-rate control. (a) Transient response of P_{INV} and P_{ES}. (b) Transient response of PoC voltage.

The transient response of P_{INV} with and without the ramp-rate control is shown in Fig. 9(a) that shows how P_{INV} maintains the desired value with the appropriate amount of power discharged by the battery storage. The transient response of P_{ES} resulting from the step change command by the storage controller is also observable in Fig. 9(a). The transient response of the PoC voltage a given in Fig. 9(b) that shows how the rapid voltage change is minimized with the help of the ramp-rate control of the storage device.

V. Conclusions

This paper has investigated the added benefit of the PV inverter ramp-rate control using an integrated battery energy storage system for the mitigation of unacceptable rapid voltage variations in distribution networks due to passing clouds. The amount of power delivered or stored by the energy storage device is appropriately controlled at each instant of time to maintain the inverter ramp-rate at a desired level. This results in mitigating the steep voltage variations at the inverter connection point. An analysis performed on a typical weak distribution feeder in Australia using captured PV outputs and measured load demands revealed that, without any support, the voltage variation can increase beyond the international standard limit. However, with a proper ramp-rate control, the variation can be maintained well below the limits.

References

[1] E. C. Kern, E. M. Gulachenski, and G. A. Kern, "Cloud effects on distributed photovoltaic generation: slow transients at the Gardner, Massachusetts photovoltaic experiment," *IEEE Transactions on Energy Conversion,* vol. 4, pp. 184-190, 1989.

[2] H. Sugihara, K. Yokoyama, O. Saeki, K. Tsuji, and T. Funaki, "Economic and Efficient Voltage Management Using Customer-Owned Energy Storage Systems in a Distribution Network With High Penetration of Photovoltaic Systems," *IEEE Transactions on Power Systems,* vol. 28, pp. 102-111, 2013.

[3] "IEC Standard 61000-2-2: Electromagnetic compatibility (EMC) - Part 2-2: Environment - Compatibility levels for low-frequency conducted disturbances and signalling in public low-voltage power supply systems ", ed: International Electrotechnical Commission, 28 March 2002.

[4] *Voltage Fluctuations in the Electric Supply System* [Online]. Available: http://www.elec.uow.edu.au/apqrc/content/technotes/technote7.pdf.

[5] "IEEE Application Guide for IEEE Std 1547, IEEE Standard for Interconnecting Distributed Resources with Electric Power Systems," *IEEE Std 1547.2-2008,* pp. 1-207, 2009.

[6] N. Kakimoto, H. Satoh, S. Takayama, and K. Nakamura, "Ramp-Rate Control of Photovoltaic Generator With Electric Double-Layer Capacitor," *IEEE Transactions on Energy Conversion* vol. 24, pp. 465-473, 2009.

[7] J. Traube, L. Fenglong, D. Maksimovic, J. Mossoba, M. Kromer, P. Faill, S. Katz, B. Borowy, S. Nichols, and L. Casey, "Mitigation of Solar Irradiance Intermittency in Photovoltaic Power Systems With Integrated Electric-Vehicle Charging Functionality," *IEEE Transactions on Power Electronics,* vol. 28, pp. 3058-3067, 2013.

[8] T. Kinjo, T. Senjyu, N. Urasaki, and H. Fujita, "Output levelling of renewable energy by electric double-layer capacitor applied for energy storage system," *IEEE Transactions on Energy Conversion,* vol. 21, pp. 221-227, 2006.

[9] T. Monai, I. Takano, H. Nishikawa, and Y. Sawada, "A collaborative operation method between new energy-type dispersed power supply and EDLC," *IEEE Transactions on Energy Conversion,* vol. 19, pp. 590-598, 2004.

[10] T. D. Hund, S. Gonzalez, and K. Barrett, "Grid-Tied PV system energy smoothing," in *Photovoltaic Specialists Conference (PVSC)*, Honolulu, HI, USA, 20-25 Jun., 2010, pp. 2762-2766.

[11] M. J. E. Alam, K. M. Muttaqi, and D. Sutanto, "A Novel Approach for Ramp-Rate Control of Solar PV Using Energy Storage to Mitigate Output Fluctuations Caused by Cloud Passing," *IEEE Transactions on Energy Conversion,* vol. 29, pp. 507-518, 2014.

[12] M. J. E. Alam, K. M. Muttaqi, and D. Sutanto, "Mitigation of Rooftop Solar PV Impacts and Evening Peak Support by Managing Available Capacity of Distributed Energy Storage Systems," *IEEE Transactions on Power Systems,* vol. 28, pp. 3874-3884, November 2013.

[13] NREL. *1-Second Global Horizontal Irradiance Oahu, Hawaii* [Online]. Available: http://www.nrel.gov/midc/oahu_archive/

Energy Storage for Rooftop Solar Photovoltaic Systems to Reduce Peak Demand

B. C. Teitzel,[1] M. H. Haque,[1*] and R. Inwood[2]

[1]School of Engineering, University of South Australia, Mawson Lakes, Australia

[2]Tindo Solar, Mawson Lakes, Australia

[*]Mohammed.Haque@unisa.edu.au

Abstract—**More than one in five homes in South Australia have a rooftop solar photovoltaic (PV) system and this number is expected to continually increase in the future. This paper investigates the possibility of storing the solar power generated during daytime and strategically releasing it during peak period to relieve stress on the national power grid. A small scale experimental setup with a battery storage is designed and developed for this purpose. The experimental results found are systematically described. The peak demand problem for South Australia power network and the challenges of reshaping the load curve through storing and releasing solar power are also discussed.**

***Index Terms*-Energy storage, load management, reshaping load curve, solar PV systems.**

I. Introduction

Adverse effects of greenhouse gases on environment have motivated the world towards reducing the use fossil fuels and increasing the utilization of renewable energy sources. Solar and wind are the fastest growing renewable energy technologies in the world. Both solar and wind are emission free renewable energy sources and are abundant in many parts of the world.

In South Australia (SA), the weather conditions are favourable for solar photovoltaic (PV) systems. With generous government subsidy, the installed capacity of rooftop solar PV systems onto residential houses in SA has increased exponentially over the past few years as can be seen in Fig. 1 [1, 2]. The total installed capacity of rooftop PV panels in SA reached 400 MW (as June 2013) and that contributes about 3.7% of the state's electrical energy demand [1]. It is expected that the capacity of rooftop PV system may exceed 1000 MW by 2023 [1].

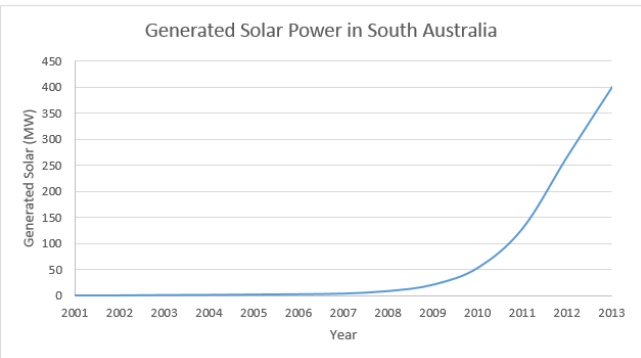


Fig. 1. Installed rooftop PV capacity in SA from 2001-2013.

The ratio of peak demand to average demand in SA is relatively high and this requires significant generation capacity and transmission/distribution facilities to meet the peak demand. However, with existing solar PV systems, an opportunity exists to reduce the peak demand by storing solar energy during daytime and strategically releasing the energy during peak period which usually occurs in early evening.

There are various forms of energy storage such as batteries, fly-wheels, super capacitors, compressed air, pumped hydro, etc. However, with present technologies, battery energy storage is possibly the most effective and efficient for small scale applications. Lead-acid batteries are commonly used in off-grid energy storage with solar PV systems and are also used in this study.

Numerous studies have been conducted on energy storage system. Giraud and Salameh [3] considered a 44-kWh of battery storage for a 2.5-kW solar PV and a 1.5-kW wind turbine system to power a grid-connected residential load during peak demand period. It is suggested that a hybrid system consisting of solar PV and wind generators is more reliable than a purely solar or wind system. Tina and Pappalardo [4] used a 200-Ah, 48-V battery storage system for a 2-kW solar PV system to power a grid-connected residential load. It is claimed that the system can control the amount of energy drawn from or injected into the grid and that allows the plant owner to save electricity cost by exploiting the time of use tariff. In addition, the system can assists to reduce system demand during peak periods. Siddique et al. [5] studied a rooftop hybrid system that stores solar and wind energy to use during peak period. It is encouraged to install solar and wind hybrid system to minimize the problems associated with peak demand. The system implemented was a 440-W solar PV system and 2x400-W wind turbines combined with 2x360-Ah, 12-V battery storage system. The above three systems have the capability of discharging the storage energy during peak demand period.

This paper analyses the peak demand problem of South Australia region and discusses the opportunity of exploiting the state's existing rooftop solar PV systems to reduce the peak demand. An experimental setup is build-up for this purpose and the results found are carefully analysed and discussed. The cost benefit analysis of the experimental setup is also outlined.

II. Minimizing Peak Demand

In an ideal electrical power system, the demand should remain more or less constant for better utilization of system assets. However, the electricity demand for most of the systems or regions is not constant but varies widely. In general, a system has low demand in the early morning hours and peak demand in the early evening hours. In order to cater the peak demand, the system should have adequate generation

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

capacity as well as transmission and distribution facilities and that contribute significantly to overall cost of electricity.

In SA, the peak demand usually occurs around 7pm, when most household appliances are in use. The peak demand occurs after sunset, when the solar panels cannot generate electricity. Thus, the entire demand is transferred to the traditional power plants of the national grid. Fig. 2 shows the average demand in 2008 (blue) and 2012 (red) in SA. The data of the figure is obtained from Australian Energy Market Operator (AEMO) website [6]. It can be noticed in the figure that during the day hours (8am to 6pm), the demand in 2012 is significantly lower than that in 2008. This is due to sudden start of installing rooftop solar PV systems at the beginning of 2008 and increasing exponentially each year (see Fig. 1) due to attractive financial incentive in the form feed-in tariffs and rebates [1]. Since the amount of installed solar PV system in 2008 was negligible compared to 2012 (as can be seen in Fig. 1), the area between the two curves (2008 and 2012) can be considered as the average daily energy supplied by solar PV systems in 2012 and was calculated as 1239 MWh. Note that load growth in SA between 2008 and 2012 was not significant.

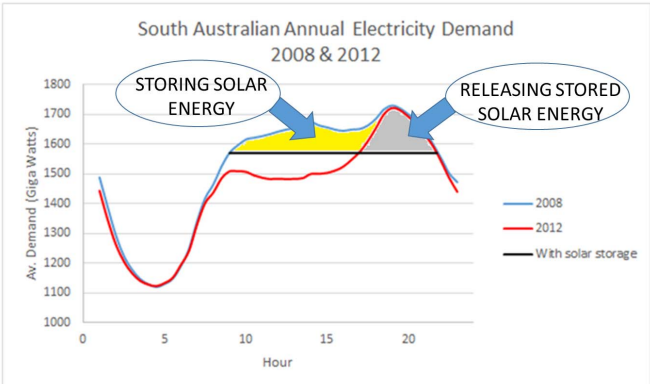


Fig. 2 Electricity demand in SA for years 2008 and 2012.

The objective of this study is to store some of the energy generated by solar PV systems during the daytime and strategically releasing it during the peak period to reduce the system stress and consequently flattening the peak demand. A simple calculation indicated that the state requires about 800 MWh of battery storage (represented by yellow area in Fig. 2) to flatten the peak demand. In practice, once the batteries are fully charged during the day, the PV systems can either supply power to the homes or feed into the grid, depending on the owner's preference and the configuration of the control circuitry. There are potential benefits for both the PV system owners and the power utilities.

The PV system owners would be benefited most if the anticipated smart metering system is introduced. The owners can manage their energy consumption and exploit the benefit of time-of-use tariff. In general, the cost of electricity is high during peak period. With sufficient PV energy storage systems, the owners may not need to pay such a high price and thus the system may eventually payoff itself or even save money. In addition, taking up of solar energy storage systems by a large number of customers would reduce the peak demand significantly and thus the utilities may not require the same level of generation capacity and transmission/ distribution facilities. This potentially means that the costs of electricity for the consumers may decrease or at least will not continue to increase at the same rate as in the recent years.

The main challenge for a solar PV energy storage is to develop a system that is cost effective with a relatively short payback period. The PV panels are becoming more cost effective each year as manufacturing techniques improve both the quality and efficiency of the panels. Energy storage is an emerging technology and does not yet have the same competitive price as the solar PV panels. However, with increased research and development programs in energy storages, it is anticipated that the energy storage system would follow a similar economic trend as the solar PV panels have seen. In fact, industry professionals are predicting that solar storage will be the new solar boom [7].

III. Experimental Setup and Results

In this study, a small-scale rooftop solar PV storage system is designed and developed to efficiently store solar energy during the daylight hours and release the energy during the peak hours. Fig. 3 shows a schematic diagram of the system with main components such as PV panel, battery storage, maximum power point tracker (MPPT), inverter, loads, and measuring instruments. A photograph of a part of the system is shown in Fig. 4.

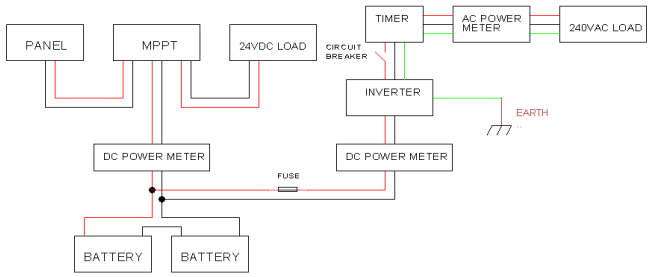


Fig. 3. Schematic diagram of the experimental setup.

Fig. 4. A photograph of the exerimental setup (excluding the PV panel).

One Karra-250 solar PV panel was used in the experimental setup. The panel has a rated output power of 250-W under standard test condition (STC) of irradiance 1000 W/m^2, solar spectrum of AM 1.5, and module temperature of 25°C. The panel was installed facing towards North at an optimal tilt angle of 30° for the site (Adelaide, South Australia, 34°55' 44.4'S 138°36'3.6'E) during experimental period. At the rated power, the output voltage and current of the panel are specified as 30.89 V and 8.10 A, respectively.

The solar irradiance of Adelaide, South Australia was downloaded from Bureau of Meteorology website [8]. The monthly average solar exposure of Adelaide is shown in Fig. 5 and it indicates that the average daily solar exposure in 2012 was 17.9 MJ/m^2 which corresponds to 4.97 kWh/m^2. That is the average peak sun hour (PSH) of the site is 4.97 hours/day.

The average daily energy yield for the panel (E_{PV}) can be obtained as

$$E_{PV} = P_{STC} \times \eta_m \times \eta_{dirt} \times H_{PSH} \quad (1)$$

Here P_{STC} is the PV panel output power under STC, η_m is the manufacture's specified output efficiency, η_{dirt} is the efficiency factor for dirt/dust on panel, and H_{PSH} is the average peak sun hour. For $\eta_m = \eta_{dirt} = 0.95$, the daily average output energy of the 250-W panel is found as 1.12 kWh/day. The temperature de-rating factor is not considered as it is assumed that the average annual day temperature of the site is around the standard test temperature of 25^0C.

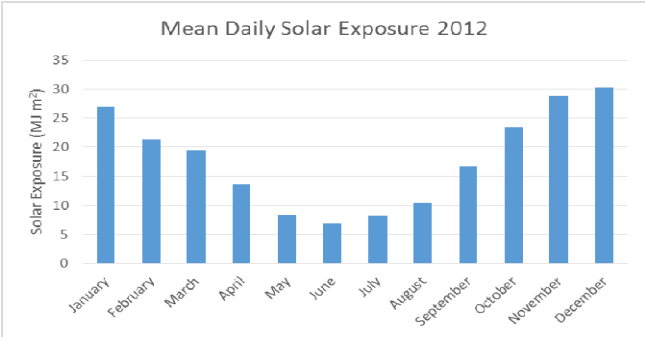


Fig. 5. The average daily solar exposure in 2012.

The efficiency of the panel (η_{PV}) can be defined as

$$\eta_{PV} = \frac{E_{PV}}{1000 A_{PV} H_{PSH}} \times 100\% \qquad (2)$$

Here A_{PV} is the area of the PV panel in m^2. The manufacturer's specified area of the panel is $1.73m^2$ and that provides an efficiency of the panel as 13.03%.

The size of the battery depends on the average daily energy generated by the PV panel and depth of discharge (DoD) of the battery. The DoD is directly related to lifespan of the battery. In this study, the battery size is determined based on 80% DoD (shorter lifespan and smaller capacity) as well as 50% DoD (longer lifespan and larger capacity). For a 24-V system, the minimum battery size of the system is found as 58.33 Ah for 80% DoD and 93.33 Ah for 50% DoD.

In this study, two 12-V, 38-Ah gel cell deep-cycle lead acid batteries are used. The batteries are connected in series to configure as 24-V to make compatible with the solar PV panel and the battery charger. Due to budget limitations, lower battery capacity was used compared to the calculated values of 58.33 Ah for 80% DoD, and 93.33 Ah for 50% DoD. The battery bank is housed in a wooden box and is sheltered from weather with appropriate ventilation. Since the charging voltage changes with temperature, the batteries have a temperature sensor attached so that the MPPT charger can continuously adjust the charge voltage according to the battery temperature to increase the battery life.

A microprocessor controlled MPPT battery charger with a current rating of 30 A is used. The charger is suitable for both 12V dc and 24V dc systems. The battery charge level determines the charging procedure is to be followed. The various charging procedures are bulk charging, absorption charging, and float charging which are performed automatically. The MPPT battery charger also has a deep discharge protection feature which protects the batteries and assists to increase the battery lifetime.

To convert the dc stored energy into ac energy (at 240 V), a modified sine wave inverter is used. The inverter is rated at 400-W for continuous load and 1200-W for surge load. The inverter has protection circuits such as a battery low voltage alarm and a thermal shut down with an alarm.

A total 400-W, 240-V incandescent light globes were used as the base load for this study. The load was limited by the inverter rating. Different loads such as fluorescent lights, water pump and water filter were also used to investigate the discharge times.

The system was monitored by a number of meters. DC meters were used in conjunction with a data logger to record the necessary data such as power, voltage, and current over time on the dc side. A 240 V ac meter was also used to monitor the quantities on the ac side of the system.

The experimental setup successfully stored the solar energy during the daytime hours and released it during the peak hours. With the measured data at various points, the efficiency of the battery storage and the dc/ac inverter was found as 79.69% and 90.11%, respectively. For a sunny day the system provided more than 725 Wh of energy, and for an overcast day the system provided less than 490 Wh of energy.

The inbuilt low voltage cut-off function in the dc/ac inverter is designed to protect the batteries against deep discharging (more than 44%). However, to further discharge the batteries, a 24V dc load was directly connected to the batteries, which in a real application could represent a 24V dc lighting circuit or something similar.

The undersized capacity of the deep-cycle batteries meant that the batteries would be fully charged before the end of the day's sunlight. To overcome this issue and get the maximum benefit, different operating modes of the system were considered which are discussed in the following.

Fig. 6 shows the power output from the MPPT battery charger during the daylight hours for a sunny day. Since the battery capacity is undersized, the batteries are fully charged by middle of the day and after that the MPPT battery charger is not extracting the maximum power from the panel. It is simply extracting enough power to float charge the batteries. Fig. 7 shows the power output form the MPPT battery charger during the daylight hours for a sunny day when the battery is allowed to discharge during the day. Unlike Fig. 6, the MPPT continues to extract the maximum power from the solar PV panel for the entire daylight hours as the batteries are fully charged only at the end of the day.

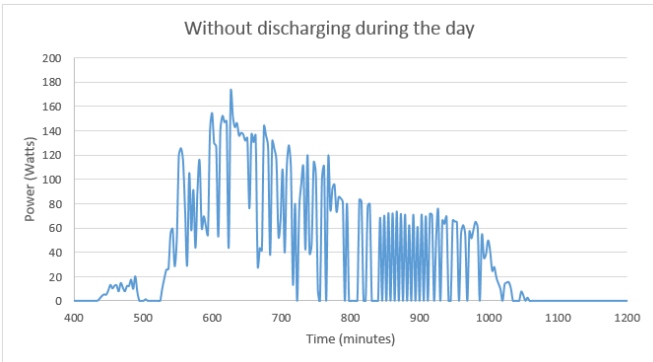


Fig. 6. Output power of MPPT without discharging the batteries.

To overcome the undersized battery issue, three operating modes of the system are defined. (a) For a low solar radiance day, the system can operate as charging the batteries during the daylight hours and discharging it during the peak hours, (b) For a high solar radiance day, the system can operate as both charging and discharging modes during the daylight hours and ensuring that there is enough stored energy to discharge during the peak hours, (c) For a high solar radiance day, the

system can operate as charging the batteries and powering a 24-V dc load during the daylight hours, and discharging the batteries during the peak hours.

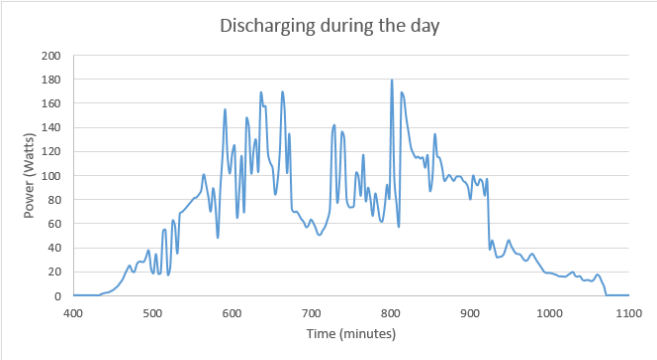


Fig. 7. Output power of MPPT when discharging the batteries.

IV. Cost Benefit Analysis

The experimental setup was a part of an undergraduate final year project and was designed as a small scale because of budget limitations. The system can deliver a power of 270 W during the peak demand for a period of 108 minutes (or 1.8 hours). With an electricity price of 30c/kWh, the system can save 14.6c/day. The cost of the components (excluding metering equipment) was A$865 and that translates to a simple payback period of 16.23 years which is longer than the life expectancy of the batteries. However, it is expected that the return would be better for a typical installation of about 3-kW size PV system. The indirect benefit of battery storage system is the reduction of system peak demand and hence utility installed capacity including transmission and distribution facilities.

The average Australian home consumes about 10-15 kWh of energy per day and the average South Australian 3-kW solar PV system produces about 12.6 kWh of energy per day [9]. Table-I summarizes the energy consumption of a typical Australian household during peak period (5pm to 8pm) [9].

TABLE 1
Typical Australian household load during peak period

Appliance	Power (W)	Hr. usage during peak (hrs)	Energy used during peak (Wh)
Kitchen Appliances			
Refrigerator/freezer	300	1.5	450
Oven	750	0.5	375
Stove	2000	0.5	1000
Microwave	1200	0.25	300
General Household			
Light – CFL (7 @ 12W)	84	3	252
LCD television	200	3	600
Laptop computer	30	2	60
Assorted electronics(10 @ 3 W)	30	3	90
Cloths washer	250	0.75	187.5
Air conditioning (2kW split)	2000	1.5	3000
Miscellaneous (10% of load)	---	---	631
	Total		6945.5

Considering the inverter efficiency of 85% and the deep-cycle lead acid battery efficiency of 80%, the minimum size of the energy storage system (ESS_{min}) required to deliver the energy during peak period can be calculated as

$$ESS_{\min} = \frac{E_{peak}}{\eta_{inv} \times \eta_{batt} \times DoD} \qquad (3)$$

Here E_{peak} is the energy consumption during peak period, η_{inv} is the inverter efficiency, η_{batt} is the battery efficiency, and DoD is depth-of-discharge of the battery. Using (3), the minimum energy storage system EES_{min} per household is found as 12.77 kWh. For a 24-V battery system, the size of the battery required is 532 Ah.

A typical 3-kW fully installed rooftop solar PV system with a 25 year warranty costs about A$4,500. A 15-kWh battery bank costs about A$6,000 (at a rate of A$400/kWh). Assuming the installation cost of the battery storage system to integrate it with the rooftop solar PV system is A$1,500, thus the total cost for the rooftop solar PV storage system becomes A$10,000. The average daily energy production of a 3-kW system in SA is 12.6 kWh and with the battery storage efficiency of 80%, the system can deliver 10.08 kWh of energy. With the current electricity price of abut 30c/kWh, the simple pay-back period of the system is 9.06 years. This is around the expected life time of the battery bank and presently does not seem to be overly cost effective. However, as solar storage technology develops, the above prices will come down and the payback period will become increasingly more attractive to perspective buyers.

V. Conclusions

More than 20% of Australian homes have rooftop solar PV systems. This paper investigates the possibility of storing solar PV energy during daytime and strategically releasing it during peak period to reduce system stress during peak demand. A small scale experimental set has been built as a final year project to demonstrate some of the experimental results. With declining trend of the cost of battery energy storage, the system may become more attractive in future. The addition benefit of the system is the reduction of peak demand and consequently reduction of system generation capacity as well as transmission and distribution facilities.

References

[1] AEMO, 'South Australian electricity report 2013', viewed 10 October 2013, http://www.aemo.com.au/Electricity/Planning/South-Australian-Advisory-Functions/South-Australian-Electricity-Report

[2] Clean Energy Council, 'Consumer guide to buying household solar panels', viewed 06 May 2013, http://www.cleanenergycouncil.org.au/resourcecentre/Consumer-Info/solarPV-guide.html

[3] F. Giraud and Z. M. Salameh, 'Steady-State Performance of a Grid-Connected Rooftop Hybrid Wind-Photovoltaic Power System with Battery Storage', IEEE Transactions on Energy Conversion, Vol. 16, No. 1, pp. 1-7, 2001.

[4] G. M. Tina and F. Pappalardo, 'Grid-Connected Photovoltaic System with Battery Storage System into Market Perspective', Proc. of the 1st IEEE-PES/IAS Conference on Sustainable Alternative Energy, pp. 1-7, Valencia, Spain, 2009.

[5] M. R. A. Siddique, S. J. Nibir and A. Hoque, 'Rooftop Hybrid System for Minimizing Peak Demand with an Effective Feed-In Tariff Structure', Proc. of the 2012 Students Conference on Engineering and Systems, Allahabad, Uttar Pradesh, India, pp. 1-4, 2012.

[6] AEMO, http://www.aemo.com.au/Electricity.

[7] Energy Matters, 'Deep cycle battery and state of charge', viewed 06 May 2013, http://www.energymatters.com.au/renewable-energy/batteries/battery- voltage-discharge.php

[8] Bureau of Meteorology, 'Daily global solar exposure', viewed 05 Aug 2013, http://www.bom.gov.au/jsp/ncc/cdio/weatherData/av?p_nccObsCode=203&p_display_type=dataFile&p_startYear=&p_c=&p_stn_num=023 0 11

[9] G. Parlinso, 'Rooftop Reshapes Energy Market in South Australia', Reneweconomy, viewed 28 March 2013, http://reneweconomy.com.au/2013/rooftop-solar-reshapes-energy-market-in-south-australia-18272.

Testing and Performance Analysis of Charge Controllers for Solar Home System

Manik Dautta*, Syed Mohammad Sifat Morshed Chowdhury, Md. Rashed Hassan Bipu, Md. Zulkar Nain and Shahidul Islam Khan

Department of Electrical and Electronic Engineering,
Bangladesh University of Engineering and Technology, Dhaka, Bangladesh
*mnkdautta8@gmail.com

Abstract— **Charge controller is the heart of a Solar Home System (SHS) which controls the charging and discharging of battery from photovoltaic (PV) module. This paper analyzes test results of fifteen charge controllers available in Bangladesh according to the specification of Infrastructure Development Company Limited (IDCOL). The different parameters are battery high voltage disconnect (HVD), low voltage disconnect (LVD)/ reconnect, short circuit protection etc. It is found that eight charge controllers meet the technical specifications of IDCOL. There is also a study of different features of the charge controllers. Finally a brief discussion on further improvement of charge controller is presented.**

Index Terms— **Solar Photovoltaic, Solar home system, Charge controllers, Testing, IDCOL.**

I. Introduction

A solar charge controller is needed in virtually all solar PV systems that utilize batteries. The job of the solar charge controller is to regulate the power going from the solar panels to the batteries. The most basic function of a charge controller is to prevent battery overcharging. If battery is allowed to routinely overcharge, their life expectancy will be dramatically reduced [1]. A charge controller will sense the battery voltage, and reduce or stop the charging current when the voltage gets high enough. This is especially important with sealed lead acid battery where we cannot replace the water that is lost during overcharging [2]. There are some other functions that a charge controller performs such that preventing battery over-discharge, protecting from electrical overload, and maximum power tracking etc.

To fulfil the basic electricity requirement in the rural areas of Bangladesh, IDCOL has introduced solar program. With support from International Development Association (IDA) and Global Environment Facility (GEF), IDCOL provides both grant and refinancing to solar energy projects in rural areas. IDCOL provides grants to lower costs of solar home system (SHS) and also offers soft loans to the partner organizations. To coordinate the program, Technical Standard Committee (TSC) has been formed with members from Bangladesh University of Engineering and Technology (BUET), Rural Electrification Board (REB), Local Government Engineering Department (LGED) and IDCOL. TSC determine the technical standards for the equipments such as solar panel; charge controller, battery etc. of solar home system. The equipment that is not approved by TSC is ineligible for IDCOL grants and refinance. In local market all available charge controller are not approved by IDCOL. Collected charge controllers are tested according to the standards given by IDCOL.

II. Charge Controller Tests And Analysis

For a solar system, the controller must comply with the technical standards provided by TSC of IDCOL in Bangladesh. The sample charge controllers are taken from Partner Organizations (PO) of IDCOL. Some foreign manufactured charge controllers are also collected for testing purpose. Performances of these sample charge controllers are measured to inspect the conformity to the specifications provided by the IDCOL [3]. Low voltage disconnect /reconnect(LVD), Battery high voltage disconnect, current handling capacity, self current consumption, short circuit protection, over current protection etc. were measured.

A. Low Voltage Disconnect (LVD)

If deep discharge occurs, it will cause some battery damage. Every time this happens, both the capacity and the life of the battery will be reduced by a small amount. If the battery sits in this over discharged state for days or weeks at a time, it can be ruined quickly. So, LVD is necessary to protect the battery.

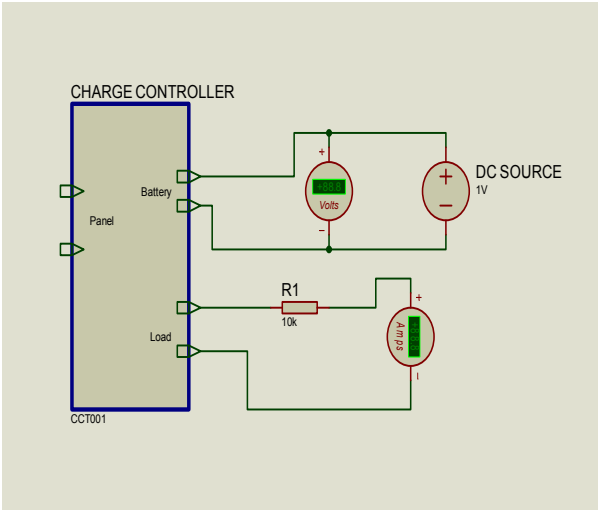


Fig. 1 Circuit connection for LVD disconnect/reconnect test

For measurement of LVD disconnect/reconnect voltage circuit was connected as Figure 1.

1) *Disconnect Voltage*: When the state of charge of battery is reduced below the specified level then charge controller disconnects load from battery. Test result of different samples is shown graphically in Figure 2.

According to the IDCOL standard, LVD disconnect voltage is 11.6±0.1 V. So the sample no.04, sample no.09 and sample no.12 did not comply with the specification.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

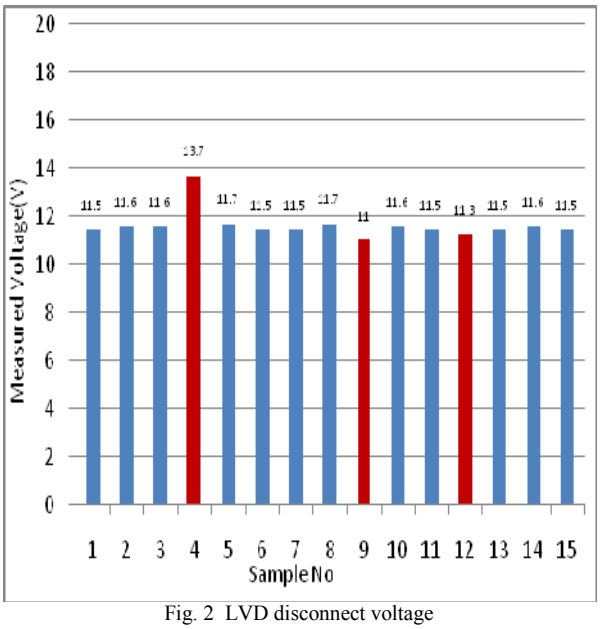


Fig. 2 LVD disconnect voltage

2) *Reconnect Voltage*: When battery voltage rises up to a specified level then charge controller automatically reconnect the load with battery. Reconnect voltage for different samples is shown in Figure 3.

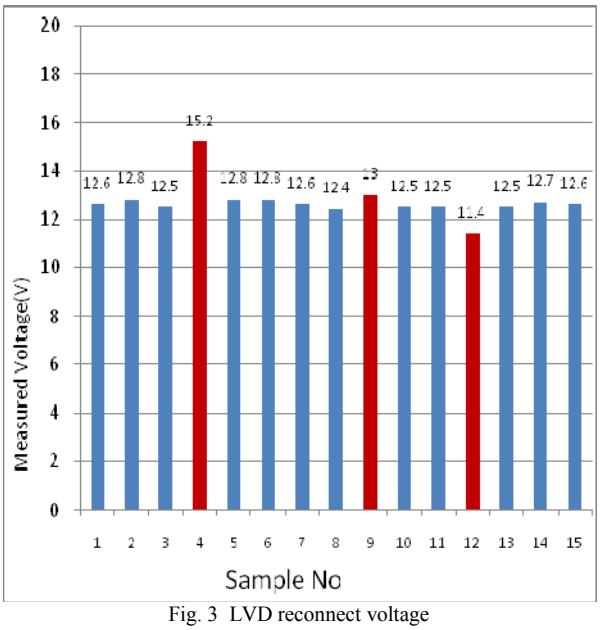


Fig. 3 LVD reconnect voltage

According to IDCOL standard, LVD reconnect voltage is 12.6±0.2 V. So the sample no.04, sample no.09 and sample no.12 did not comply the specification.

B. High Voltage Disconnect (HVD)

When a battery reaches full charge, it can no longer store incoming energy. If energy continues to be applied at the full rate, the battery voltage gets too high. There is excessive loss of water, and a chance that the gasses can ignite and cause a small explosion. The battery will also degrade rapidly and may possibly overheat. Preventing overcharge is simply a matter of reducing the flow of energy to the battery. That is why at a specified battery voltage the battery is disconnected from solar panel to stop current flow for charging.

The circuit was connected as Figure 4. The DC source voltage was higher than the battery and supplying sufficient current. When the current reading of Ammeter was 0Amp, reading of the Voltmeter was taken.

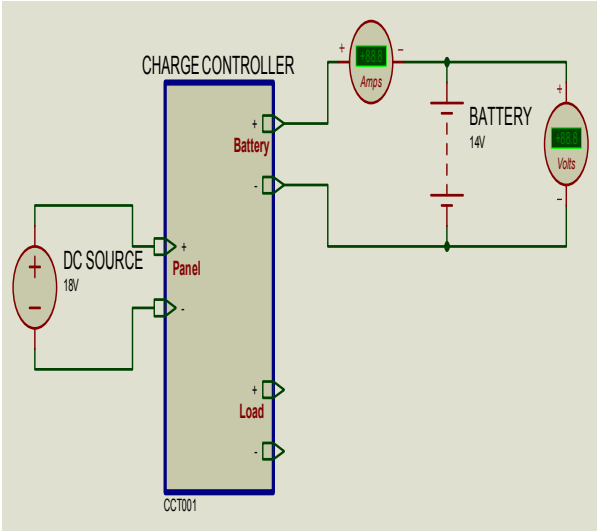


Fig. 4 Circuit connection for HVD test

Measured battery high voltage disconnect voltages are shown in Figure 5.

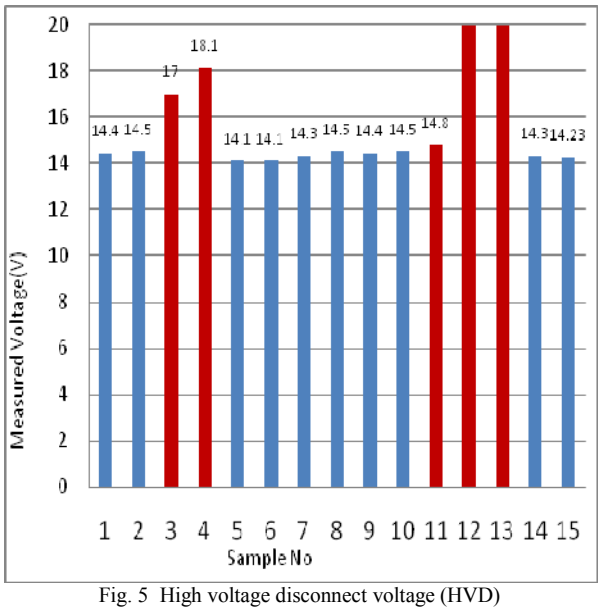


Fig. 5 High voltage disconnect voltage (HVD)

According to the IDCOL specification, HVD is 14.3±0.2 V. From graphical representation it is seen that sample no.03, sample no.04, sample no.11, sample no.12 and sample no.13 didn't comply this. Although sample no.03, sample no.04 and sample no.11 did not comply with the specification they have high voltage disconnect. But sample no.12 and sample no.13 did not have HVD feature.

C. LVD Maximum Current Handling Capability

According to IDCOL standard, The LVD must be capable of handling at least 150 percent of the maximum expected continuous load. From the analysis it has been found that 150 percent of maximum expected continuous load was attained at a slightly higher voltage than LVD voltage for all the sample charge controllers. This complies with the standard given by TSC.

Circuit was connected as Figure 6. Keeping the DC source voltage at 11.7 V the Pot1, Pot2 and the Current knob of DC

source were varied until the Ammeter reading became 150% of the maximum expected load. Then the charge controller was observed whether it was capable of carrying this load.

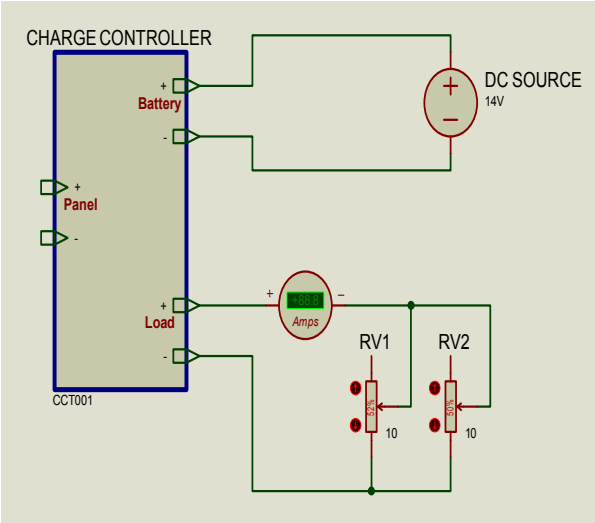


Fig. 6 Circuit connection for measuring current flow

All samples complied with IDCOL standard.

D. Over Current Protection

There must be a protection against excessive current flows through the load to protect it from damage. If not, then the circuit of charge controller will be damaged.

Resistances of both the potentiometer of Figure 6 were decreased so that total Resistance became very low. An Ammeter reading was taken at the moment of Load Out.

Testing results are shown in Figure 7.

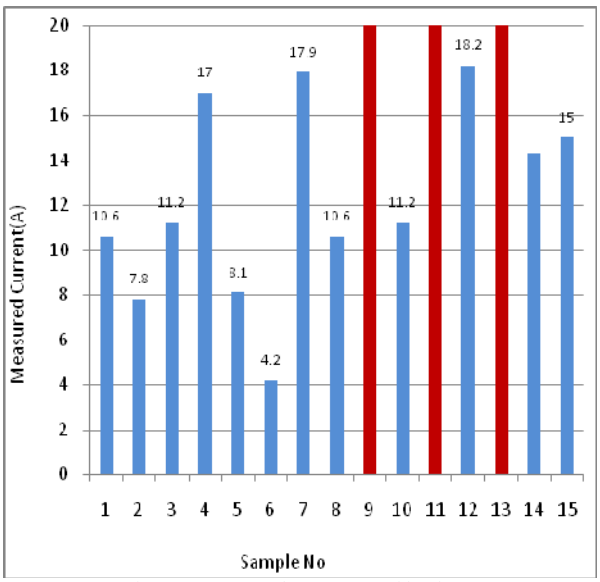


Fig. 7 Current at the moment of load out

As current rating for different charge is different, scale is not unique. From graphical representation it was seen that sample no.09, sample no.11 and sample no.13 did not have over current protection.

E. Current Handling Capability

According to IDCOL standard charge controller must handle 120% of rated current for 1 hour without any damage.

Circuit was connected as Figure 6. Keeping the DC source voltage at 14 V, Pot1, Pot2 and the Current knob of DC source were varied so that the Ammeter reading became 12A. Observation was made for 1 hour.

Over current protection for sample no.02, sample no.03 and sample no.04 were below the 120% of rated current. So they didn't comply with this standard but they have set points at different value.

F. Self Current Consumption

Every electrical device having MOSFET, BJT requires some current for internal circuitry. For charge controller self current consumption should be less then 20mA when no LED is on and it should be less then 50mA when one LED is on.

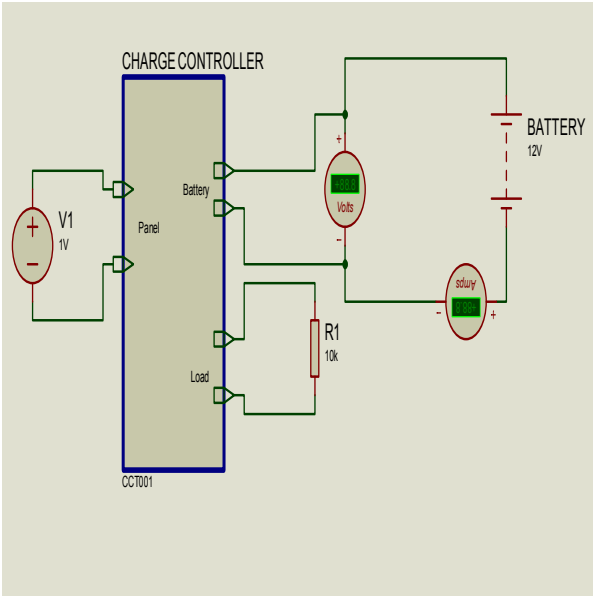


Fig. 8 Circuit connection for measuring self current consumption

Voltage at the panel side of Figure 8 was increased to maximum limit at which no LED was on and reading of the ammeter was taken. Then it was increased further so that one LED was on and reading of the ammeter was taken.

All samples complied with this specification.

G. Short Circuit Protection

Short circuit can cause overheating and can even be a fire hazard. Short circuit protection is mandatory at the load terminal according to the IDCOL standard. If load is short circuited then charge controller disconnects the load.

Circuit was connected as Figure 4 where load is short circuited by a thick wire having low resistance. The Load was out for satisfactory protection.

To protect load as well as charge controller unit every sample should include short circuit protection. Sample no.09 sample no.12 and sample no.13 didn't comply with this specification.

H. Reverse Polarity Protection

There may be opposite polarity connection at the panel side and at the battery side. IDCOL specifies it as a mandatory to have protection against reverse polarity so that no current flows from any direction.

The circuits were connected as Figure 9 and the Charge controller was observed whether there was any damage.

All samples complied with this specification.

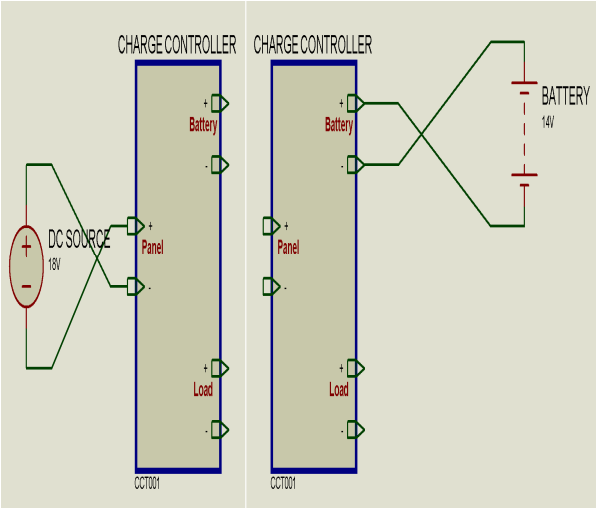


Fig. 9 Circuit connection for reverse polarity protection

I. *Reverse Current Leakage Protection*

PV works by pumping current through battery in one direction. When battery voltage goes higher than panel side voltage, the panels may pass a bit of current in the reverse direction, causing a slight discharge from the battery.

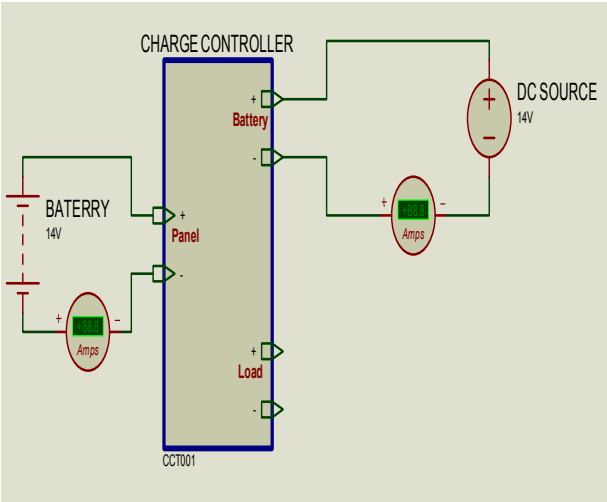


Fig. 10 Circuit connection for reverse current leakage protection

The DC source voltage of Figure 10 was higher than the battery voltage. Reading of both Ammeters were taken. Zero or very low reading indicates the protection was OK on both Panel and Battery side. According to IDCOL this leakage current should be 0Amp or very low.

All samples complied with this standard of IDCOL.

III. Features Of Observed Charge Controllers

According to connection, charge controllers can be of two types such as shunt controller and series controller. The series charge controller is connected in series between the solar panel and battery. The controller disconnects the battery from the panel by opening the switch. But shunt controller is connected in parallel to the battery and the load. When battery is fully charged, the controller interrupts current supply to the battery by making the panel short-circuited [1]. Among the charge controller we tested, all except sample no. 2 were series controllers. Sample no. 2 was shunt controller.

There are three types of charge controller such as ON-OFF, PWM and MPPT based charge controller. After testing the charge controller we found that, sample no. 5, 6 and 8 are PWM controllers. Others are On-off controllers.

Some extra features were observed during the testing process. Sample no. 5 has acoustic load disconnect pre-warning which sounds an alarm before disconnecting the load. Sample no. 6 has a built in cell phone charger which is very useful in the area where grid electricity is not available. Sample no. 6 and 7 are built using surface mounted technology which promises higher reliability and performance.

IV. Further Improvement

Among the tested charge controllers, only three were PWM charge controllers, where others are ON/OFF controller. Literature [4], [5] shows that PWM charging increases the charging efficiency and battery life. So, this technique should be implemented in the charge controllers. Also it was found that in some controllers, negative terminals of the panel and the battery have to be shorted for starting the controller. This common negative grounding feature helps to detect wrong connection at any of the terminals. This feature may be included in all the charge controllers.

It was also found that locally built charge controllers do not have maximum power point tracking (MPPT) system. This can increase the output of the solar panel from 10 to 30 percent. Though it needs an efficient DC-DC converter and capable microprocessor which are costly, it may be cost effective for larger systems.

V. Conclusions

Solar Home System (SHS) has become an effective and viable solution for the current energy crisis in Bangladesh. As a major component of solar home system, the performance of charge controller plays a vital role in the growth and popularity of solar PV system. In this paper, we randomly collected fifteen charge controllers available in the local market produced by both local and foreign manufacturer. We systematically tested these charge controllers and compared against the specification set by IDCOL. The testing data are presented in this paper. Some deviations were found from the standard values which can cause severe underperformance of solar home system. It was found that efficiency and performance of PWM charge controller is better in Bangladesh. Future work should incorporate large scale testing and evaluation of the locally available charge controllers at regular interval.

References

[1] Dr. Md. Saifur Rahman, Dr. Shahnawaz Ahmed, Dr. S.I. Khan et al, *Solar Home System*, 1st ed., Dhaka, Bangladesh, German technical Cooperation (GIZ), March 2013..

[2] S.A. Khan, R. Rahman and A. Azad, "Solar Home System Components Qualification Testing Procedure and Its Effect in Bangladesh Perspective," IEEE Global Humanitarian Technology Conference (GHTC), pp. 381 – 386, October 2012, Seattle, WA.

[3] IDCOL (2012), Technical Specifications for Solar Home System (SHS), Available at: http://www.idcol.org/publications.php , (Accessed: 18 December 2013).

[4] Lam, L.T., et al, "Pulsed-current charging of lead/acid batteries-a possible means for overcoming premature capacity loss" CSIRO, Australia, Journal of Power Sources 53, 1995.

[5] Hund, Tom, "Battery Testing for Photovoltaic Applications," Sandia National Laboratories, Albuquerque, NM, presented at 14th NREL Program Review, Nov. 1996.

Study and Analysis of Existing Solar PV System in Urban Area of Bangladesh

*Shammya Shananda Saha, Mohammad Noor Bin Shaheed, M. Mahmudul Hasan Sajeeb,
Tasnuva Hamid, Shahidul Islam Khan
Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology, Dhaka–1000, Bangladesh
Email: *shammya_eee@mail.com

***Abstract*—The fossil fuel is decreasing rapidly and excessive use of such fuels is also hazardous for the environment. Hence, renewable energy usage is the inevitable alternative to meet up electricity demand for the developing country like Bangladesh. Although solar technology has nearly been successful in rural areas where most of the technologies are adopted based on Solar Home System(SHS), it has not yet been effective in urban areas after the imposed rule of installing solar system to get new electricity connection from power utility for various reasons. In this paper, the investigation result of the installed solar system in an area in Dhaka is presented. The overall problems regarding the connection of the system components are analyzed. Efficiency calculation for the active solar home systems is provided through cost comparison analysis by HOMER software. Some proposals are provided which can be implemented to make the concept of solar home system popular in urban ares of Bangladesh to decrease the pressure on our national grid.**

***Index Terms*—HOMER, COE , solar, urban, survey.**

I. Introduction

Energy crisis is regarded as one of the widely discussed and vastly concerned issues that is causing acute controversy as presently used conventional fossil fuels are the greatest pollutants and largest contributors of greenhouse gases. Rapid depletion of fossil fuels and enhanced environmental degradation has resulted in to the extensive use of renewable power sources for electric power generation. In developing country like Bangladesh, it is therefore necessary to investigate the prospect of renewable energy sources especially solar energy that do not pollute environment, is clean and inexhaustible.

Lying in semi-tropical region of the north eastern part in South Asia, Bangladesh is prone to abundant sun light round the year which is very favorable for solar PV based electricity production. Understanding the future prospect and many diverse opportunities of solar technology, Government of Bangladesh along with other concerned organizations are trying to comprehend and address the problem of energy scarcity through promoting and popularizing solar home systems (SHS) in both the off grid and on grid areas of the country.

According to the Renewable Energy Policy as per Ministry directive to all the utilities, all new connections with demand above 2KW now require installation of solar panels. For domestic purposes, solar energy should meet at least 2% of the demand, for commercial 7% and for industrial 10%.

Proper inspection of the system is always a must to ensure that the imposed rule is making the people benefited.It is an irony of fate that a government survey [1] recently pointed out that installation of solar panel on rooftop in urban areas would not add any additional power in the system rather it would create serious havoc in system management and cause huge financial loss to the consumers.The survey also pointed out that solar panel failed to catch public interest as the distributing agencies did not take any awareness program for the consumers as well as they faced serious obstacle to get service from the agencies. Considering the results from the survey, a detail survey was done on Lalmatia Area, Dhaka to inspect the working status of the installed solar systems. Some of the houses were DPDC (Dhaka Power Distribution Company) listed and others were not. Necessary possible data was measured from panel and battery. User feedback was also collected from every possible individual. The data collected was used to differentiate the effective and ineffective systems. The efficiency calculation was carried out using the modified parameters of HOMER software.

II. Details of the Survey

23 houses of Lalmatia were inspected for surveying purpose. Of those, only 6 of the houses were found having active solar home system.The inactive systems contains problems like mismatching of voltage levels of solar panels, incorrect connection of battery, oversized inverter with respect to load etc. Table I shows the data collected from the houses who have installed solar home system. Here

- Panel Voltage refers to voltage of the solar panel measured at Charge Controller (CC) terminal.
- Panel Current refers to incoming current from panel through CC.
- Battery Voltage refers to voltage of the battery measured at CC terminal.
- Battery Current refers to outgoing current to battery through CC.
- Inverter Voltage refers to AC output voltage of the inverter.
- NM, ND, NC stands for Not Measurable, No Data and No Connection respectively.

Here the comma separated values in the first column of Table I represent the voltage values for panels of different power and the comma separated values in the third column represent the voltage values for separate battery.

III. Discussion regarding the Inspected Houses

Specific problems were found out regarding the inspected houses like on arrangement of battery and solar panels and circuit connection through the CC.From Table I following conclusions can be made :

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

TABLE I
COLLECTED DATA FROM THE HOUSES

Sl no.	Panel Voltage		Panel Current	Battery Voltage		Battery Current	Inverter Voltage
	Panel Bank 1	Panel bank 2		Battery Bank 1	Battery Bank 2		
1	58.25V	-	.1A	Damaged		Damaged	0V
2	88.2V	-	0.9A	54.4V	-	0.3A	0V
3	13.7V	-	1.47A	12.72V	-	.32A	off
4	38.4V	-	.17A	12.54V	13.79V	0	0V
5	ND		ND	ND		ND	0V
6	41.25V	-	0.37A	Damaged		Damaged	0V
7	26.32V	-	2.54A	NM		2.30A	240V
8	13.92V	-	NM	13.14V	-	2.2A	230V
9	28.4V	-	5.5A	27.66V	-	5A	230V
10	31.24V	-	12.85	30.06V	-	12.13A	230V
11	38.4V	-	.18A	27V	-	3.92A	230V
12	36.8V	-	.35A	27.32V	-	0.19A	230V
13	19.32V	-	0.1A	NM		0.03	NM
14	13.32V	-	1.09A	NM		0.11A	254V
15	ND		ND	ND		ND	ND
16	12.41V	-		12.30V	-	3.1A	NM
17	14.75V	19.13V	3.4A	13.99V	13.9V	2.98,0	0V
18	38.8V	-	0.22A	inconsistent		0.1A	0V
19	17.56V	10.85V	0.24A	28.51V	-	0.2A	0V
20	ND		ND	ND		ND	ND
21	78.4V	-	0A	49.4V	-	0A	0V
22	NC		NC	NC		NC	NC
23	45V	-	0A	0V	-	0A	0V

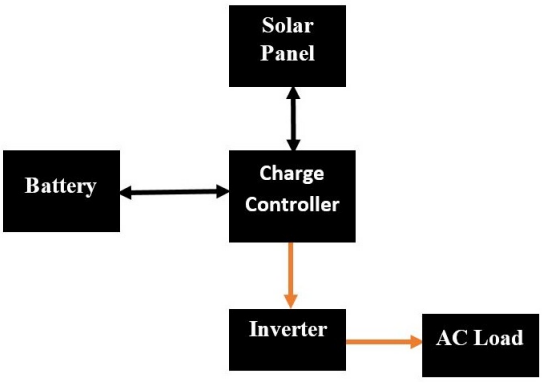


Fig. 1. Appropriate Connection among SHS Components

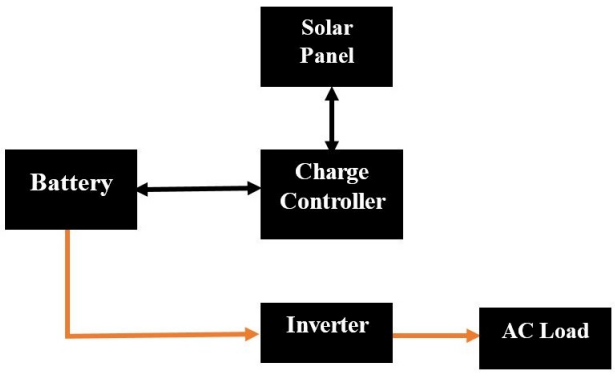


Fig. 2. Typical Connection found among SHS Components

- Only 6 of the 23 houses have active solar home system which is only 26.08 % of the total number of samples. Only 1 of the 6 samples had almost flawless system while the other ones had some design problems.
- For solar home system, a charge controller controls the charging and discharging level of a battery requiring a circuit connection between components shown in Fig. 1. But every inspected house has a circuit connection like Fig. 2, causing no control over battery discharging .
- Battery of different Ampere Hour (Ah) were found to be both in parallel and series configuration which is detrimental to battery. To increase the input DC voltage to inverter different ampere hour battery were found in series configuration where as different ampere hour battery was found in parallel configuration for increasing the total capacity. Figure 3 shows 6 130Ah batteries connected in a combination to match the input voltage of the inverter which is detrimental for the upper two batteries (Bat5 and Bat6).

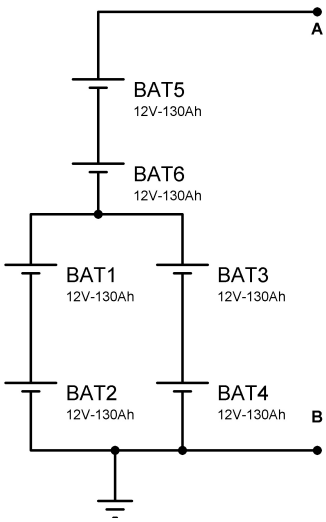


Fig. 3. Connection Found Between Batteries

- An inverter has to produce its no load voltage whether load is being served it or not. But most of the inverters were found off during the inspection time.
- From the observation of the systems, it is found that the charge controller and the inverter are oversized for that specific house. The installer company used an already available CC with an already produced inverter irrespective of the amount of the load the inverter has to handle. Due to this reason, most of the houses were manually operated by a person who had the responsibility of switching on and off the load which is not an efficient system and also responsible for battery damage.
- Panels of two different open circuit voltage were found out in parallel combination in some houses causing the solar panel of low open circuit voltage acting as load.

IV. Analysis of Active Systems & Efficiency Calculation

From the collected data, it was found that houses corresponding to serial no 8,9,10,11,12 and 16 of Table I had active solar home systems. But the systems were not designed in an efficient way. Moreover, some connection problems were also found regarding the connection of the solar panel. Every active houses were simulated in HOMER software and efficiency was calculated through detail analysis of the system.

A. Modification of HOMER Software

As this paper work has a purpose to design the Urban Solar System by using the local panels and batteries it is necessary to make the simulation software HOMER compatible to the available panels and batteries in Bangladesh. So, to start over the simulation work, necessary battery models were made in HOMER.

1) Battery Modeling in HOMER: HOMER has a built in battery resource but the batteries found are not available in the Bangladesh market. So it is necessary to remodel the batteries according to the availability of the batteries. The simulation process is performed after modeling the battery parameters to match with the performance of the 80Ah,100Ah and 130Ah batteries available in Bangladesh market.Capacity curve modeling was performed according to [3], [4] and the lifetime curve was modeled by collecting data from [2].Thus the battery modeled (Fig. 4)suits the present economic situation and is suitable for cost estimation of the investigated system.

2) Setting the Required Parameters: Table II shows the considered specifications for the purpose of efficiency calculation by HOMER.

After a market survey, battery and panel pricing were collected and are used for calculating the efficiency of the system. Inverter parameters were not found for all the installed systems. In such cases, the lowest required inverter capacity for meeting up the load is used which was being found by sensitivity analysis.

B. Defining Efficiency

For high watt load HOMER is used to measure the efficiency of the installed solar home systems in the active house. The efficiency of the system is measured by following the steps described.

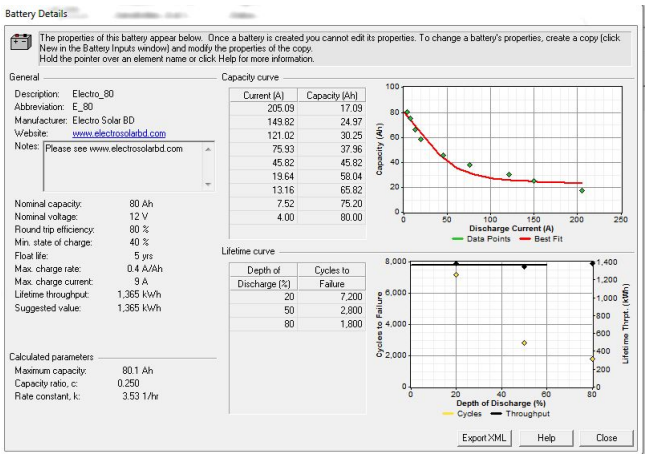


Fig. 4. Battery parameters for 80Ah Battery

TABLE II
Specifications used for Effeciency Calculation for Active Systems

Panel Lifetime	20 years
Derating Factor	80.00%
Slope	23.12
Ground reflectance	20%
Nominal operating cell temperature	25
Efficiency at standard condition	13.00%
Fixed Capital Cost	$100.00
Annual Interest Rate	6.00%
Inverter Efficiency	75.00%
Battery Lifetime	5 years
Inverter Lifetime	10 years
Inverter Price	$158 per kW
O&M Cost of Battery	$10

1) From the installed panel and battery of the house and the present running load, the COE (Cost of Energy) is calculated.
2) In the HOMER software, simulation is run by varying the load but keeping the battery and solar panel exactly same to the installed system with a purpose of finding the maximum possible load which can be run by the installed panel and battery. The simulation thus produces a minimum COE with a maximum possible load which can be served.
3) The relative change of COE in % returns the percentage by which the present system is inefficient.
4) The final efficiency of the system is = (100 − Relative Change of COE) %.

C. Efficiency Analysis for Sample No 10

1) Load specification: From the discussion with in-charge personnel of the system, it was found that some of the garage and stair case lights are PV-system driven and the load connected is approximately around 190W.

2) Battery and Panel Specification: Each battery is a 12V, 80 Ah one. Price of each battery is considered to be $127. There are 4 batteries in the system; two strings of parallel battery, each string containing two batteries in series . There were 28 panels each of 50W (each costs $45)resulting in a total panel size of 1400 W. These exact parameters are provided as input for the simulation purpose.

3) Efficiency Calculation: Figure 5 shows that 760Wh load is daily used in this house with a COE of $1.126 per

Pri. Load 1 (kWh/d)	PV (kW)	E_80	Conv. (kW)	Initial Capital	Operating Cost ($/yr)	Total NPC	COE ($/kWh)	Ren. Frac.
0.770	1.4	4	0.3	$ 2,002	138	$ 3,583	1.112	1.00
0.780	1.4	4	0.3	$ 2,002	138	$ 3,583	1.097	1.00
0.790	1.4	4	0.3	$ 2,002	138	$ 3,583	1.083	1.00
0.800	1.4	4	0.3	$ 2,002	138	$ 3,583	1.070	1.00
0.810	1.4	4	0.3	$ 2,002	138	$ 3,583	1.057	1.00
0.820	1.4	4	0.3	$ 2,002	138	$ 3,583	1.044	1.00
0.830	1.4	4	0.3	$ 2,002	138	$ 3,583	1.031	1.00
0.840	1.4	4	0.3	$ 2,002	138	$ 3,583	1.019	1.00
0.850	1.4	4	0.3	$ 2,002	138	$ 3,583	1.007	1.00
0.860	1.4	4	0.3	$ 2,002	138	$ 3,583	0.996	1.00
0.870	---	---	---	---	---	---	---	---
0.880	---	---	---	---	---	---	---	---
0.890	---	---	---	---	---	---	---	---
0.900	---	---	---	---	---	---	---	---
0.760	1.4	4	0.3	$ 2,002	138	$ 3,583	1.126	1.00

Fig. 5. Sensitivity Analysis in Tabular Form for Sample No 10

kWh(marked in red box in Fig. 5). But simulation results shows that maximum 860Wh load can be served by the installed solar home system with a COE of $0.996 per kWh (marked in green box in Fig. 5).

Efficiency of the sample=$100 - \frac{(1.126 - .996) \times 100}{.996} = 86.94\%$

Similar efficiency calculation is performed for the other active systems and presented in the Table III.

TABLE III
EFFICIENCY CALCULATION FOR THE ACTIVE SYSTEMS

	Serial No				
	8	9	10	11	12
Present Load(Wh)	400	680	760	600	800
Present COE($/kWh)	1.36	0.89	1.13	1.29	1.23
Optimized Load(Wh)	520	750	860	780	1150
Optimized COE($/kWh)	1.047	0.759	0.996	1.003	0.859
Efficiency (%)	70.11	83.34	86.94	71.09	56.81

V. PROPOSALS FOR IMPROVING THE PRESENT SCENARIO

- Feed In Tariff implementation is a path to success to improve the present condition. But as the necessary money that the Government has to pay is too high [6], Meter tempering and and corruption preventive steps will be required.
- For small load system,DC wiring can be successfully implemented using DC loads only [5]. But in case of distant loads the copper loss will be of significant amount.So inclusion of Optimizer (Boost Converter incorporated in a CC) can be a solution to such problems.
- Necessary schemes can be designed according to user needs like described in so that a person of an urban area is well known of the system being run by him. Commercial software can be used to include the shading effect and proper sizing of panels and battery can be made to meet the assigned load.
- A dedicated organization like IDCOL can be formed so that regular monitoring can be done and necessary actions can be taken against any irregularities to avoid any mismanagement .

VI. CONCLUSIONS

The paper presents a complete scenario of the existing solar home system in urban areas of Bangladesh. The irregularities found are described in a comprehensive way. Efficiency calculation has been provided with respect to the actual parameters found in Bangladesh. Some proposals are provided so that the urban solar home system can be made effective to fight against the power crisis in Bangladesh.

REFERENCES

[1] (2010) Independent BD. [Online] Available:http://www.theindependentbd.com/index.php?option=com_content& view=article& id=208590:installing-rooftop-solar-panels-not-mandatory& catid=110:business-others& Itemid=156

[2] (2010) Electro Solar. [Online] Available:http://www.electrosolarbd.com

[3] (2010) Homer Knowledgebase.[Online] Available:http://support.homerenergy.com/Knowledgebase/

[4] (2010) Homer Knowledgebase.[Online] Available:http://support.homerenergy.com/Knowledgebase/

[5] Khan, S.I. and Hague, S.M.M., "Economic analysis of solar PV system for urban areas of Bangladesh," in *Proc. 2012 7th Int. Conf. Electrical Computer Engineering*, 2012, pp. 892-895.

[6] Roy, P. and Arafat, Y. and Upama, M.B. and Hoque, A., "Technical and financial aspects of solar PV system for city dwellers of Bangladesh where green energy installation is mandatory to get utility power supply," in *Proc. 2012 7th Int. Conf. Electrical Computer Engineering*, 2012, pp. 916-919.

Solar Radiation Approximation Using Sunshine Hour at Patenga, Bangladesh

Shuvankar Podder,[1] Raihan Sayeed Khan,[2] Shah Md. Ashraful Alam Mohon, Md. Jubair Hussain, and Enamul Basher
Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology, Dhaka-1205, Bangladesh
E-mail: [1]shuvankar039@gmail.com, [2]raihansayeedkhan@gmail.com

***Abstract*—Solar radiation at Patenga, Chittagong is predicted by using record of bright sunshine hours of that place. Monthly average solar radiation is calculated from Angstrom-Prescott equation. The coefficients of that equation is found as a = 0.1688 and b= 0.5869. Patenga is a coastal area of Bangladesh and has promising potential of hybrid solar-wind based electricity. Solar radiation data is required for planning and designing photovoltaic power plant. But Bangladesh Meteorological Department has no record of radiation data of that place. So, attempt has been made to estimate solar radiation from other available meteorological parameters. Accuracy of estimated radiation is determined by comparing with data obtained from NASA Surface Meteorology and Solar Energy (SSE). Also estimated radiation is compared with five widely accepted models used for calculating solar radiation. These models were developed considering wide range of geographical area so can be applicable for present location. Present model provide better error correction scheme compared to other considered models by generating more accurate coefficient for that place.**

***Index Terms*—solar radiation, sunshine hour, patenga, renewable energy**

I. Introduction

Bangladesh is situated between 20.30° and 26.38° north latitude and 88.04° and 92.44° east longitude with an area of 147,500 square km. Daily average solar radiation varies between 4 to 6.5 kWh per square meter [1]. So there is a good potential of producing electricity using solar energy conversion system. The knowledge of solar energy radiation is very vital for the optimal design and performance analysis of any solar energy conversion system for electricity generation. At present, solar energy data can be found from the following sources:

- Renewable Energy Research Centre (RERC), Dhaka University; it has recorded long-term hourly solar irradiance data of Dhaka city with Eppley Precision Pyrometer.
- Bangladesh Meteorological Department (BMD); it has 35 sunshine recording stations [2] situated generally in towns and cities.
- NASA Surface Meteorology and Solar Energy (SSE); it has recorded world-wide solar radiation data covering the period from 1983 to 2005 [3].

Solar radiation measuring technology is expensive and has instrumental hazard for developing countries. Like most of the developing countries, Bangladesh does not have adequate number of radiation measuring stations. Therefore, it is essential that some reliable mathematical models be developed to estimate the solar radiation for places where measurements records are not available. Solar radiation reaching the earth's surface depends upon climatic conditions of that place. Thus a mathematical model can be developed relating these climatic factors with solar radiation. A number of studies [4-6] compute solar radiation from observation of cloud cover. Other studies [8-10] estimate solar radiation from sunshine hours. In this study, solar irradiance at Patenga, Chittagong is estimated. Patenga is a coastal area and has good potential for solar-wind hybrid renewable energy based electricity generation. Solar irradiance data at Patenga is not available at Bangladesh Meteorological Department(BMD). But BMD keeps record of daily sunshine hour in Patenga. For this reason, in this study, solar radiation at Patenga is estimated using sunshine duration. The result of estimation is also compared with five other commonly used correlation models.

II. Methodology

A. Angstrom Model

Angstrom [15] was first to propose a linear relation relating ratio of monthly average daily solar radiation to clear day radiation and ratio of bright sunshine hours to total day length in hours, which is given by

$$\frac{\overline{H}}{\overline{H_c}} = a + b\frac{\overline{n}}{\overline{N}} \tag{1}$$

Where $\overline{H}$=monthly average daily global radiation (Wh/m^2/day), $\overline{H_c}$= monthly average clear sky daily global radiation for the location, $\overline{n}$= monthly average daily maximum bright sunshine duration in hours, $\overline{N}$= actual sunshine duration in a day in hours, and a, b are empirical coefficients. These coefficients are location specific coefficients and known as fractions of extraterrestrial radiation on overcast days and on average days respectively.

A basic difficulty in this model is to determine $\overline{H_c}$, clear sky radiation. To avoid this difficulty a modified model was presented by Prescott [7] in 1940.

B. Angstrom-Prescott Model

Popularly known Angstrom-Prescott model is given by

$$\frac{\overline{H}}{\overline{H_0}} = a + b\frac{\overline{n}}{\overline{N}} \tag{2}$$

Where, $\overline{H}$, $\overline{n}$, $\overline{N}$, a and b are same as equation (1) and $\overline{H_0}$=monthly average daily extraterrestrial radiation at the specific location. The ratio of solar radiation at the surface of the Earth (H) to extraterrestrial radiation ($\overline{H_0}$), that is, $\overline{H}/\overline{H_0}$, is called the clearness Index and the ratio n/N is referred to as the cloudless index.

Monthly average daily extraterrestrial radiation is calculated from following equation

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

$$H_0 = \frac{24}{\pi} G_{sc} \left(1 + \cos\frac{360n}{365}\right) \times \left(\cos\varphi\cos\delta\sin\omega_s + \frac{\pi\omega_s}{180^o}\sin\varphi\sin\delta\right) \quad (3)$$

Where, G_{sc}= the solar constant = 1.367 kW/m^2
n= the day of a year (a number between 1 to 365, starting from 1st January)
φ= the latitude in degree
δ= the solar declination in degree
ω_s=the sunset hour angle in degree.

The solar declination is calculated according to the following equation:

$$\delta = 23.45^o \sin\left(360^o \frac{284+n}{365}\right) \quad (4)$$

The sunset hour angle is calculated using the following equation:

$$cos\omega_s = -tan\varphi tan\delta \quad (5)$$

The average H_0 for the month is calculated as follows:

$$H_{0,avg} = \frac{\sum_{n=1}^{N} H_0}{N} \quad (6)$$

Where, $H_{0,avg}$=the average extraterrestrial horizontal radiation for the month in kWh/m^2/day
N=the number of days in the month

The maximum possible sunshine duration N in hours for a horizontal surface is given by

$$N = \frac{2}{15}\omega_s \quad (7)$$

In this study, monthly average bright sunshine hours, n is collected from Bangladesh Meteorological Department for the period 1983 to 2012. H_0 is calculated from equation (3). N is calculated from equation (7) and compared with data obtained from WeatherSpark beta [14], a weather website. Regression coefficient a and b are obtained from following equations

$$a = \frac{\sum(\frac{H}{H_0})}{m} - b\frac{\sum(\frac{n}{N})}{m} \quad (8)$$

$$b = \frac{m\sum(\frac{n}{N})(\frac{H}{H_0}) - (\sum(\frac{n}{N}))(\sum(\frac{H}{H_0}))}{m\sum(\frac{n}{N})^2 - (\sum(\frac{n}{N}))^2} \quad (9)$$

Where, m is number of month. As measured data of monthly average solar radiation for Patenga is not available from BMD, the values of H are taken from NASA Surface Meteorology and Solar Energy (SSE) [3].
In this study, the best five sunshine duration based radiation models [9-13] are compared with our new model in predicting solar radiation to target place.

III. Models

The five models that are used to compare are

A. Gopinathan's model

This model recommends the following equations for calculating regression coefficient a and b of Anngstrom-type correlation to predict solar radiation at any location around the world [10]

$$a = -0.309 + 0.539\cos\varphi - 0.0693h + 0.290\left(\frac{\overline{n}}{\overline{N}}\right) \quad (10)$$

$$b = 1.527 - 1.027\cos\varphi + 0.0926h - 0.359\left(\frac{\overline{n}}{\overline{N}}\right) \quad (11)$$

Where, φ is the latitude in degrees, h is the elevation of the location above sea level in kilometers.

B. The model of Benson et el

Benson et el proposed two different set of value for regression coefficient a and b for two intervals of a year. His proposed models are [11]

$$\frac{\overline{H}}{\overline{H_0}} = 0.18 + 0.6\frac{\overline{n}}{\overline{N}} \quad (12)$$

for the months January-March and October-December.

$$\frac{\overline{H}}{\overline{H_0}} = 0.24 + 0.53\frac{\overline{n}}{\overline{N}} \quad (13)$$

for the months April-September.

C. Rietveld's model

This correlation is believed to be applied with world-wide validity [9]

$$\frac{\overline{H}}{\overline{H_0}} = 0.18 + 0.62\frac{\overline{n}}{\overline{N}} \quad (14)$$

D. The model of Ögelman et el

Ögelman et el used daily measured data for a period of 3 years at Ankara in Turkey to develop following eqution for predicting solar radiation [12]

$$\frac{\overline{H}}{\overline{H_0}} = 0.204 + 0.758\frac{\overline{n}}{\overline{N}} - 0.250\left(\frac{\overline{n}}{\overline{N}}\right)^2 + \sigma^2_{n/N} \quad (15)$$

In case that, the original data is not available and only available data is monthly average (n/N) but not $\sigma_{(n/N)}$ they established another correlation

$$\sigma^2_{n/N} = 0.035 + 0.326\frac{\overline{n}}{\overline{N}} - 0.433\left(\frac{\overline{n}}{\overline{N}}\right)^2 \quad (16)$$

Relating two equations they suggested the final quadratic relation

$$\frac{\overline{H}}{\overline{H_0}} = 0.195 + 0.676\frac{\overline{n}}{\overline{N}} - 0.142\left(\frac{\overline{n}}{\overline{N}}\right)^2 \quad (17)$$

E. McCulloch's model

McCulloch proposed this model that considers the latitude effect of a location. This is given by [13]

$$\frac{\overline{H}}{\overline{H_0}} = 0.29\cos\varphi + 0.52\frac{\overline{n}}{\overline{N}} \quad (18)$$

TABLE I
ERROR COMPARISON IN DIFFERENT MODELS

Model	a	b	MBE	RMSE	MAPE	r
Gopinathan's	$-0.309 + 0.539cos\varphi - 0.0693h + 0.290(n/N)$	$1.527 - 1.027cos\varphi + 0.0926h - 0.359(n/N)$	54.13%	62.61%	12.10%	0.849
Benson's	0.18 From January- March and October- December 0.24 From April - September	0.6 From January- March and October- December 0.53 From April - September	33.47%	42.02%	7.89%	0.8968
Rietveld's	0.18	0.62	28.12%	34.89%	6.10%	0.9492
Ögelman's	$0.195 + 0.676(n/N) - 0.142(n/N)^2$		27.60%	36.15%	6.80%	0.9126
McCulloch's	$0.29cos\varphi$	0.52	59.66%	65.12%	13.30%	0.8926
Our	0.1688	0.5869	0.77%	18.81%	3.49%	0.9494

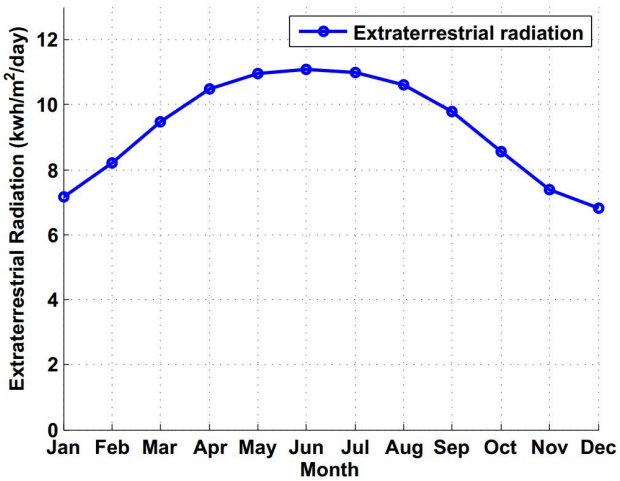


Fig. 1. Monthly extraterrestrial radiation at Patenga

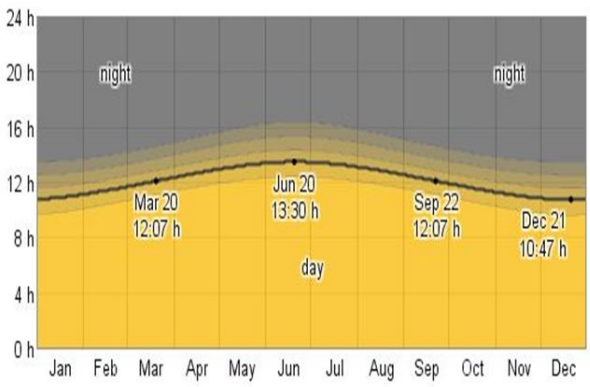


Fig. 2. Monthly average day length

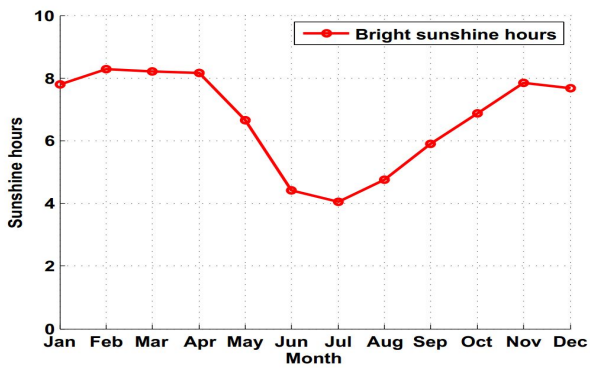


Fig. 3. Monthly average bright sunshine hours

IV. LOCATION AND DATA

Solar irradiance at Patenga, Chittagong in Bangladesh is estimated. The latitude and longitude of Patenga is 22.30^0 North and 91.80^0 East respectively. Extraterrestrial radiation over target location, obtained using equation (3) are shown in Fig. 1. Monthly average day length, N is shown in Fig. 2 [14]. Monthly average bright sunshine hours, obtained from BMD, are shown in Fig. 3.

V. COMPARISON TECHNIQUE

To make performance analysis and obtain degree of accuracy, Mean Bias Error(MBE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error(RMSE) and correlation r for twelve months is calculated for each of the five models mentioned above. MBE, MAPE, RMSE and correlation r are given by following equations

$$MBE = \left(\sum_{i=1}^{n=12} (H_{ic} - H_{im}) \right) /n \tag{19}$$

$$MAPE = \left(\sum_{i=1}^{n=12} \frac{|H_{ic} - H_{im}|}{H_{im}} \right) /n \tag{20}$$

$$RMSE = \left\{ \left(\sum_{i=1}^{n=12} (H_{ic} - H_{im})^2 \right) /n \right\}^{1/2} \tag{21}$$

$$r = \frac{\sum_{i=1}^{n=12} (H_{ic} - \overline{H}_c)(H_{im} - \overline{H}_m)}{\sqrt{\left(\sum_{i=1}^{n=12} (H_{ic} - \overline{H}_c)^2\right)\left(\sum_{i=1}^{12} (H_{im} - \overline{H}_m)^2\right)}} \tag{22}$$

Where, H_{ic} and H_{im} are the estimated and measured monthly average solar radiation for the ith month respectively. The average of the deviations $E = (H_{ic} - H_{im})$ is MBE and gives information about the long- term performance of the correlations. MAPE is a measure of the goodness of each correlation, while RMSE measures the short-term prediction quality of the correlations.

VI. RESULTS AND DISCUSSION

From equations 8 and 9 regression coefficient a and b are calculated using MATLAB simulation. Final correlation equation of this study with this coefficient is

$$\frac{\overline{H}}{\overline{H_o}} = 0.1688 + 0.5869\frac{\overline{n}}{\overline{N}} \tag{23}$$

a and b values thus estimated are employed along with the fraction of bright sunshine duration to compute the monthly

average global solar radiation on a horizontal surface for target location. The accuracy of the estimated data of solar radiations is evaluated by calculating MBE, RMSE and MAPE values. The estimated data are also compared with the models proposed by Gopinathan, Benson et el, Rietveld, Ögelman et el and McCulloch. Percentage errors in predicting global solar radiation at target place by different models are shown in Table I .

As seen from the table, correlation is quiet high in each model which indicates that considered models can predict the pattern of monthly variation of solar radiation on target place quiet satisfactory. But except our model rest of the five models suffer from large %MBE, %RMSE. In present model %MBE and %MAPE is very low which indicates long-term global solar radiation for target place using this model can be predicted with high degree of accuracy. But %RMSE in present model is comparatively high than other errors, which indicates that short term prediction suffers from lack of accuracy. A comparison between estimated solar radiation data using present model and measured data from NASA SSE is shown in Fig 4.

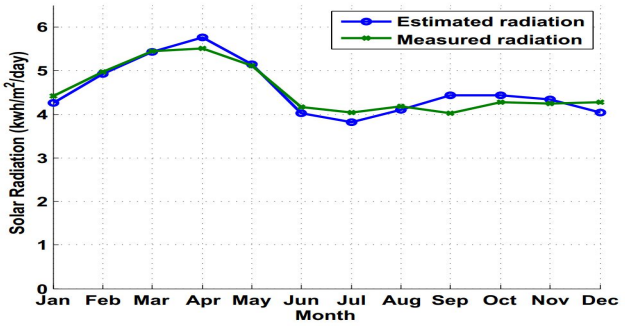


Fig. 4. Estimated and Measured solar radiation at Patenga

Monthly average global solar radiation estimated using five considered models along with present model is shown in Table II

TABLE II
MONTHLY AVERAGE SOLAR RADIATION ESTIMATED USING DIFFERENT MODELS

Month	$H_{measured}$	Gopinathan's Model	Benson's Model	Rietveld's Model	Ögelman's Model	McCulloch's Model	Our Model
January	4.4207	4.5136	4.4117	4.5157	4.3795	4.6327	4.2629
February	4.9811	5.1976	5.1019	5.2227	5.0544	5.3486	4.9304
March	5.4428	5.8436	5.6276	5.7583	5.627	5.9466	5.4354
April	5.5048	6.302	6.1263	6.1101	6.0206	6.3598	5.7668
May	5.1137	5.912	5.6048	5.4509	5.5228	5.8644	5.1421
June	4.1595	4.853	4.6089	4.2739	4.4733	4.8912	4.0277
July	4.0356	4.631	4.4192	4.0617	4.269	4.7019	3.827
August	4.1825	4.913	4.643	4.3618	4.5334	4.9088	4.1117
September	4.0227	5.1442	4.8646	4.7045	4.7903	5.0987	4.4374
October	4.2823	4.9524	4.5999	4.7019	4.685	4.9524	4.4369
November	4.2493	4.6227	4.4995	4.6051	4.4757	4.7329	4.3472
December	4.2811	4.2871	4.1848	4.2834	4.1569	4.3968	4.0435
Average	**4.5563**	**5.0977**	**4.8914**	**4.8375**	**4.8323**	**5.1529**	**4.5641**

It is evident from Table II that there is a very fine agreement between measured and estimated values obtained from present correlation model. A comparison between measured solar radiation by NASA SSE and estimated solar radiation by using all of these models considered in Table II is shown in Fig. 5.

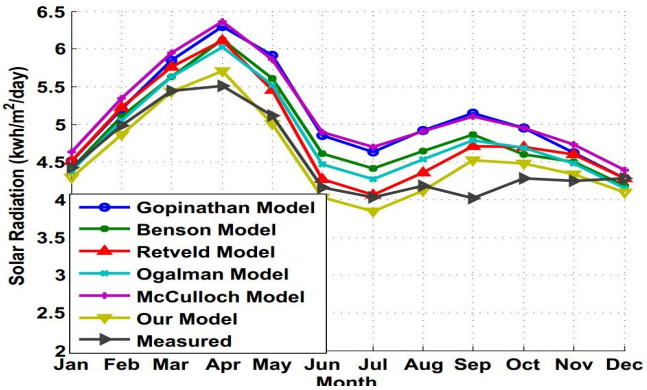


Fig. 5. Prediction of monthly solar radiation by different model

VII. CONCLUSION

In the present work, the low percentage error suggests that this model can be used with confidence for Patenga, Chittagong and also for other locations with similar climatic conditions. This statement is also supported by the high value of correlation coefficient. Solar irradiance is maximum at March-April which is dry season. Solar irradiance is minimum during June-July which is rainy season. This result shows that the prediction is practical. The accuracy of the model can be further improved by considering the fog density, cloud cover and atmospheric scattering effect.

REFERENCES

[1] Mazharul Islam, Assessment of renewable energy of Bangladesh,www.sdnbd.org/sdi/issues/energy/publications/shakti-ebook1.pdf
[2] Country Report: Bangladesh Meteorological Department 2013; available from http://www.jma.go.jp/jma/en/Activities/RIC_Workshop_2013/documents/pre/01.%20presentation%20(Bangladesh).pdf
[3] NASA Surface meterology and Solar Energy; from https://eosweb.larc.nasa.gov/cgi-bin/sse/sse.cgi
[4] S. Rangarajan, M. S. Swaminathan and A. Mani, Computation of solar radiation from observations of cloud cover. *Solar Energy*, Vol. 32, No. 4, pp. 553-556, (1984)
[5] H R Ghosh, S M Ullah, S K Khadem, N C Bhowmik and M Hussain, Measurement and estimation of sunshine duration for Bangladesh. Renewable Energy Research Center, University of Dhaka, Bangladesh
[6] A. Mani, O. Chacko, V. Krishnamurthy and V. Desikan, Distribution of global and net radiation over the Indian Ocean and its environments. *Arch. Met. Geoph. Biokli*, B15(1-2), 82 (1967).
[7] J. A. Prescott, "Evaporation from water surface in relation to solar radiation", *Transactions of The Royal Society of South Australia*, vol. 40, pp.114-118 (1940).
[8] A. De Miguel, J. Bilbao, S. Salson and A. Lage, Solar radiation and sunshine hour maps in Castilla and Leon region (Spain), *Renewable Energy*, Vol. 4. No. 8, pp.933-940, (1994).
[9] M. R. Rieltveld, A new method for estimating the regression coefficients in the formula relating solar radiation to sunshine. *Agr. Meteorol*. vol - 19, pp243-252 (1978).
[10] K. K. Gopinathan, A general formula for computing the coefficient of the correlation connecting global solar radiation to sunshine duration. *Solar Energy*, Vol. 41. No. 6, pp. 499-502. (1988).
[11] P. R. Benson, M. V. Paris, J. E. Serry and C. G. Justus, Estimation of daily and monthly direct, diffuse and global solar radiation from sunshine duration measurement , *Solar Energy*, Vol. 32, No.4, pp:523-535 (1984).
[12] H. Ögelman, A. Ecevit and E. Tasdemiroglu, A new method for estimating solar radiation from bright sunshine data. *Solar Energy*, Vol. 33, No.6, pp. 619-625, (1984).
[13] J. Glover and J. D. G. McCulloch, The empirical relation between solar radiation and hour of bright sunshine. *Q. J. Roy, Met. Soc.* 84, 172 (1958).
[14] WeatherSpark beta, a weather website http://weatherspark.com/averages/33927/Chittagong-Chittagong-Division-Bangladesh
[15] A. Angstrom, Solar and terrestrial radiation, *Quarterly Journal of the Royal Meteorological Society*, vol. 50, no. 210, pp. 121-126, 1924.

Study and Protection of Lightning Overvoltage on DC Cables of Solar Power Generation

Md. Alamgir Hossain,* and Md. Raju Ahmed

Department of Electrical & Electronic Engineering, Dhaka University of Engineering & Technology (DUET)
Gazipur-1700, Bangladesh
* alamgir_duet@hotmail.com

Abstract— **In this paper, lightning induced overvoltage on DC cables of solar power panel, which was laid on the ground and an underground level with proper insulation thickness, was investigated when a lightning strike at a nearby place and its proper protection scheme was verified. The result shows that on ground cable voltage is approximately 28.83% more than that of the underground cable having soil resistivity 100Ω-m. Moreover, when soil resistivity increases to 1000Ω-m then the underground cable voltage is less about 48.50% than that of the on ground cable. The induced voltage caused by lightning electromagnetic interference on DC cables of solar power system at power conditioning subsystem (PCS) is analysed using electromagnetic field analysis approach. The FDTD method was applied to calculate induced voltage. Virtual Surge Test Lab (VSTL) is used for FDTD calculations, which solve Maxwell's electromagnetic equations with least approximations. The model of solar power system was represented with thin wires.**

Index Terms— **lightning induced voltage, DC cables, PCS, protective device.**

I. Introduction

The photovoltaic (PV) generation is increasing whole over the world dramatically because of its attractive contribution to power generation, no CO_2 emission, cost effective and easy installation to anywhere in the place. Keeping hand in hand with the world, Bangladesh is also progressing well in this sector because of lack of fossil fuel in the upcoming future [1]; however the number of increasing PV panels is now concern of its secure operation and safety issues for unexpected weather condition like lightning. The overvoltage in solar power system can be either cause of direct or indirect lightning effect, sudden load changing, electrical switching, etc., which are dangerous for damaging photovoltaic generator and its ancillary equipment. In most of the case, lightning is major reason for damaging electrical and electronic equipment in a system [2].

Lightning phenomena has been studied intensively in many researches [3-6] and found its fairly general behaviour, nonetheless precise illustrations are not foreseeable for specific incidents. Lightning induced voltage on solar power system was investigated in Martzloff work [7], and [8]. They conclude that lightning voltage is proportional to peak value of lightning current. In Z. Beesova et all [9] work they show that strike point carries major influence components for inducing voltage besides shape of lightning current and constructional size of solar panel. In order to optimize the solar protection system, lightning induced voltages need further study to design solar power system for secure, reliable, and economical operation.

To analyse lightning phenomena of solar power system, numerical electromagnetic analysis is applied successfully. In this application, The Finite Difference Time Domain (FDTD) [10] was utilized to compute induced voltages. For simulation purpose, frame of PV panels and DC wires are represented with thin wire model [11] and frequency dependency of soil is not considered.

To generate electricity commercially from solar; large size of PV panel is required, which is exposed in open field, have a chance to be affected by directly/indirectly lightning that lead to high voltage between DC line and Power Conditioning Subsystem (PCS) that is enough to damage the PCS [12]. The most of the damage occurred in electrical and electronic devices owing to indirect lightning. In our study, we analysed the indirect lightning induced voltage between DC cables of solar generator and PCS with varying soil resistivity.

II. Analysis Methodology

In this paper, numerical simulations are carried out using FDTD method which directly solves the Maxwell equation in the time domain form for solving electromagnetic propagation problems [13]. Base on the FDTD method the Virtual Surge Test Lab (VSTL) [14, 15] code was developed by the Central Research Institute of Electric Power Industry (CRIEPI) to analyse the surge related electromagnetic field problems. To represent open space to all the boundaries in the analytic space, second-order Liao's method was adopted in VSTL code for getting precise results and using less memory compared to other methods [16].

Lightning stroke channel having propagation velocity of return stroke is 1/3 of the speed of light [17, 18] is represented by engineering model built in VSTL. The current waveform used in this model was adopted from the analytical expression of ref. [19].

$$i(0,t) = \frac{I_0}{\eta}\frac{(t/\tau_1)^n}{1+(t/\tau_1)^n}\exp\left(-\frac{t}{\tau_2}\right) \quad (1)$$

Where,

$$\eta = \exp\left[-\left(\frac{\tau_1}{\tau_2}\right)\left(\frac{n\tau_2}{\tau_1}\right)^{\left(\frac{1}{n}\right)}\right] \quad (2)$$

I_0 = Amplitude of the channel-base current
τ_1= Front time constant
τ_2= Decay time constant
η= Amplitude correction factor
n = Exponent (2 ... 10)

The sum of two functions given in Eq. (1) was chosen so as to obtain the overall wave shape of the current as observed in typical experimental results.

In FDTD method the cell size and time step must satisfy the Courant stability condition [20]. Time step can be calculated using the following equation [14] when space step is known.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

$$\Delta t = \Delta s\sqrt{\frac{\mu\epsilon}{3}}(1-\alpha)$$

Here,

α is small positive value, <1, which is used to avoid instability of the numerical integration due to rounding off error.

III. Analysis Model

A solar panel was modelled in VSTL shown in Fig. 1 having the height of 2.5m above the ground, width of 3m, length of 20m, and depth of grounding rods is 0.5m. In this model, panel wiring resistance of 2Ω has been considered in series connection with DC line. Power was transmitted to the Power Conditioning Subsystem (PCS), which was installed 17m away from the solar panel and 22m apart from DC cables, with 30m long insulated DC cables. Cables was placed on the ground with 0.1m thickness around it whose relative permittivity is 3 and separated by 0.1m distance having end open at the PCS side. The lightning channel is placed at the 17m away from the solar panel and 22m apart from PCS to represent as lightning stroke at nearby place.

In the analysis of VSTL, the simulation dimensions are 35.0m× 35.0m × 30.0m in the X, Y, and Z direction respectively. This space was divided into 270 ×170× 80 cells to shrink the calculation time. Soil was postulated homogeneously with resistivity 100Ωm and relative permittivity 10 in the Z direction of 10m height. In the six planes of the analysis space, Liao's second order absorbing boundary condition was established to keep open space [16]. The cell sizes for point of interest are kept 10cm and are increased gradually larger as they are further away from a point of interest. Therefore its ranges vary from 0.1 m to 1 m.

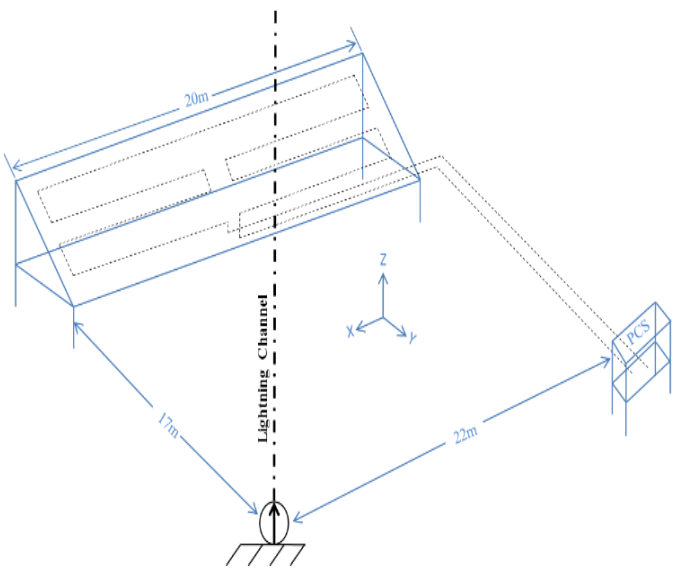


Fig. 1 Solar power system and lightning channel analytical model in VSTL

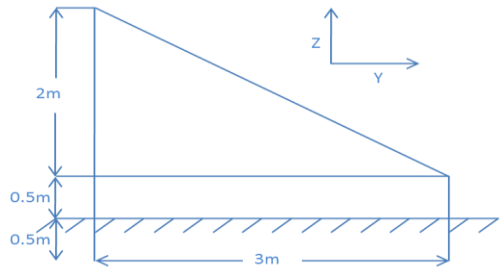
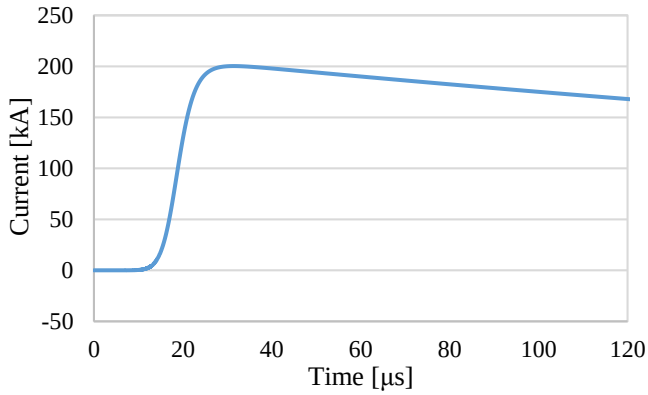
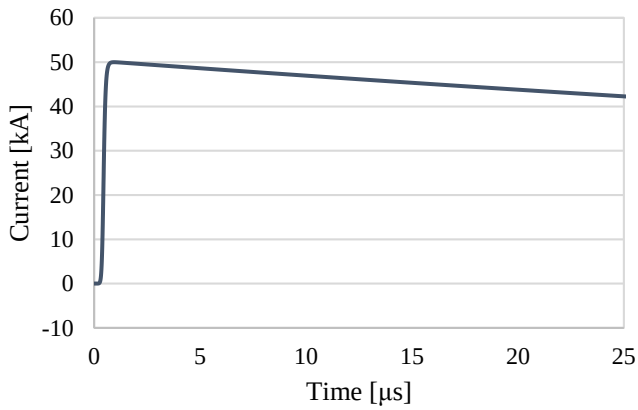


Fig. 2 Dimensions of solar panel installation from side view



(a) First return-stroke current



(b) Subsequent return-stroke current

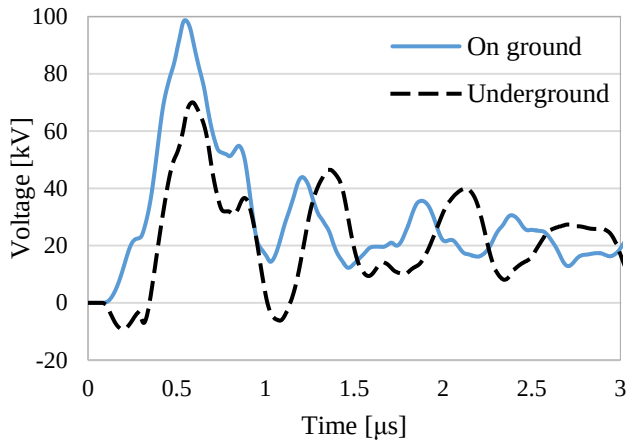
Fig. 3 Recommended test current [21] waveform for lightning (a) first return stroke current, 10/350μs, and (b) subsequent return-stroke current, 0.25/100μs.

Fig.3 shows the recommended test current waveform for lightning stroke. In our analysis, the subsequent return-stroke current was adopted because of high rising current i.e. the rate of change of current is very high compared to the first return-stroke current. Though the current magnitude of the first return stroke is 4 times that of the subsequent return-stroke, however induced voltage of subsequent return stroke current is severer than the first stroke current.

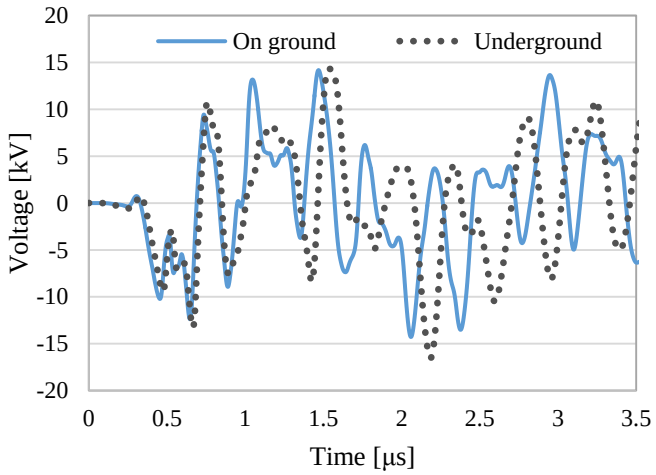
IV. Result and Discussion

A. *Effect of DC Power Cable Laying*

DC power cable which transfer power from solar panel to the PCS end can be set up on air or underground, for this reason the effect of induced overvoltage cause by nearby lightning was investigated on laying cables. The length of cable which was modelled with thin wire is 30m surrounded by 0.1m insulated cover. The insulation thickness is determined by the cell size in FDTD analysis.



(a) Line to PCS induced voltage (soil resistivity is 100**Ω**-m)



(b) Line to line induced voltage

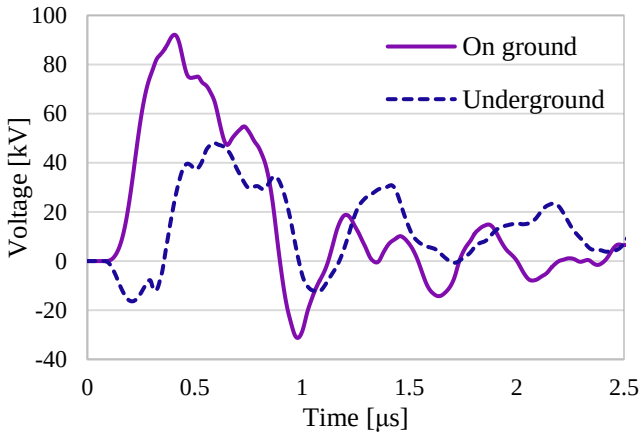
Fig. 4 Induced voltages on ground and 1m underground DC cables having soil resistivity 100**Ω**-m (a) Line to PCS and (b) Line to line.

The induced voltage on the cable depends on where it is placed; if the cables are placed on ground the voltage is higher than that of cable which is laid at 1m underground level. In Fig. 4(a), the maximum induced voltage on ground cable between line and PCS is 98.36kV at time 0.56µs with oscillating frequency 3.44MHz where peak induced voltage for underground cable is around 70kV which is 71.17% of the on ground cable voltage. Therefore underground cable reduces overvoltage 28.83% of the on ground cable induced voltage.

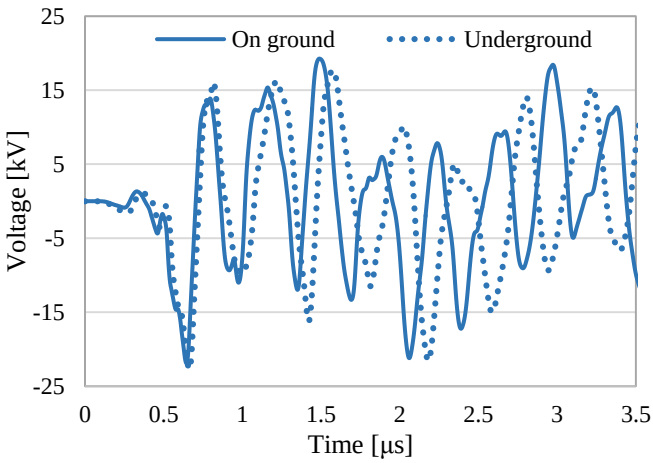
The line to line induced voltage for cable placed on ground and underground is more or less same where maximum voltage magnitude was recorded for underground is 16.5kV with oscillating frequency 4.55MHz. The reason behind is that voltage induced in the wires is similar nature except magnitude changes.

In Fig. 4 it is observed that line to PCS voltage is oscillating with high frequency i.e. MHz ranges because of sharp rising (0.25µs) injected current in the lightning channel as well as electric field reflection from the ground is more at the high rising time front of lightning test current.

In Fig 5 (a) indicates that induced overvoltage between line and PCS having soil resistivity 1000**Ω**-m is low comparatively with soil resistivity 100**Ω**-m. The line to PCS maximum overvoltage for cable placed on ground is found at time 0.41µs is around 92kV with oscillating frequency 3.13MHz, where underground cable voltage is only 47.38kV. The induced voltage of 1m underground cable is 51.5% of the on ground cables which shows voltage reduction 48.5%.



(a) Line to PCS induced voltage (soil resistivity is 1000**Ω**-m)



(b) Line to line induced voltage

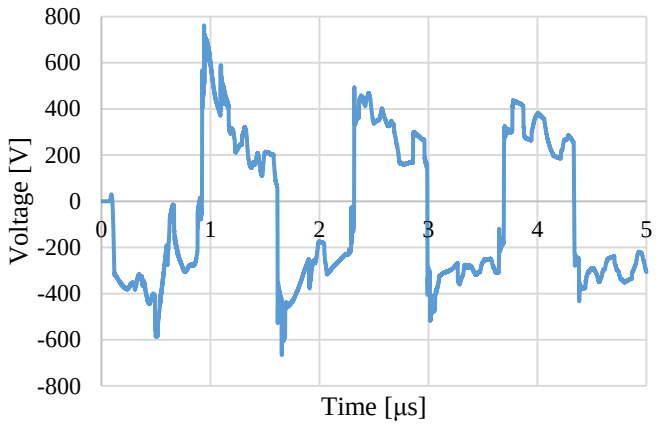
Fig. 5 Induced voltages on ground and 1m underground DC cables having soil resistivity 1000**Ω**-m (a) Line to PCS and (b) Line to line.

The line to line induced voltage is also more or less similar for both wires laying. The maximum magnitude of induced voltage for underground cable was observed at time 0.66µs is 22.62kV with oscillating frequency 3.12MHz.

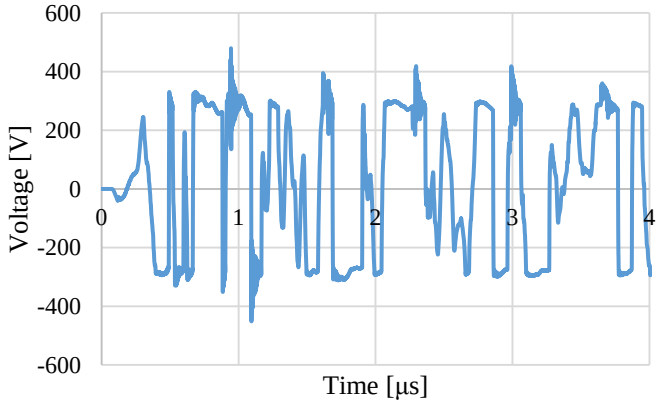
It is noticeable that with the increasing soil resistivity the induced voltage for line to PCS is reduced for both on ground and underground cables. Why? With the increasing soil resistivity; the behaviour of soil, which is inductive, is dominated that reflect the electrical field from the surface of the ground i.e. the penetration of electric field in the ground is less when soil resistivity is high. This is the reason why voltage is found in reduced form.

B. Design of Protection System

Surge Protection Device is used to limit the induced voltage/Overvoltage in the lines or electrical devices by passing surge current associated with surge voltage. In our analysis, SPD was installed between the wires of DC cables and between wires of DC cable and ground. The SPD model used in our study have an operating voltage V_{th}= 300V, resistance for preventing normal current to flow through SPD is 1M**Ω** and resistance that helps to flow current through SPD is 0.1**Ω**.



(a) Line to PCS induced voltage



(b) Line to line induced voltage

Fig. 6 After installing surge protection device in between line to line and line to ground for 1m underground cables having soil resistivity 100**Ω**-m, (a) Line to PCS induced voltage and (b) Line to line induced voltage.

Fig.6 (a) shows that induced voltage between line and PCS is reduced in normal voltage ranges where the maximum induced voltage because of nearby lightning was about 70kV at time 0.6μs with oscillation frequency1.43MHz for 1m underground cable with soil resistivity 100**Ω**-m. The maximum voltage amplitude between line and PCS after installing SPD is only 762V at time 0.94μs with 0.725MHz oscillation frequency, which abridged to 98.91% of the maximum voltage. In the case of line to line voltage, the maximum voltage amplitude was appeared at time 2.18μs is 16.52kV with oscillating frequency 4.55MHz for 1m underground cables with soil resistivity 100**Ω**-m; this voltage magnitude, after installing arrester, has been reduced to 480V with higher oscillating frequency11.11MHz. Hence, the voltage reduction in line to line is 97.10% after setting SPD between two DC power lines.

V. Conclusion

The effect of nearby lightning was investigated in two cases and their protection system was described with the help of VSTL. It is found that when DC cables are placed on ground the induced voltage is more, approximately 28.83%, than that of the underground cables while soil resistivity is 100**Ω**-m. The line to PCS induced voltage for underground cables having soil resistivity 1000**Ω**-m is found in extra reduction form nearly 48.50% less than that of the on ground cables. Therefore it can be concluded that underground cable line is preferable to limit the induced voltage from a nearby lightning strike. Moreover, we analysed protection scheme (Surge Protection Device, arrester) to overcome the induced overvoltage of the line to PCS and line to line for secure operation of the solar power system, and it is recommended that to use arrester in between line and ground to pass the overvoltage safely.

References

[1] M. A. Hossain and M. R. Ahmed, "Present energy scenario and potentiality of wind energy in Bangladesh," International Journal of Electrical, Electronic Science and Engineering, vol. 7, no. 11, pp. 1001-1005, 2013.

[2] C. A. Nucci, F. Rachidi, M. Ianoz and C. Mazzetti, "Lightning-induced Voltage on Overhead Lines," IEEE Trans. on Electromagnetic Compatibility, vol. 35, no. 1, pp. 75-86, 1993.

[3] H. –. Stern and H. C. Karner, "Lightning Induced EMC phenomena in photovoltaic," in IEEE International Symposium on Electromagnetic Compatibility 1993, 1993.

[4] V. Rakov, M. Uman, M. Fernandez, C. Mata, K. Rambo, M. Stapleton and R. Sutil, "Direct lightning strikes to the lightning protective system of a residential building: Triggered lightning experiments," IEEE Trans. on Power Delivery, vol. 17, no. 2, 2002.

[5] P. Hasse, Overvoltage protection of low voltage systems, 2nd edition ed., London IEE press.

[6] F. A. M. Rizk, "Modeling of transmission line exposure to direct lightning strokes," IEEE Trans. on Power delivery, vol. 5, no. 4, pp. 1983 - 1997, 1990.

[7] T. Youping, C. Zhang, J. Hu, S. Wang, W. Sun and H. Li, "Research on lightning overvoltages of solar arrays in a rooftop photovoltaic power system," in 7th Asia Pacific International Conference on lightning, 2011.

[8] S. Yokoyama, K. Miyake, H. Mitani and A. Takanishi, "Simultaneous measurement of lightning induced voltages with associated stroke currents," IEEE Trans. on Power Apparatus and Systems, Vols. PAS-102, no. 8, pp. 2420-2429, 1983.

[9] Z. Benesova, R. Haller, J. Birkl and P. Zahlmann, "Overvoltages in photovoltaic systems induced by lightning strikes," in 2012 International Conference on Lightning Protection (ICLP), 2012.

[10] K. Tanabe, "A method for computational analysis of transient resistance for grounding systems based on the FDTD method," Trans. IEE of Japan, Vols. 120-B, no. 8/9, pp. 1119-1126, 2000.

[11] T. Noda and S. Yokoyama, "Thin wire representation in finite difference time domain surge simulation," IEEE Trans. on Power Delivery, vol. 17, no. 3, pp. 840-847, 2002.

[12] N. E. a. I. t. D. O. (NEDO), "Technical analysis and evaluation activities related to the status of measures to reduce lightning damage of photovoltaic systems," 2009.

[13] K. Yee, "Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic material," IEEE Transactions on Antennas and Propagation, Vols. AP-14, no. 3, pp. 302-307, 1966.

[14] T. Noda and S. Yokoyama, "Development of general Surge analysis program based on the FDTD method," Trans. IEE of Japan, Vols. 121-B, no. 5, pp. 625-632, 2001.

[15] T. Noda, A. Tatematsu and S. Yokoyama, "Improvements of an FDTD-based surge simulation code and its application to the lightning overvoltage calculation of a transmission tower," in Proc. Int. Conf. on Power Systems Transients (IPST)-2005, Montreal, Canada..

[16] Z. P. Liao, H. L. Wong, B. P. Yang and Y. F. Yuan, "A transmitting boundary for transient wave analyses," vol. A27, no. 10, p. 1063 – 1076, 1984.

[17] B. F. J. Schonland, "The Lightning Discharge," Handbuch der Physic, vol. 22, pp. 576-628, 1956.

[18] V. P. Idone and R. E. Orville, "Lightning return stroke velocities in the thunderstorm research internatinal program (TRIP)," Journal of Geophysical Research, vol. 87, no. C7, pp. 4903-4915, 1982.

[19] F. Heidler, "Traveling current source model for LEMP calculation," in 6th International Symposium on Electromag. Comp., Zurich, Switzerland,, 1985.

[20] A. Quarteroni, R. Sacco and F. Saleri, Numerical Mathematics, Springer, 2000.

[21] "Protection against Lightning – Part 1: General Principles," International Electrotecnical Commission (IEC) , 2010-12.

[22] K. Kunz and R. J. Luebbers, The finite difference time domain method for elctromagnetics, CRC Press, 1993.

Implementation of Maximum Power Point Tracking in a Photovoltaic Inverter using Incremental Conductance Technique

Nawrin Islam Natasha[1], Wahida Taskin Bhuiyan[1], M. A. Razzak[2]
[1]Department of Electrical & Computer Engineering, North South University, Dhaka, Bangladesh
[2]Department of Electrical and Electronic Engineering, Independent University, Bangladesh, Dhaka, Bangladesh
E-mail: natz_1313@hotmail.com, wahida.taskin@gmail.com, razzak@iub.edu.bd

***Abstract*—Photovoltaic is a sustainable and environment friendly energy source. However due to dynamic changes in solar irradiance and temperature, this infinite and free energy cannot be converted with high efficiency. This paper illustrates a control strategy to design and implement Maximum Power Point Tracking (MPPT) in a photovoltaic system using Incremental Conductance algorithm. The PSIM simulation results verify proper functioning of the proposed MPPT sub-circuit. A constant 312V DC output is generated from irregular voltage of solar panel by changing the duty cycle of the MOSFET of the 24V-312V boost converter. The filtered output waveform of the SPWM driven H-bridge inverter via the L-C low pass filter is found to be a pure sine-wave of 312V_{peak} (220V_{rms}) and 50 Hz frequency with a total harmonic distortion (THD) of 0.002 which is significantly lower than the IEEE519 standard.**

***Keywords*—MPPT, Boost converter, PWM, H-bridge inverter, PSIM, SPWM, IncCond**

I. INTRODUCTION

The steep increase of the world's energy intake and the degradation of environmental conditions due to various energy generation processes, raise a burning concern for green and efficient energy production. Solar energy is one of the best solutions for energy crisis as it is inexhaustible and has a clean nature. As the solar panel outputs are irregular and vary with changes in irradiance, photovoltaic (PV) systems are continuously being made more efficient. Maximum Power Point Tracking (MPPT) techniques play a vital role to maximize the PV array output power in fluctuating conditions. MPPT is a fully electronic system that tracks the maximum power point of a PV module and tries to keep it at that point. The concept is to continuously monitor the terminal voltage and current, calculate the change in power simultaneously and then to take necessary steps to achieve Maximum Power Point [1]. Several algorithms exist for tracking MPP like Perturb and Observe (P&O), Incremental Conductance (IncCond), Neural Network, Current Sweep, Fuzzy Logic Method, etc [2], [3]. However, IncCond is more ideal as P&O method has certain disadvantages like the oscillations around MPP which result in tracking the MPP in the wrong direction during varying atmospheric conditions [4].

This paper demonstrates a control circuit for MPPT in a photovoltaic system using the IncCond method due to its simple implementation and high performance characteristics as described in section IIB. The proposed photovoltaic system is composed of a solar panel, a DC-DC boost converter driven by MPPT sub-circuit, an SPWM driven H-bridge inverter [5] and a low pass L-C filter to remove harmonics and produce a pure sine wave output with fundamental frequency of 50 Hz.

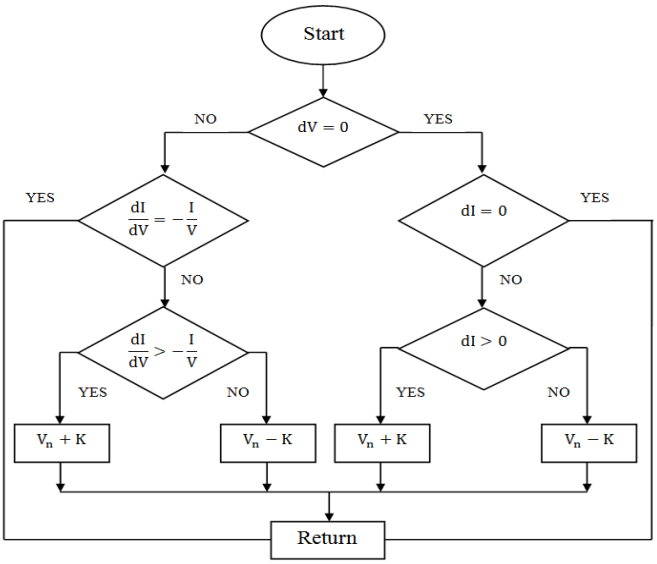


Fig. 1. Flow chart of InCond algorithm.

The topology is called incremental conductance because adjustment of the instantaneous voltage is done according to the MPP voltage which eventually is based on the incremental and instantaneous conductance of the PV module. The basic concept is that the slope of the P-V module power curve is zero at the MPP (dP/dV=0), greater than zero on the left side of MPP (dP/dV>0) and less than zero (dP/dV<0) on the right hand side of the MPP. Applying the chain rule for the derivative of products yields to dP/dV = [d(VI)]/ dV. As dP/dV=0 at MPP, the above equation could be written in terms of array voltage V and array current I and derive into dI/dV = -I/V.

The flowchart of the IncCond MPPT algorithm is shown in Fig.1. Initially, the values of instantaneous current (I_n) and voltage (V_n) are considered and then judged whether the change in voltage (dV) is zero, and if it is true then whether

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

change in current (dI) is equal to zero. If both are true then it means that the PV is operating at the MPP, so no change needs to be done [6]. However, if dI is positive then a constant voltage **K** is added to V_n and if dI is negative then a constant voltage **K** is deducted from V_n and this adjustment is done to move the array terminal voltage towards the MPP voltage.

On the other hand, if dV is not zero then it is checked whether dI/ dV is equal to –I/V. If it is true then no alteration needs to be done since this is the suitable condition at MPP. A constant voltage **K** is to be added or deducted to or from V_n if dI/dV is greater or less than –I/V respectively.

II. PROPOSED INVERTER

A. Photovoltaic Panel Modeling

Photovoltaic panel modeling is done using two physical models of solar modules and a triangular DC voltage source as shown later in Fig.6. The modules are positioned in series using capacitors and reverse parallel diodes to provide the desired input to the system. Furthermore, a triangular voltage source of 24 V_{p-p} with a diode in series is put in parallel with the panels to represent the fluctuating nature of solar panel's voltage.

B. MPPT Sub-Circuit

Here, the IncCond MPPT method is chosen to track the maximum power and the PSIM sub-circuit is divided into two parts for clear explanation. To perform IncCond algorithm, a series of values are needed initially and Fig.2 shows the first part. At first, this sub-circuit measures the change in voltage (dv) and it is turned into an absolute value which is eventually feed into the positive terminal of the comparator. If dv is not equal to zero then logic 1 will be the output otherwise logic 0. Correspondingly, with the same strategy it is determined whether di is equal to zero or not and whether di is greater or less than zero. For the last two cases a voltage of 1v is used in the negative terminal as the reference voltage instead of ground to obtain less oscillation. The fourth comparator determines whether di/dv>-i/v or di/dv<-i/v and the summer with the not gate calculates whether di/dv=-i/v.

Fig.3 shows the second part of the sub-circuit. Here A1, B1, C1 and D1 are the AND gates with inputs of no change condition and a NOT gate is used in front of each of them so that the value is inverted and no alteration is done if it is true. A3, B3, C3 and D3 will be 1 only when "no change conditions" are 0 and "change conditions" are 1. The input logic combination activates either V_n+K or V_n-K and this turns on the desired switch for either addition or subtraction of voltage to or from the V_n. The total voltage is then passed on to the fifth comparator which compares it with a DC reference voltage. The output of this comparator is the PWM control signal which is used to operate the boost converter. The proposed MPPT sub-circuit helped to achieve the 312V_{peak} ac at the inverter output from the fluctuating solar panel voltage (~24V) by varying duty cycle of the boost converter.

Figure 4(a) shows the 24 V_{p-p} half-wave triangular input voltage (V_n) which represents the fluctuating voltage of the solar panel. Here half-wave input is chosen so that the condition for no change can be examined. Fig. 4(b) shows the input current (I_n).The output voltage from the summer (V_{sum}) and a DC reference voltage of 15V(V_{ref}) are demonstrated in Fig. 4(c). For clear verification the frequency of V_n is chosen to be low as 25Hz and the timescale is fixed to 0.08s. Here **K** is considered as a high value just to verify clearly the addition and subtraction of voltage K from the instantaneous input voltage. Depending on the combination result the graph in Fig. 4(c). The PWM signal generated by the process is shown in Fig. 4(d), which is used for switching the boost converter.

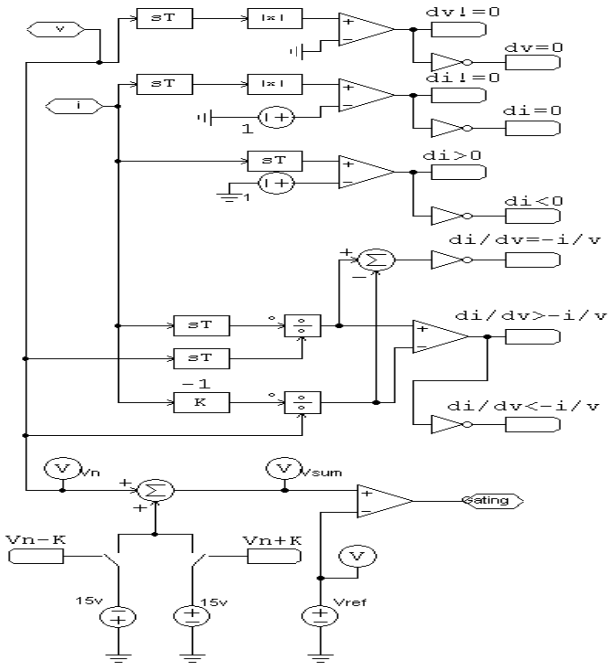


Fig. 2. Schematic diagram of MPPT sub-circuit in PSIM.

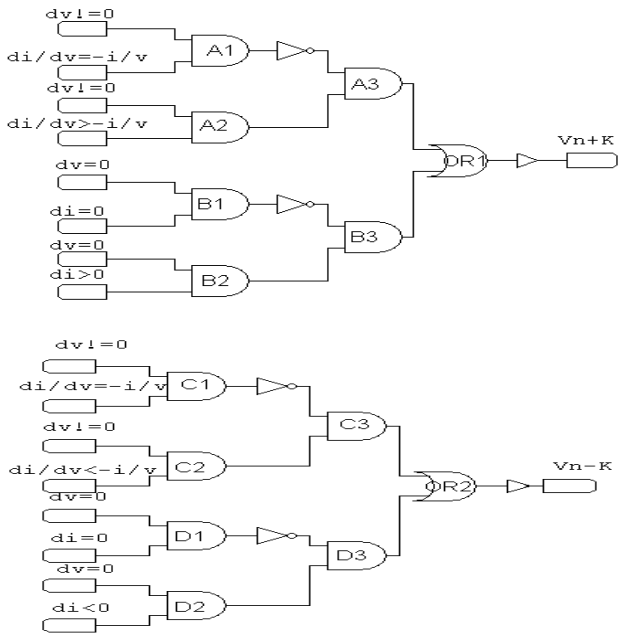


Fig. 3. Logic circuit of IncCond PSIM sub-circuit.

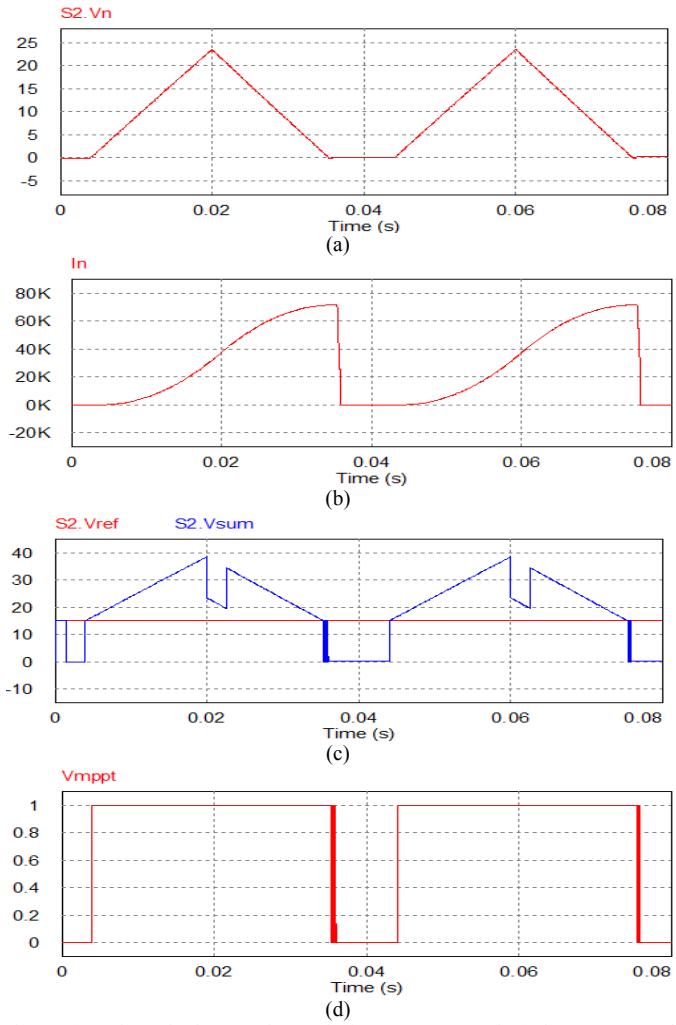


Fig. 4. (a) Triangular input voltage, (b) input current, and (c) the summer and reference voltages (d) PWM pulses to the boost converter.

C. DC-DC Boost Converter

To step up the unregulated voltage of the photovoltaic panel (~24V dc) to a desired output voltage for grid application (312V ac or 220V_{rms}) a boost converter has been designed and implemented. The designed parameters for the boost converter are listed in Table-I. The 312V output (V_{boost}) of the boost converter from a varying (~24V) dc source is demonstrated in Fig.5. Pulse width modulation (PWM) of gating signal is applied to the switch which is produced using the MPPT control strategy as described in section IIB. As the voltage fluctuates, the duty cycle also changes accordingly to produce the desired output voltage.

TABLE I. BOOST CONVERTER'S PARAMETERS

Symbol	*Actual Meaning*	*Value*
V_{in}	Given input voltage	24V
V_{out}	Desired average output Voltage	312V
f_s	Minimum switching frequency	21KHz
I_{LMax}	Maximum inductor current	230A
ΔI_L	Estimated inductor ripple current (2.6% of I_{LMax})	6A
ΔV_{out}	Desired output voltage ripple (0.1% of V_{out})	0.35V
I_{out}	Maximum output current(V_{out}/R)	10.4A

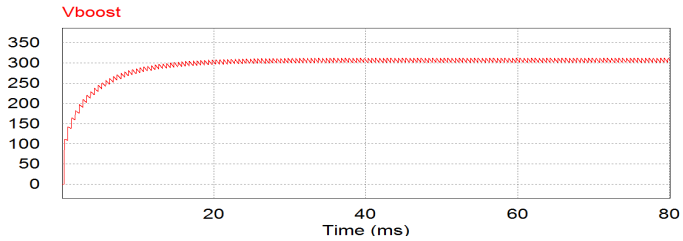


Fig. 5. Boost converter output.

D. H-bridge Inverter

An H-bridge inverter [6] is used to convert the 312V DC voltage of the boost converter to 312V_{peak} AC (220V_{rms}) voltage which matches the specifications for grid connection. Sinusoidal pulse width modulation (SPWM) is applied for operating the four transistors A,B,C and D.

E. Low Pass Filter

To match the output of the proposed PV system with the specification of the national grid of Bangladesh [6], an LC low pass filter is integrated in the design. The inductor and capacitor values of 95.95 mH and 106uF respectively are calculated using equations (1) and (2) for a characteristic impedance of 30Ω and a cut-off frequency of 50 Hz.

$$R = X_c = \frac{1}{2\pi f C} \quad (1)$$

$$f_c = \frac{1}{2\pi\sqrt{LC}} \quad (2)$$

III. RESULTS AND DISCUSSIONS

Figure 6 depicts the entire PSIM schematic of the proposed PV system. The non-sinusoidal output of the H-bridge inverter is shown in Fig.7(a), which contains high frequency components with a total harmonic distortion (THD) of 0.733. To achieve harmonics free output, a low pass L-C filter is incorporated with the proposed PV system. The 220V_{rms} and 50Hz pure sinusoidal output which emerges from low-pass LC filter is shown in Fig.7(b). After filtering, the THD significantly reduces from 0.733 to 0.002 which is much lower than the IEEE519 standard. The equation which is used to calculate THD is:

$$THD = \frac{\sqrt{Harmonics^2}}{Fundamental\ Voltage} \quad (3)$$

Fig. 8(a) and (b) shows the Fast Fourier Transform (FFT) of the inverter output before and after filtering. From the FFT analysis, it can be seen that extra harmonic spectra is present together with fundamental component at 50Hz. Nevertheless, after filtering the higher harmonics are seen to be vanished with only the 50 Hz component still remaining. This confirms that a pure sinusoidal wave has been produced.

The performance of the proposed photovoltaic inverter is tested by varying the output load from 5Ω to 100Ω at an interval of 5Ω as shown in Fig.9. The testing result shows that the output current remains almost constant with changing load, which confirms the applicability of the proposed PV system for dynamic loads.

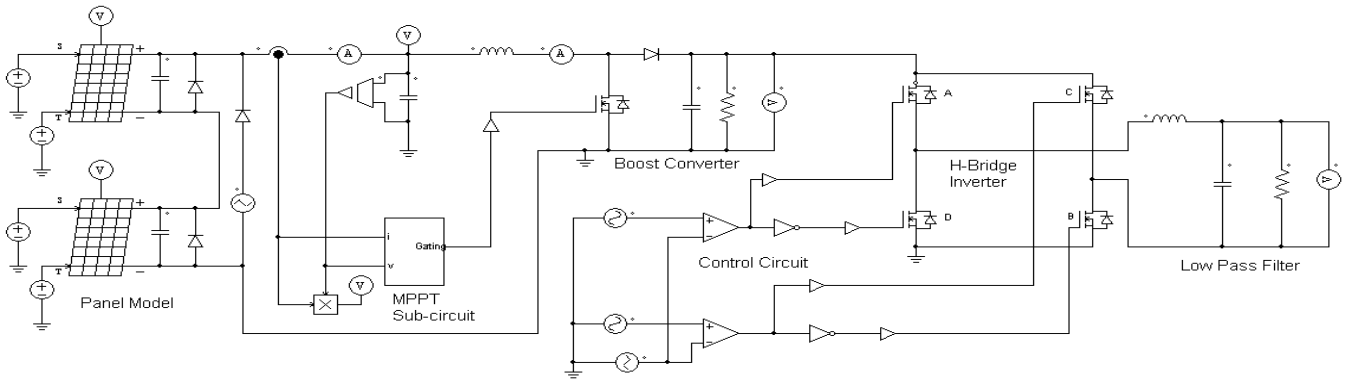
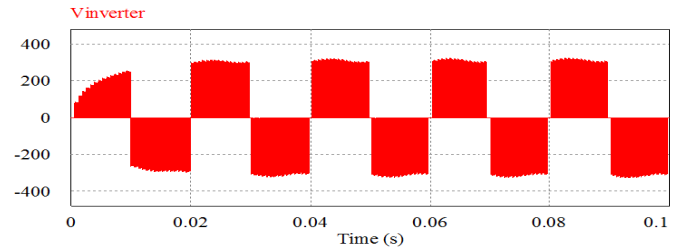
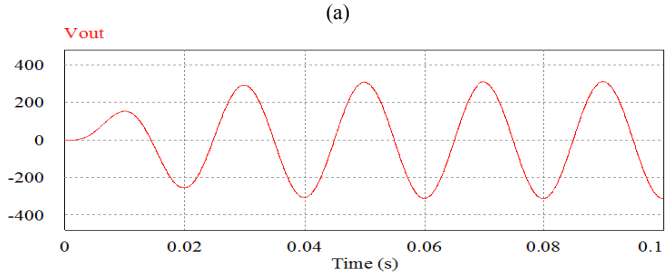


Fig. 6. Complete PSIM circuit of proposed PV system.

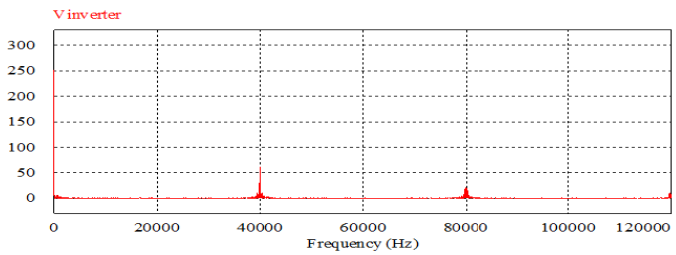


(a)

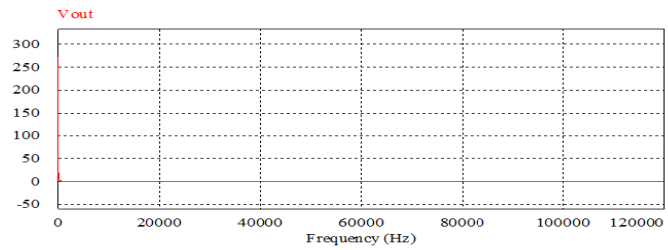


(b)

Fig. 7. (a) Output voltage before filtering and (b) after filtering.



(a)



(b)

Fig. 8. FFT of the inverter voltage (a) before filtering and (b) after filtering.

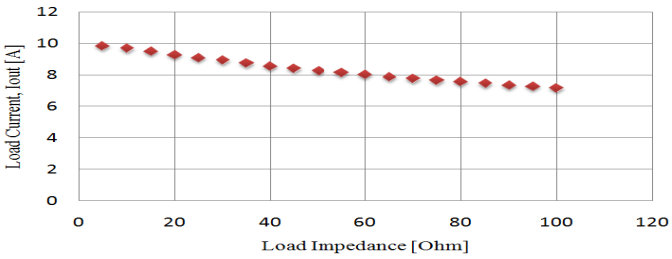


Fig. 9. Output current variations with increasing load.

IV. Conclusion

This paper demonstrates the design and simulation of a transformer-less PV system with an incorporated maximum power point tracking (MPPT). The designed MPPT circuit satisfies all the conditions of the IncCond algorithm and accurately adds or subtracts K voltage to the input voltage as per requirement of the situation. A perfect sinusoidal wave with 220 V_{rms} and 50 Hz is successfully achieved. After filtering, the THD is found to be 0.002 which is a lot lower than the IEEE 519 standard. The efficient functioning of the filter can be seen from the FFT of the inverter output prior to and after filtering. With varying load the output current remains more or less constant. This confirms the application of the proposed system for dynamic loads.

References

[1] D. Sera *et al.*, "Improved MPPT algorithms for rapidly changing environmental conditions," Proc. 12th International Conference on Power Electronics and Motion Control, 2006, pp. 1614-1619.

[2] J. A. Santiago-Gonzalez *et al.*, "Thre phase induction motor drive using flyback converter and PWM inverter fed from a single photovoltaic panel," Proc. IEEE PES General Meeting, pp. 1-6, 2011.

[3] T. Esram *et al.*, "Comparison of photovoltaic array maximum power point tracking techniques," IEEE Transactions On Energy Conversion, Vol. 22, No.2, June 2007 .

[4] N. Femia *et al.*, "Optimizing sampling rate of P&O MPPT technique," in Proc. IEEE PESC, 2004, pp. 1945- 1949.

[5] Zhou Yan *et al.*, "Study on realizing MPPT by improved incremental conductance method with variable step-size," IEEE 2008, pp. 547~550.

[6] A. S. K. Chowdhury *et al.*, "Design and implementation of a highly efficient sine-wave inverter for photovoltaic applications,"Proc. 2nd Intl Conference on Informatics, Electronics & Vision (ICIEV), May 2013.

An Endeavour to Find Two Unequal False Coins

Joydeb Ghosh[1], Piyali Datta[2], Arpan Chakraborty[2], Ankita Nandy[2], Lagnashree Dey[2], Rajat Kumar Pal[2], and Ranjit Kumar Samanta[1]

[1]*Department of Computer Science and Application, North Bengal University, Darjeeling – 734 013, West Bengal, India*
joydeb009, rksamantark@gmail.com

[2]*Department of Computer Science and Engineering, University of Calcutta*
92, A. P. C. Road, Kolkata – 700 009, West Bengal, India
piyalidatta150888, arpan250506, yoursankita.nandy09, rose2009mail, pal.rajatk@gmail.com

***Abstract*—The counterfeit coin problem is well-known and truly interesting in Computer Science, Game theory, and also in Mathematics. In this problem the objective is to detect the fake coin(s) of identical appearance but different weight in minimum number of comparisons. The word *counterfeit* most frequently describes forgeries of currency or documents, but can also describe software, pharmaceuticals, clothing, and more recently, motorcycles and cars, especially when these result in patent or trademark infringement. Finding one fake coin among *n* coins is tricky enough and complex. The problem becomes rigorous when there are two fake coins, as the false coin pair may form several different combinations that make the problem particularly tricky and complex to solve. In this paper we have developed a new algorithm for solving two versions of the two counterfeit coins problem in O(log *n*) time, where *n* is the number of coins given.**

Keywords: Coin counterfeiting; Problem; Weighing; Algorithm; Complexity; Applications.

I. Introduction

Computing a solution of the counterfeit coin problem has immense importance not only in theoretical domain but in commercial sphere as well to prevent forgery in different fields. The objective of a minimum number of weighing for which it is sufficient to determine the defective coin(s) in a set of n coins given that are identical in appearance but different in weight (either heavier or lighter than a true coin) using only an equal arm balance, when precisely the number of odd coins is known. In addition, the complexity of the problem increases with the increment of the number of false coins.

If p is the number of false coins in a set of n coins, it is not only sufficient to consider whether they are heavier or lighter than a true coin individually, but we must take into account their relative weights as well. In this paper we consider that there are two false coins in a set of n coins. By comparing the weight of the coins we have to detect the odd coins. Moreover, the objective is to identify the counterfeit coins using a minimum number of comparisons (or weighing) and also recognizing whether they are lighter or heavier in comparison to a genuine coin.

II. Literature Survey

In paper [4], the problem has been introduced in two ways. In the first case we do not know if there is a fake coin in the given set. The process is supposed to check it first, and if yes, then identify the targeted coin by means of a minimum number of weighing. In the second case, it is told that there is a counterfeit coin and the objective is only to find the coin through minimum number of weighing. At times, a standard coin may also be given.

In the first case, if a lighter coin is there in the given set S of n coins, then it is proved that the least number of weighing to find out the fake coin satisfies $3^{d-1} < n \leq 3^d$ for some unique value of d, where $|S| = n$, the cardinality of S. In the second case, we are given a set S of n coins plus a standard coin, where only one coin in S is of different weight. Then it is proved that $(3^{d-1}-1)/2 < n \leq (3^d-1)/2$.

Let S be a set of more than two coins, out of which only one is a counterfeit coin. In this case, the least number of weighing to find out the odd coin satisfies $(3^{d-1}-3)/2 < n \leq (3^d-3)/2$, for some sole value of d.

The paper claims that a lighter coin can be found in $\lceil \log_3 n \rceil$ weighing, if there are n coins in the set. Moreover, according to the paper, $\lceil \log_3(2n+1) \rceil$ is the complexity of the algorithm for solving the problem, if an additional standard coin is given; otherwise, the complexity is $\lceil \log_3(2n+3) \rceil$.

In paper [1], the problem has been addressed as an application of dynamic programming and the associated analysis has been made through optimal and suboptimal testing policy. Here also only one coin is defective out of n given coins. This technique always assumes $k < n$ coins in each pan for each weighing, where the value of k in effect depends on the value of n. If the two groups balance, the defective coin must be in the remaining $n-2k$ coins; otherwise, the false coin is in one of the k groups. After each weighing, the number of coins to be examined reduces, but the problem remains the same. This allows the authors to apply dynamic programming to this problem.

One classical solution is available in the form of a *decision tree* that represents a set of all possible decisions by which we can acquire the desired solution(s) of the problem [2, 3]. In this solution, each internal vertex (that is not a leaf vertex) symbolizes a comparison between a pair of equal sets of coins using an equal arm balance. Here the problem under consideration is more generalized; the fake coin can either be heavier or lighter. So, for n given coins, there are $2n$ leaf vertices in the tree as probable solutions.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

In paper [5], the problem of ascertaining the minimum number of weighing that suffices to determine the false (heavier) coins in a set of n coins of the same appearance, given a balance scale and the information that there are exactly two heavier coins present, has been considered. Both the heavier coins are of equal weight and they are not heavier by 1½ times than a true coin. The paper developed an algorithm with lower bound $\lceil \log_3({}^nC_2) \rceil$ on p, where p is the maximum number of comparisons required to find out the two false coins that are equally heavier. In this paper, an infinite set of n coins has been determined for which this lower bound is reached, and the upper bound is at most one more than the lower bound attained.

III. Problem Formulation and the Algorithm

In this section, we formulate and develop a comprehensive algorithm to solve two counterfeit coins problem. Here we are to devise an algorithm to find two false coins among n coins. Two issues are to be considered in this context. The minimum value of n as given coins is important such that we can find out the two false coins among them and the kind of false coins as heavier and/or lighter are also computed without using any additional standard coin.

As a counterfeit coin means that it is either heavier or lighter than a true coin, finding out two false coins introduces several cases, i.e., different combinations of false coin pairs, such as heavier-heavier, lighter-lighter, heavier-lighter, etc. In our paper we shall consider the case where one false coin is heavier and another is lighter than a true coin. Let us consider that a true coin is denoted as T and its weight is w(T), whereas a heavier and a lighter coin are denoted as H and L, respectively. Then, $w(H)-w(T) = \Delta H$ and $w(T)-w(L) = \Delta L$ convey the following cases that may arise in practice: $\Delta H > \Delta L$, $\Delta H < \Delta L$, and $\Delta H = \Delta L$.

In this paper, the first two cases are considered to solve the problem of two coins counterfeiting. By the way, the following lemma tells about the minimum number of n given coins out of which $p < n$ coins are fake coins.

Lemma 1: To find p coins counterfeiting among n given coins, n must be as minimum as $2p+1$.

Proof: If there is one false coin out of two coins, a true coin must be provided to detect the false coin unless the weight of the correct coin is given. So if the number of false coin(s) $p = 1$, the minimum value of $n = 3 = 2p+1$. Now, if there are two false coins we can identify them from four coins if and only if the two false coins are of different weight. If they are of same weight, we cannot conclude which pair of coins are true, as there are equal number of false and true coins. So, for $p = 2$, the minimum value of $n = 5$ (i.e., same as $2p+1$).

In general, if there are p counterfeit coins we can detect them from $p+2$ coins only if all the p coins are of distinct weights, but if at least two false coins are of same weight we cannot distinguish them among the set of genuine coins and false coins. So, to satisfy the above cases, specially the case where all the false coins are of same weight we need at least one more true coin than the false coin, i.e., the minimum number coins required is $p+p+1 = 2p+1$. □

Our algorithm proceeds by dividing the coins into three subsets $K1$, $K2$, and $K3$ such that the cardinality of set $K1$ is equal to the cardinality of set $K2$ are equal. At first these sets are put on the two pans of an equal arm balance for weighing. Depending on the result of this weighing and the specification of the counterfeit coins, i.e., both are equally heavier or equally lighter or one is heavier than the other, etc. which we get, we decide the subsequent set(s) to be compared next.

The algorithm starts by considering the fact that any positive integer n can be assumed into either of the three categories as follows: (i) n is divisible by 3, i.e., $n|3$, (ii) $n+1$ is divisible by 3, i.e., $(n+1)|3$, and (iii) $n-1$ is divisible by 3, i.e., $(n-1)|3$. Thus, the set of n coins can easily be checked to which group it belongs to and precisely which variant of the algorithm can be applied on the given set of coins. We may presume that the coins are indexed by natural numbers 1 through n (on any sequence of n coins). For the first case, $|K1| = |K2| = |K3| = n/3$. For the second case, $|K1| = |K2| = (n+1)/3$ and $|K3| = n-2(n+1)/3 = (n-2)/3$. So, there is a difference of one coin between $K1$ (or $K2$) and $K3$. For the third case, $|K1| = |K2| = (n-1)/3+1 = (n+2)/3$ and $|K3| = n-2(n+2)/3 = (n-4)/3$. So, there is a difference of two coins between $K1$ (or $K2$) and $K3$.

After creating the sets $K1$, $K2$, and $K3$, at the beginning sets $K1$ and $K2$ are compared. Depending on the outcome, three versions of the algorithm proceed towards the next weighing taking different sets for finding out the false coins. At each internal node, some weighing is performed considering the result of its parent node. At a leaf node we perform two operations: OCPH(K_i) (OCPL(K_i)) that denotes the operation to locate one heavier (lighter) counterfeit coin in set K_i. We call it whenever we are sure that there is only one heavier (lighter) false coin in K_i. When we are convinced that there are two false coins in K_i, then the algorithm is recursively applied to set K_i, considering its steps of operation as TCP(K_i) in general; here OCP (TCP) ≡ one (two) counterfeit coin(s) problem.

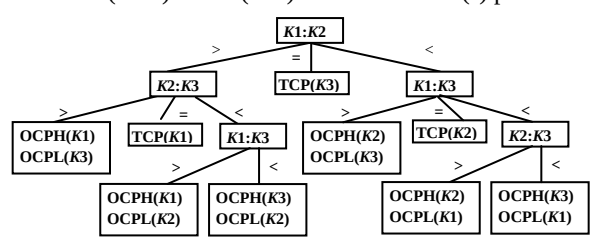
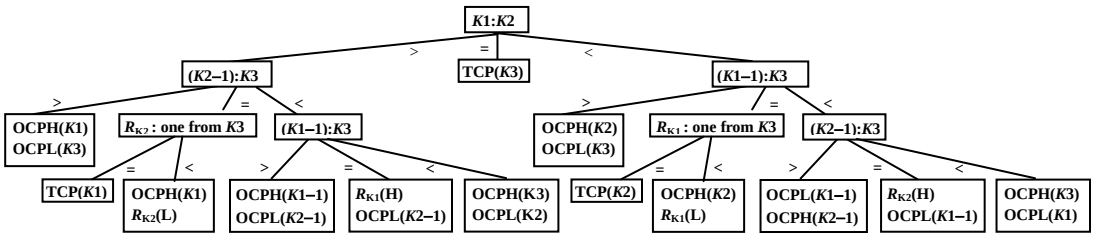


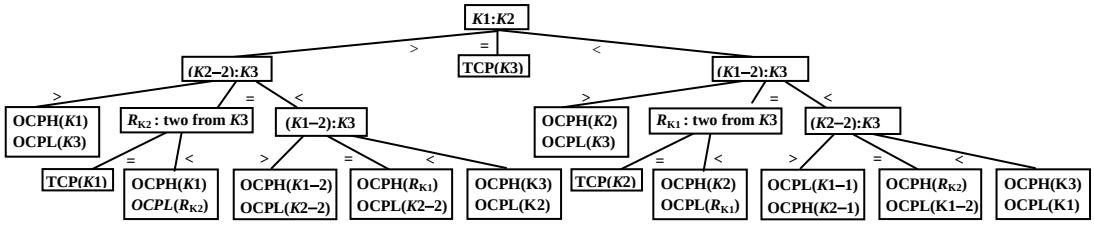
Fig. 1: The decision tree of the algorithm for $n|3$.

If $n|3$, then the decision tree that we obtain following the algorithm is shown in Fig. 1. At the root node, $K1$ and $K2$ are compared and the result is analyzed as follows. If $w(K1) = w(K2)$, then we are certain that they contain only true coins; this is because if any (one) pan contains at least one false coin, they would never equalize. Hence we can conclude that $K3$ contains both the false coins. So we perform TCP on the set $K3$ of coins. On the other hand, if $w(K1) > w(K2)$, then the possibilities are:

- $K1$ contains the heavier coin while $K2$ contains the lighter one and $K3$ contains only the true coins.



(a)



(b)

Fig. 2: (a) The decision tree of the algorithm for $(n+1)|3$. (b) The decision tree of the algorithm for $(n-1)|3$.

- Both the false coins are in $K1$, as the sum of the two false coins exceeds the sum of two true coins being $\Delta w(H) > \Delta w(L)$.
- $K1$ contains the heavier coin and $K3$ contains the lighter coin while $K2$ is the set of true coins only.
- $K1$ contains only true coins and $K2$ contains the lighter coin while the heavier coin is in $K3$.

Since $K2$ does not contain both the false coins (as $w(K1) > w(K2)$), next we check $K2$ against $K3$. If they are equal, $K1$ contains both the false coins and we call TCP($K1$). On the other hand, if $w(K2) > w(K3)$, the heavier coin must be in $K1$, as in the previous level of comparison $K1$ is heavier than $K2$ and the lighter coin is in $K3$. Accordingly, we call OCPH($K1$) and OCPL($K3$) at the leaf. If $w(K2) < w(K3)$, we are sure that the lighter coin is in $K2$ as this set is lighter than both $K1$ and $K3$, but to be definite of the heavier coin, we have to compare $K1$ and $K3$ now. If $w(K1) > w(K3)$, the heavier coin is in $K1$; otherwise, this is in $K3$. Following the result of this weighing, we either call OCPH($K1$) and OCPL($K2$), or OCPH($K3$) and OCPL($K2$) at the associated leaves.

When we call OCPL(K_i) or OCPH(K_i), the algorithm proceeds dealing with the set of coins(K_i) and we divide it into two equal subsets for further weighing. If this set contains even number of coins we put half of them on the left pan and the remaining half on the right pan and weighing (i.e., comparison) is performed. Again, if we are searching for a heavier coin at this stage, i.e., in case of OCPH(), after the weighing we deal only with the coins in the heavier pan. If the number is odd, we divide them into two equal halves and one coin remains out of weighing. If the weighing results in inequality, we focus on the pan with more weight for subsequent assessment.

The second variant of the algorithm is applicable when $(n+1)$ is divisible by 3. In this case the algorithm follows the same logic as for the version n|3 except the second level of comparisons, it takes $(|K2|-1)$ coins instead of $|K2|$ and $(K1-1)$ coins instead of $|K1|$. To make the algorithm deterministically faster, we always take the coins from set K_i except the last coin in the set. The comparisons in the successive levels are shown in the form of a decision tree in Fig. 2(a).

There is another difference of this version of the algorithm. In the left (or right) branch from the root of the decision tree, $K2-1$ (or $K1-1$) coins are compared to that of $K3$, and if it results in equality, we cannot conclude whether both the false coins are in $K1$ (or $K2$), though we are sure that $K3$ contains only true coins. The reason is that one coin from set $K2$ (or $K3$) is still remaining to be compared. So, in the third level, the remaining coin is compared to the first coin of $K3$ (which is obviously a true coin). Depending on the result of this comparison we can conclude that either both the false coins are in $K1$ (or $K2$) or a false coin is in $K1$ (or $K2$) and the other false coin is in $K2$ (or $K1$).

For $(n-1)|3$, the algorithm proceeds in the same way as for the version $(n+1)|3$ with a difference in the second level of comparison in the left and right branch of the decision tree. As $K3$ contains two coins less than that of $K2$ (or $K1$), it takes $(|K2|-2)$ (or $(|K1|-2)$) coins instead of $|K2|$ (or $|K1|$). The operations in the consecutive levels are depicted in the form of decision tree as shown in Fig. 2(b).

So far we have developed an algorithm to find out two false coins among a set of n given coins where one is heavier and the other is lighter satisfying the condition $\Delta H > \Delta L$. For the variant, i.e., $\Delta H < \Delta L$, the algorithm works as well. The only difference lies within the fact that the sum of both the false coins is less than the weight of two true coins. Considering this fact, the algorithm chooses the set to be weighed at each level of comparison.

IV. Experimental Results

In this section, we decide to accept some values for n such that it covers all the three categories for the subdivision of

n and calculate the average number of comparisons required. Here, we have considered all the possibilities, i.e., any coin may be a counterfeit coin, either heavier or lighter, and it could be placed at any indexed location of the given sequence of n coins. To compute the average case complexity we have to consider all the possible false coin pairs. Thus for a given value of n, there are ${}^nC_2$ pairs of combinations as for coin pair $\{i, j\}$, the ith coin could be heavier whereas the jth one is lighter, or the reverse. Hence, there are $2\times{}^nC_2$ leaves in the decision tree.

The results show that on an average the number of comparisons required is O(log n).

Table 1: The average number of comparisons to find out two counterfeit coins, for some values of n.

Number of coins, n	Total number of comparisons, S	Possible number of false coin combinations, C ($2\times{}^nC_2$)	Average number of comparisons, AVG = S/C
5	72	20	3
9	330	72	4
10	484	90	5
11	536	110	4
18	1974	306	6
20	2354	380	6
27	5346	702	7
29	6064	812	7
36	10572	1260	8
46	17966	2070	8
54	27324	2862	9
56	28172	3080	9
72	52572	5112	10
82	66314	6642	9
100	113582	9900	11
108	133056	11556	11
144	252852	20592	12
198	520602	39006	13
200	517606	39800	13

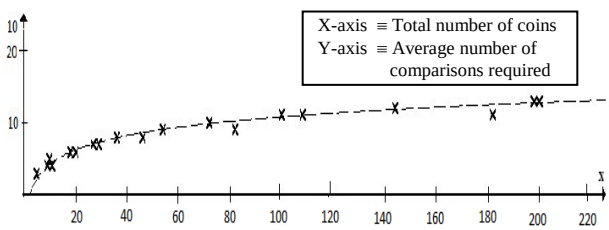


Fig. 3: Plot of average number of comparisons required against the number of coins given.

V. Computational Complexity

At each iteration, n is divided into nearly three equal parts, and the cardinality of the set of coins on which the operations are performed, always reduces by a factor of three. Let us consider the case $n|3$. As we see at the ith level each set is of cardinality $n/3^i$. Now if we reach to a set with five coins, then we can solve it using only four comparisons. So, let at ith level of comparison the cardinality of the set reduces to five. Thus, $n/3^i = 5$, i.e., $3^i = n/5$. Hence, $i = \log_3(n/5)$. Again, if TCP(K_i) is applied at each iteration before reaching a set with five coins, then $2\times i$ comparisons are required resulting in a total of $2\times i+4$ comparisons.

If OCPH(K_i) or OCPL(K_i) is applied at the kth level of comparison, then it is sure that before this iteration TCP() is applied for (k–1) times. We know that OCPH() or OCPL() requires $\lceil \log_2 n \rceil$ comparisons and at the kth level it is to be applied on $n/3^k$ coins. Hence, it would require total number of $2|K|+2\log_3(n/3^k)$ comparisons. So, in the worst case it would take $O(2|K|+2\log_3(n/3^k))$ + $O(2\times\log_3(n/5)+4)$, i.e., O(log n) comparisons as a whole.

VI. Applications of the Problem

The application of the problem under consideration is not only limited to coins, but to any kind of valued article that are frequently counterfeited, like ornaments, watches, metal (like gold), jewellery, etc. The word *counterfeit* most frequently describes forgeries of currency or documents, but can also describe pharmaceuticals, software, clothing, and more recently, motorcycles and cars, especially when these result in patent or trademark infringement.

Coin counterfeiting occurs regularly in the antique coin market, but there are various modern forgeries that also make it into general circulation [3, 4]. Counterfeit antique coins are generally made to a very high standard so that they often fool collectors; this is not easy and many coins still stand out.

VII. Conclusion

In this paper we have introduced an algorithm to detect two false coins among a set of n coins with identical appearance. In this case we have assumed that one of the false coins is heavier than the true coin while the other false coin is lighter. Two variations of the problem, i.e., the weight difference between the heavier and the true coin may be greater (less) than the difference in weight of the lighter and the true coin, may be solved using this algorithm. The algorithm solves the problem with time complexity O(log n). The very fact that the decision tree structure can be used to solve such problems of large size, by eliminating a part of the solution domain after each step of decision making, is of utmost importance. Especially as our algorithm works for any value of n, it does not matter if the value of n is not known a priori. This is, in fact, a new innovative work, and important in solving problems using the decision tree structure.

References

1. R. Bellman and B. Gluss, On Various Versions of the Defective Coin Problem, *Information and Control*, vol. 4, no. 2-3, pp. 118-131, 1961
2. J. Ghosh, P. Senmajumdar, S. Maitra, D. Dhal, and R. K. Pal, A Generalized Algorithm for Solving *n* Coins Problem, *Proc. of 2011 IEEE International Conference on Computer Science and Automation Engineering (CSAE 2011)*, Shanghai, China, vol. 2, pp. 411-415, Jun. 10-12, 2011.
3. J. Ghosh, P. Senmajumdar, S. Maitra, D. Dhal, and R. K. Pal, Yet Another Algorithm for Solving *n* Coins Problem, *Assam University Journal of Science & Technology: Physical Sciences and Technology*, ISSN: 0975-2773, vol. 8, no. II, pp. 118-125, 2011.
4. B. Manvel, Counterfeit Coin Problems, *Mathematics Magazine*, Mathematical Association of America, vol. 50, no. 2, pp. 90-92, 1977.
5. R. Tošić, Two Counterfeit Coins, *Discrete Mathematics*, North-Holland, vol. 46, pp. 295-298, 1983.

The First Algorithm for Solving Two Coins Counterfeiting with ω(ΔH) = ω(ΔL)

Joydeb Ghosh[1], Arpan Chakraborty[2], Piyali Datta[2], Lagnashree Dey[2], Ankita Nandy[2], Rajat Kumar Pal[2], and Ranjit Kumar Samanta[1]

[1]*Department of Computer Science and Application, North Bengal University, Darjeeling – 734 013, West Bengal, India*
joydeb009, rksamantark@gmail.com

[2]*Department of Computer Science and Engineering, University of Calcutta*
92, A. P. C. Road, Kolkata – 700 009, West Bengal, India
arpan250506, piyalidatta150888, rose2009mail, yoursankita.nandy09, pal.rajatk@gmail.com

***Abstract*—Counterfeit coin problem is of utmost importance and it is truly interesting in Computer Science and Game theory as well as in Mathematics. In this problem the objective is to detect the fake coin(s) of identical appearance but of different weight in minimum number of comparisons. The word *counterfeit* is most frequently applicable to forgeries of currency or documents, but can also describe software, pharmaceuticals, clothing, and more recently, motorcycles and other vehicles, especially when these result in patent or trademark infringement. In this paper we have developed a new algorithm for solving two counterfeit coins problem in linear time, where *n* is the total number of coins given. However, this is the first algorithm that identifies and solves the problem, given the false coins with type ω(ΔH) = ω(ΔL), i.e., one false coin is heavier and another is lighter than a true coin, and their difference in weight from the true coin is equal. However, this is the degenerate case in the field of two counterfeit coins problem.**

Keywords: Counterfeit coin problem; Two coins counterfeiting; Weighing; Algorithm; Complexity; Applications.

I. Introduction

Computing a solution of the counterfeit coin problem has immense significance and relevance not only in theoretical domain but in commercial sphere as well to prevent forgery in different. The objective of a minimum number of weighing for which it is sufficient to determine the defective coin in a set of n coins given that are identical in appearance but different in weight (either heavier or lighter than a true coin) using only an equal arm balance, when precisely there is one odd coin present. By comparing the weight of the coins we have to detect the odd one. In this paper, our objective is to identify two false coins using a minimum number of comparisons when the surplus amount in a heavier coin is same as the amount deficient in a lighter coin comparing a true one.

II. Literature Survey

In paper [1], the problem has been introduced as an application of dynamic programming and the associated analysis has been made through optimal and suboptimal testing policy. It also considers single counterfeit coin problem. This technique always assumes $k < n$ coins in each pan for each weighing, where the value of k essentially depends on the value of n. If the two groups balance, the flawed coin must be in the remaining $n–2k$ coins; otherwise, the false coin is in one of the k groups. After each weighing, the number of coins to be examined reduces, but the problem remains the same. This allows the dynamic programming concept to solve this problem.

In paper [4], the author introduces the problem in two ways. In the first case, it is not sure whether there is a fake coin in the given set. The algorithm is supposed to check it first, and if any, then identify the false coin by means of a minimum number of weighing. In the second case, it is known that there is a counterfeit coin and the objective is to find the coin through minimum number of weighing. At times, a standard coin may also be given. In the first case, if a lighter coin is there in the given set S of coins, then it is proved that the least number of weighing to find out the fake coin satisfies $3^{p-1} < |S| \leq 3^p$ for some unique value of p, where $|S| = n$ is the cardinality of set S. In the second case, we are given a set S of coins plus a standard coin, where only one coin in S is of different weight. Then it is proved that $(3^{p-1}–1)/2 < |S| \leq (3^p–1)/2$.

Let S be a set of more than two coins, out of which only one is a counterfeit coin. In this case, the least number of weighing to find out the odd coin satisfies $(3^{p-1}–3)/2 < |S| \leq (3^p–3)/2$, for some sole value of p. The paper claims that a lighter coin can be found in $\lceil \log_3 k \rceil$ weighing, if there are k coins in a set. According to the paper, $\lceil \log_3(2k+1) \rceil$ is the complexity of the algorithm for solving the problem, if an additional standard coin is given; otherwise, the complexity is $\lceil \log_3(2k+3) \rceil$.

In papers [2] and [3], one classical solution for single counterfeiting is available in the form of a *decision tree* that represents a set of all possible conclusions by which we can acquire the desired solution(s) of the problem. In this tree, each internal vertex (that is not a leaf vertex) symbolizes a comparison between a pair of same number of coins using an equal arm balance. Here the problem under consideration is more generalized; the fake coin can either be heavier or lighter. So, for n given coins, there are $2n$ leaf vertices in the tree as probable solutions.

In paper [5], the problem of discovering the minimum number of weighing that suffices to determine the forged (heavier) coins in a set of n coins of the same appearance,

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

given a balance scale and the information that there are exactly two heavier coins present, has been considered. Moreover, both the heavier coins are of equal weight and they are not heavier than 1.5 times than a true coin. If P is the maximum number of comparisons required to find out two false coins (equally heavier), the paper introduces an algorithm which has the lower bound $\lceil \log_3({}^nC_2) \rceil$. In this paper an infinite set of n coins, has been determined, for which this lower bound is reached, and the upper bound computed is merely one more than the lower bound.

III. Problem Formulation and the Algorithm

In this section, we formulate an effective algorithm to solve two counterfeit coins problem, i.e., to find two false coins among n coins. Two issues are to be considered in this context. The minimum value of n such that we can find two false coins among them without using any spare genuine coin and the kind of false coins, i.e., heavier and/or lighter than a true coin through their mutual comparisons of weight. The term 'mutual comparison' is explained as follows.

By a false coin we mean that it is either heavier or lighter than a true coin. Finding out two false coins introduces several cases, i.e., different combinations of false coin pairs. Let T denote a true coin whose weight is w(T). Also heavier and lighter coins are denoted as H and L, respectively. Then, $\omega(H)-\omega(T) = \omega(\Delta H)$ and $\omega(T)-\omega(L) = \omega(\Delta L)$. Two coins counterfeiting introduces several other cases but when $\omega(\Delta H) = \omega(\Delta L)$, it becomes very much troublesome in the sense that the false coin pair always gets the weight-sum of two true coins. Thus at any level of weighing if the false coin pair is at one side of the balance, from that weighing we possibly fail to conclude about the pan where both the false coins may reside. In this paper we consider the case where one false coin is heavier and the other is lighter than a true coin and their individual difference in weight from a true coin is equal, which is the most typical case in the realm of two counterfeit coins problem as we are even puzzled about a set of coins that may contain them.

The algorithm begins by dividing the coins into three sets $K1$, $K2$, and $K3$ such that the cardinality of set $K1$ and $K2$ are equal. At first $K1$ and $K2$ are put on the arms for weighing. Depending on the result(s) of this weighing and the specification of the counterfeit coins, we decide to choose the set that (may) contain the false coin(s).

The algorithm also considers the fact that any positive integer n can be categorized into either of the following three cases: (i) n is divisible by 3, i.e., $n|3$, (ii) $n+1$ is divisible by 3, i.e., $(n+1)|3$, and (iii) $n-1$ is divisible by 3, i.e., $(n-1)|3$. Thus, the set of n coins can easily be checked to which group it belongs to and precisely which variant of the algorithm can be applied on the given set of coins. We assume that the coins are indexed by natural numbers 1 through n. For the first case, $|K1| = |K2| = |K3| = n/3$. For the second case, $|K1| = |K2| = (n+1)/3$, and $|K3| = n - 2(n+1)/3 = (n-2)/3$. So, there is a difference of one coin between sets $K1$ (or $K2$) and $K3$. For the third case, $|K1| = |K2| = (n-1)/3+1 = (n+2)/3$ and $|K3| = n-2(n+2)/3 = (n-4)/3$. Hence, there is difference of two coins between $K1$ (or $K2$) and $K3$.

After creating the sets $K1$, $K2$, and $K3$, $K1$ and $K2$ are first compared. Depending on the outcome, three versions of the algorithm proceed towards the next weighing taking different sets. At each internal vertex some weighing are performed considering the result of its parent vertices. At each leaf vertex we perform two operations: OCPH(K_i) (OCPL(K_i)) that denotes the operation to find one heavier (lighter) counterfeit coin in set K_i being sure that there is only one heavier (lighter) coin in K_i . When we know that there are two false coins in K_i, the algorithm is recursively applied to K_i, considering its version TCP(K_i) that tells that set K_i contains both the counterfeit coins.

If $n|3$, then the decision tree we obtain for this version of the algorithm is shown in Fig. 1. At the root of the tree, $K1$ and $K2$ are compared and the result is analyzed as follows. If $\omega(K1) > \omega(K2)$, there are several possibilities:

- The heavier coin may be in $K1$ whereas the lighter coin may be in $K2$.
- The heavier coin may be in $K1$ but $K2$ contains only true coins.
- The lighter coin may be in $K2$ but $K1$ contains only true coins.
- Both the false coins are definitely not on the same pan.

Now we like to check the type of the coin set $K3$ by comparing it with set $K2$. So in the second level of comparison if we obtain $\omega(K2) > \omega(K3)$, again there are some possibilities:

As set $K2$ can never contain a heavier coin, from the outcome of this comparison it is clear that set $K2$ is the set of true coins only. Hence, $K3$ certainly contains a lighter coin and $K1$ definitely contains a heavier coin. Thus, OCPL($K3$) and OCPH($K1$) are applied.

Again, if at the second level of comparison we find that $\omega(K2) < \omega(K3)$, then no further information is gained for $K2$ as at the root level comparison it is clear that it can contain either only true coins or a lighter coin, and $K3$ contains either only true coins or a heavier coin. Thus we go on the third level, and compare $K1$ and $K3$.

If $\omega(K1) > \omega(K3)$, $K3$ cannot contain a heavier coin and thus it is the set of true coins only and a heavier coin is in set $K1$ and $K2$ contains a lighter coin. So, we apply OCPH($K1$) and OCPL($K2$). If $\omega(K1) < \omega(K3)$, then $K1$ contains only true coins as it would never contain a lighter one. So, a heavier coin is in set $K3$ and a lighter coin is in set $K2$, and hence we apply OCPH($K3$) and OCPL($K2$).

Now we check the root level comparison when $\omega(K1) < w(K2)$; we compare $K1$ and $K3$ at the second level.

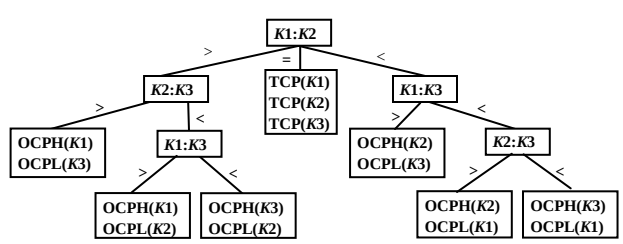


Fig. 1: The decision tree of the algorithm when $n|3$.

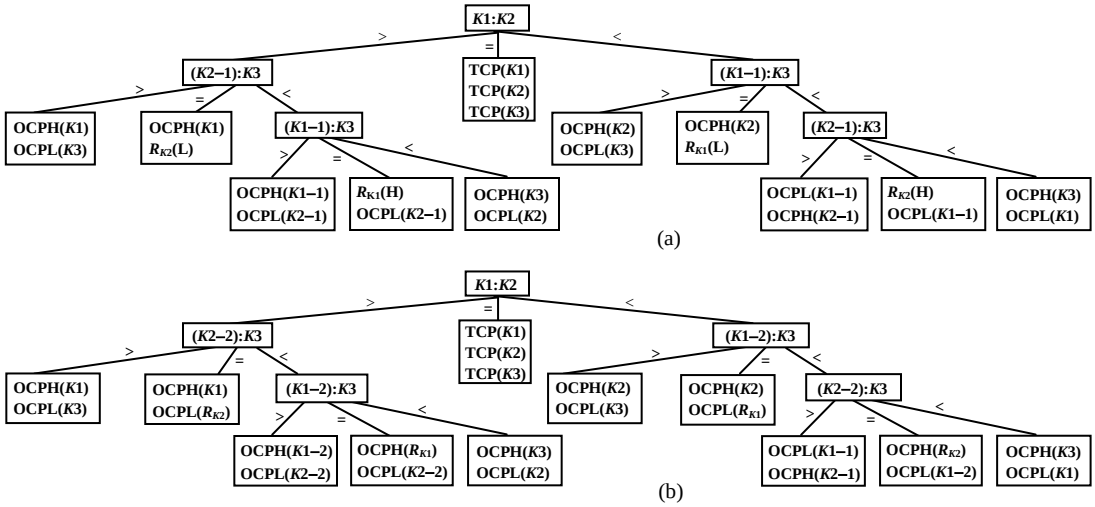


Fig. 2: (a) The decision tree of the algorithm for $(n+1)|3$. (b) The decision tree of the algorithm for $(n-1)|3$.

As previously addressed, we consider the possibilities and apply the same logic to reach at a leaf vertex where OCPH() and OCPL() are called for the respective sets of coins in K_i.

The expected most critical situation occurs if at the root level we find $\omega(K1) = \omega(K2)$, then we know both $K1$ and $K2$ are the true coin sets (and $K3$ contains both the false coins), or both the false coins are in any one of the sets $K1$ and $K2$. So, there is no scope in the reduction of sample space. Hence, we have to apply TCP() on the three sets $K1$, $K2$, and $K3$ one after the other. It may so happen that TCP(K_i) reports both the false coins or no existence of the false coins in K_i. We also apply TCP(K_{i+1}) if and only if TCP(K_i) reports that K_i contains no false coins.

If $(n+1)|3$, the algorithm does the same thing as for the version n|3 except the second level of comparisons; it takes only $(|K2|-1)$ coins instead of $|K2|$ and $(|K1|-1)$ coins instead of $|K1|$. For the sake of ensuring determinism we always take the first $|K_i-1|$ coins from the set K_i. The operations or tasks executed in the consecutive levels are depicted in the decision tree as shown in Fig. 2(a).

There is another difference of this version from the previous one. In the left (or right) branch from the root, $K2-1$ (or $K1-1$) coins are compared to that of $K3$, and if they result in equality, a false coin is in $K1$ (or $K2$) and the other false coin is in set $K2$ (or $K1$). If $\omega(K2-1) < \omega(K3)$, we are certain that the lighter coin is in set $K2$ or set $(K2-1)$. To be definite about the heavier coin, we compare $(K1-1)$ coins with that of set $K3$ at the third level. If $\omega(K1-1) > \omega(K3)$, the heavier coin is in $(K1-1)$; or else, it is in $K3$. If this is not the case, the remaining coin of set $K1$ that has not been weighed in the third level of comparison, is the heavier coin. Accordingly, OCPH() is called on the corresponding set of coins.

The algorithm for the values of n satisfying $(n-1)|3$ proceeds in the same way as for the version $(n+1)|3$ with a little difference in the second level of comparison in the left and right branches of the decision tree. As $K3$ contains two coins less than that of $K1$ or $K2$, it takes $(|K2|-2)$ coins instead of $|K2|$ and $(|K1|-2)$ coins instead of $|K1|$. The operations performed in the consecutive levels are revealed in the decision tree shown in Fig. 2(b).

IV. Experimental Results

In this section, we settle on using some relevant values of n such that it covers all the three categories for the subdivision of n and calculate the average number of comparisons required. Here, we have considered all the possibilities, i.e., any coin may be a counterfeit coin, either heavier or lighter, and it could be placed at any indexed location of the given sequence of n coins. To compute the average case complexity, we have considered all the possible false coin pairs. Thus, for a given value of n, there are $2\times{}^{n}C_2$ combinations as ith coin heavier and jth coin lighter is different from the combination ith coin lighter and jth coin heavier. Hence, there are $2\times{}^{n}C_2$ leaves in the decision tree. The results in Table 1 show that on an average the number of comparisons required is O(log n).

Table 1. Average number of comparisons for different values of n.

Number of coins, n	Total number of comparisons, S	Possible number of false coin combinations, $C = 2\times{}^{n}C_2$	Average number of comparisons, AVG = S/C
3	16	6	2
5	54	20	2
10	483	90	5
18	2064	306	6
26	5130	650	7
37	15777	1332	11
45	26436	1980	13
53	33780	2756	12
64	47468	4032	11
72	64704	5112	12
80	77502	6320	12
91	166586	8190	20
99	253965	9702	26
107	359674	11342	31
127	492976	16002	30
144	525873	20592	25
161	549668	25760	21
181	717268	32580	22
197	863108	38612	22
198	872526	39402	22
200	884954	39800	22

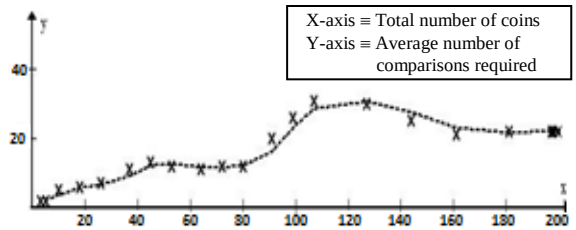


Fig. 3: Variation of average number of comparisons required against the number of coins.

V. Computational Complexity

At each iteration, n is divided into three nearly equal parts, and the cardinality of the set of coins on which the operations are essentially performed, always reduces by a factor of three. We notice that at the ith level each reduced set has cardinality $n/3^i$. The best case occurs when the root level comparison results in inequality. The subsequent comparisons lead to a leaf node where only OCPH() and OCPL() are applied on two different sets and each of them takes O(log n) time. OCPH() or OCPL() can be applied at one of the leaves that takes at most three comparisons from the root vertex which is a constant. Hence the best case complexity is O(log n).

On the other hand, if the root level comparison results in equality, all the coin sets are suspected to contain the false coin pairs. Thus the sample space to be further considered is not reduced. In our algorithm, TCP(K_i), TCP(K_{i+1}), and TCP(K_{i+2}) are applied sequentially only when the first comparison shows equality in each recursion of TCP(). As at the beginning we divide the initial set into $K1$, $K2$, and $K3$, we call TCP($K2$) if and only if TCP($K1$) reports 'no false coin', and we call TCP($K3$) if and only if both $K1$ and $K2$ contain no false coin. Now if we reach to a set with only 5, or 4, or 3 coins, then we may conclude whether the corresponding set contains either both the false coins or none of them. It takes at most four comparisons. Thus, in the worst case to find the two false coins we have to apply TCP() on $K1$, $K2$, and $K3$ that requires O(n) time.

Lemma 1: The algorithm for two coins counterfeiting takes linear time, when $\omega(\Delta H) = \omega(\Delta L)$.

Proof: We assume that the algorithm uses only equal number of coins on both the pans. The case $\omega(\Delta H) = \omega(\Delta L)$ leads to a confusion as the sum of the weights of two false coins of that kind results in the weight-sum of two correct coins. If there are n coins in a set, there are at most $2\times{}^{n}C_2$ combinations of the two false coins' position in any indexed sequence of n coins. Now at any instant of the algorithm, there might have equal number of coins shared by two pans of the equal arm balance, and hence both the false coins may reside in any one of the pans. The purpose of this algorithm is to verify two sets of coins together and depending on the weighing it proceeds to the next step by reducing the number of coins in each set next to be considered for further checking. However, if both the false coins are always in one set, i.e., in one pan, the checking of the sets results in equality. If we divide the given set of coins into $p \leq n/2$ subsets and weigh them pairwise, we may get all these sets equal in weight. Thus, the weighing process may result in equality until we reach a case where a set of coins under concern has only one false coin that certainly results in inequality; here the false coins may share the pans or at least one of them must be in one pan. Hence, the sample space cannot be reduced below a factor of n. So, running time of the algorithm is linear. □

VI. Applications of the Problem

The word *counterfeit* most frequently describes forgeries of currency or documents, but can also describe software, pharmaceuticals, clothing, watches, effigies, and recently bikes and automobiles, especially when these result in intellectual property like patent or trademark infringement. Coin counterfeiting occurs regularly in the antique coin market, but there are various modern forgeries that also make it into general circulation [1, 3]. Counterfeit antique coins are generally made to a very high standard so that the collectors often become fool; this is not easy and many coins still be conspicuous. The application of the problem under consideration is not only limited to coins, but to any kind of valued article that are frequently counterfeited, like ornaments, timepieces, metal (like gold), jewellery, etc.

VII. Conclusion

In this paper we have introduced an algorithm to detect two false coins among a set of n coins with identical appearance. In this case we have assumed that one false coin is heavier and the other is lighter than a true coin and their difference in weight from the true coin is equal. We addressed that this is the most critical case in the domain of two coins counterfeit for its inherent characteristics. This is the first algorithm that considers this type and solves the problem in linear time. The very fact that the decision tree structure can be used to solve such problems of large size, by eliminating a part of the solution domain after each step of decision making, is of utmost upshot. Especially as our algorithm works for any value of n, it does not matter if the value of n is not known a priori. This is, in fact, a new innovative work, and important in solving problems using the decision tree structure.

References

[1] R. Bellman and B. Gluss, On Various Versions of the Defective Coin Problem, Information and Control, vol. 4, no. 2-3, pp. 118-131, 1961

[2] J. Ghosh, P. Senmajumdar, S. Maitra, D. Dhal, and R. K. Pal, A Generalized Algorithm for Solving n Coins Problem, *Proc. of 2011 IEEE International Conference on Computer Science and Automation Engineering (CSAE 2011)*, Shanghai, China, vol. 2, pp. 411-415, Jun. 10-12, 2011.

[3] J. Ghosh, P. Senmajumdar, S. Maitra, D. Dhal, and R. K. Pal, Yet Another Algorithm for Solving *n* Coins Problem, Assam University Journal of Science & Technology: Physical Sciences and Technology, ISSN: 0975-2773, vol. 8, no. II, pp. 118-125, 2011.

[4] B. Manvel, Counterfeit Coin Problems, Mathematics Magazine, Mathematical Association of America, vol. 50, no. 2, pp. 90-92, 1977.

[5] R. Tošić, Two Counterfeit Coins, Discrete Mathematics, North-Holland, vol. 46, pp. 295-298, 1983.

Evaluating the Performance of the Practical Lattice Reduction Algorithms

Md Mokammel Haque,[1,*] Josef Pieprzyk,[2]
[1]Department of Computer Science and Engineering, Chittagong University of Engineering and Technology
Chittagong 4349, Bangladesh
[2]School of Electrical Engineering and Computer Science, Queensland University of Technology
Brisbane, QLD 4000, Australia
*malin405@yahoo.com

***Abstract*—We investigate the current state-of-art performance of lattice reduction algorithms. We perform experiments to examine the limit of the practical lattice reduction algorithms such as Recursive-BKZ. We report a better attack on SWIFFT lattices based on a recent improvement of BKZ reduction. We also participate in several lattice challenges that evaluate the performance of the lattice-reduction algorithms. Our experimental results thus compare and sometimes concretize the previous similar work in this area.**

***Index Terms*—Cryptosystem, Lattice Reduction, Lattice Challenge, BKZ, SWIFFT.**

I. Introduction

The *Lattices* we are interested in are integer or point lattices; discrete subgroups of $\mathbb{R}^n$. A lattice L is defined as all integer linear combinations of some linearly independent vectors $(\mathbf{b}_1, ..., \mathbf{b}_n)$ of the form: $\sum_{i=1}^{n} x_i \cdot \mathbf{b}_i$, where $x_i \in \mathbb{Z}$ for all i. The vectors $(\mathbf{b}_1, ..., \mathbf{b}_n)$ are called a basis of the lattice. The number of vectors n is the dimension of the lattice L.

The goal of lattice reduction is to find a "good" basis for the applications which is usually the one that contains short vectors. Clearly, it is an interesting problem to find a short basis or even to try to find a shortest basis. The SVP or Shortest Vector Problem is the most common and fundamental algorithmic lattice problem on which many cryptosystems rely. It is an NP-hard problem under randomized reductions [2].

Lattice problems are considered to be building blocks in many public-key cryptographic constructions. The most significant examples of cryptographic constructions that use lattices include: public key encryption based on knapsack problem; variant of RSA and DSA signatures schemes; encryption schemes based on LWE (Learning with Error) problems [3]; fully homomorphic cryptosystems [4]; the SWIFFT hash function [5] as well as many others. If one knows how to solve lattice problems, then one can break some cryptosystems.

Lattice reduction algorithms are one of the few exciting recent techniques in cryptography. The most popular lattice reduction algorithms are the LLL (Lenstra-Lenstra-Lovász) [6] and BKZ (Block Korkine-Zolotarev) [7] algorithms and their variants. They are called *approximate algorithms*; another category is known as *exact algorithms* includes Enumeration.

In this paper, we perform extensive experiments with the variants of BKZ algorithm. We study similar experimental work and summarize their results to better understand the current state of the practical lattice reduction algorithms. We evaluate one of the best variants of BKZ (aka Recursive-BKZ) through our participation in several Lattice Challenges[1]. Finally, we use the algorithm to improve the attack time in one of the SHA-3 contestant[2], SWIFFT hash function.

II. Related Experimental Work

In this section we review some of the experimental work on lattice reduction for various types of lattices. The first two works are on random lattices, followed by the work on SIS lattice and ideal lattice.

A. Gama-Nguyen Experiments

In 2008, Gama and Nguyen reported the results of their experiments on the actual behaviour of lattice reduction algorithms. The result published in [8] has been widely accepted since then and cited in several similar works. The experiments reveal the state of lattice reduction algorithms in terms of their efficiency and the accuracy of solutions. The work helps to close the gap between the theory and practice. One of the conclusions is that in a higher dimension, say $n > 100$, the Hermite factor (quality of reduction) achievable is no better than $\sim 1.01^n$. Interestingly, It is still more or less true for current lattice reduction techniques.

In their experiments, Gama and Nguyen used random lattice of instances at least 20 to get an average behaviour (reduction quality and time). The experiment is conducted on the NTL package [9] with version-5.4.1 on 1.7 GHz (64-bit processor) machine. The average Hermite factor they achieved for BKZ-28 is approximately 1.0109^n. They determined the experimental complexity for both the exact algorithm (exhaustive search in the enumeration [7]) and the approximate algorithm (BKZ). Using enumeration they were able to solve SVP within an hour for dimension 60 with BKZ-25 preprocessing. But the graph they constructed demonstrates that the algorithm fails to solve SVP instances of the dimension bigger than 100. For the BKZ algorithm they experiment with dimension (n) as high as 200 and block size, $k \in (10, 40)$. A summary of their results is given in Figure 12 and Figure 13 in [8]. It shows the average runtime required for higher dimension (≥ 100) with moderate block size (≤ 30) and for smaller dimension (< 100) with higher block size (> 30). They noticed a sudden increase of runtime for block around 20 to 25 in high dimension and this is because of the huge number of calls to the enumeration routine. This invalidates the use of higher block size with a high dimension for the original BKZ algorithm.

[1]www.latticechallenge.org
[2]http://csrc.nist.gov/groups/ST/hash/sha-3/

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

B. Schneider-Buchmann Experiments

The goal of the experiments performed by Schneider and Buchmann in [10] was to go beyond the results of Gama-Nguyen [8]. They experiment with the same BKZ lattice reduction algorithm using random lattices. Another goal was to find the right parameter (blocksize etc.) for the BKZ algorithm. In particular, they revisit a common practice in BKZ reduction, namely the increasing or progressive blocksize approach. The idea is to start from a smaller block and increase the blocksize successively on the same full-basis until a desired Hermite factor is achieved. They conduct an experiment to compare the performance of this approach with the normal approach (i.e. to start with a fixed higher blocksize). Their experiment shows that the direct or fixed block approach performs slightly better than the increasing blocksize approach for both quality and runtime.

We know that for a high dimension, the use of a large block size is far from being practical for current algorithms which is also mentioned in [8]. Schneider and Buchmann provide evidence confirming the observation by running experiments for 5 instances of random lattices using the NTL software. For dimension 90 and blocksize 50, on the average the required runtime is approximately 12 days.

They also show an approach to obtain the Hermite factor achievable for a particular block size. Using curve fitting and extrapolation on the graph of Hermite factor versus block size, they find the following function:

$$\delta(k) = 1.01655 - 0.000196185 \cdot k \tag{1}$$

where $\delta(k) = \left(\frac{\|\mathbf{b}_1\|}{\mathrm{vol}(L)^{1/n}}\right)^{1/n}$ is the root-Hermite factor for blocksize k and dimension n.

C. Buchmann et al. Experiments

In [11] Buchman et al. proposed a method to generate hard lattice instances. Their proposal is based on the popular construction of Ajtai [12] and therefore also relates to the SIS problem. They first launched the *lattice challenge* (see next section) site based on these hard lattices (*aka* SIS lattice) to solve the SVP problem. The goal was to evaluate the performance of the current lattice reduction algorithms. Based on the experiment of Micciancio and Regev in [13], they generated a table (Table 1 in [11]) to find the necessary root-Hermite factor δ_m required for a parameter set (lattice dimension m, reference dimension n and a modulus q). The table showed that for a SIS-lattice of dimension 700, the required δ_m is equivalent to 1.0055. They performed an experiment with NTL (version 4.3.2) on AMD Opteron 2.6 GHz single-core processor. They found that for dimension 700 using BKZ-20, the short vector length is approximately 148, which is equivalent to a root-Hermite factor of ≈ 1.0062. The corresponding runtime is $\approx \mathrm{e}^{8.5}$ in seconds.

D. Plantard-Schneider Experiments

Plantard and Schneider proposed the third challenge in this series on ideal lattices. In [14], they describe the rules of participating in this challenge with some practical experiments on ideal lattices. Recall that an ideal lattice is defined as a polynomial ring over integers. It has special properties and therefore experimental evidence is required to investigate this effect for major algorithms. They experiment with multiple instances of ideal lattices on AMD Opteron platform with 2.3 GHz processor.

TABLE I
SUMMARY OF RECENT PUBLISHED RESULTS ON BKZ REDUCTION

Reference	Lattice	Dim. (n)	Block (k)	Hermite Factor	Runtime [s]
Gama-Nguyen [8]	Random	120	28	$\sim 1.0109^n$	$\sim$ 32500
Schneider-Buchmann [10]	Random	300	15	$\sim 1.0137^n$	$\sim$ 5700
Buchmann et al. [11]	SIS	700	20	$\sim 1.0062^n$	$\sim$ 5500
Plantard-Schneider [14]	Ideal	100	30	$\sim 1.011^n$	$\sim$ 80000

There are two ways of entering into the hall of fame of this challenge. In the SVP version one has to find a vector of norm $\|\mathbf{b}_1\| \leq 1.05 \cdot GH(L)$. In more relaxed version one has to solve α-SVP, that is find a longer vector of norm $\|\mathbf{b}_1\| \leq n \cdot \mathrm{vol}(L)^{1/n}$.

For SVP, they experiment with the extreme pruning implementation of [15] with BKZ-35 preprocessing. For a dimension 80, the best result they have achieved is the root-Hermite factor approximately 1.0107 with the runtime around 300 seconds on the average. This also satisfies the bound for SVP in the challenge.

For α-SVP instances with dimensions smaller than 300, Plantard and Schneider experiment with the BKZ algorithm implemented in NTL. The experiment demonstrates that a root-Hermite factor approximately of 1.011 can be achieved with a block size 30 on dimension 100. But the required α-SVP bound in this case is slightly less than 1.01. For a high dimension, the Hermite factor becomes far worse than the required bound, showing the impracticality of running the algorithm in higher blocksizes. This gives the limitation that BKZ reduction in higher dimensions is intractable for ideal lattices as well.

Finally, Table I gives a short summary of the above experiments for BKZ lattice reduction.

III. RECURSIVE-BKZ

In this section, we describe the current, one of the best variants of BKZ algorithm, the Recursive-BKZ (RBKZ in short) reduction algorithm. In [1], the algorithm for the RBKZ reduction is called BKZ2.0. We would like, however, to emphasize the role of preprocessing, as the preprocessing notion has not been fully explained in previous publications. In lattice reduction algorithm, preprocessing is required to give a relatively bad basis a good shape. Actually, the most classical form of preprocessing we always observe is with polynomial time LLL algorithm to work with a slow enumeration-based algorithm like BKZ. In BKZ reduction, LLL is used before the enumeration routine to produce a preprocessed basis. Then the enumeration will run more efficiently. The preprocessing in the RBKZ can be defined as follows:

Definition 1 (RBKZ preprocessing): Let M_i be the projected block basis of index i such that $1 \leq i \leq n-1$ where n is the lattice dimension, parameter s is a step size and k_{min} is the minimum blocksize for recursion then before running enumeration on each block lattice k, run BKZ-(k-s) on the block if $k > k_{min}$ (otherwise, do as in original BKZ, i.e. only LLL preprocessing of the local block lattice).

TABLE II
HIGHEST INDIVIDUAL RANKING IN DIFFERENT LATTICE-CHALLENGE SITES

Challenge	Lattice Type	Dim. (n)	Hermite Factor	Algorithm Used	Contestant
Latticechallenge	SIS	825	1.0049^n	BKZ 2.0	Chen, Nguyen
SVP challenge	Random	126	1.0085^n	Enum, BKZ	Chen, Nguyen
Ideal lattice challenge	Ideal	128	1.0083^n	–	Ishiguro et al.

TABLE III
SOME SVP CHALLENGE DIMENSIONS FOR WHICH WE SOLVE THE BOUND AND REACH THE CHALLENGE HALL OF FAME.

Dim. (n)	BKZ block	Seed	$\|\mathbf{b}_1\|$	Approx. of GH	Hermite Factor
87	35	1112	~ 2408	1.0115	1.0099^n
83	30	5123	~ 2384	1.0257	1.0103^n
77	30	190	~ 2271	1.0091	1.0104^n
75	30	1000	~ 2220	1.0030	1.0104^n
71	30	1000	~ 2195	1.0197	1.0109^n

We first give the compact RBKZ algorithm in the form of Algorithm 1 in this text. We used an optimized version of this Algorithm in the later experiments for Lattice Challenge and for the cryptanalysis of SWIFFT hash function.

Algorithm 1 The RBKZ(n, k) basis reduction algorithm

Input: a basis $B = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ of a lattice L with blocksize $k \in (2, n)$, minimum blocksize k_{min}, step size s, the GS triangular matrix μ and a vector $\mathbf{c}$ to hold all $\|\mathbf{b}_i^*\|^2_{(1 \le i \le n)}$
Output: a BKZ-k reduced basis $(\mathbf{b}_1, \ldots, \mathbf{b}_n)$ with updated μ and $\mathbf{c}$

if $k < k_{min}$ **then**
  return the input basis
else
  LLL(B, c, μ, δ)
  $z \leftarrow 0$, $i \leftarrow 0$
  compute $\tau(n, k)$
  while $z < \tau(n, k) \cdot (n - 1)$ **do**
    $i \leftarrow i + 1$, $j \leftarrow \min(i + k - 1, n)$
    if $(i = n)$ **then**
      $i \leftarrow 1$, $j \leftarrow n$
    end if
    compute the local projected block basis M_i of dim. $d \leftarrow \min(k, n - i + 1)$
    for $(\frac{1}{p_{succ}})$ times **do**
      apply randomization on M_i $\Rightarrow$ output is a randomized basis Z_i
      $n \leftarrow d$, $k \leftarrow (d - s)$
      call RBKZ(n, k) on Z_i $\Rightarrow$ output is a preprocessed basis P_i
      apply pruned enumeration on P_i
    end for
    compute $\mathbf{b}_{new}$ as a shortest vector found among all iterations of M_i
    $h = \min(j + 1, n)$
    $B_h = (\mathbf{b}_1, \ldots, \mathbf{b}_{i-1}, \mathbf{b}_{new}, \mathbf{b}_i, \ldots, \mathbf{b}_h)$
    LLL(B_h, c, μ, δ); $z \leftarrow z + 1$
  end while
end if

IV. LATTICE CHALLENGE

The goal of the lattice challenge sites is to assess and compare lattice reduction algorithms. The cited results give an indication of what modern lattice reduction can achieve in practice. The first lattice challenge has been introduced by the TU Darmstadt team. The type of lattice used in this challenge is based on the hard lattice instances reported in [11]. These hard lattices related to the SIS problem are generated by some parameters: lattice dimension m, reference dimension n and a modulus q. The target of all participants is to test their lattice algorithm by running them on these lattices of different higher dimensions. If one can find a short vector $\|\mathbf{b}_1\|$ such that it is less than the reference dimension n for some unique higher dimension or some dimension with shorter vector yet to be listed, then they will be able to be ranked in the challenge hall of fame. One can have the flexibility to generate a random instance from a given generator codes or pick from a list of challenge dimensions given in the site home.

The next challenge introduced by Schneider and Gama is called the SVP challenge. The challenge lattice type is called the random or regular lattice due to Goldstein and Meyer. Here the target is to find a nonzero vector of length shorter than the bound $1.05 \cdot GH(L)$, where $GH(L) = \frac{\Gamma(n/2+1)^{1/n}}{\sqrt{\pi}}$ is the Gaussian Heuristic estimate for the length of a shortest nonzero vector in L.

The most recent challenge namely the ideal lattice challenge has been introduced by Plantard and Schneider. It is for testing lattice algorithms' performance on ideal lattices. A summary of the each challenge site is given in Table II with the highest dimensional lattice and the corresponding Hermite Factor achieved by the contestants.

Table III shows some challenge dimensions that we are able to tackle in the SVP challenge using RBKZ. We can reach a root-Hermite factor of ~ 1.0099 at best using blocksize 35 for dimension 87. The fourth column gives the approximation factor of GH (should be < 1.05)) that we achieved for each particular dimension. For dimension 87, the required complexity is approximately 31.5 (in $\log_2$ number of nodes) with the number of BKZ tours equivalent to 1587 out of ~ 2361 tours. After this many tours (~ 1587), BKZ does not improve the Hermite factor required for the norm of short vector $\|\mathbf{b}_1\|$ any longer. That justifies the use of aborted-BKZ [17] in RBKZ reduction.

V. PSEUDO-COLLISION IN SWIFFT

In this section we run an experiment to improve the current attack time (reported in [16]) on the SWIFFT hash function. We use the RBKZ reduction utilizing the aborted-BKZ reduction [17] with so-called *sublattice attack*.

A. SWIFFT Hash Function

The SWIFFT hash function is defined by the following compression function:

$$\sum_{i=1}^{m} \mathbf{a}_i \cdot \mathbf{x}_i \in R$$

where $R = \mathbb{Z}_p[x]/(x^n + 1)$ is the ring of polynomials, $\mathbf{a}_i \in R$ $(\forall\, i)$ are m fixed multipliers and $\mathbf{x}_i \in R$ $(\forall\, i)$ with coefficients

TABLE IV
CHOICE OF PARAMETERS FOR FINDING *pseudo-collision* IN SWIFFT.

m	n	p	δ	d
16	64	45	1.0124	141
16	64	49	1.0121	144
16	64	53	1.0119	147
16	64	57	1.0117	150
16	64	61	1.0115	152

$\in \{0,1\}$ are the input of length mn. The two basic parameters are n and p, where n is a power of 2 and $p > 0$ is a prime modulus. The algebraic formulation of SWIFFT is equivalent to a lattice of dimension mn and the security of such function depends on the infeasibility of finding relatively short vectors in such a lattice. Finding a collision in this function means looking for a nonzero vector $\mathbf{x}_i \in R$ with coefficients in $\{-1,0,1\}$ such that $\sum_{i=1}^{m} \mathbf{a}_i \cdot \mathbf{x}_i = 0 \bmod p$. In other words, if we can find a vector by reducing the lattice of Euclidean length $\leq \sqrt{mn}$, then it is a *pseudo-collision*. We call it pseudo-collision because we consider a sublattice of dimension $m'n' \leq mn$ of the full dimensional lattice.

B. Better Attack Time

The experiment by Micciancio and Regev [13] suggests that a shortest vector of length $l = 2^{2\sqrt{n \log p \log \delta}}$ could be found in an optimal dimension $d = \sqrt{n \log p / \log \delta}$, where mn is the dimension of the lattice with modulus p and δ is the root-Hermite factor. Again, finding a sub-lattice collision[3] in the mn dimensional lattice is equivalent to finding a vector of length $l = \sqrt{mn}$. In [16], Buchmann and Lindner use the above idea to find pseudo-collision for the SWIFFT compression function. They introduced a parameter generator algorithm that generates minimal possible parameters for SWIFFT required to satisfy the pseudo-collision condition. The smallest parameter set (in Table 1 of [16]) was $m = 16$, $n = 64$, $p = 257$, compression ratio ≈ 2 with root-Hermite factor $\delta \simeq 1.0084$ in sub-dimension $d = 206$. As this amount of the Hermite factor is hard for current reduction algorithms, they come up with this less hard problem with smaller p and δ and successfully break 9 instances of this problem with the NTL implementation of BKZ, and RSR algorithm [18]. We use aborted-BKZ to break the harder 5 of these 9 instances. See the Table IV for the scale-down parameters of these 5 instances.

We search for the δ (root-HF) required from the table to break a particular instance by increasing the number of tours within a BKZ blocksize. To get average results, we use 10 random SWIFFT seeds/instances for each of these 5 instances. As soon as we found a required δ (after some tours) we stop and record time otherwise we move to the next blocksize. In Figure 1 we plot the average runtime required for this approach comparing the original BKZ and RSR algorithms. It shows that aborted-BKZ can save a considerable amount of time comparing to BKZ and RSR.

VI. CONCLUSIONS

We identified some open problems from the above study. First, we require to evaluate the performance of other type of lattices important to the cryptosystem such as SIS, NTRU, LWE etc along with random lattices. Another question is to verify if lattice reduction for ideal lattices is as hard as for random lattices. Plantard-Schneider's experiment in [14] roughly answer this question. They also point out the limitation of the theoretical analysis of the BKZ. Another problem is to make available recent improvements of the BKZ algorithm and its variants (such as RBKZ), going beyond the NTL implementation.

[3]The same idea also used in latticechallenge to allow reduced dimensional lattice

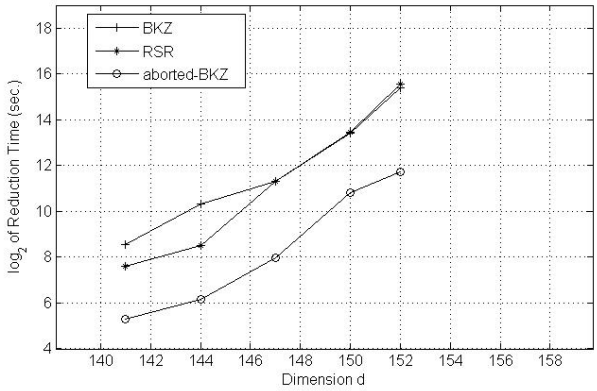


Fig. 1. Average *runtime* comparison for pseudo-collision

REFERENCES

[1] Y. Chen and P. Q. Nguyen. *BKZ 2.0: Better Lattice Security Estimates*. In ASIACRYPT'11, LNCS 7073, pp. 1-20, IACR, 2011.

[2] M. Ajtai. *The shortest vector problem in L_2 is NP-hard for randomized reductions (extended abstract)*. In STOC'98, ACM, pp. 284-293, 1998.

[3] O. Regev. *On lattices, learning with errors, random linear codes, and cryptography*. Journal of ACM, vol 56, 2009.

[4] C. Gentry. *Fully homomorphic encryption using ideal lattices*. In STOC'09, ACM, pp. 169-178, 2009.

[5] V. Lyubashevsky, D. Micciancio, C. Peikert and A. Rosen. *SWIFFT: A Modest Proposal for FFT Hashing*. In Fast Software Encryption(FSE) 2008, LNCS, pp. 54-72. Springer-Verlag, 2008.

[6] A. K. Lenstra, H. W. Lenstra, Jr., and L. Lovász. *Factoring polynomials with rational coefficients*. Math. Ann., 261(4): 515-534, 1982.

[7] C.-P. Schnorr and M. Euchner. *Lattice basis reduction: improved practical algorithms and solving subset sum problems*. Math. Programming, 66:181-199, 1994.

[8] N. Gama and P. Q. Nguyen. *Predicting lattice reduction*. In EUROCRYPT, 2008, LNCS 4965, pp. 31-51. Springer, 2008.

[9] V. Shoup. NTL: A library for doing number theory. Available at http://www.shoup.net/ntl/.

[10] M. Schneider and J. Buchmann. *Extended Lattice Reduction Experiments Using the BKZ Algorithm*. Sicherheit, pp. 241-252, 2010.

[11] J. Buchmann, R Lindner and M. Rückert. *Explicit Hard Instances of the Shortest Vector Problem*. In PQCrypto'08, LNCS, vol.5299, pp. 79-94, 2008.

[12] M. Ajtai. *Generating hard instances of lattice problems (extended abstract)*. In STOC'96, ACM, pp. 99-108, 1996

[13] D. Micciancio, O. Regev. *Lattice-based Cryptography*. In Post-Quantum Cryptography, D. J. Bernstein, J. Buchmann, E. Dahman (Eds), pp. 147-191, Springer, 2009.

[14] T. Plantard, M. Schneider. *Creating a Challenge for Ideal Lattices*. In IACR Cryptology ePrint Archive, vol. 2013 pp. 39, 2013.

[15] P.-C. Kuo, M. Schneider, O. Dagdelen, J. Reichelt, J. Buchmann, C.-M. Cheng and B.-Y. Yang. *Extreme enumeration on GPU and in clouds: how many dollars you need to break SVP challenges*. In CHES'11, Springer-Verlag, pp. 176-191, 2011.

[16] J. Buchmann and R. Lindner. *Secure Parameters for SWIFFT*. In INDOCRYPT'09, pp. 1-17, 2009.

[17] G. Hanrot, X. Pujol and D. Stehlé. *Analyzing Blockwise Lattice Algorithms using Dynamical Systems*. In CRYPTO'11. LNCS, vol. 6841, pp. 447-464. Springer, 2011.

[18] C. P. Schnorr. *Lattice reduction by random sampling and birthday methods*. In STACS'03, Springer, vol. 2607, pp. 145-156, 2003.

Application of Association Rule Mining for Replication in Scientific Data Grid

Md. S. Q. Zulkar Nine[1], Md. Abul Kalam Azad[2], Saad Abdullah[3], Mohammad Alaul Haque Monil[4], Ibna Zahan[5] Abdulla Bin Kader[6], Rashedur M Rahman[7]

Department of Electrical and Computer Engineering, North South University, Dhaka, Bangladesh
[1]engrbiruni@gmail.com, [2]ak.azad@live.com, [3]saad149@gmail.com, [4]monil01@gmail.com, [5]ibzprince@gmail.com, [6]babu_pixel@yahoo.com, [7]rashedur@northsouth.edu

***Abstract*—Grid computing is the most popular infrastructure in many emerging field of science and engineering where extensive data driven experiments are conducted by thousands of scientists all over the world. Efficient transfer and replication of these peta-byte scale data sets are the fundamental challenges in Scientific Grid. Data grid technology is developed to permit data sharing across many organizations in geographically disperse locations. Replication of data helps thousands of researchers all over the world to access those data sets more efficiently. Data replication is essential to ensure data reliability and availability across the grid. Replication ensures above mentioned criteria by creating more copies of same data sets across the grid. In this paper, we proposed a data mining based replication to accelerate the data access time. Our proposed algorithm mines the hidden rules of association among different files for replica optimization which proves highly efficient for different access patterns. The algorithm is simulated using data grid simulator, OptorSim, developed by European Data Grid project. Then our algorithm is compared with the existing approaches where it outperforms others.**

Keywords—Scientific Data Grid; Replica Optimization; Dynamic Replication; Association rule mining

I. INTRODUCTION

The term scientific data grid [1] refers to the massive network of computational infrastructure used by a wide range of researchers who use petabytes of data for experimentation all over the world. It is used for large scale data sharing and simulation based experimentation in wide range of scientific and engineering domains – such as *Large Hadron Collider (LHC)*[8], *Earth System Grid (ESG)*[6], *Laser Interferometer Gravitational Wave Observatory (LIGO)* [7] *etc*. All these grids use massively distributed heterogeneous computing platform where reliable and fast access of huge data sets is a challenging task. To ensure data access requirements same data sets need to be replicated in several regions of the grid. An efficient replication strategy is required to make data sets available so that it can be accessed efficiently by all the users and to reduce the number of replicas in an optimal level. There are many techniques those are described in literature which address the problem. But they have their own problems. Least Frequently Use (LFU) approach decides about replica by deleting the least frequent item extracted from the access history. However, another popular approach is Least Recently Used Optimizer, which makes decision based on temporal information. It deletes the least recent files to make room for the new files and assumes that new files are more important than the old ones. Many economy based approaches are described in literature. To address the problem we have used data mining based history mining [9] to capture the hidden structure of files. We structured the grid using Bandwidth Hierarchy Replication (BHR) standard to simulate our Grid. It ensures file locality in the regional sites. Our algorithm collects site access history to optimize the files regionally. This process ensures locality based optimization for most efficient use of local network and storage elements. We compared our algorithm with all other existing approaches. To make comparison we have used some statistical measures- such as Mean Job Execution Time (MJET), Total number of replication, number of local file access, number of remote file access, Average Storage Usage (ASU), Network Usages (NU). Our proposed model successfully reduces the mean job execution time. It also reduces the overall local network usage to only 20% and simultaneously maximizes the local file access.

This paper is organized as follows: Section II explains our proposed model with algorithms. Experimental evaluation is described in Section III. Results and discussion is explained in Section IV.

II. PROPOSED ALGORITHM

Grid computing contains a massive collection of computing and network resources dedicated to perform complex computing and simulation on large scale data sets. Whole Grid architecture is a collection of numerous sites responsible for computing and data storage. Computational and Data storage resources are called *Computing Element (CE) and Storage Element (SE)* respectively. The grid site that has Computing Element is called Computing Site. Computing sites are used to run jobs that use data sets stored in Storage Elements. When user issues a job to be performed in Grid, the request is first handled by resource Broker. Resource broker is responsible for scheduling the job into a computing site to ensure proper load balancing. It is also responsible for analyzing the user requirements to make appropriate decision. When a particular job is scheduled in a computing site, it tries to get the required data sets and contacts the Replica manager. Replica manager is responsible for administering the replication process. It manages the replica copy, deletion, replacements etc. Replication can be both static and dynamic. In Static replication the location of data is predefined, however, in dynamic replication the decision is made in real time to cope with the changing access pattern and data distribution. Replica Optimizer is responsible for optimizing the number of replicas

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

of a file. The main goal of an optimizer is to make minimum number of replicas in such a way that can provide maximal data access facility among the Grid sites.

The computing elements need to access numerous numbers of files with different access patterns. To ensure the data locality it is essential to make available all the files in regional repositories called Regional Header. Regional headers are chosen based on bandwidth and storage efficiency, so that it can work as a pseudo-master repository for that region.

Algorithm 1.1 Data Mining Based Replica Optimization (DMRO)

$P = Number\ of\ jobs,$
$N = Number\ of\ sites$
$S_i = Replica\ sites\ [i = 1, \dots, N]$

Input : *Grid Topology,*
Jobs, $J_K\ (K = 1, \dots\dots\dots, P)$,
Bandwidth details, $BW_{i,j}\ [i = 1, \dots, N; j = 1 \dots, N]$

Output: *Mean Job Execution Time,* $MJET_j\ [j = 1, \dots, N]$
Average Storage Used, $ASU_j\ [j = 1, \dots, N]$
Networ Usage, $NU_j\ [j = 1, \dots, N]$
Number of Replications, $NR_j\ [j = 1, \dots, N]$
Local File Access $LFA_j\ [j = 1, \dots, N]$
Remote File Access $RFA_j\ [j = 1, \dots, N]$

Method:

1. *Master Repository replicates data to regional Header.*
2. *Regional Header has all info about Regional Sites.*
3. *User execute a job,* J_K
4. *Job* J_K *scheduled to site* S_i *by the Job Scheduler*
5. *Replica optimizer decides to get best replica and remove the least important replica;*
6. *Execute the job* J_K
7. *Compute statistical measures -* $MJET_j, ASU_j, NU_j, NR_j, LFA_j, RFA_j$ *for all site* S_j

An overview of the proposed model is depicted in Algorithm 1.1. After producing the raw data from the experiments, these are stored in the master repository *(Step1)*. Then the master repository replicates all those files to Regional Header *(Step 2)*. When a user requests to execute a job, J_K (*k* is the number of jobs), Job scheduler reserves an appropriate CE to execute the job *(Step-3)*. Replica optimizer is used when data file is not available in local site and replication of a specific file is required. The goal of replica optimizer is to reduce the replication of file and concurrently ensures best access time for the file *(Step-5)*. When file becomes available in local site CE starts to execute the job *(Step-6)*. Then the algorithm computes some statistical measures shown in Step-7. The replica optimizer is detailed below:

A. Replica Optimizer

At first, Replica optimizer checks whether it is available in local storage *(Step-1)*. When the file exists in local storage, optimizer simply executes a *getFile* command to make it available for the job *(Step-2)*. Then the TTL (Time to Live) values of all other files are reduced by one and proceeded to perform the *job (Step-3-4)*. However, in case when file does not exist in local storage, Replica optimizer starts querying the local storage whether it has enough available space to replicate the requested file *(step-6-7)*. In case of storage availability a replica selection method is used to replicate the file from Regional header and other regional sites *(Step-8-9)*. The file replacement becomes mandatory in situations when file is not found in local storage which does not have enough space to replicate the file *(Step-10)*. Here we have used a data mining based approach to capture the hidden access pattern of files in local storage. We have used apriori association mining discover the high confidence rules contained in *Frequent File Set (FFS)* and (Step-12). Then files are ranked based on their confidence value (Step-13). At step-14, Non Frequent File set, *NFS* is computed by set difference between *Files in Local Set (FLS) and Frequent File Set (FFS)*. This NFS file set is sorted according to their TTL value (Step-15). Finally, replica selection procedure is executed to get the replica file. The concept of Apriori rule mining is described in Algorithm 1.2

Algorithm 1.2 Data Mining Based Replica Optimizer

$F_i\ where\ i = 1, \dots, m$
$FFS = Frequent\ File\ Set$
$FLS = Files\ in\ local\ Site$

1:If (F_i *is available in local site* S_j *) {*
2: getFile(F_i*);*
3: reduce TTL value of all other files
4: proceed job J_K *execution*
}
5: else {
6: query local site S_j *for storage availability12;*
7: if (S_j *storage available) {*
8: replicaSelection(BW_{ij}*,latency,sites,* S_i*)*
9: *getFile(*F_i*);*
}
10: else {
11. Accumulate (accessHistory, AH_j*,* S_j*);*
12: $FFS = Apriori\ (AH_j, minSup, minConf)$
13: Rank files based on confidence
14: $NFS = \{FLS - FFS\}$
15: Sort based on NFS, TTL *value*
16: Remove files from NFS *with lowest* TTL *value*
17: $replicaSeceltion(BW_{ij}, latency, sites, S_i)$
18: getFile(F_i*);*
}
}

B. Association Rule Mining

At first, **In Algorithm 1.1** minsup is provided as input to find out the k-frequent item set. At step-1, k is initialized as 1. Step-2, initial frequent item set is prepared by taking one item. From step-3 loop is started to find out all (k+1)-frequent item set, considering k-frequent item set as input. At step-5, *apriori-gen* function gives the candidate item set of k-frequent item set. At step-7, for each transaction, candidate items of that transaction are determined and at step-9, support count is measured. After going through all the transactions, at step-12, items having minsup as support count are identified. For all possible combination the algorithm is run and all frequent item set are stored together at F_s.

***Algorithm 1.3: Frequent-item set generation Apriori Algortithm* [9]**

Input: minsup
Output: F_s *Frequent-item-set*

1) $k=1$
2) $F_k = \{i \mid i \in I \wedge \sigma(\{i\}) \geq N * minsup\}$
3) $Repeat$
4) $k = k + 1$
5) $C_k = apriori - gen(F_{k=1})$
6) ***For*** *each transaction* $t \in T$ ***do***
7) $C_t = subset(C_k, t)$
8) ***For*** *each candidate itemset* $c \in C_t$ ***do***
9) $\sigma(c) = \sigma(c) + 1$
10) *End* ***for***
11) *End* ***for***
12) $F_k = \{c \mid c \in C_k \wedge \sigma(c) \geq N * minsup\}$
13) $until\ F_k = \phi$
14) $F_s = \cup\ \phi$

In Algorithm 1.2, frequent items set and minconf are given as input. At step-1, loop is started for each k-frequent item set where k is greater than 2. One item consequent is taken from the frequent item set and *ap-genrule()* is called. **Algorithm 1.3** describes the *ap-genrule()*function. As input, minconf and frequent item set is given and support count of all frequent item set are given. At step-1 and 2, size of frequent item set and rule consequent is determined. At step-4, *apriori-gen()* generates candidates sets from H_m. For each rule, at step-6, confidence is measured from. A rule is stored if it meets the minconf threshold and deleted if threshold is not met. These are done in step-7, 8, 9 and 10. Then *ap-genrule()* is called for next set of rule consequents at step-13.

***Algorithm 1.4: Rule generation of Apriori Algorithm* [9]**

---

1) ***For*** *each* $frequent\ k - itemset\ f_k, k \geq 2$ ***do***
2) $H_1 = \{i \mid i \in f_k\}$
3) $call\ ap - genrules(f_k, H_1)$
4) *End* ***for***

Algorithm 1.4: Procedure $ap - genrules(f_k, H_1)$. [9]

Input: minconf
Output: Rule set

1) $k=|f_k|$
2) $k=|H_m|$
3) ***if*** $k > m + 1$ *then*
4) H_{m+1}=*apriori-gen*(H_m)
5) ***for*** *each* $h_{m+1} \in H_{m+1}$ ***do***
6) *Conf* $= \sigma(f_k) / \sigma(f_k - h_{m+1})$
7) ***if*** $conf \geq minconf$ ***then***
8) ***output*** *the rule*$(f_k - h_{m+1}) \rightarrow h_{m+1}$
9) ***else***
10) ***delete*** h_{m+1} *from* H_{m+1}
11) *end* ***if***
12) *End* ***for***
13) *call ap-genrules*(f_k, H_{m+1})
14) *End* ***if***

III. EXPERIMENTAL EVALUATION

OptorSim [2] serves as a framework to simulate the grid environment. Inputs to OptorSim can be controlled by using four configuration files, namely the parameter configuration file, grid configuration file, job configuration file and bandwidth configuration file. We also have used the CMS topology shown in Fig.1.We have used CMS Data Challenge 2004 [4] to simulate our proposed model. It submits jobs according to a Gaussian distribution of jobs entered around 3 pm. Here total number of job submitted each day is under 700. The basic simulation parameters are set up by means of the parameter configuration file. The most important parameters includes the number of jobs, the choice of scheduling algorithms, the choice of replication algorithms, access patterns for the job, initial file distribution job delay and the job submission pattern by which the user submit jobs to the resource broker. The job configuration file contains information for the simulated jobs. It also specifies the job each site will accept, the probability each job runs and the files needed by each job. There are six job types, with no overlap between the set of files each job accessed.

TABLE I. OPTORSIM RESULTS

	No Replication	LRU	LFU	EMO Zipf-like	EMO Replicates	Proposed Model
Number of Jobs Remaining	0	0	0	0	0	0
MJET of all Jobs on Grid	105237	27682	32190	34910	16547	13320
Total number of Replications	0	1984	2341	931	735	754
Total Number of Local File Accesses	0	8901	9421	7812	8125	10536
Total Number of Remote File Accesses	9275	0	0	3490	4189	1053
ASU	15.37%	40.91%	40.91%	40.91%	40.91%	56.23%
Effective Network Usage	1.0	0.25	0.27	0.28	0.29	0.20

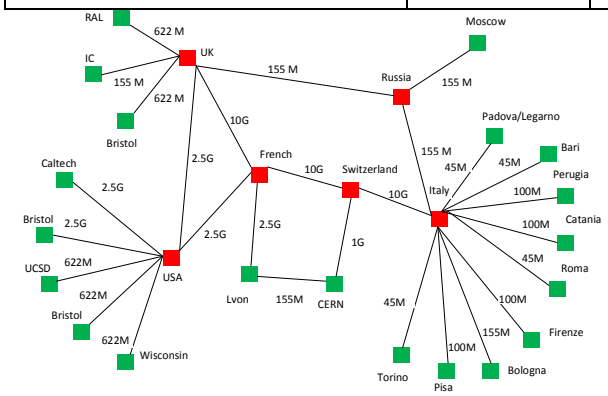


Fig. 1. CMS Data Grid Testbed

IV. RESULT AND DISCUSSION

After the simulation, we have compared different optimizers – such as LRU, LFU, EMO-Zipf-like, EMO Replicates with our novel algorithm. We have used several statistical measures, for example Mean Job Execution Time (MJET), Average Storage Used (ASU), Effective Network Usage (ENU) MJET is defined as the total time to execute all the jobs divided by the number of jobs completed. The total time includes the time that elapses from when a job enters the queue in a site to await execution until the time when the job finishes its processing and leaves the site. It is computed as shown in Eq. (1)

$$MJET = \sum_{i=1}^{n} \frac{Job\ Arrival\ time - Job\ departure\ time}{No\ of\ jobs\ completed} \quad (1)$$

ASU is the storage used specifies the amount of space used by files in MB. Monitoring the use of storage resources in the grid sites can also be a valuable source of information. Storage usage can be calculated for each site as a percentage of the capacity reserved by files out of the total capacity of the underlying storage. The average of all the storage elements in the grid can reflect the total system storage cost, which is computed as shown in Eq. (2)

$$ASU = \sum_{i=1}^{n} \frac{\frac{u}{c}(Site\ i)}{n} * 100\% \quad (2)$$

Replicating a file takes time and uses network bandwidth. This cost is effectively the ratio of files transferred to files requested, so a low value indicates that the optimization strategy used is better at placing files in the right places. It can be measured by using Eq. (3)

$$ENU = (Nremote\ file\ accesses + Nfile\ replication)/(Nremote\ file\ accesses + Nlocal\ file\ accesses) \quad (3)$$

Summary of all those optimizers are recorded in Table I. Our proposed model makes replication decision based on the local request pattern and the bandwidth between the source and destination of the file. It can be seen that the no replication policy takes the highest mean job execution time as it has to rely completely on the network bandwidth when it tries to access a file that is not available locally. However, local mining and hierarchical information sharing about popular data file proved efficient against other approaches. According to our approach, corresponding files of a popular file are replicated in a hierarchical order to the upper level storages and reduces the mean job execution time to 13320 s. It also reduces the network usage to only 20%. Total number of replication is also decreased due to the hierarchical order of file replication and it is approximately same as EMO Replicates Algorithm. Remote file access is highest for no replication as it has to request every file that is not locally available. LRU and LFU policy does not have remote file access option they just copy the file that is needed in the local site. However, our algorithm reduces the remote file access to only 1053 where EMO based approaches needs 3 or 4 times higher remote access.

REFERENCES

[1] B. Allcock, J. Bester, J. Bresnahan, A.L. Chervenak, I. Foster, C. Kesselman, S. Meder, V. Nefedova, D. Quesnal, S. Tuecke, Data management and transfer in high performance computational grid environments, Parallel Computing Journal 28 (3) (2002) 749–771.

[2] W. H. Bell, et al. "Optorsim: A grid simulator for studying dynamic data replication strategies." *International Journal of High Performance Computing Applications* 17.4 (2003): 403-416.

[3] R. Kavitha, A. Iamnitchi, I. Foster, Improving data availability through dynamic model driven replication in large peer-to- peer communities, in: Proceedings of Global and Peer-to-Peer Computing on Large Scale Distributed Systems Workshop, Berlin, Germany, May 2002.

[4] CMS Data Challenge, 2004: http://www.uscms.org/s&c/dc04/.

[5] The European DataGrid Project, http://www.edg.org.

[6] The Earth System Grid Project, http://www.earthsystemgrid.org/.

[7] Laser Interferometer Gravitational Wave Observatory; http://www.laserglow.com/

[8] large hardon collider, http://home.web.cern.ch/about/accelerators/large-hadron-collider.

[9] Tan, Pang Ning, Vipin Kumar, Introduction to data mining. Pearson Education, 2007.

Effectiveness of Minimization and Measurement of Schema Complexity Based on Weighted graph

G. M. M. Bashir, [1,*] A. Howlader, [2,] I. Shemu, [2] and E. Ety [2]

[1] Dept. of CCE, Patuakhali Science and Technology University
Patuakhali, Bangladesh

[2] Faculty of CSE, Patuakhali Science and Technology University
Patuakhali, Bangladesh

[*]murad98csekuet@yahoo.com

Abstract— **The complexity of data requirement increasing day by day; therefore minimization of schema becomes an important role for equi distribution or equalization of two or more problems. Schema complexity is measured by transmitting schema to a graph with weight or weighted graph. Selected special fundamentals from schema for weight the schema graph are described as major part. Minimization of schema ensures easy to measure the schema complexity. Some consideration for this approaches are normalization and de-normalization. We also find the effect of normalization and de-normalization on minimization of schema of real life data model as well as synthetic data model. The purpose of this research is to examine the effectiveness of schema complexity based on schema minimization with weighted graph.**

I. Introduction

The definitive goal of minimization of schema is to build a good relational schema for database system. There are a number of design approaches for relational and object relational databases [2]: ER [4], Extended ER [4], Object Oriented [3] and Spatial ER [3]. Relational database can be formed by the following design approach: ER → Relational schema → normalization →denormalization. At the core of modern information, object / relational database management systems (ORDMSs) improve new object storage capabilities to the relational systems [3]. A relational schema, at the logical level of database design and describes how real world entities are modeled in the database. Schema complexity is measured by transforming it to direct weighted graph and minimization of schema ensures easy to measure the schema complexity. Normalization is good for optimization method of relational schema and reduces large table into smaller tables but time consuming for accessing data. Whether denormalization improve query performance with reducing number of table and speed up data retrieval but slow data modification process. Graph theory is an established field where users represent data as a series of nodes and edges [10]. A node represents an object of interest and an edge indicates a relationship between two nodes. A graph is one of the vital data concepts in computer science. Now a days graph theory also use in Graph databases model. Graph database models are useful in fields where facts about data interconnectivity or topology is more significant as, or as important as the data itself [11]. Graph theoretical ideas are highly utilized by computer science applications. On the basis of our research measurement, we are using the weighted graph for our measurement purposes of schema in database field.

II. Trade off Normalization and Denormalization:

To minimize redundancy, database normalization is the procedure of establishing the fields and tables of a relational database. The originator of the model, Edgar, presented the perception of normalization; 1NF in 1970. Codd also define the 2NF and 3NF in 1971. Codd and Raymond F. Boyce defined the Boyce-Codd Normal Form (BCNF) in 1974. The theoretical system contains a total of 6 paradigms (That is Normalization Function, referred to as NF) [5]: 1NF, 2NF, 3NF, BCNF, 4NF and 5NF. These are the elementary approaches of schema optimization based on normalization theory, 5NF is the ultimate pattern. An example:

Table I: Unnormalized Form

SID	Name	Instructor ID	Subject	Grade	Grade point
1	Ram	101 102 103	C C++ DS	A+ A A+	10 9 10
2	Shy	101 104 105	C DM EM	B+ A A	8 9 9
3	Mahe	101 102 103	C C++ DS	A+ A+ A+	10 10 10

With 1NF, there's really no reason for duplicate columns and non-unique entries and remove repeating groups of data. 2NF is a good idea for all but the smallest and most casual of databases and removes partial dependency.

Table II: 3NF -3rd Normalized Form

SID	Name
1	Ram
2	Shy
3	Mahe

Grade	Grade point
A+	10
A	9
B+	8
B	7
C+	6
C	5
D	4
F	0

SID	Instructor Id	Subject	Grade
1	101	C	A+
1	102	C++	A
1	103	DS	A+
2	101	C	B+
2	104	DM	A
2	105	EM	A
3	101	C	A+
3	102	C++	A+
3	103	DS	A+

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

Normal 3 is necessary, at least in the extended term, for any business task or any other task where the data is acute and removes transitive dependency. While higher level normal forms are available, normalization up to BCNF is often found to be adequate for business data and removes multiple overlapping candidate keys.

Table III: BCNF –Boyce code Normalized Form

SID	Name
1	Ram
2	Shy
3	Mahe

Grade	Grade point
A+	10
A	9
B+	8
B	7
C+	6
C	5
D	4
F	0

SID	Instructor id	Grade
1	101	A+
1	102	A
1	103	A+
2	101	B+
2	104	A
2	105	A
3	101	A+
3	102	A+
3	103	A+

Instructor id	Subject
101	C
102	C++
103	DS
104	DM
105	EM

4NF isn't always necessary [9] but is required to remove non trivial multivalued dependency. 5NF is needed to remove join dependency. At least 2NF is for tiny organizations and at least 4NF is for any operation where the data's critical [9]. Most 3NF tables are permitted of insertion, update, and deletion anomalies. A relational DB table is regularly described as "normalized" as it is in the 3NF [8]. In 3NF database most of the business rule and the real world description can be implemented, where BCNF or 4NF is only used in some rare cases that are not mostly needed all intervals [2]. Again for database storage analysis, insert data in the three normal from and get a table and a bar chart. The space gain as of 1NF to 2NF and 2NF to 3NF is not same and linear. The increment of database size varies with the increment of record in the database.

Table IV: Database Size with Varied Number of Records

SL No.	No.of records	Table size in MB		
		1NF	2NF	3NF
1	10000	15	8	8
2	100000	120	57	56
3	200000	229	110	82
4	500000	493	272	264

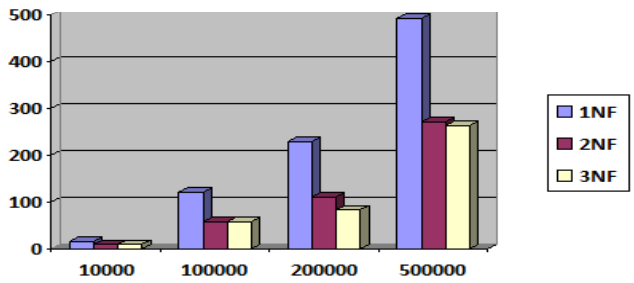
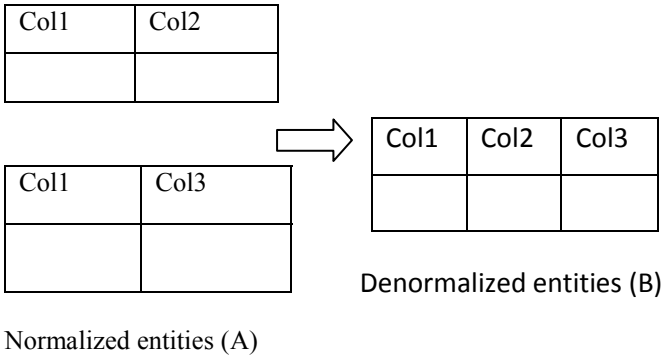


Fig. 1. Database Size vs Numbers of Record

A standard design management is that the designer should first construct an entirely normalized design; then selective denormalization can be achieved for performance reasons by terminating join. A denormalized data model is not the identical as a data model that has not been normalized [9]. Denormalization should only apply after a reasonable level of normalization has been established because Denormalizing can progress database access speed. Collapsing of One- to- One relationships are one of the most common and important techniques. This situation arises while each row of entity A, there is only one related row in entity B. To see Col 1, Col 2 and Col 3 at the similar time, collapse the two tables into one shown in figure.

Table V: Collapsing Table for Denormalization



We can apply this in BCNF table 3. Again the response time for point query is described in the following tables for fully normalized database and tuned database schema [2].

Table VI : Response Time for Point Query

Number of records	Query Type	Response time in second	
		Fully Normalized	Tuned Data base
44296	Point Query	3.57	1.95
221480	Point Query	14.67	1o.17
487256	Point Query	67.6	16.96

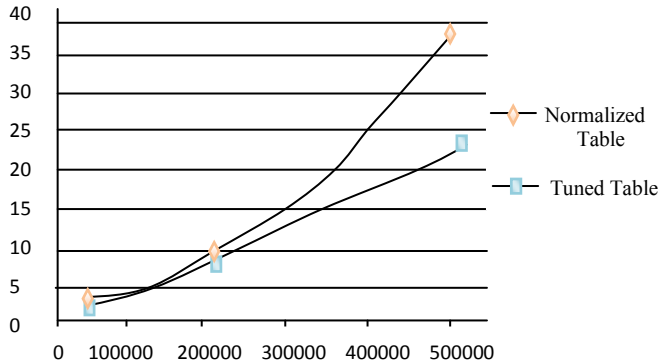


Fig. 2.Response time for point query

The response time gain by denormalization process is shown in above figure. Whether for large number records, exclusive joining processes make delay in the normalized database. The representation of data in the form of a graph lends itself well to structured data with a dynamic schema [12]. Relational database is losing its importance due to its dependence on a rigid schema which makes it difficult to add new relationships between the objects. The available data is growing manifolds; it is becoming

difficult to work with relational model as joining large number of tables is not working efficiently [13].

III. RESULT AND DISCUSSION:

Students only gather knowledge on weighted graph from mathematical fields. At present weighted graph are used in data structure section and many other section of engineering. On the basis of our research measurement, we can say that the weighted graph can also be used for schema in database field. Schema complexity is measured by transforming it to directed weighted graph and minimization of schema ensures easy to measure the schema complexity. Those minimized schema whose process discused before plays a vital role for transforming schema to weighted graph .

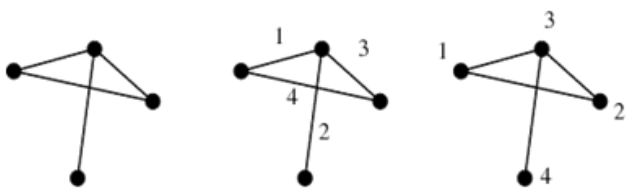
A. *Graph:*

Representations of a set of several pairs of objects are connected by links known as graph. A graph [6] is an object consisting of two sets called its vertex set and its edge set. The connected objects are denoted by mathematical notions called vertices, and the links that attach several pairs of vertices are called edges. A weighted graph is *a* graph in which each branch is given a numerical weight. Usually, the edge weights are nonnegative integers. Weighted graphs may be either directed or undirected. Such weights might represent, for example, costs, lengths or capacities, etc.

B. *Weighted graph:*

A weighted graph [7] is a graph in which a number w, called its weight, is assigned to each edge e or vertex v. A weighted graph is:

- A graph where vertices have some weights or values.
- A graph where edges have some weights or values.
- A graph where both edges and vertices have some weights or values.
- A graph where neither edges nor vertices have any weights or values.



Unlabeled graph edge-labeled graph vertex- labeled graph

Fig. 3. Labeled and unlabeled graph

C. *Graph for schema:*

Weights are usually real numbers or numerical values. Weighted graphs may be either directed or undirected.

For transforming schema to graph we choose directed weighted graph. We consider four fundamentals to weight the directed weighted graph of schema.

1. PK (primary key)
2. Fields
3. Tables
4. FK (foreign key)

We consider following schema:

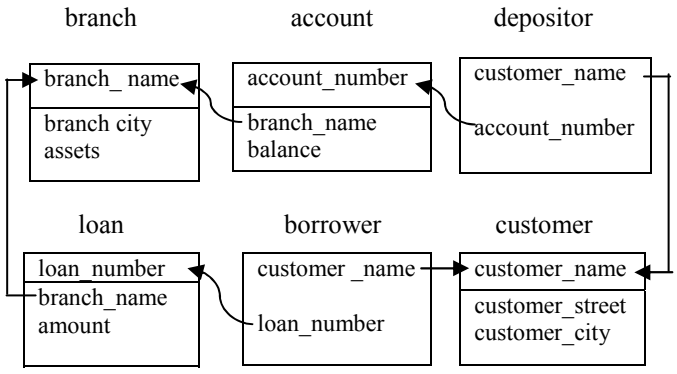


Fig. 4. A schema diagram

We represent the above schema through directed weighted graph. In our analysis representation of graph where vertices have some weights or values. Consider primary key as value for weight, the weight value is always one or zero because in every table there is only one primary key, value one or only foreign key, value zero. Consider field as value for weight, where value always more than zero because of containing enormous fields. Due to fragile performance multiply with .1 to form less weight. Tables are joined through foreign keys. Because of rare necessity, omit the consideration of table as weight value. Consideration of foreign key signifies the join between tables and present reasonable weight. A primary key of a table joined with one or more tables as a FK. So weight is based on this joined of FK. For calculating vertex weight use the concept of probability distribution. After combining above weight of all fundamentals the directed weighted graph is given below:

Vertex weight = (.7)*PK + (.2)*FK + (.1)* Fields

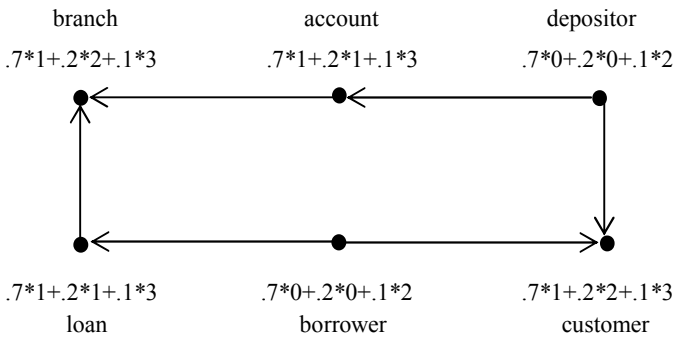


Fig. 5. Directed Weighted Graph with Value for Vertices

For schema representation vertices are denoted tables name and edges are denoted the connection between tables. The following equations are to determine total weight and the weight of each vertex of the weighted graph:

Total weight:

$$V = \sum_{W=1}^{n} v_w$$

Weight of each vertex:

V_w= (.7) * fields no. of PK + (.2)* total fields involved as FK + (.1)*no. of fields of particular table

Now the weight value for total weight and for each vertex is given by the following table:

Table VII: Weight value for vertices in table

Problem No.	Vertices Name	Fundamentals			Total weight on each vertex
		PK	FK	Field	
1	branch	.7*1	.2*2	3*(. 1)	1.4
	account	.7*1	.2*1	3*(. 1)	1.2
	depositor	.7*0	.2*0	2*(. 1)	0.2
	customer	.7*1	.2*2	3*(. 1)	1.4
	borrower	.7*0	.2*0	2*(. 1)	0.2
	loan	.7*1	.2*1	3*(. 1)	1.2
				Total weight = 5.6	

Through this paper we assure that the graph can be also used in database system efficiently. But we represent total schema to a weighted graph from which we are able to understand the number of FK in a table, connections between tables, number of fields in a table. We can't measure the complexity of schema through individual fundamental. For this reason we combine all fundamentals and find a perfect weight value. For schema minimization, is it necessary to add other fundamentals or not, is not verified. Here this is our method to measure the schema complexity on the basis of weight value. Based on this weight, our process will measure the complexity of schema manually which is time consuming.

IV. CONCLUSION

Based on the analysis presented in this paper, this study introduces the measurement of complexity and improves performance of database schema. To minimize the schema use normalization which produces minimal redundancy with larger number of smaller tables, minimized update anomalies and de-normalization which reduces join between tables, speed up data retrieval. This paper also provides concepts when not to normalize or should stop normalization in the database. Using those minimized schema, we are capable of transforming schema to directed weighted graph. Selected special fundamentals from schema are described in our research as major part for measuring complexity of schema. There may be some other considerations for complexity measurement which are not fully defined or understood. Complexity measurement of schema is partially implemented in our analysis though working on it is still going on. The significant properties of this analysis are that it is easy to understand, simple to measure and less time consuming. It fulfills most of the properties of a decent measurement of complexity. We are pretty sure that upcoming progress on this section will produce a great complexity measurement.

REFERENCES

[1] Codd, E.: "A Relational Model of Data for Large Data Banks". Communication of the ACM, Volume 13, No.6, Pages 377-387, 1970.

[2] Md. Abdul Lahel Shafey: "Performance Evaluation of Database Design Approaches for Object Relational Data Management", IICT, 2008.

[3] Lisbeth Bergholt, Jacob Steen Due, Thomas Hohn, other: "Database management systems: Relational, Object-Relational and Object-Oriented Data Models", 1998.

[4] Chen, Peter Pin: "The Entity-Relationship model –toward a unified view of data", Association for Computing Machinery (ACM), 1976.

[5] Dong Yu-Jie, Li Fu-Guo: "Research on Optimization Strategy of Relational Schema based on Normalization Theory", ISCSCT, August 2010.

[6] Trudeau, Richard J: "Introduction to Graph Theory",New York: Dover Pub. p. 19, 8 August 2012.

[7] Fletcher, Peter; Hoyle, Hughes; Patty, C. Wayne: "Foundations of DiscreteMathematics (International student ed. Ed.)." Boston: PWS-KENT Pub. Co. p. 463.ISBN 0-53492-373-9.

[8] C.J. Date.: "An Introduction to Database Systems",p. 290, Addison-Wesley (1999).

[9] Melissa Hollingsworth: "Data Normalization, Denormalization, and the Forces of Darkness".

[10] J. Pearl. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, (Morgan Kaufmann Publishers), 1988.

[11] Darshana Shimpi, Sangita Chaudhari: "An overview of Graph Databases", International Conference in Recent Trends in Information Technology and Computer Science (ICRTITCS - 2012) Proceedings published in International Journal of Computer Applications® (IJCA) (0975 – 8887.

[12] Mike Buerli: "The Current State of Graph Databases", Department of Computer Science, Cal Poly San Luis Obispombuerli@calpoly.edu, December 2012.

[13] Shalini Batra, Charu Tyagi: "Comparative Analysis of Relational And Graph Databases", (IJSCE), ISSN: 2231-2307, Volume-2, Issue-2, May 2012.

A Comprehensive Approach towards Guard Zone Computation Detecting and Excluding the Overlapped Regions

Ranjan Mehera, Piyali Datta, Arpan Chakraborty, and Rajat Kumar Pal

Department of Computer Science and Engineering, University of Calcutta
92, A. P. C. Road, Kolkata – 700 009, West Bengal, India
ranjan.mehera, piyalidatta150888, arpan250506, and *pal.rajatk@gmail.com*

***Abstract*— The guard zone computation problem finds immense applications in the field of VLSI physical design automation and design of embedded systems, where one of the major purposes is to find an optimized way to place a set of two-dimensional blocks on a chip floor. In VLSI layout design, the circuit components (or the functional units / modules or groups / blocks of different sub-circuits) that may be viewed as a set of polygonal regions on a two-dimensional plane, are not supposed to be placed much closer to each other in order to avoid electrical (parasitic) effects among them [12]. Each (group of) circuit component(s) C_i is associated with a parameter δ_i such that a minimum clearance zone of width δ_i is to be maintained around C_i. If the guard zonal regions overlap, we have to remove the overlapped regions in order to compute the resultant outer guard zone (sometimes inner guard zones are also an issue to be considered). The location of the guard zone (of specified width) for a simple polygon is a very important problem for resizing the (group of) circuit components. In this paper, we have developed an algorithm to compute the guard zone of a simple polygon as well as to exclude the overlapped regions among the guard zonal segments (if any) in O(*n* log *n*) time, where *n* is the number of vertices of the given simple polygon.**

Keywords: Simple polygon; Safety zone; Notch; Convex hull; False hull edge; Convolution; Minkowski sum.

I. Introduction

The guard zone computation problem occupies vast place of interest in the field of VLSI physical design automation and design of embedded systems, as resizing is an important problem in VLSI layout design as well as in embedded system design. This motivates us to compute a guard zone of a simple polygon. Given a simple polygon P, its guard zone G (of width r) is a closed region consisting of straight line segments and circular arcs (of radius r) bounding the polygon P such that there exists no pair of points p (on the boundary of P) and q (on the boundary of G) having their Euclidean distance $d(p,q)$ less than r. If more than one polygonal regions are closed enough, their safety zones may overlap, violating the minimum separation constraint among them as shown in Fig. 1.

In computing the guard zone of a simple polygon [7], we have at most O(n) straight line segments and O(n) circular arcs for a polygon of n vertices. So, to find out all the intersection points and exclude the overlapped region(s) accordingly (in order to get the desired guard zone only, including holes, if any), we may execute an O(n^2) algorithm for each pair of such segments, among all straight line segments and circular arcs. Indeed, this algorithm is greatly expensive.

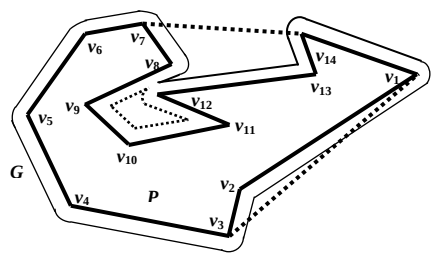


Fig. 1: A guard zone G of a simple polygon P.

In this paper, we have developed an algorithm to find and exclude that overlapped regions using coordinate as well as computational geometry, which reduces the time complexity of the algorithm to O(n log n).

II. Literature Survey

In the context of guard zone computation, several different algorithms have been proposed so far. The most discussed tool for guard zone computation is the Minkowski Sum [3]. Apart from Minkowski sum, convolution can also be used as a tool for guard zone computation. Essentially, Minkowski sum between a line (as polygonal segment) and a point (perpendicularly at a distance r apart) with the same x- and y-coordinates gives a line parallel to the given one. But the question arises is whether the parallel line is inside or outside the polygon. Here the definition of Minkowski sum [3] can be extended as below.

If A and B are subsets of R^n, and $\lambda \in R$, then $A+B = \{x+y \mid x \in A, y \in B\}$, $A–B = \{x–y \mid x \in A, y \in B\}$, and $\lambda A\ \{\lambda x \mid x \in A\}$. Note that $A+A$ does not equal $2A$, and $A–A$ does not equal 'zero' in any sense.

The convolution between a polygon and a circle of radius r may give us the desired solution of a guard zone. But the circles need to be drawn in every possible point of the polygon and thus the time complexity of the algorithm increases. The complexity of computing Minkowski sum of two arbitrary simple polygons P and Q is O(m^2n^2) [3], where m and n are the number of vertices of these two polygons, respectively. In particular, if one of the two polygons is convex, the

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

complexity of Minkowski sum reduces to O(mn). In [5], a number of results are proposed on the Minkowski sum problem when one of the polygons is monotone.

In this context, a linear time algorithm is devised for finding the boundary of the minimum area guard zone of an arbitrarily shaped simple polygon in [11]. This algorithm uses the idea of Chazelle's linear time triangulation algorithm and requires space complexity of O(n) as well [2], where n is the number of vertices of the polygon. After having the triangulation step, this algorithm uses only dynamic linear and binary tree data structures. Again, a time-optimal sequential algorithm for computing a boundary of guard zone has been developed in [7, 8, 9] that uses simple analytical and coordinate geometric concepts. The algorithm can easily be modified to compute the regions of width r outside the polygon as guard zone, and also inside the polygon.

III. Problem Formulation and the Algorithm

In this section, we formulate and develop a comprehensive algorithm to compute guard zone of a simple polygon while detecting and excluding the overlapped regions (if any) of the guard zone. The case becomes most simple, if the polygon is convex as there is no overlapping among the guard zonal regions. Thus, the guard zone for a convex polygon can be computed in linear time [7]. But the problem arises when the polygon has one or more notches in it.

A *notch* is a polygonal region outside the polygon that is starting and terminating between two consecutive convex vertices of the polygon that are not adjacent. That means a notch starts with a convex vertex, goes through one or more concave vertices, and terminates with a convex vertex, along the edges (and vertices) of the polygon (in one particular direction). But for some cases a simple polygon may contain a notch (or more) and guard zonal overlapping may also occur. Figs. 2(a) and 2(b) show two different cases of overlapping.

We can assume another case where a part of a polygon is shown Fig. 2(c), where B is a convex vertex but A is concave, and the guard zones of these two regions may overlap. Again the guard zones of two linear portions BD and BE of a simple polygon may also overlap with the linear portions of the guard zones AC and AF, respectively. All the above cases may occur for a simple polygonal guard zone. Now, as our objective is to compute the intersection points and exclude the overlapped regions (if any), we first compute the guard zone G (without excluding overlapped portions) of the given simple polygon P in linear time [7, 9].

Instead of checking all pairs of segments to detect the intersection points (if any), we apply the line sweep algorithm to reduce the time complexity to O(n log n) from O(n^2) [10]. But the line sweep algorithm can only be applied to a set of line segments as input, while a guard zone of a polygon is not only a set of line segments, but there are circular arcs as well. In case of guard zonal line segments we can compute the intersections of the guard zones (if any) by using the line sweep algorithm [10]. But in case of circular arc it cannot be applied directly. So, we have to do some pre-processing to make the algorithm run in time O(n log n).

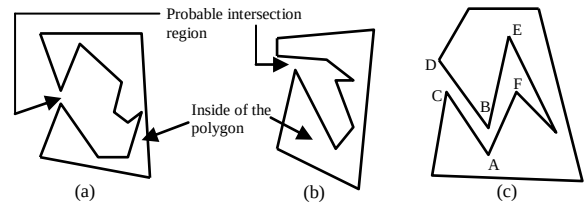


Fig. 2: Different kinds of probable intersections between overlapping of guard zonal regions. (a) Probable overlapping of guard zone of two convex regions. (b) Probable overlapping of guard zone of one convex region with a straight line segment. (c) A polygon consisting of a number of convex (e.g., B, C, D, F) and concave (e.g., A and E) vertices whose guard zonal regions are probable to overlap.

We know that a circular arc can be assumed to be the aggregation of a number of infinitesimally small ordered line segments. Thus if we can subdivide the arc into line segments, we can then apply the line sweep algorithm for detecting the intersection points between these derived line segments (instead of the circular arcs) with others. Though to break an arc into line segments, we have to take some precision, but that does not affect the output of the algorithm as it definitely reports all the intersections existing among the guard zonal segments [10].

Let us consider a simple polygon P and its related guard zone G [7, 8, 9]. To exclude the intersecting regions we have to find the intersections among the guard zonal segments. This application of the line sweep algorithm replaces each circular arc with its corresponding set of line segments and hereby converting the guard zone of P into a collection of line segments only. To convert the computed guard zone of P into a set of line segments, we only focus on each convex vertex of the given polygon P, as only here the consequent guard zone is a circular arc of radius (say) r. In Fig. 3, v is a convex polygonal vertex, also uv and vw are the adjacent edges of v, whose guard zonal segments are two straight line segments $u'v'$ and $v''w'$, where $u'v' \parallel uv$ and also $v''w' \parallel vw$, and the perpendicular distance between the parallel lines for both the pairs is same as r.

In order to obtain the smallest possible line segments, we follow a binary recursive procedure. First of all, we bisect $\angle v'vv''$ (or $\angle uvw$) by a bisector bs_1 whereon $vp_1 = r$. We join $v'p_1$ and p_1v''; so, $v'p_1v''$ is an approximated guard zone (for $p = 1$) of the circular arc. Next to make this approximation finer, we further bisect $\angle v'vp_1$ by bs_2, such that $vp_2 = r$ and bisect $\angle p_1vv''$ by bs_3 whereon $vp_3 = r$. Then we join $v'p_2$, p_2p_1, p_1p_3, and p_3v''; so, $v'p_2p_1p_3v''$ is an even finer approximation of the guard zone (for $p = 2$) than the previous one (i.e., $v'p_1v''$) of the

circular arc we usually draw. Now it is easy to realize that for $p = 3$, we get even smaller line segments $v'p_4 = p_4p_2 = p_2p_5 = p_5p_1 = p_1p_6 = p_6p_3 = p_3p_7 = p_7v''$, which is closer to the circular arc we usually draw as a part of guard zone for a convex polygonal vertex.

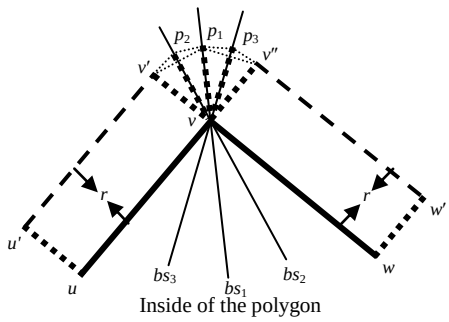


Fig. 3: Recursive division of a convex polygonal angle formed at vertex v wherefrom equal length smallest possible chords are computed that in a group replaces the circular arc that is computed as a part of guard zone outside the polygon up to a desired level of precision of an angle that is formed at the convex vertex for each such smallest possible chord.

This process of bisection is continued till the value of each bisectional angle becomes 0.50° or 0.25°, or up to some precision of angle that makes the straight line segments as chords of the circular arc reasonably very small. Needless to mention that all the points p_i over the bisections outside the polygon are the points on the circular arc as part of G, and each small line segment (whose length tends to zero for a smaller value of r) is an approximation of its associated arc for which it is the largest chord. So, further bisection of each of the bisected angles in subsequent levels of recursion and spotting a point p_i on each of the bisections at distance r from v outside the polygon helps to achieve more points on the said circular arc that are consecutively equidistant and closer to each other. Hence, up to some desired level of precision, we may obtain a set of equal length straight line segments collectively that replaces the circular arc we liked to draw as a part of G for vertex v.

Now, the line sweep algorithm is applied where the input is a set of line segments associated with their starting and ending coordinates as event points. The sweep line is traversed through the sorted list of event points depending on x- or y-coordinates. Here starting point, ending point, and point of intersection (if any) are the three types of event points at which insertion, deletion, and update of neighbouring operations are performed, respectively, during the sweep line traversal.

Here the data structures used are of the highest importance to accomplish it in $O(n \log n)$ time. This algorithm maintains two binary search trees, one for storing the event points and another for the line segments to keep track of their neighbouring information, which is known as the query tree. All the three operations are performed on the query tree. As an intersection occurs only between the neighbours, this data structure reduces the search space for intersection checking to the set of neighbours of a line segment. In addition, it is proved that line sweep algorithm reports all the intersection points [10]. After getting all these intersection points, we traverse the original guard zone and depending on the intersection points we exclude the intersected or overlapped region(s), and report the outer and inner guard zone [6], if they occur.

IV. Experimental Results

In this section, we cite a relative study based on the polygon shown in Fig. 1, in terms of time and detecting intersections, between the naive algorithm where the intersections are identified by checking each pair of segments and our proposed method on line sweep algorithm after repetitive bisecting the circular arcs.

Table 1: A study of comparison of detecting intersection(s).

Variables to be considered	Naïve algorithm	Repetitive bisection and line sweep algorithm
Number of edges of the polygon	14	14
Number of convex vertices of the polygon	9	9
Segments to be considered for checking intersections	14 (line segments) + 9 (circular arcs)	14 (original guard zonal line segments) + $9 \times p$ (derived line segments from the circular arcs), where p is a constant
Computation required	$(14+9)^2 = 529$	$(23+I)(\log_2 23) \approx 104.042 + 4.524 \times I$, where I is the number of intersections
Time complexity	$O(n^2)$	$O(n \log n + I \log n)$

V. Computational Complexity

If the number of edges in the original polygon is n, then the number of edges after the preprocessing of the polygon by replacing all the circular arcs by a set of (smaller) line segments is also $O(n)$ as we require a constant number of operations for each of the convex vertices. The next step is to apply the line sweep algorithm taking the set of all line segments, original as well as derived segments by preprocessing. The algorithm starts by creating the event queue by sorting the starting and ending points of the line segments, which takes $O(n \log n)$ time using any standard sorting algorithm. Initializing the status structure takes constant time. The handling of status structure consists of three operations, insertion, deletion, and interchange of positions that takes $O(\log n)$ time each. Now $m = n+I$, where I is the number of intersection points. Hence, the complexity of line sweep algorithm is $O(m \log_2 n)$ [10].

Lemma 1: If $p \geq 1$ be the number of times a circular arc is divided to replace the arc by a set of equal (smaller) line segments, then p must be a constant.

Proof: To prove the lemma we consider two extreme cases, i.e., the circular arcs for the convex vertices with two extreme values of the angles. If we can prove that p is a constant for these two cases, it is true for all the intermediate values. As circular arcs are introduced in the guard zones for convex vertices and the convex angle to be subdivided is minimum when it tends to 180° and maximum when it tends to 360°. Figs. 4(a) and

4(b) show these two cases. If the internal angle tends to 180º, the external angle also tends to 180º. Hence, we draw perpendiculars to the two adjacent line segments at the convex vertex O of the given polygon, where $OA = OB = r$ (see Fig. 4(a)). So, $\angle AOB$ is very close to 0º, and we get the line segment corresponding to the circular arc by connecting points A and B that may not be bisected further.

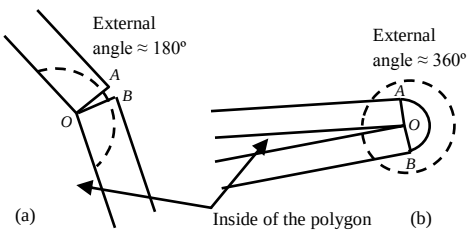


Fig. 4: (a) The external angle tends to 180º, when the angle associated with the circular arc to be subdivided tends to 0º. (b) The external angle tends to 360º, when the angle associated with the circular arc to be subdivided tends to 180º.

On the other hand, if the external angle tends to 360º (see Fig. 4(b)), the angle $\angle AOB$ to be subdivided tends to 180º. When an arc makes an angle less than or equal to 4º, we may consider it to be a line segment. So, if we apply a recursive algorithm to subdivide $\angle AOB$ up to a level of height six (or any constant number more than six), we may get a collection of straight line segments that may replace the circular arc up to a level of desired precision; hence the value of p is proved as constant. □

VI. Applications of the Problem

Now in brief we like to point out the importance and motivation of the problem as follows. In case of VLSI physical automation design, given a set of isothetic non-overlapping polygonal regions and a common resizing parameter δ, the objective is to compute another set of closed regions resizing each polygon by an amount of δ. If two or more polygons are closed enough so that their guard zones overlap, indicating the violation of least separation constraint among them, then the polygons have to move relatively to overcome this breach [12].

The guard zone problem finds another important application in the automatic monitoring of metal cutting tools. Here a metal sheet is given and the problem is to cut a polygonal region of some specified shape from that sheet. The cutter is like a ballpoint pen whose tip is a small ball of radius δ, and it is monitored by a software program. If the holes inside the notch also need to be cut, our algorithm can easily be tailored to satisfy that requirement too.

The Minkowski sum is an essential tool for computing the free configuration space of translating a polygonal robot [1]. It also finds application in the polygon containment problem and in computing the buffer zone in geographical information systems [4], to name only a few.

VII. Conclusion

As we have discussed earlier, resizing of electrical circuits is an important problem in VLSI layout design as well as in embedded system design, while accommodating the (groups of) circuit components on a chip floor, this problem motivates us to compute a guard zone of a simple polygon. In this paper we have considered the problem of computing a guard zone of a (2D) simple polygon, and developed a sequential algorithm for computing the same that uses the concepts of analytical and coordinate geometry to detect overlapped region(s) within the guard zone (if any) and accordingly exclude that region to report the resulting outer guard zone. Our algorithm can easily be modified to compute the regions of width r (as guard zonal distance) outside the polygon, and also inside the polygon (if necessary), which may find several applications in practice. This work can also be extended for computing a guard zone of a three-dimensional simple solid object as a problem of probable future work.

References

[1] Bajaj C. and M.-S. Kim, Generation of Configuration Space Obstacles: The Case of a Moving Algebraic Curves, *Algorithmica*, vol. 4, no. 2, pp. 157-172, 1989.

[2] Chazelle B., Triangulating a Simple Polygon in Linear Time, *Discrete Computational Geometry*, vol. 6, pp. 485-524, 1991.

[3] Harnandez-Barrera A., Computing the Minkowski Sum of Monotone Polygons, *IEICE Trans. on Information Systems*, vol. E80-D, no. 2, pp. 218-222, 1996.

[4] Heywood I., S. Cornelius, and S. Carver, *An Introduction to Geographical Information Systems*, Addison Wesley Longman, New York, 1998.

[5] Lee I.-K., M.-S. Kimand, and G. Elber, Polynomial / Rational Approximation of Minkowski Sum Boundary Curves (Article No.: IP970464), *Graphical Models and Image Processing*, vol. 60, no. 2, pp. 136-165, 1998.

[6] Mehera R., A. Chakrabarty, P. Datta, and R. K. Pal, A 2D Guard Zone Computation Algorithm for Reassignment of Subcircuits to Minimize the Overall Chip Area, *Proc. of the International Doctoral Symposium on Applied Computation and Security Systems (ACSS-2014)*, Apr. 18-20, 2014.

[7] Mehera R., S. Chatterjee, and R. K. Pal, A Time-Optimal Algorithm for Guard Zone Problem, *Proc. of 22nd IEEE Region 10 International Conference on Intelligent Information Communication Technologies for Better Human Life (IEEE TENCON 2007)*, CD: Session: ThCP-P.2 (Computing) (Four pages), Taipei, Taiwan, 2007.

[8] Mehera R., S. Chatterjee, and R. K. Pal**,** Yet another Linear Time Algorithm for Guard Zone Problem. *The Icfai Journal of Computer Sciences*, vol. II, no. 3, pp. 14-23, Jul. 2008.

[9] Mehera R. and R. K. Pal, A Cost-Optimal Algorithm for Guard Zone Problem, *Proc. of 10th International Conference on Distributed Computing and Networking (ICDCN 2009)*, pp. 91-98, Hyderabad, India, 2009.

[10] Goodrich M. T., Intersecting Line Segments in Parallel with an Output-Sensitive Number of Processors. *Society of Industrial and Applied Mathematics*, vol. 20, no. 4, pp. 737-755, Aug. 1991.

[11] Nandy S. C., B. B. Bhattacharya, and A. Hernandez-Barrera, Safety Zone Problem, *Journal of Algorithms*, vol. 37, pp. 538-569, 2000.

[12] Sherwani N. A., *Algorithms for VLSI Physical Design Automation*. Kluwer Academic, Boston, 1993.

GreMuTRRR: A Novel Genetic Algorithm to Solve Distance Geometry Problem for Protein Structures

Md. Lisul Islam[1], Swakkhar Shatabda[1] and M Sohel Rahman[2]
Department of Computer Science and Engineering, United International University
House 80, Road 8A, Dhanmondi, Dhaka-1209, Bangladesh
[2]Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology
Palashi, Dhaka-1000, Bangladesh
Email: {lisul,swakkhar}@cse.uiu.ac.bd, msrahman@cse.buet.ac.bd

***Abstract*—Nuclear Magnetic Resonance (NMR) Spectroscopy is a widely used technique to predict the native structure of proteins. However, NMR machines are only able to report approximate and partial distances between pair of atoms. To build the protein structure one has to solve the Euclidean distance geometry problem given the incomplete interval distance data produced by NMR machines. In this paper, we propose a new genetic algorithm for solving the Euclidean distance geometry problem for protein structure prediction given sparse NMR data. Our genetic algorithm uses a greedy mutation operator to intensify the search, a twin removal technique for diversification in the population and a random restart method to recover stagnation. On a standard set of benchmark dataset, our algorithm significantly outperforms standard genetic algorithms.**

***Index Terms*—Genetic Algorithms, Protein Structures, NMR, Euclidean Distance Geometry, Greedy Mutation**

I. Introduction

The function of a protein depends on its *native conformation* or the stable three dimensional structure with minimum free energy in a particular environment. Knowledge about this native structure is of paramount importance and can have an enormous impact on the field of rational drug discovery. Nuclear magnetic resonance (NMR) spectroscopy is one of the most widely used techniques to predict native structure of proteins. NMR machines provide inter-atomic distance data for a given protein. Thus, the molecular distance geometry problem (MDGP) arises in this context: given the set of Euclidean distances between the atoms in a protein, MDGP asks to find the Cartesian coordinates of the atoms.

However, in practice, NMR machines are able to produce the inter-atomic distances of only a subset of the pairs of atoms that are spatially close and this data too lacks accuracy. As a result, we are given the upper and lower bounds of only a subset of the Euclidean distances. Thus we are left to solve a variant of the MDGP problem with incomplete and inaccurate data. Many computational approaches have been applied to solve MDGP problem with sparse and inaccurate data on real instances [1], [2], [3], [4], [5], [6]. However, complete search methods like spatial branch and bound (sBB) and stochastic methods like variable neighborhood search (VNS) have been able to solve the problem for proteins with only upto 50 atoms [7] and fail to converge quickly for larger number of atoms.

In this paper, we present GreMuTRRR (pronounced grey matter), which is a scalable genetic algorithm to solve the MDGP problem for sparse and inaccurate NMR data. Much of the success of our approach comes from a greedy mutation operator, sude to intensify the search, a twin removal technique to diversify the population and a random restart method to recover from stagnation. Experimental results on proteins with number of atoms ranging from 50–2147 shows that our algorithm outperforms standard genetic algorithms and thus obtains state-of-the-art results.

II. Preliminaries

In this section, we provide a formal definition of the MDGP problem and a brief description of genetic algorithms.

A. Distance Geometry Problem

Molecular distance geometry problem can formally be defined as followed [8]: find a set of Cartesian coordinates $c_1, c_2, \cdots, c_n \in \mathbb{R}^3$ of atoms of a molecule such that $l_{ij} \leq d_{ij} \leq u_{ij}, \forall (i,j) \in E$, where l_{ij} and u_{ij} are the lower and upper bounds of the Euclidean distance $d_{ij} \equiv ||c_i - c_j||$ between a pair of atoms $(i,j) \in E$. Notably, in the context of NMR data, E is sparse. Here, Cartesian coordinate c_i of an atom i corresponds to a three dimensional point (x_i, y_i, z_i) in the Euclidean space.

B. Genetic Algorithm

Genetic Algorithms are population-based search methods that resemble the natural phenomena of biological evolution. Genetic Algorithms are widely used for different search optimization problems in different fraternity. It basically starts with a pool of initial random solutions which is called the initial population. Each individual in the population are encoded by a set of properties which are called chromosome or genotype which can be altered for attaining diversification in the population. It then follows an iterative process in which each of these iteration is called a generation. In each generation, the population are then allowed to evolve using different operators which also mimics the natural process of biological evolution like mutation, recombination or survival of the fittest. In each generation, the fitness of each individual is evaluated. Generally the fitness is the value of the optimization function being considered to be solved. The more fit individuals are usually selected to breed and generate new fitter individuals in the population. Thus a new generation of population is 'breeded' and are used in the next generation. This process of evolution are continues until a sufficient number of generations has been produced or a satisfactory level of fitness value has been attained.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

III. Related Works

Euclidean distance geometry problem and its variants are applied to many problems in various fields including wireless sensor network localization [9], inverse kinematic problem [10], multi dimensional scaling [11], protein structure determination [6] etc. The variant of the MDGP problem, when we know distances for all pairs $(i, j) \in E = \{1, 2, \cdots\}^2$ and $d_{ij} = l_{ij} = u_{ij}$ has a polynomial algorithm to produce exact solution [8]. Even when some of the pairwise distances are unknown, the problem is solvable by a linear time algorithm [3]. However, the variant of MDGP with sparse and inaccurate data is shown to be NP-hard by More and Wu [12]. A recent survey of computational methods applied to solve this variant of MDGP can be found in [2].

Among the general purpose methods, spatial branch and bound [13] and variable neighborhood search (VNS) [14] methods are not scalable [7]. Smoothing based methods like DGSOL [6], [12] also fail for large instances of the problem. In [15], VNS was combined with DGSOL approaches which provided better results for larger instances but resulted into a slow algorithm. A combinatorial build up algorithm was proposed in [16]. It is important to note that all these methods were tested on dense instances only. Among other methods applied to this problem graph decomposition methods [1] and NLP formulations [17] are notable.

IV. Our Method

GreMuTRRR is formally presented in Algorithm 1. It starts with a population initialized randomly and terminates at convergence. Convergence is achieved if the search does not improve the quality of the global solutions for a given period of time. In each generation, individuals are selected from the population for cross-over to produce new individuals. Mutation operations are performed probabilistically on the newly found individuals. Global best solutions are updated in each iteration and are kept in the new populations to maintain elitism. Periodical twin removal and random restart operations are activated to ensure diversification and recover stagnation. Rest of this section describes the various components of GreMuTRRR.

A. Search Space

A Protein structure is formed by the interaction among different amino acids tied with each other by chemical bonds. A pair of these amino acid cannot reside closer than 3.8Å to each other to avoid steric clash. If all the amino acids in a certain protein structure align in a straight line, we have a chain of amino acids of length $3.8 \times V$ in each directions, where V is the number of amino acids in the protein structure. Hence, we have taken the upper bound of our 3-D search space to $3.8 \times V$.

B. Encoding

In our method, we have encoded every individual in a population by $3 \times V$ number of genes. Here, V is the number of atoms in a protein structure. Each of these genes in an individual is initialized randomly from an uniform distribution of the range $[0, 3.8 \times V]$. So, an individual X can be represented as the ordered list of $3V$ number of genes:

$$X = \left\{ \underbrace{x_1, y_1, z_1}_{1^{st} Atom}, \underbrace{x_2, y_2, z_2}_{2^{nd} Atom}, \ldots, \underbrace{x_i, y_i, z_i}_{i^{th} Atom}, \ldots, \underbrace{x_V, y_V, z_V}_{V^{th} Atom} \right\}$$

Algorithm 1: GreMuTRRR()

```
nonDiverseSteps = 0
nonImprovingSteps = 0
GreedyMutationRate = 0.8
Intialize the population, P randomly
while Termination criterion are met do
    P_new = {globalBest}
    for each Individual X ∈ P do
        ⟨X_1, X_2⟩ = tournamentSelection(P)
        X_new = crossOver(X_1, X_2)
        add X_new to P_new
    end
    for each Individual X in P_new do
        if rand(0,1) ≤ GreedyMutationRate then
            greedyMutate(X)
        else
            randomMutate(X)
        end
    end
    find the individual X_best ∈ P_new with best fitness
    if fitness(globalBest) < fitness(X_best) then
        globalBest = X_best
        nonImprovingSteps = 0
    else
        nonImprovingSteps++
    end
    if nonDiverseSteps ≥ twinRemovalInterval then
        activate twinRemoval(P_new) procedure
        nonDiverseSteps = 0
    else
        nonDiverseSteps++
    end
    if nonImprovingSteps ≥ randomRestartInterval then
        activate randomRestart(P_new) procedure
        nonImprovingSteps = 0
    end
    P = P_new
end
return globalBest
```

C. Fitness Evaluation

We have calculated the euclidean distance between each pair of points present in the individual's chromosome where each non-overlapping consecutive subsequence of length three in the chromosome represents the position of an atom in the protein structure in a 3-D space. We have assumed an upper bound, u_{ij} and lower bound, l_{ij} of the distances between pair of amino acids. The fitness of an individual X is defined by as Equation 1 below:

$$Fitness(X) = \left(\frac{1}{|E|} \sum_{(i,j) \in E} e_{ij}^2 \right)^{1/2} \tag{1}$$

where, $e_{ij} = max\{l_{ij} - \|c_i - c_j\|, \|c_i - c_j\| - u_{ij}, 0\}$ is the error associated to the constraints $l_{ij} \leq \|c_i - c_j\| \leq u_{ij}$ and $|E|$ denotes the number of distance pairs given. This is known as the *Largest Distance Error (LDE)* in the literature [2].

D. Genetic Operators

Genetic algorithms are guided by different genetic operators to help in keeping a balance of the exploitation and exploration of the search process. We use three types of genetic operators in our search: random mutation, greedy mutation and cross-over.

1) Random Mutation: Genetic diversity is a necessity for the process of evolution which is generally attained by mutation operator. Mutation helps the evolutionary process to guide through and look for different avenues and solutions in the search space. It also helps to avoid local optima by preventing the individuals from becoming too similar to one another. We have used uniform mutation in our method. We have mutated the value of each gene into a new value within the pre-specified bound of the search space($[0, 3.8 \times V]$). Whether an individual will be mutated or not depends on a mutation rate which is kept to a lower value of 0.015 to avoid primitive random search. A sketch of the pseudo-code is given in Algorithm 2.

Algorithm 2: Random Mutation (individual X)

$mutationRate = 0.015$
for *each gene X(i) in the genotype of* X **do**
  if $rand(0,1) \leq mutationRate$ **then**
    $X(i) = U(0,1) * (3 * V)$;
  end
end
return X

2) Greedy Mutation: In our proposed method, we have also used a greedy mutation where we greedily choose the new value of a particular gene in an individual. We randomly alter a certain gene and try r different random values for the gene temporarily. The value of gene that gives the best fitness of that individual is retained and used as the final new value of the gene. As, greedy mutation is computationally expensive, we make selection between the random mutation and greedy mutation with a probability, $GreedyMutationRate(= 0.8)$. The algorithm for greedy mutation is given in Algorithm 3.

Algorithm 3: Greedy Mutation (individual X)

set $r = 20$
for *each gene* $X(i)$ *in the genotype of* X **do**
  S = set of r random values for gene $X(i)$ from the range $[0, 3.8 * V]$
  find $v \in S$ for which fitness(X) is minimum
  set, $X(i) = v$
end
return X

3) Cross-over: Cross-over operators help genetic process to exploit the better solutions found thus far along the evolution process and regenerate new individuals by combining the genetic information of the individuals with better fitness value. It uses the better historical genetic information to guide the search into the search space regions with solutions having better fitness. By recombining the fitter individuals to generate new offspring, this operator is likely to produce more fitter individuals. We have used tournament selection with tournament size being equal to 5 to select two individual parents to take part in the cross-over operation. We then have recombined genes of the two parents and have generated a new set of genes for each of the offsprings with equal probability for each gene of being selected from any of the parents. The algorithm is outlined in Algorithm 4.

E. Twin Removal

We have removed and reinitialized individuals with identical genetic information. We have defined the similarity measure

Algorithm 4: Cross-Over (individual X_1, individual X_2)

X_{new} be a new offspring
$uniformRate = 0.5$
for *each gene X(i) in the genotype of* X_{new} **do**
  if $rand(0,1) \geq uniformRate$ **then**
    $X_{new}(i) = X_1(i)$
  else
    $X_{new}(i) = X_2(i)$
  end
end
return X_{new}

between two individuals as follows:

$$Similarity(X_1, X_2) = e^{-\frac{\|X_1 - X_2\|^2}{2\sigma^2}} \quad (2)$$

Here, the value of σ has been chosen to be 75. The $Similarity(X_1, X_2)$ function will return a value in the range [0,1]. The more the value of the similarity function is closer to 1, the more genetically similar are X_1 and X_2. We have chosen 0.8 as the threshold to define similarity between two individual. If $Similarity(X_1, X_2) >= 0.8$, we declare that X_1 and X_2 are twins and reinitialize randomly one of them. We have run the this Twin Removal procedure after every 100 generations of evolution.

Algorithm 5: Twin Removal

$similarityThreshold = 0.8$
for *each pair of individuals* (X_i, X_j) *in the population* **do**
  if $Similarity(X_i, X_j) \geq similarityThreshold$ **then**
    declare X_i and X_j as Twins
    reinitialize X_j
  end
end

F. Random Restart

If the algorithm does not show enough improvement within a significant number of generations, we reinitialize some individuals of the population with random values within the search space. We have evaluated the improvement over the immediate past 50 generations. If the improvement in the last 50 generations is less than or equal to a threshold, $t = 0.001$, we initiate the random restart procedure.

V. Experimental Results

We have implemented GreMuTRRR in Java programming language using JDK 1.6 and have run our experiments on an Intel 3.3GHz core $i3$ machine with 2GB RAM. We compared the performance of GreMuTRRR with a basic genetic algorithm without considering the greedy mutation, twin removal and random restart. We denote this algorithm as 'Basic GA' henceforth. Basic GA differs with GreMuTRRR given in Algorithm 1 in Lines 27–36 since it does not contain twin removal and random restart and in Line 3, as the parameter $greedyMutationRate = 0.0$ for Basic GA. All other parameters were kept same for the sake of fair comparison.

A. Benchmark

We have considered two sets of protein instances for our experiments: one considering only the backbone atoms (backbone only) and another containing all atoms (full atomic).

These are larger protein benchmarks introduced in [18]. We extracted the structures from PDB website [19] and have calculated the distance among pair of atoms. Lower and upper bounds of the distances are calculated using the following equation:

$$l_{ij} = (1-\epsilon)\hat{d}_{ij}, u_{ij} = (1+\epsilon)\hat{d}_{ij} \quad (3)$$

Here, $\hat{d}_{ij}$ is the real distance between point c_i and point c_j in the known structure of the protein sequence and $\epsilon = 0.8$. To make the dataset sparse, we took only 30% of the distances that are less or equal to $6\mathring{A}$.

B. Results

The largest distance error (LDE) values for our algorithm and Basic GA for backbone only instances and full atomic instances are given in Table I and Table II respectively. Each protein is reported with its corresponding PDB id and number of atoms in the structure considered.

TABLE I
BEST FITNESS FOR POPULATION SIZE 50 WITH RUN OF 2000 GENERATIONS(WITH BACKBONE ATOMS ONLY)

Protein Id	V	Basic GA	GreMuTRRR
1PTQ	50	0.49546	0.44977
1LFB	77	0.18060	0.13360
1F39	101	0.06595	0.02051
1AX8	130	0.06064	0.00253
1RGS	264	0.12380	0.00297
1TOA	277	0.12993	0.00156
1KDH	356	0.19119	0.00458
1BPM	481	0.34434	0.00441
1MQQ	679	0.58863	0.00894

TABLE II
BEST FITNESS FOR POPULATION SIZE 50 WITH RUN OF 1500 GENERATIONS(INCLUDING ALL ATOMS)

Protein Id	V	Basic GA	GreMuTRRR
1PTQ	402	0.71000	0.17224
1LFB	641	0.97121	0.31463
1F39	767	1.81091	0.16805
1AX8	1003	2.52303	0.43211

C. Analysis

From the results presented in the tables it is evident that for both of the instances GreMuTRRR achieves better solutions compared to the Basic GA algorithm for the proteins. However, the improvement is more clearly shown in Figure 1. GreMuTRRR significantly lowers the objective function LDE before both algorithms gets stuck and can not improve further. Its interesting to note thatGreMuTRRR is able to converge very early ($\leq$ 50 generation) compared to the Basic GA.

VI. CONCLUSION

In this paper, we presented a new genetic algorithm to solve the molecular distance geometry for protein structure determination problem using NMR data. We use a greedy mutation operator to intensify the search, a twin removal technique for diversification in the population and a random restart method to recover stagnation. On a standard set of benchmark dataset, our algorithm significantly outperforms standard genetic algorithms. In future, we want to implement a web service based on our method and make it public to be used by biologists.

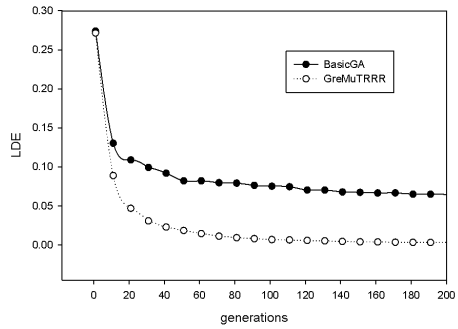


Fig. 1. Search progress for two algorithms for the protein 1AX8.

REFERENCES

[1] M. Souza, C. Lavor, A. Muritiba, and N. Maculan, "Solving the molecular distance geometry problem with inaccurate distance data," *BMC bioinformatics*, vol. 14, no. Suppl 9, p. S7, 2013.

[2] L. Liberti, C. Lavor, N. Maculan, and A. Mucherino, "Euclidean distance geometry and applications," *SIAM Review*, vol. 56, no. 1, pp. 3–69, 2014.

[3] D. Wu and Z. Wu, "An updated geometric build-up algorithm for solving the molecular distance geometry problems with sparse distance data," *Journal of Global Optimization*, vol. 37, no. 4, pp. 661–673, 2007.

[4] A. Mucherino, L. Liberti, C. Lavor, and N. Maculan, "Comparisons between an exact and a metaheuristic algorithm for the molecular distance geometry problem," in *Proceedings of the 11th Annual conference on Genetic and evolutionary computation*. ACM, 2009, pp. 333–340.

[5] L. Liberti, C. Lavor, A. Mucherino, and N. Maculan, "Molecular distance geometry methods: from continuous to discrete," *International Transactions in Operational Research*, vol. 18, no. 1, pp. 33–51, 2011.

[6] J. J. Moré and Z. Wu, "Distance geometry optimization for protein structures," *Journal of Global Optimization*, vol. 15, no. 3, pp. 219–234, 1999.

[7] C. Lavor, L. Liberti, and N. Maculan, "Computational experience with the molecular distance geometry problem," in *Global Optimization*. Springer, 2006, pp. 213–225.

[8] G. M. Crippen and T. F. Havel, *Distance geometry and molecular conformation*. Research Studies Press Taunton, UK, 1988, vol. 74.

[9] C. Savarese, J. M. Rabaey, and J. Beutel, "Location in distributed ad-hoc wireless sensor networks," in *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on*, vol. 4. IEEE, 2001, pp. 2037–2040.

[10] D. Tolani, A. Goswami, and N. I. Badler, "Real-time inverse kinematics techniques for anthropomorphic limbs," *Graphical models*, vol. 62, no. 5, pp. 353–388, 2000.

[11] J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[12] J. J. Moré and Z. Wu, "Global continuation for distance geometry problems," *SIAM Journal on Optimization*, vol. 7, no. 3, pp. 814–836, 1997.

[13] L. Liberti and S. Kucherenko, "Comparison of deterministic and stochastic approaches to global optimization," *International Transactions in Operational Research*, vol. 12, no. 3, pp. 263–285, 2005.

[14] L. Liberti and M. Drazic, "Variable neighbourhood search for the global optimization of constrained nlps," in *Proceedings of GO Workshop, Almeria, Spain*, vol. 2005, 2005.

[15] L. Liberti, C. Lavor, N. Maculan, and F. Marinelli, "Double variable neighbourhood search with smoothing for the molecular distance geometry problem," *Journal of Global Optimization*, vol. 43, no. 2-3, pp. 207–218, 2009.

[16] Q. Dong and Z. Wu, "A geometric build-up algorithm for solving the molecular distance geometry problem with sparse distance data," *Journal of Global Optimization*, vol. 26, no. 3, pp. 321–333, 2003.

[17] B. Hendrickson, "The molecule problem: Exploiting structure in global optimization," *SIAM Journal on Optimization*, vol. 5, no. 4, pp. 835–857, 1995.

[18] P. Biswas, K.-C. Toh, and Y. Ye, "A distributed sdp approach for large-scale noisy anchor-free graph realization with applications to molecular conformation," *SIAM Journal on Scientific Computing*, vol. 30, no. 3, pp. 1251–1277, 2008.

[19] H. M. Berman, J. D. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, "The protein data bank," *Nucleic Acids Research*, vol. 28, no. 1, pp. 235–242, 2000.

Empirical Prediction of Optical Transitions in Metallic Armchair SWCNTs

G. R. Ahmed Jamal[1,*] and S. M. Mominuzzaman[2]

[1]Department of Electrical and Electronic Engineering, University of Asia Pacific
Dhaka, Bangladesh

[2]Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology
Dhaka, Bangladesh

[*]ahmed.eee@uap-bd.edu

Abstract—**In this work, a quick and effective method to calculate the second and third optical transition energies of metallic armchair single wall carbon nanotubes (SWCNT) is presented. In this proposed method, the transition energy of any armchair SWCNT can be predicted directly just by knowing one of its chiral index. The predicted results are compared with recent experimental data and found to be accurate over a wide diameter range. The empirical equation proposed here is also compared with that proposed in earlier works. The proposed way may help the research works or applications where information of optical transitions of armchair metallic nanotubes is needed.**

Index Terms—**SWCNT, chiral index, metallic, optical transition energy.**

I. Introduction

Electronic and optical properties of single-wall carbon nanotubes (SWCNT) are directly associated with their geometrical structures [1, 2] which are uniquely specified by a pair of chiral index (n, m). A SWCNT (n, m) will be metallic if its n-m = 3k (k is integer), i.e. mod(n-m, 3) = 0 and it will be semiconducting if its n-m ≠ 3k, i.e. mod(n-m, 3) = 1 or 2 [3]. This relation is always found true except for SWCNT with very small diameter, where curvature effect dominates its properties [4]. This relation indicates that theoretically two third of the total SWCNTs are semiconducting and one third are metallic. Metallic nanotubes are a potential choice for nanoscale electrodes and 1-D quantum wires [5], transparent conductors and semi-transperant conductive coating [6, 7] and many similar nano device applications.

The one-dimensionality of the nanotubes gives rise to a discrete set of singularities (splitting into 1D subbands instead of one wide electronic energy band), the so called *van Hove singularities* (vHS) in the nanotube Density of States (DOS) [2,8,9]. Each SWCNT (*n*, *m*) has a unique set of interband energies E_{ii} denoting the energy differences between the *i*th van Hove singularities in the conduction and valence bands and optical transitions associated with the vHSs can only occur between these mirror subbands [2]. This vHS is an unique feature of nanotubes and also primarily responsible for many distinguished optical and electronic properties of SWCNTs.

According to the electronic band theory, the vertical interband transition energy (E_{ii}) between the *i*-th pair of VHSs in both the valence and the conduction bands of a semiconducting (S) or metallic (M) nanotube is represented by E_{ii} which is inversely proportional to nanotube diameter and corresponds to the first, second, third,…… interband transitions (E_{11}^S, E_{22}^S, E_{44}^S, ….) of semiconducting SWCNTs when *i* = 1, 2, 4,… and to the first, second, third interband transitions (E_{33}^M, E_{66}^M, E_{99}^M…) of metallic SWCNTs when *i* = 3, 6, 9,… respectively [2]. This inverse proportional trend of transition energy with SWCNTs diameter is also observed from well known Kataura plot [10].

Theoretical model derived from electronic band theory fails quantitatively to predict experimentally observed values of optical transitions for both semiconducting and metallic SWCNTs. Nanotube's 'curvature effect' [4, 11-13] and 'trigonal warping effect' [14] cause deviations in the electronic properties of nanotubes derived from the simple π-orbital graphene picture. Curvature causes π and σ states of nanotube's chemical bond to mix. This leads to hybridization between σ and π orbitals. The degree of hybridization becomes larger as the diameter of a SWCNT gets smaller [12]. The σ–π hybridization effect has been considered and calculated in the literature [11-13]. The main result is that nanotubes satisfying |n−m| = 3k (k=0, 1, 2….) develop a small curvature-induced bandgap, and hence become quasi-metallic or small gap semiconductor [15]. Armchair nanotubes (n, n) are an exception because of their special symmetry, and only they remain truly metallic for all diameters [15]. Saito *et al* [14] investigated the 'trigonal warping effect' analytically and derived expressions to estimate the corresponding deviation in optical transitions of metallic and semiconducting SWCNTs. They showed that trigonal warping causes splitting of the density of State (DOS) peaks of metallic SWCNTs except for armchair tubes. Thus only armchair SWCNTs are unaffected due to curvature and trigonal warping and remain truly metallic.

As armchair nanotubes are the most pure kind of metallic SWCNTs, the concentration of this work will be only on armchair tubes. Sometimes, researchers need to find or measure the optical transitions in metallic nanotubes. Many authors [16-19] studied the properties of metallic SWCNTs and discussed first, second and third optical transition energies of metallic SWCNTs. From these Liu *et al* [19] studied the matter more recently and more extensively than anyone before. They carried a series of optical experiments and measured optical transition energies of numerous semiconducting and metallic SWCNTs. The experimental data of optical transitions of metallic armchair SWCNTs used in this work is primarily based on the work of Liu *et al* [19]. Though sometimes researchers need the value of optical transitions in metallic nanotubes, but there is no easy way to calculate or predict those transitions except experimental procedure. Some complex empirical relations are proposed by some authors

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

[16,18,19] but they are not simple to calculate and involve lot of parameters. Here, in this work, a concise empirical relation is devised to predict the second and third optical transitions in armchair metallic nanotubes which directly relates the optical transition value with any of the chiral index of armchair tubes. The proposed relation can give a quick as well as highly accurate prediction of second and third optical transitions in armchair metallic SWCNTs.

II. Method and Results

Optical transition energies for metallic armchair (n, n) SWCNTs with chiral index (15,15) to (35,35) were closely studied here. There are total 21 armchair SWCNTs within these two chiral index range having diameter from 2 nm to 4.8 nm. Values of optical transition energies of these tubes were recorded from various experimental reports [16-19] with a special focus on the work of Liu *et al* [19] which is most recent and provides most comprehensive data. From these experimental reports, second optical transition energy of 16 armchair tubes ranging from chiral index (15,15) to (30,30) and third optical transition energy of 13 armchair tubes ranging from chiral index (23,23) to (35,35) were found.

In their work, Liu *et al* [19] proposed following empirical equation to fit their experimental values of different optical transitions of a number of semiconducting and metallic nanotubes:

$$E_p(k) = 2\hbar v_F(p)\times k + \beta\times k^2 + \eta(p)\times k^2\cos(3\theta)$$

Here, p is the transition index, k is the magnitude of wave vector in the grapheme Brillouin zone that varies with nanotube chirality (n, m) and transition index p. Magnitude of k is given by p× 2/(3d), where d is the nanotube diameter. $E_p(k)$ is the effective dispersion for transition p, $\hbar$ is the reduced Planck constant = 6.582×10^{-16} ev.s and β=-0.173 ev. nm^2. Rest two parameters v_F and η have different values for different transition index p.

For second optical transition energy (E_{66}) of metallic tubes, p=6, so, k= 4/d. Liu *et al* [19] proposed v_F (p) =1.236×10^6 m/s and η(p) = 0.058 ev.nm^2 for p=6. Using their proposed values and using the fact that Θ=30° for armchair tubes, the expression for second optical transition energy (E_{66}) of metallic armchair tubes reduces to,

$$E_{66} = \frac{6.5083}{d} - \frac{2.768}{d^2} \qquad (1)$$

For third optical transition energy (E_{99}) of metallic tubes, p=9, so, k= 6/d. Liu *et al* [19] proposed v_F (p) =1.248×10^6 m/s and η(p) = 0.047 ev.nm^2 for p=9. Using their proposed values and using the fact that Θ=30° for armchair tubes, the expression for third optical transition energy (E_{99}) of metallic armchair tubes reduces to,

$$E_{99} = \frac{9.8572}{d} - \frac{6.228}{d^2} \qquad (2)$$

So, these are the empirical expression of E_{66} and E_{99}, simplified from the work of Liu *et al* [19]. One can easily note that lot of parameters are involved in the overall calculation of empirical values of E_{66} and E_{99} from their prarent equation which increases computational efforts. Hence, an attempt will be made here to devise a simple empirical relation between the chirality and optical transition of armchair metallic tubes so that the value of second or third optical transition of these tubes can be predicted directly from just one of their chiral index n (for armchair tubes, n=m, so, knowing one index will be enough).

After a careful study of the symmetry in the structure of armchair tubes, their optical transitions and their relation with corresponding chiral index, we devised following empirical formula to predict the optical transitions from corresponding chiral index **n** of any armchair nanotube (n, n).

For second optical transition energy (E_{66}) of armchair SWCNTs:

$$E_{66} = \frac{50}{n\exp(4.3/n)} \qquad (3)$$

For third optical transition energy (E_{99}) of armchair SWCNTs:

$$E_{99} = \frac{75}{n\exp(6/n)} \qquad (4)$$

Empirical values of second and third optical transition energies of these armchair SWCNTs with chiral index (15,15) to (35,35) were calculated using empirical equations (1) and (2) as well as using empirical equations (3) and (4). Both results agree well with experimental values of second and third optical transitions of corresponding metallic armchair SWCNTs. Nevertheless, it was observed that the overall prediction accuracy of equations (3) and (4) as proposed in this work is much better than that of equations (1) and (2) as simplified from the parent equation proposed by Liu *et al* [19]. Besides, equations (3) and (4) gives quite simpler expressions to calculate the value of optical transitions directly from one chiral index in compared to the expression of equations (1) and (2) that requires additional calculation to know the value of diameter of corresponding armchair tube. Moreover, it has to be noted that the simplified expression of equations (1) and (2) results from its parent equation only after knowing the values of a number of parameters. Hence, empirical equations (3) and (4) as proposed in this work is better than simplified empirical equations (1) and (2) from the work of Liu *et al* [19].

Experimental [19] and empirical values of optical transition energies along with chiral index (n, n) and diameter of these armchair tubes are shown in Table I and Table II. In Table I, empirical result for second optical transition (E_{66}) from this work using Eqn. (3) and the empirical result from the work of Liu *et al* [19] using Eqn. (1), both are provided so as to compare them. Similarly, in Table II, empirical result for third optical transition (E_{99}) from this work using Eqn. (4) and the empirical result from the work of Liu *et al* [19] using Eqn. (2), both are provided so as to compare them. From these two tables it can be observed clearly that the predicted values from Eqn. (3) and (4) are much closer to the experimental values than those predicted by Eqn. (1) and (2). For second optical transition (E_{66}) the average absolute error in prediction using Eqn. (3) was found to be only 0.009 ev and corresponding % average absolute error is only 0.53%. Similarly, for third optical transition (E_{99}) the average absolute error in prediction using Eqn. (4) was found to be only 0.006 ev and corresponding % average absolute error is only 0.27%. Thus, the empirical equations proposed here can predict the value of

TABLE I
Experimental [19] and Empirical values of Second optical transition energy of 16 armchair SWCNTs

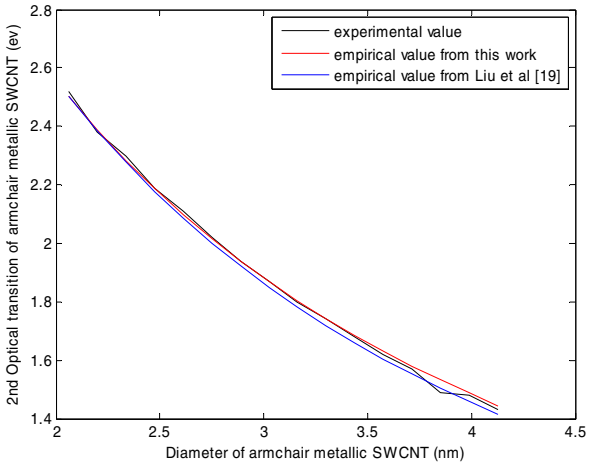
n	m	Diamater d_t (nm)	Exp E_{66} (ev)	Emp E_{66} (ev) from Eqn. (3)	Emp E_{66} (ev) from Eqn. (1)
15	15	2.063	2.52	2.503	2.505
16	16	2.200	2.38	2.389	2.386
17	17	2.338	2.30	2.284	2.277
18	18	2.475	2.19	2.188	2.177
19	19	2.613	2.11	2.099	2.085
20	20	2.750	2.02	2.016	2.000
21	21	2.888	1.94	1.940	1.922
22	22	3.026	1.87	1.869	1.849
23	23	3.163	1.80	1.803	1.781
24	24	3.301	1.74	1.742	1.718
25	25	3.438	1.68	1.684	1.659
26	26	3.576	1.62	1.630	1.604
27	27	3.713	1.57	1.579	1.552
28	28	3.851	1.49	1.531	1.503
29	29	3.988	1.48	1.487	1.458
30	30	4.126	1.43	1.444	1.415

TABLE II
Experimental [19] and Empirical values of third optical transition energy of 13 armchair SWCNT

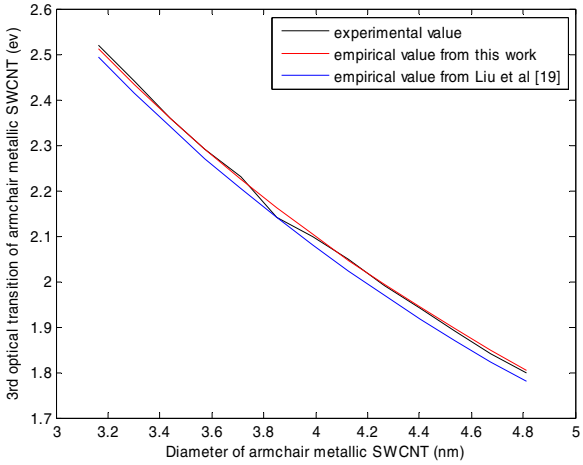
n	m	Diamater d_t (nm)	Exp E_{99} (ev)	Emp E_{99} (ev) from Eqn. (4)	Emp E_{99} (ev) from Eqn. (2)
23	23	3.163	2.52	2.512	2.494
24	24	3.301	2.44	2.434	2.415
25	25	3.438	2.36	2.360	2.34
26	26	3.576	2.29	2.290	2.27
27	27	3.713	2.23	2.224	2.203
28	28	3.851	2.14	2.162	2.14
29	29	3.988	2.10	2.103	2.08
30	30	4.126	2.05	2.047	2.023
31	31	4.263	1.99	1.994	1.969
32	32	4.401	1.94	1.943	1.918
33	33	4.538	1.89	1.895	1.87
34	34	4.676	1.84	1.849	1.823
35	35	4.813	1.80	1.805	1.779

optical transition energies of armchair tubes with higher accuracy.

Figure 1 (a) and (b) shows the plot of experimental and empirical values of second and third optical transition energies, respectively, of metallic armchair SWCNTs with respect to their diameters. Here also, empirical results for second and third optical transition (E_{66} and E_{99}) from this work using Eqn. (3) and (4) and the empirical result from the work of Liu *et al* [19] using Eqn. (1) and (2), both are plotted so as to compare them. Again it can be observed from these two plots that, for both optical transitions, empirical values predicted by Eqn. (3) and (4) are closer to the experimental values in compared to those predicted by Eqn. (1) and (2) over the full diameter range under consideration. This curve can be extrapolated using above two empirical relations to predict other armchair



(a)



(b)

Fig. 1 Experimental value(black), empirical value from this work(red) and empirical value from earlier work (blue) of optical transition vs diameter of armchair SWCNTs, a) second optical transition b) third optical transition.

tubes beyond this diameter range.

Another interesting observation from the second and third optical transition energies of armchair SWCNTs is, the ratio of third transition to second transition of same armchair tubes are equal to a constant whose value is around 1.4. This observation further facilitate the calculation of higher optical transitions in metallic armchair tubes as one can know the third transition energy value of a tube by knowing the value of second transition multiplied by that constant. This relation also indicates the built in high symmetry in armchair SWCNTs which causes their higher optical transition energy values to grow in a systematic order. This relation may also help to study the behaviour of metallic armchair tubes in a more systematic way.

III. Conclusions

In this work, an effective method to calculate the second and third optical transition energies of metallic armchair single wall carbon nanotubes (SWCNTs) is presented. Second and third optical transitions of metallic armchair (n, n) SWCNTs are expressed through a concise empirical formula that directly relates the transition energy with chiral index **n**. Thus, in this method, the transition energy of any armchair SWCNT can be predicted quickly and directly just by

knowing one of its chiral index. The predicted or calculated results are compared with recent experimental data and found to be accurate over a wide diameter range. The empirical equations and corresponding result found in this work are also compared with empirical works proposed by others. It was found that the empirical expression proposed here is much simpler than earlier works and corresponding result provides better accuracy for armchair metallic SWCNTs than earlier works. The proposed way may help the research works or applications where information of optical transitions of armchair metallic nanotubes is needed.

REFERENCES

[1] V. N. Popov, "Carbon nanotubes : properties and application", *Materials Science and Engineering R*, 43, pp. 61–102, 2004.

[2] T. W. Odom, J. L. Huang, P. Kim, and C. M. Lieber, "Structure and Electronic Properties of Carbon Nanotubes", *J. Phys. Chem. B*, *104,* pp.2794-2809, 2000.

[3] N. Hamada, S. Sawada, and A. Oshiyama, "New one-dimensional conductors: graphitic microtubules," *Phys. Rev. Lett.*, Vol.68, No.10, pp.1579-1581, 1992.

[4] V. Zólyomi and J. Kürti, "First-principles calculations for the electronic band structures of small diameter single-wall carbon nanotubes", Phys. Rev. B 70, 085403, 2004.

[5] S. J. Tans, M. H. Devoret, H. Dai, A. Thess, R. E. Smalley, L. J. Geerligs and C. Dekker, "Individual single-wall carbon nanotubes as quantum wires", *Nature,* 386, 474, 1997.

[6] M. Kociak, A. Yu. Kasumov, S. Guéron, B. Reulet, I. I. Khodos, Yu. B. Gorbatov, V. T. Volkov, L. Vaccarini, and H. Bouchiat , "Superconductivity in Ropes of Single-Walled Carbon Nanotubes" *Phys. Rev. Lett.* 86, 2416, 2001.

[7] A. A. Green, M. C. Hersam, "Colored semitransparent conductive coatings consisting of monodisperse metallic single-walled carbon nanotubes", *Nano Lett.* 8, pp. 1417–1422, 2008.

[8] J. W. Mintmire and C. T. White, "Universal density of states for carbon nanotubes", Phys. Rev. Lett., Vol. 81, No.12, 1998.

[9] S. Reich and C. Thomsen, "Chirality dependence of the density-of-states singularities in carbon nanotubes", Phys. Rev. B, Vol 62, No. 7, 2000.

[10] H. Kataura, Y. Kumazawa, Y. Maniwa, I. Umezu, S. Suzuki, Y. Ohtsuka, and Y.Achiba, "Optical properties of single - wall carbon nanotubes," *Synthetic Met.*, Vol.103, pp.2555, 1999.

[11] V. N. Popov, "Curvature effects on the structural, electronic and optical properties of isolated single-walled carbon nanotubes within a symmetry-adapted non-orthogonal tight-binding model", *New Journal of Physics*, Vol. 6, 2004.

[12] H. Zeng, H. F. Hu, J. W. Wei, Z. Y. Wang, L. Wang, and P. Peng, "Curvature effects on electronic properties of small radius nanotube," *Appl. Phys. Lett.*, Vol.91, No.3, pp.033102, 2007.

[13] O. G¨ulseren, T. Yildirim and S. Ciraci, "A systematic ab-initio study of curvature effects in carbon nanotubes", *Phys. Rev.* B 65, 153405, 2002.

[14] R. Saito, G. Dresselhaus, and M. S. Dresselhaus, "Trigonal warping effect of carbon nanotubes," *Phys. Rev. B*, Vol.61, No.4, pp.2981-2990, 2000.

[15] Francois Leonard, "The physics of carbon nanotube devices", Ed: Jeremy Ramsden, *William Andrew Inc.* Norwich, New York. 2009.

[16] M. S. Strano, S. K. Doorn, E. H. Haroz, C. Kittrell, R. H. Hauge and R. E. Smalley, "Assignment of (*n*, *m*) Raman and Optical Features of Metallic Single-Walled Carbon Nanotubes", *Nano Lett.*, Vol.3, No.8, pp.1091-1096, 2003.

[17] C. Fantini, A. Jorio, M. Souza, M. S. Strano, M. S. Dresselhaus and M. A. Pimenta, "Optical Transition Energies for Carbon Nanotubes from Resonant Raman Spectroscopy:Environment and Temperature Effects", *Phys. Rev. B*, Vol. 93, No. 14, 2004.

[18] J. Maultzsch, H. Telg, S. Reich, and C. Thomsen, "Radial breathing mode of single-walled carbon nanotubes Optical transition energies and chiral-index assignment", *Phys. Rev. B* 72, 205438, 2005.

[19] Liu K, Deslippe J, Xiao F, Capaz RB, Hong X, Aloni S, Zettl A, Wang W, Bai X, Louie SG, Wang E, Wang F., "An atlas of carbon nanotube optical transitions", *Nat. Nanotech. 7*, 325–329, 2012.

Dynamics of Fullerene Self-Insertion into Carbon Nanotubes in Water

M. M. Rahman, M. M. Chowdhury, and M. K. Alam*

Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology
Dhaka-1205, Bangladesh
*kawsaralam@eee.buet.ac.bd

Abstract—**The dynamics of fullerene self-insertion into carbon nanotubes in water are studied by molecular dynamics simulations. It is revealed that the van der Waals interaction between the nanotube and fullerene introduces a spontaneous insertion of fullerene molecule inside the nanotube in the presence of water molecules when the initial separation between the nanotube and the fullerene is less than a critical distance. It is also demonstrated that the insertion process is strongly dependent upon the radius of the nanotubes and does not show a significant dependence upon the temperature or the length of the nanotubes. The study suggests possible bottom–up self-assembly of novel nanostructures like peapod (one or more fullerene molecules inside a carbon nanotube), which has great potential in the fields of nano-oscillator and futuristic quantum information processor.**

Index Terms— **Carbon nanotube, fullerene, molecular dynamics, water.**

I. Introduction

In the past few decades, carbon nanotubes (CNTs) have drawn a great deal of attention as a model system for studying many important physical phenomena because of its extraordinary mechanical and electronic properties [1, 2]. These properties along with the rigid structure make it ideal for an integrated chip design [3], biosensors [4, 5] and fuel storage [6]. Besides, these nonpolar rigid structures can be used as unique molecular channels for water and protons, with tunable conductivity by changing the local channel polarity and solvent conditions [7, 8]. Recently, CNT in water system has drawn a lot of interest because of its emerging applications in superfast water transport [9] and water supply [8]. The possibility of tuning the properties of CNTs by inserting molecules of different sizes and doping has improved the scope of using CNTs in nanoelectronics [3]. Dynamics of different molecules like CNTs, DNA, fullerene translocating through CNTs as nanofluidic channels both in vacuum and in different media have been reported [10, 11]. Recently, fullerene peapod (i.e. an array of fullerene molecules in a CNT) has got extensive attention for its potential use in nano-oscillator [12]. Peapod is also considered as a possible candidate for quantum information processor [13, 12]. Xue *et al.* [11] has presented that molecules translocating through CNTs in vacuum are encapsulated by the nanotubes in an oscillating manner whereas in the presence of aqueous media, the oscillation is strongly impeded. Khlobystov *et al.* [10] has also reported the possibility of assembling fullerene arrays inside CNTs in supercritical fluids at low temperatures. However, a detail analysis on the interaction of CNTs and fullerene in water has not been reported yet. A comprehensive study is required to fully understand the physics and dynamics of these nanostructures and to explore their usage for novel applications.

In the present study, the dynamics of self-insertion of fullerene through single-walled carbon nanotube (SWCNT) is investigated in water using molecular dynamics (MD) simulation [14, 15]. The influence of van der Waals (vdW) interaction between the CNT and the fullerene has been studied. In addition, effect of the CNT chiralities and length on the self-insertion time has been investigated. We have also varied the initial center of mass (COM) distance between the CNT and the fullerene to observe its impact on the insertion time and mechanism. The effect of temperature on the insertion process has been studied as well.

II. Model and Simulation Methods

CNTs are characterized by two indices (n, m) corresponding to two unit vectors ($\boldsymbol{a_1}$ and $\boldsymbol{a_2}$) in the crystal lattice of graphene. The structures with n = m are known as armchair and the structures with m = 0 are known as zigzag. Throughout this study, armchair CNTs (with chiral indices (10, 10), (12, 12) and (15, 15)) have been used along with the buckyball (C_{60}) fullerene molecule. For molecular dynamics simulation [14, 15], CNTs and the fullerene molecules can be described by a Morse bond potential, a harmonic cosine potential, a twofold cosine potential for torsion and a Lennard-Jones (LJ) 12-6 potential for nonbonding interaction [16]:

$$U\left(r_{ij},\theta_{ijk},\varphi_{ijkl}\right)=K_{Cr}\left[\exp\left(-\gamma\left(r_{ij}-r_C\right)\right)-1\right]^2$$
$$+\frac{1}{2}K_{C\theta}\left(cos\theta_{ijk}-cos\theta_C\right)^2+\frac{1}{2}K_{C\varphi}\left(1+\cos\left(2\varphi_{ijkl}-\varphi_C\right)\right)$$
$$+4\varepsilon_{CC}\left[\left(\frac{\sigma_{CC}}{r_{ij}}\right)^{12}-\left(\frac{\sigma_{CC}}{r_{ij}}\right)^{6}\right] \quad (1)$$

where, the bonding parameters K_{Cr}, r_C and γ are the parameters for the Morse bond potential; $K_{C\theta}$ and θ_C are the parameters for the bending angle potential; $K_{C\phi}$ and ϕ_C are the torsion parameter [16]. The nonbonded parameters ϵ_{CC} and σ_{CC} are the Lennard-Jones interaction for carbon-carbon interactions and ϵ_{CO} and σ_{CO} are Lennard-Jones interaction for carbon-water interactions [17]. The values of the interaction potential parameters for carbon atoms are given in Table I.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

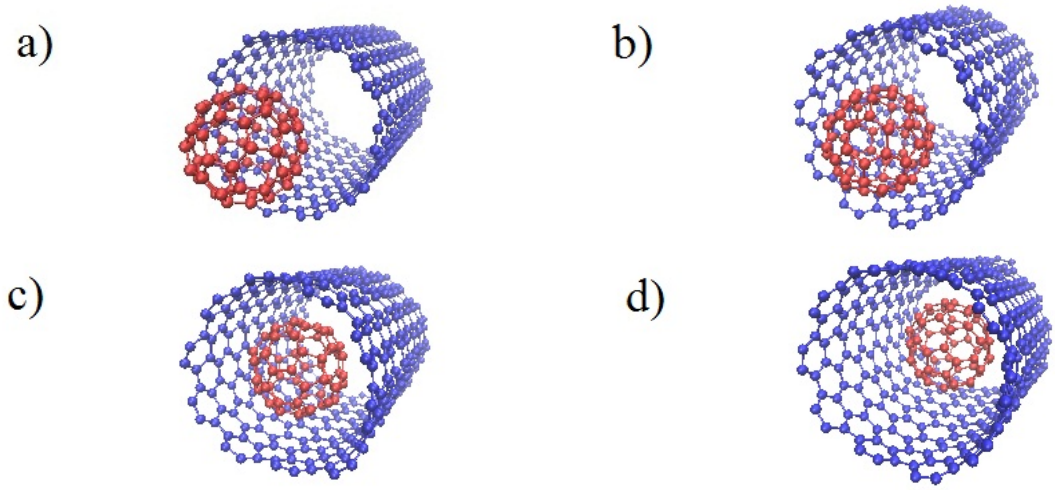


Fig. 1. Simulation snapshots of the self-insertion process of a fullerene inside a (10, 10) nanotube at (a) $t = 0$ ps, (b) $t = 2$ ps, (c) $t = 5$ ps and (d) $t = 100$ ps (water molecules omitted for clarity). The molecular dynamic simulation showing the movement of the molecules up to 20 ps can be found in the video files uploaded on – http://teacher.buet.ac.bd/kawsaralam/res.html.

TABLE I
PARAMETERS FOR INTERACTION POTENTIALS USED IN THE SIMULATIONS

Interaction Potential Parameters	
K_{Cr} = 478.9 kJ mol^{-1}	r_C = 0.1418 nm
$K_{C\varphi}$ = 25.12 kJ mol^{-1}	γ = 21.867 nm^{-1}
$K_{C\theta}$ = 562.2 kJ mol^{-1}	θ_C = 120°
ε_{CC} = 0.3601 kJ mol^{-1}	φ_C = 180°
ε_{CO} = 0.4787 kJ mol^{-1}	σ_{CC} = 0.3400 nm
	σ_{CO} = 0.3275 nm

The CNT-water and fullerene-water interactions have been described by a Carbon-Oxygen LJ potential [17]:

$$U\left(r_{ij},\theta_{ijk}\right)=\frac{1}{2}K_{Wr}\left(r_{ij}-r_W\right)^2+\frac{1}{2}K_{W\theta}(cos\theta_{ijk}-cos\theta_W)^2 \qquad (2)$$

where, r_W and θ_W are the reference O-H bond length and H-O-H angle, and the K_{Wr} and $K_{W\theta}$ are the corresponding force constants. The well-known TIP3P water model [18] has been used to represent water molecules. The values of the interaction potential parameters for water are given in Table II.

TABLE III
PARAMETERS FOR THE TIP3P WATER MODEL [19]

Interaction Potential Parameters	
K_{Wr} = 502416 kJ mol^{-1}	r_W = 0.09572 nm
$K_{W\theta}$ = 628.02 kJ mol^{-1}	θ_W = 104.52°
ε_{OO} = 0.6367 kJ mol^{-1}	σ_{OO} = 0.3151 nm
q_O = -0.834e	q_H = 0.417e

All simulations were performed at a constant pressure of 1 bar and the Berendsen thermostat [20] was used to keep the temperature constant with a coupling time of 0.1 ps. The cut-off radius for the Lennard-Jones interaction as well as Born electrostatics was 1.2 nm. The electrostatic interaction was evaluated using the particle-mesh-Ewald (PME) method [21]. To achieve a structure with a natural orientation, energy minimization was performed. Periodic boundary condition was applied in all directions to avoid problems with boundary effects caused by finite size of the system and no position restraint was applied in the simulation.

III. RESULTS AND DISCUSSION

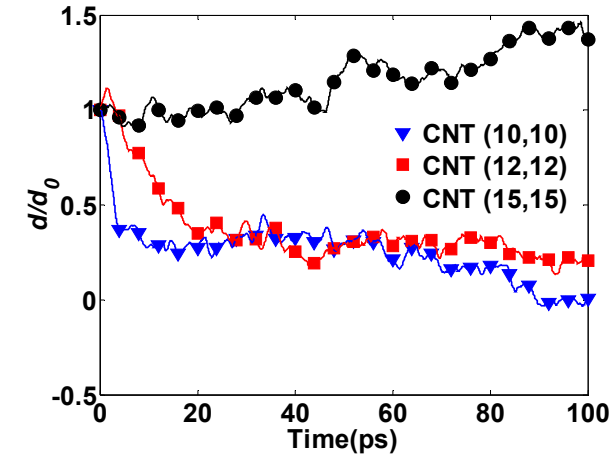
To explore the self-insertion process, we have performed a series of MD simulations with different CNT chirality, length, temperature and initial COM separation between the CNT and the fullerene. Fig. 1 shows the snapshots of the self-insertion process of a fullerene inside a (10, 10) nanotube at $t = 0$, 2, 5, 10 ps. For a better understanding of the simulation results, an illustration video of the MD simulation for 20ps is uploaded on http://teacher.buet.ac.bd/kawsaralam/res.html.

We start with the simulation of an armchair CNT having chirality (10, 10) and a fullerene molecule separated by a distance of 0.1 nm. The length and radius of the CNT is 1.95 nm and 0.678 nm. Blue triangle marker in Fig. 2(a) shows the time vs. normalized COM distance (d/d_0) of this simulation (where d_0 is the initial COM distance between the CNT and the fullerene, and d is the COM distance at different time). It shows an extremely fast insertion process of the fullerene into the CNT within 4 ps. After the insertion process is completed, the fullerene buckyball oscillates slightly inside the nanotube due to the inertia effect. By bringing together more fullerene molecules close to this structure, it is possible to create a peapod (one or more fullerene molecules inside a carbon nanotube).

The physical mechanisms of the insertion process lies within the CNT-fullerene vdW energy which is plotted in Fig 2(b) [blue triangle marker] against time. The CNT-fullerene vdW energy decreases rapidly as a function of the insertion time and reaches a stable value after a while. If we change the chirality of the CNT to (12, 12) [corresponding radius of the nanotube is 0.814 nm], the insertion process takes longer time which is shown in Fig.

2(a) (red square marker). This indicates the dependence of insertion time with the radius of CNT. If the chirality of the CNT is further changed to (15, 15) [corresponding radius of the nanotube is 1.017 nm], the fullerene does not enter into the nanotube [Fig. 2(a) black circle marker]. This is because of the fact that the presence of the carbon atoms (of CNT molecule) too far away from the fullerene results in lower vdW force. To support this argument, another simulation was done (not shown here) with the fullerene initially placed close to the edge of the (15, 15) tube with the same separation of 0.1 nm. This time the fullerene gets into the CNT, confirming the essence of our argument. It is important to mention that there exists a lower limit for the radius of the CNT for a successful self-insertion too. The radius of the fullerene must be smaller than the radius of the CNT added to the interaction parameter σ_{CC}. This is caused by the repulsive short-range interactions between the CNT and the fullerene carbon atoms [15].

a)



b)

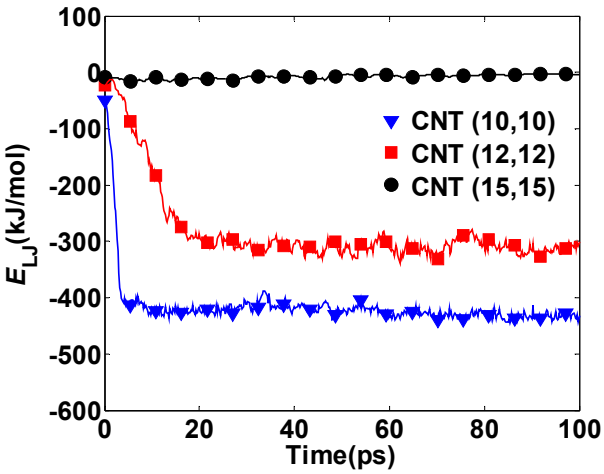


Fig. 2. The dynamic assembly process of fullerene and (10, 10), (12, 12) and (15, 15) CNTs. (a) Normalized COM distances between the fullerene and the CNTs as a function of time. (b) The Lennard-Jones potential for the CNT-fullerene interaction as a function of time.

If the initial separation between the CNT (10, 10) and the fullerene is changed from 0.1 to 0.2 nm the normalized COM distance vs. time does not show significant change (Fig. 3 red square marker). However, when it is increased to 0.3 nm, it shows a slow insertion (black circle marker). The fullerene shows a slower insertion with the initial separation further increased to 0.4 nm (magenta diamond marker). Interestingly, if the separation is further increased to 0.5 nm, the insertion process is stopped completely (green asterisk marker). At this critical distance, the fullerene is too far away from the CNT to feel the vdW force and hence fail to perform the self-insertion mechanism.

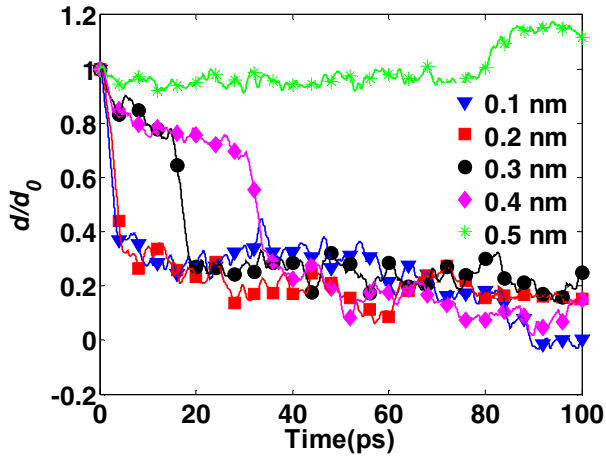


Fig. 3. The effect of initial position variation (between the CNT (10,10) and the fullerene) in the self-insertion process. The self-insertion process is stopped if the separation between the CNT and the fullerene is above 0.4 nm.

We have also examined the effect of temperature on the insertion process (for an initial separation of 0.1 nm between the (10, 10) CNT and the fullerene). It has been found that the insertion time has a very little dependence on the temperature from 200 K to 500 K (Fig. 4). However, at increased temperature of 500 K (green asterisk marker) after getting inserted into the CNT, the fullerene moves back and forth inside the nanotube due to the increased molecular mobility.

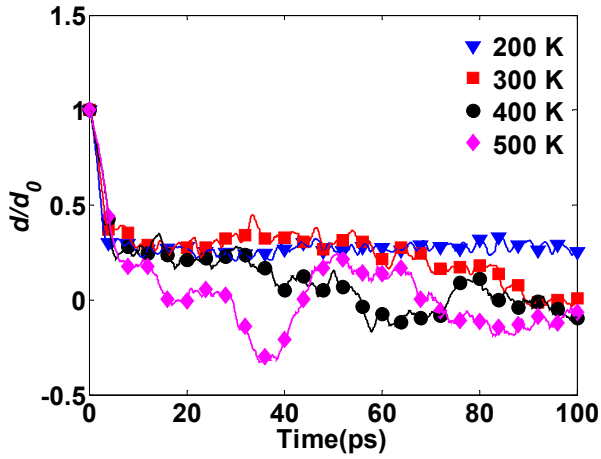


Fig. 4. The effect of temperature on the self-insertion process.

It has also been found that the length of the CNT does not play a significant role in the insertion time (Fig. 5a). This can be easily understood by looking at the change in the Lennard-Jones potential for the CNT and the fullerene (Fig. 5b). For an initial separation of 0.1 nm, the Lennard-Jones CNT-Fullerene interaction goes to a stable value at almost the same time for the CNT lengths of 1.47, 1.95 and 2.44 nm. If we increase the tube length, the carbon

atoms far away (e.g. atoms at the right side of the CNT in Fig. 1) from the fullerene molecule do not have a significant influence on the insertion process. The atoms (of the nanotube) which are near to the fullerene play the major role here. Therefore, the tube length does not affect the fullerene insertion time considerably (Fig. 5).

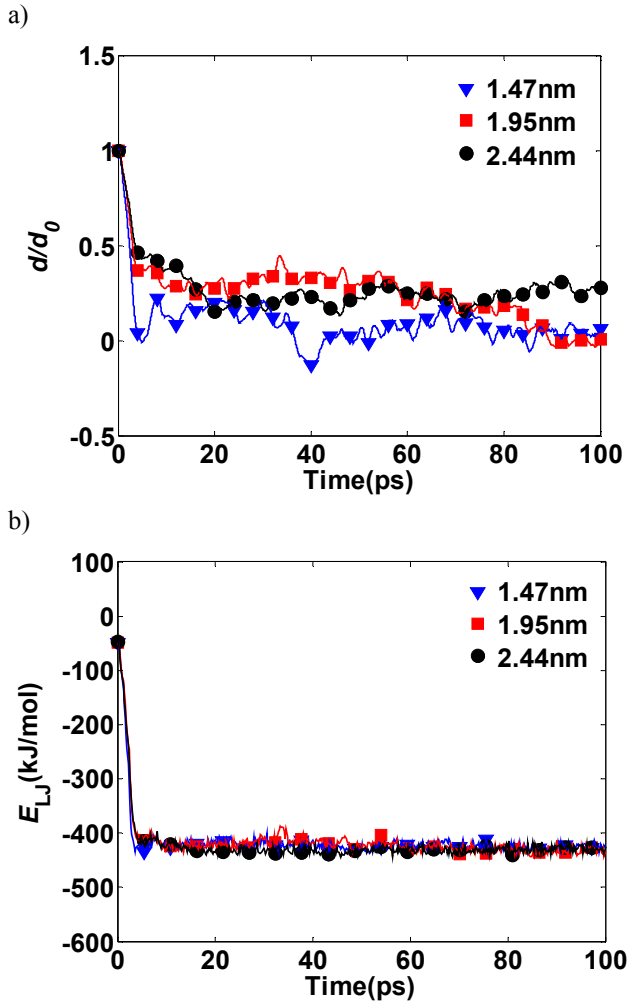


Fig. 5. The effect of variation in the length of the CNT. (a) The normalized COM distance between the CNTs and the fullerene as a function of time for CNT lengths of 1.47, 1.95 and 2.44 nm. (b) The Lennard-Jones potentials for the corresponding CNT-fullerene interaction as a function of time.

IV. Conclusions

In summary, we have conducted molecular dynamic simulation of the self-insertion of fullerene into carbon nanotubes in water. It has been observed that the insertion time of the fullerene is dependent upon the radius of the CNT as well as the initial separation between the fullerene and the CNT. However, after a critical distance the self-insertion process has been found to be completely stopped. It has also been observed that the insertion time does not show a significant dependence upon the temperature as well as the length of the CNT. To investigate further with more accuracy and precision, a combined quantum mechanics/molecular dynamics or the *ab initio* study could be performed.

References

[1] S. Iijima, "Helical microtubules of graphitic carbon", *Nature*, vol. 354, pp. 56-58, Nov. 1991.

[2] R. H. Baughman, A. A. Zakhidov, W. A. de Heer, "Carbon nanotubes--the route toward applications", *Science*, vol. 279, pp. 787-792, Aug. 2002.

[3] A. Javey, J. Guo, Q. Wang1, M. Lundstrom and H. Dai1, "Ballistic carbon nanotube field-effect transistors", *Nature*, vol. 424, pp. 654-657, Aug. 2003.

[4] C. R. Martin, P. Kohli, "The emerging field of nanotube biotechnology", *Nat. Rev. Drug Discovery*, vol. 2, pp. 29-37, Jan. 2003.

[5] F. Balavoine, P. Schultz, C. Richard, V. Mallouh, T. W. Ebbesen, T. W.; Mioskowski, C. Angew, "Helical crystallization of proteins on carbon nanotubes: a first step towards the development of new biosensors", *Angew. Chem. Int.*, vol. 38, pp. 1912–1915, Jul. 1999.

[6] Y.L. Chena, B. Liua, J. Wua, Y. Huang, H. Jiang, K.C. Hwanga, "Mechanics of hydrogen storage in carbon nanotubes", *J. Mech. Phys. Solids*, vol. 56 , pp. 3224–3241, 2008.

[7] G .Hummer, J. C. Rasalah and Noworyta, "Water conduction through the hydrophobic channel of a carbon nanotube", *Nature*, vol. 414(6860), pp. 188-90, Nov. 2001.

[8] H. Y. Yang, Z. J. Han, S. F. Yu, K. L. Pey, K. Ostrikov & R. Karnik, "Carbon nanotube membranes with ultrahigh specific adsorption capacity for water desalination and purification", *Nat. Commun.*, vol. 4, Aug. 2013.

[9] Q. Zhang, W.-Zhou Jiang, Jian Liu, Ren-De Miao, and Nan Sheng, "Water transport through carbon nanotubes with the radial breathing mode", *Phys. Rev. Lett.*, vol. 110, p. 254501, 2013.

[10] A. N. Khlobystov, D. A. Britz, J. Wang, S. A. O'Neil, M. P. and G. A. D. Briggsa, "Low temperature assembly of fullerene arrays in single-walled carbon nanotubes using supercritical fluids", *J. Mater. Chem.*, vol. 14, pp. 2852–2857, July 2004.

[11] Y. Xue and M. Chen, "Dynamics of molecules translocating through carbon nanotubes as nanofluidic channels", *Nanotechnology*, vol. 17, pp. 5216–5223, Sept. 2006.

[12] M. Wang and C. M. Li, "An oscillator in a carbon peapod controllable by an external electric field: a molecular dynamics study", *Nanotechnology*, vol. 21, p. 035704, Dec. 2009.

[13] W. L. Yang, Z. Y. Xu, H. Wei, M. Feng, D. Suter, "Quantum information processing architecture with endohedral fullerenes in a carbon nanotube", *Phys. Rev. A*, vol. 81, p. 032303, Mar. 2010.

[14] B. Hess, C. Kutzner, D. V. D. Spoel, E. Lindahl, "GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation", Journal of chemical theory and computation, vol. 4, pp. 435-477, March 11, 2008.

[15] D. van der Spoel, E. Lindahl, B. Hess, A. R. van Buuren, E. Apol, P. J. Meulenhoff, D. P. Tieleman, A. L. T. M. Sijbers, K. A. Feenstra, R. van Drunen and H. J. C. Berendsen, Gromacs User Manual version 4.6-beta1, www.gromacs.org (2010)

[16] J. H. Walther, R. Jaffe, T. Halicioglu, P. Koumoutsakos, "Carbon nanotubes in water: structural characteristics and energetics", *J. Phys. Chem. B*, vol. 105, pp. 9880-9887, Sept. 2001.

[17] G. Hummer, J. C. Rasaiah, J. P. Noworyta, "Nanopores: Water flow at the flip of a switch", *Nat. Nanotechnol.*, vol. 6, pp. 759-760, Dec. 2011.

[18] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, "Molecular dynamics simulation of liquid water: hybrid density functionals", *J. Chem. Phys.*, vol. 100, pp. 3685-3691, Dec. 2005.

[19] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, "Comparison of simple potential functions for simulating liquid water", *J. Chem. Phys.*, vol. 79, pp. 926– 935, Apr. 1983.

[20] Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., Haak, J.R., "Molecular dynamics with coupling to an external bath", *J. Chem. Phys.*, vol. 81, pp. 3684–3690, Oct. 1984.

[21] T. A. Darden, D. M. York and L. G. Pedersen, "Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems", *J. Chem. Phys.*, vol. 98, pp. 10089-10092, Apr. 1993.

Low temperature synthesis of carbon nanotubes and carbon nanospheres

Md. Golam Kibria, [1,*] and A. K. M. Bazlur Rashid[1]

[1]Materials and Metallurgical Engineering, BUET
Dhaka, Bangladesh
[*]kibria.buet004@gmail.com

Abstract—**Recently nanomaterial have gained a great interest due to their unique electrical and optical properties. The area of carbon materials have since been dominated by studies of carbon nanotubes. So far numerous methods have been developed for the synthesis of CNTs; in this work we have synthesized CNTs via a simple chemical route. Carbon nanotubes (CNTs) are successfully synthesized by heating a mixture of ferrocene and ammonium chloride at temperature as low as 250 ^{0}C. Carbon nanospheres are also observed here. Carbon spheres (CSs) of different sizes and shapes are present randomly in the sample. Field emission scanning electron microscopy is performed to examine the morphology of the sample and Energy Dispersive X-ray Spectroscopy (EDX) is performed to examine the composition.**

Index Terms—**nanotube; nanosphere; ferrocene; ammonium chloride.**

I. Introduction

The discovery of fullerenes in the mid-1980s has led to an upsurge in interest in the structures and shapes of carbon containing materials, in particular when the carbons are in the 'nano' regime. In the past two decades many differently shaped carbon materials (SCMs) have been synthesized and characterized and their properties in a range of different environments studied. These shaped materials include tubes and fibers, onions, spheres, horns, calabashes, flasks, etc [1]. Carbon nanotubes (CNTs) took a great interest among the carbon materials due to their various physical and chemical properties. CNTs were discovered by Iijima [2] in 1991. CNTs can be viewed as sheets of graphite rolled into seamless cylinders.Carbon Nanotubes can be categorized by their bonding i.e. c-CNTs and α-CNTs. CNTs can be divided also based on number of walls i.e. Single-wall nanotubes (SWNTs) and Multi-wall Nanotubes (MWNTs). The multi walled carbon nanotubes can be considered as an assembly of concentric SWNTs of different diameters. Various methods were developed such as arc discharge method, chemical vapor deposition method, laser vaporization method, pyrolysis, plasma-enhanced deposition method, template assisted method, solvothermal method, etc. High synthesis temperature and pressure, complicated processing steps, catalyst supports, longer synthesis period or expensive costs are the major drawbacks from the above-mentioned techniques [3]. Defects in CNTs give outstanding electrical properties which are used in field-emission display devices. Rakitin et al. [4] investigated both theoretically and experimentally the nanotubes comprised mainly of disordered graphene and found that this type of nanotube displayed a semiconductor band gap that is inversely proportional to the diameter.

Carbon nanospheres (CSs) have inspired great interest due to their special structure among carbon materials. Spherically shaped carbon materials have been given many names. These include carbon balls, carbon nanospheres, carbon microbeads, carbon blacks, onions, mesoporous microbeads, etc [1].

The synthesis methods used to produce carbon spheres include chemical vapor deposition, the catalytic-reduction route, the sol-gel method, the solid template approach, and the solvothermal method [5]. For example, Bao et al. have used hexachlorobenzene and sodium to prepare porous carbon materials; after an annealing process in N_2 flowing atmospheres above 1400^{0}C, carbon spheres were finally obtained [6]. Liu et al. synthesized carbon spheres via the reactions between ferrocene and ammonium chloride in a graphite crucible at 700^{0}C under an Ar atmosphere of 2.8 MPa[7]. A co-pyrolysis method was also developed to prepare high yield and proportion of carbon spheres (composed of solid and hollow one) by pyrolyzing a C_2C_{l4}/Fe-$(C_2H_5)_2$ mixture at 550^{0}C [8].

In this paper, CNTs and carbon nanospheres are synthesized simultaneously via a simple chemical route i.e. reaction of ferrocene and ammonium chloride. Carbon nanotubes synthesis from this reaction were previously reported by many authors [3,10].The low temperature synthesis is capable of eliminating the potential disadvantages associated with the aforementioned growth techniques[3]. Characterization is done by Field emission scanning electron microscopy to examine the morphology of the sample. EDX is also done to analysis the composition of the CNTs and carbon nanospheres.

II. Experimental Procedure

A. Mixing

1g ferrocene and .5g ammonium chloride was taken in a mortar and mixed thoroughly for 30 min. Mixture was taken in alumina boat and covered using alumina foil to prevent the sublimation of ferrocene.

B. Heating

Temperature was set to 250 ^{0}C in air furnace. After attaining the equilibrium temperature the boat was inserted carefully into the furnace and heated for 30 min. After 30 min of heating sample was removed and allowed to cool naturally.

C. Remixing

The content of sample was taken in a mortar again and mixed homogenously for 30 min .

D. Reheating

The mixed sample was heated again in air furnace for 30 min. After 30 min heating sample was removed for cooling.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

E. Purification

After second time natural cooling black powder was washed consecutively by dilute HCl and distilled water to remove the presence of trace iron .Sample was filtered and dried for 24 hr at 80c.

F. Characterization

Then sample was taken for Field emission scanning electron microscopy (FESEM) and EDX for characterization.

III. Results And Discussion

Field emission scanning electron microscopy (FESEM) was employed to observe the structure and morphology of the samples. The FESEM was operated at an accelerating voltage of 15 kV and 5 KV, with a working distance of 8.5 mm and 8.1 mm.

Fig. 1,2 presents both low and high magnification FESEM images of nanotubes after the treatment with hydrochloric acid. Nanotubes are agglomerated heterogeneously and different sizes of Nano-tubes were produced.

The full understanding of the growth mechanism of this kind of structure is still under investigation. It is interesting to mention that Liu et al. [9,11] reported the formation of carbon nanotubes following similar kind of experiments. According to them, ferrocene was first oxidized to ferrocenium cation which in turn acted as electron acceptor and ultimately produced $(C_5H_5)_2FeCl$, which further oxidized by dehydrogenation and reacted with ammonium chloride. The following chemical reaction equations have been proposed to explain how the CNTs were synthesized

$$4(C_5H_5)_2Fe + O_2 + 4NH_4Cl \rightarrow 4\,(C_5H_5)_2FeCl + 4NH_3 + 2H_2O \quad (1)$$

$$(C_5H_5)_2FeCl + 3O_2 + 2NH_4Cl \rightarrow 10C + FeCl_3 + 2NH_3 + 6H_2O \quad (2)$$

During this process pentagon unit of carbon in ferrocene constitute the tube structure. Also the diameter determining factor is not known, as according to them the iron particles do not act as a catalyst but only act a reducing agent. Obviously, further research is necessary for clearly understanding the growth mechanism [10, 11].

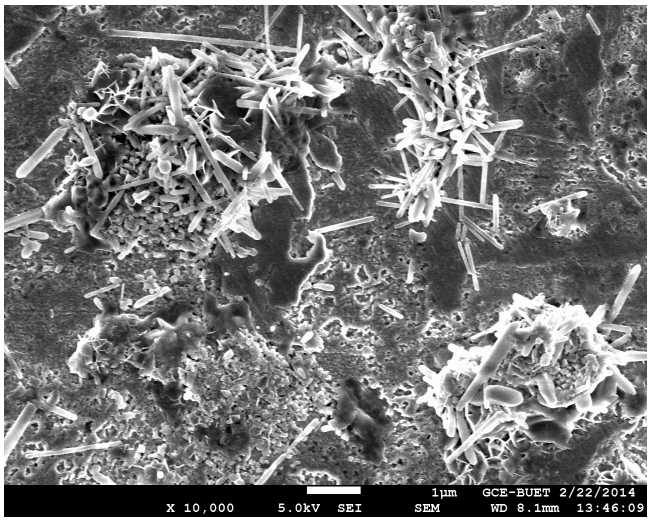


Fig. 1 FESEM of the obtained CNTs after 1g ferrocene and .5g ammonium chloride reaction at 250^0C (10000X)

Fig. 2 FESEM of the obtained CNTs after 1g ferrocene and .5g ammonium chloride reaction at 250^0C(20000X).

Fig. 3 represents the carbon nanospheres. Spheres are dispersed randomly in the sample .They are in different sizes and irregular in shape. It is observed in fig. 4 that nano spheres are growing and some nanofibres as well.

Liu at al. [7] reported the formation mechanism of carbon nanospheres. According to this author the reaction rate between ferrocene and ammonium chloride is slow for the formation of carbon spheres.

The chemical reactions took place in this temperature range are proposed as follows

$$(C_5H_5)_2Fe \rightarrow C + Fe + H_2 + CH_4 + C_5H_6 + \text{other hydrocarbon} \quad (3)$$

$$Fe + 2NH_4Cl \rightarrow Fe(NH_3)_2Cl_2 + 2H_2 \quad (4)$$

$$Fe(NH_3)_2Cl_2 + 3NH_4Cl \leftrightarrow (NH_4)_3FeCl_5 + 3NH_3 \quad (5)$$

$$Fe(NH_3)_2Cl_2 + 4NH_3 \leftrightarrow Fe(NH_3)_6Cl_2 \quad (6)$$

ferrocene will decompose to form carbon and iron which will react with ammonium chloride to form $Fe(NH_3)_2Cl_2$ .Thus the reaction rate should be dominated by the decomposition of ferrocene. In this experiment, as the decomposition of ferrocene is a slow process, excessive NH_4Cl will further react

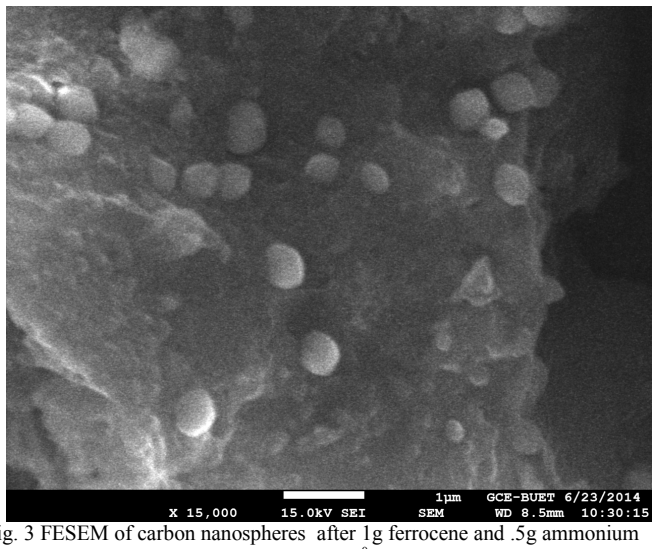


Fig. 3 FESEM of carbon nanospheres after 1g ferrocene and .5g ammonium chloride reaction at 250^0C(15000X)

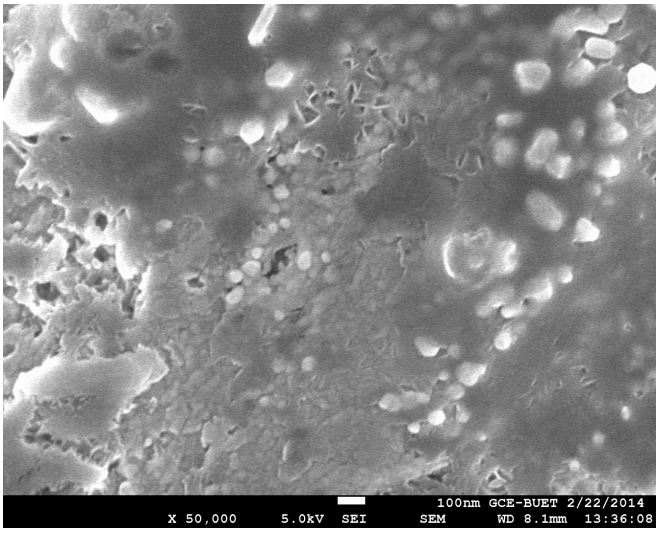


Fig. 4 FESEM of the growing process of carbon fibres and carbon spheres simultaneously after 1g ferrocene and .5g ammonium chloride reaction at 250^0C(50000X)

with $Fe(NH_3)_2Cl_2$ to form $(NH_4)_3FeCl_5$ and ammonia. Moreover, the atmosphere in the boat is a quasi-static atmosphere, so the ammonia generated by Eq. (5) could not quickly diffuse out of the alumina boat to the muffle furnace. Therefore, the generated ammonia will simultaneously react with $Fe(NH_3)_2Cl_2$ to form $Fe(NH_3)_6Cl_2$.

The EDX results of CNTs and carbon nanospheres are shown in fig. 5, 6. The peak represents the higher percentage of carbon but some N, Cl are present also in CNTs and carbon spheres. N and Cl came from ammonium chloride .Some traces of iron came from ferrocene. Oxygen in carbon nanospheres came from air due to the long time exposure in air.

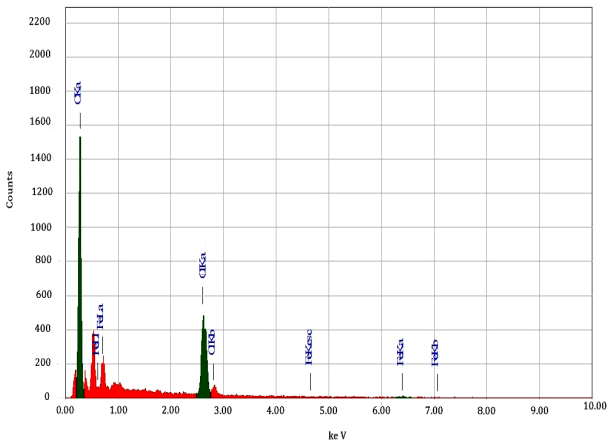


Fig. 5 EDX spectra for CNTs

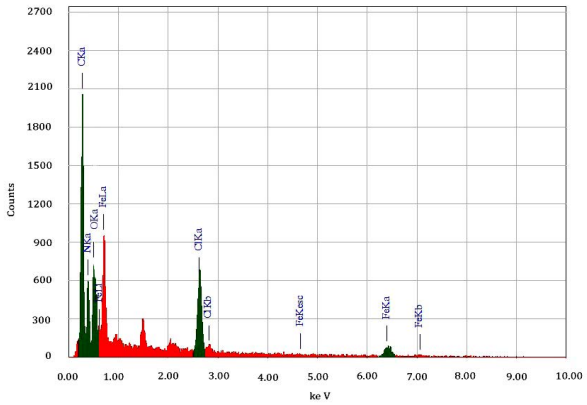


Fig. 6 EDX spectra for carbon nanospheres

IV. Conclusions

CNTs and carbon nanospheres were produced successfully after the reaction of 1g ferrocene and .5g ammonium chloride at 250^0C. FESEM and EDX confirm that the produced samples are CNTs and carbon nanospheres. Yield of CNTs were small in this procedure. Details of controllable preparation and mechanism studies are in progress for the large scale production of carbon nanotubes. This chemical synthesis for carbon spheres is very cost effective and simple for mass production.

Acknowledgment

The authors would like to thank the Glass and Ceramics Engineering department (GCE), BUET for giving the opportunity to perform FESEM and EDX analysis.

References

[1] Amit A. Deshmukh, Sabelo D. Mhlanga and Neil J. Coville, "Carbon spheres," *Materials Science and Engineering R*,vol. 70, pp. 1–28, 2010.

[2] Sumio Iijima, "Helical microtubules of graphitic carbon," *Nature,* vol. 354, pp. 56 - 58, Nov. 1991.

[3] Kim Han Tan , Roslina Ahmad, Bey Fen Leo, Ming Chian Yew, Bee Chin Ang and Mohd Rafie Johan, "Physico-chemical studies of amorphous carbon nanotubes synthesized at low temperature," *Materials Research Bulletin,*vol.47, pp.1849 − 1854,2012.

[4] A. Rakitin, C. Papadopoulos, and J. M. Xu, "Electronic properties of amorphous carbon nanotubes," *Phys. Rev. B* ,vol. 61, pp. 5793, Feb. 2000.

[5] Guangda Li, Chunli Guo, Changhui Sun, Zhicheng Ju, Lishan Yang, Liqiang Xu and Yitai Qian, "A Facile Approach for the Synthesis of Uniform Hollow Carbon Nanospheres," *J. Phys. Chem. C*,vol. 112, pp. 1896-1900, 2008.

[6] Gang Hu,a Ding Ma,ab Mojie Cheng,a Lin Liua and Xinhe Bao, "Direct synthesis of uniform hollow carbon spheres by a self-assembly template approach," *Chem. Commun*., pp. 1948-1949, Nov. 2005.

[7] Liu B. Y., Jia D. H., Meng Q. C. and Rao, J. C., "A novel method for preparation of hollow carbon spheres under a gas pressure atmosphere ," *Carbon*,vol. 45, pp. 668−670, 2007.

[8] S. Zhang, C. Zhu, J. K. O. Sin and P. K. T. Mok, "A novel route to hollow and solid carbon spheres," *Carbon*,vol. 43, pp. 1090–1092, 2005.

[9] Boyang Liu, Dechang Jia, Yu Zhou, Haibo Feng, Qingchang Meng, "Low temperature synthesis of amorphous carbon nanotubes in air," *Carbon*,vol. 43, pp. 1090–1092, 2005.

[10] D. Banerjee a, A.Jha a and K.K.Chattopadhyay"Low temperature synthesis of amorphous carbon nanotubes in air," *Carbon*,vol. 43, pp. 1090–1092, 2005.

[11] Boyang Liu, Dechang Jia , Yu Zhou, Haibo Feng and Qingchang Meng, "Low temperature synthesis of amorphous carbon nanotubes in air ," *Carbon*,vol. 45, pp. 1696–1716, 2007.

Characterization of Electrodeposited Multiwall Carbon Nanotube Films on Silicon Substrates

Mohammad Wahidur Rahman[1*], Muhammad Athar Uddin[1], Ahmedullah Aziz[1], Tanvir Mustofa[1], Zahid Hasan Mahmood[2], T. Soga[3] and Sharif M Mominuzzaman[1]
[1]Department of Electrical & Electronic Engineering, Bangladesh University of Engineering & Technology, Dhaka-1000, Bangladesh
[2]Department of Applied Physics, Electronics and Communication Engineering, University of Dhaka, Dhaka-1000, Bangladesh
[3]Department of Environmental Technology and Urbane Planning, Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya, 466-8555, Japan
*kawsar.eee.buet@gmail.com

Abstract—**Multiwall Carbon Nanotube (MWCNT) films were deposited on Silicon (Si) substrate. The effect of 0.1% and 0.2% solutions of MWCNT incorporation with methanol were investigated. pH of various solutions before and after deposition has been analyzed. pH of the solutions was increased after deposition. Change in the variation of current density as a function of applied potential was observed. Current density was increased as the percentage of MWCNT was increased. The electrodeposited MWCNT films were characterized by SEM, UV-VIS-IR and Raman spectroscopy. In SEM micrographs, changes were observed in the electrodeposited films compared to pure Si substrates. Raman measurement proved the amorphous nature of MWCNT in these thin films. From the UV-VIS-IR measurement it was observed that the absorption of light has been increased by the increase of MWCNT. So the optoelectronic properties of the electrodeposited films are tunable by electrodeposition using MWCNT on Si substrates.**

***Keywords*— Thin Film, Electrodeposition, Multiwall Carbon Nanotube.**

I. Introduction

There have recently been important advances in the science of crystalline carbon, once considered exotic— fullerenes are now commonly synthesized and used in research. They include buckyballs (C_{60}), single-walled carbon nanotubes (SWCNT), multi-walled carbon nanotubes (MWCNT), carbon nanofibers etc. Carbon nanotubes exhibit remarkable mechanical and electronic properties and are, therefore, being regarded as a new functional material for next generation electronics. Nevertheless, several obstacles still exist for an application in industry. The control of carriers in carbon nanotubes is of critical importance prior to an industrial application in transistors [1].

Within materials science, the optical properties of carbon nanotubes refer specifically to the absorption, photoluminescence, and Raman spectroscopy of carbon nanotubes. Spectroscopic methods offer the possibility of quick and non-destructive characterization of relatively large amounts of carbon nanotubes. There is a strong demand for such characterization from the industrial point of view: numerous parameters of the nanotube synthesis can be changed, intentionally or unintentionally, to alter the nanotube quality. Raman spectroscopy and optical absorption photoluminescence allow quick and reliable characterization of this "nanotube quality" in terms of non-tubular carbon content, structure (chirality) of the produced nanotubes, and structural defects. Those features determine nearly any other properties such as optical, mechanical, and electrical properties.

Carbon nanotubes are unique "one dimensional systems" which can be envisioned as rolled single sheets of graphite (or more precisely graphene). This rolling can be done at different angles and curvatures resulting in different nanotube properties. The diameter typically varies in the range 0.4–40 nm (i.e. "only" ~100 times), but the length can vary ~10,000 times reaching 18.5 cm. Thus the nanotube aspect ratio, or the length-to-diameter ratio, can be as high as 132,000,000:1, [2] which is unequalled by any other material. Consequently, all the properties of the carbon nanotubes relative to those of typical semiconductors are extremely anisotropic (directionally dependent) and tunable.

Whereas mechanical, electrical and electrochemical (supercapacitor) properties of the carbon nanotubes are well established and have immediate applications, the practical use of optical properties is yet unclear. The tunability of properties is potentially useful in optics and photonics. In particular, light-emitting diodes (LEDs) [3]- [4] and photo-detectors [5] based on a single nanotube have been produced in the lab. Their unique feature is not the efficiency, which is yet relatively low, but the narrow selectivity in the wavelength of emission and detection of light and the possibility of its fine tuning through the nanotube structure. In addition, bolometer [6] and optoelectronic memory [7] devices have been realized on ensembles of single-walled carbon nanotubes.

Precursors and method of deposition of carbon films are the two dominating factors that dictate the optical and electrical properties of the film. Therefore, researchers on finding alternative precursor materials and simple method of deposition have been getting priority all the time. In connection with this research, MWCNT has been used as an alternative precursor material.

There is experimental evidence that most materials that can be deposited from the vapour phase can also be deposited in the liquid phase using electroplating techniques and vice versa. Enlightened by this conclusion, Namba [8] first attempted to grow diamond phase carbon films in the liquid phase with the aid of an organic solution such as ethanol at a temperature less than 70°C. DLC films had been obtained there, but the choice of a suitable solution was limited. Then Suzuki et al. [9] made an attempt to deposit carbon films by electrolysis of a water-ethylene glycol solution. Nevertheless, ethylene glycol is a viscous solution, which will cause some difficulty in cleaning the substrate after deposition. Hao Wang

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

et al. [10] deposited film by using methanol solution. Methanol is selected because its polarizability and conductivity are stronger than those of ethanol and the structure of methanol is even closer to that of diamond. Mominuzaman *et al.* [11] reported semiconducting carbon films by pulsed laser ablation obtained from camphor ($C_{10}H_{16}O$), a natural source. In connection to this few electrodeposion works have been reported [12]-[15] where carbonic thin films were deposited on silicon (Si), cupper (Cu) and aluminium (Al) substrates by electrolysis of methanol with camphor incorporation. Teresa *et al.* deposited [16] CNTs on Si substrates by CVD where CNT growth rate was possible even at low pressure. Tomoya *et al.* electrodeposited [17] carbon nanotubes on Zinc Oxide where CNTs were embedded on ZnO films.

In the present article considering all these continuations from electrodeposition in methanol, deposition of CNT on Si substrate by CVD and electroplating of CNTs on ZnO an attempt has been made for electrodeposition of MWCNT on Si substrate. Varying concentration of MWCNT materials were mixed to prepare carbon nanotube-methanol solution. As per our knowledge this type of work has been reported first time.

II. Experimental Details

For the deposition of MWCNT films on Si substrate an experimental setup was prepared. Si substrate was mounted on the negative electrode. Prior to deposition the substrates were cleaned in electronic grade acetone (CH_3COCH_3) and methanol (CH_3OH) for 5 minutes at 55°C. The distance between the substrate and positive electrode was set to 1 cm. The pH measurements were done before and after deposition. MWCNT with varying amount (0.1% and 0.2%) was mixed in methanol solvent to prepare the electrolytes. Variations of current densities as a function of applied potentials were compared.

III. Results and Discussions

The experiment was performed for 0.1% and 0.2% of multiwall carbon nanotubes (MWCNT) in methanol solution. pH was measured before and after deposition. There was a sharp change in the pH which indicates variation of ionic concentration in the solution and reveals possibility of electrodeposition. Table-1 shows the changes in pH due to film deposition for the incorporation of MWCNT.

TABLE I
Change in pH Due to Film Deposition

Carbon Materials	pH Before Deposition	pH After Deposition
0.1% MWCNT	2.3	4.7
0.2% MWCNT	2.3	4.4

Current densities as a function of applied voltage for different percentage of MWCNT in methanol solution were compared. Higher current density indicates more ionized charge particles to move from solution to electrode, which indicates the effect on the growth rate of film. Current densities were measured in the voltage range of 100V~600V.

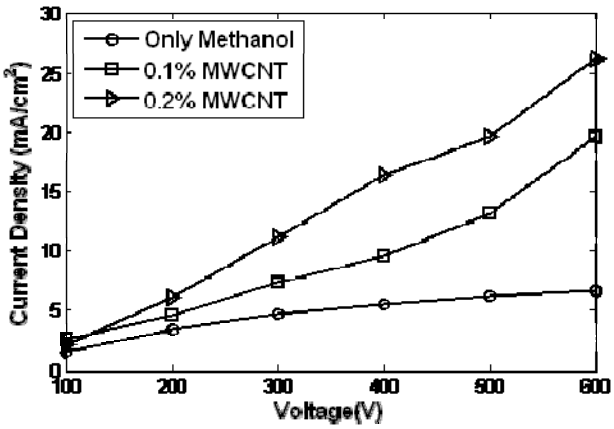


Fig. 1 Comparison of current densities as a function of applied potential for different percentage of MWCNT with methanol.

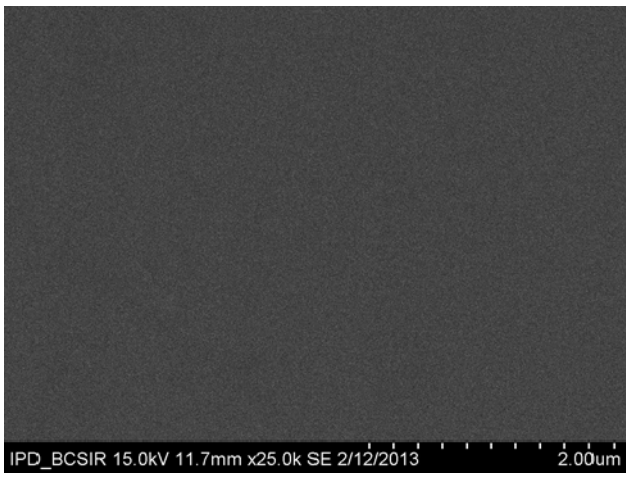


Fig. 2 Pure Si substrate observed in SEM.

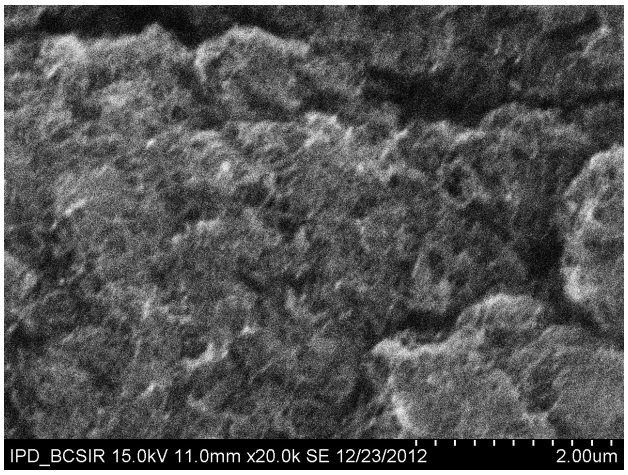


Fig. 3 Nanotube films deposited on Si using 0.1% MWCNT observed in SEM.

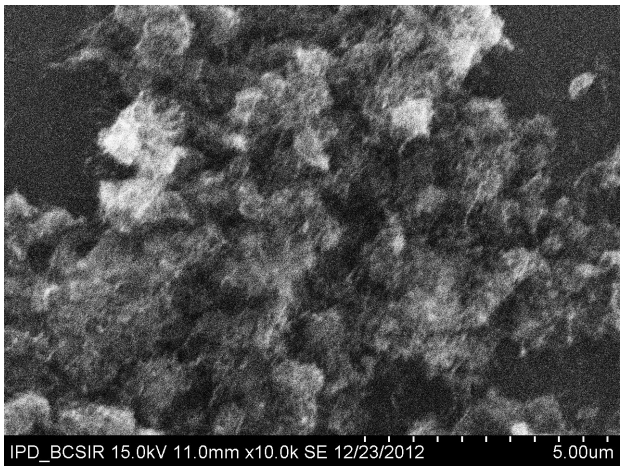


Fig. 4 Nanotube films deposited on Si using 0.2% MWCNT observed in SEM.

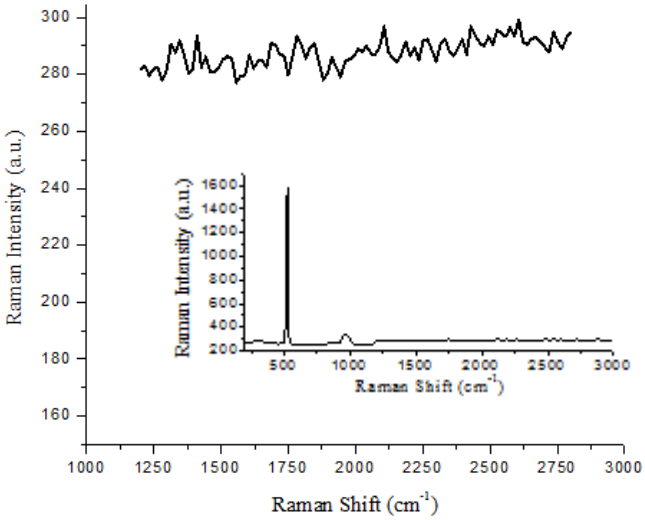


Fig. 5 Raman Spectra of Pure Si substrate

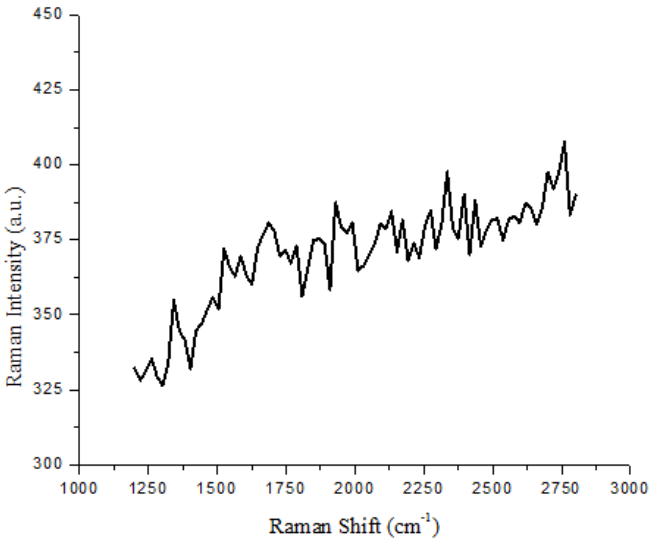


Fig. 6 Raman Spectra of Nanotube films deposited on Si using 0.1% MWCNT

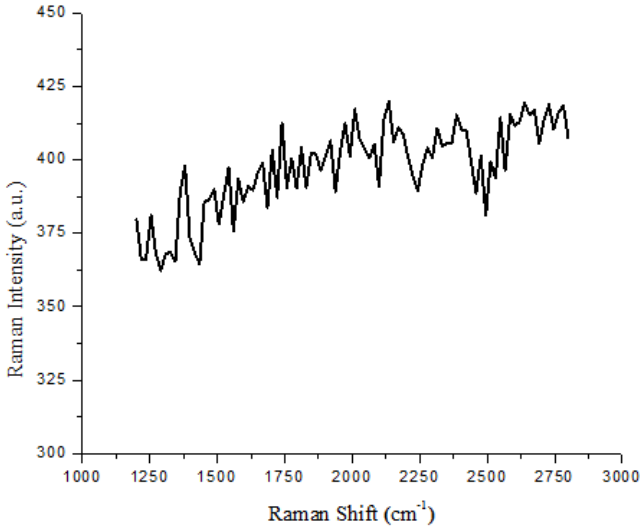


Fig. 7 Raman Spectra of Nanotube films deposited on Si using 0.2% MWCNT

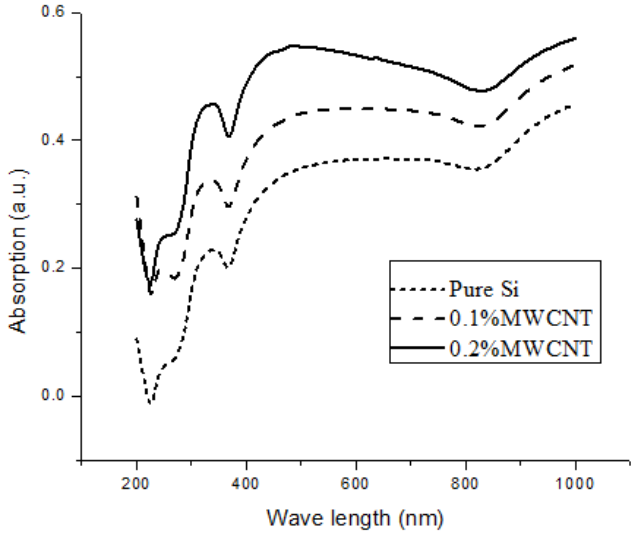


Fig. 8 Absorption of light as a function of wavelength for MWCNT films electrodeposited on silicon substrate.

Fig. 1 show the comparison of current densities of only methanol solution and that of 0.1% and 0.2% MWCNT in methanol solutions as a function of applied potential in the voltage range of 100V~600V. Fig. 2, 3 and 4 show the SEM micrographs of pure Si substrate, 0.1% MWCNT electrodeposited film and 0.2% MWCNT electrodeposited film respectively.

Raman measurements were performed using a He-Ne laser operating at a wavelength of 632.8nm (1.96 eV) and input power of ~300μW at the sample surface. A confocal microscope experimental setup [18] was used for Raman signal acquisition. Fig. 5, 6 and 7 show the Raman Spectra of pure Si substrate, 0.1% MWCNT electrodeposited film and 0.2% MWCNT electrodeposited film respectively.

Fig. 8 shows the comparative study of absorption of light as a function of wavelength in UV-VIS-IR region for pure Si substrate, 0.1% and 0.2% MWCNT electrodeposited films.

Current densities for MWCNT were higher than that of only methanol solution. Also the current densities were increased by the increase of MWCNT (from 0.1% to 0.2%).

Change in the variation of current density as a function of applied potential was observed with different percentage of carbon nano-materials. For example at 600V, the current densities for only methanol, 0.1% MWCNT and 0.2% MWCNT were 7mA/cm^2, 20mA/cm^2 and 27mA/cm^2 respectively. So higher conductivities were observed for MWCNT. It is possible to grow MWCNT films by electrodeposition at lower voltages compared to carbonic thin films with methanol [13].

In SEM images, changes were observed in the electrodeposited films compared to as-obtained Si substrates. The distinction could be ascertained by 0.1% and 0.2% MWCNT films. These images shows that MWCNT and amorphous carbon are exist in combined form.

In Raman spectroscopic images, only 1200 cm^{-1} to 2800 cm^{-1} Raman shift was considered as the peak at 300 cm^{-1} corresponds to the radial breathing mode (RBM) which is unique to CNTs, is not seen in case of MWCNT [19]. But there will be peaks explicitly appears at ~1360 cm^{-1} and ~1580 cm^{-1} for D band and G band respectively for MWCNT [20]. These peaks are not prominent as compared with SWCNT [19]. Furthermore MWCNT is combined with amorphous carbon (Fig. 3, 4) which makes the peak more broaden. So no prominent peaks for the samples were found.

From the comparative study of the spectrum of light absorption as a function of wavelength in UV-VIS-IR region in Fig. 8 it was observed that, absorption of light increased by the increase of MWCNT (from 0.1% to 0.2%). So the optoelectronic properties of the electrodeposited films are tunable by electrodeposition using MWCNT on Si substrates.

Existence of MWCNT combined with amorphous carbon in thin film certainly increases the electrical conductivity and optical absorption in comparison with only amorphous carbon based thin film [13], [15].

IV. Conclusions

MWCNT films were deposited on Si substrates. Effect of variation of MWCNT in methanol was observed. pH of the solutions was increased after deposition, thereby indicated variation of ionic concentration as well as formation of MWCNT on Si substrates. Current densities for MWCNT were higher with respect to methanol and increased as the percentage of MWCNT was increased. SEM micrographs shows the existence of MWCNT with amorphous carbon. Raman spectroscopy proved the amorphous nature of thin films. Absorption of light in UV-VIS-IR spectroscopy were increased by the increase of MWCNT. So the electrical and optoelectronic properties of the electrodeposited films are tunable by electrodeposition using MWCNT on Si substrates.

V. Acknowledgements

The authors would like to thank Dr. Abdul Gafur, for providing UV-VIS-IR spectroscopy and SEM facility at BCSIR and also thank to Professor Dr. B M Arora for Raman spectroscopy facility under Indian Nanoelectronics Users Program (INUP) at Indian Institute of Technology Bombay (IITB).

.

References

[1] Woo Jong Yu, Young Hee Lee, "Strategy for Carrier Control in Carbon Nanotube Transistors," *ChemSusChem,* vol. 4, no. 7, pp. 890-904, 2011.

[2] Xueshen Wang et al., "Fabrication of Ultralong and Electrically Uniform Single-Walled Carbon Nanotubes on Clean Substrates," *Nano Letters*, vol. 9, issue 9, pp. 3137-3141, 2009.

[3] J. A. Misewich et al., "Electrically Induced Optical Emission from a Carbon Nanotube FET," *Science* vol. 300, issue 5620, pp. 783-786, 2003.

[4] J. Chen et al., "Bright Infrared Emission from Electrically Induced Excitons in Carbon Nanotubes," *Science*, vol. 310, issue 5751, pp. 1171-1174, 2005.

[5] M. Freitag et al., "Photoconductivity of Single Carbon Nanotubes," *Nano Letters*, vol. 3, issue 8, pp. 1067–1071, 2003.

[6] M. E. Itkis et al., "Bolometric Infrared Photoresponse of Suspended Single-Walled Carbon Nanotube Films," *Science*, vol. 312, issue 5772, pp. 413-416, 2006.

[7] A. Star et al., "Nanotube Optoelectronic Memory Devices," *Nano Letters*, vol. 4, issue 9, pp. 1587-1591, 2004.

[8] Y. Namba, "Attempt to grow diamond phase carbon films from an organic solution," *J. Vac. Sci. Technol. A*, vol. 10, pp. 3368-3370, 1992.

[9] T. Suzuki, Y. Marita, T. Yamazaki, S. Wada, and T. Noma, "Deposition of carbon films by electrolysis of a water-ethylene glycol solution," *J. Mater. Sci.,* vol. 30, pp. 2067-2069, 1995.

[10] Hao wang, Ming-Rong Shen, Zhao-Yuan Ning, and Chao Ye., "Deposition of diamond-like carbon films by electrolysis of methanol solution," *Appl. Phys. Lett.,* vol. 69, pp. 1074-1076, 1996.

[11] S. M. Mominuzzaman, T. Soga, T. Jimbo, and M. Umeno, "Camphoric carbon soot: a new target for deposition of diamond-like carbon films by pulsed laser ablation," *Thin Solid Films*, vol. 376, pp. 1-4, 2000.

[12] Md. Mahmud Hasan, Muhammad Athar Uddin, S. M. Mominuzzaman, "An Experimental Approach of DLC Film Deposition on Metal Substrates," *5th International Conference on Electrical and Computer Engineering ICECE 2008,* pp. 803-806, 20-22 December 2008, Dhaka, Bangladesh.

[13] Muhammad Athar Uddin, S. M. Mominuzzaman, "Semiconducting Carbon Thin Film Deposition on Silicon by Electroplating," *5th International Conference on Electrical and Computer Engineering ICECE 2008*, pp. 846-850, 20-22 December 2008, Dhaka, Bangladesh.

[14] Muhammad Athar Uddin, Md. Shamimul Haque Choudhury and Sharif M Mominuzzaman, "Characterization of Carbon Thin Films Electrodeposited on Aluminum, Copper and Silicon Substrates," *6th International Conference on Electrical and Computer Engineering ICECE 2010*, pp. 206-209, 18-20 December 2010, Dhaka, Bangladesh.

[15] Muhammad Athar Uddin, M. Shamimul Haque Choudhury and Sharif M Mominuzzaman, "Optical properties of electrodeposited carbon thin films using camphor," *Bangladesh Journal of Physics*, vol. 9 & 10, pp. 9-19, 2011

[16] Teresa de les Arcos, Peter Oelhafen and Daniel Mathys, "The importance of catalyst oxidation for the growth of carbon nanotubes on Si substrates", *J. of Carbon*, vol. 47, issue 8, pp. 77-82, 2009.

[17] Tomoya Kuwajima, Yusuke Nakanishi, "Fabrication of Carbon Nanotube/Zinc Oxide Composite Films by Electrodeposition", *Jpn. J. Appl. Phys.*, vol. 50, pp. 085504-085506, 2011.

[18] Ying Geng, David Fang, and Lei Sun. (2006, Dec. 18). *Determining Carbon Nanotube Properties from Raman Scattering Measurements* [online]. Available: http:/ www.photonics.ethz.ch/fileadmin/user_upload/optics/Courses/Nano Optics/ProjectReports/swntRaman.pdf.

[19] S. Costa, E. Borowiak-Palen, M. Kruszynska, A. Bachmatiuk, and R. J. Kalenczuk. "Characterization of carbon nanotubes by Raman spectroscopy." *Mater Sci-Poland,* vol. 26(2), pp. 433-441, 2008.

[20] M. Zdrojek, W. Gebicki, C. Jastrzebski, T. Melin, and A. Huczko. "Studies of multiwall carbon nanotubes using Raman spectroscopy and atomic force microscopy," *Solid State Phenomena,* vol. 99 pp. 265-268, 2004.

Humidity Sensor Using Surface Adsorbed Channel Modulated Graphene Nanoribbon: NEGF Approach

S. R. Shakil[1*], F. T. Zohra[1], P. Pramanik[1], R. I. Tushar[1], A. K. Saha[1], M. B. H. Bhuian[1]

[1]School of Engineering and Computer Science, BRAC University, Bangladesh

[*]shifur.shakil@gmail.com

Abstract— **Electronic transport properties of Graphene nanoribbon (GNR) are explored by self-consistent numerical simulation. The effects of vapour (H_2O) adsorption in GNR sheets on its current-voltage (I-V) characteristics, conductivity, electrostatic difference potential and device density of states were studied with necessary physical insight. We found that conductivity of GNR sheet increases in some particular energy range with the increase in number of H_2O molecules adsorbed. Then we encounter the effects of metal-GNR contact to investigate the practical performance of GNR as a medium of humidity sensing. Later we propose a simple schematic model for GNR Humidity Sensor (GNRHS) to investigate its performance in applied regime.**

Index Terms— Humidity sensor; Graphene; NEGF; GNR; GNRHS.

I. Introduction

Flexible, highly sensitive and low cost humidity sensors are highly enviable in future generation sensor technology. As a result of recent year's research conducted concerning the applicability in gas and humidity sensors, graphene is reported to be an appropriate sensing material for this purpose [1-3]. The inimitable structural, mechanical and electronic properties of graphene has attracted extensive attention of scientists, therefore the successful synthesis of novel two-dimensional (2D) graphene and the experimental observation of Dirac fermions in unpatterned graphene devices has been increasing rapidly [4]. Gas and vapour adsorption on these devices could change their electronic properties which can be used as sensing parameter. The previous and recent researchers have concentrated to focus on the interaction of small gas molecule, such as CO, NO_2 and NH_3 with the pristine graphene in theory [5] and CO_2 sensing using graphene sheet in experiment [6].

The interaction between small molecules and graphene can be highly enhanced by introducing adsorption into 2D nanosheets of graphene. Again, it can be shown that water (H_2O) molecules adsorbed on graphene cause defects in it which facilitate to electron-tunnelling of the band gap and cause widening of graphene band gap to 0.206 eV [1]. However, the changes in resistance are directly influenced by the amount of water vapour adsorbed by the graphene [3].

Here we report a simulation based study of humidity sensing ability of patterned graphene adsorbing vapour (H_2O) and investigate electronic and quantum transport properties of these system such as Device Density of States (DDOS), Electrostatic Effective Potential (EDP), Conductivity (G) and Current-Voltage (I-V) characteristics. Transport simulation is based on Non-equilibrium Green's Function (NEGF) formalism. An unpatterned graphene sheet is a zero bandgap material with linear energy dispersion relation. Patterning graphene sheet can open up bandgap. In this paper we work on patterned graphene sheet called Graphene nanoribbon (GNR). We consider the edge atoms as hydrogen (H) passivated. Later we propose a simple schematic model to characterize device performance which incorporates the effect of metal contact resistance with GNR.

II. Device Geometry and Simulation

The electronic bandstructure of graphene was calculated very early on [7]. The valence electrons of carbon atoms in graphene are sp^2 hybridized with the remaining p_z carbon orbitals forming an extended π-electron system that is responsible for the low energy transport and optical properties of graphene. The bonding π-states form the valence band and the antibonding π*-states form the conduction band. These two bands touch at six points, the so called Dirac or neutrality points show zero bandgap semimetal characteristics. Whereas nano-patterning of graphene sheet can open up a bandgap upto ~1.4eV [8], depending on the chirality of nanoribbon it can be Armchair or Zigzag. Armchair GNR (A-GNR) and Zigzag GNR (Z-GNR) shows semiconducting or semimetallic behaviour depending on the number of carbon atoms, N along the width of the nanoribbon [9].

In general humidity sensing through GNR is a process of vapour (H_2O) adsorption in graphene surface where the interaction is limited to π-electrons of GNR and electrons of oxygen ion in H_2O. In this paper we investigate the change in electronic transport in GNR due to H_2O adsorption. Fig. 1 shows an A-GNR (N=10, E_g=~1.1eV) without H_2O adsorption (Fig. 1(a)) and with adsorption of four H_2O molecules (Fig. 1(b))

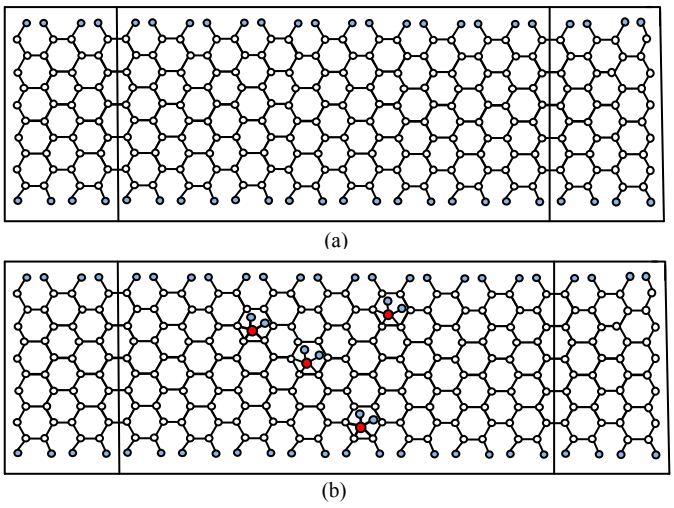


Fig. 1 A-GNR (N=10) surface (a) without H_2O adsorption (b) after H_2O adsorption. Middle portion is the sensing medium and side portions are extended electrodes.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

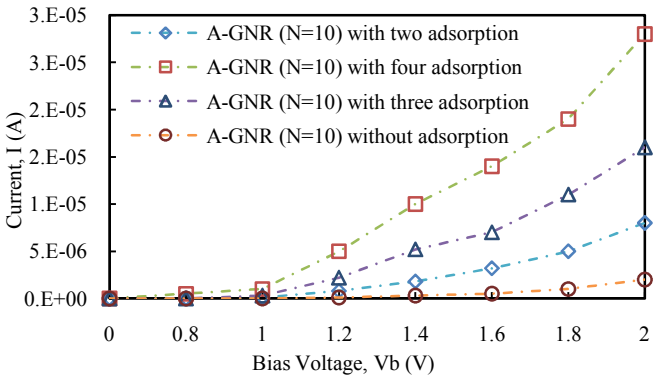


Fig. 2 I-V curves of semiconducting A-GNR (N=10) with and without H_2O adsorption.

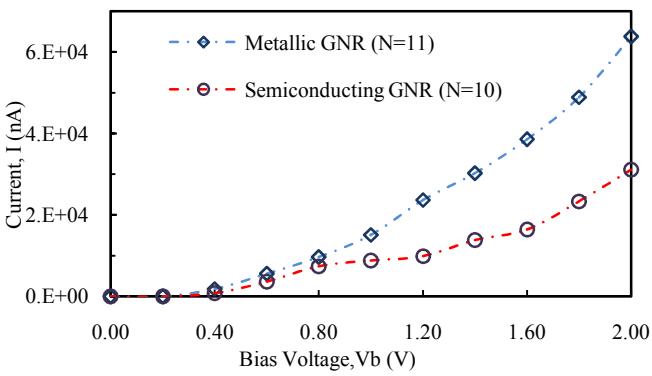


Fig. 3 I-V curves of metallic A-GNR (N=11) and semiconducting A-GNR (N=10) with metallic A-GNR electrodes with four H_2O molecules adsorbed.

The electronic transport calculations were performed by implementing non-equilibrium Green's function (NEGF) formalism [10, 11]. Ballistic current through GNR sheet considering self-consistent field imposed by H_2O molecules is calculated using Launder's equation.

$$I = \frac{2q^2}{h}\int_{-\infty}^{\infty} dE\, T(E)\,(f_L^+ - f_R^+)$$

Where, q=charge of electron and h=Plank's constant. f_L^+ is the Fermi function of left electrode and f_R^+ is the Fermi function for right electrode.

$$f_{L,R}^+ = \frac{1}{1 + e^{(E - E_F^{L,R})/K_B T}}$$

Where, K_B =Boltzmann's constant. $E_F^{L,R}$ is the electro-chemical potential at left or right contact and T(E) is the transmission spectrum.

$$T(E) = Tr[\Gamma^L G^R \Gamma^R G^A]$$

$$G^R(E) = [G^A(E)]^\dagger = [(E + i\delta_+)S - H - \Sigma^L(E) - \Sigma^R(E)]^{-1}$$

Here, $G^R(E)$ is Retarded Green's function where S and H are overlap integral matrix and device Hamiltonian matrix respectively. Self-energy terms are represented by $\Sigma^L(E)$ and $\Sigma^R(E)$ respectively for left electrode and right electrode. Here 'Tr' represents Trace of a matrix and $\Gamma^L = i[\Sigma_L - \Sigma_L^\dagger]$ and $\Gamma^R = i[\Sigma_R - \Sigma_R^\dagger]$. δ_+ is an infinitesimally positive number. Matrix S and H provides the atomistic description of GNR which is dependent on tight-binding P_Z orbital hopping parameter, previously calculated by Reich et al. [12]. In our simulation we have calculated Device Density of States (DDOS), conductance (G) and Electrostatic difference potential (EDP) using following equations.

$$DDOS(E) = \frac{1}{2\pi} Tr[G^R(\Gamma^L + \Gamma^R)G^A]$$

$$G(E) = \frac{2q^2}{h} Tr[\Gamma^L G^R \Gamma^R G^A]$$

$$EDP\left(\frac{r}{\max(r)}\right) = \frac{U(r) - \max(U(r))}{\max(U(r))}$$

Where, U(r) is the electrostatic potential calculated from Poisson's equation using finite difference algorithm and r represents two dimensional space coordinates (x, y, z). EDP is an unit less and normalized parameter.

III. Results and Discussion

At first, we have analysed homogenous semiconductive (N=10) A-GNR sheet as sensing medium. Then, we have analysed transport after the adsorption of two and four H_2O molecules on that GNR sheet. Fig. 2 shows the simulated I-V curves. Typically the GNR sheet doesn't show significant conduction at bias voltage of below 1V and then its conduction increase with the increment of bias voltage, Vb, whereas adsorption of H_2O increase the current more significantly depending on number of adsorbed H_2O molecule.

Later we have used two other GNR structures; homogeneous metallic A-GNR (N=11) and cascaded Hetero-GNR structure, where sensing medium is semiconducting A-GNR (N=10) which is connected to metallic A-GNR (N=11) electrodes. Fig. 3 shows I-V curves of these two types of structure. Here we found that the cascaded structure shows significant conduction even in lower bias voltage (Vb=0.4V) than a homogeneous semiconducting A-GNR (Vb=1V). Moreover, metallic A-GNR is less sensible in current increment adsorbing H_2O molecules at low bias voltage. To get the physical insight of such sensing phenomenon we have to investigate the change in device density of states (DDOS), conductance (G) and electrostatic difference potential (EDP) with the amount of H_2O adsorption.

Fig. 4 and Fig. 5 show DDOS and G respectively of a semiconducting A-GNR (N=10) for different H_2O adsorption conditions. Energy, E=0 signifies Fermi energy when GNR sheet is at equilibrium (without external bias voltage). For unadsorbed semiconducting A-GNR sheet bandgap region contains no energy states and so that conduction is zero within this energy range (Fig. 4(a) and Fig. 5(a)). Later due to adsorption of H_2O numbers of energy states are suppressed at higher energy level, whereas some electrons become localized due to heavy nuclei of oxygen atom from H_2O molecules. Because of this electron localization some small potential barriers are generated within the GNR sheet. These locally generated potential barriers can be seen from EDP. Fig. 6(a) shows the difference in sheet potential in three different directions of the A-GNR. Along the length of A-GNR or in transport direction potential is zero everywhere. With the adsorption of two H_2O molecules Fig. 6(b) shows two peaks in potential profile working as potential barriers for electrons. These potential barriers give rise to quasi quantized energy states at lower energy range which is evident from Fig. 4(b). These low energy states increase conductance of the GNR sheet at lower energy (Fig. 5(b)) which in turns enable significant current increment even in lower bias voltage, Vb.

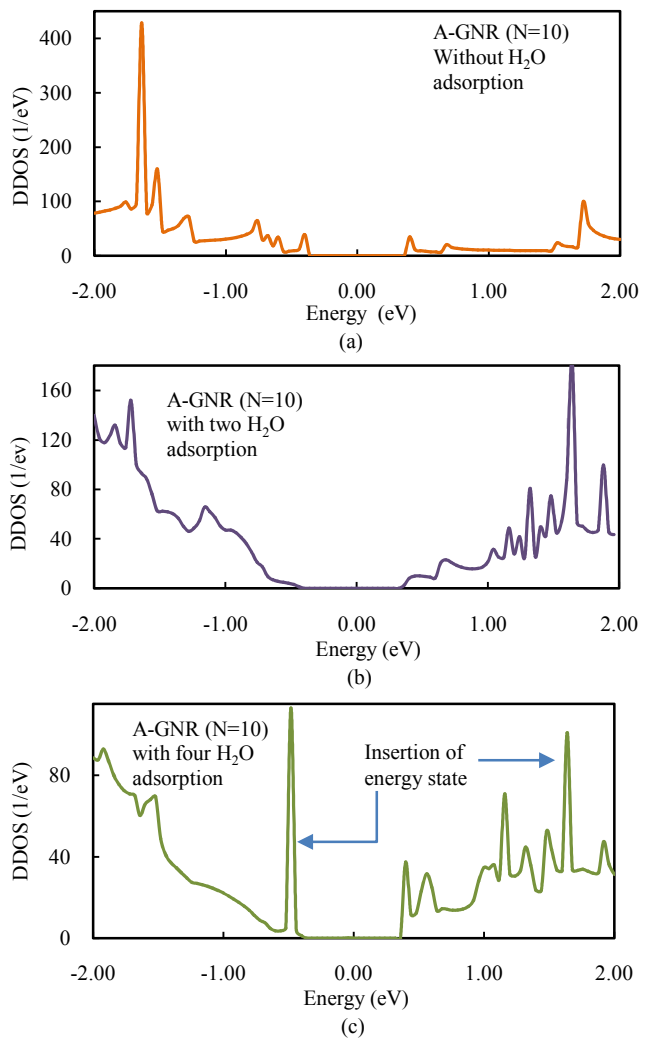


Fig. 4 Device Density of States (DDOS) (a) Without H_2O adsorption (b) Two H_2O adsorbed (c) Four H_2O adsorbed on GNR surface.

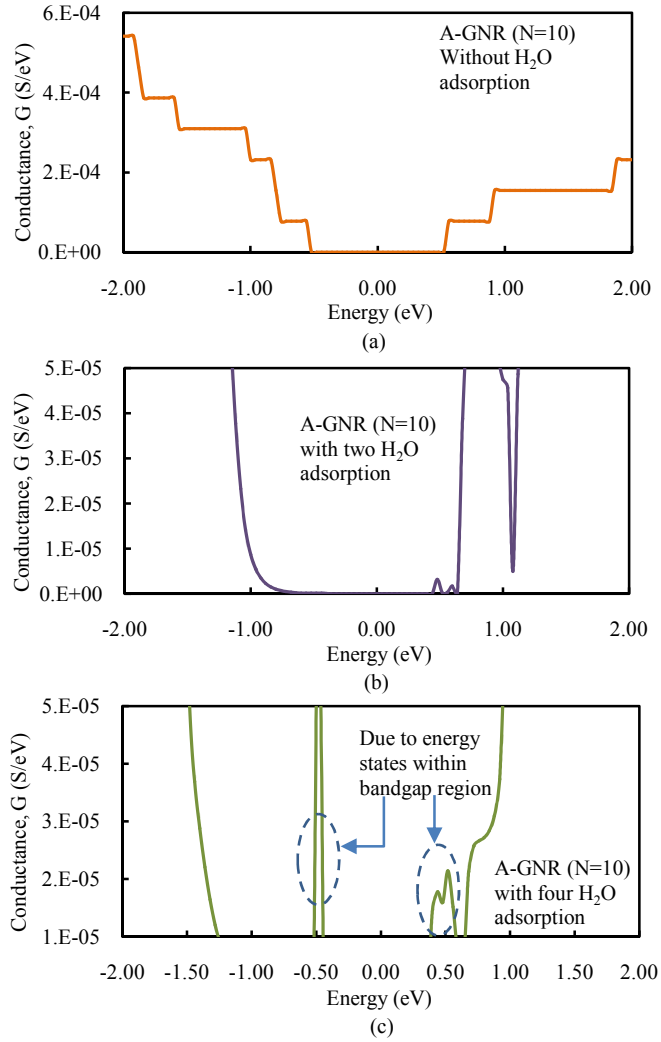


Fig. 5 Conductance of GNR sheet (a) Without H_2O adsorption (b) Two H_2O adsorbed (c) Four H_2O adsorbed.

With the increment of number of adsorbed H_2O molecules the number of potential barriers increased. At Fig. 6(c) there are four peaks in potential profile of GNR along the transmission direction because of the adsorption of four H_2O molecules. Because of such closely packed but relatively lower height of potential barriers the ground states of quantized energy states become lower and fall within the bandgap energy region, which enable electronic conduction within the bandgap energy region of the semiconducting A-GNR sheet shown at Fig. 4(c) and Fig. 5(c). For insertion of such inter-bandgap electronic conduction, current start to increase in GNR even at lower bias voltage, Vb.

On the other hand metallic A-GNR shows significant electronic conduction at very low energy because of the absence of bandgap, which enable current flow at very low bias voltage, Vb. So increment of current in metallic A-GNR is less significant than semiconducting A-GNR with adsorption of H_2O molecules.

Finally we can make a statement that the adsorption process actually has two impacts on conductivity of GNR sheet. Firstly, it lowers higher energy electronic conduction of the GNRs and secondly, it enables low energy conduction through quantized energy state (even through the bandgap region of the semiconducting A-GNR). So, for low voltage application semiconducting A-GNR will be a better choice over metallic A-GNR as the sensing medium though metallic GNR gives higher current but less sensitive to H_2O adsorption.

IV. Effect of Metal Contact

So far we have considered only the effects on electronic conduction of GNRs due to H_2O adsorption. Here we will discuss the effect of metal contact with the GNRs which is realistic to encounter the practical performance of such sensing device. The metal-GNR contact can be ohmic contact or schottky contact depending on the width of A-GNR and the difference in work functions, $\Delta\Phi$ of the metal and A-GNR. Recent works suggest that metal-GNR schottky barrier height (SBH) can be modulated through vertical potential which works as electrostatic doping in GNRs [13]. Their numerical simulation investigated that metal-GNR contact becomes ohmic with a steeper decrement in contact resistance, R_C with the decrement of GNR width, W_{GNR}. For semiconducting A-GNR of width, W_{GNR}=1.1nm (N=10) the value of normalized contact resistance, (R_C/ W_{GNR}) is ~11Ω-µm or a resistance (R_C) of ~10kΩ for $\Delta\Phi$=460meV, whereas the international roadmap for semiconductor calls for a normalized contact resistance of 80Ω-µm [14], which can be achieved by contacts to GNR materials under the ideal conditions with ballistic transport. For Gold (Au) to A-GNR contact ($\Delta\Phi$=~440meV) assumption of a contact resistance of 10kΩ is reasonably realistic in low voltage range. Fig. 7(a) shows the effect of metal contact on the sensing performance of GNR through its decrement in current due to some voltage drop across the contact to GNRs potential barrier.

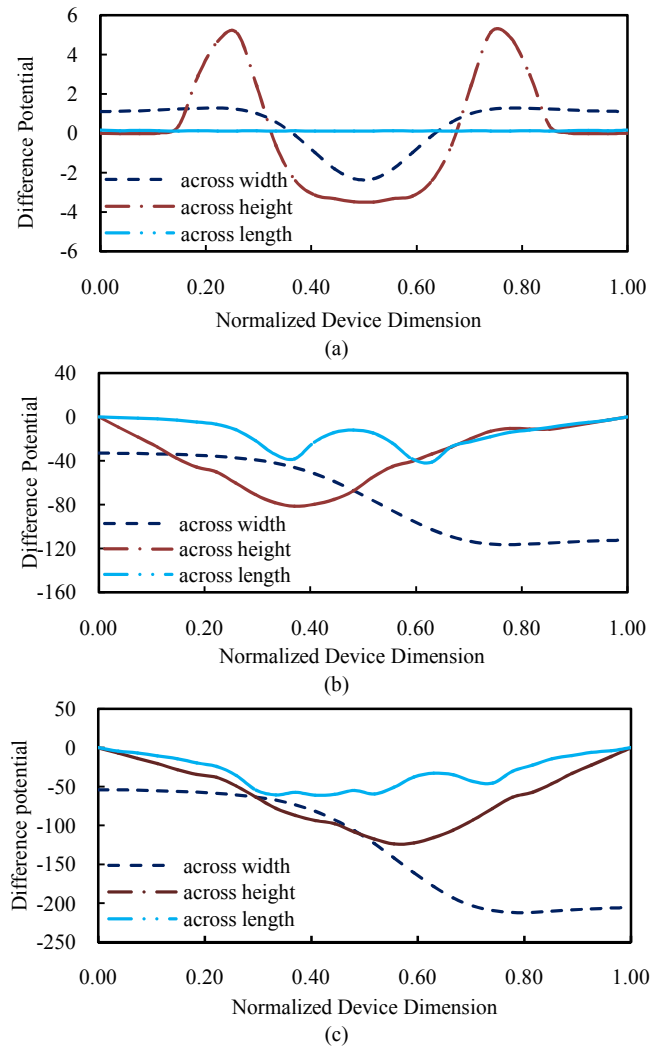


Fig. 6 Electrostatic difference potential of GNR sheet (a) Without H_2O adsorption (b) Two H_2O adsorbed (c) Four H_2O adsorbed.

Finally we have proposed a simple schematic model (Fig. 7(b)) for GNRHS approximated with linear operational range. Here the sensing medium is modelled by two parallel resistances named $R_{graphene}$ and R_{vapour}, where $R_{graphene}$ represent the GNR sheet without H_2O adsorption and R_{vapour} imposes the change in conductance due to H_2O adsorption. R_{vapour} decreases with the increment of number of adsorbed H_2O molecules.

$$I = \frac{V_b - 1}{\left[2R_c + \frac{R_{graphene}R_{vapour}}{R_{graphene}+R_{vapour}}\right]}, \quad R_{graphene} = 1M\Omega$$

Fig. 7(a) shows I-V curves for such schematic model which shows a good match with the I-V curves we get from NEGF simulation for different number of adsorbed H_2O.

V. Conclusions

In summary, we have investigated changes in electronic transport of A-GNR due to H_2O adsorption through self-consistent NEGF and electrostatics. Different transport properties show significant changes due to adsorption of H_2O on A-GNR surface, which become dominant in transport mechanism with increasing number of adsorbed H_2O molecules. Finally, we proposed a schematic model which incorporates this sensing phenomenon along with the effect of contact resistance. This work may provide useful guideline for future work in the field of GNR humidity sensor (GNRHS).

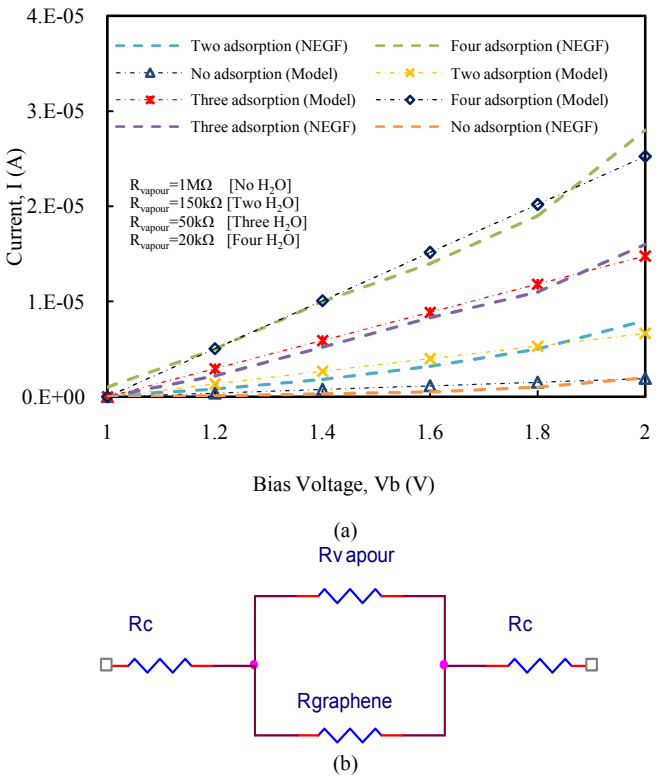


Fig. 7 (a) I-V curves for semiconducting A-GNR sheet with Gold (Au) contact for different number of adsorbed H_2O molecules (b) Schematic model of a GNR humidity sensor (GNRHS).

References

[1] J. Berashevich and T. Chakraborty, "Tunable bandgap and magnetic ordering adsorption of molecules on graphene," *Phys. Rev.* B, vol. 80, p. 033404, Jul. 2009.

[2] F. Yavari, C. Kritzinger, C. Gaire, L. Song, H. Gullapalli,T. Borca Tasciuc, P. M. Ajayan, and N. Koratkar, "Tunable Bandgap in Graphene by the Controlled Adsorption of Water Molecules," *Small*, vol. 6, no. 22, pp. 2535–2538, Nov. 2010.

[3] B. H. Chu, C. F. Lo, J. Nicolasi, C. Y. Chang, V. Chen, W. Strupinski, S. J. Pearton, F. Ren, "Hydrogen detection using platinum coated graphene grown on SiC," *Sensor and Actuators B: Chemical*, vol. 157, no. 2, pp. 500-503, Oct. 2011.

[4] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, A. A. Firsov, "Electric Field Effect in Atomically Thin Carbon Films," *Science*, Vol. 306, no. 5696, pp. 666-669, Oct. 2004.

[5] O. Leenaerts, B. Partoens, F. M. Peeters, "Adsorption of H_2O, NH_3, CO, NO_2, and NO on graphene: A first-principles study," *Phys. Rev. B*, vol. 77, no. 12, p. 125416, Mar. 2008.

[6] H. J. Yoon, D. H. Jun, J. H. Yang, Z. Zhou, S. S. Yang, M. M. C. Cheng, "Carbon dioxide gas sensor using a graphene sheet," *Sensors and Actuators B: Chemical*, vol. 157, no. 1, pp. 310-313, Sep. 2011.

[7] P. R. Wallace, "The band theory of graphite," *Phys. Rev.*, vol. 71, pp. 622–634, 1947.

[8] Y. W. Son, M. L. Cohen, and S. G. Louie, "Energy gaps in graphene nanoribbons," *Phys. Rev. Lett.*, vol. 97, no. 21, p. 216803, Nov. 2006.

[9] L. Yang, C. Park, Y. Son, M. L. Cohen, and S. G. Louie, "Quasiparticle Energies and Band Gaps in Graphene Nanoribbons," *Phys. Rev. Lett.*, vol. 99, no. 18, p. 186801, Nov 2007.

[10] S. Datta, "Nanoscale device modeling: The Green's function method," *Superlattices Microstruct.*, vol. 28, no. 4, pp. 253–278, Oct. 2000.

[11] D. A. Areshkin and B. K. Nikolić, "Electron density and transport in top-gated graphene nanoribbon devices: First-principles Green function algorithms for systems containing a large number of atoms," *Phys. Rev. B*, vol. 81, p. 155450, Apr. 2010.

[12] S. Reich, J. Maultzsch, C. Thomsen, and P. Ordejón, "Tight-binding description of graphene," *Phys. Rev. B*, vol. 66, p. 035412, Jul. 2002.

[13] G. Seol and J. Guo, "Metal contact to graphene nanoribbon," *Appl. Phys. Lett.*, vol. 100, no. 6, p. 063108, Feb. 2012.

[14] The National Technology Roadmap for Semiconductors, Semiconductor Industry Association, 2010.

Effect of Biaxial Strain on Structural and Electronic Properties of Graphene / Boron Nitride Hetero Bi-layer Structure

Md. Hasibul Alam,[1]Khalid Ibne Masood, and Quazi D.M. Khosru

Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology
Dhaka-1000, Bangladesh

[1]hasib.alam@ieee.org

Abstract—**In this work, structural and electronic properties of a bilayer structure composed of one single layer of graphene on top of a two dimensional hexagonal boron nitride (h-BN) has been investigated using first principle calculations based on density functional theory (DFT). Three different configurations are possible for this bilayer system namely B1, B2 and B3. In this paper biaxial strain was applied to B2 configuration and structural and electrical properties were observed. A small band gap opens for this bilayer structure though graphene is a zero gap material. Some important material parameter such as band gap, effective mass and Fermi velocity were investigated with the variation of biaxial strain. It was observed that the material remains to be a direct band gap material for the applied strain range from -12% to +12%. Interestingly, with the application of compressive strain effective mass decreases and Fermi velocity increases which is very crucial for high frequency RF devices.**

Index Terms—**Graphene, Boron Nitride, 2-D Material, Biaxial Strain, Density Functional Theory, Fermi velocity**

I. Introduction

As the dimensions of modern silicon metal-oxide semiconductor field-effect transistors (MOSFETs) have been shrinking according to the prediction of Moore's law, the conventional planar structures face severe short-channel effects [1]. Scaling theory predicts that a FET with a thin channel material like graphene will be robust against short-channel effects [2]. But graphene is not suitable for logic applications as it lacks proper band gap. However, it is possible to open a band gap in three ways: i) constraining large-area graphene in one dimension to form graphene nano ribbons (GNRs) [3]-[5] ii) biasing bilayer graphene [6]-[7] and iii) applying strain to graphene [8]-[9].GNRs have rough edges that cause band gap difference with different edge geometries [4]. If an electric field is applied in bilayer graphene, a band gap of 200-250meV opens [6]-[7] which is not suitable for application of transistors in logic circuits. The effect of uniaxial strain on the band structure has been simulated [10]-[11].However, opening a gap in this way will require a global uniaxial strain exceeding 20%, which will be difficult to achieve practically. On the other hand, though experimentally biaxial strain has been applied on graphene [12] but little is known about the influence of this strain on the band structure of graphene. However, once the problem of creating band gap is resolved in graphene, there comes the issue of substrate related mobility degradation for SiO_2/Si substrate [13]. Recently, graphene on h-BN has shown enhanced mobility, reduced carrier inhomogeneity and reduced intrinsic doping than on SiO_2 [14]. So to reduce SiO_2/Si substrate induced performance degradation, graphene/h-BN hetero bilayer composite material can be a promising channel material for future logic FETs. Several methods for tailoring band gap in this composite material has been proposed in recent literature [15]-[16] which varies with (a) the stacking order, (b) the separation distance of graphene from the h-BN, and(c) the externally applied perpendicular electric field. Also variation of band gap with biaxial strain was reported for grapheme/h-BN bilayer structure [17]. But detailed analysis for all the possible configurations of this bilayer system is still missing in literature.

In this work, structural and electrical properties of B2 configuration of bilayer graphene/h-BN were studied using DFT. Using Hellmann-Feynman theorem with a Broyden–Fletcher–Goldfarb–Shanno(BFGS)quasi-Newton optimization algorithm, optimized crystal structure for bilayer graphene/h-BN is computed. Then, the band structure of this hetero bilayer material was computed. With the help of band structure some important material parameter such as band gap, effective mass and Fermi velocity were computed. Biaxial strain was applied to the material and its effect on these parameters was also investigated.

II. Computational Details

Three distinct structures for graphene/h-BN super lattice are discussed in literature in which they are usually referred as B1, B2 and B3 [17]. In Fig. 1 we have produced the geometry of B1 and B2 structures schematically with Xcrysden [18]. The B1 structure is thoroughly discussed in [17]. Here we have conducted a detailed treatment on the B2 structure.

The electronic properties of the proposed super lattice are investigated using density-functional theory (DFT) with a plane-wave basis set. Perdew-Zunger variant LDA (Local Density Approximation) was performed using the Ab inito code PWSCF package of Quantum Espresso [19]. To include the electron-ion interaction we have used norm conservative pseudo potentials for all the atoms.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

Our calculation methodology started with structural optimization of the B2 super lattice structure and then a precise self consistent field (scf) analysis was done to

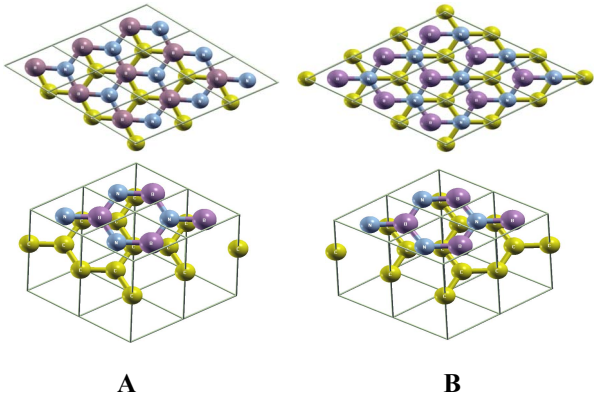


Fig. 1 Two structures of grapheme/h-BN super lattice. (A) B1, (B) B2. The blue ball represents Nitrogen, purple ball represents Boron and the yellow ball is carbon.

determine the band structure of the super lattice. For the structural optimization the plane wave basis cut off was set at 45 Ry and the convergence threshold on force was 10^{-3} Ry/a.u. A 15×15×1 Monkhorst-Pack k-point grid was used. To avoid the interaction between repeating unit cells a large vacuum space of 20 Å thickness was maintained throughout the optimization process. Then the lattice parameters were varied for different interlayer distances to determine the optimum unit cell for the B2 structure.

In the next part of scf calculation a 30×30×1 k-points Monkhorst-Pack grid was employed for better accuracy. The convergence threshold for self consistency energy was set at 10^{-8} Ry. All atomic position and unit cell parameters found as the final output of geometry optimization and relaxation process were used for scf and band-structure calculation. The band structure is plotted on the lines joining the high-symmetry points Γ, M, K and Γ.

III. ATOMIC STRUCTURE AND GEOMETRY OPTIMIZATION

At first the B1 structure was reviewed using PWSCF to compare the simulated results with the reported results published by G. Mukhopadhyay et al. to emphasize the accuracy and validity of our calculation. These results are summarized in Table 1 and Fig. 2. While Table 1 illustrates the optimized parameters obtained from both PWSCF and ELK code, Fig. 2 shows the band gap vs. percentage of biaxial strain applied to the super lattice. Both Table 1 and Fig. 2 indicate the accuracy of PWSCF package.

Having verified the compatibility of PWSCF package with the B1 structure we investigated the B2 structure. The optimization process was performed iteratively by varying both interlayer distance and lattice parameter and observing the minima of the potential energy surface. A built in variable cell relaxation function of PWSCF was run to verify the iterative process. The result is shown in Fig. 3, from which it is seen that the lattice parameter of the B2 is 2.48 Å and the inter layer distance is 3.4 Å.

A more precise relaxation process reveals that the optimized lattice parameter is 2.4805 Å and the interlayer distance is 3.44 Å, which agree with the lattice parameter of 2.47 Å and interlayer distance of 3.4 Å reported by [20] and [15] respectively.

TABLE I
LATTICE PARAMETER AND BAND GAP FOR B1 STRUCTURE

	Simulated Result	**Reported by G. Mukhopadhyay et.al. [17]**
Lattice parameter (a = b) (Å)	2.481	2.47
Interlayer Distance (Å)	3.241	3.217
Band gap (mev)	59	59

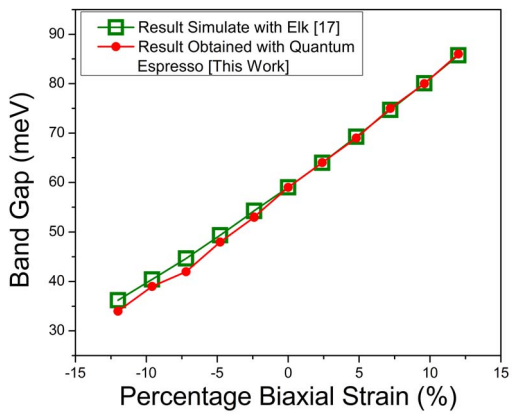


Fig. 2 The band gap of B1 structure as a function of biaxial strain applied to the super lattice. The red line shows the results obtained from PWSCF and the green line is obtained from the result published by G. Mukhopadhyay [17]

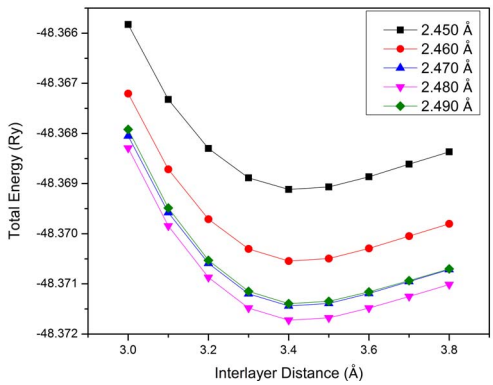


Fig. 3 The total energy of B2 structure as a function of lattice parameter and interlayer distance.

IV. ELECTRONIC STRUCTURE

The band structure of the super lattice was calculated along the line joining the points Γ, M, K and Γ in the Brillouine Zone. A total of 804 points was taken along the line using the Xcrysden. Care was taken to calculate the empty bands with the same accuracy as for the filled bands. The band structure for the optimized and relaxed super lattice is shown in Fig. 4 and a direct band gap of 54 meV was obtained for this case.

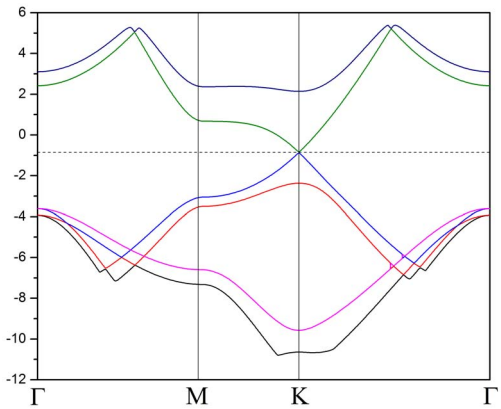


Fig. 4 Band structure of the B2 super lattice. The direct band gap of about 54 meV was found in the K symmetry point. The dashed horizontal line shows the Fermi energy.

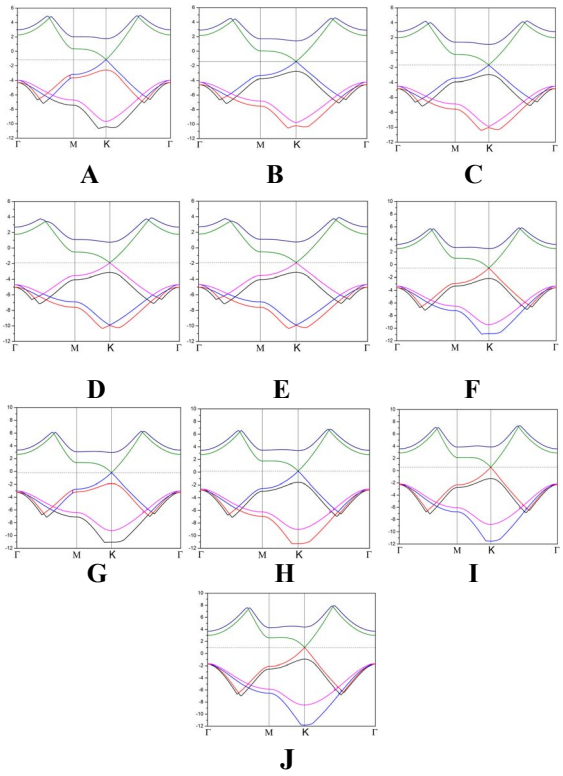


Fig. 5 The band structures of the strained super lattice. A to E shows the band structure of super lattice when +2.4%, +4.8%, +7.2%, +9.6% and +12% strain was applied. F to J shows the band structure when the super lattice was compressed to 2.4%, 4.8%, 7.2%, 9.6% and 12% respectively.

Furthermore biaxial strain was applied at the super lattice and each time the structure was relaxed before the scf analysis to conduct the band structure determination. The interlayer distance remained almost constant when expanded and slightly increased during compression. The band gap remained direct at the K point and during compression to expansion the band gap increased monotonically. Fig. 5 shows all the band structures of the strained super lattice.

From Fig. 5 it is observed that upon expansion the Fermi energy decreases gradually. Again when the structure is compressed, the Fermi energy increases and becomes positive.

The variation of band gap with the biaxial stress applied is shown in Fig. 6, which also compares the band gap between the B1 and B2 structures. This figure shows that with B2 structure offers less band gap than B1.

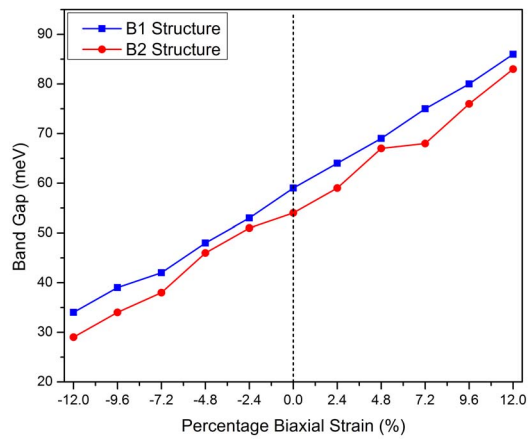


Fig. 6 The band gap of B2 (red line) and B1 (blue line) as a function of percentage biaxial strain applied to the super lattice.

In the next part we have calculated the longitudinal effective mass of both super lattices using the well-known formula: $m^* = \dfrac{\hbar^2}{\dfrac{\partial^2 E}{\partial k^2}}$ (1)

along the Γ to K path. The second derivative was evaluated on the lowest conduction band minima using five point stencil method [21], [22].The energy dispersion near the lowest conduction band minima is almost linear like graphene. Previous works [16], [17] have exploited this linearity to extract the electron effective mass using the formula

$$m^* \simeq \frac{E_g}{2 * v_f^2} \quad (2)$$

Where, E_g is the band gap and v_f is the Fermi velocity. But here we have tried to consider the meagre non-linearity present in the conduction band minima region and evaluate the effective mass from a more general formula (1). To appreciate our method we have presented a comparison between the effective mass calculated with equation (1) from Quantum Espresso results and the results published by G. Mukhopadhyay et al. with equation (2) [17] for B1 structure in Fig. 7. The results are in fine agreement. In Fig. 8 we have shown the electron effective mass for both B1 and B2.

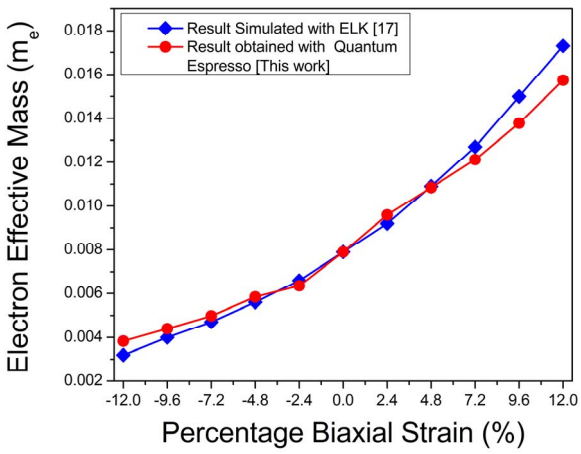


Fig.7 Longitudinal electron effective mass of B1 structure as a function of biaxial strain. Result obtained from equation (2) and with Elk simulation is shown in blue line. Electron effective mass calculated using equation (1) with Quantum Espresso is shown in red line.

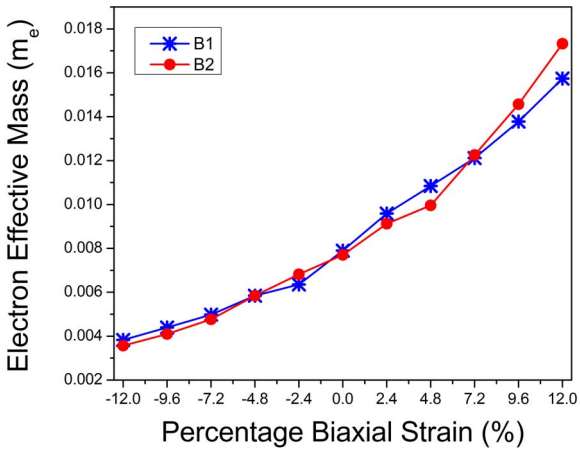


Fig. 8: Longitudinal electron effective mass of B1 structure (blue line) and B2 structure (red line) as a function of biaxial strain.

Fig. 8 indicates that the effective mass of the B1 and B2 structure is almost same except for the higher strain regions. At higher compression B2 structure prevails, while at higher expansion, B1 structure prevails.

Finally the Fermi velocity is calculated by fitting a straight line through the lowest conduction band minima and nearby almost linear points and then its slope was determined to calculate the group velocity using the equation:

$$v_g = \frac{\frac{\partial E}{\partial k}}{\hbar} \tag{3}$$

Using this formula the Fermi velocity for optimized B1 structure was found to be 0.864×10^6 m/s, closely matched with 0.8×10^6 m/s reported by [17]. For optimized B2 structure the value is 0.851×10^6 m/s almost equal to the value for B1.

The Fermi velocity for B1 and B2 is shown Fig. 9

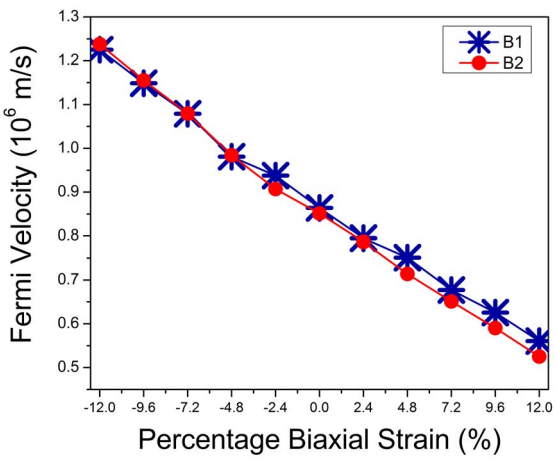


Fig. 9 Fermi velocity for B1 structure (blue line) and B2 structure (red line) as a function of biaxial strain

V. Conclusions

In this work, structural and electrical properties of B2 super lattice structure of graphene/h-BN hetero bilayer system were investigated. For optimized structure very high Fermi velocity (0.851×10^6 m/s) and extremely low effective mass ($0.0077m_e$) were observed. In addition, our study reveals that by applying compressive biaxial strain Fermi velocity can be increased and effective mass be decreased further which is the basic requirements of channel material for high frequency radio frequency (RF) applications. Moreover, the finding of non-zero band gap of the optimized structure (54 meV) that can be increased with tensile strain makes it also promising for channel material in logic transistors.

References

[1] I. Ferain, C. A. Colinge, and J. P. Colinge, "Multigate transistors as the future of classical metal-oxide-semiconductor field-effect transistors," *Nature*, vol. 479, no.7373, pp. 310-316, 2011.

[2] D. J. Frank, Y. Taur, and H. S. Wong, "Generalized scale length for two-dimensional effects in MOSFETs," *IEEE Electron Device Lett.*, vol. 19, no. 10, pp. 385-387, 1998.

[3] M. Y. Han, B. Özyilmaz, Y. Zhang, and P. Kim, "Energy band-gap engineering of graphene nanoribbons," *Phys. Rev. Lett.*, vol. 98, no.20, pp. 206805, 2007.

[4] M. Evaldsson, I. V. Zozoulenko, H. Xu, and T. Heinzel, "Edge-disorder-induced Anderson localization and conduction gap in graphene nanoribbons,"*Phys. Rev. B*, vol. 78, no.16, pp. 161407, 2008.

[5] F. Cervantes-Sodi, G. Csanyi, S. Piscanec, and A. C. Ferrari , "Edge-functionalized and substitutionally doped graphene nanoribbons: Electronic and spin properties," *Phys. Rev. B*, vol.77, vol.16, pp.165427,2008.

[6] E. V. Castro, K. S. Novoselov, S. V. Morozov, N. M. R. Peres,J. L. Dos Santos, J. Nilsson and A. C. Neto (2007), "Biased bilayer graphene: semiconductor with a gap tunable by the electric field effect," *Phys. Rev. Lett.*, vol. 99, no.21, pp. 216802, 2007.

[7] P. Gava, M. Lazzeri, A. M. Saitta, and F. Mauri , "Ab initio study of gap opening and screening effects in gated bilayer graphene," *Phys. Rev. B*, vol.79, no.16, pp. 165431, 2009.

[8] S. Y. Zhou, G. H. Gweon, A. V. Fedorov, P. N. First, W. A. De Heer, D. H. Lee, and A. Lanzara , "Substrate-induced band gap opening in epitaxial graphene," *Nature materials*, vol. 6,no. 10, pp. 770-775,2007.

[9] A. Bostwick, T. Ohta, T. Seyller, K. Horn, and E. Rotenberg, "Quasiparticle dynamics in graphene," *Nature Physics*, vol. 3, no.1, pp.36-40, 2007.

[10] V. M. Pereira, A. C. Neto, N. M. R. Peres , "Tight-binding approach to uniaxial strain in graphene," *Phys. Rev. B*, vol. 80, no. 4, pp. 045401,2009.

[11] Z. H. Ni, T. Yu, Y. H. Lu,Y. Y. Wang, Y. P. Feng, and Z. X. Shen, "Uniaxial strain on graphene: Raman spectroscopy study and band-gap opening ,"*ACS nano*, vol. 2, no.11, pp. 2301-2305, 2008.

[12] W. Pan, J. Xiao, J. Zhu, C. Yu, G. Zhang, Z. Ni, and X. Wang, "Biaxial compressive strain engineering in graphene/boron nitride hetero structures," Scientific reports, 2., 2012.

[13] J. H. Chen, C. Jang, S. Xiao, M. Ishigami, and M. S.Fuhrer, "Intrinsic and extrinsic performance limits of graphene devices on SiO_2," *Nature Nanotech.*, vol. 3, no. 4, pp.206-209, 2008.

[14] C. R. Dean, A. F. Young, I. Meric, C. Lee, L. Wang, S. Sorgenfrei, and J. Hone, "Boron nitride substrates for high-quality graphene electronics," *Nature Nanotech.*, vol. 5, no.10,pp. 722-726,2010.

[15] J. Sławińska, I. Zasada, and Z. Klusek. "Energy gap tuning in graphene on hexagonal boron nitride bilayer system." *Phys. Rev. B,* vol. 81, no. 15, pp. 155433, 2010.

[16] Fan, Yingcai, M. Zhao, Z. Wang, X. Zhang, and H. Zhang. "Tunable electronic structures of graphene/boron nitride heterobilayers."*App. Phys. Lett.,* vol. 98, no. 8, pp. 083103, 2011.

[17] Behera, Harihar, and G. Mukhopadhyay. "Strain-tunable band gap in graphene/h-BN hetero-bilayer." Journal of Physics and Chemistry of Solids 73, no. 7, pp. 818-821, 2012.

[18] A. Kokalj, "XCrySDen--a new program for displaying crystalline structures and electron densities," *J. Mol. Graphics Modeling*, vol. 17, pp. 176-179, 1999.

[19] P. Giannozzi et al., "QUANTUM ESPRESSO: a modular and opensourcesoftware project for quantum simulations of materials"J. Phys.Condens. Matter, vol. 21, pp. 395502, 2009.

[20] Giovannetti, Gianluca, P. A. Khomyakov, G. Brocks, P. J. Kelly, and J. V. D. Brink. "Substrate-induced band gap in graphene on hexagonal boron nitride: Ab initio density functional calculations." *Phys. Rev. B* 76, no. 7, pp. 073103, 2007.

[21] Y. H. Lee, X. Q. Zhang, W. Zhang, M. T. Chang, C. T. Lin, K. D. Chang, and T. W. Lin , "Synthesis of Large-Area MoS2 Atomic Layers with Chemical Vapor Deposition." *Advanced Materials*, vol. 24, no.17, pp. 2320-2325, 2012.

[22] Effective Mass Calculator, A. Fonari, C. Sutton, (2012).

Human-centered 2D/3D Video Content Analysis and Description

K. Papachristou,[1] N. Nikolaidis,[1,*] I. Pitas,[1,†] A. Linnemann,[2] M. Liu,[2] and S. Gerke[2]
[1]Department of Informatics, Aristotle University of Thessaloniki, Thessaloniki, Greece
[2]Image Processing Department, Fraunhofer Institute HHI, Berlin, Germany
{*nikolaid,†pitas}@aiia.csd.auth.gr

***Abstract*—In this paper, we propose a way of using the AudioVisual Description Profile (AVDP) of the MPEG-7 standard for stereo video and multichannel audio content description. Our aim is to provide means of using AVDP in such a way, that 3D video and audio content can be correctly and consistently described. Since AVDP semantics do not include ways for dealing with 3D audiovisual content, a new semantic framework within AVDP is proposed and examples of using AVDP to describe the results of analysis algorithms on stereo video and multichannel audio content are presented.**

***Index Terms*—AudioVisual Description Profile (AVDP), MPEG-7, stereo video, semantic content description**

I. Introduction

Automatic analysis of video and audio includes various tasks e.g., shot/scene boundary detection, person detection/tracking/recognition, facial expression recognition, music/speech segmentation, speaker diarization and music genre/mood characterization. The relevant algorithms are used in various applications, such as semantic description of video content for archival, indexing and retrieval, implementation of better audiovisual editing tools, intelligent content manipulation, etc. Recorded, broadcasted and webcasted data increase exponentially with time and research is focused on finding ways to analyse, describe and annotate video content in an efficient and automatic manner.

A rather new trend in multimedia is the use of stereoscopic video. Many of the recent film productions have their 3D versions [1]. Analysis of stereoscopic video has the advantage of benefiting from the additional available information, namely depth/disparity, which can boost the performance of analysis algorithms, such as the ones mentioned above. In addition, analysis of stereoscopic video can derive information that cannot be inferred from single-view video, such as 3D object position. Finally, the particularities of 3D video call for analysis algorithms that can characterize 3D video quality or its conformance to the rules of 3D cinematography [2].

For handling audiovisual content annotation and description, MPEG-7 standardizes a set of Descriptors (Ds), Description Schemes (DSs), a description definition language (DDL) and a description encoding [3]. A considerable amount of effort has been invested over the last years to improve MPEG-7 ability to deal with semantic content description. Nevertheless, 3D audiovisual content description has not yet been investigated in the MPEG-7 context. Although some description and description schemes have been proposed to model 3D information, they are only explicit descriptors for geometrical information and not for 3D video content.

The AudioVisual Description Profile (AVDP) has been recently adopted as a new profile of the MPEG-7 standard [4]. This profile consists of a subset of the original MPEG-7 standard and aims at describing the results of most of the known audiovisual analysis tasks (e.g., shot detection, face detection/tracking), in a normative manner. AVDP was designed to benefit both broadcasters and the digital media industry in order to create a normative layer between media content production and consumption. In this paper, we propose to store 3D video and multichannel audio content (e.g., 3DTV content) semantic analysis results to an XML file compatible to the specifications of the AVDP. Our aim is to show that AVDP can be used, by properly utilizing its descriptors and description schemes, for describing and storing the results of 3D audiovisual content analysis.

II. The Use of the AudioVisual Description Profile for 3D Content Description

The AudioVisual Description Profile (AVDP) provides a normative way to store high and low level information, extracted from the analysis of video and/or audio content. We have selected a subset of the decomposition types (namely TemporalDecomposition (TD), MediaSourceDecomposition (MSD), SpatioTemporalDecomposition (STD), SpatialDecomposition (SD)) available in the AVDP, in order to create various segment types (AudioVisualSegment (AVS), VideoSegment (VS), AudioSegment (AS), StillRegion (SR), MovingRegion (MR)) for representing the output of various analysis algorithms. In our approach, a content entity is used as the root of the description for a specific channel. Such a channel can be a color or disparity channel (left/right) of a stereo video segment or the audio content description, which unlike video is stored in a single content entity despite the fact that multichannel audio is present. All the relevant information of a channel obtained by various media analysis algorithms is stored within the content entity representing the corresponding channel by employing the above mentioned decomposition and segment types. Information that does not refer to a specific video channel (e.g., left or right), but to the 3D video as a whole, is stored to a reference channel, namely left channel. In the following sections, we will describe a number of 3D video and audio content analysis algorithms and the way their results can be stored using AVDP. Figure 1 illustrates the proposed AVDP usage for 3D content description.

III. Description of 3D Video Analysis Results

A. Scene/Shot Boundary Description

A scene/shot boundary detection algorithm can detect boundaries of scenes/shots in 2D or stereo video content, resulting in a temporal decomposition of a multimedia content into different scenes/shots [5], [6]. In terms of AVDP, a

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

TemporalDecomposition of an AudioVisualSegment describing the entire channel of a 3D video content is generated. Each resulting AudioVisualSegment represents a scene/shot. Moreover, with the same approach we can handle the case of shot transitions, such as fade-in, fade-out, dissolve etc. In all the cases, the type of the segment (e.g., scene, shot, transition of a certain type) is stored using the StructuralUnit element within the AudioVisualSegment. Additionally, semantic characterizations of scenes/shots, such as comfortable for 3D viewing, wide, close-up or popup, is provided using the How element of the StructuralAnnotation type. In the following partial XML example, a shot with its time and shot type information (medium shot) is described:

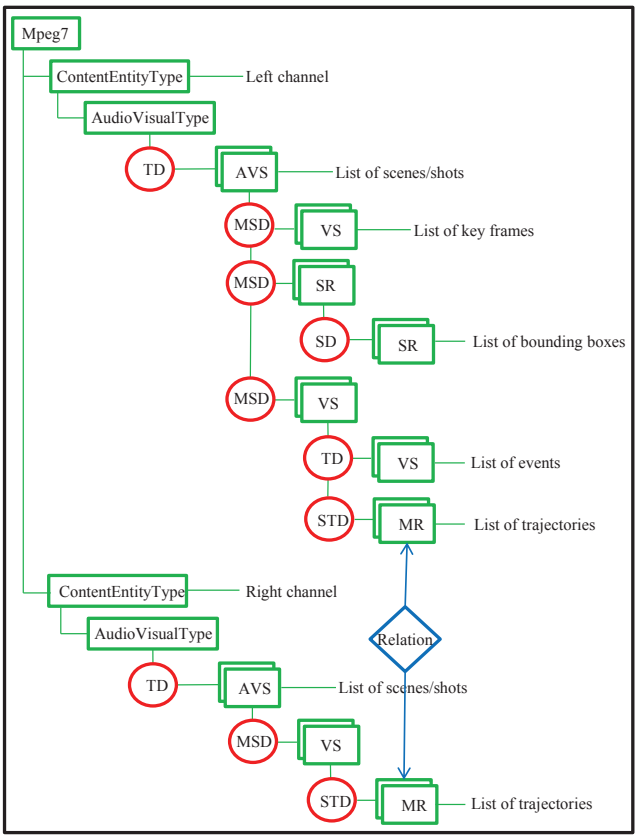


Fig. 1. Schematization of AVDP usage for 3D content description.

```
<AudioVisualSegment id="Shot_25">
 <StructuralUnit href="StructuralUnitCS#shot"/>
 <TextAnnotation>
  <StructuredAnnotation>
   <How href="SceneShotCharacterizationCS#medium"/>
  </StructuredAnnotation>
 </TextAnnotation>
 <MediaTime>
  <MediaRelIncrTimePoint>50</MediaRelIncrTimePoint>
  <MediaIncrDuration>192</MediaIncrDuration>
 </MediaTime>
</AudioVisualSegment>
```

B. Key Frame and Key Video Segment Description

Key frame or key video segment extraction algorithms produce characteristic video frames/segments summarizing a video segment [7]. Both color and depth information can be used for 3D video summarization. Key video frames and segments provide a video summary that can be used e.g., for fast browsing of query results in a 3D video asset management system. To describe such a summary, the MediaSourceDecomposition type is used within each respective shot, in order to generate a list of VideoSegment types representing each a key frame/video segment with the corresponding frame duration. Information regarding the respective multimedia data (e.g., image files for the key frames) can be stored using the MediaLocator type. A key frame example is given below:

```
<MediaSourceDecomposition
    criteria="DecompositionCS#key_segments">
 <VideoSegment id="KeyFrame_1">
  <MediaLocator>
   <MediaUri>/kf_4.jpg</MediaUri>
  </MediaLocator>
  <StructuralUnit href="StructuralUnitCS#key_frame"/>
   <MediaTime>
    <MediaRelIncrTimePoint>4</MediaRelIncrTimePoint>
    <MediaIncrDuration>1</MediaIncrDuration>
   </MediaTime>
  </VideoSegment>
</MediaSourceDecomposition>
```

C. Event Detection

An event is a semantically important concept having a certain duration in the audiovisual stream, such as a dialogue among people or a car accident. For describing events occuring within a shot, we use the TemporalDecomposition type of the VideoSegment, representing the visual information of the shot. The criteria attribute is set to “events”, to create a series of VideoSegments, each corresponding to a single event. The specific type of an event is stored within the StructuralUnit element of VideoSegment. An example of a dialogue occuring in the time interval [100, 50] is shown below:

```
<TemporalDecomposition
    criteria="DecompositionCS#events">
 <VideoSegment id="Event_1">
  <StructuralUnit href="StructuralUnitCS#event.dialogue"/>
  <MediaTime>
   <MediaRelIncrTimePoint>100</MediaRelIncrTimePoint>
   <MediaIncrDuration>41</MediaIncrDuration>
  </MediaTime>
 </VideoSegment>
</TemporalDecomposition>
```

D. Semantic 3D Video Quality Descriptions

3D video quality is highly related to the depth perception and the visual comfort while watching such a video [2], [8]. Semantic qualitative descriptions of 3D quality are more intuitive than quantitative descriptions and thus are preferred. Such descriptions can be at the shot level, pre frame or refer to a segment of the video. Shot level 3D quality semantic descriptions include information related to e.g., depth continuity, syncness, or depth stress. Annotation of the VideoSegment representing a shot is done by assigning the 3D quality semantic terms via the How element of StructuralAnnotation. Frame level 3D quality descriptions include information related to e.g., colorimetric or contrast mismatch, and are stored using the same approach as above within the StillRegion that corresponds to the frame. The semantic terms are defined in a respective Classification Scheme file. 3D video quality defects such as stereoscopic window violations, bent window effects or depth jump cuts refer to a segment of the video and the corresponding characterizations are stored as events (see Section III-C). An example of 3D quality description at shot level is given below:

```
<VideoSegment id="vis.shot_0">
 <StructuralUnit href="StructuralUnitCS#vis.shot"/>
 <TextAnnotation> <StructuredAnnotation>
   <How href="SemanticFeat3dCS#DisparityMap.low"/>
   <How href="SemanticFeat3dCS#DepthBudget.medium"/>
   </StructuredAnnotation> </TextAnnotation>
 <MediaTime>
  <MediaRelIncrTimePoint>100</MediaRelIncrTimePoint>
  <MediaIncrDuration>199</MediaIncrDuration>
 </MediaTime>
</VideoSegment>
```

E. Person/Face/Object Description in Stereo Video

The description of an object (e.g., a person, a face, a car, a ball) is performed after object/person detection and tracking that provide the location of an object over time. In the case of humans, additional information can be included, such as depicted activity or facial expression. Semantic characterizations of object/person position and motion can also be stored.

In more detail, object detection is the process of finding the location of a predefined object (e.g. a face, a car, a ball etc) in a per-frame basis [9]. Since the object detector usually detects a specific object or a specific object category, this information can be used to semantically annotate the detected object type. To store the location of the detected object(s), usually in terms of a bounding box, as well as other relevant information, a StillRegion type is used for each detected object, where the specific type (e.g., face, car) is stored, within the StructuralUnit element, while the SpatialLocator type is used to hold the coordinates of the bounding box.

Object/person tracking produces the trajectory of a predefined object (e.g., a face) in a sequence of video frames [10], [11]. Usually, object tracking is initialized by an object detector and performed in a video segment in a frame-by-frame basis [12]. The spatial position of the tracked object (usually in the form of a bounding box) over time forms the object trajectory. The type e.g. face or ball, and the coordinates of the bounding boxes of a trajectory are stored within the StructuralUnit and SpatioTemporalLocator elements of a MovingRegion, respectively.

The resulting StillRegions (detection) and MovingRegions (tracking) are stored within the VideoSegment representing the visual information of the corresponding shot. As can be seen in Figure 1, we use a MediaSourceDecomposition type, to decompose this VideoSegment into a list of StillRegion elements, where each of them represents an entire frame of the video segment. Subsequently, we decompose each frame into further StillRegions representing detected objects through a SpatialDecomposition type. Similarly, we use a SpatioTemporalDecomposition type to decompose the VideoSegment, representing the visual information of the corresponding shot, into a list of MovingRegion elements where each of them represents an object trajectory. In 3D video content, correspondences between StillRegions (e.g. bounding boxes) or MovingRegions (e.g. trajectories), that correspond to the same depiction of the object in two channels, e.g. between the right and left video channel, should be established. To denote such a correspondence, we use the Relation type in order to connect two Segment types, namely StillRegions and MovingRegions across channels.

After object/person detection and tracking, various video analysis algorithms may be used for extracting semantic descriptions for these entities, e.g., information regarding person identity, facial expression or activity and/or object motion. More specifically, human activity recognition can be used to recognize specific predefined human activities, such as run or walk on a person trajectory. Such information can be stored using the WhatAction element of the StructuredAnnottion of the respective StillRegion or MovingRegion. Moreover, facial expression recognition labels, such as happiness, anger, fear etc., can be stored in the WhatAction element. The value “affect” is used within the StructuredAnnotation of the respective StillRegion or MovingRegion, whereas the How element holds the recognized expression.

By using geometrical reasoning algorithms, one can also annotate 3D video content with information related to the geometric or motion properties of objects or object ensembles [13]. Such properties may refer to the geometrical position of an actor or object in the 3D world, the size of an actor/object, the location of a displayed object with respect to the screen or the stereoscopic comfort zone, the direction of object motion etc. In the case of stereo video, object position and motion characterization in the depth domain can be obtained. Such geometric descriptions are stored using appropriate elements of the StructuralAnnotation type of StillRegions and MovingRegions. Specifically, the Where, WhatObject and How elements are used to store labels related to the position (e.g., “left”, “near”, “in front of screen”), size and movement (e.g., “forward”) of an object, respectively.

Image regions depicting faces (facial images) can be clustered into clusters of actors though a facial image clustering algorithm [14]. For storing the results of face/object clustering (i.e., labels such as Actor_1, Actor_2,...), we update appropriately the Who/WhatObject element, respectively, in the structured annotation of each involved segment type (StillRegion and MovingRegion).

It should be noted that in all above cases, the href attribute of the various elements (namely Who, WhatObject, How, Where) of StructuralAnnotation is updated by utilizing appropriate terms from corresponding Classification Schemes.

Relations between faces and real persons may be obtained through manual annotation or through a face recognition algorithm. For example, a face, that is automatically detected and is given an abstract face term (e.g., Actor_1) after face clustering, may be annotated as depicting a certain real person (e.g., Jack Smith). Hence, relations between abstract face terms and real names may be obtained. Each such relation is represented by a StructuredAnnotation which contains two ’Who’ elements: the first ’Who’ element holds the corresponding abstract face term and the second a term from a Classification Scheme containing real names of actors. The resulting StructuredAnnotations are stored in the top level structure for content description, namely the AudioVisual.

In the following XML example, a face trajectory, where the depicted person is sad and running, as well as the relation between Face_1 and John Smith is described:

```
<Description xsi:type="ContentEntityType">
 <MultimediaContent xsi:type="AudioVisualType">
  <AudioVisual id="CHANNEL_left" mediaTimeUnit="PT1N25F">
   <StructuralUnit href="StructuralUnitCS#channel.left_view"/>
   <TextAnnotation><StructuredAnnotation>
     <Who href="PersonCS#John_Smith"/>
     <Who href="FaceCS#1"/>
    </StructuredAnnotation></TextAnnotation>
   <MediaTime>
    <MediaRelIncrTimePoint>0</MediaRelIncrTimePoint>
    <MediaIncrDuration>1000</MediaIncrDuration>
   </MediaTime>
   <TemporalDecomposition
   criteria="DecompositionCS#shots" id="ShotSet_1">
   ...
       <SpatioTemporalDecomposition
            criteria="DecompositionCS#humans">
        <MovingRegion id="MovingObject_97">
         <StructuralUnit
           href="StructuralUnitCS#face.trajectory"/>
         <TextAnnotation>
          <StructuredAnnotation>
           <WhatAction href="ActionCS#run"/>
          </StructuredAnnotation>
          <StructuredAnnotation>
           <WhatAction href="ActionCS#affect"/>
```

```
          <How href="AffectCS#sadness"/>
         </StructuredAnnotation>
         <StructuredAnnotation>
          <Who href="FaceCS#1"/>
         </StructuredAnnotation>
        </TextAnnotation>
        <Relation strength="1.0" target="MovingObject_121"/>
        <SpatioTemporalLocator>
         <ParameterTrajectory motionModel="still">
          <MediaTime>
           <MediaRelIncrTimePoint>188</MediaRelIncrTimePoint>
           <MediaIncrDuration>1</MediaIncrDuration>
          </MediaTime><InitialRegion><Polygon>
            <Coords mpeg7:dim="2 4">38 33 0 -33 19 0 35 0
            </Coords></Polygon></InitialRegion>
         </ParameterTrajectory>
        </SpatioTemporalLocator>
       </MovingRegion>
      </SpatioTemporalDecomposition>
      ...
   </TemporalDecomposition>
  </AudioVisual>
 </MultimediaContent>
</Description>
```

F. Soccer Analysis in Stereo Video

3D Soccer analysis [15] extracts diverse semantic informations, such as camera view class, players positions and identities and interesting highlights [16]. More specifically, the goal of camera view classification for football videos is to recognize different camera views. Usually, four different camera view classes can be observed during a broadcast football match: (i) overview, which is the main camera placed usually on the stands, (ii) close-up, these are views taken from cameras around the field showing a single player. (iii) Medium views are also taken from cameras around the field but show more than one player. (iv) Out-of-field denotes those shots not showing any field action but either audience or the coaching bench. Such class label is stored as a How element within the StructuralAnnotation of the VideoSegment representing the visual information of the respective shot.

Soccer highlights such as corners, goals, etc can also be annotated. This is sensible, as soccer broadcasts usually follow a characteristic cinematic model. For example, most goal scenes start with the shot on the goal shown by an overview camera, followed by medium and/or close-up shots of celebrating players, coaches and audience. Goal scenes typically end with a replay, showing the goal shot from different perspectives. Soccer highlights can be stored using the same approach used for annotating events (see Section III-C).

Finally, players can be detected and tracked for the duration of their appearance. For identifying players, the most obvious way of performing player identification is by recognizing jersey numbers. Similar to other subjects (e.g., actors in a movie), a MovingRegion is used to represent a player's trajectory, while its name is stored in the Who element of the StructuralAnnotation type. In the following example, a shot of a soccer match is characterized as overview shot and a corner is stored as an event:

```
<VideoSegment id="vis.shot_1">
 <StructuralUnit href="StructuralUnitCS#vis.shot"/>
 <TextAnnotation><StructuredAnnotation>
   <How><Name>Overview</Name></How>
  </StructuredAnnotation></TextAnnotation>
 <MediaTime>
  <MediaRelIncrTimePoint>149</MediaRelIncrTimePoint>
  <MediaIncrDuration>91</MediaIncrDuration>
 </MediaTime>
 <TemporalDecomposition
     criteria="DecompositionCS#events">
 <VideoSegment id="EVID_0">
  <StructuralUnit href="event.corner"/>
  <MediaTime>
   <MediaRelIncrTimePoint>250</MediaRelIncrTimePoint>
   <MediaIncrDuration>49</MediaIncrDuration>
   </MediaTime>
  </VideoSegment> </TemporalDecomposition> </VideoSegment>
```

IV. Conclusions

In this paper we have presented a new way of using the AVDP profile of the MPEG-7 standard for 3D video content description that can accommodate multiple audio and video (left/right, color/disparity) channels. We have detailed how we can store the results of several audiovisual analysis algorithms within such a context and how specific 3D metadata can be incorporated in the framework. The proposed framework can be used for storing the analysis results in a 3DTV media asset management (MAM) system. Such a system can then be queried and return the video e.g., where actor X appears in the comfort zone in a 3D context. The proposed XML description can be easily extended to support multiview video coming from multiple cameras.

Acknowledgment

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement number 287674 (3DTVS). This publication reflects only the author's views. The European Union is not liable for any use that may be made of the information contained herein.

References

[1] A. Smolic, P. Kauff, S. Knorr, A. Hornung, M. Kunter, M. Muller, and M. Lang, "Three-dimensional video postproduction and processing," *Proceedings of the IEEE*, vol. 99, no. 4, pp. 607–625, 2011.

[2] B. Mendiburu, *3D Movie Making - Stereoscopic Digital Cinema from Script to Screen.* Focal Press, 2009.

[3] B. Manjunath, P. Salembier, and T. Sikora, *Introduction to MPEG-7: Multimedia Content Description Language*. J. Wiley, 2002.

[4] I. T. .-A. 1:2012, "Audiovisual description profile (avdp) schema," 2012.

[5] Z. Cernekova, I. Pitas, and C. Nikou, "Information theory-based shot cut/fade detection and video summarization," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 1, pp. 82–91, 2006.

[6] C. Cotsaces, N. Nikolaidis, and I. Pitas, "Video shot boundary detection and condensed representation: A review," *IEEE Signal Processing Magazine*, vol. 23, no. 2, pp. 28–37, 2006.

[7] P. Kelm, S. Schmiedeke, and T. Sikora, "Feature-based video key frame extraction for low quality video sequences," in *10th Workshop on Image Analysis for Multimedia Interactive Services*, 2009, pp. 25–28.

[8] S. Delis, N. Nikolaidis, and I. Pitas, "Automatic 3d defects identification in stereoscopic videos," in *20th IEEE International Conference on Image Processing*, Sept 2013, pp. 2227–2231.

[9] N. Nikolaidis, M. Krinidis, G. Stamou, and I. Pitas, "Motion tracking in video," in *The Essential Guide to Video Processing*, A. Bovik, Ed. Elsevier, 2009.

[10] O. Zoidi, A. Tefas, and I. Pitas, "Visual object tracking based on local steering kernels and color histograms," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 5, pp. 870–882, 2013.

[11] H. Baltzakis, A. Argyros, M. Lourakis, and P. Trahanias, "Tracking of human hands and faces through probabilistic fusion of multiple visual cues," in *Proceedings of the 6th International Conference on Computer Vision Systems*. Springer-Verlag, 2008, pp. 33–42.

[12] O. Zoidi, N. Nikolaidis, and I. Pitas, "Appearance based object tracking in stereo sequences," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 2434–2438.

[13] N. Papanikoloudis, S. Delis, N. Nikolaidis, and I. Pitas, "Semantic description in stereo video content for surveillance applications," in *Int Workshop on Biometrics and Forensics*, April 2013, pp. 1–4.

[14] G. Orphanidis, N. Nikolaidis, and I. Pitas, "Facial image clustering in 3d video using constrained ncut," in *21st European Signal Processing Conference*, September 2013.

[15] T. D'Orazio and M. Leo, "A review of vision-based systems for soccer video analysis," *Pattern Recognition*, vol. 43, no. 8, pp. 2911–2926, 2010.

[16] A. Linnemann, S. Gerke, S. Kriener, and P. Ndjiki-Nya, "Temporally consistent soccer field registration," in *20th IEEE International Conference on Image Processing*, Sept 2013, pp. 1316–1320.

Emergence of THz Technologies and Design and Optimisation Low-loss Waveguides and Devices

B.M.A. Rahman,[1,*] M. Uthman[1], A. Quadir[1], K. T. V. Grattan[1], C. Markides[2] and C. Themistos[2]

Dept. of Electrical and Electronic Engineering,
City University, London, UK
School of Engineering and Applied Sciences
Frederick University, Nicosia, Cyprus
*b.m.a.rahman@city.ac.uk

***Abstract*— THz is an emerging technology with many important applications in imaging and sensing, but due to lack of suitable low-loss waveguides future progress can be limited. A rigorous full-vectorial modal solution approach based on the computationally efficient finite element method is used to find the propagation properties of THz waveguides. Design approaches are presented to reduce the modal loss of such waveguides. Designs of several THz devices, including quantum cascade lasers, power splitters and narrow-band filters are also presented.**

***Index Terms*— Finite Element Method, Terahertz Waveguides, THz Devices.**

I. Introduction

The emerging terahertz (THz) region occupies a large portion of the electromagnetic spectrum, located between the microwave and optical frequencies and normally is defined as the band ranging from 0.1 to 10.0 THz. In recent years, this intermediate THz radiation band has attracted considerable interest, because it offers significant scientific and technological potential for applications in many fields, such as sensing [1], imaging [2] and spectroscopy [3]. However, waveguiding in this intermediate spectral region is a major challenge and strong dielectric and conductive losses in the terahertz frequency range have been a major problem for the development of practical low-loss waveguides. Due to the lack of suitable low-loss waveguides most of the present day THz systems uses free-spaces transmission. The availability of low-loss waveguides and followed up by THz guided-wave devices can improve the functionality and reliability of future THz systems. The conventional guiding structures exemplified by microstrips, coplanar striplines and coplanar waveguides are highly lossy and dispersive. However, so far the most promising dielectric waveguides have been the use of photonic crystal fibers at terahertz frequencies [4, 5] and metal coated guides [6] at terahertz frequencies. In this paper, various types of practical dielectric and metal coated waveguides are evaluated and design optimization of Quantum Cascade Lasers, MMI-based power splitters and narrow-band filters are presented, by using full-vectorial finite element method [7].

II. Page Layout

Dielectric waveguides in silica, silicon, polymer, and other semiconductors materials have been widely used at optical frequencies due to their low-loss values: however most of these materials are very lossy at the lower terahertz frequency range and not so suitable to fabricate THz waveguides. Particularly, silica has been extensively used to fabricate a new class of optical fibre, the photonic crystal fibres (PCF), however, as its material loss is prohibitively high at THz frequencies, only recently, Han *et al*. [4] have fabricated a PCF for THz using alternative high-density polyethylene (HDPE) with modal loss values of 0.2 cm^{-1} and Goto *et al*. [5] have reported a PCF-like waveguide using Teflon tubes and filaments with loss values of 0.5 cm^{-1}, showing their potential. A typical PCF cross section with a triangular array of holes is shown in Fig. 1. The structure predominantly consists of such periodic air-holes with diameter, *d*, and the pitch length between the two nearest holes is Λ. In this work, a refractive index value n_g = 1.444 is considered, at the operating frequency 1.2 THz (wavelength 0.25 mm). Variations of the effective indices for both the fundamental H^x_{11} and the second H^x_{21} modes for d/Λ = 0.9 are shown in Fig.2. It can be noted that for both the modes, the effective indices reduce monotonically as the pitch length is reduced. The variation of the n_{fsm} value with the pitch length is also shown by a chained line for d/Λ = 0.9 at 1.0 THz. The n_{fsm} represents a frequency dependent equivalent index of the perforated cladding region and this was calculated by solving a unit cell problem with periodic boundary conditions implemented around all the four sides. It can be observed that the effective index curve for the H^x_{21} mode crosses the n_{fsm} line at a pitch length, Λ_c = 0.24 mm, so when the pitch length is reduced below 0.24 mm, the second mode cannot be guided anymore. On the other hand, the effective index for the fundamental, H^x_{11} mode also crosses the n_{fsm} line, but at Λ_c = 0.11 mm. So, a PCF with d/Λ = 0.9 and pitch length between 0.11 mm to 0.24 mm would be strictly single moded. By controlling the d/Λ ratio and the pitch length the modal properties of such low-loss PCFs at THz operating frequency can easily be controlled. It is possible to broaden the single mode operating region or even design an endlessly single mode PCF by reducing the d/Λ value.

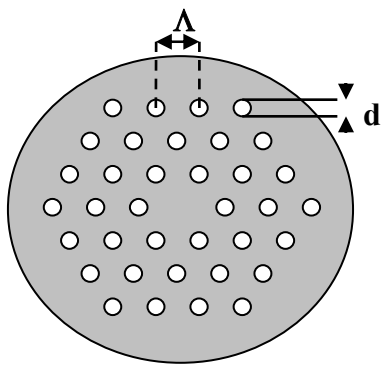


Fig. 1. Cross section of a Photonic Crystal Fiber

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

The modal loss of a guided mode in a PCF is due to the combination of the material loss and the leakage loss. The leakage loss arises due to the modal index being lower than the surrounding high index cladding regions. For this study a PCF with parameter d/Λ=0.5 is used. To consider the effect of the loss tangent, three different values of the n_i, (the imaginary part) have been considered. Additionally, Perfectly Matched Layers (PML) were added around the conventional waveguide boundaries to calculate the leakage losses. Figure 3 shows the loss mechanism in a PCF. In the absence of any material loss, for $n_i = 0$, the leakage loss is shown by a solid line. It can be observed that as the pitch length is reduced, the leakage loss increases almost linearly from a very low loss value. It should be noted that Perfectly matched Layers (PML) region around orthodox computational boundary should be introduced to obtain leakage loss of any waveguide. In the case of n_i = 0.00119, the total loss included both the leakage loss and the material loss, which is shown by a dashed line. At a higher pitch value, when the leakage loss is negligible, the total loss is mainly attributable to the material loss. However, it can be observed that modal loss is more than 100 dB/m, this is only because lack of suitable dielectric material in this frequency range.

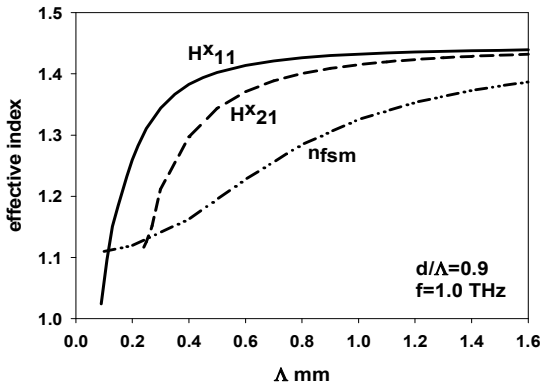


Fig. 2. Variation of the effective indices with the pitch length, Λ.

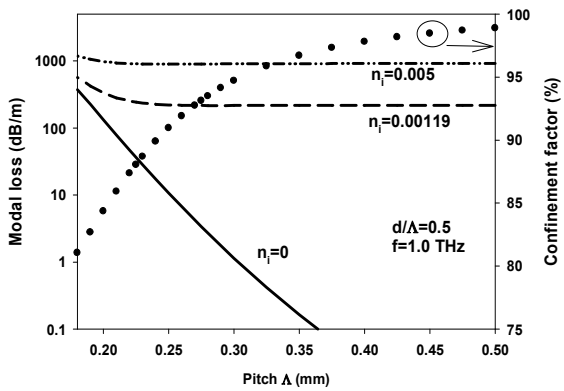


Fig. 3. Material and leakage losses of PCF

Since at present the available dielectric materials at THz frequencies are considerable lossy, a novel design approach is considered next, where besides the porous cladding with air-holes, additionally a porous core PCF is considered. As the waveguides dimensions for THz frequency are considerable bigger than that for the optical wavelength, it would be relatively easy to fabricate such microstructured core. Two different d/Λ values are used for core (d/Λ_i) and cladding (d/Λ_o), and for cladding this value has to be smaller to have higher equivalent index in the core for wave guidance. Variation of the power fraction in the air-region of the porous core is shown in Fig.4. As shown in this figure, the power confinement can be increased to 35% in the low-loss air-holes of the cores, and additionally another 25% in the cladding air-holes [8], but this is not shown here. So, in this design, the overall modal loss mainly arises due to the material loss of the polystyrene, and this can be reduced by 60% but using such perforated core besides air-holes in the cladding region. The overall reduction is significant, but still need more work to be done to make the more suitable for practical applications.

Amongst the various THz waveguides that have been suggested, the metal-clad waveguides supporting surface plasmon modes show the greatest promise as low-loss waveguides for use both in active components and as passive waveguides. Several waveguide structures incorporating metallic layers have been reported, such as low-loss and flexible hollow polycarbonate waveguide with copper and dielectric inner coatings, deposited by using a liquid chemistry approach [9].

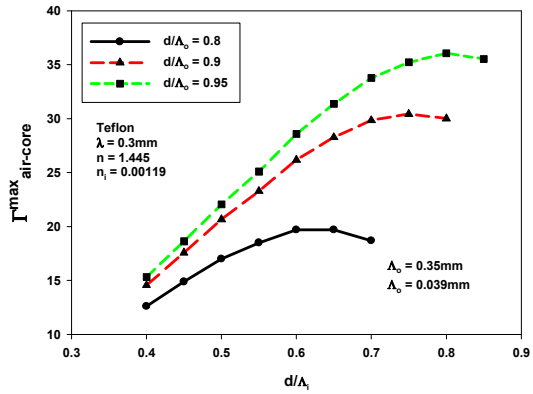


Fig. 4. Variation of power confinement in the porous air-holes in the core region with the pitch length, Λ

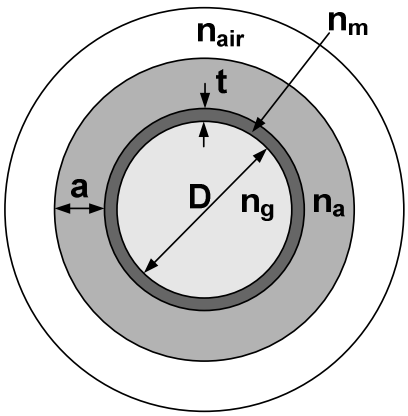


Fig. 5. Metal-coated hollow glass waveguide

A metal-coated hollow glass waveguide (HGW) [10] with an inner silver/polystyrene-coating, as shown in Fig.5, is considered for the better understanding of the various loss mechanisms and subsequently design optimization of a low-loss THz waveguide. For this waveguide the thickness of the silica tube (*s*) is taken as 0.5 mm and the bore diameter (*D*) of the HGW as 4 mm. Inside the hollow-core first metal and then polystyrene are coated. The thickness of the silver cladding (*t*) and polystyrene layer (*p*) are taken as t µm and p µm, respectively. The complex refractive index of the polystyrene,

the silver metal cladding layers and the silica ring are taken as n_p=1.58 –j0.0036, n_m=308-j532 and n_s=1.96 –j0.0061, respectively at an operating frequency of 2.5 THz. At this frequency, as the air is also not absolutely loss-free, to represent this loss factor its complex refractive index is taken as 1.0 - j 1.1x10^{-6}.

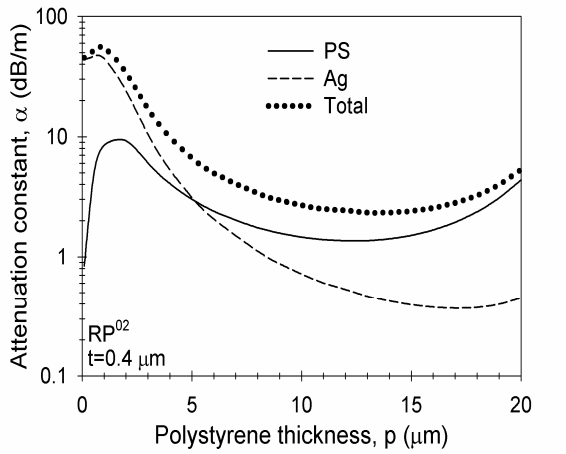


Fig. 6. Attenuation constant with the polystyrene thickness, *p*, for *t*=0.4μm

There are two metal/dielectric interfaces, which can support surface plasmon modes (SPM), one the outer silver/silica boundary and the other at the inner silver/polystyrene (or air, when p = 0 μm) boundary. This waveguide supports two SPMs along these two metal/dielectric interfaces. The refractive indices of the inner and outer cladding materials being very different, the two SPMs have widely different propagation constants and they do not interact with each other. However, at the right and left hand sides of the metal/dielectric interfaces, when the same electric-wall boundary condition is imposed, the H_x field is forced to be zero at the metal boundary and no SPM exists. Another mode with the dominant H_y field, would form a similar SPM; however, at the left and the right interfaces. The H_y field profile of this mode is similar to the H_x field profile, but rotated by 90 degrees. These two modes have identical propagation constants and being degenerate, they can be superimposed to form radially polarized RP-like modes.

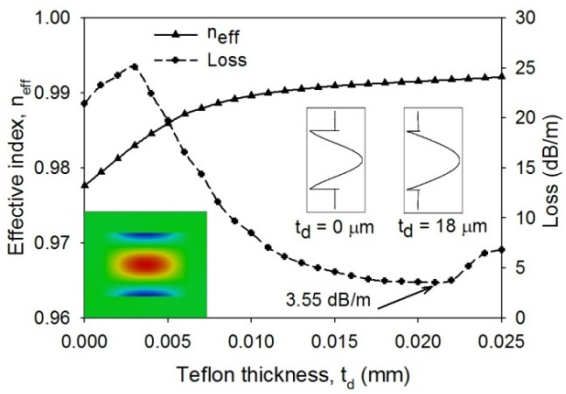


Fig. 7. Effective index and loss with Teflon thickness for the H^x_{10} mode

The loss values of the fundamental plasmonic mode increases with the PS thickness and not shown here, as this mode was more lossy. The attenuation characteristics of the RP^{02} mode, with the variation of the polystyrene thickness, for a silver thickness t = 0.4 μm, is shown in Fig. 6, where the loss contribution of the polystyrene and silver layers has also been examined. As can be seen from the above characteristics, the total attenuation shows a maximum and a minimum loss at a polystyrene thickness of about 1 μm and 13 μm, respectively. The attenuation curves due to the polystyrene and the silver layer exhibit similar trend with the total attenuation. Throughout the range of polystyrene thicknesses examined, the optical power confinement in the inner air-core is of the order of 99.9%, thus contributing a constant attenuation of about 0.25 dB/m. For a polystyrene thickness lower than 5 μm, the total attenuation is affected mainly due to the metal attenuation but as the polystyrene thickness increases above 5 μm the total attenuation is mainly governed by the loss in the polystyrene layer. This mode shows a greater promise to achieve low-loss guidance through a metal clad dielectric waveguides. It would also be easier to couple this mode since the field profile is also very close to a Gaussian shape [10]. The modal loss of this waveguide, when design is optimized, is significantly lower than most of the THz waveguides reported so far, as most of the power is being guided in the central air-hole region.

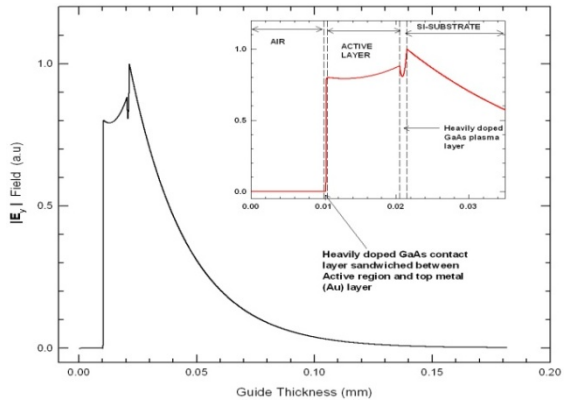


Fig. 8. Field profile of $\mathbf{E}_y$ mode in a THz Quantum Cascade Laser. The field inside the confinement layers is shown on the inset.

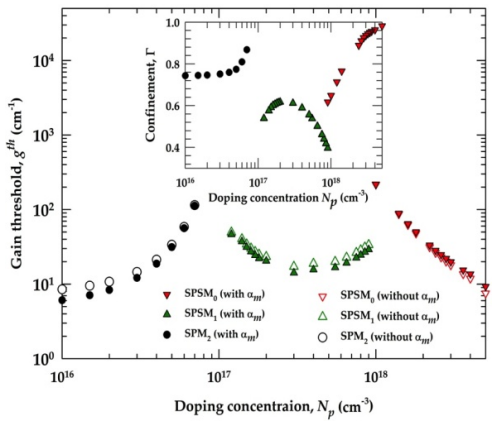


Fig. 9. Gain threshold of several plasmonic modes.

Next this novel approach of incorporating additional PS layer to draw away power from lossy metal layer is considered to design and optimize a low-loss rectangular core metal waveguide. If we consider its height and widths are different then the two polarized modes will not be degenerate. In our more recent work [11], we have shown that similarly a polarization maintaining rectangular core air-core dielectric-

clad metal-coated waveguide can also be less lossy. A thin metal coating would support plasmonic modes, but these are relatively lossy. However, a Teflon coating on the gold layer can draws field away from the lossy conducting layer and loss may reduce considerably. Figure 7 shows the variation of the loss value with the Teflon thickness for the H^x_{12} mode in an air-core 1 mm x 0.6 mm rectangular waveguide with 0.7 µm gold coating at 2.5 THz. It can be seen that at the optimum 21 µm Teflon thickness, the loss value can be 3.5 dB/m, one of the lowest reported so far [11]. The evolution of third order mode for no Teflon coating to a near Gaussian profile for 18 µm Teflon coating are shown as insets. In this guide also, the fundamental TE_{10} or TE_{01} mode was more lossy and its loss value rather increased with the polystyrene coating. Further improvement in the waveguide design to reduce loss and dispersion can encourage the development of THz guided-wave systems.

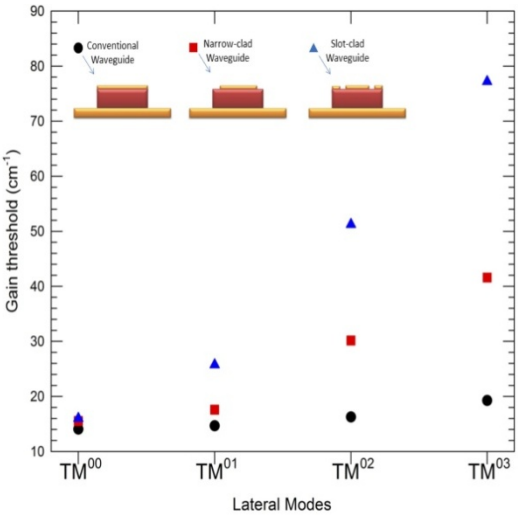


Fig. 10. Gain Threshold of the lower order modes for different waveguides

We also need good sources for THz generations. There are some convergences for the development THz sources from microwave engineering to increase the frequency and from photonics to reduce the frequency. Quantum cascading lasers [12] are emerging as efficient high power THz source for many important applications, such as imaging and sensing. However, for the THz frequency, it is not possible to grow the semiconductor materials comparable to the operating wavelength. For this reason, for THz QCL often smaller height needs to be considered and plasmonic confinement is used in that direction. The plasmonic confinement in the vertical direction is shown in Fig.8, which clearly shows the mode formation at the metal-dielectric interfaces. However, it is easily possible to have a wider guide, so mostly dielectric confinement is used in horizontal direction. The gain threshold for such QCL is shown in Fig.9. Because of the wider guides, the gain threshold difference between the fundamental and higher order (lateral) modes are very small. This allows possibility of mode hopping for any external changes. A novel design approach is considered [13] using slotted upper metal clad, to enhance the gain threshold of the higher order modes. Figure 10 clearly shows that gain threshold of the higher order modes are increased for slotted-electrode designs. This would reduce mode competition and the resulting output beam would be more stable with the environmental variations. Besides reducing waveguide loss of QCL, it may also be possible to increase the operating temperature of QCL, another critical issue in the further development of room-temperature QCLs.

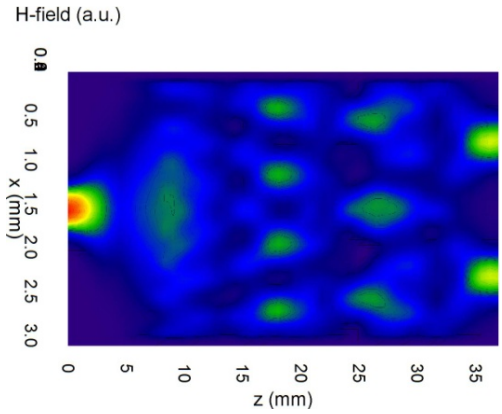


Fig. 11. FDTD Simulation of the MMI 3dB coupler

For the future THz system, it is essential to design various integrated guided-wave components. In that spirit, it is shown here that a compact power splitter can be designed by using the MMI principle. In Fig. 11, it is shown here that an efficient power splitter can be designed by using a 35 µm long multimoded section [14]. A Finite Difference Time Domain (FDTD) approach is used here to simulate this structure.

Next, a Terahertz frequency range band-stop filter for molecular sensing [15], where two 5 µm wide band-stop filter stubs with a length of 192 µm and 83 µm are placed at a 400 µm apart along the direction of propagation, as shown in the right inset of Fig.12, has been considered. The structure of the microstrip is also shown as an inset on the lower left side. Initially the above device has been simulated without a polysterene film on top of the metal layer, using the FDTD approach and the variation of the insertion loss with the frequency is presented in Fig.12. As can be seen from the above frequency response, the device exhibits two resonant frequencies due to the stubs at about 600 and 800 GHz, with a minimum insertion loss of about -55 and -30 dB. These resonant peaks will shift due to presence of any external materials on the microstrip and measuring this frequency shift, a THz sensor can be developed.

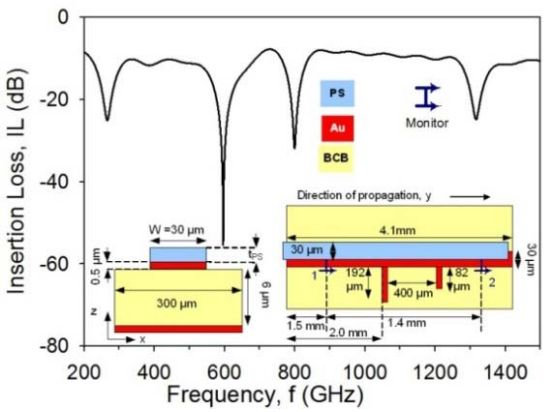


Fig. 12. Insertion Loss with frequency for the microstrip filter

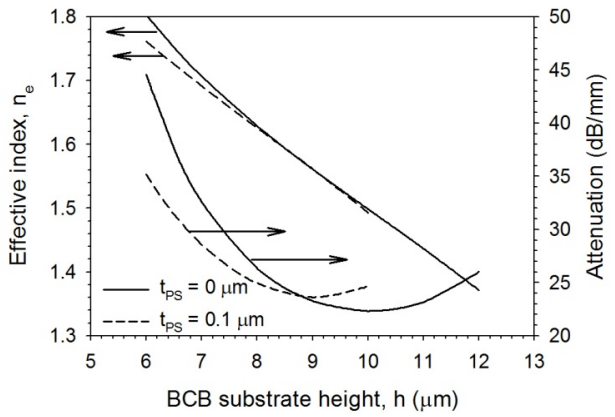


Fig. 13. Variation of the effective index and attenuation with BCB substrate height of the microstrip line.

Finally, the BCB substrate height, h, has then been varied and the effective index and attenuation of the microstrip structure with no polysterene layer on top of the metal and that with a 0.1 μm film thickness of the above material have been examined and presented in Fig.13. As can be seen from the above characteristics, as the substrate height increases both the effective index and the attenuation decrease, with the corresponding values of the structure with the polysterene film being lower than those obtained for the device without it. At a substrate height of about 9 μm, the above corresponding values are about the same for both the structures examined. However, the device with the polysterene film exhibits minimum attenuation at the above substrate height and as the substrate height increases further the attenuation of the structure with the polysterene layer becomes higher than that obtained for the structure without the above material.

III. Conclusions

A finite-element approach, based on a full-vectorial **H**-field formulation, has been used to study the detailed modal properties of dielectric and metal clad waveguides operating in the THz frequency range. It is also shown here that by using porous core the effect of material loss can be reduced significantly. It is also shown by using a thin but optimized dielectric over-layer plasmonic loss in a hollow-core waveguide can also be reduced. It is also shown here that by using a novel slot-type electrode, the differential loss of the higher order modes can be significantly increased to reduce mode hopping. Finally, simple guided-wave devices such as power splitters and band-pass filters are also presented here. The full-vectorial numerically efficient finite element based design approach used here can be extended to optimize not only THz waveguides but also more advanced guided-wave devices for future THz integrated circuits (TIC).

References

[1] R. H. Jacobsen, *et al*., "Chemical recognition of gases and gas mixtures with terahertz waves," *Optics Lett*., **21**, no.24, pp. 2011-2013, 1996.

[2] Q. Chen, *et al*., "Near-field terahertz imaging with a dynamic aperture," *Optics Lett*., **25**, no.15, pp.1122-1124, 2000.

[3] J. Q. Zhang and D. Grischkowsky, "Waveguide terahertz time-domain spectroscopy of nanometer water layers," *Optics Lett*., **29**, no.14, pp.1617-1619, 2004.

[4] H. Han, H. Park, M. Cho, and J. Kim, "Terahertz pulse propagation in plastic photonic crystal fiber," *Appl. Phys. Lett*, **80**, no.15, pp.2634-2636, 2002.

[5] M. Goto, *et al*., "Teflon photonic crystal fiber as terahertz waveguide," *Jpn. J. Appl. Phys*, Part 2 **43**, L317-L319, 2004.

[6] R. W. McGowan, *et al*., "Propagation of ultrawideband short pulses of Terahertz radiation through submillimeter diameter circular waveguides," *Opt. Lett*, vol. **24**, no. 20, pp. 1431-1435, Oct 1999.

[7] B. M. A. Rahman and J. B. Davies, "Finite-element analysis of optical and microwave waveguide problems," *IEEE Trans. Microwave Theory Tech*, **MTT-32**, no.1, pp.20-28, 1984.

[8] M Uthman *et al.*, Design and characterization of low-loss porous-core photonic crystal fiber, IEEE Photonics J, vol. 4, no. 6, pp.2314-2325, Dec 2012.

[9] J. A. Harrington, *et al*., "Hollow polycarbonate waveguides with inner Cu coatings for delivery of terahertz radiation," *Opt Express*, vol. **12**, no.21, pp. 5263-5268, Oct. 2004.

[10] C. Themistos, *et al*., "Characterization of Silver/Polystyrene (PS)-coated hollow glass waveguides at terahertz frequency," *J. Lightwave. Technol.*, vol. **25**, pp. 2456-2461, no. 9, September 2007.

[11] B.M.A. Rahman, A. Quadir, H. Tanvir and K.T.V. Grattan, "Characterization of Plasmonic Modes in a Low-Loss Dielectric-Coated Hollow Core Rectangular Waveguide at Terahertz Frequency," *IEEE Photonics Journal,* vol.3, pp.1054-1066, 2011.

[12] J. Faist, *et al*., "Quantum cascade laser," *Science*, vol.**264**, pp. 553-556, 1994.

[13] H. Tanvir, B.M.A. Rahman and K.T.V. Grattan, "Impact of "Ghost" Mode Interaction in Terahertz Quantum Cascade Lasers," *IEEE Photonics Journal*, vol.3, pp.926-935, 2011

[14] C. Markides, *et al*., "Multimode Interference 3dB splitters in hollow core metallic waveguides for low loss THz wave transmission", *IEEE J. Selected Topics in Quantum Electron*. Vol. 19, ID 8500606, Jan. 2013.

[15] J. Cunningham, C. Wood, A. G. Davies, I. Hunter, E. H. Linfield, and H. E. Beere, "Terahertz frequency range band-stop filters," *Applied Physics Letters,* vol.86, p.213503, 2005

On the Optimal Contention Window Size in IEEE 802.11 Based WLANs Under Co-channel Interference

Tasnova Nasrin Munni[1], Nazia Mahbub[1], Kaim Iftahaj Nirjhor[1] and Md. Forkan Uddin[2,*]
[1] Dept. of Electrical, Electronic and Communication Engineering, Military Institute of Science and Technology
Dhaka, Bangladesh
[2] Dept. of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
†mforkanuddin@eee.buet.ac.bd

***Abstract*—Throughput in IEEE 802.11 based wireless local area networks (WLANs) significantly depends on the minimum contention window (CW) parameter of CSMA/CA medium access control protocol. The throughput-optimal value of the minimum CW is modeled by G. Bianchi for a single WLAN without considering co-channel interference as $N\sqrt{2T_c}$, where N is the number of contending nodes in the WLAN and T_c is the collision time in mini-slots [1]. In this paper, the throughput-optimal configuration of the minimum CWs of the CSMA/CA based WLANs is studied under co-channel interference via extensive simulations. From the simulation results, it is investigated that in a WLAN the optimal value of the minimum CW is 4% to 25% higher than the Bianchi optimal CW due to co-channel interference. However, in WLANs, the optimal throughput is very close to the throughput obtained at the Bianchi optimal CW. In WLANs, very close to the optimal throughput can be obtained by configuring each WLAN independently with the simple Bianchi model even when the co-channel interference is considered.**

***Index Terms*—Local area networks, throughput, medium access control, co-channel interference, the minimum contention window.**

I. Introduction

In the last two decades, the popularity of wireless local area networks (WLANs) increases dramatically as wireless technology has evolved to meet the fundamental requirements of all categories of users. The WLANs are rapidly deployed in homes, public places, universities, offices and enterprises. Most modern WLANs are based on IEEE 802.11 standard which use carrier sense multiple access with collision avoidance (CSMA/CA) medium access control (MAC) protocol for managing interference [2]. IEEE 802.11 based WLANs use two unlicensed frequency spectrum bands: 2.4 GHz Industrial, Scientific, and Medical (ISM) band and 5 GHz Unlicensed National Information Infrastructure (UNII) band [3]. There are 14 channels available in IEEE 802.11 b/g. However, only three channels are non-overlapping among the total 14 available channels [4], [5]. Due to the limited number of non-overlapping channels, WLANs reuse the same frequency spectrum and co-channel interference exists in WLANs. The throughput performance in WLANs severely degrades due to co-channel interference [6].

Throughput performance of WLANs are determined by the configuration of the physical layer and MAC layer parameters [7]. A threshold level of power known as carrier sensing threshold (CST) is used in CSMA/CA MAC protocol to sense whether the medium is free or not. The spatial reuse as well as throughput in WLANs depends on the joint configuration of node transmit power and CST. In [8], it is shown that the optimal value of CST is related to the node transmit power and for a particular data rate the product of transmit power and CST is constant. The effect of transmit power and CST on the throughput performance in WLANs is also studied in [5], [7], [9], and [10]. All these studies advocate to configure the transmit power and CST appropriately to achieve a good throughput performance in WLANs.

Appropriate configuration of transmit power and CST is not enough to achieve the optimal throughput in CSMA/CA based WLANs since the efficiency of the MAC protocol depends on its contention resolution technique. CSMA/CA MAC protocol uses slotted binary exponential back off (BEB) algorithm for contention resolution and the minimum contention window (CW) is the main parameter of the BEB algorithm on which the throughput of WLANs depends mostly. In WLANs, collision occurs due to concurrent transmissions and interference. If the minimum CW of the nodes is set to a low value, the collision rate will be very high and the throughput performance will be very poor. Collision rate reduces and the throughput increases with increasing the value of the minimum CW of the nodes from a low value. On the other hand, if the minimum CW is set to a very large value, the nodes pass a significant amount of time in backing off and hence, the throughput performance will be poor. Thus, there is an optimal value of the minimum CW for the nodes for which throughput in a WLAN is the optimal.

Under the *protocol interference model* without considering co-channel interference, an analytical model of throughput in a single WLAN is provided by G. Bianchi [1]. The expression of the optimal value of the minimum CW for a single WLAN is also derived as $N\sqrt{2T_c}$, where N is the number of contending nodes in the WLAN and T_c is the collision time in mini-slots. Throughout the paper this configuration of the minimum CW is called as Bianchi optimal CW. Recently, in [11], it is shown that the changes in the optimal value of the minimum CW is negligible for using the *physical interference model* instead of the *protocol interference model* in a single WLAN. So far the optimal configuration of the minimum CW is well understood for a single WLAN where there is no co-channel interference. However, there is no clear indication in the literature on how to configure the minimum CW sizes of the nodes in WLANs to achieve the optimal throughput when co-channel interference exists.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

In this paper, the basic CSMA/CA MAC protocol of IEEE 802.11 and a single transmission rate are considered to address the problem of configuring the minimum CW in WLANs under co-channel interference. Via extensive simulation campaign, it is demonstrated that in a WLAN the optimal value of the minimum CW is 4% to 20% higher than the Bianchi optimal CW due to co-channel interference. However, the optimal throughput is very close to the throughput obtained by the Bianchi optimal CW. The main contribution in this paper is to provide the insight that a near optimal throughput is achievable in WLANs by configuring the minimum CW of each WLAN independently using the simple Bianchi model.

The rest of the paper is organized as follows. System model is presented in detail in Section II. Simulation results and the engineering insights are presented in Section III. This work is concluded in Section IV.

II. System Model

A. Network Topology

Consider a WLAN which has six co-channel WLANs. The WLANs are hexagonal in structure with cluster size of 3. Each WLAN has an AP which is located at the center. The users are randomly distributed in the WLANs and their positions are known. The total number of nodes including the APs in the WLANs is N and their set is denoted by $\mathcal{N}$. Each node of the WLANs has an omnidirectional antenna and cannot transmit and receive at the same time. The transmission power of all the nodes is assumed to be equal and fixed. Only uplink communication is considered in the WLAN system, i.e., only the users access their APs to send their data packets. Each node uses the basic CSMA/CA MAC protocol of IEEE 802.11. The CST for all the nodes is assumed to be fixed and equal. A single rate system is considered where all the users use the same modulation and coding scheme. The users are assumed to be in saturation, i.e., they always have packets to send.

B. Propagation Model

For simplicity, log-normal path loss model is considered in this study. For a transmission of node n_1, the received power at a node n_2 is $P_t G_{n_1,n_2}$, where P_t is the node transmit power and G_{n_2,n_2} is the channel gain between nodes n_1 and n_2. The channel gain is assumed to be quasi time-invariant and G_{n_1,n_2} is given by $(\frac{d}{d_0})^{-\eta}$, where d is the distance between the nodes, d_0 is the far-field cross-over distance and η is the path loss exponent of medium.

C. Interference Model

The *physical interference model* based on signal to interference plus noise ratio (SINR) is considered for wireless interference because it is more realistic interference model for wireless interference [12]. According to this model, a transmission is successful if the received SINR at the receiver is not less than a minimum required SINR. Let the minimum required SINR at the receiver for successful packet decoding is γ. Thus, a packet sent by a transmitter n_1 will be successfully received by a receiver n_2 if

$$\frac{G_{n_1,n_2} P_t}{N_0 + \sum_{n' \in \mathcal{N} \setminus n_1} G_{n',n_2} P_t Y_{n'}} \geq \gamma \tag{1}$$

where N_0 is the noise power and $Y_{n'}$ is a binary variable taking value 1 if node n' transmits and 0 otherwise.

D. Basic CSMA/CA MAC Protocol

In the basic CSMA/CA MAC protocol, each node senses the medium for a time equal to the distributed interframe space (DIFS) before initiating a transmission of a packet. A node n senses the medium as busy if the received power is higher than CST, i.e., if

$$\sum_{n' \in \mathcal{N} \setminus n} P_t G_{n',n} Y_{n'} + N_0 \geq CST. \tag{2}$$

If the medium is sensed idle by a node in DIFS period it transmits its data packet. Otherwise, the node sets a value of CW randomly choosing from a uniform distribution (0, CW_{min}-1), where CW_{min} is the minimum CW of the nodes. After the DIFS time, the time is divided into mini-slots and each node senses the channel in each mini-slot if they do not transmit. If a node senses the channel as busy in a mini-slot, the node does not reduce its CW. If the channel is found as idle in a mini-slot, the node reduces its CW by one. This is called back-off process. Each node continues the back-off process until their CWs become zero. A node transmits its data packet when its CW becomes zero. Since all the nodes randomly chose the values of CWs, more than one node might have the same CW. In that case, they will start transmission at the same time and the success or failure of a transmission will depend on the interference model. The data packet sizes of all the nodes are assumed to be equal. If a data packet transmission is successful, the receiver will send an acknowledgement (ACK) packet after a short inter frame space (SIFS) time. If a node receives an ACK packet which has already transmitted a data packet, it will attempt to send the next data packet according to the procedure mentioned above.

If transmission of a data packet fails, the packet will be retransmitted. During retransmission, the value of CW is randomly chosen from a uniform distribution (0, $2^m CW_{min}$-1), where m is the number of retransmission attempts to send the packet. This is called BEB mechanism. When a data packet (re)transmission is successful and an ACK packet is received by the sender, for the next data packet, CW will be reset by randomly choosing from a uniform distribution (0, CW_{min}-1). If a packet is not successful for five retries, the data packet is dropped.

III. Simulation and Results

In this section, the optimal configuration of the minimum CW of the nodes in WLANs is studied under co-channel interference via simulation.

A. Case Study - Equal Number of Users in Each WLAN

1) Networks and Algorithm Parameters: A WLAN and its six co-channel WLANs (Network-1) are considered as shown in Fig. 1. The radius of each WLAN is taken to be 100 m. The AP of each WLAN is labeled by a rectangular box. The number of users in each AP is taken to be 7. The physical layer and MAC layer parameters are given in Table I. The path loss exponent is taken to be 3.5 considering the environment of indoor. Another network (Network-2) similar to the network-1 is also considered.

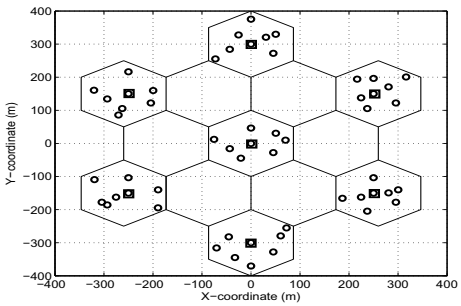


Fig. 1. The positions of the nodes in WLANs (Network-1).

TABLE I
PHYSICAL LAYER AND MAC LAYER PARAMETERS

Layer	Parameter	Value
Physical	SINR threshold, γ (dB)	6.4
	Noise power, N_0 (dBm)	-100
	Far-field crossover distance, d_0 (m)	1
	Data rate, C (Mbps)	2
MAC	SIFS (μs)	10
	DIFS (μs)	50
	a slot time, δ (μs)	20
	Data packet size, L_d (byte)	1058
	ACK packet size (byte)	14
	PHY + MAC header (byte)	24

2) Simulator Setup: A custom simulator written in C++ is used for simulation [11]. The commonly available network simulators are not used since they cannot provide the flexibility of tuning the network parameters. The average rates of the sources are set to be equal values and their traffic is assumed to be Poisson. The source rate of the users is kept sufficiently large to make their queues saturated. To determine the total throughput of the WLANs for a given network scenario, the network is simulated for 500 s and the total throughput of the WLAN is calculated as $1058 \times 8 \times N_{pkt}/500$ bps, where N_{pkt} is the total number of packets received by the APs of the WLANs at the end of the simulation. For a network scenario, the average throughput over five simulation runs is taken to determine the total throughput of the WLANs.

3) Configuration of Transmit Power and CST: The optimal values of transmit power and CST are used to study the optimal configuration of the minimum CWs of WLANs. These values are determined through extensive simulations by varying the node transmit power for different values of CST at the Bianchi optimal CW. For the WLAN system considered in this paper, the collision time is found to be 219 mini-slots using the following equation [1].

$$T_c = \frac{8(L_d + PHY + MAC)}{C\delta} + \frac{DIFS}{\delta} \tag{3}$$

where L_d is the data packet size, PHY and MAC are the physical layer and MAC layer headers, respectively, and δ is the duration of a mini-slot. The Bianchi optimal CW is found to be 146 for the users of each WLAN since the number of contending nodes is 7 in each WLAN. From the simulation results, the optimal values of transmit power and CST power are found to be -18 dBm and -94 dBm, respectively, for the both networks.

4) Optimal CW versus Bianchi Optimal CW: Using the optimal transmit power and CST, the total throughput of the both networks are determined via simulation by varying the minimum CW ranging from 31 to 320. The total throughput of each network is shown in Fig. 2 for different values of the minimum CW. The optimal values of the minimum CWs and the optimal throughputs of the both networks are determined from the figure. The optimal values are compared with the Bianchi optimal values in Table II, where CW_O and CW_B are the optimal value of the minimum CW and the Bianchi optimal CW, respectively and λ_O and λ_B are the optimal throughput and the throughput obtained by the Bianchi optimal CW, respectively. From the results in Table II, it is found that the optimal value of the minimum CW is 15% to 25% higher than the Bianchi optimal CW. However, the difference between the optimal throughput and the throughput obtained by the Bianchi optimal CW is less than 0.5%. Thus, the Bianchi optimal CW is sufficient to achieve a near-optimal throughput.

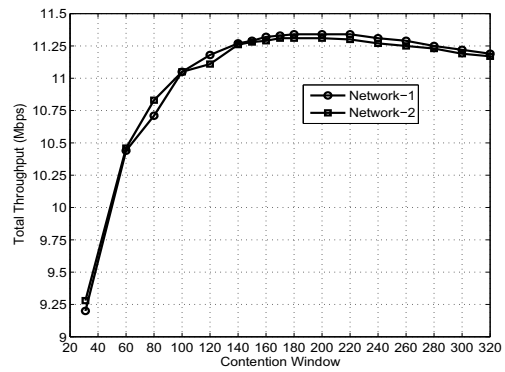


Fig. 2. Throughput performance at different values of the minimum CW.

TABLE II
COMPARISON OF THROUGHPUT PERFORMANCE

Network	CW_O	CW_B	% Diff.	λ_O (Mbps)	λ_B	% Diff
1	180	146	23.28	11.34	11.29	0.44
2	170	146	16.44	11.31	11.27	0.35

B. Case Study - WLANs with Different Number of Users

1) Networks and Algorithm Parameters: Two networks (Network-3 and Network-4) with different number of users in the WLANs are considered to study the impact of co-channel interference on the optimal configuration of the minimum CW in WLANs. One of the networks (Network-3) is shown in Fig. 3, where the WLANs are labeled as WLAN-1, WLAN-

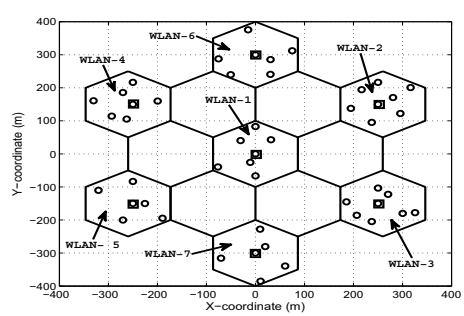


Fig. 3. The positions of the nodes in WLANs (Network-3).

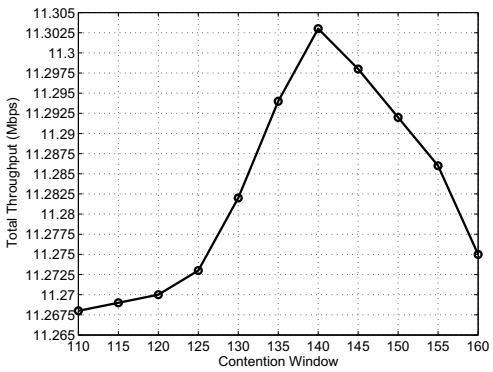


Fig. 4. Total throughput of the WLANs in Network-3 by varying the minimum CW of the users in WLAN-1.

TABLE III
COMPARISON OF THE OPTIMAL VALUE OF THE MINIMUM CW AND THE BIANCHI OPTIMAL CW

	Network-3			Network-4		
WLAN	CW_O	CW_B	% Diff.	CW_O	CW_B	% Diff
1	140	125	12.00	140	125	12.00
2	156	146	6.85	130	125	4.00
3	171	146	17.12	119	104	14.42
4	145	125	16.00	119	104	14.42
5	124	104	19.23	145	125	16.00
6	140	125	12.00	135	125	8.00
7	124	104	19.23	161	146	10.27

2, WLAN-3,... and WLAN-7. The physical layer and MAC layer parameters and the simulator setup are considered as in sub-section III-A. The optimal values of transmit power and CST are considered for the study which are determined via extensive simulations by varying the transmission power of the users at different CST power levels at the Bianchi optimal CW at each WLAN. For Network-3, the optimal values of transmission power and CST are found to be -16 dBm and -92 dBm, respectively. On the other hand, the optimal values of transmission power and CST are -18 dBm and -94 dBm, respectively for Network-4.

2) Optimal CW versus Bianchi Optimal CW: Using the optimal transmission power and CST, Network-3 as well as Network-4 is simulated by varying the minimum CW of a WLAN while keeping the minimum CW of the other 6 WLANs of the network to their Bianchi optimal CW. In Fig. 4, the total throughput of Network-3 is presented for different values of the minimum CW of WLAN-1. The number of users in WLAN-1 of Network-3 is 6 and for this WLAN the Bianchi optimal CW is calculated as 125. From Fig. 4, it is found that the optimal value of the minimum CW is 140 and the optimal throughput is 11.303 Mbps. The total throughput obtained for the Bianchi optimal CW is 11.273 Mbps. Thus, the optimal value of the minimum CW is higher than the Bianchi optimal CW. However, the optimal throughput and the throughput obtained by the Bianchi optimal CW are very close. Similar results are generated for the other WLANs of Network-3 and Network-4. In Table III, the comparison between the optimal value of the minimum CW and the Bianchi optimal CW of the both networks is presented. The comparison between the optimal throughput and the throughput obtained by the Bianchi optimal CW is presented in Table IV. From these tables, it can be observed that the difference between the optimal value of the minimum CW and the Bianchi optimal CW is 4% to 20%. However, the difference between the optimal throughput and the throughput at the Bianchi optimal CW is bounded by 0.5%. Thus, configuration of each WLAN independently with the Bianchi optimal CW provides a near optimal throughput.

TABLE IV
COMPARISON OF THE OPTIMAL THROUGHPUT AND THE THROUGHPUT OBTAINED BY THE BIANCHI OPTIMAL CW

	Network-3			Network-4		
WLAN	λ_O	λ_B	% Diff.	λ_O	λ_B	% Diff
1	11.303	11.273	0.27	11.278	11.277	0.01
2	11.307	11.288	0.17	11.284	11.278	0.05
3	11.291	11.285	0.053	11.286	11.276	0.09
4	11.370	11.320	0.44	11.284	11.275	0.08
5	11.306	11.285	0.19	11.282	11.272	0.09
6	11.292	11.285	0.06	11.281	11.277	0.04
7	11.297	11.285	0.11	11.285	11.277	0.07

IV. CONCLUSIONS

In this paper, the effect of co-channel interference on the optimal configuration of the minimum CWs in IEEE 802.11 based WLANs is studied via simulation. From the simulation results, it is found that the optimal value of the minimum CW is higher than the Bianchi optimal CW due to co-channel interference. However, the difference between the optimal throughput and the throughput at the Bianchi optimal CW is less than 0.5%. As it is very difficult to determine the optimal values of the minimum CWs in WLANs considering co-channel interference, this research advocates to use the simple Bianchi model to configure the minimum CW in each WLAN independently to obtain a very close throughput to the optimal throughput.

REFERENCES

[1] G. Bianchi, "Performance Analysis of the IEEE 802.11 Distributed Coordination Function," *IEEE Journal On Selected Areas in Communications*, vol. 18, vo. 3, March 2000.
[2] IEEE Std. 802.11, "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification," *IEEE*, New York, USA, Nov. 1997.
[3] S. Chieochan, E. Hossain, and J. Diamond,"Channel assignment schemes for infrastructure-based 802.11 WLANs: A survey," *IEEE Communications survey tutorial*, vol. 12, no. 1, First Quarter 2010.
[4] M. Haidar, R. Ghimire, H. Al-Rizzo, R. Akl, and Y. Chan, "Channel assignment in an IEEE 802.11 WLAN based on signal-to- interference ratio," in *Proc. CCECE*, pp. 1169-1173, May-2008.
[5] A. Vasan, R. Ramjee, and T. Woo,"ECHOS: Enhanced Capacity 802.11 Hotspots," in *Proc. IEEE INFOCOM*, 2005.
[6] P. K. Tiwary, N. Maskey, S. Khakurel, and G. Sachdeva, "Effects of co-channel interference in WLAN and cognitive radio based approach to minimize it", in *Proc. ICARTCC*, pp.158-160, 2010.
[7] M. S. Mahmud and M. F. Uddin, "Unfairness problem in WLANs due to asymmetric co-channel interference and its mitigation," in *Proc. IEEE ICCIT*, 2013.
[8] T. Kim, H. Lim, and J. C. Hou,"Improving spatial reuse through tuning transmit power, carrier sense threshold, and data rate in multihop wireless networks," in *Proc. MobiCom*, pp. 23- 26, Sept. 2006.
[9] V. P. Mhatre and K. Papagiannaki,"Optimal design of high density 802.11 WLANs," in *Proc. ACM CoNEXT*, 2006.
[10] J. A. Fuemmeler, N. H. Vaidya, and V. V. Veeravalli, " Selecting transmit power and carrier sense threshold for CSMA protocols," *Technical Report*, Oct. 2004.
[11] M.A. Parvej, S. Chowdhury, N. I. Hia, and M. F. Uddin, "Capture effect on the optimal contention window in IEEE 802.11 based WLANs" in *Proc. IEEE ICIEV*, 2013.
[12] A. Iyer, C. Rosenberg, and A. Karnik, "What is the right model for wireless channel interference?" *IEEE/ACM Trans. on Networking*, vol. 8, no. 5, May 2009.

Performance Analysis of CSMA/CA Based MU-MIMO WLANs

Md. Moin Uddin Chowdhury, Md. Asif Tanveer and Md. Forkan Uddin*
Dept. of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
*mforkanuddin@eee.buet.ac.bd

***Abstract*—In this paper, the performance of carrier sense multiple access with collision avoidance (CSMA/CA) based multi-user MIMO (MU-MIMO) uplink wireless local area networks (WLANs) is analysed. Analytical models are provided to compute the throughput of each user and the total energy consumption in a WLAN with a single user MIMO (SU-MIMO) technique as well as a MU-MIMO technique. By means of the analytical models, the throughput, fairness and energy consumption performances are studied for several WLAN scenarios. The numerical results show that the MU-MIMO technique provides higher throughput than the SU-MIMO technique only when the number of MIMO communication antennas is high and the medium access parameters of the users are configured properly. Further, the SU-MIMO technique is found to be more energy efficient than the MU-MIMO technique. However, the MU-MIMO technique provides better fairness over the SU-MIMO technique.**

***Index Terms*—Medium access control, wireless local area network, multi-user MIMO, interference, throughput, fairness.**

I. Introduction

Multiple-Input Multiple-Output (MIMO) technology is a wireless technology that uses multiple antennas at transmitter and receiver to transfer multiple data streams at the same time. In the recent years, MIMO transmission schemes have been adopted by many systems to achieve high spectral efficiency [1]-[2]. In these schemes, aggregate data rate can be increased by transmitting several independent data streams simultaneously through different antennas, which is called spatial multiplexing. If multiple data streams are originated from a single transmitter and transmitted to a single receiver then it is called single-user MIMO (SU-MIMO) technique. If a transmitter is allowed to transmit only one data stream by one of its antennas, a set of transmitters can transmit to a particular receiver simultaneously such that the receiver can receive multiple streams using its antennas. Such a technique is known as multi-user (MU-MIMO) [3], [4]. Presently, the researchers in wireless communications are showing tremendous interest to the MU-MIMO technique due to its advantages over the SU-MIMO technique.

Wireless local area networks (WLANs) are widely deployed in most of the counties to provide high speed Internet access to the users. The current WLANs are based on IEEE 802.11 standard and use carrier sense multiple access with collision avoidance (CSMA/CA) medium access control (MAC) protocol [5]. Like the others recent standards, the working group of IEEE 802.11 is also preparing MU-MIMO based physical layer standard IEEE 802.11ac to provide data rate up to 7 Gbps [3].

Throughput, fairness and energy consumption are the most important performance metrics in any wireless network. Throughput is a measure of traffic carrying capacity while fairness measures the fair share of system resources among the users. In a WLAN, a user near the access point (AP) may prevent a user located far away from the AP from achieving sufficient throughput. Moreover, with the rapid and radical evolution of information and communication technology (ICT), the total energy consumption in wireless networks is also growing at a staggering rate. So it is necessary to study the throughput, fairness and energy consumption performances in MU-MIMO based WLANs.

In [6], an analytical model is provided by G. Bianchi to calculate the total throughput in a single-input single-output (SISO) based WLAN. In [3], analytical model of total throughput is provided for a MU-MIMO WLAN and it is shown that the MU-MIMO scheme over performs the SU-MIMO scheme in term of total throughput. To the best of our knowledge, fairness and energy consumption performances of CSMA/CA based MU-MIMO WLANs are not yet studied.

To study the fairness among the users in a WLAN, analytical model to compute the throughput of each user of the WLAN is necessary. In [9], Libin and Jean propose a quasi-reversible distributed CSMA model and provide an explicit throughput formula of each user in SISO based multi-hop wireless networks. To study energy consumption in wireless networks, an analytical model is presented in [11] to predict energy consumption in saturated IEEE 802.11 single-hop ad hoc networks under ideal channel conditions. Analytical models of throughput of each user and the total energy consumption for MU-MIMO based WLANs are necessary to study fairness and energy consumption performances.

In this paper, a CSMA/CA based multi-rate uplink MU-MIMO WLAN with zero forcing (ZF) [7] MIMO decoding is considered. The *physical interference model* based on signal to interference plus noise ratio (SINR) is assumed for the study since it is a more realistic interference model for wireless interference [8]. The contributions in this paper are as follows.

- Analytical models are provided to compute the throughput of each user and the total energy consumption in a CSMA/CA based MU-MIMO WLAN.
- By means of the proposed analytical models, numerical results on throughput, fairness and energy consumption are provided for several WLAN scenarios. From the numerical results, the following engineering insights are obtained: (i) the MU-MIMO technique is beneficial in term of throughput only when the number of MIMO communication antennas is high, (ii) the MU-MIMO technique provides better throughput fairness than the SU-MIMO technique, and (iii) to achieve the same total throughput, the energy consumption in the SU-MIMO technique is significantly lower than that in the MU-MIMO technique.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

The rest of this paper is organized as follows. In Section II, the system model is described in detail. The throughput and energy consumption models are developed in Section III. In Section IV, numerical results are presented to compare the performance of the SU-MIMO and MU-MIMO techniques. Section V concludes this paper.

II. System Model

A. Network Topology

A WLAN consisting of U stationary users and an AP is considered. The set of users is given as $\mathcal{U} = \{1, 2, ..., U\}$. The transmission power of each user is P_t. The number of antennas at the AP and at each user is considered to be N. The antennas are omni-directional and cannot transmit and receive simultaneously. Only uplink communication is considered, i.e., only the users transmit data to the AP. It is also assumed that all the users are in saturation, i.e., they always have packets to send.

B. MIMO Schemes

1) SU-MIMO: In this scheme, only one user communicates to the AP at a time through its N antennas. The transmitter first encodes the input bits with one encoding block, generates N data streams and transmits through N antennas to the AP. The transmission power at each antenna for each data stream is P_t/N.

2) MU-MIMO: In this scheme, a user transmits using only one of its antennas and the transmission power at the transmitting antenna is P_t. Thus, the AP can successfully retrieve N data streams from N users at a time.

C. Propagation and Interference Model

The channel gain between the user of data stream i and the AP is given by $g_i = \left(\frac{d_i}{d_0}\right)^{-\eta}$; where, d_i is the distance between the user of data stream i and the AP, η is the path loss exponent, and d_0 is a reference distance in the far field of the antenna. Due to its simplicity, ZF MIMO decoding [7] technique is considered at the AP. Let β be the minimum required SINR at the receiver for successful reception of data stream i. It is assumed that the AP knows the channel gains of the desired data streams. According to [4], the data stream i is successful if $\frac{g_i E\left[|s_i|^2\right]}{N_0 + \sum_{j=1}^{J} g_j E[|s_j|^2]} \geq \beta$; where, $E\left[|s_i|^2\right]$ is the transmitted power of the received signal s_i of the data stream i, $g_j E\left[|s_j|^2\right]$ is the interference power at the AP from a undesired data stream j and N_0 is the background noise power.

D. Medium Access Control Protocol

The basic CSMA/CA protocol is considered for MAC operation where all the stations enter a back-off state when the shared channel is sensed busy in a specific distributed inter frame space (DIFS) time. Each station randomly chooses an integer value between 0 and (CW-1) as back-off counter value for a data packet, where CW is the contention window size. The initial value of CW is set to a minimum value CW_{min}. The time is divided into mini-slots after the DIFS period. A station decreases its back-off counter value by one whenever the medium is sensed idle in a mini-slot time. It keeps the back-off counter value unchanged if the medium is sensed busy in a mini-slot. The carrier sensing threshold of the nodes is set to such a value that all the other nodes sense the medium as busy if a node transmits. A station transmits its data stream(s) according to its MIMO scheme when its back-off counter reaches to zero. For simplicity of analysis, it is assumed that a station knows immediately whether its transmitted data stream(s) is successful or not. This assumption is valid since acknowledgement (ACK) packet transmissions by the AP will be always successful. If a packet transmission fails it is retransmitted and hence, the value of CW is chosen according to the binary exponential back-off mechanism of IEEE 802.11 standard [5].

E. Rate Selection for a Data Stream

It is assumed that a user can transmit at R different rates. The set of the rates $\mathcal{R}$ is given by $\mathcal{R} = \{r_1, r_2, ..., r_R\}$. The set of the required SINRs for the available data rates is given by $\mathcal{B} = \{\beta_1, \beta_2, ..., \beta_R\}$, where β_j is the SINR threshold for rate r_j. It is assumed that a user selects the highest possible rate for a data stream based on the signal to noise ratio (SNR) of the data stream at the AP, i.e., the rate of a data stream i is r_j if $\beta_j \leq \frac{g_i E\left[|s_i|^2\right]}{N_0} \leq \beta_{j+1}$. The transmission times of the data packets of each user are considered to be equal to L mini-slots regardless the data transmission rate.

III. Models for Throughput and Energy Consumption

A. Discrete Time Markov Chain Model of the CSMA/CA MAC Protocol

In the CSMA/CA MAC protocol, time is divided into equal length mini-slots or states where transition from one state to another state forms a discrete time Markov chain (DTMC). The steady state probability of transmission of each user in a mini-slot is considered as τ if the medium is idle. The relation between the steady state probability τ and CW_{min} can be approximated as $\tau = \frac{2}{CW_{min}}$ [6]. A state w of the CSMA/CA system can be expressed as $w := \{v, l_k \quad \forall k \in v\}$, where, v is the set of transmitting users and $l_k \in \{1, 2, 3, \ldots, L\}$ is the slot number of the transmitting data stream(s) of user k. The stationary distribution of a state w can be obtained as [9]

$$p(w) = \frac{1}{E}\tau^{|v|}(1-\tau)^{U-|v|} \tag{1}$$

where, $|v|$ is the cardinality of set v and E is the normalizing term given by $\sum_w p(w) = 1$.

The null set of v is denoted by ϕ. Due to carrier sensing mechanism of the CSMA/CA protocol, a set of users will start their transmissions simultaneously and end the transmissions at the same time. Since there are L states for a packet transmission and the probability in each state is equal, the fraction of time that the users of set v, $v \neq \phi$ transmit simultaneously is given by

$$\pi(v) = \frac{L\tau^{|v|}\left(1-\tau\right)^{U-|v|}}{E}. \tag{2}$$

Note that $\pi(\phi) = \frac{(1-\tau)^U}{E}$. Let $\mathcal{I}$ be the possible sets of v in the WLAN including the null set ϕ. The normalizing term E can be obtained as $E = L\sum_{v\in\mathcal{I}\setminus\phi}\tau^{|v|}\left(1-\tau\right)^{U-|v|} + \left(1-\tau\right)^U$. Since $\sum_{v\in\mathcal{I}}\tau^{|v|}\left(1-\tau\right)^{U-|v|} = 1$, E can be obtained as

$$E = L - (L-1)(1-\tau)^U. \tag{3}$$

B. Throughput Model

Let the sets of user-set v which contains the user k is $\mathcal{I}_k$ and the sets of user-set in which the transmissions of user k are successful is $\mathcal{I}_k^s \subset \mathcal{I}_k$. Thus, the throughput of user k is given as

$$\lambda_k = C_k \sum_{v \in \mathcal{I}_k^s} \pi(v). \tag{4}$$

where, C_k is the data rate of user k. This model can be used for the both MIMO schemes and hence, the difference will be in the set $\mathcal{I}_k^s$ and data rate C_k. The main complexity in computing the throughput of a user k is to enumerate all the successful user-sets $\mathcal{I}_k^s$. Under the exhaustive search method, for the both MIMO schemes, one has to check the condition of successful transmission of data stream(s) of a user for $\sum_{i=N'}^{U-1} \binom{U-1}{i}$ user-sets using the condition in Section II-C, where N' is number of data streams transmitted by user k depending on the used MIMO scheme. The value of C_k is given by $N'c_k$; where, c_k is the data rate of a data stream of user k. The value of c_k can be determined according to Section II-E. Thus, the complexity of computing the throughput of a user in a WLAN increases exponentially with the number of users.

C. Energy Model

Two modes are considered for a node: IDLE and TRANSMIT. In the IDLE mode, a node senses the carrier in each mini-slot and in the TRANSMIT mode, a node transmits data stream(s). Let $\frac{P_t}{\alpha}$ be the power consumption when a node is in the IDLE mode, where α is ratio of the energy consumptions in the TRANSMIT and IDLE modes. It is assumed that the energy consumption does not depend on data transmission rate. Further, the energy required for processing signal of a data stream at transmitter and receiver is neglected. The energy consumption per second for the transmissions by a user-set v is given as

$$E_v = \left(|v| P_t + (U - |v|) \frac{P_t}{\alpha} \right) \pi(v). \tag{5}$$

The total energy consumption per second is then given as

$$E = \sum_{v \in \mathcal{I}} E_v = \sum_{v \in \mathcal{I}} \left(|v| P_t + (U - |v|) \frac{P_t}{\alpha} \right) \pi(v). \tag{6}$$

Approximating $(\alpha - 1) \approx \alpha$, (6) can be written as

$$E \approx \frac{U P_t}{\alpha} + P_t \sum_{v \in \mathcal{I}} |v| \pi(v). \tag{7}$$

The number of user-sets in $\mathcal{I}$ is 2^U and hence, the computation complexity of energy computation increases exponentially with the number of users. For a given transmission probability, $\pi(v)$ is the same for the SU-MIMO and MU-MIMO schemes for a user-set v, thus the total energy consumption is the same for the both MIMO schemes for a given MAC parameter and it is independent of the number of MIMO communication antennas.

IV. Numerical Results

A. Networks and Parameters

Five WLAN scenarios each consisting of 19 users and an AP are considered. The AP is located at the centre surrounded by the users. The positions of the users are determined by a uniform random distribution. One of the WLAN scenario (WLAN-1) is shown in Fig 1, where the AP is labelled as a rectangular box. It is assumed that for a data stream a user can use one of the 10 modulation and coding schemes in an 80 MHz channel depending on the SNR of the data stream at the AP. The data rates and the SNR thresholds for the different modulation and coding schemes are illustrated in Table I. The transmission power and the noise power are considered to be 10 dBm and -100 dBm, respectively. Considering the indoor WLAN environment, the path loss exponent and the far field cross-over distance are assumed to be 4 and 1 m, respectively. The transmission time of each packet is taken to be 200 mini-slots.

TABLE I
DATA RATES AND THE CORRESPONDING SNR THRESHOLDS.

Data Rate (Mbps)	Modulation	Coding Rate	SNR Threshold (dB)
15	BPSK	1/2	8
30	QPSK	1/2	11
45	QPSK	3/4	13
60	16-QAM	1/2	16
90	16-QAM	3/4	20
120	64-QAM	2/3	24
135	64-QAM	3/4	25
150	64-QAM	5/6	26
180	256-QAM	3/4	30
200	256-QAM	5/6	32

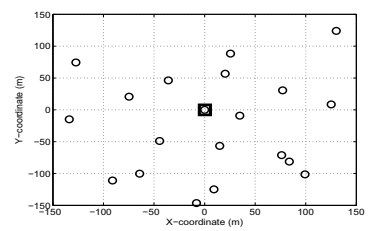


Fig. 1. Positions of the nodes in WLAN-1.

B. Comparison of Total Throughput

For each of the WLAN scenarios, the total throughputs are computed for different transmission probabilities by using the throughput model under the two MIMO schemes. The total throughput of each WLAN scenario is shown in Fig. 2, where the same marker is used for a WLAN scenario for different MIMO schemes and different number of MIMO communication antennas. It is observed that the SU-MIMO scheme provides better throughput than the MU-MIMO scheme when the number of MIMO communication antennas is 4. However, the optimal throughput of the MU-MIMO scheme is higher than that of the SU-MIMO scheme when the number of MIMO

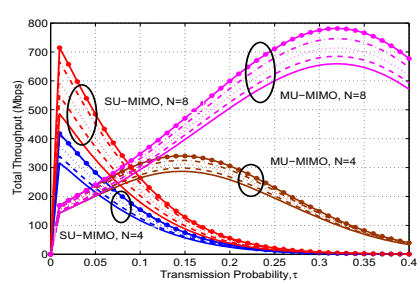


Fig. 2. Comparison of total throughput of the SU-MIMO and MU-MIMO schemes.

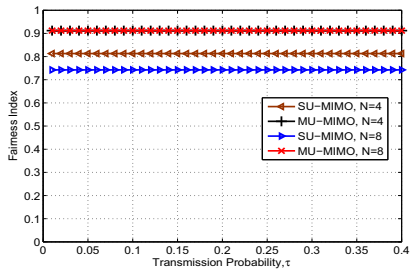


Fig. 3. Comparison of fairness index of the SU-MIMO and MU-MIMO schemes for WLAN-1.

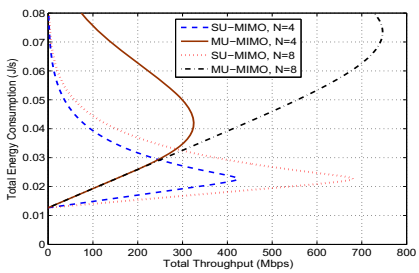


Fig. 4. Total energy consumption in the SU-MIMO and MU-MIMO schemes for WLAN-1.

communication antennas is increased to 8. The throughput optimal transmission probability of the MU-MIMO scheme is higher than that of the SU-MIMO scheme which increases significantly for the MU-MIMO scheme with increasing the number of MIMO communication antennas. This is due to the fact that a higher number of antennas at the AP allows it to receive a higher number of data streams simultaneously from a higher number of users and hence, a higher transmission probability causes to transmit a higher number of users at the same time.

C. Comparison of Fairness

To investigate the fairness among the users, for the both MIMO schemes, the throughput of each user of the WLAN-1 is computed and then the Jain's fairness index (FI) [12] is calculated as $FI = \frac{\left(\sum_{k\in\mathcal{U}} \lambda_k\right)^2}{U \sum_{k\in\mathcal{U}} \lambda_k^2}$. The FI of WLAN-1 with transmission probability is depicted in Fig. 3. It is observed that the FI of the WLAN does not change with transmission probability for the both MIMO schemes. This is due to the fact that physical layer capture is absent in the WLAN for using a large number of data rates. Further, the FI of the MU-MIMO scheme is higher than that of the SU-MIMO scheme. This is due to the fact that, in the SU-MIMO scheme, the users located close to the AP obtain the highest data rate and contribute mostly in the total throughput. On the other hand, in the MU-MIMO scheme, the contribution of each user in total throughput is more or less very close in value since multiple users can transmit their data streams successfully at the same time. Similar fairness results are also generated for the other 4 WLAN scenarios and found the similar insights.

D. Comparison of Energy Consumption

The total energy consumption of the WLAN-1 is computed for different values of transmission probability under the both SU-MIMO and MU-MIMO schemes. The value of α is assumed to be 15 in computation of energy. Since for a given transmission probability, the energy consumption is the same for the both MIMO schemes and the total throughput is different, the energy consumptions of the both MIMO schemes to achieve the same throughput are calculated for the WLAN-1 and plotted in Fig. 4. From the plots it can be observed that the SU-MIMO scheme is more energy efficient than the MU-MIMO scheme. We attribute this to the fact that the same throughput is achieved by the MU-MIMO scheme at a higher value of transmission probability (See Fig. 2) and the energy consumption increases with increasing the value of transmission probability for increasing the attempt rate of the users. Similar results are also generated for the other 4 WLAN scenarios and found the similar conclusions.

V. Conclusion

In this paper, the total throughput, fairness and energy consumption performances are studied in CSMA/CA based uplink WLAN under SU-MIMO and MU-MIMO techniques. Analytical models are provided to compute the throughput of each user and the total energy consumption of a WLAN. Through numerical results of several WLAN scenarios, it is demonstrated that the MU-MIMO technique is more beneficial than the SU-MIMO technique in term of throughput only when the number of MIMO communication antennas is high and the MAC parameters of the users are chosen properly. However, the energy consumption increases significantly for achieving the higher throughput with the MU-MIMO technique.The MU-MIMO technique provides better fairness over the SU-MIMO technique.

References

[1] T. Kailath and A. Paulraj, "Increasing capacity in wireless broadcast systems using distributed transmission/directional reception," US Patent 5345599, Sept. 1994.

[2] K. Raoof, M. Khalighi, and N. Prayongpun, "MIMO systems: Principles, iterative techniques and advanced polarization," in *Adaptive Signal Processing for Wireless Communication*, CRC Press, vol. 2, pp. 95-134, Aug. 2008.

[3] H. Jin, B. C. Jung, and D. K. Sung, "A trade off between single-user and multi-user MIMO schemes in multi-rate uplink WLANs," *IEEE Trans. on Wireless Commun.*, vol. 10, no. 10, pp. 3332-3342, 2011.

[4] M. F. Uddin, "Joint routing and scheduling in fixed multi-user MIMO wireless multi-hop networks," in *Proc. IEEE/CIC ICCC*, pp. 745-751, 2013.

[5] IEEE Std. 802.11, "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification," *IEEE*, New York, USA, Nov. 1997.

[6] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE Journal on Selected Areas in Communications,* vol. 18, no. 3, pp. 535-547, March 2000.

[7] C.-J. Chen and L.-C. Wang, "On the performance of the zero-forcing receiver operating in the multiuser MIMO system with reduced noise enhancement effect," in *Proc. IEEE GlobeCom*, 2005.

[8] J. Luo, C. Rosenberg and A. Girard, "Engineering wireless mesh networks: Joint scheduling, routing, power control, and rate adaptation," *IEEE/ACM Trans. on Networking*, vol. 18, no. 5, pp. 1387-1400, 2010.

[9] L. Jiang and J. Walrand, "Approaching throughput-optimality in a distributed CSMA algorithm with contention resolution," *Technical Report* UCB/EECS-2009-37, EECS Department, University of California, Berkeley, Mar. 2009.

[10] A. Iyer, C. Rosenberg and A. Karnik, "What is the right model for wireless channel interference?" *IEEE Trans. on Wireless Commun.*, vol. 8, no. 5, pp. 2662-2671, May 2009.

[11] M. Carvalho, C. Margi, K. Obraczka and J. J. Garcia-Luna-Aceves, "Modeling energy consumption in single-hop IEEE 802.11 ad hoc networks," in *Proc. International Conference on Computer Communications and Networks (ICCCN)*, pp. 367-372, 2004.

[12] R. Jain, D. M. Chiu and R. W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer system," *Technical Report* DEC-TR-301, Digital Equipment Corporation, Maynard, Mass, USA, 1984.

HETNET Neighbour Handover Cell List Optimization Based On Cognitive Programmed Threshold Response

Sanyat Hoque, Shovon Pal, and Atiqur Rahman

Department of Electrical and Computer Engineering
North South University, Dhaka, Bangladesh

sanyat.hoque@gmail.com, shovon.pal.bd@ieee.org, atiqur@northsouth.edu

Abstract—- **The growth of mobile broadband has unleashed the necessity of advance mobile network structure. Heterogeneous Network (HetNet) technology is designed to be highly adaptable in subscribers' homes to provide high data rate communications with good quality of service. Femtocells are the indoor nodes of HetNet. Efficient handover of calls is the priority for successful femtocell/macrocell joint transmission. Handoff of femtocell-to-femtocell handover is the most visible challenge for dense deployment of nodes. This particular handover needs minimum as well appropriate neighbouring cell list which is needed for its successful operation. In this paper, we proposed an algorithm for dense femtocells deployment, i.e., an algorithm to create an optimized neighbour cell list with minimum number of femtocells (including the reduction of interfering femtocells and finding hidden ones). An appropriate and optimal neighbour femtocell list is mandatory for the much needed dense femtocellular network deployment in the near future. Therefore, reducing the size of the neighbour femtocell and making the list more efficient results in minimizing the amount of scanning and signal flow of mobile station during handover.**

Keywords: Handoff, HetNet, Femtocell, SON.

I. INTRODUCTION

The demand for high data rate for wireless communication is increasing in exponential growth [1]. Existing mobile data communication systems faces various challenges [2] to support high wide band data access. Both the coverage area and the capacity of existing cellular network systems are not sufficient to meet the expected demand of multimedia traffics. Closer transmitter and receiver of a wireless system cause increase in capacity of a wireless link and creates dual benefits of higher quality links and more spatial reuse. So among many approaches femtocell approach is one of the best approach to diverse the load from the cellular networks as well as to reduce the operating and capital expenditure costs for operators. In Heterogeneous network environment of the mobile communication network design, femtocells are taken as a home base station solution of the data offloading problem and coping up with the OPEX problem.

The HETNET Scenario comprises with macrocell along with the small cell nodes like femtocell, picocells, etc. The femtocells are enabled with SON capability [3]. Their special care with respect to SON(Self-Organizing Networks) shall be taken for mass deployment scenarios like in the case of Home eNodeB, i.e. for In and Outbound Mobility for Home eNodeB and problems caused by incorrect behaviour of Home eNodeBs. With the deployment of the Home eNodeB, the handover between femtocell and 3GPP macrocell networks is becoming more and more important, mostly in the LTE based networks. Thousands of femtocells within a macrocell area will create a large neighbour cell list and interference problem where the modifications of handover procedures for existing networks are needed.

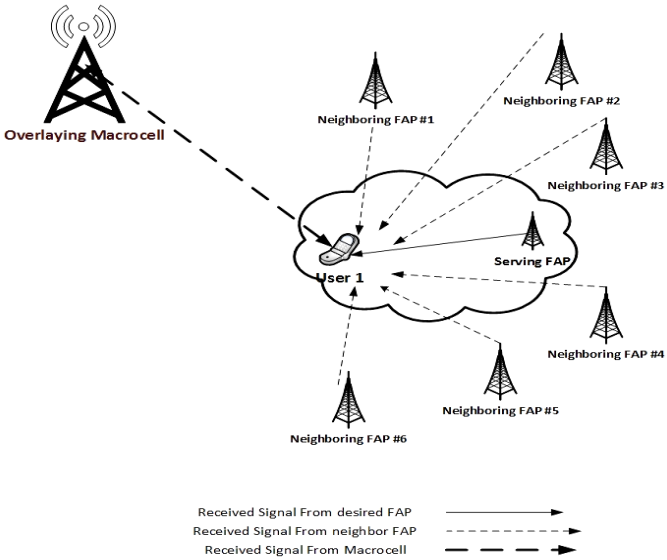


Fig. 1. Dense deployment scenario of eNodeB

In figure 1, the dense deployment scenario has been illustrated of dense deployment of small cell femtocell (eNodeB) in an under-laid macrocell environment. With the received signal indicator, the problem of unplanned deployment has been shown with a perspective view of one user. With the interference problem [4], the Handoff problem emerges as well. With the localization change of the user the neighbour cell list expands correspondingly and the efficiency of handoff decays-so the QoS of network experiences bad data signal from the adjacent femto access points (FAPs). The remainder of this paper is organized as follows: The related works are discussed in Section II. We present the system model and analyse the proposed scheme in Section III. In Section IV, cell association and handover management algorithms are proposed. The performance evaluation and analysis is shown in Section V. Section VI concludes this paper.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

II. RELATED WORK

Deployment of femtocells at a large scale is the final objective of this technology. Zhang et al. [5] presented a handover optimization algorithm based on the UE's mobility state. Soldani and Ore report results on self-optimization of NCR lists for UTRA FDD networks using DSR measurements [6]. Traditional schemes (e.g., [6,7]) based on the received signal strength indicator (RSSI) are used for the existing cellular system. But none of them provided a clear definition of optimized neighbour cell list for HeNodeB to HeNodeB handover. The neighbour eNodeB list based on only the RSSI will contain a large number of femtocells in the list. Features like self-organizing network (SON) capability and FAP management system should be added to the HetNet for improved performance.

III. SYSTEM MODEL

The existing macrocellular networks utilize centralized RNC to control their associated macro base stations (BSs). One RNC is in charge of radio resource management of approximately 100 macro BSs. Femtocells are deployed within a macrocell coverage area or within a separate zone. The recent change in the network optimization, ie. small cells like femtocells, being overlaid within the macrocell network for the purpose of data offloading and serving the cell edge users with a better data quality. This change is named as Heterogeneous network or the HETNET. In figure 1, there is an illustration of the HETENT network with a reference user in the network mode-under the couple of unplanned FAP (Femto access points) under laid in the macrocell. Initially the user is in the region of FAP#1 but as long as the user is not static it will experience the interference from the adjacent FAPS and strong interfering signal from the macro cell as well. This interference issue is not the leading topic in the research. Along with the interference the Hand off issue arises in a big questionable fact.

In a dense HETNET deployment, a lot of femtocells are deployed within small coverage area. Therefore, there may be huge interference effects. Whenever a mobile station (MS) realizes that the receive signal from the serving femto AP (FAP) is going down, the MS receive many signals from the several neighbour FAPs for the handover. The hidden FAP problem arises when a neighbour FAP is very close to the MS but the MS cannot receive the proper amplitude of signal it is supposed to due to a barrier between the MS and that FAP. So, that particular hidden FAP will be out of neighbour femtocell list if the neighbour femtocell list is designed based on the receive signals only as in the case of traditional schemes. In the case of the use of same frequency by surrounding T-FAPs and S-FAP, the corresponding T-FAPs should be avoided to avoid interference. Femtocells that uses the same frequency that matches with S-FAP should only be used that is located far enough apart. In a more recent scheme [8], two threshold levels are used in order to determine the NCL. The first threshold (S_{T0}) is a pilot signal at -90 dBm that is used to detect FAPs in the surrounding area. The second threshold, S_{T1} is higher (-75 dBm) than the first threshold which is responsible for adding the corresponding T-FAPs in the NCL. Use of stationary thresholds in these schemes prevent many potentially important FAPs to be included in the final NCL.

IV. OPTIMIZED NEIGHBOUR FEMTOCELL LIST

In my proposed algorithm, value of upper threshold depends on the density of femtocells keeping -90 dBm as the minimum signal for femtocell detection in coverage area. Therefore, by increasing the value of S_{T1} with the increasing density of femtocells (S_{T2}), the number of femtocells in the neighbor cell list can be reduced yet optimized. Chowdhury et al.[8,12] proposed a scheme of neighbor cell list based on reduction of femtocells using the same frequency as the serving FAP to reduce Co channel interferences along with the inclusion of hidden FAP detection based on the increased threshold compared to the threshold of Traditional Scheme.
In my proposed scheme, the approach starts with the RSSI level indicator from the neighbouring FAPs. The FAPs that are initially listed as the neighbour femtocell list based on the received RSSI level can be expressed as set N_1:

$$N_1 = |\{...FAPi(RSSIi),...:1 \leqslant i, RSSI_i \geqslant S_{T2}\}| \quad ; \quad (1)$$

where, FAPi(RSSIi) represents that *i-th* neighbour FAPs from where the received RSSI level is greater than or equal to S_{T2}. During the deployment of dense femtocellular network, neighbouring FAPs using the same frequency as the S-FAP should be cut off from NCL to prevent frequency interference. So from the neighbor femtocell list we deduct those femtocells which use same frequency as the serving femtocell. The femtocells those can be categorized in this group can be expressed as set N_2:

$$N_2 = |\{...FAPk(f_k),...:1 \leqslant k, N_2 \subseteq N_1, RSSI_k \geqslant S_{T2}\}| \quad ; \quad (2)$$

where, FAPk (f_k) is the *k-th* neighbor femtocell that uses the same frequency as S-FAP. Hidden FAPs are femtocells that are very close to Serving Femtocell but cannot be detected due to an obstacle. The hidden FAP problem is tackled using SON based coordination between S-FAP and its surrounding neighbouring FAPs to find out location of the hidden FAPs. Hidden femtocells are chosen where the received RSSI level is less than S_{T2}, or femtocells using the same frequency as the serving femtocell. Since the serving FAP can coordinate with nearest FAPs, those femtocells can inform the location of some hidden FAPs. Therefore, the hidden femtocell nodes within a specific range of distance can be included in the neighbour femtocell list. The femtocells those can be categorized in this group can be expressed as set M:

$$M = |\{...FAPm(RSSI_m, f_m, d_m),...:1 \leqslant m, ((RSSI_m < S_{T2}) \vee (fs \cup fm = fs)) \wedge (d_m \leq d_{max})\}| \quad ; \quad (3)$$

where, d_{max} is the predefined maximum distance between MS and mth FAP that is set by Mobile Operator depending on propagation environment.

Based on the new proposed cognitive algorithm depending on femtocell density in a coverage area, proposed neighbour femtocell list is as follows:

$$N_f = N_1 - N_2 + M \quad ; \quad (4)$$

Similarly, the corresponding exponential polynomial function is given below:

$$S_{T2} = \{-41.5 \times exp(-0.019n)\} - T_{\max} \quad ; \quad (5)$$

where,

n = Number of Femtocells in its coverage vicinity,
T_{max} = Maximum Threshold Mobile Operator can allow (-50 dBm is taken to be the maximum in this model scenario).

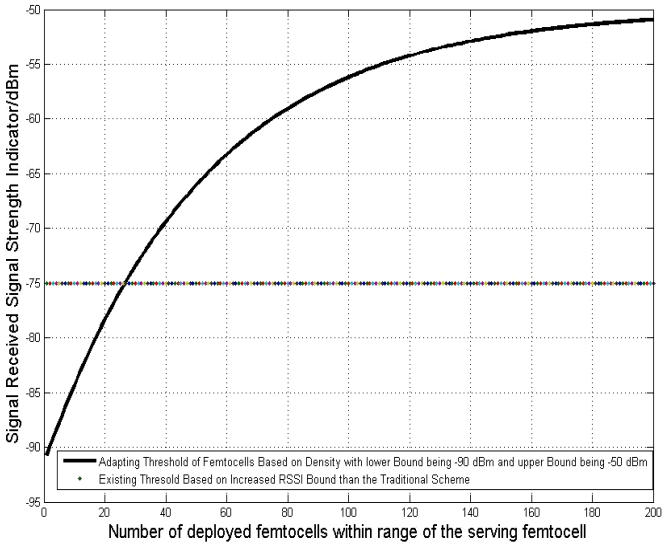


Fig. 2. Maximum and Minimum Threshold of RSSI that MS would accept.

The algorithm is illustrated in Table. 1 below:

Table. 1. Optimized Cell list algorithm

1. **Start**
2. Detection of downwards signal from S-FAP
3. While Checking signal level of #N femtocells
   if $N_f = N_f - 1$ exclusion for set N1, added on M set
   else Frequency of potential T-FAP is checked whether it is the same as S-FAP.
   else if $N_f = N_f$ Kth FAP included in N1 list
4. Checking # neighbor Femtocells of RSSI level ≤ ST2
   if d ≤ dmax, include the FAP in Set M
5. Add Set M with N_f $(N_1 - N_2)$, $N_f = N_f + M$
6. **end**

This cognitive programmed threshold density-based response function (S_{T2}) as shown in figure 2 is an exponential polynomial function that gives the final upper bound to –50 dBm (T_{max} or any value set by the operator) and makes the lower bound to be –90 dBm depending on femtocell detection in its coverage vicinity. The variable threshold rapidly rises from very low threshold of RSSI to accommodate closer femtocells with greater signal power and then passes the mark of –75 dBm rapidly after acquiring fairly adequate number of femtocells in its cell list while keeping the RSSI for femto detection constant at –90 dBm. Only after acquiring a decent number of T-FAPs in its neighbour femtocell list, threshold rises and eventually stops at –50 dBm. This is done so that new femtocells can be added by the mobile operator through increasing the transmission power of any particular T-FAP to more than –50 dBm. If increasing the desired T-FAP's propagation power and thus its RSSI level to S-FAP becomes more than –50 dBm, then that T-FAP would still be included in the neighbour femtocell list despite having significant number of femtocells that has been already added to neighbour cell list.

V. THE PERFORMANCE EVALUATION AND ANALYSIS

For simulation model, four carrier frequency were used (0.7 GHz, 1.8 GHz, 2.1 GHz, 3 GHz) which were distributed randomly along with the Poisson distribution of femtocells around the simulation area of 1000 m. Call arriving processes are assumed as Poisson. Also, a random distribution of hidden femtocells is assumed.
The indoor propagation path loss prediction model for ITU model in dBm [9], [10], [11] is:

$$PL = 20 \log_{10}(f) + N \log_{10}(d) + L_f(n) - 28 \text{ dB} \quad ; \quad (6)$$

Distance power loss coefficient, N is the quantity that expresses the loss of signal power with distance. Floor penetration loss, L_f (n), where n ranges from 1-3.

Table. 2. Parameters used for simulation

- Radius of femtocell coverage area 10 m
- Carrier frequency for S-FAP 1.8 GHz
- Maximum transmit power by a FAP 10mW
- Number of deployed femtocells in scenario 200
- Standard deviation for lognormal shadowing loss 4
- Penetration loss 20 dB
- Power Loss Coefficient Values, N 30
- Coverage Area 1000 m

Referring to coefficient values, the path loss propagation of an indoor environment is being simulated and shown in figure 3. Based on this propagation model, we extend our result for the proposed algorithm for optimized cell list.

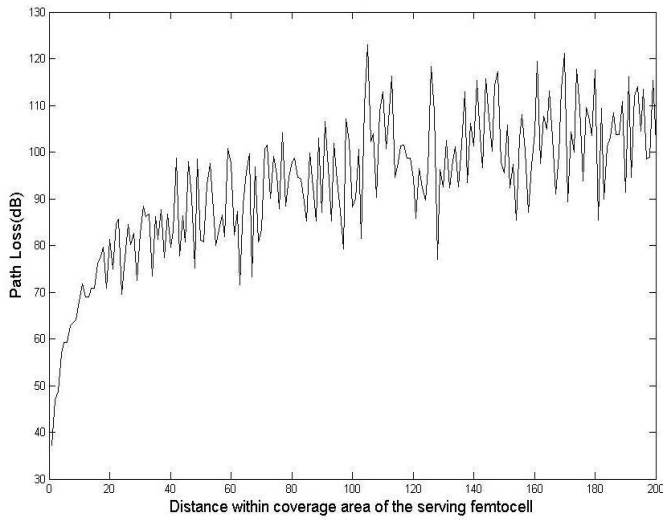


Fig. 3. Path loss with respect to distance between S-FAP and T-FAPs.

In the simulation in figure 4, we compare the performance of the traditional neighbour cell list management scheme [6,7] and the existing one [8,12] versus the proposed scheme. The result shows that neighbour femtocell list based on the proposed scheme contains only a limited number of femtocells during the handovers despite increasing number of FAPs in coverage area. The new algorithm completely optimizes the neighbour cell list. However, reducing number of femtocells in the neighbour femtocell list does not hamper handover efficiency. The new scheme rapidly adds a satisfactory number of femtocells but excludes most of the

unnecessary FAPs through: improved reduction of Co-channel interference femtocells and improved algorithm of finding Hidden FAPs, based on coverage area's density. This makes the algorithm more efficient and effective for handoff.

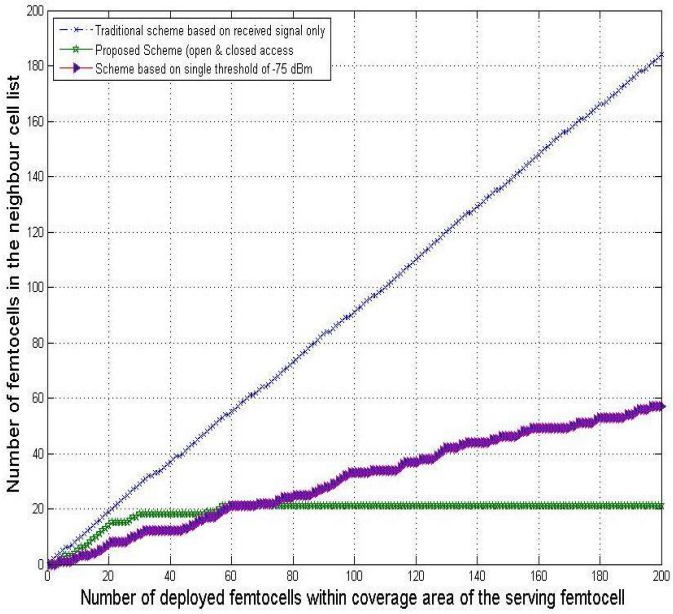


Fig. 4. Neighbour femtocell listing comparison.

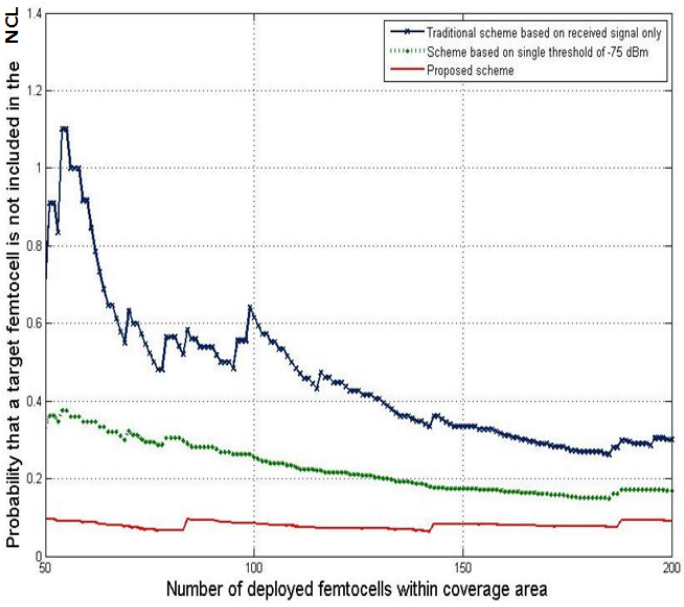


Fig. 5. Probability comparison that a target femtocell is not in its NCL.

Also, expected probability of missing target femtocell from neighbour femtocell list in traditional scheme is higher than the algorithm proposed by Chowdhury and Jang [8,12]. The proposed scheme improves the reduction of missing target femtocell probability by tackling and improving the co-channel interference problem and the Hidden FAP issue. The problem of hidden FAP is tackled in proposed scheme by checking for signals that are less than S_{T2} (i.e., less than -90 dBm initially as shown in fig. 2) at a threshold of a very small distance (10 m for this simulation model where maximum power loss occurs).

VI. Conclusions

In this paper, we proposed a novel approach to solve the mobility management concern for HETNET with dense deployment of small cells. The proposed SON-based algorithm is capable of handling large numbers of FAPs inside its coverage. On top of that, our proposed algorithm helps in improving to overcome the hidden FAP problem. The reduced neighbour cell list results in reduced power loss as well as reduced MAC overhead. The proposed handover neighbour cell list can be very effective to implement for handover processes in dense HETNET deployment.

Results found in this paper adequately show the advantage of our proposed schemes.

References

[1] S. Yeh, S. Talwar, "WiMAX femtocells: a perspective on network architecture, capacity, and coverage," *IEEE Communication. Magazine*, vol. 46, no. 10, pp. 58–65, Oct. 2008.

[2] R. Y. Kim, J. S. Kwak, K. Etemad, "WiMAX femtocell: requirements, challenges, and solutions," *IEEE Communication Magazine*, vol. 47, no. 9, pp. 84–91, Oct. 2009.

[3] M. Z. Chowdhury, Y. M. Jang, Z. J. Haas, "Network evolution and QoS provisioning for integrated femtocell/macrocell networks," *International Journal Wireless Mobile Network*, vol. 2, no. 3, pp. 1–16, Sept. 2010.

[4] M. Z Chowdhury, Y. M. Jang, Z. J. Haas, "Cost-effective frequency planning for capacity enhancement of femtocellular networks," *Wireless Personal Communication*, vol. 60, no. 1, pp. 83–104, Sept. 2011.

[5] H. Zhang, W. Ma, W. Li, W. Zheng, X. Wen, C. Jiang, "Signalling cost evaluation of handover management schemes in LTE-advanced femtocell", in Proceeding of *IEEE Vehicular Technology Conference (VTC Spring)*, pp. 1–5, May 2011.

[6] D. Soldani, I. Ore, "Self-optimizing neighbor cell list for UTRA FDD networks using detected set reporting", in *Proceeding of IEEE Vehicular Technology Conference (VTC Spring)*, Dublin, pp. 694–698, 2007.

[7] H. Olofsson, S. Magnusson, M. Almgren, "A concept for dynamic neighbor cell list planning in a cellular system," in *Proceeding of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, vol. 1, pp. 138–142, Taipei, 1996.

[8] M. Z. Chowdhury and Y. M. Jang, "Handover management in high-dense femtocellular networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 6, 2013.

[9] Femtoforum, "OFDMA Interference Study: Evaluation Methodology Document," 2008.

[10] T. Chrysikos, G. Georgopoulos, S. Kotsopoulos, "Site-Specific Validation of ITU Indoor Path Loss Model at 2.4 GHz," *IEEE Int. Sym. on World of Wireless, Mobile and Multimedia Networks & Workshops*, pp.1-6, 2009.

[11] S. Hameed, "Indoor Propagation Modelling for Wireless Local Area Network (WLAN)", *IJCCCE*, vol. 11, no. .1, 2011.

[12] M. Z. Choudhury, B. M. Trung, and Y. M. Jang, "Neighbor Cell List Optimization for Femtocell-to-Femtocell Handover in Dense Femtocellular Networks," *International Conference on Ubiquitous and Future Networks (ICUFN)*, pp 242-245, Dalian, June 2011.

Determination of Energy Efficiency of a Multi-User Wireless Network Based on Limited User Traffic Model

Md. Nurul Islam,[1] Shamim Ahmed,[1] Ummy Habiba, [1] Md. Imdadul Islam, [2] and M. R. Amin, [1,*] *Member, IEEE*

[1]Department of Electronics and Communications Engineering, East West University
Jahurul Islam City, Aftabnagar, Dhaka 1219, Bangladesh
[2] Department of Computer Science and Engineering, Jahangirnagar University
Savar, Dhaka 1342, Bangladesh
[*]raminrifat@gmail.com

Abstract— **In recent days intense research is taking place on energy efficiency (EE) of a multi-user wireless network. In this paper we propose a model to detect EE based on limited user traffic model. We observe the variation of EE against the number of users and energy signal to noise ratio engaged by the user. The impact of enhancement of subcarriers on EE and channel capacity is also analyzed to reach an optimum condition of network performance. Finally satisfaction index (SI) is plotted against signal-to-noise-ratio (SNR) for different bandwidth of subcarriers.**

Index Terms— **Energy efficiency, OFDMA, satisfaction index, Markov chain, Engset traffic.**

I. Introduction

Mobile and wireless communication system has flourished over the last few years as a result of the increasing demands for high data rates and for supporting large number of users with guaranteed quality of service (QoS). With the aim to provide higher data rates and to optimize radio resource management, next generation wireless communication systems such as Wimax and 3GPP LTE has emerged recently. The LTE system has been designed based on the orthogonal frequency division multiple access (OFDMA) where the channel bandwidth is divided into large number of orthogonal subcarriers for multi-access purposes and effective reduction of the inter-symbol interference (ISI) and inter-carrier interference (ICI). Energy consumption is one of the substantial problems for the LTE system. A number of techniques have been studied so far to reduce the energy consumption and to design an efficient power model for LTE systems [1 - 3].

However, the performance of the OFDMA systems can be greatly improved through energy efficiency which has become the fundamental issue in recent wireless technology. The influence of subcarrier numbers, modulation techniques and resource allocation techniques in OFDMA systems has been studied in [4, 5]. To be able to operate a network in an efficient manner is a very difficult and complex task. At times, optimizing energy consumption in the one part of the network can enlarge power consumption and can degrade the performance on the other side of the network. For network optimization there should be trade-off among QoS, efficiency in capacity, coverage, etc. If we can adopt an intelligent power management which can monitor the traffic and power consumption of different components of the base station, it can identify idle components and can decrease the carriers in a time slot to provide energy efficiency. The power-efficient radio resource control mechanism can also be applied to achieve energy efficiency in multi-antenna packet radio system [6]. Different aspects of energy efficiency in modern and future telecommunication networks have been outlined in [7].

Some of the earlier researches on OFDMA systems considered energy efficient link adaptation technique to maximize energy efficiency [8]. Recent research focuses on energy efficient radio resource scheduling techniques. In [9], authors have proposed QoS aware energy efficient resource scheduling for multi-user OFDMA systems based on Lagrangian model. To address this issue, we have proposed a model to determine the energy efficiency using Engset traffic model. To observe the impact of varying the number of subcarriers on energy efficiency and channel capacity, we have done analysis in terms of number of users and signal to noise ratio (SNR) level. In addition to energy efficiency, to meet the expectation of the customers the main concern of modern marketing is to build a strong customer relationship. Customer satisfaction index (SI) is one of the key parameters by which companies can estimate the customer's satisfaction level [10, 11]. In this paper, we have examined the SI for our proposed model based on the relationship among SI, bandwidth, and SNR.

This paper is structured as follows. The system model is illustrated in section II. The numerical results for the proposed model are described in section III and finally, the paper is concluded in section IV.

II. System Model

Let us consider a central base station (BS) with M mobile users and N orthogonal subcarriers in an OFDMA cell. The SNR of the system can be defined as

$$\gamma = \frac{P_m^n \left| h_m^n \right|^2}{\sigma^2}, \tag{1}$$

where P_m^n is the transmission power for user m with allocated subcarrier n and h_m^n is the multi-path fading coefficient between the transmitter and receiver which is also known as

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

the complex fading experienced by user m and subcarrier n. σ^2 is the variance of the additive white Gaussian noise (AWGN) channel on each subcarrier.

The instantaneous transmission capacity can be denoted as

$$C_m^n = B\log_2(1+\gamma), \tag{2}$$

where B is the channel bandwidth. From equations (1) and (2), we get

$$C_m^n = B\log_2\left(1+\frac{P_m^n\left|h_m^n\right|^2}{\sigma^2}\right). \tag{3}$$

Let us now consider a binary variable $X_m^n \in \{0,1\}$. It indicates the assignment of subcarrier in a multi-user system, where $X_m^n = 1$ indicates n^{th} subcarrier is used by the mth user and $X_m^n = 0$ indicates n^{th} subcarrier is not used by the mth user.

The QoS requirement of a user m can be defined as

$$\sum_{n=1}^{N} C_m^n X_m^n \geq \overline{C}_m, \tag{4}$$

where $\overline{C}_m$ is the maximum rate required for user m.

The power consumed by a base transceiver station (BTS) can be defined as

$$P_{BTS} = \in_0 P_{tx} + \overline{P}_o, \tag{5}$$

where $\in_0$ is a coefficient, that is scaled with radiation power P_{tx}, such as feeder loss, amplifier efficiency etc., $\overline{P}_o$ is the power offset derived from signal processing, battery backup etc. and independent with radiation power.

Now, the maximized energy efficiency can be defined as [12]:

$$EE = \frac{\sum_{n=1}^{N}\sum_{m=1}^{M} C_m^n X_m^n}{\in_0 \sum_{n=1}^{N}\sum_{m=1}^{M} P_m^n X_m^n + \overline{P}_o}. \tag{6}$$

Again, let us consider a limited user network with M number of users. Fig. 1 shows the network model by a Markov chain where λ is the call arrival rate per user and μ is the call termination rate. Here P_0, P_1, P_2 … etc are the probability of arrival of no call, one call, two calls and so on. Basic assumption of Markov chain is that probability of arrival of next step depends on present state but not on previous states. Any new call arrival rate λ will make transition of a probability state P_i to P_{i+1}; similarly, any call termination rate μ will make transition from P_i to P_{i-1}.

Applying cut equations [12] in Fig. 1, we get

$$P_1 = P_0 M\left(\frac{\lambda}{\mu}\right), \tag{7}$$

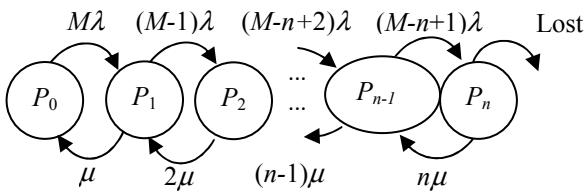


Fig. 1. State transition chain of limited user.

$$P_2 = P_0\left({}^M C_2\right)\left(\frac{\lambda}{\mu}\right)^2, \tag{8}$$

$$P_3 = P_0\left({}^M C_3\right)\left(\frac{\lambda}{\mu}\right)^3, \tag{9}$$

… … … … … … …

and so on. The probability state in general form can be written as

$$P_x = P_0\left({}^M C_x\right)\left(\frac{\lambda}{\mu}\right)^x. \tag{10}$$

In case of the entire sampling space, the following normalization condition must be satisfied:

$$P_0 + P_1 + P_2 + P_3 + \ldots\ldots\ldots + P_n = 1. \tag{11}$$

From the above equations (7-11), the probability of arriving no call P_0 can be derived as

$$P_0 = \frac{1}{\sum_{i=0}^{n} {}^M C_i\left(\frac{\lambda}{\mu}\right)^i}. \tag{12}$$

Now, substituting equation (12) into equation (10), we get

$$P_x = \frac{{}^M C_x a^x}{\sum_{i=0}^{n} {}^M C_i a^i}, \tag{13}$$

where the offered traffic per user is $a = \lambda/\mu$.

The expected number of subcarriers used in a cell is

$$\overline{n} = \sum_{x=0}^{N} x P_x \tag{14}$$

and expected number of deprived user is

$$\overline{m} = \sum_{x=N}^{M-1} x P_x. \tag{15}$$

Hence, the energy efficiency of a multi-user wireless network can be defined as

$$\eta = \frac{\overline{n}C}{\overline{n}P_m^n + \overline{m}\overline{P}_o} = \frac{B\log_2\left(1+\frac{P_m^n\left|h_m^n\right|^2}{\sigma^2}\right)\times\left(\sum_{x=0}^{N} x P_x\right)}{P_m^n\left(\sum_{x=0}^{N} x P_x\right) + \overline{P}_o\left(\sum_{x=N}^{M-1} x P_x\right)}, \tag{16}$$

where $\overline{P}_o$ is the power of the paging signal.

The satisfaction index of the system is defined as

$$SI = \frac{1}{M}\sum_{m=1}^{M}\frac{C_m}{\overline{C}_m} = \frac{B}{\overline{C}_m}\log_2\left(1+\overline{\gamma}_m\right), \tag{17}$$

where $\gamma_1 = \gamma_2 = \ldots\ldots\ldots\ldots = \gamma_m = \overline{\gamma}_m$.

III. Results

We consider an OFDMA cell where the channel bandwidth is 4kHz, transmission power is 45W, $\overline{P}_o$ =12W multipath fading coefficient is 0.78 and variance is chosen as 0.89. Fig. 2 shows the variation of energy efficiency against different number of users considering the number of subcarriers N = 5, 10, 15, and 20. It is visualized from the Fig. 2 that energy efficiency decreases as the number of users increases. Again

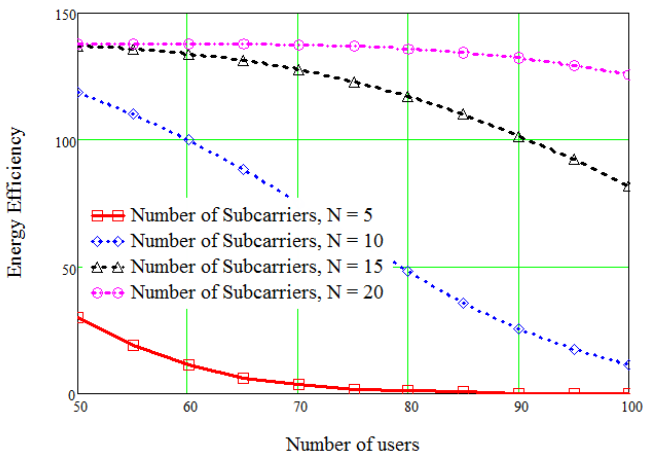


Fig. 2. Variation of energy efficiency against number of users.

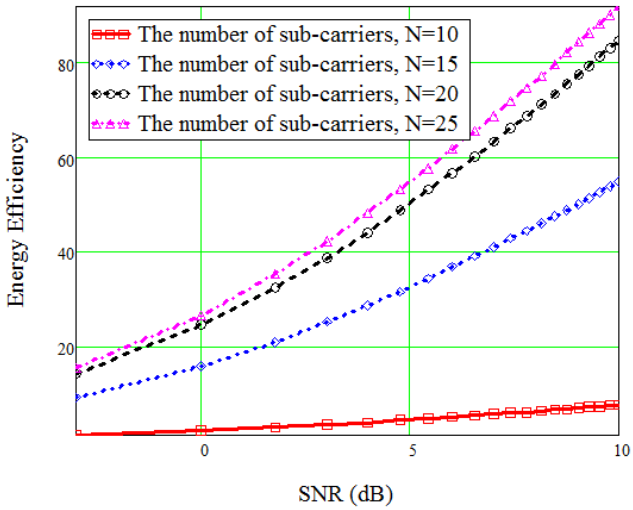


Fig. 3. Variation of energy efficiency against SNR (dB).

energy efficiency increases with increasing number of subcarriers that is the case of enhancement of licensed bandwidth.

From equation (6) it is noticed that energy efficiency is proportional to SNR, taking other parameters constant. It indicates that if the value of SNR increases, the value of energy efficiency also increases. Fig. 3 shows the exact scenario as we have discussed; that is, energy efficiency increases with the increase of the SNR, considering $M = 100$ and $N = 10$, 15, 20, and 25. Instead of using complex algorithm based on Lagrangian model, similar trend found in [9] can be achieved by using our simple state transition model for limited user traffic model.

Fig. 4 illustrates the relation between SI and SNR. It shows that, SI increases as SNR increases for data rate 150kbps, 130kbps, and 120kbps. By showing the relationship between SI and SNR we tried to indicate the user groups whose SI responds with users' data rate with fixed channel capacity.

IV. Conclusions

Previously, energy efficiency in cellular network was analyzed with Lagrangian model where as this paper we have calculated the energy efficiency of a network using limited user traffic model. In particular, we have used Engset traffic model with Markov chain. We have plotted energy efficiency against number of users and SNR taking the number of subcarriers as a parameter. Our results resemble to the conventional Lagrangian model. This implies that using the

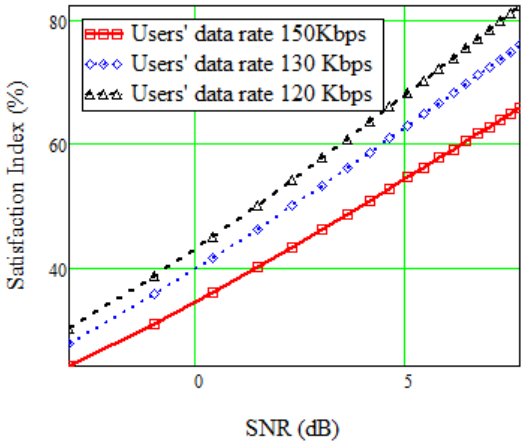


Fig. 4. Variation of satisfaction index against SNR(dB).

simple state transition model, the energy efficiency can be greatly improved for multi-user wireless network. The entire work can be extended incorporating an intelligent network management system to measure the energy efficiency over a period to take necessary measures to allow or block certain users group.

References

[1] S. Deng and H. Balakrishnan, "Traffic-Aware Techniques to Reduce 3G/LTE Wireless Energy Consumption," in *proc. of CoNEXT 2012*, ACM, New York, pp. 181-192, 2012.

[2] J. Huang, F. Qian, A. Gerber, Z. M. Mao, S. Sen and O. Spatscheck, "A Close Examination of Performance and Power Characteristics of 4G LTE Networks," in *Proc. of MobiSys 2012*, ACM, New York, pp. 225-238, 2012.

[3] T. T. Tesfay, R. Khalili, L. B. J. L. Boudec, F. Richter and A. J. Fehske, "Energy Saving and Capacity Gain of Micro Sites in Regular LTE Networks: Downlink Traffic Layer Analysis," in *Proc. PM2HW2N 2011*, ACM, New York, pp. 83-92.

[4] I. Orovic, N. Zaric, S. Stankovic, I. Radusinovic and Z. Veljovic, "Analysis of Power Consumption in OFDM Systems," *MIPRO, 2011 Proc. of the 34th International Convention*, pp. 653-657, May 2011.

[5] O. R. Afolabi, A. Dadlani and K. Kiseon, "Multicast Scheduling and Resource Allocation Algorithms for OFDMA-Based Systems: A Survey," *Communications Surveys & Tutorials, IEEE*, vol. 15, no. 1, pp. 240, 254, 2013.

[6] S. Nonchev and M. Valkama, "Advance Radio Resource Management For Multi Antenna Packet Radio Systems," *International Journal of Wireless Networks (IJWMN)*, Vol.2, pp. 1-14, May 2010.

[7] G. Koutitas and P. Demestichas, "A review of energy efficiency in telecommunication networks", in *Proc. of Int. Telecom. Forum (TELFOR)*, vol. 2, No. 1, 2010.

[8] G. Miao, N. Himayat and G. Y. Li, "Energy-efficient link adaptation in frequency-selective channels," *IEEE Transactions on Communications*, vol. 58, no. 2, pp. 545-554, February 2010.

[9] X. Xiao, X. Tao and J. Lu, "QoS-Aware Energy-Efficient Radio Resource Scheduling in Multi-User OFDMA Systems," *IEEE Communications Letters*, vol. 17, no. 1, pp. 75-78, January 2013.

[10] B. Angelova and J. Zekiri, "Measuring Customer Satisfaction with Service Quality Using American Customer Satisfaction Model (ACSI Model)," *International Journal of Academic Research in Business & Social Sciences*, vol. 1, no. 3, pp. 232-258, October 2011.

[11] M. Dachyar and M. Noviannei, "Customer Satisfaction Index Telecommunication Industry in Indonesia," in *Proc. of World Academy of Science, Engineering and Technology*, no. 69, pp. 135-136, 2012.

[12] M. I. Islam and L. J. Rozario, *Telecommunications Traffic and Network Planning*, Dhaka: Christian Communications Centre Pratibeshi Prokashoni, pp. 15-17, 2008.

Effect of Timing Jitter on the BER Performance of a M-PPM FSO Link over Atmospheric Turbulence Channel

K. M. Nazrul Islam,[1,*] and S. P. Majumder[2,†]

[1]Department of Electrical, Electronic and Communication Engineering, Military Institute of Science and Technology
Dhaka-1216, Bangladesh

[1, 2]Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology
Dhaka-1000, Bangladesh

[*]nazrul_4620@yahoo.com

[†]spmajumder@eee.buet.ac.bd

Abstract—**An analytical approach is presented to evaluate the impact of timing jitter on the bit error rate performance of optical M-ary pulse position modulation (M-PPM) free space optical (FSO) link under the influence of atmospheric turbulence modelled by Log-normal distribution. The expression of the conditional bit error rate (BER) for a given timing jitter is developed and the average BER is found by averaging the conditional BER over the probability density function of the timing jitter which is considered to be zero mean Gaussian. The results are evaluated at a bit rate of 2.4 Gbps over weak turbulence with Avalanche Photodiode (APD) receiver for different order of M-PPM modulation. It is found that there is significant deterioration in BER performance with increase in timing jitter variance and results in BER floor at higher value of jitter variance for a given M-PPM order and APD gain. Further, it is noticed that higher order M-PPM system suffer less than lower order PPM system at a given system BER. The analysis can also be useful for evaluating the impact of slot timing jitter for M-PPM FSO system with strong turbulence.**

Index Terms*— *Bit Error Rate (BER), Free-Space Optical (FSO), Pointing Error, Timing Jitter, M-PPM, Log-Normal Distribution, Gamma-Gamma Distribution, Hermite Polynomial etc.

I. INTRODUCTION

Free Space Optical (FSO) communications have gain sufficient interests in recent years as a serious alternative to radio frequency (RF) links due to the fact that FSO equipment are small, less costly, and easy to deploy [1-3]. Further FSO have enormous bandwidth compared to RF communication systems. However, the performance of terrestrial FSO links are highly affected by the atmospheric effect caused by refractive index fluctuations due to temperature change which results in atmospheric turbulence classified into weak and strong turbulence [1-2]. Several statistical models are considered such as log-normal distribution, Gamma-gamma distribution and negative exponential distribution to evaluate analytically the degradation caused by atmospheric turbulence on terrestrial FSO links with intensity modulation and direct detection (IM/DD) [2-5].

Numerous woks on optical M-ary pulse position modulation (M-PPM) are also carried out to investigate the effectiveness of M-PPM in FSO links as M-PPM is an energy efficient modulation format compared to on-off keying (OOK) and other modulation schemes [6-10]. The effects of atmospheric turbulence and pointing error on the bit error rate performance of a PPM FSO link is reported in [6]. In [7] the authors have investigated the BER performance of M-PPM over FSO link with log-normal and negative exponential distribution with error correction coding and without coding. The performance of optical PPM for deep space communication is also reported recently [8] with multi-pulse position modulation with channel fading. The performance of Avalanche Photodiode (APD) based M-PPM receiver over atmospheric turbulence is also carried out considering different APD gain [9]. More recently, the effect of strong turbulence on the shot noise limited and thermal noise limited operations of M-PPM receivers for a FSO link is investigated with strong turbulence condition [10]. However, all the previous works on M-PPM FSO links are based on perfect synchronization although the performance of an optical M-PPM receiver is highly sensitive to slot synchronization error [11]. Several slot synchronizers are also reported earlier for optical M-PPM system [11-12] and the effect of imperfect slot synchronization on the performance of optical M-PPM receiver are reported earlier [11-13] without and with error correction coding.

In this paper we provide an analytical approach to evaluate the impact of imperfect slot synchronization on optical M-PPM FSO system in presence of atmospheric turbulence. Analysis is carried out for M-PPM FSO links to find the effect of timing error on the BER performance with log-normal fading considering weak and strong scintillation indices.

II. SYSTEM MODEL

The block diagram of the free-space optical (FSO) M-ary PPM link is shown in fig. 1. Input binary bits are given input to M-ary PPM modulator to generate the M-PPM symbols. The M-PPM symbols are transmitted by a laser transmitter using a collimator lens.

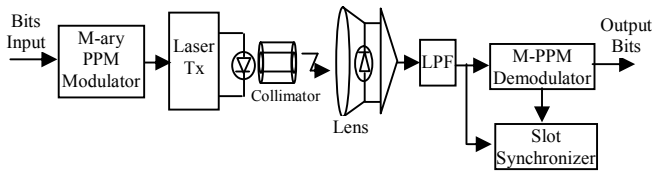


Fig. 1. Block diagram of M-PPM transmitter and photon counting receiver for a FSO link

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

At the receiver the optical beam is received by a photo-detector followed by a preamplifier and a low pass filter (LPF). The slot synchronizer generates the clock using the received M-PPM signal. The generated clock is used by the M-PPM demodulator to determine the slot count over the M-PPM slots. Based on the slot count the demodulator take decision of the transmitted M-PPM symbol and generates the output bit stream.

III. SYSTEM ANALYSIS

We consider an M-PPM slot of duration T_s and slot count over k^{th} time slot with perfect synchronization is m_k. We define the normalized timing error as $\varepsilon = \Delta T/T_s$. In presence of timing error ε, the slot count over a signal slot is given by:

$$m_k(\varepsilon) = m_k\left(\frac{T_S - \Delta T}{T_S}\right) = m_k(1-\varepsilon) \tag{1}$$

The slot duration T_s of M-PPM with the modulation order M, data rate R_b and bit duration T_b is given by:

$$T_s = (T_b \log_2 M)/M = \frac{\log_2 M}{M} R_b \tag{2}$$

In presence of log-normal distribution, the average photon per PPM slot $E[K_s]$ are function of the mean $m_k(\varepsilon)$ and variance of the log-normal channel σ_k^2 and can be represented as:

$$E[K_s] = \exp\left[\frac{\sigma_k^2}{2} + m_k\right] \tag{3}$$

where, E means expectation and σ_k^2 is a function of the scintillation index σ_{si}^2 as:

$$\sigma_K^2 = \ln\left(\sigma_{si}^2 + 1\right) \tag{4}$$

The total noise photons per PPM slot $K_n(\varepsilon)$which results from background noise, thermal noise and timing error induced signal count as:

$$K_n(\varepsilon) = \frac{2\sigma_n^2}{Gq^2} + 2F(K_b + \varepsilon . m_k) \tag{5}$$

where, K_b is the average background noise photons per PPM slot, G is the mean gain of APD and q is the electronic charge and F is the APD excess noise factor defined by:

$$F = 2 + \zeta G \tag{6}$$

where, ζ is the ionization factor for APD.

The variance σ_n^2 of the thermal noise in a PPM slot is defined as:

$$\sigma_n^2 = \frac{2KT_r T_S}{R_L} \tag{7}$$

where, T_r is the effective absolute temperature of the APD receiver, K is the Boltzmann's constant, R_L is the APD receiver load resistance.

The conditional it error rate (BER) conditioned on a given timing error ε, of an M-ary PPM over log-normal channel can be represented as [8]:

$$P_b(\varepsilon) \le \frac{M}{2\sqrt{\pi}} \sum_{\substack{i=-N \\ i\neq 0}}^{N} w_i Q\left[\frac{\exp\left\{2(\sqrt{2}\sigma_k x_i + m_k(\varepsilon))\right\}}{F\exp(\sqrt{2}\sigma_k x_i + m_k(\varepsilon) + K_n(\varepsilon))}\right] \tag{8}$$

where, M is the PPM order, w_i and x_i are the weight and zeroes of Hermite Polynomial [7]. Considering $\varepsilon = \frac{\Delta T}{T_s}$ being Gaussian distributed with zero mean and variance σ_ε^2, the pdf of ε is given by[13]:

$$p(\varepsilon) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} e^{-\varepsilon^2/2\sigma_\varepsilon^2} \tag{9}$$

The average BER can now be evaluated as:

$$\text{BER} = \int_{-\infty}^{\infty} P_b(\varepsilon)\, P(\varepsilon)\, d\varepsilon \tag{10}$$

IV. RESULTS AND DISCUSSIONS

Following the analytical approach in section III we evaluate the conditional bit error rate $P_b(\varepsilon)$ conditioned on a given value of normalized timing error ε and the average bit error rate (BER) for M-PPM FSO link with APD gain of 150 at a bit rate of 2.4 Gbps with weak turbulence with scintillation index of 0.3.

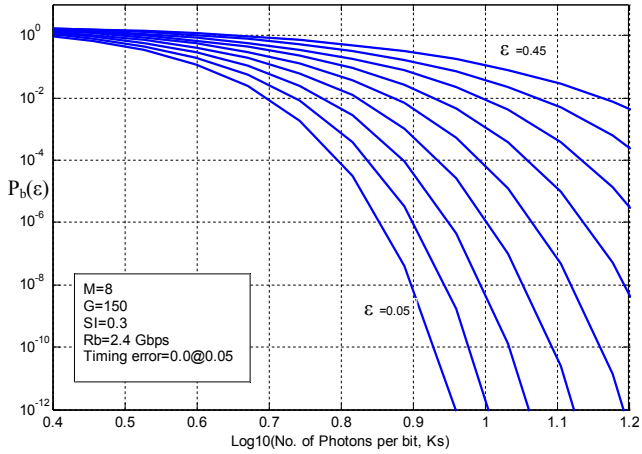


Fig. 2. Plots of conditional BER $P_b(\varepsilon)$ conditioned on a given value of normalized timing error ε versus number of photon per bit of 8-PPM FSO link with turbulence of scintillation index 0.3 at a bit rate of 2.4 Gbps and APD Gain 150.

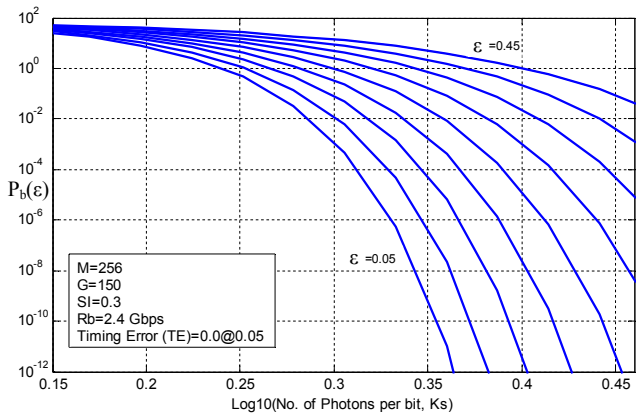


Fig. 3. Pots of conditional BER $P_b(\varepsilon)$ conditioned on a given value of normalized timing error ε versus number of photon per bit of 256-PPM FSO link with turbulence of scintillation index 0.3 at a bit rate of 2.4 Gbps and APD Gain 150.

The plots of $P_b(\varepsilon)$ as a function of number of photons per bit K_s are depicted in fig. 2 and fig. 3 for M=8 and M=256 respectively with ε as a parameter. It is noticed that M-PPM FSO System suffers due to timing error ε, which is significant at high value of ε.

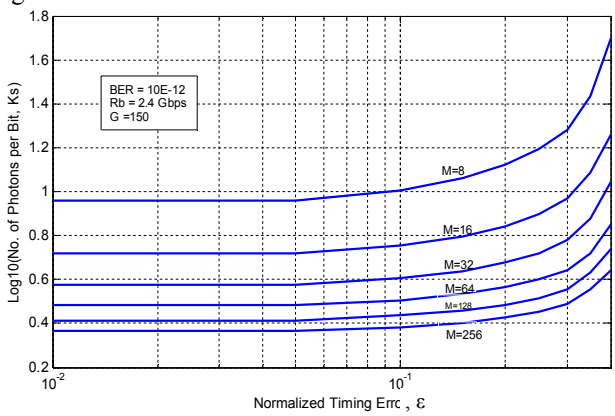


Fig. 4. Plots of receiver sensitivity in terms of number of photons per bit for a conditional BER of 10^{-12} as a function of fixed timing error ε for different M-PPM order.

The plots of receiver sensitivity in number of photons per bit due to timing error at a given $P_b(\varepsilon)$ of 10^{-12} are shown in fig. 4 for M=8, 16, 32, 64, 128 and 256. It is noticed that penalty is significant at higher timing error and is less for high PPM order.

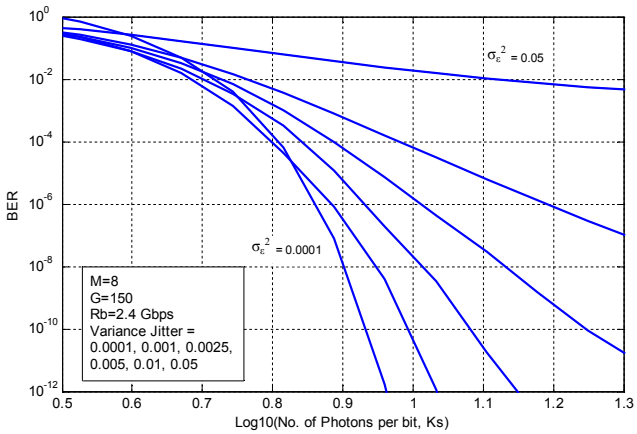


Fig. 5. Pots of average BER versus number of photon per bit of 8-PPM FSO link for different values of timing variance σ_ε^2 with turbulence of scintillation index 0.3 at a bit rate of 2.4 Gbps and APD Gain 150.

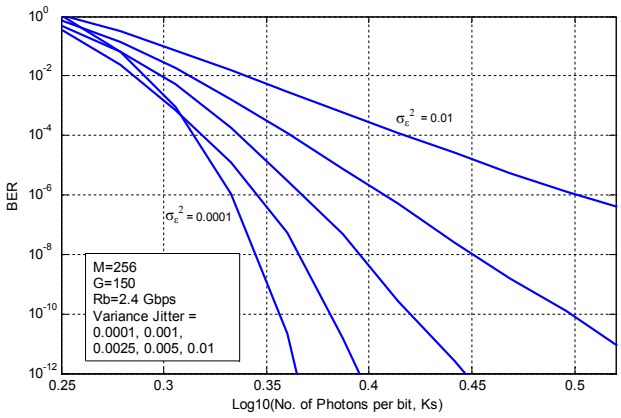


Fig. 6. Pots of average BER versus number of photon per bit of 256-PPM FSO link for different values of timing variance σ_ε^2 with turbulence of scintillation index 0.3 at a bit rate of 2.4 Gbps and APD Gain 150.

The plots of average BER as a function of number of photons per bit K_s are depicted in fig. 5 and fig. 6 for M=8 and M=256 respectively with Jitter Variance σ_ε^2 as a parameter. It is also noticed that M-PPM FSO System suffers due to Jitter Variance, which is significant at high value of σ_ε^2.

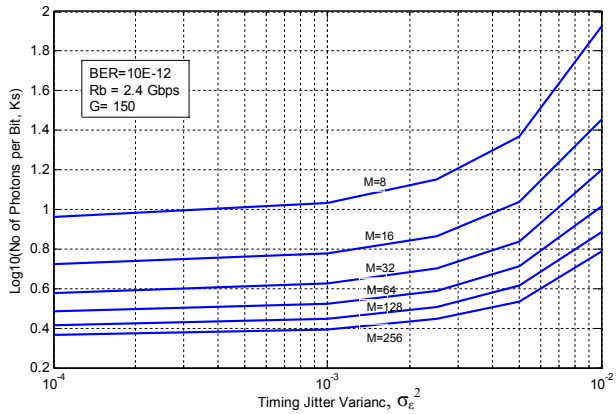


Fig. 7. Plots of receiver sensitivity in terms of number of photons per bit for an average BER of 10^{-12} as a function of variable timing jitter variance σ_ε^2 for different M-PPM order.

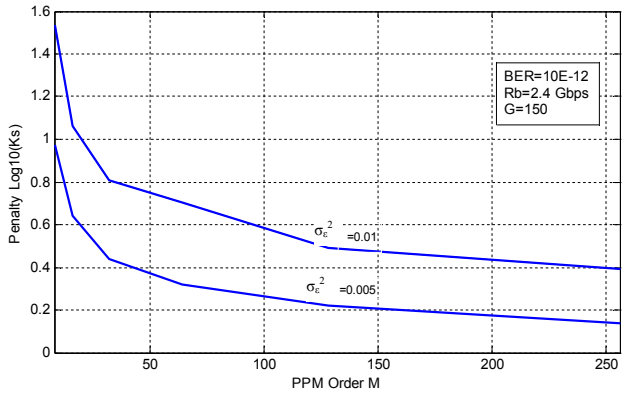


Fig. 7. Plots of penalty log10 (Ks) in terms of number of photons per bit for an average BER of 10^{-12} as a function of different M-PPM order for different timing error variance σ_ε^2 .

The plots of receiver sensitivity in number of photons per bit due to variable timing error at a given BER of 10^{-12} are shown in fig. 6 for M=8, 16, 32, 64, 128 and 256. It is noticed that penalty is significant at higher Jitter Variance σ_ε^2 and is less for high PPM order.

The plots of power penalty in number of photons per bit due to Jitter Variance and PPM modulation order at an average BER of 10^{-12} are shown in fig. 7. It is noticed that penalty is significant at higher jitter variance and is less for high PPM order.

CONCLUSIONS

An analytical development is carried out to obtain the BER of an M-PPM FSO communication system under the influence of atmospheric turbulence with log-normal distribution. Results are evaluated at a bit rate of 2.4 Gbps with weak turbulence and APD receiver. It is found that the system suffers significant power penalty due to timing jitter and BER floor occurs at higher value of jitter variance. The analysis will find application in design of high data rate M-PPM FSO system with APD based receiver.

REFERENCES

[1] S. Arnon, "Effect of atmospheric turbulence and building sway on optical wireless communication system," Optics Letter, vol. 28 no. 2, pp. 129-131, January 2003.

[2] D. Kedar and S. Arnon, "Urban optical wireless communication network: the main challenges and possible solution," IEEE Communication Mag. Vol. 42, pp. 52-57, May 2004.

[3] S. M. Navidpour, M. Uysal and M. Kaverad, "BER performance of free space optical transmission with spatial diversity," IEEE Transaction on Wireless Communications, vol. 6, no. 8, pp. 2813-2819, Aug 2007.

[4] T. A. Tsiftsis, H. G. Sandalidis, G. K. Karagiannidis, and M. Uysal, " BER Performance of FSO Links over Strong Atmospheric Turbulence Channels with Pointing Errors," IEEE Communications Letters. Vol. 12, no. 1, pp. 44-46 Jan 2008.

[5] E. Bayaki, R. Schober, R. K. Mallik, "Performance analysis of MINO Free-Space Optical System in Gamma-Gamma Fading," IEEE Transaction on Communications, vol. 57, no. 11, pp. 3415-3424, Nov 2009.

[6] W. Gappmair, S. Hranilovic and E. Leitgeb, "Performance of PPM on terrestrial FSO links with turbulence and pointing errors," IEEE Communication Letters, vol. 4, no. 5, pp. 468-470, May 2010.

[7] M. R. Abaza, N. A. Mohammed, M. H. Aly, "BER Performance of M-ary PPM free-space optical communications with channel fading," Proc. IEEE 8th International Symposium on High Capacity Optical Networks and Enabling Technologies, HONET 2011, Riyadh, Saudi Arabia, Dec 2011.

[8] H. Selmy, H. M. H. Shalaby and Z. Kawasaki, "Performance analysis of gradual multi-pulse position modulation in deep space optical communications," J. Optical Communications and Networking, vol. 4, no. 10, pp. 812-820. Jan 2012.

[9] T. T. Pham Hien and T. D. Ngoc, "A comprehensive model for Performance analysis of APD-based FSO systems using M-PPM signalling in atmospheric turbulence," REV J. on Electronics and Communications, vol. 2, no. 3-4, pp. 147-152, July-Dec 2012.

[10] A. E. Morra, H. S. Khallaf, H. M. H. Shalaby and Z. Kawasaki," Performance analysis of both shot and thermal noise limited multi-pulse PPM receivers in Gamma-Gamma Atmospheric Channels," J. Lightwave Technology, vol. 31, no. 19, pp. 3142-3150, Oct 2013.

[11] G. Ling and R. M. Gagliardi, "Slot Synchronization in Optical PPM Communications," IEEE Transaction Communications, vol. 34, no. 12, pp. 1202-1208, Dec 1986.

[12] C. Chen and C. S. Gardner, " Performance of PLL synchronized optical PPM communication systems" IEEE Transaction Communication, vol. 34, no. 10, pp. 988-994, Oct 1986.

[13] S.P. Majumder, R Gangopadhyay and G. Prati, "Effect of Imperfect Slot Synchronization on the Performance of Coded Optical M-PPM System", Journal of Optical Communications, vol. 12, no. 3, pp. 96-100, 1991.

Effect of bandwidth in cross-correlation based underwater network size estimation

*M. S. Anower[1], S. A. H. Chowdhury[2], J. E. Giti[1], and M. I. Haque[1]
[1]Department of Electrical & Electronic Engineering
[2]Department of Electronics & Telecommunication Engineering
Rajshahi University of Engineering & Technology, Rajshahi-6204, Bangladesh
*md.shamimanower@yahoo.com, arif.1968.ruet@gmail.com, monimala89@yahoo.com, ismail07rueteeee@gmail.com

Abstract—**Due to the harshness of underwater environment, size estimation of underwater network is difficult using conventional protocol techniques. A statistical signal processing approach of size estimation is proposed for this purpose using cross-correlation of Gaussian signals, which is effective for any environment networks. Limited bandwidth of underwater communications channel poses constraint on using Gaussian signal, which has infinite bandwidth. The aim of this paper is to investigate the effect of this limited bandwidth in cross-correlation based underwater network size estimation using Gaussian signals with finite bandwidth.**

Index Terms—**Cross-correlation function (CCF); effect of bandwidth; size estimation; underwater acoustic channels; underwater wireless communication network (UWCN).**

I. Introduction

UWCNs using acoustic waves, which are referred to as underwater acoustic sensor networks (UASN), have several applications, such as monitoring daily ocean life, climatic data collection, disaster alleviation [1], deep-sea archaeology, seawater data collection [2]. To successfully achieve these goals, network size information plays an important role, which ensures proper functionality of the network as well as effective data collection. So, network size is a crucial factor, which is equal to the number of operational nodes in the network and needed to estimate it accurately.

Number of tag ID estimation in radio frequency identification (RFID) systems is equivalent to size estimation of terrestrial wireless network, where different protocols [3]–[5] are used for tag estimation. For dynamic anonymous networks, consensus based size estimation methods have been proposed in [6]–[8]. Another size estimation technique has been investigated by Evans et al. [9] for Peer-to-Peer (P2P) network. Size of a network can also be estimated using distributed orthogonalization [10]. These methods are designed for RFID as well as terrestrial network and not applicable to underwater network due to large propagation latency, non-negligible capture effect, high path loss and limited bandwidth of underwater environment [11]. For underwater wireless network, a dynamic network structure estimation technique has been proposed by Blouin [12] based on intermission between the nodes. ALOHA based node estimation techniques have been investigated by Howlader et al. [13]–[15] for underwater network considering long propagation delay and capture effect. As these size estimation techniques ([12]–[15]) utilize some aspect of communications protocol(s), they suffer from protocol complexity in underwater environment.

An alternative approach of underwater network size estimation with reduced protocol complexity has been proposed by Anower et al. [16], [17] through the cross-correlation of Gaussian signals received at two sensors. A similar node estimation technique using three sensors has been investigated by Chowdhury et al. [18], [19]. In those literatures, Gaussian signals of infinite bandwidth are used for estimation and bandwidth effect is not considered. In this paper, finite band Gaussian signals are used to analyze the effect of limited bandwidth of underwater acoustic channel in cross-correlation based size estimation technique.

II. Size Estimation Procedure

To obtain insight into the cross-correlation based size estimation technique, a review of the estimation process with two sensors [16], [17] is discussed in this section. In this process, a 3D spherical region is considered as network where N nodes are uniformly distributed as shown in Fig. 1. Two sensors are placed at the middle of that spherical region to estimate the network size as shown in Fig. 1, which send probe request. In this approach, acoustic wave is considered as the underwater communication carrier for signal propagation.

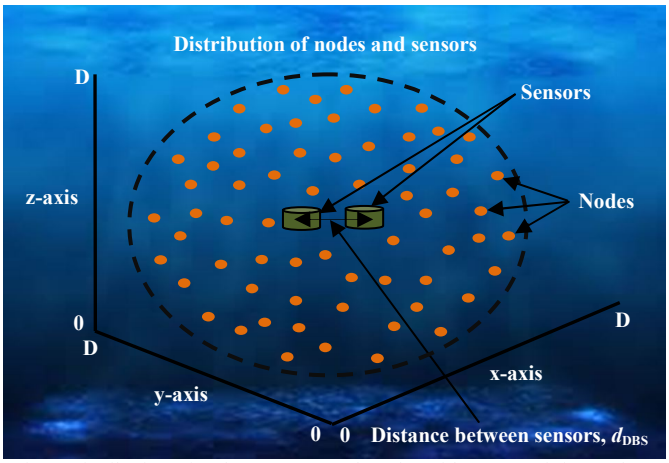


Fig. 1 Distribution of underwater network nodes with N transmitting nodes.

After receiving probe request, it is assumed that, nodes can transmit Gaussian signal using acoustic wave as the carrier. Gaussian signal has a certain property that, cross-correlation of two Gaussian signals results a delta function, which is the basic idea of this estimation approach and also the reason of using Gaussian signals as transmitted signals. Simultaneously transmitted Gaussian signals from N nodes are received by the sensors at different delay differences and summed at each

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

sensor location forming composite Gaussian signals. These two composite signals are then cross-correlated to formulate cross-correlation function (CCF), which results a series of delta functions [17]. One such CCF obtained with N (=1000) nodes is shown in Fig. 2.

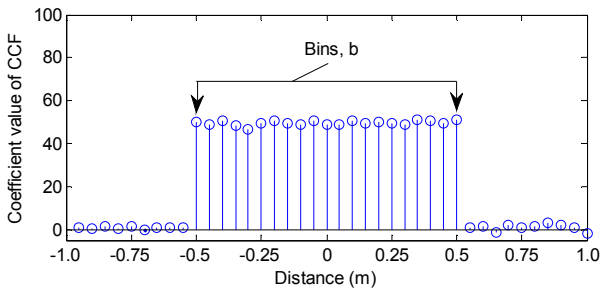


Fig. 2 Bins, b in the cross-correlation function.

Bins, b in the CCF (as shown in Fig. 2) is defined as a place occupied by a delta inside a space of a width twice the distance between sensors and that place is determined by the delay difference of the signal coming to the sensors [20]. The deltas of equal delay differences are placed in that particular bin. Number of bins, b is twice the number of samples between the sensors (NSBS), m, minus one [16] and can be expressed as [20]:

$$b = \frac{2 \times d_{DBS} \times S_R}{S_P} - 1 \tag{1}$$

where, d_{DBS} is the distance between the sensors, S_{R} is the sampling rate and S_{P} is the speed of acoustic wave propagation.

Ratio of standard deviation (σ) to the mean (μ), R of the CCF is chosen as the estimation parameter, which requires no prior knowledge of signal strength [20]. Direct calculation of this estimation parameter using statistical expression is very complex. For this reason, cross-correlation problem is reframed to a probability problem using the well known occupancy problem which follows the binomial probability distribution, where the parameters are N and $1/b$ [17]. After reframing, ratio of standard deviation to the mean, R of CCF can be expressed as [16]:

$$R = \frac{\sigma}{\mu} = \sqrt{\frac{b-1}{N}} \tag{2}$$

Form (2) network size, N is obtained as follows:

$$N = \frac{b-1}{R^2} \tag{3}$$

Using this equation we can estimate N, as we know b and R can be calculated from the CCF.

III. Effect of Bandwidth

An underwater acoustic channel is effectively considered as bandwidth limited due to the frequency dependency of absorption loss, as absorption coefficient, α is an increasing function of frequency [21]. So, absorption loss becomes more significant with the increase of bandwidth, which limits the transmission range. Typical bandwidths of underwater acoustic channel for different ranges are shown in Table I [22]. Channel bandwidth restricts the signal bandwidth which might affect the estimation performance.

TABLE I

Available Bandwidth for Different Ranges in Underwater Acoustic Channel

	Range [km]	Bandwidth [kHz]
Very long	1000	< 1
Long	10 – 100	2 – 5
Medium	1 – 10	≈ 10
Short	0.1 – 1	20 – 50
Very short	< 0.1	> 100

To analyze the effect of bandwidth, a 10kHz Gaussian signal is used in the simulation instead of infinite band signal using a low-pass filter to avoid unwanted high frequency attenuation of underwater environment. The ratio of standard deviation to the mean of CCF for finite and infinite bandwidth case is denoted as R_{finiteBW} and $R_{\mathrm{infiniteBW}}$, respectively. Now ratio of R_{finiteBW} and $R_{\mathrm{infiniteBW}}$ is plotted against N in Fig. 3 for $b = 29$. It can be seen in the figure that, R_{finiteBW} is almost the constant multiple of $R_{\mathrm{infiniteBW}}$ and the mean of those constants are 0.8093 for this case. This scaling factor (0.8093) can be used to match the results of finite (10kHz) bandwidth cases with infinite bandwidth cases.

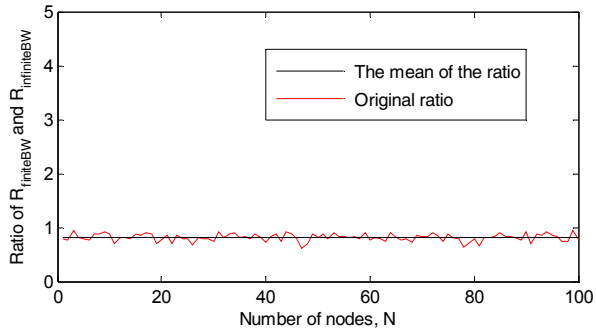


Fig. 3 Ratio of R_{finiteBW} and $R_{\mathrm{infiniteBW}}$ for 10kHz bandwidth.

Now, theoretical approximation of R_{finiteBW} can be obtained by multiplying this scaling factor with theoretical $R_{\mathrm{infiniteBW}}$. So, R_{finiteBW} can be expressed as:

$$R_{\mathrm{finiteBW}} = 0.8093 \times R_{\mathrm{infiniteBW}} \tag{4}$$

As infinite band Gaussian signals are considered in section II, R of CCF expressed in (2) is similar to $R_{\mathrm{infiniteBW}}$. Then using (2) we can write:

$$R_{\mathrm{finiteBW}} = 0.8093 \times \sqrt{\frac{b-1}{N}} \tag{5}$$

For finite (10 kHz) bandwidth case, network size, N can be estimated using the following equation:

$$N = \left(\frac{0.8093}{R_{\mathrm{finiteBW}}}\right)^2 \times (b-1) \tag{6}$$

IV. Results and Discussion

Matlab is used as the programming tool. Theoretical and simulated results of R_{finiteBW} and $R_{\mathrm{infiniteBW}}$ are shown in Fig. 4 with respect to N for $b = 29$ ($d_{\mathrm{DBS}} = 0.5$m and $S_{\mathrm{R}} = 45$kSa/s). It is obvious from Fig. 4 that, R_{finiteBW} is always less than $R_{\mathrm{infiniteBW}}$, but maintains the same shape. Due to the limited bandwidth, cross-correlation of Gaussian signals results sinc functions [23] instead of delta functions, which corrupts the CCF as well as R_{finiteBW}.

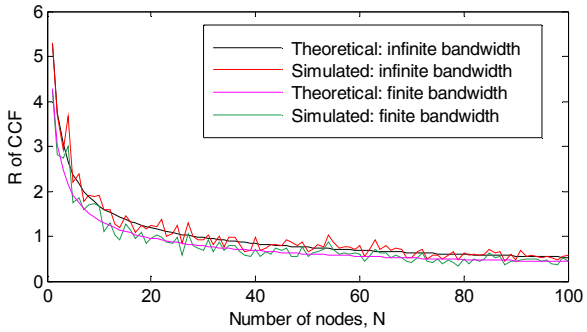


Fig. 4 R_{finiteBW} and $R_{\text{infiniteBW}}$ versus N for finite (10kHz) and infinite bandwidth cases with $b = 29$ ($d_{\text{DBS}} = 0.5$m and $S_{\text{R}} = 45$kSa/s).

For better understanding, results of $R_{\text{infiniteBW}}$ and R_{finiteBW} for $b = 29$ ($d_{\text{DBS}} = 0.5$m and $S_{\text{R}} = 45$kSa/s) are plotted against N in Fig. 5 with less points. Simulated results agree with the theoretical results in Fig. 5, which verify the effectiveness of the scaling factor. With this R_{finiteBW} and the scaling factor (0.8093), number of nodes, N can be estimated using (6).

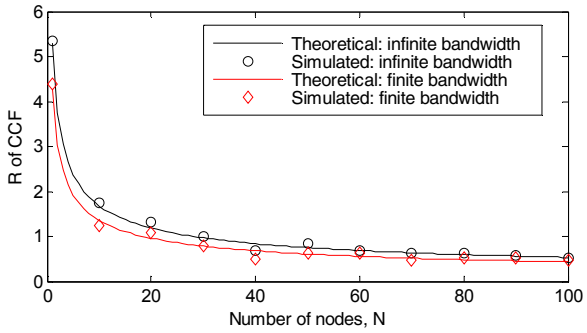


Fig. 5 R_{finiteBW} and $R_{\text{infiniteBW}}$ versus N for finite (10kHz) and infinite bandwidth cases with $b = 29$ ($d_{\text{DBS}} = 0.5$m and $S_{\text{R}} = 45$kSa/s). (with less points)

To test the validity of the scaling factor obtained in section III for different number of bins, b, theoretical results of R_{finiteBW} and $R_{\text{infiniteBW}}$ are compared with the corresponding simulated results in Fig. 6 and 7 for $b = 49$ ($d_{\text{DBS}} = 0.5$m and $S_{\text{R}} = 75$kSa/s) and 69 ($d_{\text{DBS}} = 1.0$m and $S_{\text{R}} = 52.5$kSa/s), respectively. Matching theoretical and simulated results of R_{finiteBW} in Fig. 6 and 7 signify that, for 10kHz bandwidth case, this scaling factor (0.8093) will hold for all b. All of these results are obtained based on an assumption that, received power from each node is equal, which can be achieved using probing technique. The other parameters used in the simulations are: dimension of the sphere, $D = 2000$m; $S_{\text{P}} = 1500$m/s; signal length, $N_{\text{s}} = 10^6$ samples.

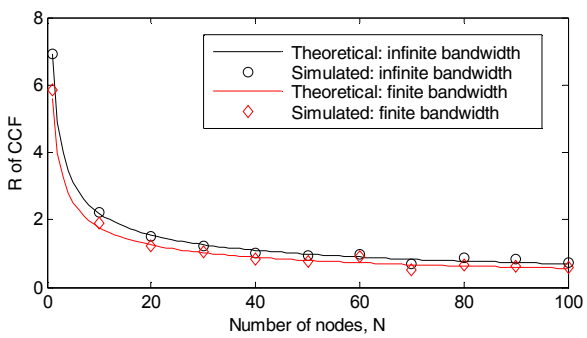


Fig. 6 R_{finiteBW} and $R_{\text{infiniteBW}}$ versus N for finite (10kHz) and infinite bandwidth cases with $b = 49$ ($d_{\text{DBS}} = 0.5$m and $S_{\text{R}} = 75$kSa/s).

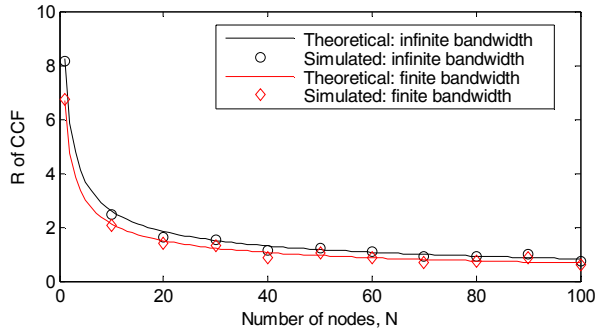


Fig. 7 R_{finiteBW} and $R_{\text{infiniteBW}}$ versus N for finite (10kHz) and infinite bandwidth cases with $b = 69$ ($d_{\text{DBS}} = 1.0$m and $S_{\text{R}} = 52.5$kSa/s).

V. Conclusions

Limited underwater acoustic channel bandwidth makes it impractical to utilize infinite band Gaussian signals in cross-correlation based size estimation technique. Effect of cross-correlating band limited Gaussian signals on estimation parameter, R of CCF is investigated in this paper using 10kHz Gaussian signals. It is shown that, scaling factor can be derived for this band limited condition to develop the corresponding theoretical results. Using this scaling factor network size can be estimated accurately with finite band (10kHz) Gaussian signals. Further investigations are required for the bandwidth less than 10kHz. It is also shown that, this scaling factor is independent of b.

References

[1] N. N. Soreide, C. E. Woody, and S. M. Holt, "Overview of ocean based buoys and drifters: Present applications and future needs," *16th International Conference on Interactive Information and Processing System (IIPS) for Meteorology, Oceanography, Hydrology*, Long Beach, California, USA, 2004.

[2] F. Baronti, G. Fantechi, R. Roncella, and R. Saletti, "Wireless sensor node for surface seawater density measurements," *Sensors (Basel)*, vol. 12, no. 3, pp. 2954–2968, March 2012. doi: 10.3390/s120302954.

[3] C. Yihong, F. Quanyuan, Z. Ma, and T. Liu, "Multiple-bits-slot reservation aloha protocol for tag identification," *IEEE Transactions on Consumer Electronics*, vol. 59, no. 1, pp. 93–100, February 2013. doi: 10.1109/TCE.2013.6490246.

[4] Y. Wang, Y. Liu, H. Leung, R. Chen, and A. Li, "A multi-bit identification protocol for RFID tag reading," *IEEE Journal on Sensors*, vol. 13, no. 10, pp. 3527–3536, Oct. 2013. doi: 10.1109/JSEN.2013.2272460.

[5] Y.-C. Lai, L.-Y. Hsiao, H.-J. Chen, C.-N. Lai, and J.-W. Lin, "A novel query tree protocol with bit tracking in RFID tag identification," *IEEE Transactions on Mobile Computing*, vol. 12, no. 10, pp. 2063–2075, Oct. 2013. doi: 10.1109/TMC.2012.176.

[6] G. Luna, R. Baldoni, S. Bonomi, and I. Chatzigiannakis, "Conscious and Unconscious Counting on Anonymous Dynamic Networks," *Distributed Computing and Networking*, vol. 8314, pp. 257–271, 2014. doi: 10.1007/978-3-642-45249-9_17.

[7] D. Varagnolo, G. Pillonetto, and L. Schenato, "Consensus based estimation of anonymous networks size using Bernoulli trials," In *Proceedings of American Control Conference (ACC)*, Monteral, QC, Canada, 27–29 June, 2012, pp. 2196–2201.

[8] H. Terelius, D. Varagnolo, and K. H. Johansson, "Distributed size estimation of dynamic anonymous networks," In *Proceedings of IEEE 51st Annual Conference on Decision and Control (CDC)*, Maui, Hawaii, USA, 10–13 Dec. 2012, pp. 5221–5227. doi: 10.1109/CDC.2012.6425912.

[9] N. Evans, B. Polot, and C. Grothoff, "Efficient and Secure Decentralized Network Size Estimation," In *Proceedings of NETWORKING 2012*, vol. 7289, pp. 304–317. doi: 10.1007/978-3-642-30045-5_23.

[10] O. Sluciak, and M. Rupp, "Network size estimation using distributed orthogonalization," *IEEE Signal Processing Letters*, vol. 20, no. 4, pp. 347–350, April 2013. doi: 10.1109/LSP.2013.2247756.

[11] L. Liu, S. Zhou, and Jun-Hong Cui, "Prospects and problems of wireless communication for underwater sensor networks," *Wireless*

Communications and Mobile Computing, vol. 8, no. 8, pp. 977–994, 2008. doi: 10.1002/wcm.v8:8.

[12] S. Blouin, "Intermission-based adaptive structure estimation of wireless underwater networks," *10th IEEE International Conference on Networking, Sensing and Control (ICNSC)*, 2013, 10–12 April. 2013, pp. 146–151. doi: 10.1109/ICNSC.2013.6548727.

[13] M. S. A. Howlader, M. R. Frater, and M. J. Ryan, "Estimating the number of neighbours and their distribution in an underwater communication network (UWCN)," In *Proceedings of Second International Conference on Sensor Technologies and Applications*, Canberra, 20–22 November, 2007.

[14] M. S. A. Howlader, M. R. Frater, and M. J. Ryan, "Estimation in underwater sensor networks taking into account capture," *IEEE Oceans'07*, Aberdeen, Scotland, 18–21 June, 2007, pp. 1–6.

[15] M. S. A. Howlader, M. R. Frater, and M. J. Ryan, "Delay-insensitive identification of neighbors using unslotted and slotted protocols," *Wirel. Commun. Mob. Comput.*, 2012. doi: 10.1002/wcm.2242.

[16] S. Anower, M. R. Frater, and M. J. Ryan, "Estimation by cross-correlation of the number of nodes in underwater networks," *Australasian Telecommunication Networks and Applications Conference (ATNAC)*, 2009, 10–12 Nov. 2009, pp. 1–6. doi: 10.1109/ATNAC.2009.5464716.

[17] M. S. Anower, M. A. Motin, A. S. M. Sayem, and S. A. H. Chowdhury, "A node estimation technique in underwater wireless sensor network," *International Conference on Informatics, Electronics & Vision (ICIEV)*, 2013, 17–18 May 2013, pp. 1–6. doi: 10.1109/ICIEV.2013.6572582.

[18] S. A. H. Chowdhury, M. S. Anower, and J. E. Giti, "A signal processing approach of underwater network node estimation with three sensors," In *Proceedings of 1st International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT 2014)*, 10–12 April, 2014, Dhaka, Bangladesh.

[19] S. A. H. Chowdhury, M. S. Anower, and J. E. Giti, "Effect of sensor number and location in cross-correlation based node estimation technique for underwater communications network," In *Proceedings of 3rd International Conference on Informatics, Electronics & Vision (ICIEV 2014)*, 23–24 May, 2014, Dhaka, Bangladesh. doi: 10.1109/ICIEV.2014.6850728.

[20] M. S. Anower, "Estimation using cross-correlation in a communications network," Ph.D. dissertation, SEIT, University of New South Wales at Australian Defense Force Academy, Canberra, 2011.

[21] M. Stojanovic, "On the relationship between capacity and distance in an underwater acoustic communication channel," *WUWNet'06*, Los Angeles, California, USA, 2006.

[22] I. F. Akyildiz, D. Pompili, and T. Melodia, "Challenges for efficient communication in underwater acoustic sensor networks," *Ad Hoc networks*, vol. 3, no. 3, pp. 257–279, 2005.

[23] M. J. Alam, M. R. Frater, and E. H. Huntington, "Improving resolution and SNR of correlation function with the increase in bandwidth of recorded noise fields during estimation of bottom profile of ocean," *IEEE Oceans'10*, Sydney, Australia, 24–27 May, 2010.

Dispersion Optimization of 160-Gb/s 2000-km Transmission by Appropriate Orientation of Chirped Fiber Bragg Grating

Shariful Islam,[1] Farzana Rahmat Zaki,[2] and Mohammad Faisal[3‡]
[1,2] Dept. of Electrical and Electronic Engineering, Eastern University, Dhaka
Dhaka-1205, Bangladesh
[3] Dept. of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka
Dhaka-1000, Bangladesh
‡mdfaisal@eee.buet.ac.bd

***Abstract*—In this paper the optimal allocation of fiber Bragg grating (FBG) as a dispersion compensator in the presence of group velocity dispersion (GVD) and third-order dispersion (TOD) has been examined. The optimization is performed by varying the compensator arrangements and span length. Four different arrangements are considered–pre-configuration, post-configuration, mid-configuration and bi-end configuration. Span length is varied from 30 to 120 km. At the receiver, Q-factor and bit error rate (BER) have been systematically compared for standard single mode fiber (SSMF) and non-zero dispersion-shifted fiber (NZDSF). The results show that NZDSF based long-haul lightwave communication system with mid compensated scheme provides better performance at 160-Gb/s.**

***Index Terms*—FBG, coupled-mode theory, dispersion compensation, BER, Q-factor.**

I. Introduction

Recent growth in telecommunication services and upcoming technologies like video conferencing, HD video streaming, interactive multimedia and a tremendous increase in the number of the end users have set a limit to the current capacity of optical fiber communication system. To meet the ever increasing demand data rate must also be increased to a higher level. With the increase in data rate the detrimental effects of chromatic dispersion has also become more prominent. Studies on dispersion compensation is thus a matter of acute importance for next generation of optical fiber communication.

Several techniques such as fiber Bragg grating (FBG) [1], dispersion compensating fibers (DCF) [2], optical phase conjugator (OPC) [3] and electronic compensation [4] have been employed to lessen dispersion at high-speed links. The conventional DCF provides high loss and increased Kerr nonlinearity as a result of its attempt to reduce core size to achieve the required negative dispersion. OPC might be a good alternative to DCF but due to the mitigation of attenuation by periodic amplification using Erbium doped fiber amplifier (EDFA), asymmetric signal power evolves in the fiber with respect to the phase conjugator position, which degrades the system performance [3]. Although today's signal processing ensures total compensation of chromatic dispersion in electronic domain, its clumsy and time consuming property increases the possibility of data packet loss.

At ultra-high speed system performance is degraded not only by group velocity dispersion (GVD) but also by third-order dispersion (TOD). Among the other competing dispersion compensators FBG is proven to be an efficient means to compensate dispersion both for GVD and TOD [1], [5]–[8]. The comparison of FBG based system with the popular dispersion compensating fiber is presented in [9]–[11]. Due to the accessibility of low loss high performance FBG in optical fiber communication system, the dispersion management technique has gone through significant improvement during the past decade. The dispersion value from −495.5 ps/nm to −1406 ps/nm and dispersion slope of −43.3 ps/nm^2 is already reported [6], [8].

Although a number of research work is found unveiling the opportunities of FBG as both GVD and TOD compensator, there is hardly any optimization routine at ultra high-speed. The optimization at 10-Gb/s was previously investigated in [7]. However the authors considered GVD compensation only for NRZ data format. The current work presents an optimization of the appropriate location of FBG in the transmission line so that GVD and TOD will be compensated substantially. A number of compensator arrangements in the presence of self-phase modulation, chromatic dispersion and polarization mode dispersion are explored. In this paper four different orientations namely, pre-configuration, post-configuration, mid-configuration and bi-end configuration are investigated thoroughly. In addition to that, repeater span length is varied to find the best performance for a given configuration.

II. Theoretical Modeling

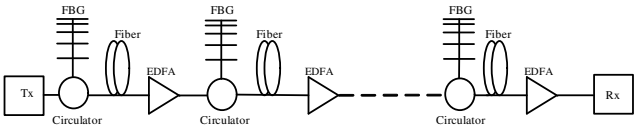
A. Mathematical Analysis

The optical pulse propagation through the fiber is simulated using non-linear Schrödinger equation. However, the transfer function of chirped FBG is modeled by solving coupled-mode equations. FBG is used in reflective mode to achieve the required negative dispersion. For reflection of a particular wavelength λ_B, it needs to satisfy the Bragg condition $\lambda_B = 2n_{eff}\Lambda_B$, where n_{eff} and Λ_B are the effective refractive index of the core and chirping period for the wavelength λ_B. Various appodization function and chirping are used to modulate the refractive index.The refractive index of apodized and linearly chirped FBG can be described as,

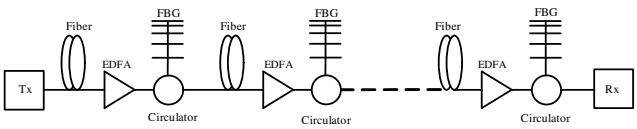
$$n(x,y,z) = n_{co}(x,y) + \delta n_{dc}(x,y,z) + \delta n_{ac} A(z) f(\Lambda(z)/\cos(\theta), z) \quad (1)$$

$$\delta n_{ac}(z) = \delta n_{ac} A(z) f(\Lambda(z)/\cos(\theta), z) \quad (2)$$

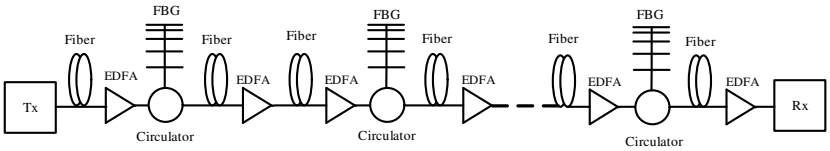
978-1-4799-4166-7/14/$31.00 ©2014 IEEE



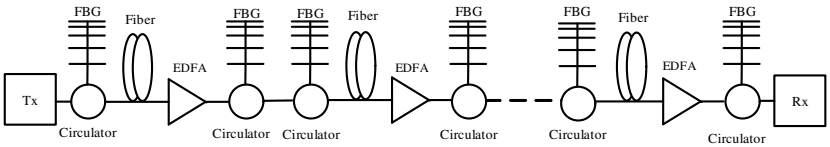
(a) pre-configuration.



(b) post-configuration.



(c) mid-configuration.



(d) bi-end configuration.

Fig. 1. Schematic diagram of the system considered.

TABLE I
PROPOSED SCHEMES

Model	Fiber				Dispersion Compensator Orientation
	Type	Attenuation (dB/km)	Dispersion (ps/nm/km)	Dispersion Slope (ps/nm^2/km)	
Scheme I	SSMF / NZDSF	0.2	17 / 4	0.07	Pre-configuration
Scheme II	SSMF / NZDSF	0.2	17 / 4	0.07	Post-configuration
Scheme III	SSMF / NZDSF	0.2	17 / 4	0.07	Mid-configuration
Scheme IV	SSMF / NZDSF	0.2	17 / 4	0.07	Bi-end configuration

where, n_{co} is the refractive index of unperturbed core, $\delta n(z)$ is the peak to peak index modulation amplitude and it includes the effect of direct index change δn_{dc}, index change due to apodization δn_{ac} and refractive index variation due to appodization and chirping. f, $A(z)$ and $\Lambda(z)$ are shape function, apodization and period chirp function respectively. The chirping function can be expressed as

$$\phi(z) = \frac{8\pi n_{co}^2 L_g^2}{\lambda_B^2 c D_f L_f} \times \frac{z^2}{2L_g^2}, \tag{3}$$

where, L_g, λ_B, D_f, L_f are length of the grating, Bragg wavelength, dispersion of the fiber and length of the fiber respectively. Coupled-mode equations are used to describe the pulse propagation along the FBG. If U and V are forward and backward propagating wave in a periodic structure it can be expressed as,

$$\frac{\partial U}{\partial z} = i\sigma U + i\kappa V \tag{4}$$

$$\frac{\partial V}{\partial z} = -i\sigma V - i\kappa U. \tag{5}$$

where,

$$\kappa = \frac{\pi \delta n_{ac}(z)}{\lambda}, \ \psi = \frac{\pi \delta n_{dc}(z)}{\lambda} \tag{6}$$

$$\delta = \frac{2n_{eff}}{c}\left(\frac{2\pi c}{\lambda} - \frac{2\pi c}{\lambda_B}\right) \tag{7}$$

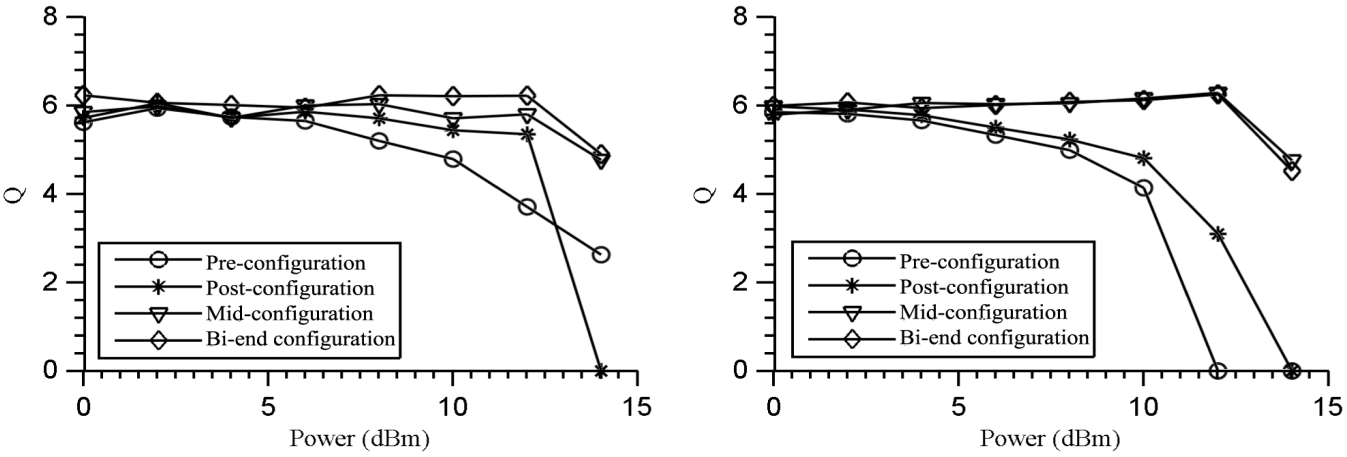
$$\sigma = \delta + \psi - \frac{1}{2}\frac{d\phi}{dz}. \tag{8}$$

The parameter κ and δ are called the alternate coupling coefficient and the detuning of the grating. The coupled-mode equations are solved using transfer-matrix method to find the transmitted and reflected wave, delay and dispersion of the FBG.

B. Description of the System

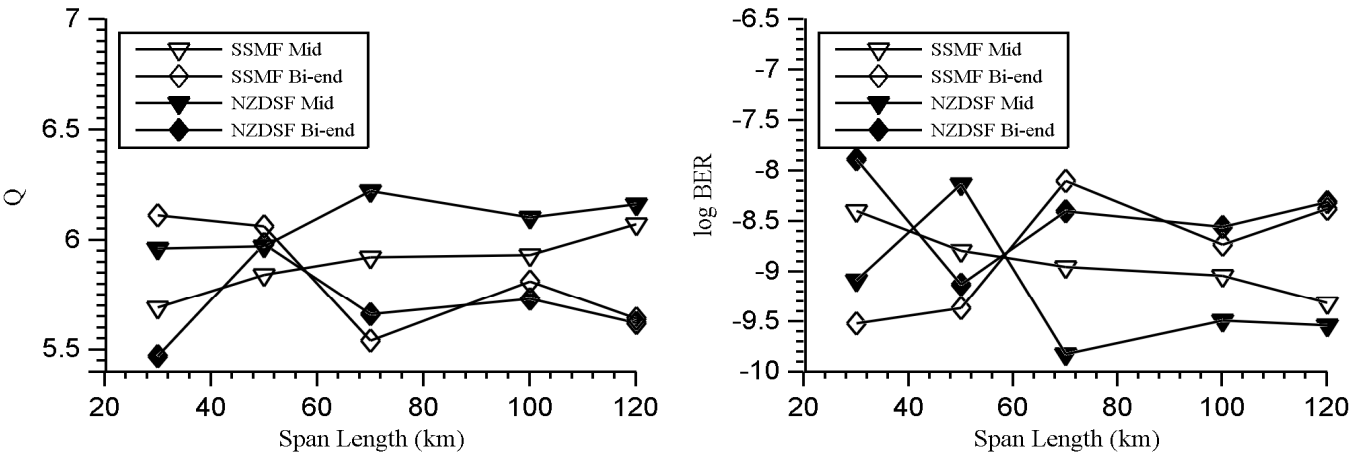
The different models considered for optimization is summarized in Table I. Four different configurations are used and named according to the location of FBG dispersion compensator in the transmission line. Scheme I is called pre-configuration where dispersion is compensated prior to transmitting the signal to the fiber using FBG. Post-configuration in Scheme II does exactly the opposite of Scheme I, it compensates dispersion after the signal go through the transmission line. Mid-configuration in Scheme III uses FBG to compensate dispersion at the mid of the line. Scheme IV refers to biend-configuration that compensate dispersion at both starting and terminating end.

The transmission model basically consists of three parts: transmitter, transmission line (that includes the fiber, dispersion compensator and amplifier) and receiver as shown in Fig. 1. The transmitter section is composed of a bit sequence generator, an electric pulse generator, an optical source (CW laser) and an optical modulator (Mach-Zehnder (MZ) modulator). Following the bit sequence generator, which generates



(a) SSMF (b) NZDSF

Fig. 2. Q-factor as a function of input power for two different types of fiber.



(a) (b)

Fig. 3. (a) Q-factor as a function of repeater span. (b) BER as a function of repeater span.

a PRBS of 2^7–1, the electric pulse generator generates binary sequence of data which is used to modulate the optical signal from CW laser at MZ modulator. Transmitter section produces RZ Gaussian pulses of 50% duty cycle at center emission wavelength of 1550 nm at 160 Gb/s.

Two different models of single channel single-mode fiber, namely standard single mode fiber (SSMF) and non-zero dispersion shifted fiber (NZDSF) have been considered in our work. Both of these fibers are individually checked with each of the configurations described in Table I. SSMF has an effective core area of $70\,\mu\text{m}^2$ and nonlinear index coefficient of 2.6×10^{-20} m^2/W whereas NZDSF an effective core area of $80\,\mu\text{m}^2$ and nonlinear index coefficient of 2.6×10^{-20} m^2/W. PMD coefficient of 0.5 ps/$\sqrt{\text{km}}$ is used in this paper. The specification of the fiber parameters are summarized in Table I. GVD, TOD and SPM are included inside the fiber. For dispersion compensation FBG is used. FBG offers low loss and negligible nonlinearity in practice, so for simulation purpose both of these effects are neglected in FBG. An EDFA is used to compensate for the attenuation due to fiber loss at the end of each repeater span. However EDFA noise is not considered. The whole length of 2000 km is covered by a recirculting loop at the repeater span.

At the receiver end the optical signal is detected by a PIN detector followed by a fourth order Bessel type low pass electrical filter with cutoff frequency 0.75×bit rate. The PIN has a responsivity of 1 A/W and a dark current of 10 nA.

III. Optimization

To begin with the optimization we first decide the best scheme from Table I. For this purpose we have tested the Q-factor of the system both for SSMF and NZDSF type of fiber. The variation is observed with the increase in input power.

Fig. 2 shows the input power versus Q-factor curve for the four proposed schemes. Fig. 2(a) shows the variation for SSMF fiber and the Fig. 2(b) shows the variation for NZDSF fiber. In case of SSMF fiber Q-factor degrades with the increase in input power. Bi-end configuration performs best among the four schemes. The Q-factor is 5.61, 5.71, 5.84 and 6.22 for the four configurations respectively at 0 dBm of input power which degrades to 2.62, 0, 4.76 and 4.89 at 14 dBm of input power. In reference to Fig. 2(b), both mid and bi-end compensations work well. At 0 dBm the Q-factor is 5.84, 5.78, 5.97 and 5.98 for pre-, post-, mid- and bi-end configuration which degrades to 0, 0, 4.75 and 4.71 respectively for the four models.

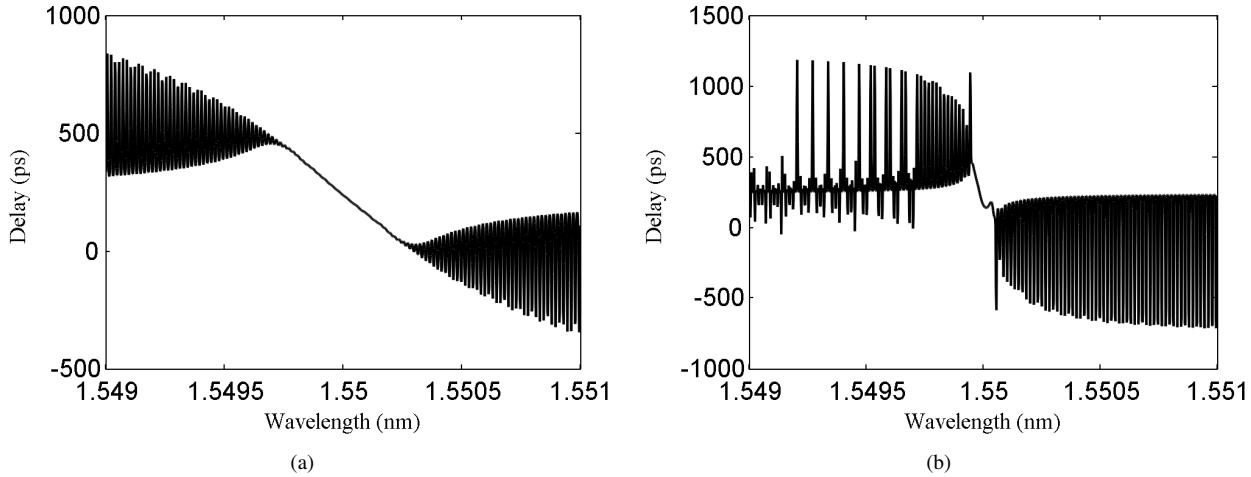


Fig. 4. (a) Group delay ripple in delay line for lower GVD. (b) Group delay ripple in delay line for higher GVD.

In comparison among the four models, mid- and bi-end configuration performs better than pre- and post-configuration. So in the next simulation of varying the repeater span we consider only the mid- and bi-end configuration. The Q-factor and BER is monitored. The results of varying the repeater span are included in Fig. 3. The repeater span is varied from 30 to 120 km with an increment of 20 km and tested for Q-factor and BER both for SSMF and NZDSF fiber. From the figures above it can be concluded that SSMF fiber gives the best performance at repeater span of 50 km (Q-factor=6.06, BER=4.25×10^{-10}) and NZDSF fiber at 70 km (Q-factor=6.22, BER=1.49×10^{-10}). It can also be concluded that mid-configuration with NZDSF fiber works slightly better than bi-end configuration with SSMF fiber.

The physical reason underlying the above conclusion can be explained with Fig. 4. In Fig. 4(a) we can see the delay line with negligible ripple inside the passband. But with the same device when we try to increase the delay to meet the required higher level of accumulated dispersion by the fiber, an enormous amount of ripple show off as shown in Fig. 4(b). We call this the group delay ripple (GDR) and this degrades the system performance of FBG as a consequence. FBG can operate good within certain amount of negative GVD. Beyond this certain limit it's performance is limited by the GDR. Mid-configuration with NZDSF fiber and bi-end configuration with SSMF fiber meet FBG within this specific limit and as a result performs better than other configurations.

IV. Conclusion

In this paper we present an optimization routine to compensate GVD and TOD using FBG at 160 Gb/s long haul optical fiber transmission. We find that received signal gets degraded with the increase in input optical power due to the increased nonlinear effects. Apparently bi-end and mid-configurations offer better performance than that of pre and post configurations. Bi-end configuration works better with SSMF fiber whereas mid-configuration with NZDSF fiber. While changing repeater span, SSMF based model performs excellent at 50 km of repeater span whereas NZDSF based model with 70 km of repeater span. The overall best performance is achieved for NZDSF based model with repeater span 70 km among the all combinations.

References

[1] B. Dabarsyah, C. S. Goh, S. K. Khijwania, S. Y. Set,K. Katoh and K. Kikuchi, "Adjustable Group Velocity Dispersion and Dispersion Slope Compensation Devices With Wavelength Tunablity Based on Enhanced Thermal Chirping of Fiber Bragg Gratings," *J. Lightwave Technol.*, vol. 25, no. 9, pp. 2711-2718, Sept. 2007.

[2] F. R. Zaki and M. Faisal, "Impact of third-order dispersion in ultra-high speed long-haul optical fiber communication system," in Proc. of International Conference on Informatics, Electronics & Vision (ICIEV), 2013, Dhaka, Bangladesh, pp. 1–5, 17-18 May 2013.

[3] J. Li, K. Xu, G. Zhou, J. Wu, and J. Lin, "Dispersion-Compensation Scheme for 160-Gb/s 1200-km Transmission by Optical Phase Conjugation,"*J. Lightwave Technol.*, vol. 25, no. 8, pp. 1986-1995, Aug. 2007.

[4] J. M. Gene, P. J. Winzer S. Chandrasekhar, H Kogelnik, "Simultaneous Compensation of Polarization Mode Dispersion and Chromatic Dispersion Using Electronic Signal Processing," *J. Lightwave Technol.*, vol. 25, no. 7, pp. 1735-1741, Aug. 2007.

[5] P. Fernandez, J. C. Aguado, J. Blas, R. Duran, J. Duran, I. de Miguel, R. Lorenzo and E. Abril, "Optimization of the apodization strength of linearly chirped Bragg grating for span dispersion compensation," *Opt. and Quantum Electron.*, vol 36, pp. 57-66, 2004.

[6] J. Kim, J. Bae, Y. G. Han, S. H. Kim, J. Jeong and S. B. Lee, "Effectively Tunable Dispersion Compensation Based on Chirped Fiber Bragg Gratings Without Central Wavelength Shift," *IEEE Photonics Technol. Lett.*, vol. 16, no. 3, Mar. 2004.

[7] K. Ennser, M. N. Zervas, R. I. Laming, "Optimization of Linearly Chirped Grating Dispersion-Compensated Systems," *IEEE Photonics Technol. Lett.*, vol. 3, pp. 120-122, Aug. 1997.

[8] J. Zhang, P. Shum, X. P. Cheng, N. Q. Ngo, and S. Y. Li, "Analysis of Linearly Tapered Fiber Bragg Grating for Dispersion Slope Compensation," *IEEE Photonics Technol. Lett.*, vol. 15, no. 10, Aug. 2003.

[9] A. Dochhan, S. Smolorz, H. Rohde, and W. Rosenkranz, "FBG dispersion compensation in a 43 Gbit/s WDM system: Comparing different FBG types and modulation formats," in 11th International Conference on Transparent Optical Networks, (ICTON 2009), Azores, pp. 1-4, June 28-July 2, 2009.

[10] D. Tripathi, P. Singh, N. Shukla, and H. Dixit, "Performance study in disperion compensation techniques with duobinary format at different bit rates," in 2nd International Conference on Power, Control and Embedded Systems (ICPCES), 2012 , Allahabad, pp. 1-5, 17-19 December 2012.

[11] M. Al-Noor, J. Loo, R. Comley, "WiMAX 54 Mbit/s over radio over fiber using DCF, SMF and FBG for fibre over 410km," in 7th International Symposium on Wireless Communication Systems (ISWCS), 2010 , York, UK, pp. 829–833, 19-22 September 2010.

Channel Capacity and Bit Error Rate Analysis of Non-Uniform Type M-Ary Optical CDMA in Presence of Self Phase Modulation and Dispersion

Rashad Kabir, Md. Maruful Haque, and Moshiour Rahaman

Department of Electrical and Electronic Engineering, Khulna University of Engineering & Technology
Khulna-9203,Bangladesh

rashad6968@gmail.com

Abstract— **The Bit Error Rate(BER) and theoretical channel capacity of OCDMA transmission has been analyzed in this paper for M-ary chip symbol. A comparison has also been made between uniform and non-uniform signalling and the result shows that channel capacity is increased for the same number of active users if the modulation level or bits per OCDMA symbol is increased for non-uniform. Finally SNR and BER simulation result has been shown for both uniform and non-uniform M-ary OCDMA transmission system in presence of Self Phase Modulation and dispersion. The simulation result shows that unlike conventional OCDMA system better performance can be achieved using M-ary Chip Symbol.**

Index Terms—**M ary Chip Symbol, Gold Sequence, Channel Capacity, Self Phase Modulation.**

I. Introduction

Optical Code Division Multiple Access (OCDMA) can provide numerous smart features such as asynchronous, flexible and decentralized networks, potential of high data throughputs, inherent data security, potential of total bandwidth utilization by all network users and many more[1]-[9]Since users in OCDMA networks are non-orthogonal and asynchronous, severe multi-user interference (MUI) has been a crucial factor which is responsible for limiting the performance of OCDMA channel transmission [4]. Besides the effect of the non-linearity such as self-phase modulation, cross phase modulation, Stimulated Brilloiun Scattering, Stimulated Raman Scattering, Four Wave mixing and various kinds of dispersion such as chromatic dispersion, Polarization Mode Dispersion, Modal Dispersion are also responsible for degrading the performance of an OCDMA transmission system[5]-[7].

This paper illustrates the results that the instead of the traditional binary OCDMA, utilizing M-ary OCDMA system; where symbol is constructed by several bits a higher spectral efficiency can be achieved [8]-[9]. The mathematical modelling and simulation has been carried out considering the effect of the dispersion and noise. We present relevant channel capacity results and simulated bit error rate (BER) performance of a coded M-ary OCDMA system to show that as much as the modulation level is increased, the data rate is also increased consequently when compared to binary OCDMA systems. As in most prior literature, single-user demodulation and decoding (SUD) has been focused in this research due to practical implementation complexity constraints and signal processing speed requirements.

The rest of this paper is organized as follows: the next section includes system models including Self Phase Modulation(SPM), noises and MUI, Section III presents numerical capacity results illustrating the achieved capacity considering M-ary OCDMA transmission, BER performance for different modulation level and in addition to SNR for different modulation level has also been analysed. Section IV concludes this paper and presents directions of future work.

II. Proposed System Model

Figure 1 illustrates M-ary OCDMA transmission architecture with single-user demodulation at the receiver where K independent users send information data simultaneously over a fiber-optic channel. In the transmitter, a user's binary data is converted to symbol by bit to symbol converter. During the modulation, transmission and demodulation processes, the optical signals are assumed to add up incoherently without beatings and optical intensity direct detection is used at the receiver. In the receiver, a particular user data symbol is recovered by the correlation operation using an optical correlator receiver. The decoded signal is incident on the avalanche photo detector. a level detector is used to determine the output signal magnitude and then the output symbol is recovered.

In our paper it is assumed that the data signal and the chip signal has no relative delay although the signals from various users may have a certain amount of delay. This delay is not shown in the simulator. But its effect on BER is calculated. The product signals depicted in the block diagram of the simulator is the multiplication of data sequence and chipping sequence of each user. The product signal for k-th user can be represented by

$$S_k(t-\tau_k) = V\alpha_k(t-\tau_k)b_k(t-\tau_k)\mu_k(t-\tau_k) \quad (1)$$

Here $b_k(t)$ is the data sequence for user k, $\alpha_k(t)$ is the spreading sequence (chip) sequence for user k, $\tau_k$is the delay of user k relative to some user reference user 0, V is the amplitude of the signal and $\mu_k(t)$ is the Gaussian function. Both $\alpha_k(t)$ and $b_k(t)$ are binary sequence with values -1 and +1 . In this simulator, it is assumed that the total chipping sequence is repeated for each bit period. The chip sequence $\alpha_k(t)$ is of the form

$$\alpha_k(t) = \sum_{j=-\infty}^{\infty}\sum_{i=0}^{N-1}\alpha_{k,i}\prod\left(\frac{t-(i+jN)T_c}{T_c}\right) \quad (2)$$

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

N is the spreading factor of the number of the chips ber bit, T_c is the chip period, NT_c is the repetition period of the chipping sequence = Bit period of the users (T_B).

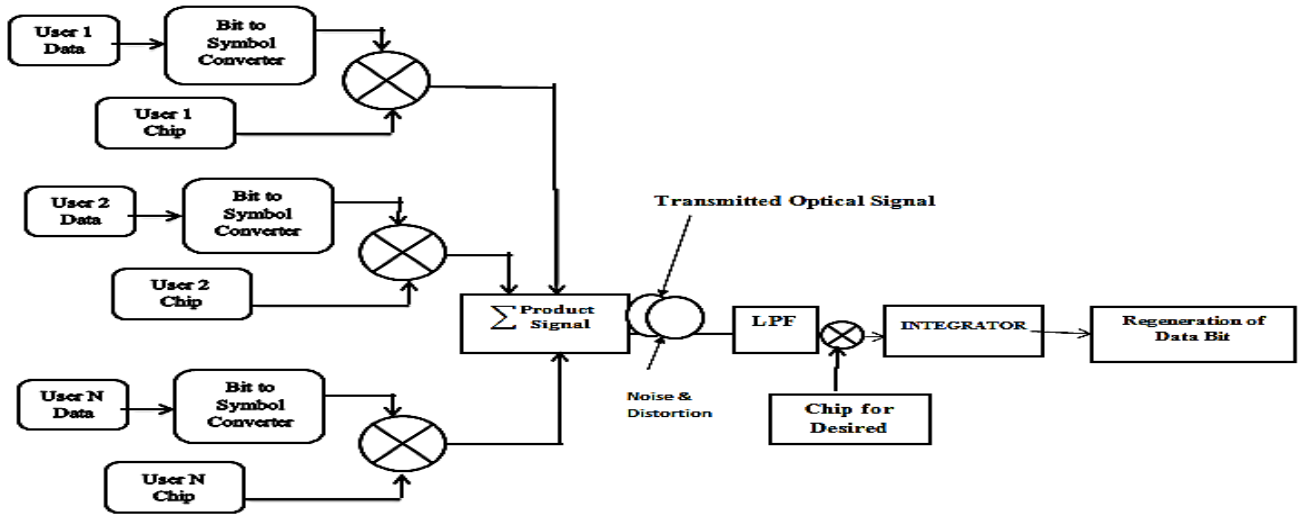


Fig 1 Block Diagram of the System Model.

$\prod(t)$ represents the unit impulse function and i is an index to donate the particular chip within a PN chip sequence

$$\prod = \begin{cases} 1 & 0 < t < 1 \\ 0 & otherwise \end{cases} \quad (3)$$

The user data sequence $b_k$is given by

$$b_k(t) = \sum_{j=-\infty}^{\infty} \sum_{i=0}^{N-1} b_{k,i} \prod \left(\frac{t-(i+jT_b)}{T_b}\right) \quad (4)$$

The Gaussian function $\mu_k(t)$ is given by

$$\mu_k(t) = \exp\left(-\frac{1-i\alpha}{2} \cdot \left(\frac{t}{t_0}\right)^{2m}\right) \quad (5)$$

If the total number of users is k then the compound signal that is to be transmitted through the optical fiber is:

$$r(t) = \sum_{k}^{K-1} S_k(t-\tau_k) \quad (6)$$

Now if the transfer function of the optical fiber be$H(f)$, then the Fourier transform of the fiber output signal becomes:

$$G(f) = r(f)H(f) \quad (7)$$

Therefore the fiber output signal is

$$g_{out}(t) = F^{-1}(G(f)) \quad (8)$$

To observe the effect of noise in a digital communication system in this simulator there is a provision for adding a particular amount of noise. Now the noisy signal becomes

$$g_{noisy(t)} = g_{out}(t) + g_{noise}(t) \quad (9)$$

Where g_{noise} is considered as a summation of thermal noise, shot noise, bulk current noise, dark current noise. Now if the Fourier transform of the noisy signal $g_{noisy}(t)$ to be $G_{noisy}(f)$ and the transform function of the low pass filter $H_{filter}(f)$, then inverse Fourier Transform of the output of the fiber will be

$$r_{out}(t) = F^{-1}[G_{noisy}(f).H_{filter}(f)] \quad (10)$$

If the output of the low pass filter be $r_{out}(t)$ then the decision statistic for k'th user is

$$Z_0 = \int_{jT_B}^{(j+1)T_B} r_{out}(t)\,\alpha_k(t) \quad (11)$$

The output symbol magnitude of k-th user is

$$Z_0 = |y_{out}(t)| \quad (12)$$

III. Simulation Result And Analysis

A. Channel Capacity for OCDMA Transmission

As one seeks ways of improving the throughput of an OCDMA system, it is important to know the absolute maximum throughput. This gives insight in handling a suitable channel coding strategy to allow reliable communication. The concepts in information theory .can be applied to any channel model to determine the maximum rate at which data can be sent across a channel with error probability approaching zero as the block length is made arbitrary large. Fundamentally, the channel capacity is given by the source probability that maximizes the mutual information of the channel. For a discrete-memory less channel, the channel capacity C in bits per channel use is given by

$$C = maxI(X;Y) = \max[H(Y) - H(Y|X)] = max[H(X) - H(X|Y)] \quad (13)$$

To evaluate the performance of channel model under the condition of interface the following near-optimal channel input Probability Mass Function (PMF) is used :

$$\rho_0 = 1 - \frac{1}{K} and\ \rho 1 = \rho 2 = \cdots = \rho M - 1 = \frac{1}{K(M-1)} \quad (14)$$

It I observed form Fig. 2 that, when uniform signaling is used i.e all M-ary inputs are used with equal probability and the number of users K increases, the OCDMA spectral efficiency limit converges to 0.72 bits per OCDMA chip, despite the fact of the modulation levels M increases from M = 2 to M = 4,8

or even 16. Another important factor that should be noticed from the figure is that when non-uniform signaling is considered for analysis, as the modulation level increases the OCDMA Network throughput i.e bits/OCDMA chip also increases. Whereas capacity remains around 0.87 bits/OCDMA chip for M=2, it becomes 2.1 bits/OCDMA chip when modulation level M=16 is used which means that the capacity is increased near about 2.5 times.

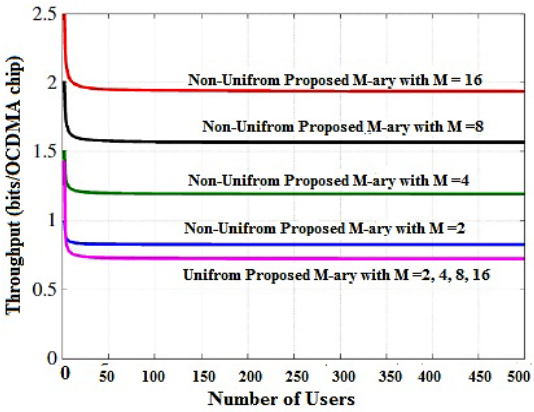


Fig. 2 Theoretically achievable rates for an OCDMA network transmission both for uniform and nonunifrom signal.

Another key factor which limits the performance is the Multi User Interference (MUI).It is expected that increasing the number of modulation levels is expected to increase the amount of MUI, but the use of the non-uniform input PMF from Eq. (14) keeps the MUI reduced. Figure 3illustrates the reduced MUI obtained with non-uniform channel input PMF and compares it to the much larger amount of MUI resulting from uniform channel input signaling. This reduction of the MUI through non-uniform signaling allows increase of the theoretically achievable spectral efficiency to almost 2.1 bits per chip for M =16.

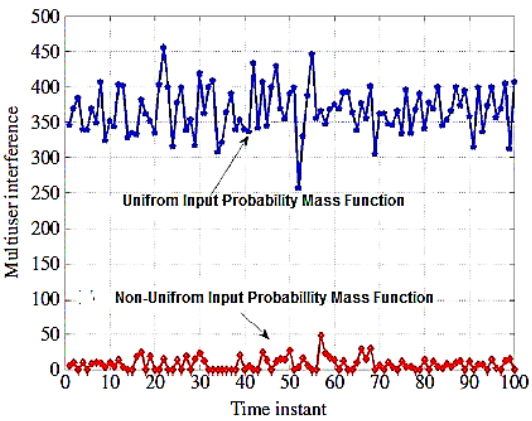


Fig. 3 Comparison of the amount of OCDMA MUI in the cases of uniform and non-uniform channel input PMF.

Another key observation is that the MUI, when using the non-uniform channel input PMF from Eq. (14), turns out to be strongly reduced and *non-Gaussian*in nature, as illustrated in fig. 4 for an OCDMA network transmission.

The diagram has shown for modulation level M=16 obtained from equation (14). On the contrary, conventional OCDMA network transmission with single-user decoding using uniform chip symbols, and the MUI converges in distribution to a Gaussian random variable. The reason behind this is the effect central limit theorem. In such a case, the transmitted data end up being severely corrupted by the MUI, which is also shown for illustration purposes in Fig. 4.

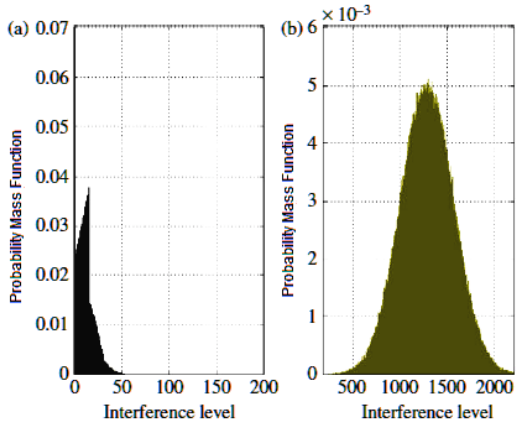


Fig. 4 Histogram of MUI for 16-ary (M = 16) OCDMA transmission (a) Non-Gaussian MUI for non-uniform signalling (b) Gaussian interference with uniform PMF

From this observation we can conclude that, since non-uniform channel input signaling leads to a reduced MUI and larger achievable transmission rates, we will consider OCDMA transmission with non-uniform channel input.

B. Performance Analysis in terms of BER and SNR

Bit error rate is defined as the rate at which errors occur in a transmission system. This can be directly translated into the number of errors that occur in a string of a stated number of bits. The definition of bit error rate can be translated into a simple formula:

$$Bit\ Eror\ Rate\ BER = \frac{Numbers\ of\ Error}{Total\ Numbers\ of\ bit\ Sent}$$

In practice, the bit error rate of a system for optical data transmission (e.g. a fiber-optic link) can be increased by noise influences (particularly in the receiver, but also in the transmitter and in amplifiers), by optical losses, and chromatic and other types of dispersion. On the other hand, Signal to noise ratio (SNR) is a measure of the amount of good signal divided by the amount of noise being received.

The Bit Error Rate (BER) or Probability of Error is the probability to occur an error during the transmission of a bit. It is the sum of the probabilities of all the ways that an error can occur. BER for transmission of binary digits in presence of a Gaussian distributed noise having a noise variance σ^2 is

$$P_{B=}Q\left(\frac{\alpha_1-\alpha_2}{2\sigma_0}\right) \quad (15)$$

Where $Q(x)$ is the complementary error function or co-error function and $\alpha_1$and $\alpha_2$be the voltage levels for two bits.

Again Signal to Noise Ratio (SNR) $=(\frac{S}{N})_T=\left(\frac{\alpha_1-\alpha_2}{\sigma_0}\right)^2$ (16)

Therefore $$P_B = Q\left(\frac{\sqrt{(\frac{S}{N})_T}}{2}\right) \quad (17)$$

For Matched Filter $$(\frac{S}{N})_T=(\frac{S}{N})_{T,max} = \frac{2E_d}{N_o} \quad (18)$$

Where E_d = Average Differential Energy for filter input. Differential energy is taken to include the effect of bipolar signals in consideration. Therefore,

$$P_B = Q\left(\frac{\frac{2E_d}{N_o}}{2}\right) = Q\left(\sqrt{\frac{E_d}{2N_o}}\right) \quad (19)$$

But $$Q(x) = \frac{1}{2}\, erfc\left(\frac{x}{\sqrt{2}}\right) \quad (20)$$

Therefore $$P_B = \frac{1}{2}\, erfc\left(\frac{\left(\sqrt{\frac{E_d}{N_o}}\right)}{2}\right) \quad (21)$$

For convection Filters,

$$\left(\frac{S}{N}\right)_T = \frac{\alpha_i^2}{\sigma_0^2} \quad (22)$$

Where signal output $\alpha_i = F^{-1}$ (Filter Transfer Function * F (input of Filter Signal)) = Output signal of Filter. To include also the effect of $\alpha_i$is taken as follows

$$\alpha_i = mean\big(abs(s_1 - s_2)\big) \quad (23)$$

Where s_1 and $s_2$be the voltage levels of two bits. Therefore,

$$\left(\frac{S}{N}\right)_T = \frac{\left[mean\big(abs(s_1 - s_2)\big)\right]^2}{\sigma_0^2} \quad (24)$$

$$P_B = Q\left(\frac{mean\big(abs(s_1 - s_2)\big)}{2\sigma_0}\right) \quad (25)$$

$$= \frac{1}{2}\, erfc\left(\frac{mean\big(abs(s_1 - s_2)\big)}{2\sqrt{2}\sigma_0}\right) \quad (26)$$

Figure 5 illustrates the BER performance versus Received Optical Received Power(ORP) for up to M=64 modulation level. We observe again that BER decreases when the ORP and the number of users is increased.

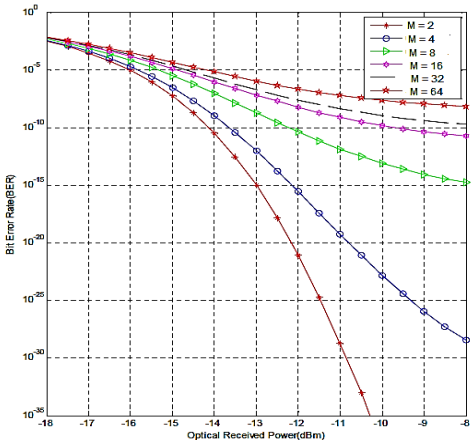


Fig. 5 Performance Comparison – BER vs. Received Optical Power

For instance, when we consider a 10-5 BER, the ROP is -14.8 dBm for M=32 modulation level while for M=64, this becomes -14.2 dBm.

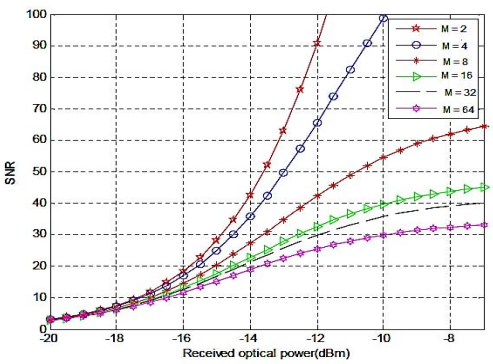


Fig. 6. Performance Comparison – SNR Vs. Received Optical Power

Similarly, Signal to Noise Ratio influences the receiver section of the OCDMA system. If SNR is increased, it will increase the performance of the system. In our paper performance of our proposed M-ary system has been analyzed in terms of SNR. Figure 7 presents the SNR versus ROP of the OCDMA system for up to 64 modulation level.

For instance, it is observed, that when the SNR is 30 dB, the ROP is -12 and -10 dBm for M-=32 and M=64 modulation level respectively.

IV. Conclusions

In this paper, the channel capacity limits on optical CDMA transmission is evaluated, when single-user detections used at the receiver. It has been seen that the *M*-ary modulation for OCDMA systems can be used to increase the capacity limitations of the conventional binary optical CDMA. It is illustrated that the channel capacity is increased with the increase of the modulation level and increases up to 2.1 bits/OCDMA chip for M=16.Furthermore, the proposed scheme is robust against several impairments which limit the performance such as Self Phase Modulation, MUI, Poisson shot noise and AWGN thermal noise.

It is attention-grabbing to note that the considered non-uniform M-ary OCDMA transmission scheme achieved a spectral efficiency which is significantly better compared to the best uniform M-ary OCDMA. Finally with the analysis of performance based on BER and SNR, it can concluded that the improved performance achieved by the designed M-ary OCDMA system..

Acknowledgment

The authors are extremely thankful to Communication and Simulation laboratory of Khulna University of Engineering & Technology (KUET), Khulna for providing the necessary support in this research.

References

[1] *Coding in Optical CDMA Network with M-ary Modulation: Impact of Selected Physical Imperfections,"* 0-7803-8886-0/05/$20.00©2005 IEEE, CCECE/CCGEI, Saskatoon, May 2005.

[2] M. J. Islam and M. R. Islam. "*Impact of GVD and SPM onthe Performance of DS-OCDMA,''* Academy Publisher Journal of Networks, vol. 06, no. 01, Jan. 2011, pp. 12-17.

[3] S.P. Majumder, Afreen Azhari, and F.M. Abbou, *"Impact offiber chromatic dispersion on the BER performance of an opticalCDMA IM/DD transmission system," IEEE photon. technol, lett.*,vol. 17, no. 6, June 2005, pp. 1340-1342.

[4] A. A. Garba and J. Bajcsy, *"Using turbo codes and non-Gaussian interference to improve spectral efficiency in M-ary optical CDMA networks,"* in Proc. Int. Symp. Turbo Codes Relat. Topics, Apr. 2006, pp. 1–6.

[5] A. A. Farid and S. Hranilovic, *"Channel capacity and non-uniform signaling for free-space optical intensity channels*,"IEEE J. Sel. Areas Commun., vol. 27, no. 9, pp. 1553–1563, Dec. 2009.

[6] E. E. Narimanov and P.Mitra, "*The channel capacity of a fiberoptics communications system: perturbation theory," J. Light wave Technol.*, vol. 20, no. 3, pp. 530–537, 2002.

[7] T. Ohtsuki, "*Performance analysis of direct-detection optical asynchronous CDMA systems with double optical hard limiters,"* J. Lightw. Technol., vol. 15, no. 3, pp. 452–457, Mar. 1997.

[8] M. R. Dale and R. M. Gagliardi, "*Channel coding for asynchronous fiber optic CDMA communications," IEEE Trans. Commun.*, vol. 43, no. 9, pp. 2485–2492, 1995.

[9] S K.-I. Kitayama, "Code division multiplexing lightwave networks based upon optical code conversion," IEEE J. Sel. Areas Commun.,vol. 16, pp. 1309–1319, 1998.Aminata A. Garba and Jan Bajcsy.

Design of a High Speed OFDM Transmitter and Receiver

Foisal Ahmed,[1,*] Md. Liakot Ali,[2] and Mohammad Imam Hasan Bin Asad[2]

[1]Dept. of Electronics and Telecommunication Engineering
Prime University, Dhaka-1216, Bangladesh
[2]Institute of Information and Communication Technology (IICT)
Bangladesh University of Engineering and Technology (BUET), Dhaka-1000, Bangladesh
[*]faisal_ete@yahoo.com

Abstract— **System–on-a-chip (SoC) is now a trend in digital design because it gives a lot of advantages over discrete electronic based product such as higher speed, lower power consumption, smaller size, lower cost etc. Reconfigurable platforms such as FPGA, CPLD, and PLD are now being used for designing and implementing SoC due to their low cost, high capacity and tremendous speed. In this project single chip Orthogonal Frequency Division Multiplexing (OFDM) transmitter and receiver have been designed using Verilog HDL. OFDM is a multi carrier modulation technique used in the various digital communication systems like 3G GSM, WiMAX and LTE etc. The main advantage of this transmission technique is its robustness to channel fading in wireless communication environment. There are many applications of OFDM in communication such as digital audio broadcasting, Asynchronous Digital Subscriber Line (ADSL) and High bit-rate Digital Subscriber Line (HDSL) systems. In OFDM, two algorithms digital signal processing algorithm Fast Fourier transform (FFT) and Inverse Fast Fourier Transform (IFFT) are mainly involved. The 8-point IFFT/FFT Decimation-In-Frequency (DIF) with radix-2 algorithm has been analyzed and incorporated in the design. The design has been simulated on the FPGA platform with Altera's Quartus II simulator. Simulation results show that each of the modules of the proposed OFDM is working as desired. The test output achieved from the simulation result of the OFDM has been verified with that of the MATLAB output.**

Keywords—OFDM, FFT, IFFT, FPGA, Verilog HDL

I. INTRODUCTION

Reconfigurable computing is a computer architecture, which combines software with the high performance hardware by very flexible high speed computing fabrics like field-programmable gate arrays (FPGAs).This reconfigurable platform is now being used for designing and implementing SoC due to its low cost and high capacity and tremendous speed [1].

Designing various modern communication standards require digital signal processing throughput. Modeling theses standards on general purpose CPUs can be extremely slow with traditional software. Besides this, for some real time applications the test time can become even longer when noise or real time disturbances are considered.

By transferring the processing technique to FPGAs, test time can be reduced significantly and real time application could be implemented. By using large amount of parallel processing, FPGAs can be targeted and reconfigured to test any communication system like OFDM scheme. In present, the application of FPGAs to speed up computation tasks has been researched widely. Some of these researches have focused on the benefits of using FPGAs for communications processing [2-5]. We have implemented an optimized 8-point FFT or IFFT processor for OFDM system in the FPGA platform using Verilog HDL.

A. *OFDM System Overview*

The Orthogonal Frequency Division Multiplexing (OFDM) is developed to support high data rate and can handle multi carrier signals. Its specialty is that, it can minimize the Inter Symbol Interference (ISI) much more compared to other multiplexing schemes [6]. It is more likely improved Frequency Division Multiplexing (FDM) as FDM uses guard band to minimize interference between different frequencies which wastes a lot of bandwidth. But OFDM does not contain inter-carrier guard band which can handle the interference more efficiently than FDM. So, this is the perfect choice for WiMAX as it can help to satisfy the requirements of efficient use of spectrum and minimize the transmission cost [10]. The Discrete Fourier Transform (DFT) is the novel technique to transmit the FDM signal [9]. In OFDM, two digital signal processing algorithm Fast Fourier transform (FFT) and Inverse Fast Fourier Transform (IFFT) are mainly involved. However, the main constraints nowadays for FFT processors used in wireless communication systems for OFDM are execution time and lower power consumption [11]. The main issue in FFT/IFFT processors is complex multiplication, which is the most prominent arithmetic operation used in FFT/IFFT blocks. By using The Cooley–Tukey FFT algorithm, the complexity can be reduced to O ($N.\log_2 N$) [13].

The main objective of this work is to design an OFDM Transmitter and Receiver using FPGA. FFT or IFFT are the main building blocks of OFDM, which operates on finite sequences. The OFDM signal is generated by implementing the Inverse Fast Fourier Transform (IFFT) at the transmitter and Fast Fourier Transform (FFT) at the receiver side. Firstly we derive an algorithm for the IFFT and FFT. A variable may correspond to a wire or a register depending on its application and sometimes an operator can be mapped to hardware like adder, latches, multiplexers etc.

The mathematical model of OFDM symbol defined by Inverse Discrete Fourier Transform IDFT which would be transmitted as given bellow:

$$x(n) = \frac{1}{N}\sum_{k=0}^{N-1} X(K) e^{\frac{j2\pi k}{N}} \quad (1)$$

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

x(n) represents the IDFT time domain output at the *n-th* spectral point where n ranges from *0 to N-1*. The *N* represents the number of sample points in the IDFT data frame. The quantity *X(k)*represents the *k-th* time sample, where *k* also ranges from *0 to N-1*.And the corresponding Discrete Fourier Transform (DFT) operates on sample time domain signal which is periodic. The sequence x(n) gives a sequence X(k) defined only on the interval from *0 to N-1* .The equation for DFT is:

$$X(k)=\sum_{n=0}^{N-1} x(n)e^{-\frac{j2\pi k}{N}} \tag{2}$$

The computation of each point of DFT requires (N-1) complex multiplication and (N-1) complex addition. Thus to compute N points in DFT require N(N-1) complex multiplication and N(N-1) complex addition. The DFT equation (2) can be divided into two equal halves and after that the two equations were formed, one is for even signal terms and another for odd signal terms. DFT can be subdivided until only two points are left in each DFT. Since the outputs were subdivided to obtain this algorithm, it is referred to as decimation-in-frequency (DIF) FFT algorithm.

B. *FPGA Processing*

Modern FPGAs have a tremendous amount of parallel processing power. The latest family of FPGAs released by Xilinx (Vertex Ultra Scale) has up to 4,407,480 logic cells and 4,268 GMACs of DSP processing throughput. This processing power exceeds the latest Multi-Core Multi-Gigahertz processor. In addition to processing power, FPGAs are highly reconfigurable. The user can simply recompile and download a new design and they have a new custom piece of digital processing hardware.

C. *Contributions*

We make the following contributions in this work:

- Demonstration of FPGA as a communication design platform
- Design and Implementation of OFDM transmitter and receiver on the FPGA device
- A comparison of Verilog HDL and MATLAB simulation

D. *Paper Organization*

Section II: Implementation of FFT and IFFT blocks and their equations are optimized in Verilog Hardware Description Language

Section III: Simulation result using Altera Quartus II

Section IV: Comparison Verilog HDL simulation and MATLAB simulation

Section V: Conclusion

II. Implementation of An 8-Point FFT/IFFT Processor

The implementation of IFFT and FFT is simple as compared to DFT and its 8-point DIF output is derived from the input directly. Multiplication of the twiddle factor is skipped to avoid redundancy and reduce computation time. Thus, IFFT and FFT implementation is optimized.

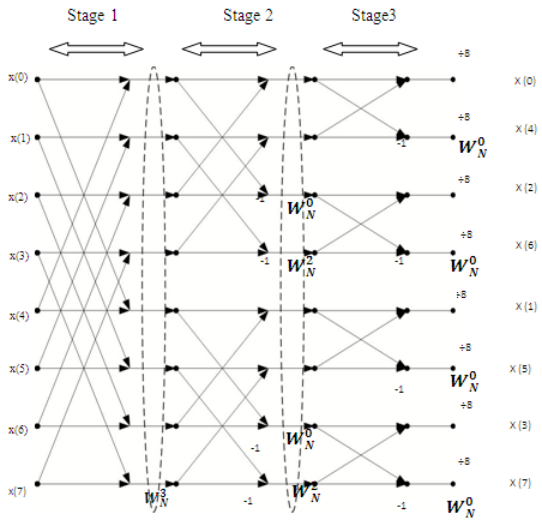


Fig.1 Stage 3 Computation Flow Chart of an 8-point IFFT Computation

The Fig. 1 shows the 8-point IFFT computation. There are three stages. In stage 1 the input data is accepted directly. The outputs of Stage 1 are feed as the inputs of the Stage 2. In stage 2 computations take place and this process repeats at the final stage 3. The complexity of the output equations increases as the Stage number increases because twiddle factor computations are involved. It includes multiplication and additions operations.

Fig. 2 shows an 8-point IFFT block diagram and their interconnections. This module passes the inputs to the sub-modules that process the IFFT computations.

The Pass module consists of 8 D type flip-flop registers. The function of Path 0 and Path 4 is to compute and display the result of these computations.

The arithmetic operation for Xout(0) is summation. The Xout(4) involves summation, subtraction and division. The arithmetic operation for Xout(2) and Xout(6) involves real and imaginary operation. They are performed separately. The twiddle factor for this output is either j or –j which contributes to the imaginary components for this path.

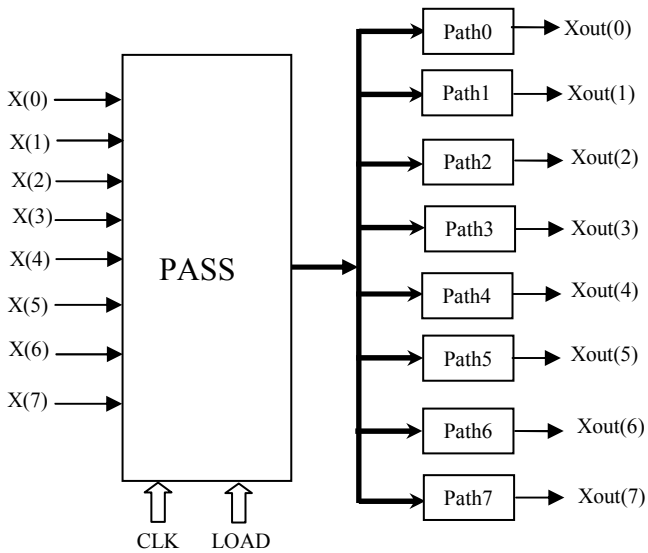


Fig. 2 Block diagram of an 8 point IFFT processor

Same arithmetic operations are involved for Xout(1), Xout(3), Xout(5) and Xout(7) modules including real and imaginary calculation. They are performed separately. The output of the twiddle factor is approximated to 0.70703125 or in binary 0.110110101, including a most significant bit (MSB) to indicate the twiddle factor is a positive number. Any decimal value after summation is approximate to integer '1'.

A. *Expanding Direct Method of an 8-point IFFT*

Using direct method, the equations of every module have been optimized for an efficient implementation. The twiddle factors are represented by 8 bit binary number. There will be slight error percentage in this design due to the approximation of the twiddle factor values. By using Verilog HDL the final equations are implemented to make an 8-point IFFT processor which is shown in below.

TABLE I
THE FINAL EQUATIONS FOR AN 8-POINT IFFT PROCESSOR

X(0) = x(0)+ x(4)+ x(2)+ x(6)+ x(1)+ x(5)+ x(3)+ x(7)
X(4) = x(0)+ x(4)+ x(2)+ x(6)- x(1)- x(5)- x(3)- x(7)
X(2) = x(0)+ x(4)- x(2)- x(6)+ jx(1)+ jx(5)- jx(3)- jx(7)
X(6) = x(0)+ x(4)- x(2)- x(6)- jx(1)- jx(5)+ jx(3)+ jx(7)
X(1) = x(0)- x(4)+ jx(2)- jx(6)+ 0.7071x(1)+ j0.7071x(1) - 0.7071x(5)- j0.7071x(5) - 0.7071x(3)- j0.7071x(3) + 0.7071x(7) + j0.7071x(7)
X(5) = x(0)- x(4)+ jx(2)- jx(6)- 0.7071x(1)- j0.7071x(1) + 0.7071x(5)+ j0.7071x(5) + 0.7071x (3) + j0.7071x (3) - 0.7071x(7) - j0.7071x(7)
X(3) = x(0)- x(4)- jx(2)- jx(6)- 0.7071x(1)+ j0.7071x(1) + 0.7071x(5)- j0.7071x(5) + 0.7071x(3)- j0.7071x(3) - 0.7071x(7)+ j0.7071x(7)
X(7) =x(0)- x(4)- jx(2)- jx(6)+ 0.7071x(1)- j0.7071x(1) - 0.7071x(5)+ j0.7071x(5) - 0.7071x(3)+ j0.7071x(3)+ 0.7071x(7)- j0.7071x(7)

The equations for FFT are similar to IFFT equations except for the negative sign in the twiddle factor and the scaling factor. Thus, the algorithm developed for the IFFT in the previous section can be used for FFT algorithm development with minor modification.

B. *IFFT and FFT Algorithm*

Both the processors IFFT and FFT are divided into subsequent modules. These are pass module, Path0, Path1, Path2, Path3, Path4, Path5, Path6 and Path7 modules. In pass module the input data is passed at the positive edge of the clock to the different modules of FFT processor. In that condition, the load signal is active high. It requires only one clock signal to pass the data.

The implementations of Path0 and Path4 are almost identical except the mathematical operators. There is no imaginary component present at the output. But Path2 and Path6 modules have imaginary component. Thus, they have more complex mathematical expressions.

Path1, Path3, Path5 and Path7 are the most complicated modules among all the modules in the FFT processors because they involve number of mathematical operations, such as addition, subtraction, and multiplication. The outputs contain real and imaginary components. The imaginary components resulted from the twiddle factor which involves sin 45 degree and cosine 45 degree. This value is approximated to 0.70703125 which is equivalent to 0.10110101 in binary form. The IFFT processor has more complex mathematical operations because it has a scaling factor (1/N). In the digital domain, this translates a division operation. The working of all the modules of IFFT is same as in the FFT processor.

III. SIMULATION RESULTS FOR IFFT AND FFT PROCESSOR

The results of IFFT and FFT Processors are obtained from the Quartus-II simulation with random input samples. Each of the input samples contains 8 bits of input. The accuracy of the output is also verified to the output from MATLAB simulation. The result is divided into 2 different sections, for FFT processor and IFFT processor. The output from each of the modules is shown and followed to the overall output.

A. *Resources Used*

The design of the FFT and IFFT processor are compiled and simulated separately using Altera Quartus-II simulator software. The following resources used in the compilation and simulation process are shown in the Table II.

TABLE II
RESOURCE USED IN CYCLONE II FAMILY

Resource	**Usage**
Device Family:	Cyclone II
Device:	EP2C35U484C8
Total Logic elements:	745 (out of 33216) for FFT and 1146 for IFFT
Total combinational function:	641 (out of 33216) for FFT and 967 for IFFT
Dedicated log registers:	349 for FFT and 632 for IFFT
Total registers:	349 and 632 for IFFT
Total Pin:	178 (out of 362) for FFT and 179 for IFFT
Total memory bits:	0
Embedded Multiplier 9 -bit elements	3

Fig. 3 shows the final simulation of the 8-point FFT processor where we see that when we give data through the eight input point it will give eight output data respectively. These simulations also show that Path0 and Path4 give only real output and rest of paths give both real and imaginary value. So, this simulation results is very close to the mathematical calculation and further we will compare these data with MATLAB simulation.

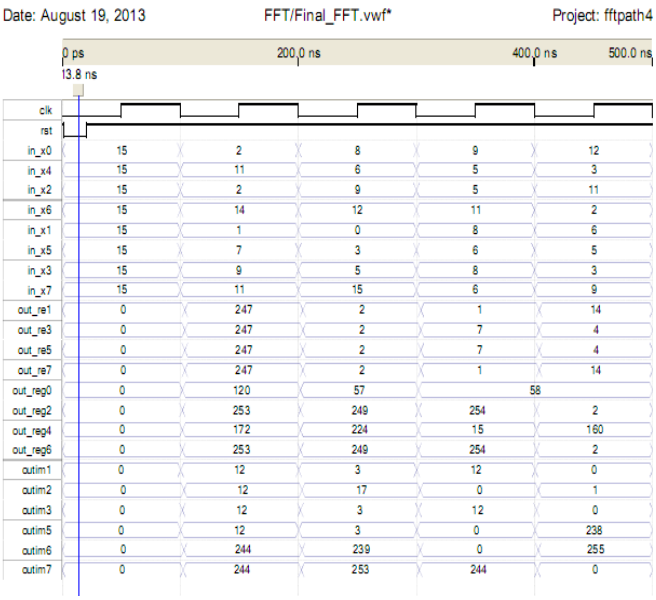


Fig. 3 FFT processor output

IV. Verification Of Verilog HDL Simulation Output With Matlab

In this section, the result of Verilog HDL has been verified. The Verilog HDL output and the MATLAB simulation output using same arbitrary input number are compared to measure the accuracy of the result. In Table III, the accuracy of the Verilog HDL simulation output has been rounded to the nearest integer because only 8 bit is used to represent the output value.

TABLE III
Comparison Between Quartus-II Vs Matlab Output For 8-point FFT

Sl No	Input	Quartus II Out put		MATLAB Out put	
-	---	Real	Imaginary	Real	Imaginary
1	2	60	0	60	0
2	1	247	12	246	19
3	2	253	15	253	15
4	9	247	12	248	252
5	11	224	0	254	0
6	7	247	12	248	4.22
7	14	253	241	253	241
8	14	247	244	246	237

V. Conclusion

In this modern era, OFDM (Orthogonal Frequency Division Multiplexing) is one of the best technologies which support the 3G and 4G standards of wireless communication. OFDM is mainly used in the WiMAX, LTE, WLAN etc. and providing tremendous speed.

The core processing blocks of an OFDM system i.e. the FFT and the IFFT processors have been designed and implemented successfully using common pass module and multiple path modules. Direct mathematical method was adopted as it was found to be an efficient and optimized method implementation which is based on butterfly operation instead of the structural. At the end functional simulation was performed from Altera Quartus II to generate the design net list file and to translate the design into the target Altera FPGA device. The output of the functional simulation matched the data calculated from pre- synthesis MATLAB simulations.

References

[1] Z. Beiling, and G. Xuan, "A novel reconfigurable architecture for generic OFDM modulator based on FPGA," IEEE 16th International Conferenece on Advanced Communication Technology (ICACT), pp. 851-853, Feb. 16-19, 2014.

[2] M. Mefenza, and C. Bobda, "FPGA Implementation of Subcarrier Index Modulation OFDM Transceiver," IEEE 27th International Parallel and Distributed Processing Symposium Workshops & PhD Forum, 2013.

[3] A. Tolmachev, M. Orbach, M. Meltsin, R. Hilgendorf, T. Birk, and M. Nazarathy, "Real-time FPGA implementation of efficient filter-banks for digitally sub-banded coherent DFT-S OFDM receiver," IEEE Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference (OFC/NFOEC), pp. 1-3, 2013.

[4] J. Mar, C. Kuo, and F. Tsai, "Realization of DBF-OFDM Transceiver for Vehicular Communication Using FPGA Chip," IEEE 75th Vehicular Technology Conference (VTC Spring), pp. 1-5, May. 6-9, 2012.

[5] A. Hechri, M. Abdellatif, "FPGA implementation of an OFDM baseband transmitter," IEEE International Conference on Communications, Computing and Control Applications (CCCA), pp. 1-4, Mar. 3-5, 2011.

[6] H. Ali, X. Zhou, and K. Iqbal, "FPGA architecture for OFDM Software Defined Radio with an optimized Direct Digital Frequency Synthesizer," IEEE Saudi International Electronics, Communications and Photonics Conference (SIECPC), 2013.

[7] L. Ali, R. Sidek, I. Aris, M. A. M. Ali, and B. S. Suparjo, "Challenges and Directions for IC testing," Integration, the VLSI Journal, Elsevier Science, vol. 37(1), pp. 17-28, Netherland, 2004.

[8] C. Dick, and F. Harris, "FPGA Implementation of an OFDM PHY," Thirty-seventh Asilomar Conference on Signals and Systems, Nov. 2003.

[9] L. Dai, Z. Wang, and Z. Yang, "Time-Frequency Training OFDM with High Spectral Efficiency and Reliable Performance in High Speed Environments," IEEE Journal in Communications, vol. 30, no.4, 2012.

[10] F. Ahmed, M. A. Zaman, and S. Das, "Performance Evaluation of different level Modulation schemes for an efficient Mobile WiMAX system in OFDM under Rayleigh Fading and AWGN channel," International Journal of Mobile and Adhoc Networking (IJMAN), (IFRSA) ISSN (Online) 2231:6825 ISSN (Print) 2249–202X. vol. 3, Issue: 4, Nov. 2012.

[11] Weidong. Li, "Studies on implementation of lower power FFT processors" Linköping Studies in Science and Technology, Thesis No. 1030, ISBN 91-7373-692-9, Linköping, Sweden, Jun. 2003.

[12] S. B. Weinstein, and P. M. Ebert, "Data transmission by frequency division multiplexing using the discrete Fourier transform," IEEE Transactions on Communication Technology, pp. 628-634, Oct. 1971.

[13] J.W. Cooley, and J. W. Tukey, "An algorithm for the machine calculation of complex Fourier series," Math. Computation, pp. 297-301, 1965.

[14] N. T. Hieu, N. T. Thanh, V. H. Nguyen, T. T. Nguyen, and H. P. Bui, "FPGA implementation of MIMO-OFDM STBC systems," IEEE International Conference on Control, Automation and Information Sciences (ICCAIS), 2012.

Maximum PEP and ICI Over Coset Representatives for 32 Subcarriers Reed-Muller Coded OFDM

M Maksud Alam,[1,*] Nusrat J. Disha,[2] Md. Ataur Rahman,[3] and Besma Smida[4]
[1,2,3]Electrical & Computer Engineering Department, North South University, Bangladesh
[4]Electrical & Computer Engineering Department, Purdue University, IN, USA
*maksud@northsouth.edu

***Abstract*—Orthogonal Frequency Division Multiplexing (OFDM) is an excellent choice of high speed data transmission systems. Since a large number of subcarriers uses in most of the broadband data communications standards, high PAPR is also a great drawbacks. On the other hand, subcarriers orthogonality is very sensitive to carrier frequency offset (CFO) in between transmitter and receiver oscillators. Golay complementary sequences (Coset Representatives) with Reed-Muller code gives maximum $3\ dB$ PMEPR [20] and reduces ICI, due to the CFO, significantly [1]. In this work, our contributions are two-fold: 1) we obtain all, total 1024, possible coset representatives for 32 subcarriesrs BPSK modulated Coded-OFDM (COFDM), 2) we simulated maximum Peak Envelope Power (PEP) and Inter-carrier Interference (ICI) over each coset representatives.**

***Index Terms*—PEP, ICI, Coset Representatives, Coded-OFDM.**

I. Introduction

A. Motivation

Orthogonal frequency division multiplexing (OFDM) systems are characterized by large number of closely spaced orthogonal subcarriers that carry data from a single source. OFDM systems, hence, have high spectral efficiency and excellent ability to cope with various channel condition e.g. narrow-band interference and frequency selective fading cause of multipath environment [3]. That's the reason for considering the OFDM as air-interface by many high-data-rate wireless transmission systems. OFDM systems can however experienced two major drawbacks: high peak-to-mean envelope power ratio (PMEPR) due to summing up of large number of subcarriers data together and inter-carrier interference (ICI) because the orthogonal properties are easily broken down by carrier frequency offset (CFO) errors caused by inaccuracies between transmitter and receiver oscillators, Doppler effects, etc.

B. Basic COFDM Architecture

A basic COFDM systems architecture is represented in Fig.1. As the blocks show that information data, output of a source-encoder, are passed through a channel-encoder which adds redundancy bits for error control due to severe channel conditions. In the modulation phase, the bits are mapped into a symbol, and they are phase shift keying (PSK) modulated so as to maintain simplicity and constant amplitude patterns. The data is then converted into parallel from serial followed by inverse discrete fourier transform (IDFT) process. IDFT transmutes the domain of data to frequency from time and coherently sums up to get OFDM symbol. Following some required manipulation, the OFDM symbol is transmitted through the AWGN channel. At the receiver end inverse operations of the transmitter end takes place.

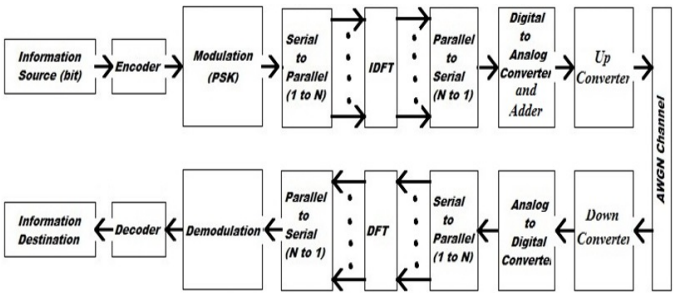


Fig. 1. A basic architecture of Coded-OFDM systems.

C. Related Works

It is clear from the OFDM architecture given in the Fig.1, to create an OFDM symbol, all subcarriers signals are added together after IDFT has been done. The process results high fluctuations in signals power levels and the detrimental aspects well known as peak-to-mean envelope power ratio (PMEPR). Also recognizing as peak to average power ratio (PAPR), some techniques have been proposed in literature to reduce the effect. The PAPR reduction methods are categorized as follows: Selected mapping (SLM) and partial transmit sequences (PTS) [14], [15], [16], [17], Clipping and filtering techniques [13] and coding [20]. SLM and PTS approaches reduce the probability of generating a signal with large PMEPR, by generating multiple statistically independent OFDM symbols for a given data frame and transmitting the symbol with the lowest peak power.

In Clipping and filtering techniques, a pseudo-maximum amplitude has been determined to cut-off the peaks of the OFDM time domain signals followed by the filtering process for minimizing the out of band radiation, generated by clipping. Last approach is to use coding to reduce PMEPR. The theoretical aspects of the relation between the code rate, minimum Euclidean distance of the code and its block length is provided in [18] as two fundamental theorems. The first one proves a lower bound for PAPR based on the three aforementioned parameters.

To get the advantage of spectrum efficiency, the orthogonality between sub-carriers must be maintained. In the literature, two types of approaches have been proposed to mitigate the frequency-offset problem. The first approach focuses on estimating and then eradicating the frequency offset [4]. The second approach focuses on reducing the sensitivity to CFO with no need to precisely estimate the CFO.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

The first approach is beyond the scope of this paper. We assume here that a coarse-offset estimation that estimates the part of frequency offset that is multiple of the subcarrier spacing is carried on successfully.

The second approach employs a combination of signal processing and coding to reduce the sensitivity to the residual frequency offset [5]. An effective method, named ICI self-cancelation, proposed by Zhao and Haggman in [6] and later extended in [7], [8], [9], uses polynomial coding in the frequency domain. The self-cancellation method modulates copies of the same data symbol on r adjacent sub-carriers using optimized weights. The method reduces the ICI significantly but it lowers the transmission rate by a factor r. Some forward-error correction codes are also proposed, to remove the errors caused by the ICI, in [10]. The correlative-coding, proposed in [11], is another way of reducing the sensitivity of the ICI. More recently, Smida proposed in [12] to choose a well understood code and to rotate each coordinate of the code by a fixed phase shift to reduce ICI.

All of the above methods are useful for reducing PAPR and ICI separatetly. Very few research work has been done for remedy both the problems of OFDM systems. The most few important exceptions are Golay complementary repetition code [2] and Selective Coset (SLC)[1] which alleviate both PAPR and ICI.

D. Our Contributions

In this paper, we simulate maximum PEP and ICI over all possible coset representatives for 32 subcarriers, where 1024 Golay complementary sequences are possible. The results are important because both Golay complementary repetition codes and Selective Coset (SLC) methods use the Golay complementary sequences, well known as coset representatives.

II. Mathematical Terms and Definitions

A. Preliminaries

We consider a OFDM systems with N subcarriers where modulated time domain signals of all subchannels are added together to form an OFDM signals within a OFDM time period of T_s after IDFT operation. The transmitted signal at time t may be represented as the real part of the complex envelope,

$$S_{\mathbf{c}}(t) = \sum_{n=0}^{N-1} c_n \exp(j2\pi(f_0 + nf_s)t) \quad \text{for} \quad 0 \le t \le T_s, \quad (1)$$

where $j = \sqrt{-1}$, f_0 is the beginning frequency and f_s is the bandwidth of each subcarrier. The OFDM-symbol duration T_s is equal to $\frac{1}{f_s}$ to ensure the orthogonality among the subcarriers. The vector $\mathbf{c} = [c_0, c_1, \ldots, c_{N-1}]$ of length N is the modulating data sequence. We refer to $\mathbf{c}$ as codeword of code $\mathcal{C}$ that maps blocks of k input bits into blocks of N symbols. The data-coefficients c_n are typically taken from a constellation $\mathcal{Q}$ with 2^h elements, where h is the number of bits being mapped on each symbol.

The *peak envelop power*(PEP) of the transmitted signal is

$$\text{PEP}(\mathbf{c}) = \max_{0 \le t \le T_s} |S_{\mathbf{c}}(t)|^2, \quad (2)$$

So according to the definition, the *peak-to-mean envelope power ratio* (PMEPR) is as follows,

$$\text{PMEPR}(\mathbf{c}) = \max_{0 \le t \le T_s} \frac{|S_{\mathbf{c}}(t)|^2}{\| \mathbf{c} \|^2}. \quad (3)$$

Where, $\| \mathbf{c} \|^2$ is known as *mean* or *average* power of the OFDM signal over T_s. In case of *phase shift keying*, the magnitude of $\| \mathbf{c} \|^2$ is equal to N.

We use an additive white Gaussian noise (AWGN) environment to represent the channel and $S_{\mathbf{c}}(t)$ is transmitted through it. The receiver performs inverse of IDFT after receiving the signal $\Re(S_{\mathbf{c}}(t))$ which is the perturbed version, due to noise, of transmitted signal and estimates of $\mathbf{c}$. Then at the receiver, a useful error-correction algorithm is used to extracts the block of input codes. The received signal sample for the kth subcarrier after DFT can be written as

$$y_k = c_k b_0 + \sum_{l=0, l \ne k}^{N-1} b_{l-k} c_l + n_k = c_k b_0 + I_k + n_k, \quad (4)$$

for $k = 0, 1, \ldots, N-1$, where n_k is a complex Gaussian noise sample. The first term in the right-hand side of (4) represents the desired signal. The ICI term, attributable to the CFO, on the kth subcarrier is expressed by I_k. The ICI coefficients b_k is given by

$$b_k = \frac{\sin \pi(k+\varepsilon)}{N \sin \frac{\pi}{N}(k+\varepsilon)} \exp\left[j\pi(1 - \frac{1}{N})(k+\varepsilon) \right], \quad (5)$$

where ε is the normalized frequency offset defined as a ratio between the frequency offset (which remains constant over each symbol period) and the subcarrier spacing. Note that I_k is a function of both $\mathbf{c}$ and ε.

Sathanannthan and Tellambura [2] defined the *peak interference-to-carrier ratio* PICR as

$$\text{PICR}(\mathbf{c}, \varepsilon) = \max_{0 \le k \le N-1} \frac{|I_k|^2}{|b_0 c_k|^2}. \quad (6)$$

PICR is *maximum interference-to-signal ratio* for any kth subcarrier, where $|b_0 c_k|^2$ is the signal power of kth subcarrier. It specifies the worst-case ICI on any subcarrier. Large I_k values cause high bit errors in subcarriers. To make the PICR compatible with PMEPR, we called it the *peak interference-to-carrier power ratio (PICPR)*.

To measure the detrimental effect of ICI, we have introduced the *peak interference power* (PIP) as

$$\text{PIP}(\mathbf{c}, \varepsilon) = \max_{0 \le k \le N-1} |I_k|^2, \quad (7)$$

which is fuction of both $\mathbf{c}$ and ε, too. To reduce ICI effects, PIP should be minimized.

To improve the performance of OFDM systems both PMEPR in (3) and PICPR in (6) should be minimized.

B. Golay Complementary Sequences and Reed-Muller Codes

In [20], authors present a coding technique that connects Golay sequeces and first order binary Reed-Muller codes. The Golay sequeces transpire as cosets of first order Reed-Muller codes in the second order Reed-Muller codes. The following theorems justify the preference of the Golay complementary sequences.

Theorem 1: For unit power constellations, the PMEPR of any complementary Golay sequence is bounded by 2.

Theorem 2: For any permutation ϕ of the set $\{1, 2, \ldots, m\}$, and for any $c_k, c \in \mathbb{Z}_{2^h}$. The sequence generated by

$$u(\mathbf{x}) = 2^{h-1} \sum_{k=1}^{m-1} x_{\phi(k)} x_{\phi(k+1)} + \sum_{k=1}^{m} c_k x_{\phi(k)} + c,$$

is Golay complementary over $\mathbb{Z}_{2^h}$ of length 2^m.

For H-*phase shift keying*, where $H = 2^h$ and $h \geq 1$, each OFDM symbol period contains exactly h code bit. In the case of $H = 2, 4,$or 8 the code is called binary,quaternary and octary respectively.

Popovic proved Theorem 1 in [19] which inspired the use of Golay sequences for OFDM systems to mitigate PAPR problem. Theorem 2 explicitly describes connection between the Golay complementary sequences and RM codes [20]. The RM code over $\mathbb{Z}_{2^h}$ of order r and length 2^m, denoted by $\mathrm{RM}_{2^h}(r, m)$, consists of all linear combinations of vectors associated with the monomials of degree at most r in m variables. The Boolean monomial of m variables is defined as $f_x = x_0^{r_0} x_1^{r_1} ... x_{m-1}^{r_{m-1}}$ where $(x_0, x_1, \ldots, x_{m-1})$ is an m-tuple binary vector. Henceforth, each of $\frac{m!}{2}$ cosets of $\mathrm{RM}_{2^h}(1, m)$ in $\mathrm{RM}_{2^h}(2, m)$ having a coset representative of the form $\sum_{k=1}^{m-1} x_{\phi(k)} x_{\phi(k+1)}$ comprises $2^{h(m+1)}$ Golay sequences over $\mathbb{Z}_{2^h}$ of length 2^m. Thus, we conclude that costs of Reed-Muller codes over $\mathbb{Z}_{2^h}$, provide codewords with PMEPR that does not exceed 3 dB [20].

For example, if we consider $r = 1$ and $m = 5$ then $\mathrm{RM}_{2^h}(1, 5)$ has the generator matrix, G, given below as

$$\begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 11111111111111111111111111111111 \\ 00000000000000001111111111111111 \\ 00000000111111110000000011111111 \\ 00001111000011110000111100001111 \\ 00110011001100110011001100110011 \\ 01010101010101010101010101010101 \end{bmatrix}$$

Therefore, the Reed Muller codeword, $\mathbf{c} = [c_0, c_1, \ldots, c_{31}]$, of length 2^5 is the linear combination of the generator matrix G and $\mathbf{u} = [u_0, u_1, u_2, u_3, u_4, u_5]$ as $\mathbf{c} = \mathbf{u} .* G$, where $\mathbf{u}$ is information bits. The sign, $.*$, represents matrix multiplication. Hence the Reed Muller Golay Codeword, $\mathbf{c}' = \mathbf{c} + u(\mathbf{x})$.

III. Coset Representatives for 32 subcarriers

It has been shown in [2] that both PAPR and ICI can be reduced by using Golay complementary repetition code. It is a simple combination of Golay complementary sequences (to reduce PAPR) and repetition code (to reduce PICR). Based on the previous theorems, a more efficient code to reduce PAPR and ICI has been proposed in [1]. The main idea is to choose a subset of coset representatives. Therefore coset representatives are most important factor for both techniques.

There are $2^{m(m-1)/2} = 2^{5(5-1)/2} = 1024$ cosets for 32 subcarriers. The coset values are calculated from the matrix multiplication of a certain combination. From table I the term $x_1 x_5$ is found by array multiplication. The values of x_1 and x_5 are obtained from the generator matrix above. The binary rows of x_1 and x_5 are multiplied in order to get a coset value of a particular sequence. Therefore, $x_1 x_5 = x_1 . * x_5 =$ 00000000000000000101010101010101 is a Golay sequence. In this way all the coset values of 1024 sequences are found and in the table I all coset representatives those are having maximum PMEPR at most 2 or 3 dB have been shown.

IV. Simulation Results

The simulation results, shown in table II, obtained considering BPSK modulation, AWGN channel environment and frequency offset $\epsilon = 0.1$. The number of subcarriers for an OFDM systems has been chosen as $N = 32$, for simulatating PEP, PMEPR, PIP as well as PICPR of Golay Complementary sequences. The coset representatives those give maximum PMEPR 3 dB has been chosen to be represented in the tables.

In second table the coset representatives having maximum PMEPR 3 have been rearranged according to their PICPR values. They can be categorized in 3 types. First the values which lies within the range from -8.73 to -8.05 dB give the cosets that are best for transmission, then the range -7.97

TABLE I
Coset Representatives having PMEPR of at most 3

Coset Equations	Golay Sequences
$x_1x_2 + x_1x_3 + x_2x_4 + x_3x_5$	00000101001101100000101011000110
$x_1x_2 + x_2x_4 + x_3x_4 + x_3x_5$	00000110001101010000011011001010
$x_1x_2 + x_1x_4 + x_2x_3 + x_3x_5$	00000101000010100011011011000110
$x_1x_4 + x_2x_3 + x_2x_4 + x_3x_5$	00000101001110010011011000001010
$x_1x_3 + x_1x_4 + x_2x_4 + x_3x_5$	00000101001101100011100100001010
$x_1x_4 + x_1x_5 + x_2x_4 + x_3x_5$	00000101001101100110001101010000
$x_1x_4 + x_2x_4 + x_2x_5 + x_3x_5$	00000101011000110011011001010000
$x_1x_4 + x_2x_3 + x_2x_5 + x_3x_4$	00000011010110010011000001101010
$x_1x_2 + x_1x_4 + x_2x_5 + x_3x_4$	00000011010101100011000010011010
$x_1x_3 + x_1x_4 + x_2x_5 + x_3x_5$	00000101010100000011100101101100
$x_1x_4 + x_2x_3 + x_3x_5 + x_4x_5$	00010100000110110010011100101000
$x_1x_2 + x_2x_5 + x_3x_4 + x_3x_5$	00000110010100110000011010101100
$x_1x_3 + x_2x_4 + x_2x_5 + x_3x_5$	00000101011000110000101001101100
$x_1x_3 + x_2x_4 + x_2x_5 + x_3x_4$	00000011011001010000110001101010
$x_1x_2 + x_1x_4 + x_2x_5 + x_3x_5$	00000101010100000011011010011100
$x_1x_2 + x_2x_3 + x_3x_5 + x_4x_5$	00010100000110110001010011100100
$x_1x_2 + x_1x_3 + x_2x_5 + x_3x_4$	00000011010101100000110010100110
$x_1x_3 + x_2x_4 + x_3x_5 + x_4x_5$	00010100001001110001101100101000
$x_1x_2 + x_2x_4 + x_3x_5 + x_4x_5$	00010100001001110001010011011000
$x_1x_3 + x_1x_4 + x_2x_3 + x_2x_5$	00000000010110100011110001100110
$x_1x_3 + x_1x_4 + x_2x_4 + x_2x_5$	00000000011001100011110001011010
$x_1x_2 + x_1x_3 + x_3x_5 + x_4x_5$	00010100000101000001101111100100
$x_1x_4 + x_2x_5 + x_3x_4 + x_3x_5$	00000110010100110011010101100000
$x_1x_5 + x_2x_3 + x_2x_4 + x_3x_5$	00000101001110010101000001101100
$x_1x_2 + x_1x_3 + x_3x_4 + x_4x_5$	00010010000100100001110111100010
$x_1x_2 + x_2x_3 + x_3x_4 + x_4x_5$	00010010000111010001011100010
$x_1x_4 + x_1x_5 + x_2x_3 + x_3x_5$	00000101000010100110001101101100
$x_1x_3 + x_2x_3 + x_2x_4 + x_4x_5$	00010001001011010001111000100010
$x_1x_3 + x_1x_4 + x_2x_3 + x_4x_5$	00010001000111100010110100100010
$x_1x_2 + x_1x_4 + x_3x_4 + x_3x_5$	00000110000001100011010111001010
$x_1x_5 + x_2x_4 + x_3x_4 + x_3x_5$	00000110001101010101001101100000
$x_1x_2 + x_1x_5 + x_2x_4 + x_3x_5$	00000101001101100101000010011100
$x_1x_5 + x_2x_3 + x_3x_4 + x_4x_5$	00010010000111010100011101001000
$x_1x_2 + x_1x_3 + x_2x_4 + x_4x_5$	00010001001000100001111011010010
$x_1x_2 + x_1x_5 + x_3x_4 + x_3x_5$	00000110000001100101001110101100
$x_1x_3 + x_1x_5 + x_2x_3 + x_4x_5$	00010001000111100101001110100100
$x_1x_3 + x_2x_5 + x_3x_4 + x_4x_5$	00010010010001110001110101001000
$x_1x_2 + x_1x_4 + x_3x_5 + x_4x_5$	00010100000101000010011111011000
$x_1x_3 + x_2x_3 + x_2x_5 + x_4x_5$	00010001010010110001111001000100
$x_1x_2 + x_2x_5 + x_3x_4 + x_4x_5$	00010010010001110001001010111000
$x_1x2 + x_1x_3 + x_2x_5 + x_4x_5$	00010001010001000001111010110100
$x_1x_2 + x_1x_4 + x_2x_3 + x_4x_5$	00010001000111100010001011010010
$x_1x_5 + x_2x_4 + x_2x_5 + x_3x_4$	00000011011001010101011000110000
$x_1x_4 + x_1x_5 + x_2x_5 + x_3x_4$	00000011010101100011001010100110000
$x_1x_2 + x_1x_5 + x_2x_4 + x_3x_4$	00000011001100000101011010011010
$x_1x_3 + x_1x_5 + x_2x_4 + x_3x_4$	00000011001100000101100101101010
$x_1x_2 + x_1x_5 + x_3x_4 + x_4x_5$	00010010000100100100011110111000
$x_1x_4 + x_2x_3 + x_2x_5 + x_4x_5$	00010001010010110010001001111000
$x_1x_2 + x_1x_5 + x_2x_3 + x_4x_5$	00010001000111100100010010110100
$x_1x_3 + x_1x_4 + x_2x_5 + x_4x_5$	00010001010001000010110101111000
$x_1x_5 + x_2x_3 + x_2x_4 + x_4x_5$	00010001001011010100010001111000
$x_1x_4 + x_1x_5 + x_2x_3 + x_2x_4$	00000000001111000110011001011010
$x_1x_4 + x_1x_5 + x_2x_3 + x_2x_5$	00000000010110100110011000111100
$x_1x_3 + x_1x_5 + x_2x_3 + x_2x_4$	00000000001111000101101001100110
$x_1x_3 + x_1x_5 + x_2x_4 + x_2x_5$	00000000011001100101101000111100
$x_1x_3 + x_1x_5 + x_2x_4 + x_2x_5$	00000000011001100101101000111100
$x_1x_3 + x_1x_5 + x_2x_5 + x_3x_4$	00000011010101100101100100001100
$x_1x_5 + x_2x_3 + x_2x_5 + x_3x_4$	00000011010110010101011000001100
$x_1x_2 + x_1x_5 + x_2x_3 + x_3x_4$	00000011000011000101011010100110
$x_1x_4 + x_1x_5 + x_2x_3 + x_3x_4$	00000011000011000110010101101010

TABLE II
MAXIMUM PIP, PICPR OF COSET REPRESENTATIVES HAVING $3dB$ PMEPR FOR $\epsilon = 0.1$

Coset Equation	Max PEP	Max PMEPR (dB)	Max PIP	Max PICPR (dB)
$x_1x_2 + x_1x_3 + x_2x_4 + x_3x_5$	64	3.01	132.64	-8.73
$x_1x_2 + x_2x_4 + x_3x_4 + x_3x_5$	64	3.01	132.64	-8.73
$x_1x_2 + x_1x_4 + x_2x_3 + x_3x_5$	64	3.01	136.27	-8.62
$x_1x_4 + x_2x_3 + x_2x_4 + x_3x_5$	64	3.01	138.06	-8.56
$x_1x_3 + x_1x_4 + x_2x_4 + x_3x_5$	64	3.01	143.93	-8.38
$x_1x_4 + x_1x_5 + x_2x_4 + x_3x_5$	64	3.01	152.39	-8.13
$x_1x_4 + x_2x_4 + x_2x_5 + x_3x_5$	64	3.01	153.33	-8.10
$x_1x_4 + x_2x_3 + x_2x_5 + x_3x_4$	64	3.01	154.71	-8.06
$x_1x_2 + x_1x_4 + x_2x_5 + x_3x_4$	64	3.01	155.32	-8.05
$x_1x_3 + x_1x_4 + x_2x_5 + x_3x_5$	64	3.01	158.04	-7.97
$x_1x_4 + x_2x_3 + x_3x_5 + x_4x_5$	64	3.01	158.21	-7.97
$x_1x_2 + x_2x_5 + x_3x_4 + x_3x_5$	64	3.01	158.24	-7.97
$x_1x_3 + x_2x_4 + x_2x_5 + x_3x_5$	64	3.01	158.64	-7.96
$x_1x_3 + x_2x_4 + x_2x_5 + x_3x_4$	64	3.01	158.64	-7.96
$x_1x_2 + x_1x_4 + x_2x_5 + x_3x_5$	64	3.01	159.11	-7.94
$x_1x_2 + x_2x_3 + x_3x_5 + x_4x_5$	64	3.01	159.16	-7.94
$x_1x_2 + x_1x_3 + x_2x_5 + x_3x_4$	64	3.01	160.84	-7.90
$x_1x_3 + x_2x_4 + x_3x_5 + x_4x_5$	64	3.01	160.95	-7.89
$x_1x_2 + x_2x_4 + x_3x_5 + x_4x_5$	64	3.01	163.90	-7.81
$x_1x_3 + x_1x_4 + x_2x_3 + x_2x_5$	64	3.01	164.31	-7.80
$x_1x_3 + x_1x_4 + x_2x_4 + x_2x_5$	64	3.01	164.31	-7.80
$x_1x_2 + x_1x_3 + x_3x_5 + x_4x_5$	64	3.01	164.37	-7.80
$x_1x_4 + x_2x_5 + x_3x_4 + x_3x_5$	64	3.01	165.77	-7.76
$x_1x_5 + x_2x_3 + x_2x_4 + x_3x_5$	64	3.01	166.76	-7.74
$x_1x_2 + x_1x_3 + x_3x_4 + x_4x_5$	64	3.01	167.51	-7.72
$x_1x_2 + x_2x_3 + x_3x_4 + x_4x_5$	64	3.01	167.51	-7.72
$x_1x_4 + x_1x_5 + x_2x_3 + x_3x_5$	64	3.01	167.66	-7.72
$x_1x_3 + x_2x_3 + x_2x_4 + x_4x_5$	64	3.01	167.75	-7.71
$x_1x_3 + x_1x_4 + x_2x_3 + x_4x_5$	64	3.01	167.75	-7.71
$x_1x_2 + x_1x_4 + x_3x_4 + x_3x_5$	64	3.01	168.28	-7.70
$x_1x_5 + x_2x_4 + x_3x_4 + x_3x_5$	64	3.01	169.06	-7.68
$x_1x_2 + x_1x_5 + x_2x_4 + x_3x_5$	64	3.01	170.69	-7.64
$x_1x_5 + x_2x_3 + x_3x_4 + x_4x_5$	64	3.01	172.02	-7.60
$x_1x_2 + x_1x_3 + x_2x_4 + x_4x_5$	64	3.01	172.03	-7.60
$x_1x_2 + x_1x_5 + x_3x_4 + x_3x_5$	64	3.01	172.83	-7.58
$x_1x_3 + x_1x_5 + x_2x_3 + x_4x_5$	64	3.01	175.16	-7.53
$x_1x_3 + x_2x_5 + x_3x_4 + x_4x_5$	64	3.01	175.62	-7.51
$x_1x_2 + x_1x_4 + x_3x_5 + x_4x_5$	64	3.01	177.34	-7.47
$x_1x_3 + x_2x_3 + x_2x_5 + x_4x_5$	64	3.01	177.65	-7.46
$x_1x_2 + x_2x_5 + x_3x_4 + x_4x_5$	64	3.01	178.70	-7.44
$x_1x2 + x_1x_3 + x_2x_5 + x_4x_5$	64	3.01	182.66	-7.34
$x_1x_2 + x_1x_4 + x_2x_3 + x_4x_5$	64	3.01	183.86	-7.31
$x_1x_5 + x_2x_4 + x_2x_5 + x_3x_4$	64	3.01	189.23	-7.19
$x_1x_4 + x_1x_5 + x_2x_5 + x_3x_4$	64	3.01	189.74	-7.18
$x_1x_2 + x_1x_5 + x_2x_4 + x_3x_4$	64	3.01	193.01	-7.10
$x_1x_3 + x_1x_5 + x_2x_4 + x_3x_4$	64	3.01	193.45	-7.09
$x_1x_2 + x_1x_5 + x_3x_4 + x_4x_5$	64	3.01	193.91	-7.08
$x_1x_4 + x_2x_3 + x_2x_5 + x_4x_5$	64	3.01	194.00	-7.08
$x_1x_2 + x_1x_5 + x_2x_3 + x_4x_5$	64	3.01	195.40	-7.05
$x_1x_3 + x_1x_4 + x_2x_5 + x_4x_5$	64	3.01	195.45	-7.05
$x_1x_5 + x_2x_3 + x_2x_4 + x_4x_5$	64	3.01	199.54	-6.96
$x_1x_4 + x_1x_5 + x_2x_3 + x_2x_4$	64	3.01	200.70	-6.93
$x_1x_4 + x_1x_5 + x_2x_3 + x_2x_5$	64	3.01	200.70	-6.93
$x_1x_3 + x_1x_5 + x_2x_3 + x_2x_4$	64	3.01	201.43	-6.92
$x_1x_3 + x_1x_5 + x_2x_4 + x_2x_5$	64	3.01	201.43	-6.92
$x_1x_3 + x_1x_5 + x_2x_4 + x_2x_5$	64	3.01	201.43	-6.92
$x_1x_3 + x_1x_5 + x_2x_5 + x_3x_4$	64	3.01	207.76	-6.78
$x_1x_5 + x_2x_3 + x_2x_5 + x_3x_4$	64	3.01	208.05	-6.78
$x_1x_2 + x_1x_5 + x_2x_3 + x_3x_4$	64	3.01	212.07	-6.69
$x_1x_4 + x_1x_5 + x_2x_3 + x_3x_4$	64	3.01	213.18	-6.67

to $-7.05\ dB$ gives the cosets which are second best for our purpose. And the cosets with PICPR values within -6.96 to $-6.67\ dB$ are having high sensitivity with CFO.

V. CONCLUSION

In this paper, we repesents the cosets those are better for error correction through the simulation proces. We consider only PMEPR and PICPR to categorize the cosets. So there are 9 cosets out of 1024 those are having maximum 3 dB PMEPR and maximum $-8.05\ dB$.

REFERENCES

[1] Maksud Alam M, Smida Besma, PAPR and ICI Reduction of OFDM Signals, *in Proc. of Engineering & Technology* 2013; 3: 209-211.

[2] Sathananthan K, and Tellambura C., *Coding to Reduce Both PAR and PICR of an OFDM Signal*, *IEEE Communications Letters* 2002;6(8): 316-318.

[3] Cimini Jr. L. J. Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing. *IEEE Trans. Commun.* 1985; 33: 665-675.

[4] J. J. van de Beek J. J. Sandell M., and Borjesson P. O. ML estimation of time and frequency offset in OFDM systems. *IEEE Trans. Signal Process.* 1997; 45(7): 1800-1805.

[5] Armstrong J. Analysis of new and existing methods of reducing intercarrier interference due to carrier frequency offset in OFDM. *IEEE Trans. Commun.* 1999; 47(3): 365-369.

[6] Zhao Y. and Haggman S. G. Sensitivity to Doppler shift and carrier frequency errors in OFDM systems-the consequences and solutions. *in Proc. 46th IEEE Vehicular Technology Conf.* 1996; 1564-1568.

[7] Sathananthan K., Rajatheva R. M. A. P, Slimane S. B. Cancellation technique to reduce intercarrier interference in OFDM. *Electronics Letters* 2000; 36(25).

[8] Zhao Y. and Haggman S. G. Intercarrier interference self-cancellation scheme for OFDM mobile communication systems. *IEEE Trans. Commun.* 2001; 49: 1185-1191.

[9] Seyedi A., Saulnier G. J. General ICI self-cancellation scheme for OFDM systems. *IEEE Trans. on Vehi. Techno.* 2005; 54: 198 - 210.

[10] Sathananthan K. and Tellambura C. Forward error correction codes to reduce intercarrier interference in OFDM. *in Proc. IEEE Int. Symp. Circuits and Systems* 2001; 4: 566-569.

[11] Zhao Y. and Haggman S. G. Intercarrier interference compression in OFDM communication systems by using correlative coding. *IEEE Commun. Lett.* 1998; 2(8): 214-216.

[12] Smida B. Computation and reduction of the Peak Interference to Carrier Ratio of OFDM signals. Proc of *Globecom* 2011, Houston TX, USA.

[13] Li X. and Cimini L. J. Effects of clipping and filtering on the performance of OFDM. *IEEE Commun. Lett.* 1998; 2(5): 131-133.

[14] Muller S. H. and Huber J. B. OFDM with reduced peak-to-average power ratio by optimum combination of partial transmit sequences. *Electron. Lett.* 1997; 33: 368-369.

[15] Muller S. H., Bauml R. W., Fisher R.F.H, and Huber J.B. OFDM with reduced peak-to-average power ratio by multiple signal representation. *Ann. Telecommun.* 19917; 52: 58-67.

[16] Bauml R. W., Fischer R. F. H., and Huber M. J.B. Reducing the Peak to average power ratio of multicarrier modulation by selected mapping. *IEE Elect. Lett.* 1996; 32: 2056-2057.

[17] Cimini L. J. and Sollenberger N. R. Peak-to-average power ratio reduction of an OFDM signal using partial transmit sequences. *IEEE Commun. Lett.* 1999; 4: 511-515.

[18] Paterson K. G. and Tarokh V. On the existence and construction of good codes with low peak-to-average power ratios. *IEEE Trans. Information Theory,* 2000; 46(6): 1974-1987.

[19] Popovic B. M. Synthesis of power efficient multitone signals with flat amplitude spectrum. *IEEE Trans. Commun.* 1991; 39: 1031-1033.

[20] Davis J. A. and Jedwab J. Peak-to-mean power control in OFDM, Golay Complementary sequences and Reed-Muller codes. *IEEE Trans. Inform. Theory* 1999; 45: 2397-2417.

[21] Jayalath A. D. S. and Tellambura C. SLM and PTS Peak-Power Reduction of OFDM Signals Without Side Information. *IEEE Transactions On Wireless Communications* 2005; 4(5): 2006-2013.

Post Mitigation Performance Comparison of RZ-DPSK and CSRZ-DPSK DWDM Systems

Mohammad Faisal*, Tahsin Faruque, and M. Arafat Rahman Khan

Department of Electrical & Electronic Engineering
Bangladesh University of Engineering and Technology
Dhaka 1000, Bangladesh.
*mdfaisal@eee.buet.ac.bd

Abstract—**Performance degradation due to nonlinear Kerr effects in return-to-zero differential phase-shift keying (RZ-DPSK) and carrier suppressed returned-to-zero (CSRZ) DPSK dense wavelength division multiplexed (DWDM) systems have been investigated. The quality factor *Q* is measured at the receiving end of the transmission line. Dispersion as well as Kerr effects are mitigated. Post mitigation performance for these two systems are evaluated and compared. RZ-DPSK outperforms CSRZ-DPSK largely.**

***Index Terms*— Cross-phase modulation, CSRZ, DPSK, dispersion management, DWDM, four-wave mixing, RZ**

I. INTRODUCTION

Over the last decade traffic on Internet has increased very rapidly and various telecommunication and multimedia applications are continuously increasing the data rate and implementing suitable modulation formats to meet the demands of ever growing Internet traffic. Differential phase-shift keyed (DPSK) format, which carries the information by optical phase changing between bits, has been attracted huge research attention for optical communication systems because of its 3 dB receiver sensitivity improvement compared to on-off keying (OOK) format. It also shows more resilience to nonlinear effects and chromatic dispersion than OOK. On the other hand, return-to-zero (RZ) optical signals possess advantages in operating with high input powers, small inter-symbol interference and high receiver sensitivity compared to non-return-to-zero (NRZ). RZ-DPSK has many advantages over NRZ-DPSK with higher noise tolerance and resistance to nonlinear effects [1]. Carrier-suppressed RZ (CSRZ) signal, which is a variation of the RZ format, provides higher efficiency in spectrum utilization, more tolerance to filtering and chromatic dispersion because of its narrow spectrum [2]. Among other modulation formats which provide higher spectral efficiency like DQPSK, 8-DSPK recent works show that DPSK is more attractive than DQPSK and 8-DPSK in many cases like higher tolerance to chromatic dispersion and nonlinearity and it can handle higher bit rate and transmission reach [3]. It becomes well-known that DPSK would be the next modulation format for fiber-optic communication, however, RZ and CSRZ have been found to be the two key potential contenders as line coding.

For narrow spectrum, CSRZ could be better option for DWDM application, however, some recent works show that RZ-DPSK offers better performance than CSRZ-DPSK with dispersion-slope compensation at 40 Gb/s [4]. In presence of limited amplifier noise, RZ-DPSK has shown the best performance among NRZ/RZ/CSRZ DPSK WDM systems [5]. Combination of RZ/CSRZ and DPSK can enhance system quality [6, 7]. Our current work shows the nonlinear performance analysis of RZ/CSRZ DPSK DWDM systems at ultra-high bit rate and reveals the better option for DWDM application after mitigation of dispersion and nonlinearity.

In this work, the channel count is varied upto 33-ch for DWDM. The *Q*-factor is measured at the receiver placed at the end of dispersion-managed (DM) transmission line and different parameters like transmission reach, launch power, channel spacing, residual dispersion and bit rate are varied to observe performance. OPTISYSTEM simulation tool is used to model the system and estimate the *Q*.

II. THEORETICAL MODELING

The optical signal propagating in fiber is governed by the nonlinear Schrödinger equation (NLSE) which is solved numerically using OPTISYSTEM 12 simulation tool. The NLSE for a propagating channel can be written as

$$i\frac{\partial U}{\partial Z}+\frac{D(Z)}{2}\frac{\partial^2 U}{\partial T^2}+\gamma(Z)\left|U\right|^2 U=-i\alpha(Z)U+ig(Z)U \qquad (1)$$

where U is the slowly varying envelope of the optical pulse, Z is transmission distance, $D(Z)$, $\gamma(Z)$, $\alpha(Z)$ and $g(Z)$ represent fiber chromatic dispersion, nonlinearity, fiber loss and amplifier gain respectively. The signals for RZ-DPSK and CSRZ-DPSK are generated using one Mach-Zehnder modulator (MZM) [7].

The center channel is placed at 1552.524 nm or 193.1 THz following 100 GHz ITU grid, with the consecutive channels each having same channel spacing from the previous one. For the RZ-DPSK system, an RZ signal is achieved from the differentially encoded bit sequence by modulating the phase of carrier from laser source using MZM and a pulse carver. The CSRZ signal is achieved in a similar process but using different encoder. The DM map consists of 50 km standard single mode fiber (SSMF) followed by dispersion compensated fiber (DCF). All modulated channels are multiplexed and transmitted through the DM line. The fiber and system parameters are given in Table I. The amplifier gain is adjusted to compensate for the loss for each span. Noise figure of EDFA is 6 dB. The polarization state of all channels is assumed identical.

TABLE I: FIBER AND SYSTEM PARAMETERS

Parameter	SSMF	DCF
Attenuation, α [dB/km]	0.2	0.5
Dispersion, D [ps/nm/km]	16.75	−83.25
Fiber length, [km]	50	10
Nonlinearity coefficient, γ [1/W/km]	1.32	4.78
Effective area of core, A_{eff} [μm^2]	80	22
PMD Coefficient, [ps/√km]	0.5	0.5
Center wavelength, λ [nm]	1552.524	
Typical data rate per ch, [Gb/s]	40	

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

The 3 dB bandwidth of MUX/DEMUX filter is optimized in order to obtain highest receiver sensitivity and spectral efficiency of 0.66 b/s/Hz has been achieved.

Eye-pattern is achieved numerically at the receiver-end and Q-factor of the received signal is measured. The Q-factor is defined by,

$$Q = \frac{|m_0 - m_1|}{\sigma_0 + \sigma_1} \tag{2}$$

where, m_o and m_1 are the mean and σ_o and σ_1 are the standard deviation of the received signal measured for binary data "0" and "1", respectively. The whole system is shown in Fig. 1.

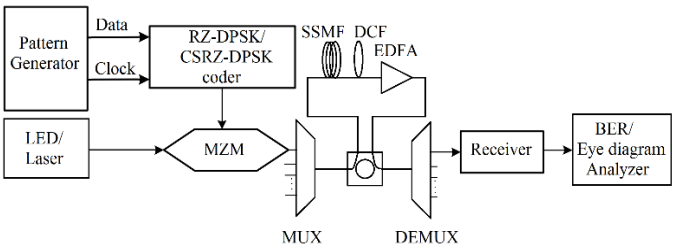


Fig. 1: System model for RZ-DPSK and CSRZ-DPSK DWDM system.

III. PERFORMANCE ANALYSES

The duty cycle for RZ-DPSK DWDM systems is 66% and typical channel spacing is taken as 100 GHz. In all cases, the worst channel is observed for measuring Q-factor using (2). In the first simulation Q-factor is evaluated by increasing the number of channels from 1 to 33. The total transmission reach is 360 km and launch power is 3mW. From Fig. 2(a) for RZ-DPSK and Fig. 2(b) for CSRZ-DPSK DWDM systems, we observe that Q-factor decreases with number of channel for uncompensated system. As we increase the number of channel, optical signal to noise ratio (OSNR) decreases due to adverse nonlinear effect i.e. addition of cross-phase modulation (XPM) and four-wave mixing (FWM). From the figure we see that Q-factor is equal to and above 6 dB (BER = 10^{-9} for Q = 6) up to 24 channels for RZ-DPSK DWDM system. For CSRZ-DPSK DWDM system it is 21 channels. After that Q goes below that margin. Nonlinear effects, mainly XPM and FWM increase with increase of number of channel, however, SPM effect might remain same as it affects single channel only. Pre-compensated, post-compensated and pre & post- compensated values for these two systems are also plotted in Fig. 2. The increase in values for Q shows that the performance degradation due to nonlinear effects are somewhat mitigated. We also find that Q values are higher (mostly in compensated cases) in RZ-DPSK than that of CSRZ-DPSK DWDM system.

Fig. 3 shows Q-factor as a function of transmission reach. For both DWDM systems, Q is equal to and above 6 dB upto 600 km for 12 channels. However, for compensated cases Q is higher in RZ-DPSK DWDM systems for short transmission reach, Q values are comparable at longer distances. In the next simulation we show the plot of Q against launch power for 24-ch DPSK DWDM system in Fig. 4. Transmission reach is taken as 360 km. Here without mitigation, Q reaches below 6 dB after 10 mW for RZ-DPSK and after 5 mW for CSRZ-DPSK. Compensated values by pre-compensation, post-compensation and pre & post- compensation are also plotted with uncompensated values. It is found that Q is above 6 dB for RZ-DPSK upto 12 mW whereas it goes below that margin for CSRZ model.

Fig. 5 shows the plot of Q vs. channel spacing for 600 km long 7-ch DWDM system. We vary the channel spacing from50 GHz to 200 GHz. At lower channel spacing like 50 GHz, Q is below 6 dB for all except RZ-DPSK with pre & post-compensation. With the increase in channel spacing, Q increases as both XPM and FWM crosstalk reduce with larger spacing.

In order to observe the effect of residual dispersion on system performance, the Q-factor is plotted as a function of residual dispersion for 21-ch DPSK DWDM system in Fig. 6. So far residual dispersion at the central channel has been kept zero. We now vary residual dispersion at central channel for both positive and negative values. At lower values of residual dispersion, RZ-DPSK DWDM system outperforms CSRZ-DPSK DWDM system. At higher residual dispersion above 30 ps/nm, CSRZ DPSK performs better. Pre & post–compensation is the best among the three mitigation techniques.

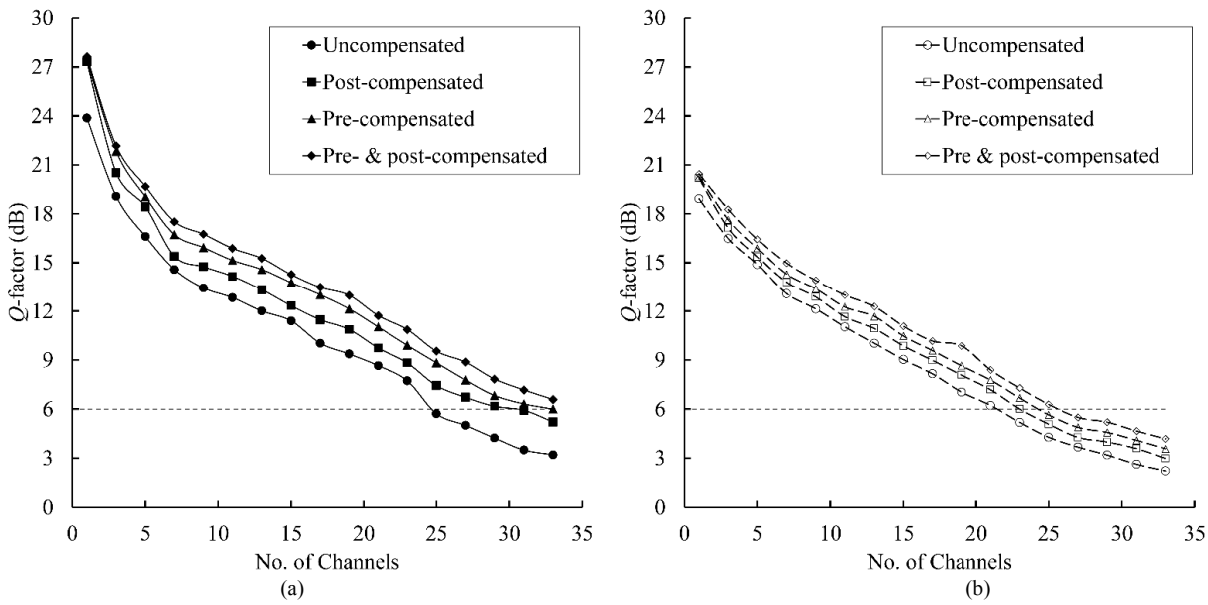


Fig. 2: Q vs. number of channels, (a) RZ-DPSK DWDM system, (b) CSRZ-DPSK DWDM system.

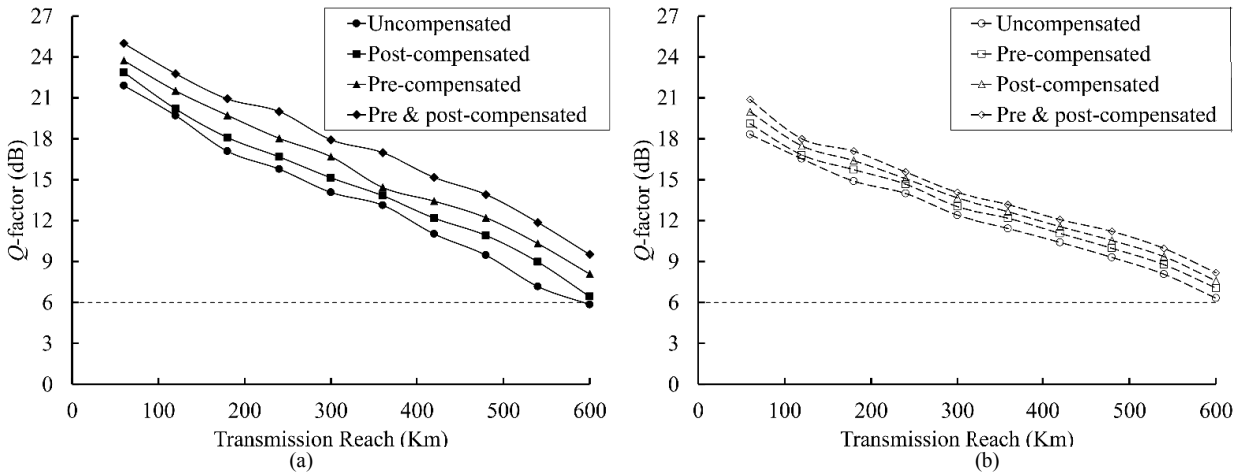


Fig. 3: Q vs transmission reach, (a) RZ-DPSK DWDM system, (b) CSRZ-DPSK DWDM system.

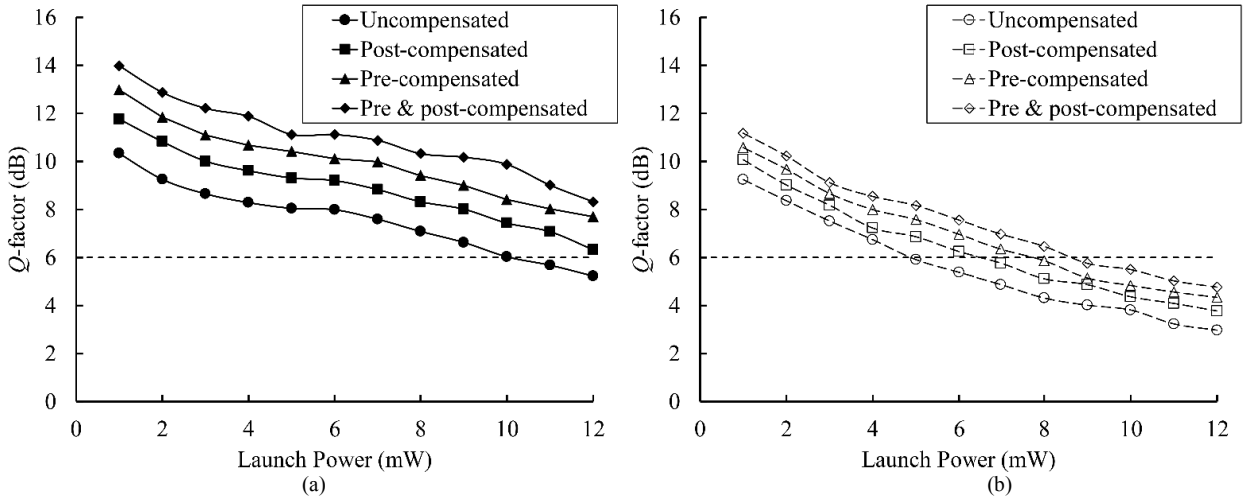


Fig. 4: Q vs. launch power, (a) RZ-DPSK DWDM system, (b) CSRZ-DPSK DWDM system.

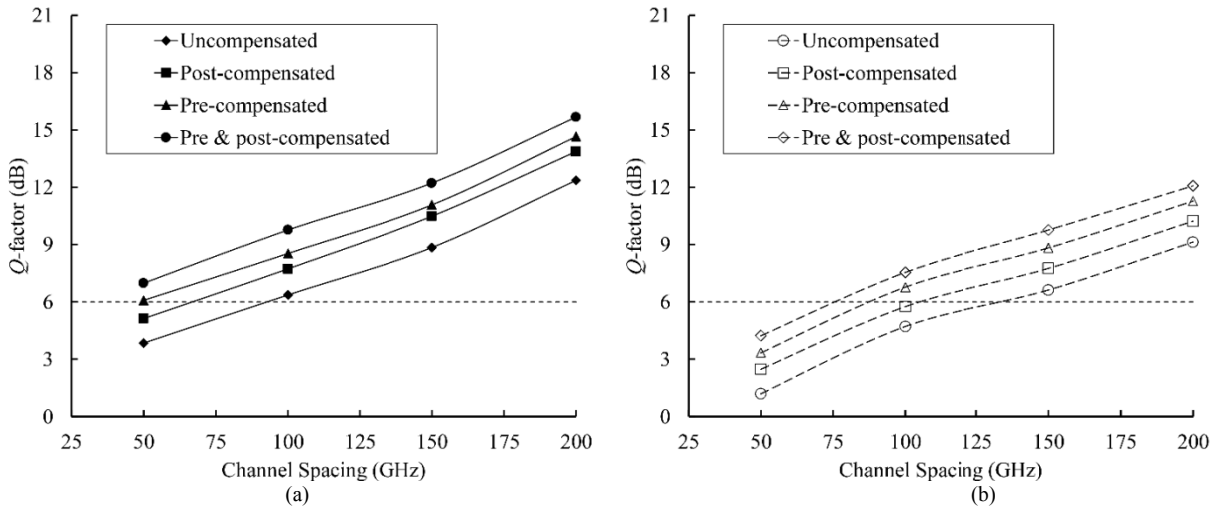


Fig. 5: Q vs. channel spacing, (a) RZ-DPSK DWDM system, (b) CSRZ-DPSK DWDM system.

The last simulation result is obtained by varying bit rate for 17-ch DPSK DWDM system. We vary bit rate per channel from 10 Gb/s to 160 Gb/s and plot the value of Q against it in Fig. 7. At lower bit rate system performance is satisfactory but degrades with increase in bit rate. The compensated values are also plotted which show that mitigation by pre & post-compensation excels the other two mitigation techniques. It is obvious that RZ-DPSK yields much improved performance compared to CSRZ-DPSK for any bit rate.

These investigations show that all the three compensation techniques reduce fiber nonlinear effects and among the three methods, pre & post-compensation is found to be the best choice for both RZ and CSRZ DPSK DWDM systems. The numerical simulation results also prove that RZ-DPSK excels CSRZ-DPSK in all cases except at higher residual dispersion. Spectrum spreads in RZ-DPSK which reduces nonlinearities and might cause the performance improvement. On the other hand, CSRZ-DPSK possesses more compact spectrum.

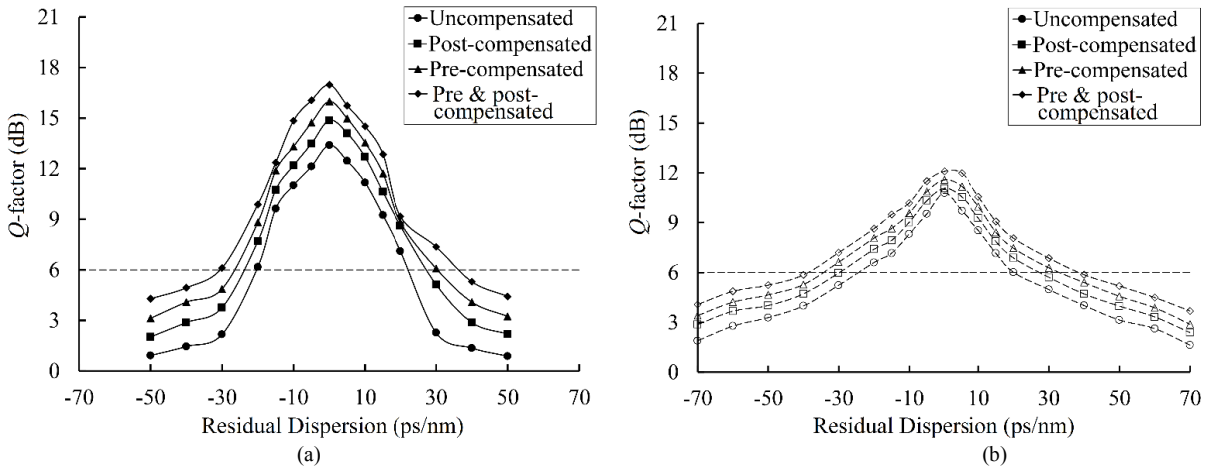


Fig. 6: Q vs. residual dispersion, (a) RZ-DPSK DWDM system, (b) CSRZ-DPSK DWDM system.

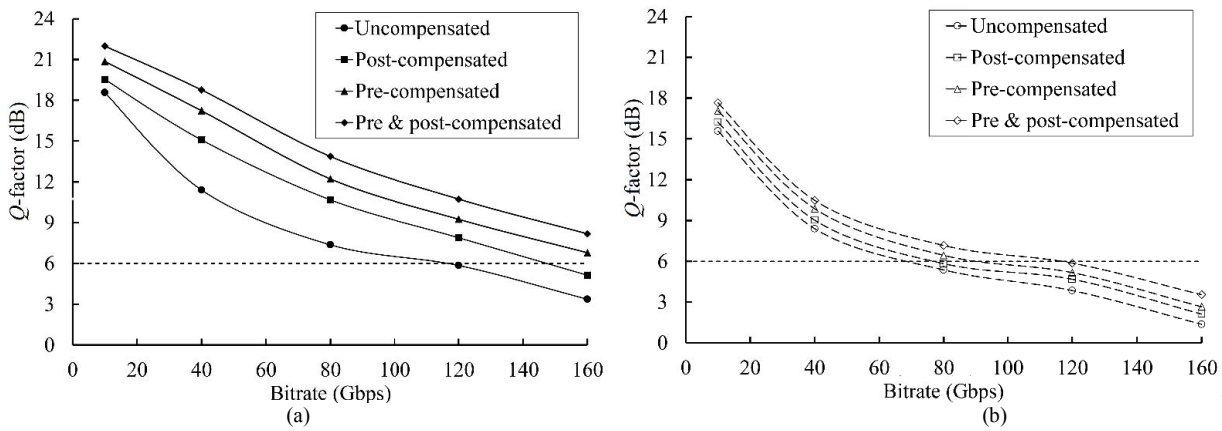


Fig. 7: Q vs. number of channels, (a) RZ-DPSK DWDM system, (b) CSRZ-DPSK DWDM system.

IV. CONCLUSIONS

In this paper, the nonlinear performance is evaluated by varying number of channels, transmission reach, launch power, residual dispersion and bit rate for RZ-DPSK and CSRZ-DPSK DWDM systems. The numerical simulation is done by solving NLSE. As far as real environment is considered, RZ-DPSK DWDM system is found to be superior to CSRZ-DPSK DWDM system in all cases except at higher values of residual dispersion. For all cases in RZ-DPSK and CSRZ-DPSK DWDM systems, pre & post-compensation mitigation technique gives the best performance improvement followed by pre-compensation and post-compensation respectively.

Future work may include experimental investigation. Intra-channel nonlinear effects and amplifier noise can be taken into account. Other dispersion and nonlinearity mitigation techniques may also be studied for DPSK DWDM systems.

REFERENCES

[1] J.-X. Cai, C. R. Davidson, D. G. Foursa, L. Liu, Y. Cai, B. Bakhshi, G. Mohs, W. W. Patterson, P. C. Corbett, A. J. Lucero, W. Anderson, H. Li, M. Nissov, A. N. Pilipetskii, and N. S. Bergano, "Experimental comparison of the RZ-DPSK and NRZ-DPSK modulation formats," in *Proc. OFC/NFOEC 2005*, vol. 4, Anaheim, CA, March 5-9, 2005, Paper OThO1.

[2] Z. Xu, K. Rottwitt, and P. Jeppesen, "Analysis of spectral efficiency and nonlinear tolerance of DPSK formats in 160-Gb/s Raman amplified systems," *IEEE Photon. Technol. Lett.,* vol. 17, no. 7, pp. 1552-1554, July 2005.

[3] M. Faisal and M. H. Rahman, "Performance comparison of various PSK modulation schemes for ultra-high speed long-haul fiber-optic communication system," in *Proc. IEEE Int. Conf. on Advanced Networks and Telecommunications (IEEE ANTS 2012),* pp. 96-98, Bangalore, India, Dec 16-19, 2012.

[4] W. T. Anderson, L. Liu, Y. Cai, D. kovsh, H. Li, A. N. Pilipetskii, M. Nissov, and N. Bergano, "Modeling of 40 Gb/s CSRZ-DPSK and RZ-DPSK trans-Atlantic transmission with dispersion slope compensation," in *Proc. OFC/NFOEC 2006*, March 5-10, 2006, CA, Paper JThB1.

[5] G. Lu, K. Xu, J. Wu, and J. Lin, "Performance evaluation of modulation formats in 40×40-Gb/s WDM repeaterless transmission systems," *Chinese Optics Lett*., vol. 4, no. 4, pp. 199-201, 2006.

[6] J. Wang, J. Sun, X. Zhang, and D. Huang, "Proposal for PPLN-Based All-Optical NRZ-to-CSRZ, RZ-to-CSRZ, NRZ-DPSK-to-CSRZ-DPSK, and RZ-DPSK-to-CSRZ-DPSK Format Conversions," *IEEE Photon. Technol. Lett*., vol. 20, no. 12, pp. 1039-1041, 2008.

[7] Y. Dong, H. He, Y. Su, and W. Hu, "RZ/CSRZ-DPSK signal generation using only one Mach-Zehnder modulator," in *proc. OFC/NFOEC 2006*, Anaheim, CA, March 5-10, 2006, paper JThB44.

Architectural Model for Next Generation Content Management System

Ankita Saha[1], S. K. Setua[2]
[1]Department of Computer Science & Engineering
University of Calcutta, Kolkata, India
arnabianki@gmail.com
[2]Department of Computer Science & Engineering
University of Calcutta, Kolkata, India
sksetua@gmail.com

***Abstract* - Content Management System (CMS) is one of the recent upcoming concepts due to its efficient and explicit use in the real world applications. The demand for growth in its diversity has been documented by many researchers. Due to the limitation of traditional database management system (DBMS), which enforces limits on the type of data items need to be stored, CMS becomes the cost-effective alternative in the scope of information management. In this ever increasing data scenario we need a framework for managing and publishing all types of data with appropriate data analytics and natural language processing. Nowadays the information, especially in the Internet, consists of structured, semi-structured and unstructured data. Therefore it is unwise to think only about the use of DBMS, rather its alternative framework must be used, keeping the objectives same. The main reason for proposing a new architectural model for a content management system is that there are several important characteristics that CMSs should have to provide opportunity to specific optimizations [1]. Our "Next Generation CMS model" includes functionalities like multi-language, copyright, multichannel publishing, natural language processing, garbage collection management which are not included in the traditional ones.**

Keywords: Fragmentation Management, Language Management, Copyright Management, Multichannel Publishing, Garbage Collection, Data Analytics(DA), Natural Language Processing(NLP)

I. INTRODUCTION

CMS is a basic framework, usually manage heterogeneous data, along with considering the security and privacy (wherever the application demands) in highly concurrent multi-user environments. CMSs are often used to run websites containing blogs, news, and shopping. Many corporate and marketing websites use CMSs. CMSs typically aim to avoid the need for hand coding, but may support it for specific requirements.

The term "content management" has several different meanings in today's marketplace. Basically, a content management system is one that stores digital content for search, browsing, access and retrieval by users in a workgroup or enterprise. The most prevalent types of content management systems are: [3]

Enterprise Content Management (ECM): systems that facilitate management of corporate documents and other types of information for use internally as well as externally with a company's business partners, customers, regulators and the general public.

Digital Asset Management (DAM): systems that manage rich media assets, often including digital audio and video clips, for retrieval and repurposing in media production environments. These systems are sometimes also called Media Asset Management (MAM).

Web Content Management (WCM): it provides page template design, editorial workflow and publishing environments specifically for Web sites and other forms of Internet content delivery.

A Document Management System (DMS) is an electronic filing system for capturing, storing and tracking document images. Access to content in a DMS should be quick and secured. Retention and Traceability should also be parts of DMS.

A Component Content Management System (CCMS) deals with content or component at a granular level rather than finished document. Each component can be represented as a topic, with a concept and a potential set of tasks defined. A component can be a chapter or a single word within a chapter. Further each component can have its own lifecycle and be tracked individually or as part of the whole document. CCMS can also be integrated with ECMs, WCMs or DMS.

Today, businesses and educational institutions are using CMS to help them for delivering targeted information for its stakeholders. Many software packages are available; however selecting a CMS is sometimes difficult because trend and evaluation data does not always coexist in the same contexts. Developing a new CMS represent great challenges especially if it copes with the organization needs. Using a process that incorporates trend analysis conducted by independent researchers and feature analysis conducted by open source organizations can often facilitate decision making. The developed CMS in KAU achieved the major requirements and needs of all customers, employee, academic members who deals with KAU either in research track or other university jobs. A lot of benefits of the developed MARZ system include: fast production and deployment of courses, reduce production and deployment cost, supporting innovation of web technology, security and privacy of contents, increased return on investment, increased security and privacy and more time and focus to learning.

Considering the different forms of CMS, we are proposing an architectural model of next generation content management system with all desirable features. The model describes the functionalities of Data Analytics (DA) module, NLP module,

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

Content Repository (CR) module along with the others in brief. We have also highlighted the measurable performance criterions of the proposed model.

The rest of the paper is structured as follows: Section II details out the architecture of the proposed model with its functional details and a brief overview of implementation part. Section III describes the performance of the model and Section IV concludes the paper with next level of development.

II. PROPOSED ARCHITECTURAL MODEL

Traditional CMS do not provide appreciable data quality. This includes timely service, quality of information and data accuracy. Nowadays all the information are either semi-structured or unstructured [2]; so there is the need for metadata [3] creation. Therefore the metadata stored in the CMS becomes the bottleneck for searching the content. But metadata based searching can be an efficient alternative. All the naive users needed technical skills for using the framework.

There are large numbers of challenges required to fulfill for maximum utilization of any CMS. CMS model must be secured from unauthorized access. A CMS must keep up -to - date strategies and maintain statistical records (required for data mining). A link from a hypertext to another location, typically activated by clicking on a highlighted word or icon att a particular location on the screen is referred as web linking. We need to avoid such switching and keep the required updated content in the CMS itself. Nowadays copyright has become an important issue in most of the real world applications. To make our CMS's performance better it would be beneficial to swipe out obsolete content. Garbage refers to the obsolete data, whose importance becomes negligible after a certain period. For example, the individual votes in polling, for selecting CEO of an organization become obsolete after the declaration of result. So we need a module to store obsolete or unused data of the CMS. Content or data security is an important issue in any CMS to protect it from intruders. One can view a single content in different formats; for example one user may need the graphical (or pictorial) format, another user may need the text (or numerical) format. Further different portion of the content must have different permission [1] for proper handling of the CMS. The possible permissions allowed can be read only, read and write only, publish only. Non-technical personnel – like marketing staff – must actively participate in managing, updating and reusing content without having to rely on technical experts.

A. ARCHITECTURE

Our CMS architectural model includes all the required functionalities discussed in section II. It is referred to as "Next Generation CMS model" because of the new modules (such as Copyright Management, Security Management, and Multichannel Publishing Management) incorporated in it. In Fig.1, pictorial representation of our architecture is shown. Two primary blocks are- Input Interface and Content Management System.

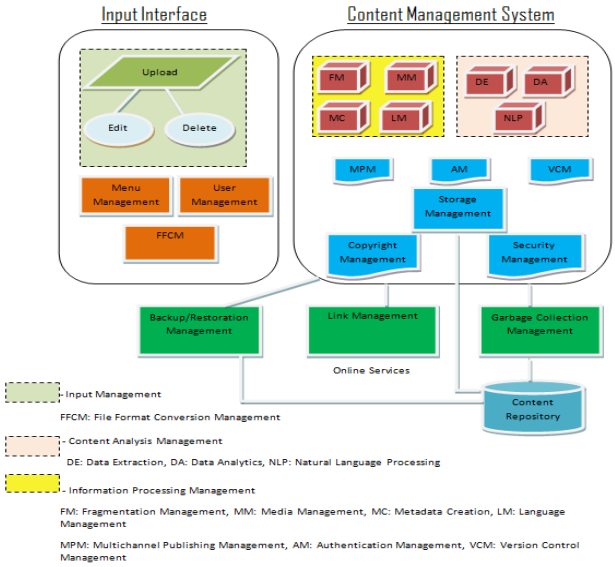


Fig.-1: Architecture of CMS

A content repository (CR) is a store of digital content with an associated set of data management, search and access methods (allowing application-independent access to the content). A CR (functions as the logical storage facility for content) thus typically forms the technical underpinning of CMS; therefore it is included in our model. With reference to any website, with every visit, visitors or customers allow you to study their browsing patterns; we can also monitor traffic, transactions, usability and even aspects like server performance. Web analytics help collect, compile and review all this data, thus providing Web managers with a crucial tool to ensure the effectiveness of the enterprise Web site and monitor conversions almost continuously. Web-based analytics typically consist of standardized analytic tool boxes through which users' can generate and view reports through a desktop or web browser. The data generated can be analyzed, graphically displayed and delivered to product development, manufacturing and senior executives who can make adjustments in real time as appropriate. Therefore Data Analytics (DA) becomes an important sub-module of Content Analysis Management. The essential attributes linked with Data Extraction (DE) are- accuracy, comprehensiveness, relevance and timeliness. It can be best explained with respect to Web CMS.

Natural Language Processing (NLP) is the ability of a computer program to understand human speech as it is spoken. The development of NLP applications is challenging because computers traditionally require humans to speak to them in a programming language that is precise, unambiguous and highly structured or, perhaps through a limited number of clearly-enunciated voice commands. NLP tasks in software programs today include: sentence segmentation, part-of-speech tagging, parsing, deep analytics, named entity extraction and co-reference resolution. "Authentication Check" is required during registration process. Each registered user will be provided a unique id for future log-in; accordingly they will be given access rights for different contents. Fragmentation Management sub-module will partition (or identify) the content of individual files into disjoint fragments of different types (audio, video, text etc.) and assign accessing pointers to them. An efficient indexing

technique will be used to index these fragments pointers. All pointers are separately stored so that the reference to content would be easier. Media Management functionality includes compression of large size files (such as a high definition image), within a fixed tolerance limit. Offline or online conversion within a stipulated time is a desirable feature of media management module. In multi-linguistic nations it is necessary to make the content available in different languages. Therefore the Language Management module will convert the content in few popular languages with the help of web link management module. The basic patent semantic for translation into different languages makes use of the Unicode. The 'moral' reasons for following copyright rules are-(1) copyright protection is essential to ensure that authors and publishers receive appropriate appreciations for their work. With Multichannel Publishing Management module, one can view a single content in different formats (for example numerical or pictorial). Garbage Collection Management module will store obsolete or unused data of the CMS. To make our CMS's performance better it would be beneficial to swipe out obsolete content. Moreover the metadata will contain a link to recover the obsolete data whenever required. Different types of garbage collection are-

Transaction-based: After termination of a project (or deal), most of the transactions related to it becomes obsolete.
Time-based: Fixed time schedule
Policy-based: Some prefixed policy will determine whether a data will be obsolete.
Poll-based (opinion poll based): Not predefined.
Context-based: Whenever a particular context goes out of scope its products become obsolete.

To protect our CMS from intruders we have two different schemes-

User level security: Uses password mechanism, each time the user logged-in.

Document level security: It can be further sub-divided into three categories:

Cryptography - Uses encryption and decryption techniques.

Trust (Reputation or Direct) - Trust towards users can be build up by satisfying few criteria over a time period (Reputation). At the initial stage we can assign trust to different users (Direct).

Policy based - Uses access control list for security specification purpose.

B. STORAGE STRUCTURE

The content, now available for the general user groups and in commercial organizations, is mostly a mixture of video, audio, text etc. So we need different structure or framework to store them. In Fig.2, A, B, C, .. are various sub domain of contents linked together; for example in an organization we will have information about personal details of employees, financial details of company(for audit purpose), customer details, etc. Within each such sub domain we will further have thousands of different files (for serving different purposes). In our problem domain, for individual file we will fragment (as shown in Fig.2) the content, so that we have mutually disjoint fragments of similar and different types. Assuming (A1,A2,A3,A4,A5), (V1,V2,V3,V4), and (T1,T2,T3,T4,T5) are the fragments pointers containing pointers to audio, video and text content respectively (in different files). Now these fragments pointers can be chained together with the help of indexing, as shown in Fig.2, and stored in the storage for efficient retrieval and access. Since buffering of different data format requires different time, systems which have different data processors (with different speed) can be utilized. For example, high speed processors can be used to access video files and low speed processors for accessing text files. Thus in a cost effective way we can have better performance.

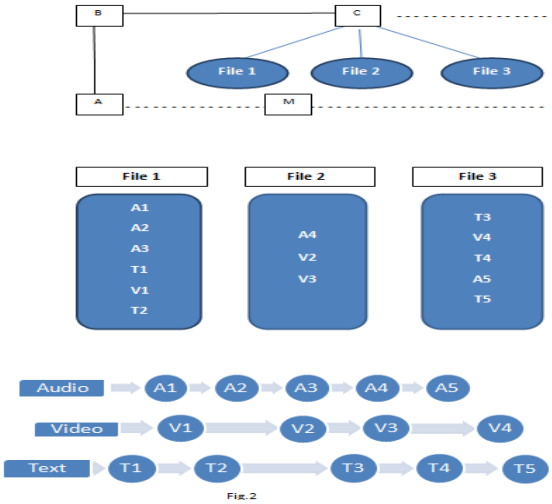


Fig.2- Storge Structure

At the initial stage we will have a readily available file format conversion engine (in FFCM); to convert different files formats (in different format) to a compatible format so that the new CMS can work on it. Autonomous and controlled CMS [1] are the two versions which will be available in our model. This will make our CMS user friendly. Since Autonomous CMS has the merits of self descriptive, scalable and self healing property, whereas the other one has reduced workload and prospective content manager's supervision (for better performance); therefore whatever an application will demand we can choose the appropriate one.

The basic operations (subject to specified access) which should be available for accessing the content in CMS are-update (insert/delete), read, publish, content's access right update, language translation, data extraction and write. When we are allowing a user to read (or update) a part of content then the user must provide few parameters. The parameters will not vary for content to content. For simplicity, the parameters would be the content name, what (keyword specific) data he wants to read or update and for how long. Internally, mapping will be done between content name and index (redirected by fragments pointers) to fetch the appropriate data. For updating, which can be either insert or delete, the user must be authenticated via some existing well defined authentication protocol. However not every content can be updated, to maintain system integrity or due to administrative constraints.

III. PERFORMANCE CRITERIA OF CMS MODEL

The proposed model should have the following performance criteria to maintain the sustainability of the system.

Validity: Content needs to get validated after a certain period of time during its life cycle. For example, digital signature needs to get renewed after a fixed interval of time. With respect to market condition or external world, some content gets stale. So we need to validate them within a predefined time. This helps in maintainability. In CMS we can group this in three categories-associated validity, time validity and context validity. In CMS we have multiple numbers of contents; if one becomes invalid, the others become invalid too, this is referred as associated validity. Time validity refers to validity change with respect to time. Whenever we predict something and the results are declared, the prediction estimate becomes obsolete after result declaration. This is referred as context validity.

Scalability: Scalability is the ability of a system to handle a growing amount of work in a capable manner or its ability to be enlarged to accommodate that growth. In CMS with increased resource, the performance must remain same (must not deteriorate aggressively).

Modularity: An efficient infrastructure must be build up to provide all the required services. Modules concept give better maintainability and proper scalability option, for accessing the content.

Upgradability: Nowadays most of the applications get upgraded to newer version. So we need each module to get upgraded to newer version, with the slightest change incorporated. For example, new file formats are introduced and put to use in a short span.

Response Time: Response time is the time a system or functional unit takes to react to a given input. In CMS it can be measured with respect to user satisfaction time. For example a particular content can be accessed in 10msecs at the first access, and it can take 20msecs in the second access. Therefore the user will be more satisfied in the first access than in the second. It can also be extended with respect to the growth of the content and rationality of the user.

Discover: Different aspects of content in CMS can be stored and combination of few aspects can become another aspect.

Flexibility: It is the ability of a system to respond to potential internal or external changes affecting its value delivery, in a timely and cost-effective manner. Universal readability (For example use of xml, awl, etc. in metadata creation) can be a key factor in CMS affecting the flexibility.

Reliability: Multiple copies (or replication) of content must be kept for making the system reliable, in case of failure (generally, of sub-components). With increased number of users and content, the performance of the CMS should not get affected. Different archival policy need to be used.

Availability: Irrespective of condition of the system, requisite information is available. MTBF (Mean Time Between Failure), MTTR (Mean Time To Repair), MTTF (Mean Time To Failure) and FIT (Failure In Time) are ways of providing a numeric value based on a compilation of data to quantify a failure rate and the resulting time of expected performance.

Fault Tolerance: Within a definite time interval some maintenance must be done to recover from a failure (here it is with respect to access ability). Failure can be with respect to-Network failure (which occurs due to site failure), User-related (when an authenticated user is not permitted to access a content; solution: Training message showing the user do not have access right, so that other users are not delayed) and Context-related (there is a one to one mapping between user group and individual content access ability; if any one changes failure occurs. Read failure occurs whenever the buffer overflows. So flow control is required. Write failure can be overcome by maintaining log files.

IV. CONCLUSIOS

In this paper we presented an architectural model for next generation Content Management System. We have proposed the extended functionalities of the desired CMS. In our future work we will extend the same to extract the semantic aspects of the related information, stored in CMS and the respective engines of different modules and sub-modules as service delivery points of the CMS. The CMS needs to be associated with its own and third party crawlers to extract the relevant updated information from the external repositories into the CMS under consideration

V. REFERENCES

[1].Rares Vasilescu: “Architectural Model for a High Performance Content Management System”, Internet Technology and Secured Transactions, 2009. ICITST 2009, pp: 1-6.

[2] Rolf Sint, Sebastian Schaffert, Stephanie Stroka, and Roland Ferstl: “Combining Unstructured, Fully Structured and Semi-Structured Information in Semantic Wikis**”,** In: 4th Workshop on Semantic Wikis (SemWiki2009) at ESWC09, Heraklion, Greece, June 2009.

[3] Khalid S. Husain: “Performance Evaluation Process of the Content Management System: Case Study of Kau Marz System”, Middle-East Journal of Scientific Research 11 (1): pp:117-127, 2012.

[4] Philip S. Yu: “Data Mining and Personalization Technologies”, Database Systems for Advanced Applications, 1999. Proceedings, 6th International Conference, pp: 6-13.

[5] Niko M¨akitalo, Heikki Peltola, Joonas Salo and Tuomas Turto: “VisualREST: A Content Management System for Cloud Computing Environment”, Software Engineering and Advanced Applications (SEAA), 2011 37th EUROMICRO Conference, pp-183-187.

ChaScript: Breaking Language Barrier using a Bengali Programming System

Arman Kamal[1], Md. Nuruddin Monsur[2], Syed Tanveer Jishan[3] and Nova Ahmed[4]

[1,2,3,4] Department of Electrical and Computer Engineering, North South University
Plot-15, Block-B, Bashundhara, Dhaka 1229, Bangladesh
[1]arman1790@gmail.com; [3]tanveer_syed@live.com; [4]nova.ahmed.2011@gmail.com

Abstract—**As technology is evolving as a driving force in our society, the need for computer programming is arising. However, since most of the computer programming languages are English-based, they can act as barriers for learning computer programming for people who are not adequate in English. In this paper we present a Non-English based educational programming system which was build with the goal to help Bengali speaking people learn and write computer programs in Bengali. We have also discussed how the system was build and why it will be useful for Bengali students to use as a platform to learn computer programming.**

Index Terms—**Non-English Based Programming Language, Educational Programming Language, ECMAScript, Jison, JavaScript, Learning Programming;**

I. Introduction

According to HOPL [8] online database of languages there are about over 8500 programming languages which are recorded and more than one third of them are build with English keywords. The prevalence use of English keywords in programming language is mainly because it is the most common language in the world. However, people who are not properly exposed to English Language may face challenges in learning and practicing computer programming using programming languages where English keywords are in use.

It is imminent that in order to understand and practice computer programming in greater depth students need to use the most common programming languages. Hence, Non-English based educational programming language will act like a bridge that will introduce users to the world of logic by solving existing problems they have probably solved in math classes or similar problem solving classes. Once the concept of logic is clear, the transition towards a full-fledged programming language would be easier.

In this paper we are presenting a Bengali educational programming language called ChaScript[1]. ChaScript was build in order to create a platform for Bengali speaking people learn computer programming without the need for English Language. To understand the need of Bengali programming language first we need to know about Social Cognitive Theory[4][6]. Social Cognitive Theory states that, learning from other people by means of observing them is an effective way of gaining knowledge and altering behaviour. In case of common programming languages it is a tough task. Not only does these programming languages have very complicated structures but their documentations are also a huge barrier for people with limited knowledge of English. This is where ChaScript[1] comes in handy. It gives a gateway for new programmers to start coding and getting the hang of it through observation and do it so without any particular need of supervision. Another aspect of Social Cognitive Theory states that a learner improves his newly learned knowledge or behaviour by putting it into practice. In common programming language cases, this might not always be easy because learners find it hard to apply them in real life scenarios. ChaScript, on the other hand, is designed in a way that allows students to use it easily for solving computational and mathematical problems.

In order to build the programming system, the grammar of the scripting language ECMAScript[9] was used and the grammar was parsed using a JavaScript parser generator, Jison. Since ChaScript contains modified parser syntax and it can be considered as a translated version of ECMAScript.

The paper is organized as follows, in section II, we present related works, section III presents implementation of the programming system, syntax of the programming system is illustrated in section IV followed by the user experiment results on section V and then the conclusion.

II. Related Works

In order to demonstrate the need for a Non-English based educational programming language, we are going to show other programming languages of similar nature as well as we explored and also explore the domains beyond the realm of computer science. We are going to present some works where it is discussed how language plays an important role in the efficiency of learning of students.

Hindawi[11] is a free, open-source, complete non-English based systems programming platform that supports all paradigms of programming languages such as assembly language, logic programming and functional programming. It takes away language barrier allowing non-English literates to take up computer sciences and learn computer programming in their mother tongue.

Mallet[10] has discussed about how a course like mathematics can be difficult to understand in a lecture conveyed in certain language that the students are not as familiar with. Mallet has also emphasized how a person who cannot understand the spoken form of the language of instruction quickly, will have difficulty understanding what is being taught without the facility of additional learning support. Novak and Langerova[2] have considered the problems arising when teaching mathematics with the aid of tools such as computer algebra systems that are not available in the students' native language. This is a problem encountered in the Czech Republic-based classroom. The authors suggested the use of a multilingual dictionary of mathematical terminology which will help overcome such issues to an extent. Bretag et al.[3] has addressed the problem of the consistent lower-than-average achievement of international students from non-English speaking

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

backgrounds in information systems courses in South Australia through an action research project aimed at facilitating improved learning outcomes for these students. Cuevas[7] had discussed that learning of mathematics needs a variety of linguistic skills of the language with which lessons are conveyed. Students who are not habituated with the language may take considerably more time to master it.

RoboMind[12] is a programming environment with its own integrated scripting language which is available in seven different language. The goal of this system is to help people learn programming without much hurdle and regardless of language and without much hurdle.

DoLittle[5] is a Japanese based educational programming language developed by Kanemune et al. This programming language fully supports Objects Oriented Paradigm. Moreover, easy syntax are provided for Japanese people to learn programming smoothly.

III. Implementation of the Programming System

We chose ECMAScript as our intermediate language to interpret and we used Jison, a JavaScript parser generator to parse the ECMAScript grammar. We used Jison to generate a JavaScript file from the ECMAScript grammar. That JavaScript file was used to parse the intermediate language. The file was used as a parser and the intermediate file was given to it as input. During the parsing, an Abstract Syntax Tree was generated. The Abstract Syntax Tree was used to maintain the semantic information of the program properly. If the parser successfully parses the file, an ECMAScript program generated from the semantic information found in the Abstract Syntax Tree is given as output which is then evaluated by the JavaScript interpreter. If the parsing procedure cannot successfully parse the file, it raises an exception and gives detail information about the parsing error and the line number where the error has occurred.

Interpreting a language gives implementations some additional flexibility over compiled implementations. Features that are often easier to implement in interpreters than in compilers include: Platform independence: ECMAScript not only gives us platform independence, but also certain level of portability. You can run an ECMAScript (as well as our code) from any platform as long as you have a web browser in it. Reflection and reflective use of the evaluator: ECMAScript has a first order eval function which lets us compile and run the code on client side without needing a backend support. Dynamic typing: Another advantage of ChaScript is we don't need to define a variable data type thanks to dynamic typing of ECMAScript. Dynamic scoping: When a symbol is referenced, the ChaScript will walk up the symbol-table stack to find the correct instance of the variable to use. This is extremely helpful for inexperienced users.

When designing the interpretation process we wanted to make the system as modular as possible so that future addition and development to the language can be easily facilitated. We also tried to maintain a certain level of standardization because we want this software to be available for future open source development.

We start by declaring 4 string arrays. These arrays contain,

- Bengali Keywords: The Bengali keywords of our programming language. E.g. যদি, নাহলে
- English Keywords: Contains the corresponding JavaScript keyword. E.g. If, Else
- Bengali Numbers: Contains Bengali Numerals(০-৯)
- Symbols: All general symbols used in ChaScript.

We take the whole code in a string and start processing. We start the interpretation process by replacing all the Bengali numerals. We do it simply by searching for each Bengali numeral and replacing it with the array index of that number. Then we look for the general symbols in the code and add whitespace to each of the symbols for tokenization purposes. We start with an empty string as our final JavaScript code. Step by step we interpret ChaScript[1] and parse it to the JavaScript code. The Interpretation process can be divided into 6 main steps. They are

a) String Handling: For string handling we keep a string stack. Than we split each line of code into string arrays based on quotation mark (“). Theoretically, every even number element of that array should be a string. We replace that string with the keyword “str” and push the string into the string stack. This is done to avoid complication in the keyword and variable handling process. After the variable handling process, we search for the keyword “str” and every time it's found, we pop an element from the string stack and parse it to the JavaScript code. The process is shown in Figure 1.

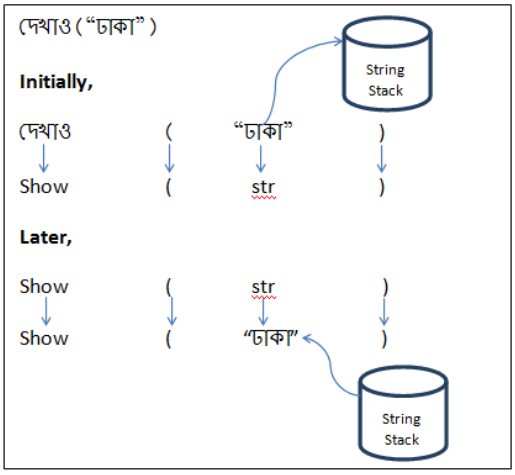


Fig. 1 String Handling Method

b) Tokenization: In this stage we simply split the whole code into several strings based on line breaks. Then each of those strings again broken down into tokens based on whitespaces. Each token is then analyzed and handled accordingly.

c) Number Handling: For number Handling we simply take each token and check all the characters within that token are a number and then parse the number to our code. If not, we go to the next stage.

d) Keyword Handling: The token later gets searched in the Bengali Keyword array, and if found the corresponding JavaScript keyword is parsed into the code. If not found in the Keyword array we go to variable handling.

e) Variable Handling: For variable handling keep two stacks. The first one keeps all the Bengali variables and the second one keeps track of their English counterpart. Whenever we find a token which is neither a keyword nor a number. We look for the token in the Bengali variable stack and if found we simply parse its corresponding English variable to our code. If not found we push the new variable to our Bengali

Variables stack. Then we randomly generate an English variable and push that in the English variable stack. Parsing that variable ends the process. The whole variable handling process is illustrated in Figure 2.

f) Input/output Handling: After the code is generated we check which platform and browser the code is running on. And implement proper JavaScript input/output function for that platform.

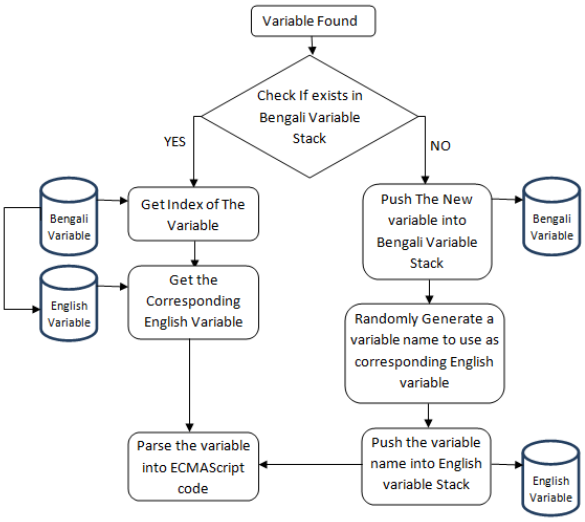


Fig 2. Variable Handling System

IV. Syntax of the Programming System

Designing ChaScript syntax our prime focus was to make the language as easy and accessible as possible. We also wanted to have maximum amount of output with minimum amount of syntax. And of course consisting with common tongue so that people new to programming language will not have much problem understanding it. If we look at the core components of our programming language we can see how easy it is compared to a generic C language .Let us start with a simple "Hello World" program which is shown in Table I.

TABLE I. Hello World Program

ChaScript	C

দেখাও ("হ্যালো ওয়ার্ল্ড")	#include"stdio.h"
	int main(){
	
	printf("Hello World");
	
	return 0;
	}

Easy variable declaration is a feature of ChaScript. There is no need for variable type declaration since type declaration is implicitly handled. If we look at the figure we can see how easily we can manipulate variables. If we tried to write the same code in C or Java, it would take a much larger set of codes and a lot of supervision. The conditional statements are almost the same as any other C-alike programming languages. However, when designing loop statement we wanted to make the concept of loop accessible to new programmers. So, instead of using conditional loop we simply run a loop and want the users to stop it manually. Although this may make the program vulnerable to infinite loop, designing the loop structure in primitive fashion will help the students to learn the logic of loop structures.

TABLE II. Fibonacci Number Generator

Using Recursion	Using Loop

ফাংশন ফিবোনাচি(ক)
{
যদি(ক == ০)
উত্তর (০)
যদি(ক ==১)
উত্তর (১)
উত্তর (ফিবোনাচি(ক - ১) +
ফিবোনাচি(ক -২))
}
 	
দেখাও(ফিবোনাচি(৩))
	ফাংশন ফিবোনাচি(ম){
	ক = ০
	খ = ১
	প=০
	গ=০
	চলবে{
	যদি (প <= ১){গ = প}
	নাহলে{
	গ = ক + খ
	ক = খ
	খ = গ
	}
	প=প+১
	যদি(প==ম){
	থামো
	}
	}
	উত্তর গ
	}
	
	দেখাও(ফিবোনাচি(৩))

In the Table II, we are representing a recursive and a loop based approach to write a Fibonacci Number generation program in ChaScript. This example highlights the syntax of this programming language where we can notice the difference in loop structure with other conventional programming languages.

V. Experiment on Non Programmers

We commenced an online survey as well as a paper based survey with approximately 110 participants. Most of these participants have not been properly exposed to computer programming before. We tested them with two simple programming problems written in ChaScript and asked them to select the right answers. Our goal was to test whether the participants can understand the programs written in ChaScript.

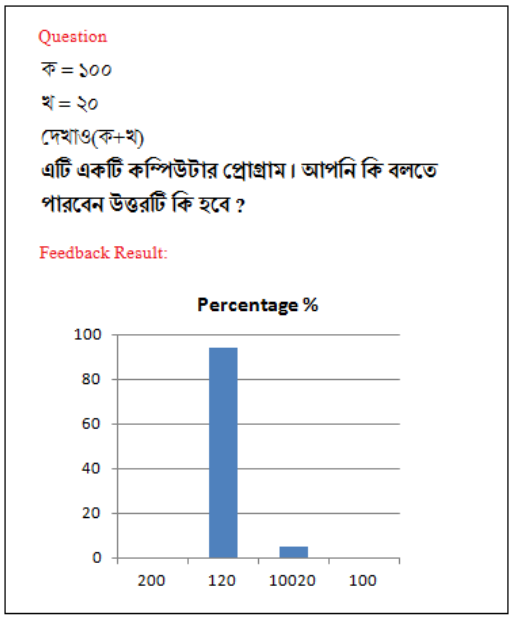


Fig. 3 User feedback on the programming question

Figure 3 shows the question that is being asked to the participants. Under the questions, a graph is presented showing the participant's responses to the multiple choice question in terms of percentages. It is evident from the graph that almost 95% of the participants chose the correct answer.

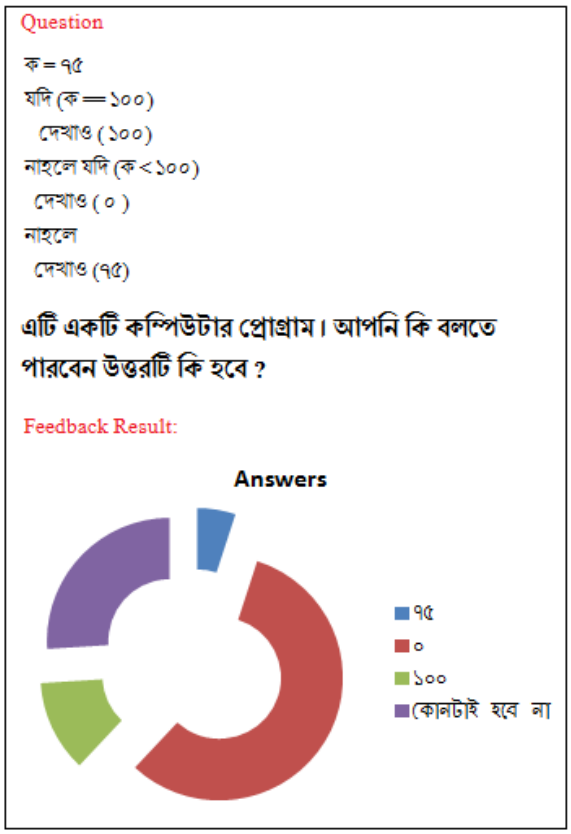


Fig 4. User feedback on another programming question

We asked the participants to answer another programming problem (Figure 4) which is slightly more difficult in nature compared to the first question. This program was also written in ChaScript and is based on conditional statements. As we can notice from the figure 4, almost 60% of the participants managed answer correctly, given that the correct answer was 0. Provided that the experiment was conducted mostly on non-programmers, getting a higher percentage of correct answers indicates that this platform can be used effectively to teach students how to program.

We would like to cordially thank the participants who agreed to take part in our survey.

VI. Conclusions

In this paper, we have introduced a programming language based on a native language that may break the initial barrier of learning the concept of programming. We have developed a full-fledged running version of this language and made it available through a web portal and the language has received positive responses from many native speakers.

As our description of ChaScript is restricted to only four pages we could not provide with full details. Although ChaScript is a well translated programming language, a lot more work is required to make it more refined and , hence, amiable to Bengali speaking individuals who are curious to learn programming but are confined because of their lack of knowledge in English language.

References

[1] ChaScript-Home, ChaScript, [online] 2014, http://www.chascript.com (Accessed:15 June 2014).

[2] M. Novák, and P. Langerová, "Raising efficiency in teaching mathematics in non-English speaking countries: an electronic bilingual dictionary of mathematical terminology", in Proceedings of 3rd International Conference on the Teaching of Mathematics at the Undergraduate Level, 2006.

[3] T. Bretag, S. Horrocks,and J. Smith, "Developing classroom practices to support NESB students in information systems courses: Some preliminary findings", International Education Journal, Vol. 3, (4), pp.57-69, 2012.

[4] A. Bandura, "Social cognitive theory: An agentic perspective" Annual review of psychology, Vol.52, (1), pp.1-26, 2001.

[5] S. Kanemune, T. Nakatani, R. Mitarai, S. Fukui, and Y. Kuno, (2004). "Dolittle-Experiences in Teaching Programming at K12 Schools", in C5 , 2004.

[6] D. Compeau, A. H. Christopher and H. Sid, "Social cognitive theory and individual reactions to computing technology: A longitudinal study.", Mis Quarterly, Vol.23, (2), 1999.

[7] G. J. Cuevas, "Mathematics learning in English as a second language.", Journal for Research in Mathematics Education, 1984, pp.134-144.

[8] *D. Piggott, HOPL: An interactive roster of programming languages.*, Perth: Murdoch University, 2006.

[9] ECMAScript, E. C. M. A., and European Computer Manufacturers Association, "ECMAScript Language Specification", 2011.

[10] D. G. Mallet, "Walking a mile in their shoes: Non-native English speakers' difficulties in English language mathematics classrooms.", Journal of Learning Design, Vol.4, (3), 2011.

[11] Hindawi-Hindi Programming System, Hindawi, [online] 2004, http://hindawi.in/en_US/ (Accessed:15 June 2014).

[12] RoboMind.net - Welcome to RoboMind.net, the new way to learn, [online] 2005, http://www.robomind.net/en/index.html (Accessed:15 June 2014).

Ontology Matching by Applying Parallelization and Distribution of Matching Task within Clustering Environment

Tanni Mittra[1,*] and Muhammad Masroor Ali[1]
[1]Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
*mittra.cse.buet12@gmail.com

***Abstract*—Recent advances of information and communication technology provides huge amount of heterogeneous information available for us. But integration of information semantically and provide machine understandable meaning to information is still a great challenge in current web technology. To overcome the challenges, ontology matching plays a vital role, which is introduced by semantic web technology. In this paper, we propose a new method of ontology matching using parallelization and distribution technique. To apply parallelism, we develop a partitioning algorithm by using property-by-class and subclass of relationship, which partitions the ontology into smaller cluster. Then the clusters from different ontology are matched based on terminological and structural similarity with semantic verification. These entire tasks of matching are handled in a parallel way and all the tasks are distributed over the computational resources. Thus, we significantly reduce the time complexity and space complexity of large scale matching task. Our proposed method reduces misaligned pairs while increasing correct aligned concepts. Validity of our claims have been substantiated through different experiments on small and large ontologies.**

***Index Terms*—Semantic web, Ontology matching, Data integration, Semantic heterogeneity**

I. Introduction

Semantic web [1], [2] is an advanced concept that has made revolutionary changes in current web technology of World Wide Web. It is the extension of current web and is built on top of it. One common reason behind the concept of semantic web is, to introduce a smart data integration agent. In current web technology, data integration is not smartly handled. For example, if we search some restaurant for particular kind of food in the web, first, we have to find out some restaurant which serves that food and then search in Google map to find out the location. We do this kind of information or data integration on the web on a daily basis, but now we are hoping that a machine will do the task for us. Another common problem of current web is that, the machine cannot understand what it represents. To solve these problems semantic web provides machine-understandable meaning and a standard way of data integration to the current web. This task of semantic web is done by constructing a vocabulary that has meaning coded inside its term.

A web page in the semantic web markup the vocabulary and thus it will get the meaning of the terms used in the vocabulary. This vocabulary is called ontology [3], [4] and it is domain specific which may be a specific subject or area such that agriculture, anatomy, education, food etc.

Since the domains have their own data vocabulary, the growth of disparate domain knowledge may cause data redundancy. To reduce data redundancy we have to manage semantic heterogeneity among them. Among the various solutions of heterogeneity problem, ontology matching [4], [5] has taken a critical place. Ontology alignment or matching is the process of determining correspondences between concepts that are semantically equivalent. It is an important operation in many traditional applications e.g. ontology merging, query answering, data translation, data warehouses etc. [4].

The evaluation in the recent years indicates that ontology matching has made a measurable improvement [4]. Considerable work has been done on automating the process of ontology alignment. System for matching and merging biomedical ontologies [6], automatic divide-and-conquer approach [7], clustering-based approach for large-scale ontology matching [8], dynamic multi-strategic matching [9], inter-matcher and intra-matcher techniques for matching [10], a session based approach for aligning large ontologies [11] and anchor wise ontology matching [12] are some of the automated processes for ontology alignment. But there are still some challenges to improve the performance of ontology matching. Among them, increasing the efficiency of matching task is still a great challenge. Because, once quality is achieved, increasing the efficiency of matching task is of prime importance. To our best knowledge, current matching tools are not optimized for resource consumption and efficiency handling. So in overall, the challenge is to come up with scalable ontology matching reference solutions. To overcome this challenge our proposed method has two major criterion. Firstly, it reduces execution time by performing the matching task by applying parallelization and distribution techniques in a clustering environment. Secondly, it performs wider automation in acquisition of reference alignments of large scale data among several ontologies. The approach presented in this paper uses cluster computing mechanism to introduce parallelization and distribution of matching tasks with available computational resources. It uses property-class relationship with subclassof relations to cluster ontology. As a result, all the cardinality relationships and instances are preserved. The issue of pairwise comparison complexity is reduced by applying only a simple version of the linguistic and structure based strategy with semantic verification. In the experimental section we compare our approach with some well established methods.

The remainder of the paper is organized as follows. Section II describes the proposed method and how it is applied in ontology matching, section III describes analytical and experimental comparison of the proposed method with some state of the art methods of matching. Finally, section IV

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

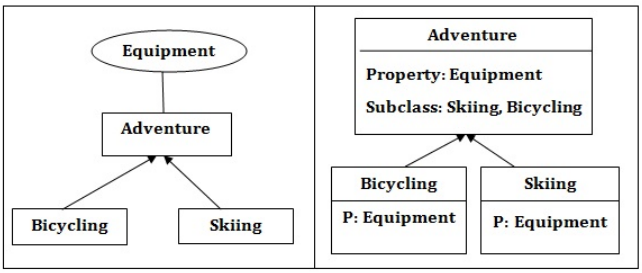


Fig. 1. Helper object model of Russia1 ontology.

provides the conclusions and future challenges of the proposed method.

II. Ontology Matching with Parallelization and Distributed Matching Task

The proposed method of ontology matching is subdivided into three basic steps. These are, 1) ontology clustering, 2) similarity matching, 3) entity alignment determination. Each of the steps sequentially produces a set of aligned concepts between two ontologies O_1 and O_2. Following subsections contain detailed description of the proposed method.

A. Ontology Clustering

Ontology may be of different sizes and from different domains. Based on this, our proposed method can take both large or small ontologies as input. One of the features of our method is that, clustering is performed for large ontologies. For small ontologies, clustering is ignored because clustering of smaller ontologies with up to three hundred classes could be a wastage of time. Classes, properties, individuals and relationships are the most common components of ontology. The first step of clustering is to create a helper object model by extracting all these necessary components and preserving the relationship that exists among them. Fig. 1 represents a small part of Russia1 ontology mentioned in [7] and a helper object model of the ontology. This ontology describes the content of a travel web site about Russia.

The helper model is used to calculate structural similarity among internal classes of ontology by using property-class and subclassof relationship, which are further elaborated hereafter. From the design principle of large ontology we come to know that an ontology is designed based on classification and modularization principle. These two issues can be easily solved by property-class relationship because the property of an ontology can provide the subclassof relationship. Suppose location is a class of an ontology and hasplace is a common property of location class. As a subclass of location class region and space get the property of hasplace. Again, ontology property ties an individual to a class and preserves cardinality relationship among entities. To handle ontologies without property declaration we use hierarchical relation among classes i.e. subclassof relationship to find structural similarity. The process of calculating structural similarity using these relationships is elaborated below.

1) Entity Vector: To calculate the similarity value the first step is, to build a property by class matrix. The key formula of building property by class matrix is property-class relationship. Assume an ontology O with $|C|$ classes and $|P|$ properties where C represents set of classes and instances and P represents set of properties. The property by class matrix is defined as (R_{ij}) and determined by using the following equation,

$$R_{ij} = \begin{cases} 1 & \text{if } C \subseteq P \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

From this property by class matrix, entity vector is calculated. Here, entity vector means vector of the classes of the ontology. As each column of the matrix represents class and each row represents property so, the entity vectors are stored in the column space of the matrix.

2) Internal Structural Similarity Calculation: After finding out the entity vector, the similarity between two entities is determined by dividing the total number of common properties by the total number of uncommon properties between them. According to this rule, two entities are similar if they share more common properties between them. This is done by using Jaccard coefficient [13], [14] which is a measurement of asymmetric information on binary variables. Our entity vector is a binary vector with 1 or 0 representation. By using entity vectors r_i and r_j we get the following equation to calculate structural similarity.

$$\mathrm{Sim}(r_i, r_j) = \frac{\#(r_i = r_j = 1)}{\#(r_i = 1) + \#(r_j = 1) - \#(r_i = r_j = 1)} \tag{2}$$

3) Partitioning Ontology: Clustering of the ontologies starts with the determination of the seed nodes by computing shared number of properties between classes. Initially all the classes are the member of seed node set. If two classes share the same properties then we remove one class from the seed node set. Otherwise, we check hierarchical relations between them. If they are connected hierarchically then we again remove one class from the seed node set. We repeat the whole process until each of the classes is checked and thus we get a seed note set.

For partitioning classes we begin our work with the remaining classes of ontology except seed classes. First, the $|S|$ seed classes are individually assigned to the $|S|$ clusters as their initial members. Here, each seed node acts as the starting point for each cluster in the cluster set S. Then we calculate the internal structural similarity value of the remaining class members with each of the classes in the cluster set. We choose the maximum value among them and assign the classes to the appropriate cluster set for which the similarity value is higher. Thus, by repeating this process, each class is assigned to a particular cluster.

B. Similarity Matching between Entities

Integration of ontology is not as simple as just mapping classes in one ontology to corresponding classes in another ontology. Finding matching by maintaining the actual meanings of classes is a great challenge for ontology matcher. So by keeping this in mind we have organized our matching task by computing structural similarity and terminological similarity with semantic verification. Description of the matching task is preseted below.

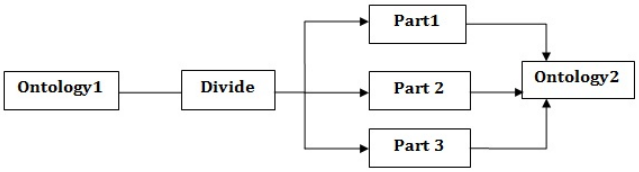


Fig. 2. Parallel matching task.

1) Terminological Similarity Calculation: The terminological similarity is calculated in two steps. At first step, it is checked that if two strings are identical or not. Here the term identical string means two strings having the same semantic significance. Again, for dissimilar strings a limit parameter is selected. If the difference between two strings are less than the limit value then we select those strings for the next step of terminological similarity calculation. The next step is to computed string similarity using WordNet lexical database [15].

But, if two entities are terminologically equal then it is not mandatory that each entity represents the same meaning in the ontologies. To ensure that each similar entity contains approximately same meaning we find out semantic similarity or proximity between them. To find out semantic proximity between each terminologically similar entity we first calculate their distance from the common super node in the class hierarchy. Distance means the depth of class in the class hierarchy. In our proposed method, if the depth of both classes is equal then they are treated as semantically equal. Then, finally we will treat them as aligned concepts. Thus, using this strategy we have found that, it improves the matching results with the inclusion of the correct matches.

2) Structural Similarity Calculation: If two entities are terminologically similar then we collect all the neighbouring entities of the aligned concept. Then we check the terminological similarity between the neighbouring concepts and find out the structurally aligned concepts. We repeat this process until each neighbour is compared.

3) Parallelization and Distribution of Matching Task: The similarity matching process described above is handled by parallel and distributed matching technique [16]. The main motivation behind this task is that, ontology matching requires a robust and scalable solution that ensures the maximal efficiency of matching operations. This is the main reason for which ontology is partitioned into small clusters. For each of the clusters, similarity matching task is performed in parallel and distributed over the computing resources. Fig. 2 describes how to divide the priority ontology and match the concepts of each part in parallel with the concepts of a selected ontology.

III. Experiments and Evaluation

We have implemented the proposed approach in Apache Jena [17], a free and open source Java framework for building semantic web and linked data applications. We have applied our proposed method on variety of datasets, including small ontologies such as Bibliography, Russia12, TourismAB and large ontologies such as Anatomical and Food ontology. For better comparison, our experimental study has been performed on data sets, considered in [7]. In this section, we discuss the result attained in our experiments in terms of scalability, memory consumption and time efficiency.

A. Evaluation Dataset

In our experiment, as mentioned above, we have used two ontologies Russia12 and TourismAB which have been mentioned in [7]. Among them, Russia1 contains 151 classes, Russia2 contains 162 classes and TourismA contains 340 classes, TourismB contains 474 classes. We have also used the Anatomy and Food ontologies[1]. Anatomy includes two ontologies, whereas Human contains 3304 entities and Mouse contains 2743 entities. Again, Food ontology contains two ontologies, whereas AGROVOC contains 28,439 entities and NALT contains 42,326 entities.

B. Experimental Methodology

We have conducted three types of experiments in order to ascertain, 1) quality of clustering, 2) correctness of alignment, and 3) compute the scalability, memory consumption, time complexity of finding alignment between ontologies.

The first experiment is done to determine the clustering quality. Because, clustering quality reflects the effectiveness of the clustering algorithm and good clustering quality indicates the possibility of discovering correct alignments. To determine clustering quality, we manually partitioned the ontology into clusters using close human intervention. After that, we again partitioned them with our proposed clustering approach. Then we compared the resultant clusters with the manual clustering result and determined the correctness of clustering.

Our second experiment is to determine the mapping quality (correctness) of the proposed method. For this reason, we evaluated our method with a popular ontology matcher, FALKON [7]. Because that matcher worked with large ontologies and used clustering method to determine alignment between entities. Moreover, in our third experiment, we experimented to observe the time efficiency, effects on memory consumption and to evaluate scalability of our proposed method.

C. Experimental Result

Experimental results, as elaborated below, of our proposed method can be judged based on the aspects which have been mentioned previously.

1) Clustering Quality: According to our first experiment of determining clustering quality the result produced by close human intervention rather than applying proposed clustering method, we found 5 clusters for Russia1; 6 for Russia2; 6 for TourismA and 7 for TourismB. On the other hand, our proposed clustering algorithm has produced almost the same number of clusters for Russia12 ontology. But Tourism AB has produced few more clusters than the number of clusters produced by human intervention. Table I shows the number of clusters that have been produced from different ontologies. In our method the clusters are not partitioned into smallest blocks like FALKON [7]. Because we build clusters automatically from ontologies by maintaining internal structural proximities and neighboring relationship between classes. Again, we kept the number of clusters smaller to reduce the computation cost of similarity mapping. Thus, our proposed approach is quite different from the previously mentioned method.

[1]http://oaei.ontologymatching.org

TABLE I
TABLE TO REPRESENT CLUSTER NUMBER

Ontology	Class Number	Property Number	Cluster Number
Russia1	151	76	6
Russia2	162	81	6
TourismA	340	97	8
TourismB	474	100	8
Mouse	2743	2	52
Human	3304	1	63

2) Time efficiency: If we consider the runtime spent by our method and FALKON during partitioning of ontologies and matching entities, we find that our method is more efficient than FALKON. Based on the current implementation, our method takes around 2 seconds to complete the comparison between Russia12 and 5 seconds for TourismAB, while FALKON spends nearly 5 seconds and 8 seconds correspondingly. Again, we also performed other different experiments over large ontologies for testing performance, which are shown in Table II. It may be argued that the machines used and hence the processing powers are different for the aforementioned methods. Still, as we have used parallelism and distribution technique for alignment measurement so our algorithm should take less time to produce alignment between entities.

3) Correctness of Alignments: The average number of comparison to find alignment between classes of our proposed method is smaller. Because, once we found alignment between two classes, then according to the nature of the cluster, alignment easily found between their neighbouring classes by collecting and comparing them with each other. Thus the average number of comparisons per aligned pair of our algorithm is approximately determined by total number of class/average number of children. The main motivation of our proposed method is to calculate one-to-many alignment between classes. As we compared the correctness of our aligned concepts with FALKON, the comparison results showed that, our method produced 45 alignment for Russia12 and 52 for TourismAB while FALKON constructs 39 mappings for Russia12 and 33 ones for TourismAB. We found that, the potential computational cost of our proposed method on discovering alignments is smaller than that of FALKON and the number of block mappings is higher than FALKON. The comparison results for other large ontologies are exhibited in Table II.

4) Memory Efficiency: We have experimented with both large and small sized ontologies. Table II contains the name of ontologies, memory consumption to store ontology model in memory. We have stored classes, properties and relationship of ontologies in a managed and structured way to minimize the memory consumption. We experimented for retrieving alignments across several pairs of ontologies and we hardly found out of memory exception. As we have used parallelization technique so each of the process is distributed within computational resources and increase memory efficiency.

TABLE II
MEMORY CONSUMPTION AND ELAPSED TIME OF COMPARISONS PER ALIGNED PAIR.

Ontology1	Ontology2	Memory(MB)	Time(s)	Aligned Pair
Russia1	Russia2	7	2	45
TourismA	TourismB	10	5	51
Mouse	Human	205	60	1091

IV. CONCLUSIONS

In this paper, we described our proposed method that can align ontologies of arbitrary size effectively, that helps our method to achieve high performance and scalability over previous alignment algorithms. Our proposed method automatically performed clustering on large ontologies and ignores clustering for smaller ones. For each ontology we created our own persistent storage with classes, properties and relations. We used both subclassof relationship and property-to-class relationship to perform clustering between entities. After clustering, we performed alignment operation in parallel for each of the entities of clusters, which reduced run time and out of memory exception. In similarity matching part, we used WordNet and string matching algorithm in our own way. We calculated both terminological and structural similarity with semantic verification between entities. We experimented our proposed method with popular ontology matchers and found that our proposed method outperformed in some aspects to other alignment systems.

In future, we will work on ontology matching that can find out similarity between ontologies of different languages. our future target includes, increasing the alignment strength by finding $m : m$ alignment between entities of ontology and implementing a tool for web interface that can find alignment between entities dynamically to work with semantic web services.

REFERENCES

[1] L. Yu, *A developer's guide to the semantic Web*. Springer, 2011.
[2] "Semantic web." http://www.w3.org/standards/semanticweb/.
[3] S. Staab and R. Studer, eds., *Handbook on Ontologies*. International Handbooks on Information Systems, Springer, 2004.
[4] J. Euzenat and P. Shvaiko, *Ontology Matching*. Springer, 2013.
[5] P. Shvaiko and J. Euzenat, "Ontology matching: state of the art and future challenges," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 25, no. 1, pp. 158–176, 2013.
[6] P. Lambrix and H. Tan, "Sambo—a system for aligning and merging biomedical ontologies," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 4, no. 3, pp. 196–206, 2006.
[7] W. Hu, Y. Qu, and G. Cheng, "Matching large ontologies: A divide-and-conquer approach," *Data & Knowledge Engineering*, vol. 67, no. 1, pp. 140–160, 2008.
[8] A. Algergawy, S. Massmann, and E. Rahm, "A clustering-based approach for large-scale ontology matching," in *Advances in Databases and Information Systems*, pp. 415–428, Springer, 2011.
[9] J. Li, J. Tang, Y. Li, and Q. Luo, "Rimom: A dynamic multistrategy ontology alignment framework," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 21, no. 8, pp. 1218–1232, 2009.
[10] Y. R. Jean-Mary, E. P. Shironoshita, and M. R. Kabuka, "Ontology matching with semantic verification," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 7, no. 3, pp. 235–251, 2009.
[11] P. Lambrix and R. Kaliyaperumal, "A session-based approach for aligning large ontologies," in *The Semantic Web: Semantics and Big Data*, pp. 46–60, Springer, 2013.
[12] M. H. Seddiqui and M. Aono, "An efficient and scalable algorithm for segmented alignment of ontologies of arbitrary size," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 7, no. 4, pp. 344–356, 2009.
[13] P. Jaccard, "Distribution de la flore alpine dans le bassin des dranses et dans quelques régions voisines," *Bulletin de la Société Vaudoise des Sciences Naturelles*, vol. 37, pp. 241–272, 1901. (in French).
[14] "Jaccard index." http://en.wikipedia.org/wiki/Jaccard_index.
[15] G. A. Miller, "Wordnet: a lexical database for english," *Communications of the ACM*, vol. 38, no. 11, pp. 39–41, 1995.
[16] R. H. Carver and K.-C. Tai, *Modern multithreading: implementing, testing, and debugging multithreaded Java and C++/Pthreads/Win32 programs*. John Wiley & Sons, 2005.
[17] "Apache Jena." https://jena.apache.org/.

Towards Privacy-aware Photo Sharing using Mobile Phones

Sadia Shamma and Md. Yusuf Sarwar Uddin
Computer Science and Engineering Department, BUET
shamma167@gmail.com, yusufsarwar@cse.buet.ac.bd

***Abstract*—Mobile phones are used these days to capture photos almost everyday everywhere. When photos are taken and shared among peers through social network sites or photo sharing portals, sometimes they pose threats to personal privacy of persons who appear in those photos, particularly when the person being photographed has reservation about sharing the same. In general, photos containing detectable human faces would be handled with care. In this paper, we propose techniques and develop a mobile app that can be used as an assistant tool in order to make people aware of their photo content. The app, integrated as a part of phone's builtin Camera app, categorically detects faces in an image preview and notify the user about detected faces (if any). The user then uses his/her rightful discretion to share or store the captured picture being respectful that no one's privacy is violated through this capture. Two possible techniques for detecting faces, namely cloud based and in-phone processing based, are proposed and a few experimental results with the app have been reported.**

***Index Terms*—Mobile Phone App, Camera, Privacy, Photo sharing**

I. Introduction

This paper proposes and develops a working solution towards privacy-aware photo sharing using mobile phones. As multimedia phones are very much available now a days, people can capture photos anywhere and everywhere. Photo sharing has become a regular phenomenon these days. Online users share photos through social network websites, such as facebook, or through photo sharing portals, such as Flickr. Due to wide penetration of Internet-enabled smart phones equipped with high resolution camera, people using those phones can easily capture images of people or scenic sites and can upload them immediately to web 2.0 portals either through standard websites or specialized apps, such as Instagram. It has been reported that facebook alone stores nearly 260 billion images and users upload around one billion new photos every week [2]. While photos are intended to capture moments of life very vividly, they also pose significant threats of personal privacy. A person can take a photo of another person where the later one may not necessarily opt in for the take. The photo can be even captured stealthily without letting the subject knowing that his/her photo is taken. Usually face reveals a person's identity quite certainly, which makes the presence of detectable faces in a photo an important issue that deserves to be treated with care. The issue becomes even more sensible in some cultures when the photo contains face of a woman whose consent was not sought while taking the photo.

In this paper, we propose techniques and develop a mobile App that can be used as an assistant tool in order to make people aware of their photo content. The app, integrated as a part of phone's builtin Camera app, categorically detects faces in an image preview and notify the user about detected faces (if any). The app can also detect the gender of the faces, specially whether there is any female face. The user then uses his/her rightful discretion to share or store the captured picture being respectful that no one's privacy is violated through this capture. We propose two possible techniques for detecting faces, one using a cloud service and another one with in-phone processing. Experimental results with the app are reported.

It is note that the app does not really solve the big problem of photo privacy. The key issue is with lack of strong enforcement. The problem is privacy breach (if at all) occurs with the person whose photo is taken, but not with the person who is taking the photo and who is actually holding the camera. This can be called *subject mismatch*. We argue that our proposed solution is mostly an advisory tool that assumes responsible and sensible attitudes to human subject guided with proper judgment. The app helps in avoiding accidental sharing, particularly for automated uploading of photos as they are shoot.

II. Related Work

Many researchers worked in different platform to protect photo privacy in web services and camera devices. Sergej Zerr, Stefan Siersdorfer and Jonathon Hare have developed a search engine for privacy oriented image PicAlert! [7] It automatically identifies potentially sensitive images in the result set and separates them from the remaining pictures. But PicAlert does not support any additional features for pictures taken with mobile phones where we expect a larger proportion of private images. Roberto Yus *et. al.* have worked on Faceblock service [6]. It takes regular pictures taken by smart phones or Google Glass as input and converts them into privacy-aware pictures. Photographer and the person who is being photographed both of them need to be Faceblock user. He Wang *et. al/* proposes techniques on Insight project which recognizes human without face recognition [5].

The closest match with our work is [3] where authors propose a system architecture for privacy-aware photo sharing over social network websites. The system allows users to set privacy preferences in their personal profiles hosted in their respective social network sites. These preferences ascertain certain policies, such as whether other people upload his/her photos without his/her consent as well as whether to notify user when someone's else upload his/her photo in any portal where the user has account in. While the work addresses privacy issues, it does so in an offline manner. In contrast, our proposed scheme is rather online in the sense that we address the privacy issue right on the spot, at the very instance when the users capture photos, not in a later time when they actually

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

upload them. This happens to be possible because we develop our scheme on mobile phones.

III. Privacy-aware Photo Sharing

We propose a service to prevent people to be photographed by camera unwillingly. If someone attempts to capture a photo, it detects if there is any face in camera preview. If any face is detected, then it notifies the user with a message that he/she is going to capture a photo with a face in it. If it is unintentional, then the user can choose his rightful actions. Face detection is computationally expensive that requires advanced vision algorithms. We use existing available services to detect faces. We apply two approaches for face detection: one is to use a cloud base service for face detection and the other one is to detect face locally in phone using available vision library, such as openCV (www.opencv.org). It is noteworthy that we do not 'recognize' individual person in photos, rather 'detect' the presence of human faces, which is apparently less expensive than recognition in terms of computation, which makes it suitable for in-phone processing. Unlike recognition, it also does not require large training dataset to learn faces.

A. Cloud based Face Detection

One of our approaches is to use cloud base service for face detection. We select a photo and send it to a Web service through REST API to check whether it contains a face. If so, the user is notified with a message. We use REST APIs from SkyBiometry for our demo application.

An advantage of using cloud base face detection is that computational overhead is transferred to cloud. Cloud itself processes the image, analyzes it and provides us the result. Our application notifies user according to that provided result. Accuracy of face detection is very satisfactory in this service as it has large data set, training set and continuous learning algorithm is applied here.

Disadvantage of using cloud base face detection is privacy issue. We need to upload image to cloud server for result. If privacy issue is highly significant, then uploading image to server violates privacy policy. We need to ensure secured transport layer for image uploading and result transferring. Moreover, we need to trust the service that it does not have any image leakage. In case of strong privacy issue user may not like this cloud base face detection system. Another disadvantage of this cloud base service is service latency. It takes some time to process in cloud and show the result according to image size. And last but not bot least, it requires a persistent (and fat) data connection which may not be so viable for all.

B. In-phone processing for Face Detection

Another approach we describe is detecting faces in phone locally. It is recommended for cases where user wants to ensure strong privacy policy. In that, users do not need to upload photo to server as it detects photo locally in phone. We apply a classifier using computer vision algorithm that detects faces locally.

Advantage of in phone local face detection is that it does not hamper privacy issues. As it processes the image in phone, there is no risk of image leaking or privacy violation. Another benefit could be quick processing as it does not involving sending the photo file over Internet and receiving response back. Disadvantage of using local face detector is, however, low accuracy. Since local face detector is tested and trained on small data set, it often fails to detect face accurately. Processing capability of phone varies according to phone configuration. So, processing time and performance may vary according to the configuration of phone.

IV. Implementation

We implemented privacy-aware photo sharing as an Android app. The app uses builtin camera application to capture photos and then analyzes to check whether the photo contains any human faces in it. If so, it prompts the users for his/her explicit consent for the actions to follow. in current implementation, the app runs as a standalone app. That means, the user can easily bypass this app to take photos with phone's builtin camera which does not have, as we know, any privacy enforcement. The app can indeed be wired into camera application so that enforcement can be activated. We do not, however, consider this option in this work.

One important aspect of the app is to detect human faces when a photo is taken by the camera. As we mentioned, for detecting faces we use two techniques. First one is to use a cloud base service and second one is to detect face locally in phone. In the following, we describe these two techniques.

A. Cloud based implementation

In our implementation, we used SkyBiometry REST API for face detection [1]. SkyBiometry is a cloud-based face detection and recognition API. It uses computer vision algorithms for face detection and recognition. This API provides facial attributes like gender, glasses, eyes state, lips state, and mode. Face detection and recognition API currently supports methods, such as faces/detect, faces/recognize, faces/train. The API uses REST interface, which means API methods are called over the Internet, using standard HTTP methods like GET and POST to api.skybiometry.com/fc/. Depending on the HTTP request parameters, server can generate response in either JSON or XML format. It also supports JSONP callbacks to simplify the application development in JavaScript and callbacks for asynchronous invoking of lengthy operations. According to the result provided by service a face detection message is shown. As this service supports gender detection, so female faces can be especially marked for additional attention.

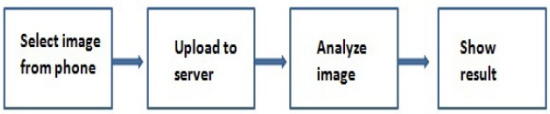


Fig. 1. Steps in cloud based face detection

B. In-phone processing

We use openCV library ported into Android for an in-phone processing of face detection. OpenCV's face detector uses a method that Paul Viola and Michael Jones published in 2001 [4]. This approach of detecting objects in images combines four key concepts: simple rectangular features, called Haar features, an integral image for rapid feature detection, the AdaBoost machine-learning method, and a cascaded classifier

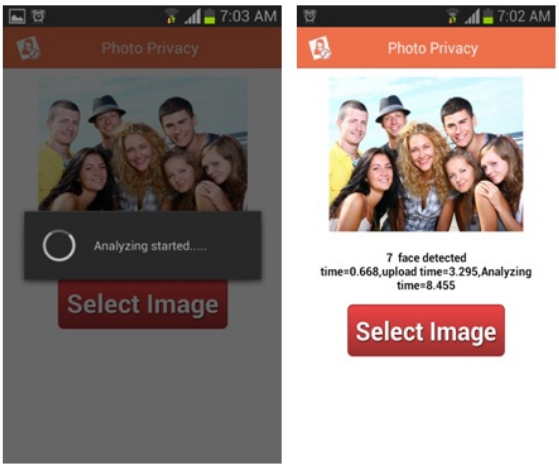


Fig. 2. Detecting face in cloud based service

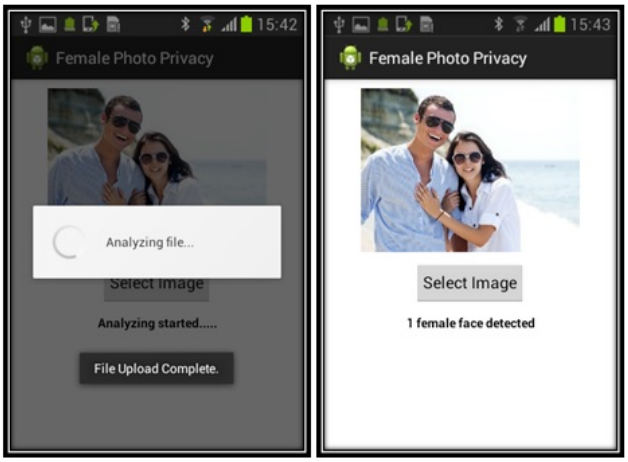


Fig. 3. Detecting female face in the cloud based service

to combine many features efficiently. It is a machine learning based approach where a cascade function is trained from a lot of positive and negative images. It is then used to detect objects in other images. OpenCV already contains many pre-trained classifiers for face, eyes, smile etc. Those XML files are stored in specific directories in openCV installation. We use openCV's face detection algorithm in our app. It detects face in camera preview and notifies user with a message when a face is detected. It then asks for user's permission to capture the photo with face. If the user wants to proceed in capturing photo with face, then it saves the image in phone memory. Otherwise, it discards the photo.

Fig. 4. In phone local face detection

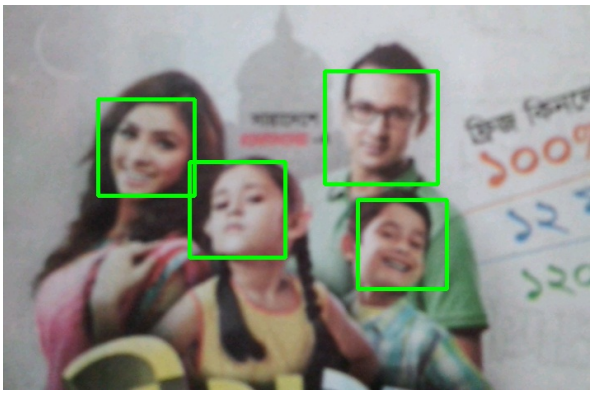


Fig. 5. In phone processing face detection

V. Experiments and Evaluation

We make experiments on our developed app with a small dataset. Firstly, we show results for the cloud based implementation. We measure the processing time for detecting faces for each photo set. We experiment with different resolution of photos as resolution affects processing time significantly. It is shown in Figure 7 that photos with higher resolution need more processing time than lower resolution. Cloud based service usually requires larger processing time as it needs transfer of image data to the server, which may constitute a large portion of processing time. We can in fact express processing time as a sum of its several components, as follows:

$$p(t) = pp(t) + u(t) + a(t)$$

where,
$p(t)$ = Processing time;
$pp(t)$ = Pre-processing time;
$u(t)$ = Data upload time (transfer image data);
$a(t)$ = Analyzing time (detecting faces);

Fig. 6. Accuracy for cloud based face detection

We also measure accuracy of our app in detecting faces. Accuracy is calculated with the following equation:

$$\text{accuracy} = \frac{\text{no of true positive + no of true negative}}{\text{total number of photos}} \quad (1)$$

Where, true position refers to cases when detection technique indeed detect faces, whereas true negative does not find

any face when there is indeed no face in the photo. False cases are those where things happen in contrary: detecting faces when there is actually none or misses a few faces to be detected. We show accuracy results in Figure 6. We see that photos with lower resolution have lower accuracy compared to photos with higher resolution. Since cloud base service uses more complex and efficient algorithm, it has better accuracy than in phone local face detector.

We also show detail accuracy results for openCV based in-phone face detection for a set of captured photos in Table I. We conclude that accuracy for cloud based service is almost 100 percent, while in-phone processing accuracy is 73 percent.

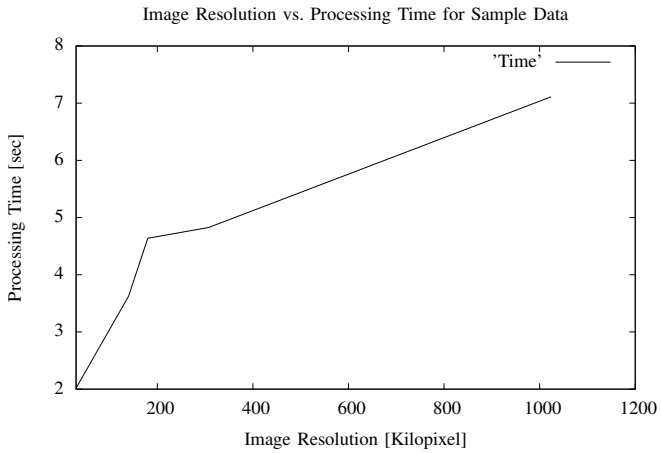


Fig. 7. Processing time for cloud based face detection

Pic ID	No of Faces	True $+ve$	True $-ve$	False $+ve$
114302	1	1		
114345	1	1		
114353	1	1		
114415	1	1		1
114435	2	2		
114503	1	0		
114513	2	1	1	
114523	5	2	3	
114554	4	3	1	
114617	1	1		
114627	0	0		
114641	1		1	
114701	1	1		1
114742	2	1	1	
114750	1	1		
114754	1	1		
114802	1	1		
114820	2	2		
115133	1	1		
115137	1	1		
115144	1	1		
115150	3	2	1	
115200	0	0		
115206	0	0		
115223	1	1		
115255	1	1		
115300	4	3	1	
sum	40	30	9	2

TABLE I
DETAIL ACCURACY CALCULATION OF IN-PHONE PROCESSING OF FACE DETECTION

VI. CONCLUSION AND FUTURE WORK

This problem of privacy breaches arising due to photo capture and sharing using mobile phones is taking much attention in recent years. This paper puts positive efforts towards privacy-aware photo sharing using mobile phones. Particularly, we propose instrumenting phones with a suitable app so that such breaches can be contained and passed to user attention immediately right after they shoot pictures of people. At this stage, the app is an independent standalone application, but it can be wired into built-in camera of devices.

Face recognition and other features can be added as an easy extension of this work. One possible thought can be as follows. Individual user can have some tagged photos of his friends/ relatives, which means these persons can be in photos without any objection. When the user attempts to capture faces other than those, the app detects a privacy breach and asks for confirmation. As a consequence, the user may avoid taking that photo or may take photo in which that unknown face is blurred. As the face area of image is detected, that area can be blurred with an appropriate measure.

Another extended version of this application can be adding features to prevent unwanted photo sharing over Internet. Photos containing unknown faces can be blocked from uploading to network. This can be integrated with gallery application of camera.

Sometime, photos in photos are automatically uploaded in some site where the user account is synced with the device. In that case, user will be asked for confirmation if image contains face before it uploads it to the respective server. If confirmation is unavailable, those photos will be kept in queue. Sometimes, phones are set of party mode, where all photos taken a particular time span are automatically shared in user's online social networks. Unwanted photos can be blocked if our scheme is activated in those cases.

REFERENCES

[1] https://www.skybiometry.com/.
[2] Doug Beaver, Sanjeev Kumar, Harry C. Li, Jason Sobel, and Peter Vajgel. Finding a needle in haystack: Facebooks photo storage. *Available at https://www.l3s.de/~siersdorfer/sources/2012/picalert.pdf*, 2010.
[3] Thorben Burghardt, Andreas Walter, Erik Buchmann, and Klemens Bohm. Primo–towards privacy aware image sharing. *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, 2008.
[4] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. *IEEE Computer Vision and Pattern Recognition*, 1:I–511 – I–518, 2001.
[5] He Wang, Xuan Bao, Romit Roy Choudhury, and Srihari Nelakuditi. Insight: Recognizing humans without face recognition. *ACM HotMobile*, 2013.
[6] Roberto Yus, Primal Pappachan, Prajit Kumar Das, Anupam Joshi Eduardo Mena, and Tim Finin. Demo: Faceblock: Privacy-aware pictures for google glass. *ACM MobiSys*, 2014.
[7] Sergej Zerr, Stefan Siersdorfer, and Jonathon Hare. Picalert!: A system for privacy-aware image classification and retrieval. *Available at https://www.usenix.org/legacy/event/osdi10/tech/full_papers/Beaver.pdf*, 2012.

Community Recommendation in Social Network Using Strong Friends and Quasi-Clique Approach

Anjum Ibna Matin[1] Sawgath Jahan[1] Mohammad Rezwanul Huq[2]

[1]Department of Computer Science and Engineering, Islamic University of Technology, Gazipur, Bangladesh
[2]Department of Computer Science and Engineering, East West University, Dhaka, Bangladesh
mrhuq@ewubd.edu

Abstract— **A social networking service is a platform to build relations among people who share interests, activities, backgrounds or real-life connections. Communities in a social network are the gathering places for the people with common interest. Social network analysis is in high demand now a days for the increasing number of users. They involve themselves into different communities. They share post, their views, what they like etc. in communities. So it is important for them to find suitable communities where they have common factors like friends, followers and their activities etc. In this paper, we propose a technique for recommending a community in social network like Facebook, Twitter etc. Finding strong friends from a user's friend list and using clique and quasi-clique concepts introduced in graph mining, we recommend suitable communities for a user in a social network.**

Index Terms— **social network analysis and mining, data mining, strong friends, social media, clique, quasi-clique.**

I. Introduction

The rapid growth and exponential use of social digital media has led to an increase in popularity of social networks and the emergence of social computing [1]. In general social networks are structures made of social entities that are linked by some specific types of interdependency (e.g. kinship, friendship, common interest, beliefs, or financial exchange).

Recommendation system like social graph generation & forecasting using social network mining [1], finding strong groups of friends [3], finding popular friends [4] in social network are well known. In this paper we develop a recommendation system for user's communities. Day by day community is increasing with the rapid growth of social users. Nowadays we see the vast activity of social life. All those community can't be suitable for a particular social users. So our recommendation system will find suitable community for a user. For example, a social user has created a new profile in social network service. He or she has some common interests with some strangers or his/her friend. So they can create a community where being friend is not important but sharing the common interest is important. Suppose a cricketer always like to join in such groups which is involved with cricket. He must have some friends in his friend list who are interested in playing cricket. We then find those friends who are connected with some community associated with cricket. Our recommendation system will then recommend the user a suitable community based on detecting his strong friends and quasi-clique techniques.

II. Problem Definition

In this section, we present the problem formulation and basic definitions for mining set of friends having higher strength and set of communities which is associated with them.

Let us consider a sample database shown in Table-I that captures some users in a social network. The table consists of the friend lists of seven users. For example, Jack has some fiends in his friend list. They are Harry, Pet, James, Oliver, Kan and Tom. The list also shows the interaction strength (e.g. connectivity of each friend of Jack). For example, "James (15)" in the list may record that the interaction between them is 15. These interactions may be of various types like posting message on wall, chatting, giving likes to photos or status for any social network like Facebook, twitter etc.

Now in Table-II, we have listed those set of communities that is associated with each user in a social community. For example, Jack is a member of C3, C4 and C5 communities.

TABLE I
Friend Database (F_{DB}) of USER (u_i) in a Social Network

User(u_i)	List of friends(f_{ui}) with interaction strength
Jack	Harry(35), Pet(25), James(15), Oliver(10), Kan(10), Tom(5)
Harry	Jack(15), Pet(5), Oliver(25)
Pet	Jack(15), Harry(10), James(5)
James	Olive(20), Pet(10), Jack(15)
Oliver	Harry(5), Jack(15), Kan(20)
Kan	Harry(5), Jack(15), Tom(20), Oliver(20)
Tom	Kan(15), Jack(20), Pet(20)

Based on this connectivity information, we recommend a suitable community for a particular user. For example, if we see Table -II, Jack is connected with community C3, C4 and C5 but he is not connected with C1, C2 and C6. But his friends, Harry, Kan, James, Oliver and Tom are connected with both C1 and C2. Pet is connected with C1. Based on the interaction with his friends we first differentiate his strong and weak friends. Then applying our algorithm we recommend a community between C1 and C2 for Jack.

In the remainder of this section, we provide formal definitions and notations required to mine friend groups and communities from a social media database.

Definition 1**:** User (u_i), we define a set U= {u1, u2… u_n} who are the users of a social network site with some friends (f_i). Example 1: from Table -I, Jack is a user in a social network. Harry, Pet, James, Oliver, Kan, Tom are Jack's friends.

Definition 2: Community (c_{ui}), a community is a group of people with a common interest among them. We define a set C= {C1, C2… C_n} that are the communities in a social network. Example 2: Pet is connected with C1, C3 communities.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

Definition 3**:** Normalized interaction strength ($nis_{ui,fj}$), interaction strength is the value or weight that defines the interaction of a user (u_i) with one of his friends (u_j) and we normalize these values based on the total number of interactions of the user (u_i). Example 3: Jack = {Harry (35), Pet (25), James (15), Oliver (10), Kan (10), Tom (5)}. Here Harry and Jack are friends and the interaction value for them is 35 and the normalized interaction strength ($nis_{ui,fj}$) is 0.35.

Definition 4: Minimum strength (min_{str}), minimum strength value (min_{str}) is the threshold or base value for recognizing strong and weak friends. In our case we set the value for min_{str} is 0.7.

Definition 5: Strong Friends (s_{ui}), we define S_{ui}= {s_1, s_2… s_n} as a set in descending order based on $nis_{ui,fj}$. When the cumulative normalized interaction strength ($nis_{ui,fj}$) exceeds the minimum strength value (min_{str}) we consider up to those friends as a strong friend set. The set indicates that the user (u_i) is most likely going to interact with these friends.

TABLE III
COMMUNITY (C_{DB}) DATABASE FOR EACH USER (u_I)

User(u_i)	Communities(c_i)
Jack	C3, C4,C5
Harry	C1, C2, C4
Pet	C1, C3
James	C1, C2,C4
Kan	C1, C2, C3, C5
Oliver	C1,C2,C3
Tom	C1,C2

Definition 6: Weak Friends (w_{ui}), we define W_i= {w1, w2… w_n} as a set of weak friends who are left in the friend list after selecting strong friends group S_{ui}. Moreover, they are such friends to which the user hardly contacts.

Definition 7: Clique, a set of vertices S is called a clique if the sub graph G(S) induced by S is complete and there is an edge between any two vertices in G(S). [7]

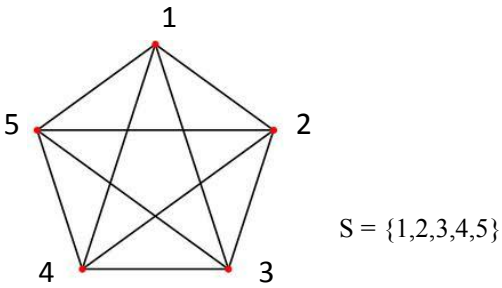


Fig. 1 Example of a clique

Definition 8: γ-quasi clique, a graph G = (V , E) is a γ -quasi-clique (0 ≤γ ≤ 1) if $\bar{G}$ is connected , and for every vertex v ∈ V , deg G(v) ≥ [γ.(|V |−1)]. [7]

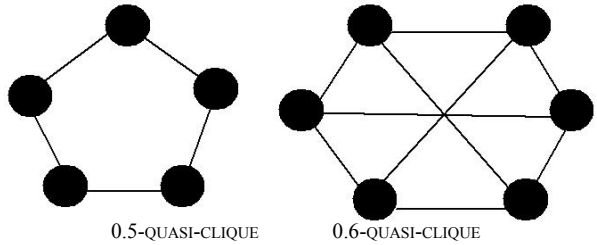


Fig. 2: Example of γ-quasi clique

III. CONSTRUCTION AND MINING.

Our proposed approach for recommending a community is based on several steps. We use clique and quasi-clique method to compare among various communities. Our algorithm first scan the friend database (Table-I) and then the community database (Table-II). After scanning, then follow some steps which are described below

STEP 1 :(Normalization) First we normalize the value of interaction strength. We calculate the normalization value of interactive strength for every friend by dividing the interaction strength with the total number of interaction to his all friends. The total value of normalized interaction strength should be equal to 1. Example: For Jack in table 1 the normalized values for his friend set (f_i) = {Harry (0.35), Pet (0.25), James (0.15), Oliver (0.10), Kan (0.10) and Tom (0.05)}. Then we arrange these normalized values in a descending order. The order of normalized interaction strength ($nis_{ui,\ uj}$) for our input dataset is given below.

TABLE IIIII
NORMALIZATION AND MIN_{STR} = 0.7

Friends of Jack	Normalized interaction strength	Cumulative normalized interaction strength
Harry	0.35	0.35
Pet	0.25	0.60
James	0.15	0.75
Oliver	0.10	0.85
Kan	0.10	0.95
Tom	0.05	1.00

STEP 2: (Recognizing s_{ui} and w_{ui}) Here we divide the friend dataset for a particular user (u_i) into two sets. One is strong friend set s_{ui} and other one is weak friend set w_{ui}. According to Table-III the result for strong friend set (s_{ui}) and weak friend set (w_{ui}) is given below.

Set of strong friends S_{ui}= {Harry (0.35), Pet (0.25), James (0.15)}

Set of weak friends W_{ui}= {Kan (0.10), Oliver (0.10), Tom (0.05)}.

STEP 3: (Creating Dense Sub graph) Here we create the dense sub graph for differentiating strong and weak group of friends. The strong friends of a user (u_i) create a sub graph and weak friend's group w_{ui} creates another. Then we calculate the γ -quasi-clique value for each sub graph. The friends of a user

who are also friend with one another from the friend database should be connected with one another. Suppose for our input data set Pet and Harry are also friends. So they should be connected.

STEP 4: (Community Recommendation algorithm):

1. The user should not be connected with the communities which will be compared for recommendation.
2. Take two communities from the list of communities of the user's friends.
3. Only consider those friends with that particular community.
4. Find the γ -quasi-clique value for each sub graph for each community and then apply our comparison algorithm for recommending a community.
5. The selected community (c_i) is then compared with another community (c_i) from the list of communities. Thus we recommend a suitable community (c_i) for and use (u_i).

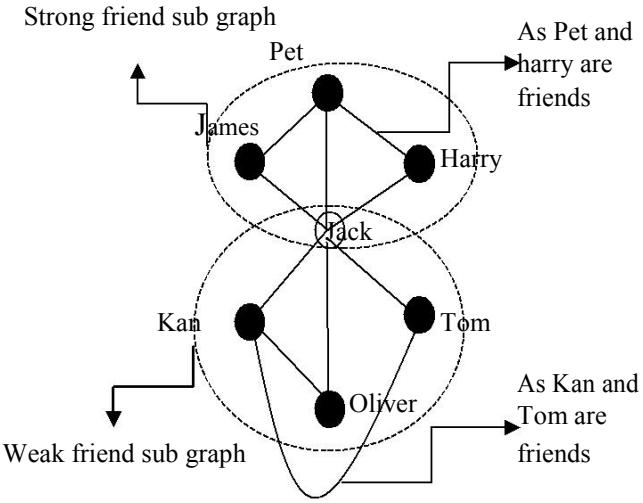


Fig. 3 Creating dense Sub graph

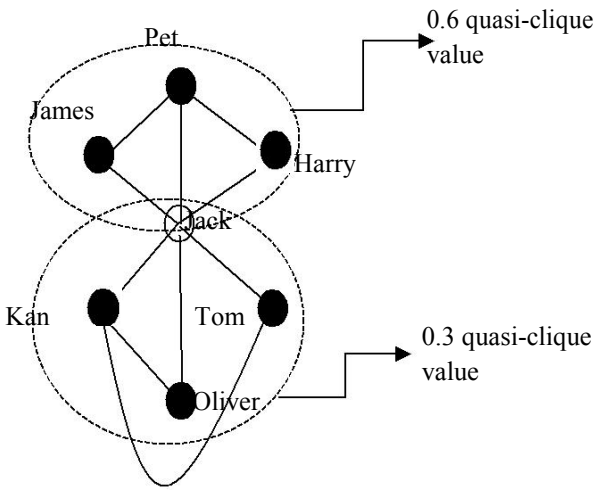


Fig. 4 γ -quasi-clique value for each sub graph of Community C1

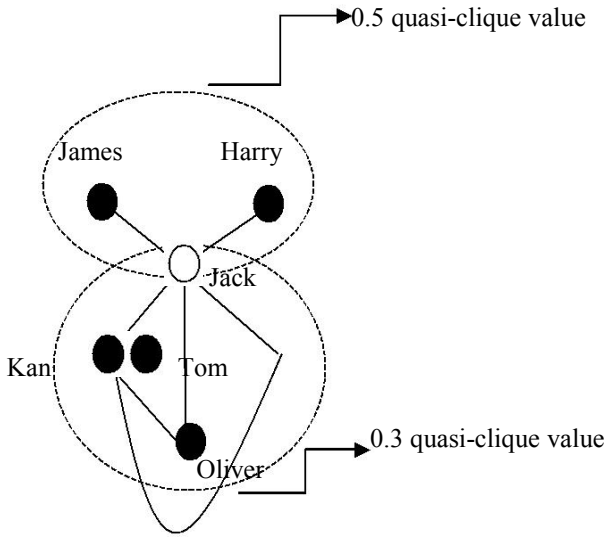


Fig. 5 γ -quasi-clique value for each sub graph of Community C2

Here q_i $[s_{ui}]$ denotes the γ -quasi-clique value for strong friends sub graph for i^{tn} community and q_i $[w_{ui}]$ denotes the γ -quasi-clique value for weak friends sub graph for i^{th} community. n_i denotes the number of nodes for each sub graph and g_i denotes the corresponding subgraph.

Community recommendation algorithm (CRA):

Input: (i) Friend database (F_{DB}), (ii) user specified minimum strength value (min_{str}), (iii) Community database (C_{DB}) Output: A recommended community (c_i)

1. If no friends are connected with the community then $q_i = 0$
2. Else if q_i $[s_{ui}] > q_{i+1}$ $[s_{ui}]$ and $n_i[s_{ui}] > n_{i+1}$ $[s_{ui}]$
3. Then $c_i =$ Recommended
4. Else if $n_i[s_{ui}] < n_{i+1}$ $[s_{ui}]$ and g_i $[s_{ui}]$ g_{i+1} $[s_{ui}]$
5. Then Recommended $c_{i+1} =$
6. Else if q_{i+1} $[s_{ui}] > q_i$ $[s_{ui}]$ and $n_{i+1}[s_{ui}] > n_i$ $[s_{ui}]$
7. Then c_{i+1} = Recommended
8. Else if $n_{i+1}[s_{ui}] < n_i$ $[s_{ui}]$ and g_{i+1} $[s_{ui}]$ g_i $[s_{ui}]$
9. Then c_i = Recommended
10. Else if q_i $[s_{ui}] = q_{i+1}$ $[s_{ui}]$ then
11. If $q_i[w_{ui}] > q_{i+1}[w_{ui}]$ and $n_i[w_{ui}] > n_{i+1}$ $[w_{ui}]$
12. Then c_i = Recommended
13. Else if $n_i[w_{ui}] < n_{i+1}$ $[w_{ui}]$ and $g_i[w_{ui}] \subset g_{i+1}$ $[w_{ui}]$
14. Then c_{i+1} = Recommended
15. Else if q_{i+1} $[w_{ui}] > q_i$ $[w_{ui}]$ and $n_{i+1}[w_{ui}] > n_i$ $[w_{ui}]$
16. Then c_{i+1} = Recommended
17. Else if $n_{i+1}[w_{ui}] < n_i$ $[w_{ui}]$ and $g_{i+1}[w_{ui}] \subset g_i$ $[w_{ui}]$
18. Then c_i = Recommended
19. Else if q_i $[s_{ui}] = q_{i+1}$ $[s_{ui}]$ and q_i $[w_{ui}] = q_{i+1}$ $[w_{ui}]$ and $n_i = n_{i+1}$

Then calculate the $\sum nis_{ui,fj}$ for each c_i

IF for strong friend sub graph,

c_i [$\sum nis_{ui,fi}$] > c_{i+1} [$\sum nis_{ui,fi}$]

Then c_i = Recommended

Else If c_{i+1} [$\sum nis_{ui,fi}$] > c_i [$\sum nis_{ui,fi}$]

Then c_{i+1} = Recommended

Else If c_i [$\sum nis_{ui,fi}$] = c_{i+1} [$\sum nis_{ui,fi}$]

Then for weak friend sub graph

If c_i[$\sum nis_{ui,fi}$] > c_{i+1}[$\sum nis_{ui,fi}$]

Then c_i = Recommended

Else if c_{i+1} [$\sum nis_{ui,fi}$] > c_i [$\sum nis_{ui,fi}$]

Then c_{i+1} =Recommended

Else both c_i and c_{i+1} recommended.

According to our CRA algorithm for community C1 in Figure-IV the γ -quasi-clique value for strong friend sub graph is 0.6 and the number of nodes is 4 and in Figure-V for community C2 the γ -quasi-clique value is 0.5 and the number of nodes is 3. And it full fills CRA's first condition. So C1 is the selected community for the user.

IV. Experimental Results

For experiments we have used a sample dataset which is given in Table-1 and Table-2 and works on various scenarios according to our survey on many users of Facebook. We have found some interesting situations like Fig. 6.

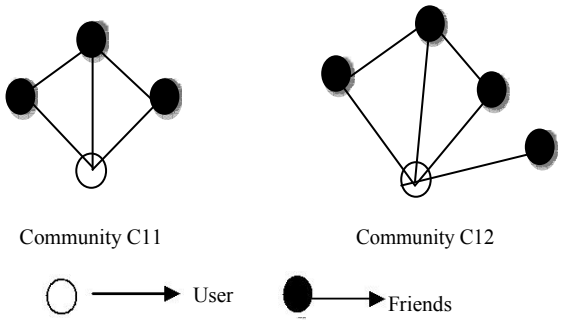


Fig. 6 Comparing communities

For community C11 the value of γ -quasi-clique is 0.6 but for community C12 the value of γ -quasi-clique is 0.25. If we only consider the γ -quasi-clique value then the resultant community is C11. But it seems that C12 is more connected. To solve this problem in our algorithm, we introduce comparing sub graph. When the γ -quasi-clique value of one sub graph is greater than the sub graph of other community, we count the number of nodes of that sub graph and compare with the other ones nodes. If the sub graph which has higher γ -quasi-clique value and has more nodes than the other one, it is recommended. But if the number of nodes is less than the other one, we check whether it's a sub graph of other one or not. If it is, then we recommend the community which has lower quasi-clique value. Here C12 is a super graph of C11 though having lower γ -quasi-clique value. So we will recommend C12 for the user. As we work on a survey of the users of Facebook, we have found our algorithm more accurate.

V. Related Work

As we work on recommending a community for an user in a social network, there are some existing works on it [1], [2]. One of those is social graph generation & forecasting using social network mining [1]. Here they used one way communication between the users and his corresponding friends. They assigned a strength value for each of the friends. And based on those values they came to an decision according their rule to recommend a community as well as the second one graph based forecasting for social networking site, communication [2], where the difference is the second one used two way directed graph to assign the strength values for defining the interactions between the users and his corresponding friends. Here we have used some basic graph techniques that is clique and quasi-clique [7] and based on those techniques we divide all the friends of an user into two groups, strong and weak, and for these we can access a portion of his friends at a time. For finding strong friends [3] or popular friends [4] the existing researches worked on finding a group of friends not on finding a single friend who is strong or weak for a particular user. But we need to detect a particular friend who is weak or strong for a particular user.

VI. Conclusion

In this paper we work through two main processes. Those are selecting strong friends and then applying quasi-clique technique to recommend a suitable community for a particular user. Usually when we work on some huge database it will take more time to check as well as to compare. But if we represent the database in a graph and work on that graph then it will be time consuming and easily understandable. As in our calculation a portion of a user's friends is accessed at a time so it is faster and we can visualize easily. Our future plan is to simplify our algorithm and add experimental results on real social network data.

References

[1] Marjaneh Safaei, Merve Sahan and Mustafa Ilkan, "Social Graph Generation & Forecasting using Social Network Mining," 33rd Annual IEEE International Computer Software and Applications Conference(COMPSAC'09), 2009, pp. 31-35.

[2] Kadge, Sanam and Bhatia, Gresha, "Graph Based Forecasting For Social Networking Site", Communication, Information & Computing Technology (ICCICT) International Conference , 2012, , pp. 1-6.

[3] Cameron, Juan J and Leung, CK-S and Tanbeer, Syed Khairuzzaman, "Finding Strong Groups of Friends among Friends in Social Networks", Ninth IEEE International Conference, 2011, pp. 824-831.

[4] Jiang, Fan and Leung, CK-S and Tanbeer, Syed Khairuzzaman, "Finding Popular Friends in Social Networks", Second International Conference on Cloud and Green Computing, 2012 , pp. 501-508.

[5] P.M. Zadeh, M.S. Moshkenani, "Mining Social Network for Semantic Advertisement", Third 2008 International Conference on Convergence and Hybrid Information Technology (ICCIT'08), 2008, Volume 1, pp. 11-13.

[6] R. Agrawal, T. Imieli´nski, and A. Swami, "Mining association rules between sets of items in large databases," in Proc. ACM SIGMOD 1993, Volume 22, pp. 207–216.

[7] Silva, Arlei and Meira Jr, Wagner and Zaki, Mohammed J, "Structural correlation pattern mining for large graphs", Proceedings of the Eighth Workshop on Mining And .Learning with Graphs, 2010.

Novel Approach for Detection of Inter-Turn Short Circuit of Induction Motor's Stator Winding through Envelope Analysis

Syed Kamruddin Ahamed[1*], Arghya Sarkar[2], Madhuchhanda Mitra[3] and Samarjit Sengupta[4]

[1*] Govt. College of Engineering & Textile Technology,
Serampore, West Bengal, India,
[2] MCKV institute of Engineering, West Bengal University of Technology,
243, G. T. Road (N), Liluah, Howrah, West Bengal, India
[3, 4] Department of Applied Physics, University of Calcutta
*syedahamed.ahamed9@gmail.com.

Abstract— **In this paper, a new approach for detection of inter-turn short in the stator winding of an induction motor is presented. Discrete Wavelet Transform using DB4 is performed on the envelopes of the windowed steady-state current signatures. The envelopes are determined by applying Hilbert Transform. Low frequency oscillations below 50 Hz were extracted from the reconstructed details at higher wavelet levels. As the envelope works on narrow band frequencies or mono-component signal, it was analyzed using higher wavelet levels which belong to the narrow band harmonics. The RMS and Mean values of the reconstructed details and Power Detail Energy defined as PDE were used as fault parameters to detect the faulty motor from the healthy one. It has been observed that the faulty motor produces higher fault parameters than the healthy one. Laboratory test results confirm the validity of the proposed method.**

Index Terms— **inter-turn short, induction motor, Discrete wavelet transform, Hilbert Transform, envelope, RMS, Mean, Power Detail Energy.**

I. INTRODUCTION

In the present day industries, induction motor plays an important role as main drive. To reduce unscheduled downtime losses, to minimise occurrence of accidents, health monitoring of three phase induction motor has become very important. According to the survey, it has been observed that almost 40% of stator winding failures start as undetected turn to turn faults in a single phase due to destruction of insulation which finally grow and terminate in major faults [1]. In technical literature, several methods have been reported based on motor current signature analysis known as MCSA due to its easy availability and non-invasive nature. Motor current was analysed to search for characteristic (spatial) harmonics in the stator current to detect fault in the winding [2-3]. Apart from masking of the harmonics due to power frequency spectral leakage, similar types of harmonics may appear due to eccentricity related faults, supply unbalance, constructional imperfection [4]. Stator winding inter-turn fault has also been diagnosed by Negative sequence current [5] and negative sequence voltage[6], but they are not immune to unbalance of supply and inherent asymmetry of the motor. To overcome these difficulties, the negative sequence impedance has been used as fault indicator [7]. Park's vector's approach are also being applied for diagnosis of stator winding fault [8]. Advanced signal processing technique, wavelet transform are also producing very promising results for identifying motor faults [9].

The present paper proposed a novel approach for diagnosis of stator inter-turn short circuit fault of the induction motor through envelope analysis of windowed steady-state motor current. The proposed algorithm relies on efficient application of Hilbert Transform (HT) and Discrete Wavelet Transform (DWT). Windowing minimizes the edge effects that result in spectral leakage in the spectrum due to truncation of the selected portion of the signals. Selection of mother wavelet is not an easy task, if not proper, serious error may occur in the results which have been overcome by application of Hilbert Transform. Low frequency oscillations below 50 Hz were extracted from the reconstructed details at higher wavelet level of the envelope. As the envelope works on narrow band signals based on instantaneous frequencies, the reconstructed details at higher wavelet level were considerd. The method works with higher resolution and higher detectability due to simultaneous application of Hilbert transform and Wavelet transform.

II. THEORETICAL OVERVIEW

A. Stator Current envelope of an Induction Motor at the presence of inter-turn short of stator winding

In the investigation of *inter-turn short*, the stator current has been generally used. The faults affect the current according to motor slip

$$f_{stator} = \left\{ \frac{n}{p}(1-s) \pm k \right\} f \qquad (1)$$

where f is the supply frequency , p is the no. of pole pairs, s is the slip, n =1, 2, 3, --, k=1,3,5,----------, respectively.

The effects cause fluctuations on the current. The modulation of the stator current is the so-called envelope and that is cyclically repeated.

B. Hilbert Transform and Envelope Detection:

Hilbert Transform works on instantaneous frequency which is phase variability given by $f_i = (1/2\Pi)\, d\Phi/dt$

The Hilbert transform is defined as [10]

$$H\left[s(t)\right] = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{s(\tau)}{(t-\tau)} d\tau \qquad (2)$$

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

Using the mean value theorem, we can evaluate

$$H\left[s(t)\right] = \frac{1}{\pi t} \otimes s(t) \qquad (3)$$

Therefore, H[s(t)] is obtained from the convolution of the function 1/(Πt) with the original function s(t). By Hilbert transform, negative frequencies are eliminated where as positive frequencies are phase shifted by Π /2 and their amplitudes are being doubled

Envelope Detection: The envelope of the signal, s(t) is defined as modulus of analytic signal z(t), given by

$$E(t) = \left|s(t) + jH\left[s(t)\right]\right| \qquad (4)$$

where $$z(t) = \left|s(t) + jH\left[s(t)\right]\right| \qquad (5)$$

It is a new dimension in the area of fault detection from the spectrum analysis of envelope which is almost free from 50 Hz frequency.

C. Discrete Wavelet Transform:

Discrete wavelet transform (DWT) decomposes a signal by successively passing it through high pass and low pass filter producing detailed (d_n) and approximate coefficients (a_n) using multi resolution analysis known as MRA [11].
Fig. 1 shows two level wavelet decomposition. Each level of decomposition corresponds to certain spectral bands given in Table I. The decomposed signal can be reconstructed as shown in Fig. 2.

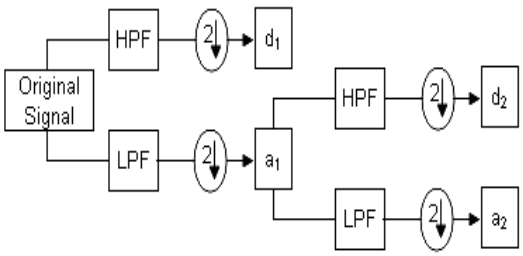


Fig. 1 Two Level multiresolution DWT decomposition

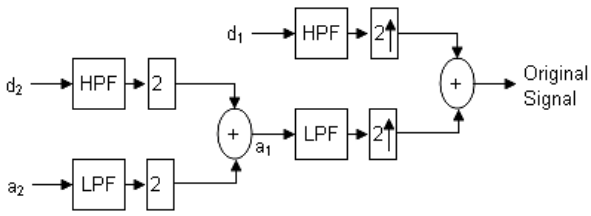


Fig. 2 Two Level multiresolution DWT reconstruction

Table 1 Spectral frequency bands at different decomposition levels

Decomposition Details	Frequency Bands(Hz)
Detail Level 1	512 - 1024
Detail Level 2	256 - 512
Detail Level 3	128 - 256
Detail Level 4	64 – 128
Detail Level 5	32 – 64
Detail Level 6	16 – 32
Detail Level 7	8 – 16
Detail Level 8	4 - 8

III. PROPOSED SCHEME

This paper presents a novel approach for detection of Induction motor stator inter-turn short circuit fault at no-load condition .The algorithm steps are summarised below:

Step 1: Use suitable window function to obtain specific size data vector from stator current of the Induction Motor. Window function is utilized to reduce the effect of transients.

Step 2: Perform Discrete Wavelet Transform in such a fashion that the reconstructed detailed coefficients at higher levels correspond to narrow band frequencies below 50 Hz.

Step 3: Compute fault parameters – RMS values, Mean values and Power Detailed Energy (PDE) of higher level wavelet coefficients. PDE is defined as

$$PDE = \sum_{1}^{N} (d_j \times d_j) \qquad (6)$$

where level of decomposition *j*, containing *N* number of samples .

Step 4: Compare the statistical parameters of the faulty motor with respect to the healthy one to obtain the information about the health of the Induction Motor.

Step 5: Go to Step 1.

This paper selects Hamming window of length 2048 (one sec. duration) samples. Then Discrete Wavelet Transform was performed using db4 mother wavelet which have the most compact support for a given number of vanishing moments, thus allowing the lowest computational loads. The spectral frequency bands at different decomposition levels are presented in Table1 for the sampling frequency 2048samples/sec.

IV. EXPERIMENTAL SET UP

Experiment was carried out on a test-rig built by Spectra Quest, USA, incorporated with a high speed data acquisition system (OROS OR35, 8 channels, 100 mbps). One induction motor with 3ph, 1HP, 415V, 50Hz, 1460 rpm with provisions for shorting from outside tapping was selected. The same motor can be used as healthy and inter-turn short (5% and 10%) in stator winding as faulty .The motors were run by direct-on-line supply at 330V, 50Hz through variac. Current signatures were captured using Hall Probe (LEM PR30 ACV 600V CATIII 30Ampac / 3Vac). The experimental set up and its schematic block diagram are shown in Fig. 3 and Fig. 4 respectively.

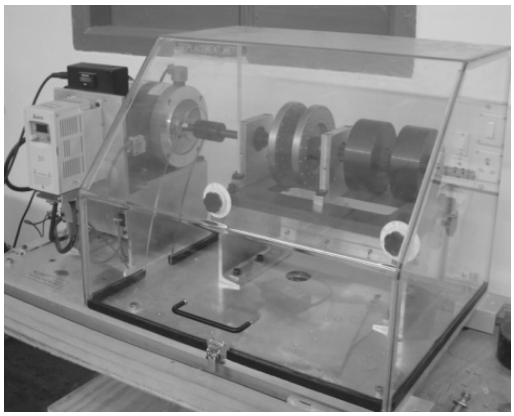


Fig. 3 Experimental setup

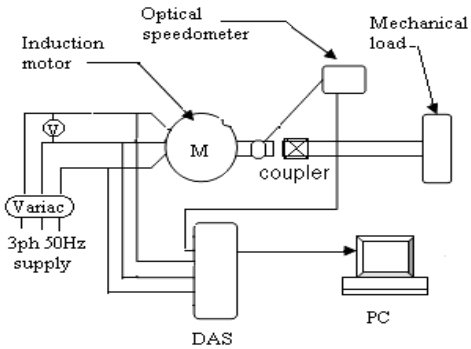


Fig. 4 Schematic block diagram of the experiment

V. Result Analysis

Discrete Wavelet Transform using db4 are performed on the the envelopes of windowed steady state signals (Sample size of duration of one sec.i.e.2048) at no-load. The signals and the envelopes for healthy, 5% and 10% inter-turn short conditions are shown in Fig. 5. Reconstructed details for the motor with the healthy and faulty conditions corresponding to 6th, 7th and 8th level below 50 Hz shown in Fig. 6, Fig.7and Fig. 8 are analysed. It is observed from the Table II that RMS, Mean values and PDE at all the three levels indicate considerable higher values for the motor with faulty condition than those for the motor with healthy condition. The most significant increases of these values are noticed at 8th level except the mean value at 7th level for 5%inter-turn short which is higher than the value at 8th level. The RMS values are 1.74 times and 6.61 times for 5% and 10% shorted turns with respect to the values for healthy condition whereas the rises in the Mean values are 1.57 and 6.22 times for 5% and 10% shorted turns. In the energy level analysis at 8th level, the increases in PDE are 3.10 and 42.84 times for 5% and 10% shorted turns with respect to the values for healthy condition.

Hence 8th level is considered most suitable for detection of stator inter-turn short.

The main advantages in this method are - the envelope analysis which makes diagnosis easier and cleaner as the power frequency is eliminated and the spectral leakage is minimized and the application of DWT on the steady state current which overcomes the difficulty of FFT analysis.

The main drawback is that once sampling frequency is selected, some ranges of frequencies remain unexplored. In the present analysis, it is (32 Hz - 50 Hz).

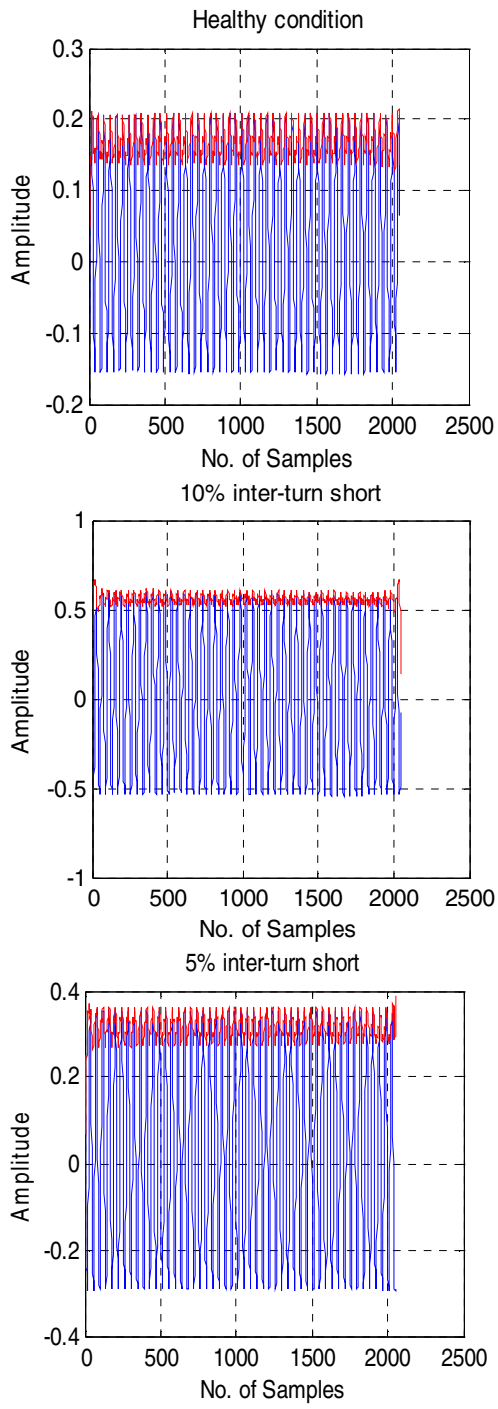


Fig. 5 Windowed Steady State Current for the (a) Healthy, (b) 10% inter-turn and (b) 5% inter-turn conditions at no load.

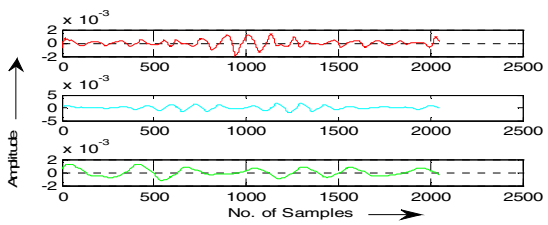


Fig. 6 The detailed coefficients corresponding) 6th ,7th and 8th level of envelope for motor at healthy condition at no load

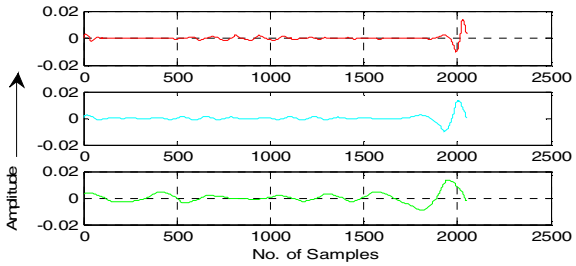


Fig. 7 The detailed coefficients corresponding) 8th ,9th and 10th level of envelope for motor at 10% at no load

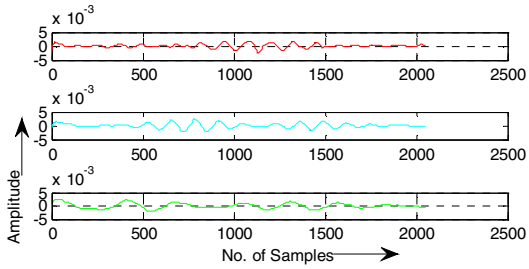


Fig. 8 The detailed coefficients corresponding to 6th ,7 th and 8th level of envelope for motor at 5% inter-turn short at no load

Table II Fault Parameters of the detailed coefficients of the Envelopes

Motor Condition	Wavelet Level	RMS value	Mean value	Power Detail Energy in joules
Healthy	6	.000496	.0000141	.000501
5% short		.000680	.0000211	.000947
10% short		.0019	.0000812	.0070
Healthy	7	.000687	.0000265	.000965
5% short		.000828	.0000378	.0014
10% short		.0024	..000148	.0115
Healthy	8	.000575	.0000409	.000677
5% short		.0010	.0000643	.0021
10% short		.0038	.0002545	.0290

VI. Conclusions

The present approach reveals that this method can successfully be employed for detection of stator winding inter-turn short. The method works with higher detectability and higher resolution due to DWT and Hilbert transform. The method is faster, less power consumption and hence less costly which makes it ideally suitable for industry

References

[1] A. Craciunescu, G. Ciumbulea, M. Media, "Stator Winding Fault Diagnostic of Induction Motor Using Instantaneous Currents' Space Phasor Approach", *Electrical Review*, ISSN 0033-2097, R. 88 NR 1a/2012.

[2] W.T.Thomson, M. Fenger, "Current signature analysis to detect induction motor Faults," *IEEE Ind. Appl. Magazine*, vol.7, no.4, pp.26-34, July/August 2001.

[3] G. M. Joksimovic and J. Penman (2000), "The detection of inter-turn short circuits in the stator windings of operating motors, *IEEE Transactions on Industrial Electronics*, vol.47, no.5, October, pp.1078-1084, 2000.

[4] S. Nandi, H. A. Toliyat, X. Li. , "Condition Monitoring and Fault Diagnosis of Electrical Motors— A Review", *IEEE Transactions on Energy Conversions*, vol.20, no.4, December, 2005.

[5] T. Kato, K. Inoue and K. Yoshida, "Diagnosis of Stator-Winding-Turn Faults of Induction Motor by Direct Detection of Negative Sequence Currents", *Electrical Engineering in Japan,* Volume 186, Issue 3, pages 75–84, February, 2014.

[6] R. M. Tallam, T. G. Habetler, R. G. Harley, "Stator winding Turn-fault detection for closed-loop induction motor drive," *IEEE Transactions Industry Applications*, vol. 39, no. 3, pp .720 –724, 2003.

[7] F. Villada , D.R. Cadavid , J.E. Aedo, B. A. Benavides, E. Velilla , " The negative impedance as fault indicator in induction motors " , *Rev. Fac. Ing. Univ. Antioquia*, no.52, pp.226-232, Marzo, 2010

[8] A. J. M. Cardoso, S. M. A. Cruz , and D. S. B. Fonseca, "Inter-turn stator winding fault diagnosis in three-phase induction motors, by Park's Vector Approach," *IEEE Trans. Energy Conversion*, vol. 14, pp. 595–598, Sept. 1999.

[9] J. Cusido, , L. Romeral , J.A. Ortega , J. A. Rosero , "Fault Detection in Induction Machines Using Power Spectral Density in Wavelet Decomposition", *IEEE Transaction on Industrial Electronics* , vol . 55 , no.2 , February, 2008.

[10] G. A. Jim´enez, A. O. Mu˜noz, M. A. Duarte-Mermoud "Fault detection in induction motors using Hilbert and Wavelet transforms" *Electrical Engineering*, vol. 89, pp. 205-220, 2007.

[11] J. Antonino – Daviu , M. Reira – Guasp , J. Roger-Folch , F. Martinez-Gimenez , and A. Peris, (2006), "Application and optimization of the discrete wavelet transform for the detection of broken rotor bars in induction machines " , *Appl. Comput. Harmon . Anal.*, **vol. 21**, no. 2 , pp .268 - 279 , Sep . 2006.

State Space Least Mean Square for State Estimation of Synchronous Motor

Arif Ahmed,[1,*] Muhammad Moinuddin,[2] and Ubaid M. Al-Saggaf[3]
[1,2,3]Centre of Excellence in Intelligent Engineering Systems
King Abdulaziz University, Jeddah, Saudi Arabia
[1,2,3] Department of Electrical & Computer Engineering
King Abdulaziz University, Jeddah, Saudi Arabia
*arif545@gmail.com

***Abstract*—Kalman filter and its variants are well known for the static and dynamic state estimation of power systems because of their accuracies. These adaptive filters generally employed for estimation purposes require high computational power when it comes to real time estimation. Therefore, in this paper we propose a computationally light yet effective estimation algorithm based on state space model which have not yet been applied to the problem of power system estimation. We propose the use of state space least mean square algorithms for the purpose of state estimation considering the problem of a two phase permanent magnet synchronous motor. The algorithms have been employed successfully in this paper in the state estimation of the highly non linear synchronous motor. We investigate the problem in the presence of Gaussian noise to show the novelty of the algorithms. Moreover, these algorithms are compared with the state estimation performance of the non linear Extended Kalman filter.**

***Index Terms*—SSLMS, SSNLMS, State Space Least Mean Square, State Space Normalized Least Mean Square, Power System Estimation, Synchronous Motor State Estimation**

I. Introduction

Adaptive filters are now widespread due to advancements in computational speed, complexity and power efficiency of digital processors. An adaptive filter adapts by self adjusting filter parameters according to the optimization algorithm utilized ([1],[2]). Two of the most widely used adaptive filtering algorithms are the least mean squares (LMS) algorithm ([1],[2],[3]) and the recursive least squares (RLS) algorithm ([1],[2],[3]). The state space (SS) version of these algorithms have been developed and presented by Malik et al. ([4],[5],[6],[7]) along with different analyses.

One of the important uses of adaptive filtering algorithms is in the area of power systems. State estimation has become an essential part of power system ([8],[9],[10],etc.) which requires the use of adaptive filtering algorithms like the Kalman filter, its variants and other algorithms ([11],[12],[13],etc.). However, these model based non linear adaptive filtering algorithms are highly computationally complex and requires more number of computations per iterations. Moreover, they don't always have the upper hand as will be presented in this paper. In this paper we deal with the state estimation of a permanent magnet synchronous motor. For effective control of the motor it is essential to know as best representation as possible of the true states of the motor. Which demands effective and efficient estimation algorithms giving us the cost benefit of having sensors. The synchronous motor is a highly non linear system and it is common practise to use non linear Kalman filtering algorithms like Extended Kalman filter (EKF), Unscented Kalman filter (UKF) ([12],[13],[14],etc.). The proposed state space least mean square (SSLMS) and the state space normalized least mean square (SSNLMS) by Malik et al. ([4],[5]) have only been applied to linear time varying systems. In this paper we propose and implement these algorithms for non linear synchronous motor state estimation and compare their performances with EKF. The proposed algorithms are superior due to the fact that they require lesser computations per iteration compared to the existing non linear model based algorithms. The performance of the algorithms in presence of Gaussian noise is investigated.

The paper is organized as such, Section II of the paper introduces the SS model along with the two phase permanent magnet synchronous magnet model. Then in Section III the overview of the existing SSLMS and SSLNMS is presented following Section IV presenting simulation results and discussion of the comparison of and proposed algorithms. And finally we conclude the paper in Section V.

II. State-Space Model

We begin by defining the general state-space model of a linear time varying system.

$$\mathbf{x}[k+1] = \mathbf{A}[k]\mathbf{x}[k] + \mathbf{B}[k]\mathbf{u}[k] + \mathbf{w}[k], \tag{1a}$$
$$\mathbf{y}[k] = \mathbf{C}[k]\mathbf{x}[k] + \mathbf{D}[k]\mathbf{u}[k] + \mathbf{v}[k] \tag{1b}$$

where $\mathbf{x} \in \Re^n$ are the process states, $\mathbf{y} \in \Re^m$ are the measured outputs such that $m \leq n$. $\mathbf{A}[k]$ is the state transition matrix, $\mathbf{B}[k]$ is the input matrix, $\mathbf{u}[k]$ is the input vector where $\mathbf{u} \in \Re^p$, $\mathbf{w} \in \Re^n$ is the process noise vector and $\mathbf{v} \in \Re^m$ is the measurement noise vector. The matrix $\mathbf{C}[k]$ is the output matrix where $\mathbf{dim}[\mathbf{C}[k]] = m \times n$, $\mathbf{D}[k]$ is the feed through matrix with $\mathbf{dim}[\mathbf{D}[k]] = m \times p$. It is assumed that the above system is observable. A special case is the unforced (autonomous) linear time varying system, represented as

$$\mathbf{x}[k+1] = \mathbf{A}[k]\mathbf{x}[k] + \mathbf{w}[k], \tag{2a}$$
$$\mathbf{y}[k] = \mathbf{C}[k]\mathbf{x}[k] + \mathbf{v}[k] \tag{2b}$$

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

The state space representation for a non linear continuous time system is

$$\dot{\mathbf{x}} = f(\mathbf{x}, \mathbf{u}, \mathbf{w}), \tag{3a}$$
$$\mathbf{y} = h(\mathbf{x}, \mathbf{u}, \mathbf{v}) \tag{3b}$$

where f and h are non linear functions and the parameters are as defined before. The two phase permanent magnet synchronous motor can be represented as ([14],[15])

$$\dot{i}_a = \frac{-R}{L} i_a + \frac{\omega\lambda}{L}\sin\theta + \frac{u_a}{L} \tag{4a}$$
$$\dot{i}_b = \frac{-R}{L} i_b + \frac{\omega\lambda}{L}\cos\theta + \frac{u_b}{L} \tag{4b}$$
$$\dot{\omega} = \frac{-3\lambda}{2J} i_a + \frac{3\lambda}{2J} i_b \cos\theta - \frac{F}{J}\omega - \frac{T_L}{J} \tag{4c}$$
$$\dot{\theta} = \omega \tag{4d}$$

where i_a and i_b are the current through the windings, R is the motor winding resistance, L is the motor winding inductance. θ is the angular position of the rotor, λ is the flux constant of the rotor, and ω is the angular velocity of the rotor. u_a and u_b are the voltages applied across the two windings, F is the viscous friction of the rotor, J is the moment of inertia of the rotor shaft and its load, and T_L is the load torque.[1]

III. Overview of the SSNLMS and SSLMS

Considering the system described by equation (2). A model based adaptive estimation process can be divided into the following two steps. Step 1, the time update which is given by

$$\bar{\mathbf{x}}[k] = \mathbf{A}[k-1]\hat{\mathbf{x}}[k-1] \tag{5}$$

Step 2, the measurement update which is given by

$$\hat{\mathbf{x}}[k] = \bar{\mathbf{x}}[k] + \mathbf{K}[k]\varepsilon[k] \tag{6}$$

where

$$\varepsilon[k] = \mathbf{y}[k] - \bar{\mathbf{y}}[k] \tag{7}$$

here $\mathbf{y}[k]$ is as mentioned in (2), **K**[k] is the gain matrix and

$$\bar{\mathbf{y}}[k] = \mathbf{C}[k]\bar{\mathbf{x}}[k] \tag{8}$$

This is basically the structure employed in all Kalman Filtering techniques ([16],[17],[18],[19], etc.). In the case of SSNLMS [4] the gain matrix $\mathbf{K}[k]$ is derived as follows

$$\mathbf{e}[k] = \mathbf{y}[k] - \hat{\mathbf{y}}[k] = \varepsilon[k] - \mathbf{C}[k]\delta[k] \tag{9}$$

where

$$\delta[k] = \hat{\mathbf{x}}[k] - \bar{\mathbf{x}}[k] \tag{10}$$

Assuming $\mathbf{C}[k]$ is full rank, $\hat{\mathbf{x}}[k]$ is chosen such that $\mathbf{e}[k] = 0$ which implies the following

$$\varepsilon[k] = \mathbf{C}[k]\delta[k] \tag{11}$$

[1]The system has been discretized and linearized for the simulation and implementation of the estimation algorithms which can be referred to in the appendix.

$\delta[k]$ is chosen as the minimum norm solution of the above equation which results in

$$\delta[k] = \mathbf{C}^T[k](\mathbf{C}[k]\mathbf{C}^T[k])^{-1}\varepsilon[k] \tag{12}$$

here

$$\mathbf{K}[k] = \mathbf{C}^T[k](\mathbf{C}[k]\mathbf{C}^T[k])^{-1} \tag{13}$$

The estimator equation from (10) and (12) is hence derived as

$$\hat{\mathbf{x}}[k] = \bar{\mathbf{x}}[k] + \mathbf{C}^T[k](\mathbf{C}[k]\mathbf{C}^T[k])^{-1}\varepsilon[k] \tag{14}$$

This is termed in ([4],[5]) as the SSNLMS algorithm. However, to avoid the non invertible scenario of $(\mathbf{C}[k]\mathbf{C}^T[k])^{-1}$, a small value γ might be added to the term and hence the estimator can be represented as

$$\hat{\mathbf{x}}[k] = \bar{\mathbf{x}}[k] + \mathbf{C}^T[k](\gamma + \mathbf{C}[k]\mathbf{C}^T[k])^{-1}\varepsilon[k] \tag{15}$$

This is termed in [4] as the SSNLMS algorithm. However, the above estimator may be unstable and there are no known conditions for stability and hence convergence is not guaranteed. For this reason a $\mathbf{G}$ matrix has been introduced in [4] to overcome this problem. In [4], it is claimed that the choice of this matrix $\mathbf{G}$ depends on the nature of the problem and one simple approach is to take $\mathbf{G}$ to be all zeros except for the first column to have non zero entries. In our investigation, we will consider the SSNLMS and also the SSLMS algorithm [4] defined as

$$\hat{\mathbf{x}}[k] = \bar{\mathbf{x}}[k] + \mu\mathbf{G}\mathbf{C}^T[k]\varepsilon[k] \tag{16}$$

here

$$\mathbf{K}[k] = \mu\mathbf{G}\mathbf{C}^T[k] \tag{17}$$

IV. Simulation Results & Discussion

Simulation results are presented here to validate the performance of the algorithms. We perform the estimation of the state parameters of the synchronous motor in the presence of Gaussian noise. For the EKF, the process and observation error covariance matrices were chosen to be similar to the noise covariances. An overview of the root mean square error (RMSE) of the states and the observations can be referred to in Table I. SSNLMS, SSLMS, and EKF were compared. The actual system is discretized at a sampling rate of 0.0002s but the algorithms sample at 0.001s to mimic the effect of real time scenario where sensors are limited by sampling rate. μ for SSLMS and SSNLMS were chosen as 0.2 and 0.4 respectively. The observation was subject to Gaussian noise of covariance $\sigma^2 = 0.01^2$ and the process noise was considered to be of covariance $\sigma^2 = 0.001^2$. The actual initial system states were considered to be $\mathbf{x}[0] = [1\ 3\ 0.5\ 2]^T$, and the initial estimate for the algorithms were chosen to be $\hat{\mathbf{x}}[0] = [0\ 0\ 0\ 0]^T$. The synchronous machine parameters can be referred to in [14] The plot of the states and observations are presented in Figure 1 and Figure 2 respectively. The square of the observation error is presented in Figure 3. It can be observed from Figure 1 that the EKF has lower error in the beginning, however, as time proceeds the SSLMS and SSNLMS converges faster than the EKF while the error of

the EKF remains higher.Moreover, the EKF performance is not acceptable when position and velocity plots are observed. A clear picture regarding the performance of the algorithms can be observed from the observation error square plots in Figure 3. It can be clearly seen that both the SSLMS and the SSNLMS converges faster and hence performs better in synchronous motor state estimation.

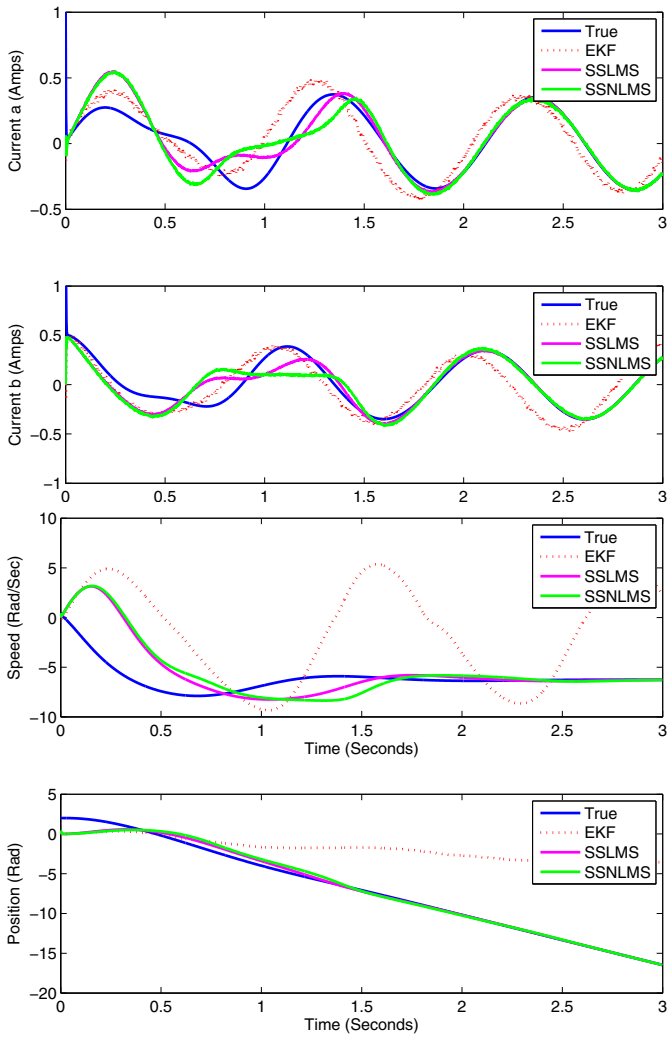


Fig. 1: State $x_1[k]$, $x_2[k]$, $x_3[k]$, and $x_4[k]$

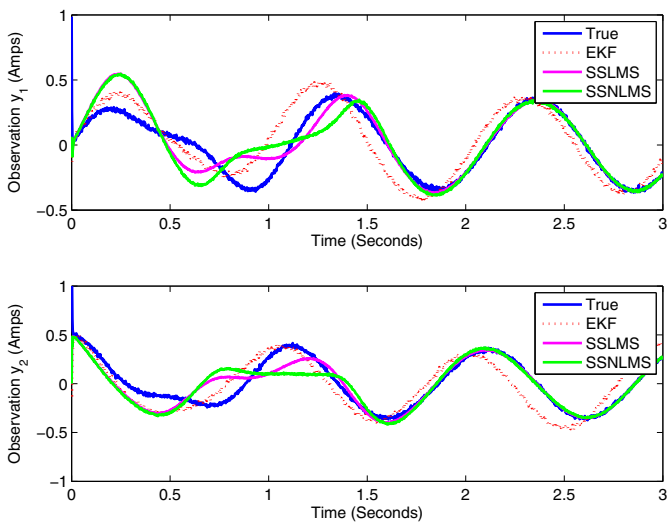


Fig. 2: Observation $y_1[k]$ and $y_2[k]$

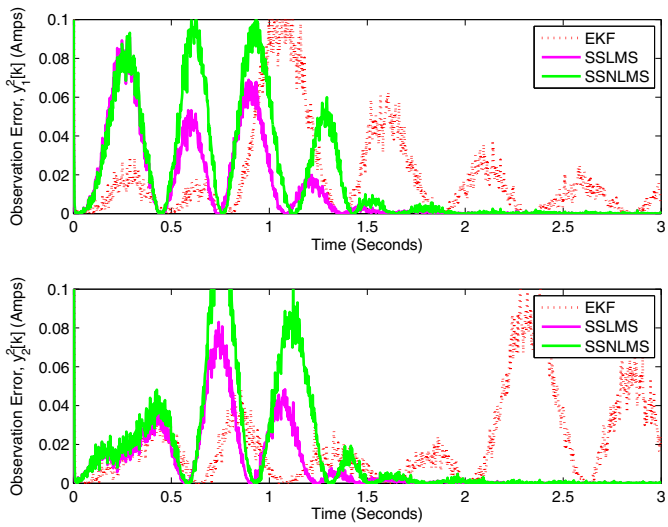


Fig. 3: Observation Error, $y_1^2[k]$ and $y_2^2[k]$

The Table presented below gives a much clearer representation of the performance of the algorithms. The root mean square error of both of the proposed algorithms remain much lower compared to the EKF and therefore these algorithms can very well be implemented and considered as novel state estimation algorithms in power system considering the fact that they have extremely low computational load.

TABLE I: Root Mean Square Error

	Root Mean Square Error		
	EKF	***SSLMS***	***SSNLMS***
State x_1	0.1389	0.1134	0.1392
State x_2	0.1644	0.1261	0.1506
State x_3	6.6513	2.2112	2.4001
State x_4	6.7208	0.5750	0.6740
Observation y_1	0.1391	0.1139	0.1397
Observation y_2	0.1644	0.1265	0.1509

V. Conclusion

In this paper the SSLMS and the SSNLMS algorithms have been implemented for the state estimation of a non linear synchronous motor. More recently, a lot of model based non linear Kalman filtering algorithms have been proposed for such estimation problems. However, it is very well known that these algorithms are complex and require high computational power in real time estimation. Therefore, we have suggested and implemented two state space least mean square algorithms which we have found to be effective by a very good margin. It is of much importance that the algorithms be computationally light which these algorithms are and therefore we claim them to be novel algorithms for non linear power system state estimation problems. We have investigated the performance of these two algorithms and compared them with EKF and found great results which is demonstrated by the root mean square error. A two phase permanent magnet synchronous motor has been investigated here in the presence of Gaussian noise. For further studies different variants of the SSLMS and SSNLMS can be developed and implemented for investigation along with time varying step size parameters for yet better results.

VI. Appendix

For the simulation of the system model, it is essential that the system be discretized. Considering the synchronous motor system described by the set of equations (4), the system is discretized as

$$\mathbf{x}[k+1] = f(\mathbf{x}[k], \mathbf{u}[k], \mathbf{w}[k]) = \mathbf{x}[k] + \begin{bmatrix} \frac{-R}{L}x(1) + \frac{\lambda}{L}x(3)\sin x(4) + \frac{u_a}{L} \\ \frac{-R}{L}x(2) + \frac{\lambda}{L}x(3)\cos x(4) + \frac{u_b}{L} \\ \frac{-3\lambda}{2J}x(1) + \frac{3\lambda}{2J}x(2)\cos x(4) - \frac{F}{J}x(3) - \frac{T_L}{J} \\ x(3) \end{bmatrix} \Delta t + \mathbf{w}[k] \tag{18}$$

$$\begin{aligned} \mathbf{y}[k] &= h(\mathbf{x}[k], \mathbf{u}[k], \mathbf{v}[k]) \\ &= \begin{bmatrix} x(1) \\ x(2) \end{bmatrix} \Delta t + \mathbf{v}[k] \end{aligned} \tag{19}$$

For the implementation of the algorithms the system was linearized which resulted in

$$\mathbf{A}[k] = f'(\hat{\mathbf{x}}[k], \mathbf{u}[k]) = \begin{bmatrix} -\frac{R}{L} & 0 & \frac{\lambda \sin \hat{x}(4)}{L} & \frac{\hat{x}(3)\lambda \cos \hat{x}(4)}{L} \\ 0 & -\frac{R}{L} & -\frac{\lambda \cos \hat{x}(4)}{L} & \frac{\hat{x}(3)\lambda \sin \hat{x}(4)}{L} \\ -\frac{3\lambda \sin \hat{x}(4)}{2J} & \frac{3\lambda \cos \hat{x}(4)}{2J} & -\frac{F}{J} & -\frac{3\lambda[\hat{x}(1)\cos \hat{x}(4) + \hat{x}(2)\sin \hat{x}(4)]}{2J} \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{20}$$

$$\begin{aligned} \mathbf{C}[k] &= h'(\hat{\mathbf{x}}[k]) \\ &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \end{aligned} \tag{21}$$

Acknowledgment

The authors acknowledge the support of King Abdulaziz University and Centre of Excellence in Intelligent Engineering Systems, King Abdulaziz University for carrying out this work.

References

[1] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, 3rd Edition, Upper-Saddle River, NJ, 1996.

[2] A. H. Sayed, *Fundamentals of Adaptive Filtering*, New York: Wiley-Interscience, 2003.

[3] B. Widrow, J. M. McCool, M. G. Larimore, and C. R. Johnson, "Stationary and Nonstationary Learning Characteristics of the LMS Adaptive Filter," *IEEE Proceedings*, Vol. 64, pp. 1151–1162, 1976.

[4] M.B. Malik, M. Salman, "State-space least mean square", Digital Signal Processing, Vol. 18, Issue 3, pp. 334-345, May 2008.

[5] M.B. Malik, Bhatti, R.A, "Tracking of linear time-varying systems using state-space least mean square," IEEE International Symposium on Communications and Information Technology, Vol. 1, pp. 582-585, Oct. 2004.

[6] M.B. Malik "State-Space Recursive Least Squares: part I," Signal Processing, vol. 84/9, 2004, pp 1709-1718.

[7] M.B. Malik "State-Space Recursive Least Squares: part II," Signal Processing, vol. 84/9, 2004, pp 1719-1728.

[8] F. F. Wu, "Power system state estimation: A survey," Int. J. Electr. Power Energy Syst., vol. 12, no. 2, pp. 80-87, 1990.

[9] A. Monticelli, "Eectric power system state estimation," Proc. IEEE, vol. 88, no. 2, pp. 262-282. 2000.

[10] A. Abur,A. G. Exposito, "Power System State Estimation: Theory and Implementation," CRC Press, 1st edition, Mar. 2004.

[11] G. Valverde and V. Terzija, "Unscented Kalman filter for power system dynamic state estimation," IET Gen., Transm., Distrib., vol. 5, no. 1, pp. 29 - 37, 2011.

[12] Ghahremani, E., Kamwa, I., "Dynamic State Estimation in Power System by Applying the Extended Kalman Filter With Unknown Inputs to Phasor Measurements," IEEE Transactions on Power Systems, pp. 2556 - 2566, Nov. 2011.

[13] Ghahremani, E., Kamwa, I., "Online State Estimation of a Synchronous Generator using Unscented Kalman Filter From Phasor Measurements Units," IEEE Transactions on Energy Conversion, pp. 1099 - 1108, Dec. 2011.

[14] D. Simon, "Using Nonlinear Kalman Filtering to Estimate Signals," Embedded Systems Design, vol. 19, no. 7, pp. 38-53, July 2006.

[15] Rached Dhaouadi, "Design and implementation of an Extended Kalman Filter fot the state estimation of a permanent magnet synchronous motor," IEEE Transactiona on Power Electronics, vol. 6, no. 3, July 1991.

[16] C. K. Chui and G. Chen, *Kalman Filtering With Real-Time Applications*, Springer, 4th Edition, 2009.

[17] Wan, E. A., and van der Merwe, R., "The unscented Kalman filter for nonlinear estimation," IEEE Proceedings of Symposium on Adaptive Systems for Signal Processing, Communication and Control, Lake Louise, Alberta, Canada, Oct. 2000.

[18] Arasaratnam, I., Haykin, Simon "Cubature Kalman Filters", Automatic Control, IEEE Transactions on, On page(s): 1254 - 1269 Volume: 54, Issue: 6, June 2009.

[19] I. Arasaratnam , S. Haykin and R. J. Elliott "Discrete-time nonlinear filteringalgorithms using Gauss-Hermite quadrature", Proc.IEEE, vol. 95, no. 5, pp.953 - 977, 2007.

Maximum Power Delivery Capability of a Three-Phase Self-Excited Induction Generator

M. H. Haque
School of Engineering, University of South Australia, Australia
Mohammed.Haque@unisa.edu.au

Abstract—**Self-excited induction generators (SEIG) are increasingly being used in remote areas to energize isolated communities. The maximum power delivery capability of a SEIG depends on various factors, such as excitation capacitance, generator parameters, operating speed and load power factor. This paper proposes a simple method of determining the maximum power that can be delivered by a three-phase SEIG for various operating conditions. The maximum power of the generator is carefully formulated as a multi-variable constrained optimization problem. The problem is then solved using a numerical based routine "*fmincon*" given in the optimization toolbox of MATLAB. The effectiveness of the proposed method is then evaluated on a 1.5-kW, 220-V induction machine operated as a SEIG. The simulation results obtained by the proposed method are also compared with the corresponding experimental values.**

***Index Terms*--excitation capacitance, induction generator, maximum power, rural electrification, self-excited induction generator, stand-alone operation.**

I. Introduction

About one-fifth of world populations do not have access to grid electricity because they live in remote areas where extension of the grid is not economically viable. However, vast amount of renewable energy sources is usually available in remote areas. One of the possibilities of energizing remote communities is by using small-scale stand-alone generating systems based on renewable energy sources or in combination of renewable and fossil fuel sources.

Induction generators are increasing being used in small power plants driven by renewable energy sources, such as wind and mini-hydro. An induction generator has many advantageous features over its counterpart synchronous generator such as lower cost, maintenance and operational simplicity, brushless and rugged construction, lack of dc excitation, etc. [1-2]. Unlike a synchronous generator, a squirrel-cage induction generator cannot deliver reactive power. In fact, it absorbs reactive power for excitation. For stand-alone operation of an induction generator, the reactive power requirement of the generator as well as the load must be supplied by local capacitances. Such a generator is called self-excited induction generator (SEIG). The characteristics of a SEIG are highly dependent on the value of the capacitance used [3-6].

The per-phase steady-state equivalent circuit of a SEIG is usually used to determine its performance characteristics [2]. The circuit is normally analyzed either by loop impedance approach or nodal admittance approach [3-9]. References [10-13] determined the minimum capacitance required to build-up voltage in a SEIG under no-load condition. However, a SEIG has very poor voltage regulation and thus its voltage decreases rapidly with load [2, 14]. References [8, 15] determined the capacitance required in a SEIG under load conditions. In order to maintain a constant terminal voltage, the capacitance needs to be adjusted continuously with load [16, 17]. Alternatively, the terminal voltage can be regulated within a pre-specified upper and lower limit using discrete capacitances [18, 19].

For a given capacitance, the load characteristic (terminal voltage versus power) of a SEIG is very similar to the well-known *P-V* curve of a load bus in a power system [20]. In other words, there is a maximum power that can be delivered by the generator. The maximum power point is also called the critical point and it depends on the value of the capacitance used [3]. Reference [21] determined the capacitance as well as the maximum power point of a wind driven three-phase SEIG. The method mainly determined the maximum power of the wind turbine but not the generator. Mahato et al [22] determined the capacitance required to obtain the maximum output power for single-phase operation of a three-phase induction generator. In general, the terminal voltage at the critical point is much lower than the nominal value. In order to guarantee a minimum pre-specified terminal voltage, the generator should be operated away from the critical point. Thus, it is very important to determine the critical power as well as the maximum power that can be delivered while maintaining a pre-specified minimum terminal voltage for successful operation of the generator.

This paper proposes a method of determining the maximum power point of a SEIG for various operating conditions. An optimization problem is formulated to determine the maximum power point while satisfying the fundamental equations of the generator as well as a pre-specified minimum voltage criterion. The effectiveness of the proposed method is then tested on a 1.5-kW induction machine operated as a SEIG. The simulation results obtained by the proposed method are also compared with the corresponding actual values found through an experimental setup.

II. Generator Model and Characteristics

The per-phase steady-state equivalent circuit of a SEIG including its excitation capacitance and load impedance is shown in Fig. 1. The stator resistance, stator leakage reactance, rotor resistance, rotor leakage reactance, core loss resistance and magnetizing reactance of the generator are represented by R_1, X_1, R_2, X_2, R_c and X_m, respectively. The load impedance and the excitation capacitance reactance are represented by $\overline{Z}_L = Z_L\angle\theta$ (= $R_L + jX_L$) and X_C, respectively. The per unit (pu) generator speed and frequency are represented by v and F, respectively. All reactances are specified at the base frequency. Note that a SEIG may not generate voltage at rated frequency and thus the circuit of Fig. 1 is normalized to the base

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

frequency by dividing all parameters and voltages by the pu frequency F [2].

All parameters of the generator, except the magnetizing reactance X_m, are considered as constant. The value of X_m depends on magnetic saturation and hence air-gap voltage to frequency ratio (V_g/F). There are several ways of expressing V_g/F in terms of X_m [2, 8, 16, 21]. In this study, V_g/F is expressed by the following third-order polynomial of X_m and is used in some previous studies [3, 18]

$$\frac{V_g}{F} = k_1 + k_2 X_m + k_3 X_m^2 + k_4 X_m^3 \tag{1}$$

The coefficients k's of the polynomial can be determined from synchronous speed test data [3].

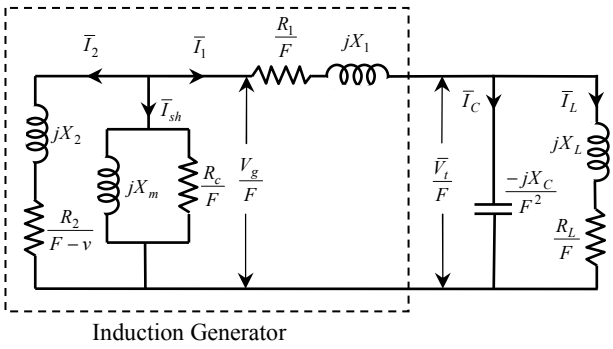


Fig. 1 Per-phase steady-state equivalent circuit of a SEIG

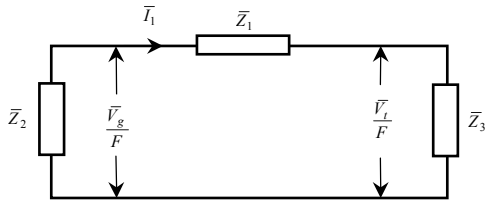


Fig. 2 Simplified representation of Fig. 1

The circuit of Fig. 1 can be represented by three series impedances ($\bar{Z}_1$, $\bar{Z}_2$ and $\bar{Z}_3$) as shown in Fig. 2. The impedances are given by

$$\bar{Z}_1 = (R_1 / F + jX_1) \tag{2.1}$$

$$\bar{Z}_2 = \left(\frac{1}{R_c / F} + \frac{1}{jX_m} + \frac{1}{R_2 /(F - v) + jX_2} \right)^{-1} \tag{2.2}$$

$$\bar{Z}_3 = \left(\frac{1}{-jX_C / F^2} + \frac{1}{R_L / F + jX_L} \right)^{-1} \tag{2.3}$$

The loop equation in Fig. 2 can be written as

$$\bar{I}_1 \left(\bar{Z}_1 + \bar{Z}_2 + \bar{Z}_3 \right) = 0 \tag{3}$$

Under normal operating condition, the stator current $\bar{I}_1$ is not zero and thus the sum of impedances must be zero.

$$\bar{Z}_1 + \bar{Z}_2 + \bar{Z}_3 = 0 \tag{4}$$

By separating the real and imaginary parts, the above phasor equation can be split into two scalar equations as follows

$$g_1 = real(\bar{Z}_1 + \bar{Z}_2 + \bar{Z}_3) = 0 \tag{5}$$

$$g_2 = imag(\bar{Z}_1 + \bar{Z}_2 + \bar{Z}_3) = 0 \tag{6}$$

Equations (5) and (6) represent the fundamental equations of a SEIG and must be satisfied for all operating conditions. Note that, for a given load power factor, the circuit of Fig. 1 has five unknowns (X_m, F, v, X_C and Z_L) but only two independent equations can be formulated [3]. Equations (5) and (6) are usually solved through a suitable iterative technique to find the values of two unknowns (X_m and F) for some assumed values of rest three unknowns (v, X_C and Z_L). By knowing the values of X_m and F, the air-gap voltage V_g can be evaluated through (1). Once V_g is known, the voltage, current and power in various parts of the circuit of Fig. 1 can easily be evaluated using standard circuit theories.

The voltage versus power curve of a SEIG can be obtained by repeatedly solving (5) and (6) in conjunction with (1) for various values of load impedance and assumed values of excitation capacitance and speed. A typical power characteristic of a SEIG is shown in Fig. 3. At no-load (when Z_L = infinity), P = 0 and the corresponding voltage is called the no-load voltage V_{NL}. When the load impedance is decreased from infinity, power increases and voltage decreases as can be seen in Fig. 3 (black line) and it represents the normal or stable operation of the generator. The above process continues until the maximum power point (marked by 'o') is reached. Further reduction of load impedance decreases both power and voltage as can be seen in Fig. 3 (red line). This is an abnormal operation and is called unstable operation. The maximum power point (marked by 'o') is also called the critical point. The power and the voltage at the critical point are represented by P_{cr} and V_{cr}, respectively.

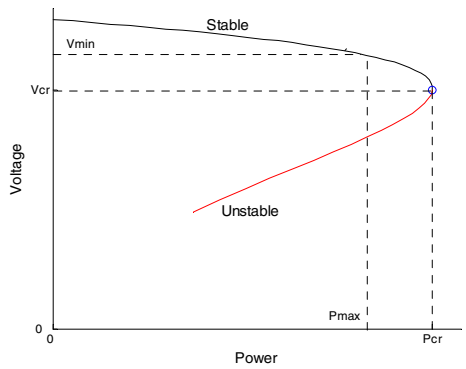


Fig. 3 A typical power characteristic of a SEIG

At the critical point, the load voltage may be far below the nominal value. When a minimum pre-specified voltage of V_{min} is maintained, the generator can deliver a maximum power of P_{max} (instead of P_{cr}) as can be seen in Fig. 3. The difference between P_{cr} and P_{max} can be considered as a power margin and is an indication of how far the operating point P_{max} is form the critical point.

III. Problem Formulation

As mentioned, the air-gap voltage V_g can be determined by solving (5) and (6) in conjunction with (1). Once V_g is known, the terminal voltage V_t in Fig. 2 can be written as

$$\bar{V}_t = \bar{V}_g \times \frac{\bar{Z}_3}{\bar{Z}_1 + \bar{Z}_3} \tag{7}$$

In order to guarantee a minimum terminal voltage of V_{min}, the following criterion must be satisfied

$$V_{\min} - \left| \bar{V}_g \times \frac{\bar{Z}_3}{\bar{Z}_1 + \bar{Z}_3} \right| \le 0 \tag{8}$$

The load current $\bar{I}_L$ in Fig. 1 can be written as

$$\bar{I}_L = \frac{\bar{V}_t}{R_L + jFX_L} = \bar{Y}_L \bar{V}_t \tag{9}$$

Here $\overline{Y}_L = 1/(R_L + jFX_L)$ is the load admittance. The active power P supplied by the generator can be written as

$$P = 3\left|\overline{I}_L\right|^2 R_L = 3\left|\overline{V}_g \times \frac{\overline{Z}_3}{\overline{Z}_1 + \overline{Z}_3} \times \overline{Y}_L\right|^2 R_L \qquad (10)$$

Determination of critical power P_{cr} involves the optimization of (10) while satisfying the fundamental equations (5) and (6). This can be achieved by varying the load impedance Z_L (or admittance Y_L) in the optimization process. In order to obtain the maximum power P_{max} while maintaining a minimum voltage of V_{min} (> V_{cr}), the inequality constraint (8) should also be included in the optimization process. Thus, for given excitation capacitance and speed, the optimization problem can be formulated as

Minimize $$f(\mathbf{X}) = -3\left|\overline{V}_g \times \frac{\overline{Z}_3}{\overline{Z}_1 + \overline{Z}_3} \times \overline{Y}_L\right|^2 R_L \qquad (11)$$

Subject to $$real(\overline{Z}_1 + \overline{Z}_2 + \overline{Z}_3) = 0 \qquad (12.1)$$

$$imag(\overline{Z}_1 + \overline{Z}_2 + \overline{Z}_3) = 0 \qquad (12.2)$$

$$V_{\min} - \left|\overline{V}_g \times \frac{\overline{Z}_3}{\overline{Z}_1 + \overline{Z}_3}\right| \le 0 \qquad (12.3)$$

Here $\mathbf{X} = [X_m \;\; F \;\; Y_L]^T$. The above problem can be solved using a numerical based optimization routine "*fmincon*" given in the optimization toolbox of MATLAB [24]. It may be mentioned here that when V_{min} in (12.3) is set to a lower value (< V_{cr}), the optimization process would provide the critical power P_{cr} (as described in Fig. 3). However, when V_{min} in (12.3) is set to a higher value (in between V_{NL} and V_{cr}), the optimization process would provide the maximum power P_{max} that occurs at V_{min} (as described in Fig. 3).

IV. Results and Discussions

The proposed method of determining the maximum power point of a SEIG was tested on a 1.5-kW, 220-V, 50-Hz, 4-pole, three-phase induction machine operated as a generator. The parameters of the generator are given in [3]. In the laboratory, the machine was driven at a constant speed of 1.0 pu by a synchronous motor. That is, the induced voltage under load condition would be slightly less than 50 Hz. The results found in the above machine are briefly described in the following.

First, the critical power (P_{cr}) at unity power factor are determined through (11) and (12) for various values of excitation capacitance. The results found through the optimization routine are shown in Fig. 4 by a solid line. The experimental results recorded in the laboratory are also shown in the figure by a symbol 'o'. It can be noticed in Fig. 4 that the critical power P_{cr} increases with the increase in capacitance. For C = 30 μF, the critical power is found as only 409 W (27.3% of rated value) and it increased to the rated value of 1500 W for a capacitance of about 50 μF.

The maximum power P_{max} that can be delivered by the generator while maintaining the rated voltage of 220 V is also determined through (11) and (12) by setting V_{min} = 220 V in (12.3). The results obtained in this case are shown in Fig. 5 and it indicates that the maximum power increases with the increase in capacitance but its value is lower than that found at the critical point (see Fig. 4) for the same value of capacitance.

In practice, the load voltage may not be kept constant at the rated value. The voltage may be allowed to drop to an acceptable minimum value (say 95% or 90% of rated value). With this in mind, the maximum power P_{max} that can be delivered by the generator while maintaining a minimum load voltage of 209 V (95% of rated value) or 196 V (90% of rated value) is also determined through (11) and (12) and the results found are shown in Fig. 6. For comparison purpose, the critical power is also shown in the figure. It can be noticed in Fig. 6 that the maximum power P_{max} increases with the increase in capacitance for all cases, as expected. For a given capacitance, the maximum power decreases as the specified load voltage is increased because the generator operates further away from the critical point.

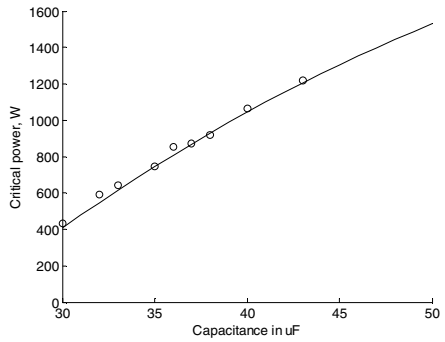


Fig. 4 Variation of critical power against capacitance.
'—' simulation results; 'o' experimental results

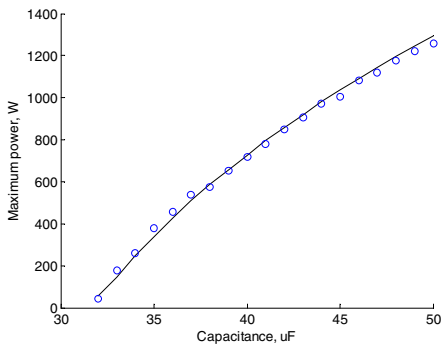


Fig. 5 Variation of P_{max} against capacitance at rated voltage (220V).
'—' simulation results; 'o' experimental results

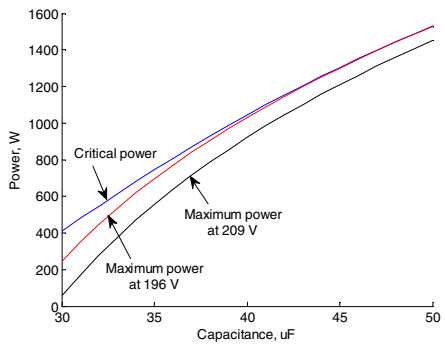


Fig. 6 Variation of critical power and maximum against capacitance.

Figure 7 shows the variation of critical power P_{cr} against capacitance for various load power factors. As mentioned, the capacitance in a SEIG supplies reactive power to both the generator and the load. For a lagging power factor, less reactive power is available for the generator excitation because of the load demand and thus the critical power decreases when compared with the unity power factor case. Finally, the variation of critical power against capacitance, for various operating speeds, is shown in Fig. 8 and it indicates that the critical power increases with the increase of speed.

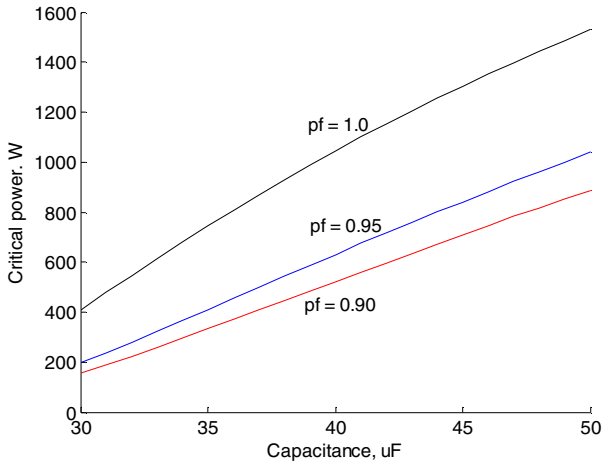


Fig. 7 Variation of P_{cr} against capacitance for various power factors.

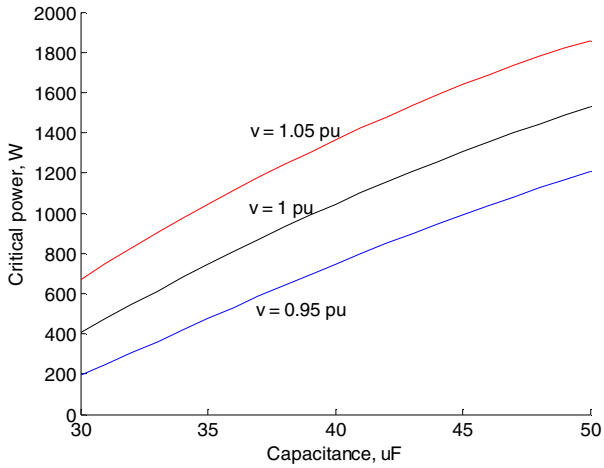


Fig. 8 Variation of critical power against capacitance for various speeds.

V. Conclusions

A simple method of determining the maximum power of a three-phase self-excited induction generator (SEIG) has been proposed in this paper. A constrained optimization problem has been formulated to find the maximum power point while satisfying the fundamental equations of the generator and a terminal voltage constraint. The formulated problem is solved using a standard numerical based routine given in MATLAB. By properly selecting the value of the minimum specified voltage, the same optimization problem can be used to determine the critical power as well as the maximum power that can be delivered while maintaining the minimum voltage. The effectiveness of the proposed method is then evaluated on a 1.5-kW, 220-V SEIG and the results found are compared with the corresponding actual values obtained through an experimental setup in a laboratory and are observed to be in very good agreement. It is found that the maximum power that can be delivered by the generator while maintaining a reasonable terminal voltage is lower than that at the critical point. It is also found that the maximum power of the generator depends not only on the value of the capacitance used but also on the load power factor and the generator speed.

Acknowledgment

The author would like to thanks Nanyang Technological University, Singapore for providing the experimental facilities.

References

[1] G. K. Singh, "Self-excited induction generator research – a survey", Electric Power Systems Research, Vol. 69, 2004, pp. 107-114.

[2] M. G. Simoes and F. A. Farret, "Renewable energy systems – design and analysis with induction generators", CRC Press, 2004.

[3] M. H. Haque, A Novel method of evaluating performance characteristics of a self-excited induction generator", IEEE Trans. on Energy Conversion, Vol. 24, No. 2, 2009, pp. 358-365.

[4] R. C. Bansal, "Three-phase self-excited induction generators: An overview", IEEE Transactions on Energy Conversion, Vol. 20, No. 2, 2005, pp. 292-299.

[5] T. F. Chan, "Steady-state analysis of self-excited induction generators", IEEE Transactions on Energy Conversion, Vol. 9, No. 2, 1994, pp. 288-296.

[6] M. H. Haque, "Comparison of steady state characteristics of shunt, short-shunt and long-shunt induction generators", Electric Power Systems Research, Vol. 79, No. 10, 2009, pp. 1446-1453.

[7] S. S. Murthy, B. P. Singh, C. Nagamani and K. V. V. Satyanarayana, "Studies on the use of conventional induction motors as self-excited induction generators", IEEE Transactions on Energy Conversion, Vol. 3, No. 4, 1988, pp. 842-848.

[8] A. K. Al Jabri and A. I. Alolah, "Limits on the performance of the three-phase self-excited induction generators", IEEE Transactions on Energy Conversion, Vol. 5, No. 2, 1990, pp. 350-356.

[9] L. Shridhar, B. Shing and C. S. Jha, "A step towards improvements in the characteristics of self-excited induction generator", IEEE Transactions on Energy Conversion, Vol. 8, No. 1, 1993, pp. 40-46.

[10] T. F. Chan, "Capacitance requirements of self-excited induction generators", IEEE Transactions on Energy Conversion, Vol. 8, No. 2, 1993, pp. 304-311.

[11] N. H. Malik and A. A. Mazi, "Capacitor requirements for isolated self excited induction generators", IEEE Transactions on Energy Conversion, Vol. 2, No. 1, 1987, pp. 62-68.

[12] C. Grantham, D. Sutanto, B. Mismail, "Steady-state and transient analysis of self-excited induction generators", IEE Proc. Part-B, Vol. 136, No. 2, 1989, pp. 61-68.

[13] O. Ojo, "Minimum airgap flux linkage requirement for self-excitation in stand-alone induction generators", IEEE Transactions on Energy Conversion, Vol. 10, No. 3, 1995, pp. 484-492.

[14] S. P. Singh, B. P. Singh and M. P. Jain, "Comparative study on the performance of a commercially designed induction generator with inductor motors operating as self excited induction generators", IEE Proc.-C, Vol. 140, No. 5, 1993, pp. 374-380.

[15] A. M. Eltamaly, "New formula to determine the minimum capacitance required for self-excited induction generator", Proc. of the IEEE 23rd Power Electronics Specialist Conference, 2002, Vol. 1, pp. 106-110.

[16] T. Ahmed, O. Noro, E. Hiraki and M. Nakaoka, "Terminal voltage regulation characteristics by static var compensator for a three-phase self-excited induction generator" IEEE Transactions on Industry Applications, Vol. 40, No. 4, 2004, pp. 978-988.

[17] B. Singh and L. B. Shilpakar, "Analysis of a novel solid state voltage regulator for a self-excited induction generator", IEE Proc.-C, Vol. 145, No. 6, 1998, pp. 647-655.

[18] M. H. Haque, "Selection of capacitors to regulate voltage of a short-shunt induction generators", IET Generation, Transmission and Distributions, Vol. 3, No. 3, 2009, pp. 257-265.

[19] M. H. Haque, "Voltage regulation of a stand-alone induction generator using thyristor-switched capacitors", Proc. of the IEEE International Conference on Sustainable Energy Technologies, Nov 2008, pp. 44-49.

[20] C. W. Taylor, "Power system voltage stability", McGraw-Hill, NewYork, 1994.

[21] S. S. Murthy and R. K. Ahuja, "Capacitive VAR and MPPT estimation of wind driven self excited induction generators for decentralized power generation using MATLAB graphical user interface based methodology" IEEE International Conference of Power Electronics, Drives and Energy Systems, Bengaluru, India, 2012.

[22] S. N. Mahato, S. P. Singh and M. P. Sharma, "Capacitors required for maximum power of s self-excited single-phase induction generator using a three-phase machine", IEEE Transactions on Energy Conversion, Vol. 23, No. 2, 2008, pp. 372-381.

[23] S. M. Alghuwainem, "Steady-state analysis of an isolated self-excited induction generator driven by regulated and unregulated turbines", IEEE Transactions on Energy Conversion, Vol. 14, No. 3, 1999, pp. 718-723.

[24] Matlab R2009b and Simulink, 2009.

Development and Performance Analysis of a Two-Phase Induction Motor in the Frame and Core of a Single-Phase Induction Motor

Md Ayubur Rahman Khan*, Md Quamrul Ahsan

Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology
Dhaka-1000, Bangladesh

*monty_ndc_08@yahoo.com

Abstract—**In this paper, a two-phase induction motor is developed in the core of a single-phase induction motor to run with a two-phase supply. Modification is done by changing the auxiliary winding of the single-phase motor thus giving a constant 90 degree phase difference between main and auxiliary windings being fed through a two-phase voltage source. The two-phase supply is obtained from the conventional three-phase supply by a Scott-T transformer. Scott-T transformer is made of two single-phase transformers with different tapping ratios. Then these two transformers are connected in such a manner which gives two-phase supplies from a three-phase source. The motor is operated in both single-phase and two-phase modes varying the loads in the motor. An electromagnetic dynamometer brake is used for loading purpose in the motor shaft. Load is increased with increasing the DC excitation voltage of the dynamometer and necessary data are collected. The performance of the two-phase mode is analysed and then compared with that of a conventional single-phase mode of induction motor. The developed two-phase mode of operation gives the single-phase induction motor to gain the advantage of starting torque, higher efficiency, less torque harmonics, higher rated power and lower copper loss.**

***Keywords*—two phase, two-phase motor, two-phase supply, scott-t transformer.**

I. Introduction

Single-phase induction motors are widely used in many commercial and domestic applications. But their performance is inferior compared to the poly-phase induction motors due to zero starting torque, low rated power, high torque harmonics and low efficiency.

As single-phase induction motors have no starting torque, they require external electrical equipment for starting. Conventional methods for starting a single-phase induction motor are quite expensive. Besides, they need mechanical switches which are the most troublesome parts of the motor. They frequently suffer from mechanical wear that affects the performance of the motor. In this regard, two-phase operation of single-phase induction motor can serve a promising way for energy saving which requires no external starting equipment.

In this paper, a practical experiment is carried out to develop a two-phase motor in the same frame and core of a single-phase induction motor. There are several ways to drive a two-phase motor. Two-phase supply can be obtained from both conventional single-phase and three-phase supplies. In [3] an inverter circuit is proposed to drive a two-phase induction motor using 8 power transistors. Here, square fed delta modulation technique is used. In [4] a conventional method using three-phase inverter is proposed to drive two-phase induction motor. There are 6 power transistors used here and this method is only proposed for driving asymmetrical two-phase motors. In [5] a technique is proposed where the main winding of a two-phase motor is supplied from the conventional single-phase supply and auxiliary winding is fed through a variable voltage magnitude and variable phase-angle voltage source. For variable frequency motor operation a matrix converter methodology is used in [6] to drive the two-phase induction motor. Another example of driving single-phase induction motor as two-phase motor using two-phase inverter is presented in [7]. Here, additional harmonics are produced by solid state inverters. As a result, an extra circuitry is required to eliminate these harmonics.

The motivation of the experiment presented here is to develop a two-phase mode of operation in the core of a single-phase motor to analyze the improved performance caused by two-phase drive. To do so, an easier and simple method is proposed here to achieve two-phase supply from conventional three-phase supply using Scott-T transformer. Through this mode of operation advantages of poly-phase induction motor such as starting torque, higher efficiency, higher rated power, lower copper losses, reduced vibration and noise become available in single-phase induction motor.

II. Design of Two-Phase Induction Motor

Conventional capacitor-start single-phase induction motor has two windings, the main winding and the auxiliary winding. A capacitor is placed in series with the auxiliary winding, thus a 90 degree phase difference in current is obtained. Consequently, a torque is developed and motor becomes a self-starting motor. After the motor starts, the auxiliary winding is disconnected usually by means of centrifugal switch that operate at about 75 percent of synchronous speed. Finally the motor runs because of the main winding. Since this is being single-phase, some level of humming noise is associated with motor during running.

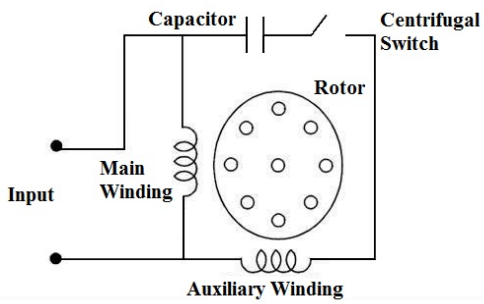


Fig. 1 Schematic diagram of capacitor start single-phase induction motor.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

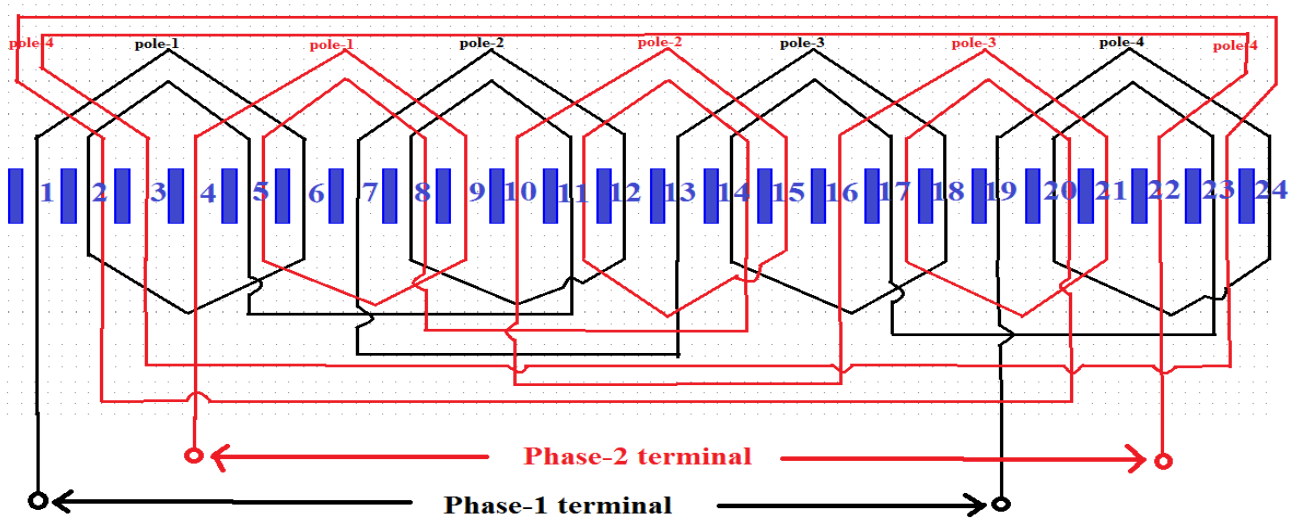


Fig. 2 winding configuration of two-phase induction motor.

To run this single-phase induction motor with two-phase supply modification is done. Capacitor and centrifugal switch is cut off first. The auxiliary winding is made of thin wire compared to that of main winding which has thick copper wire. It is because in single-phase operation auxiliary winding is connected to the supply only for a few times during staring. So, transformation is made to the auxiliary winding to have the same thick copper wire and number of turns as that of the main coil.

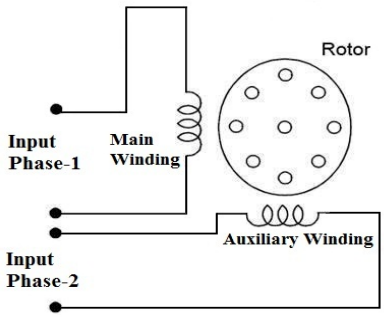


Fig. 3 Schematic diagram of two-phase induction motor.

In the developed two-phase induction motor there are 24 slots in the core of stator. Main and auxiliary coils are wound to develop a 4-pole motor shown in Fig. 2. Synchronous speed of the motor is thus 1500 rpm as supply frequency is 50 Hz.

III. TWO-PHASE SUPPLY

In order to successfully operate a single-phase induction motor as two-phase motor, two-phase power supply is required. This is done from three-phase supply using Scott-T transformer [1],[2].

The Scott-T transformer requires a three-phase power input and provides two equal single-phase outputs called Main and Teaser. The Main and Teaser outputs are 90 degrees out of phase. Scott-T transformer is made of two single phase transformers of equal power rating.

This connection consists of a centre-tapped 1:1 ratio Main transformer, T1, and an 86.6% ($0.5\sqrt{3}$) ratio Teaser transformer, T2. The centre-tapped side of T1 is connected between two of the phases on the three-phase side. Its centre tap then connects to one end of the higher turn count side of T2, the other end connects to the remaining phase. The other sides of the transformers (T1 & T2) then connect directly to the two pairs of the two-phase four wire induction motor.

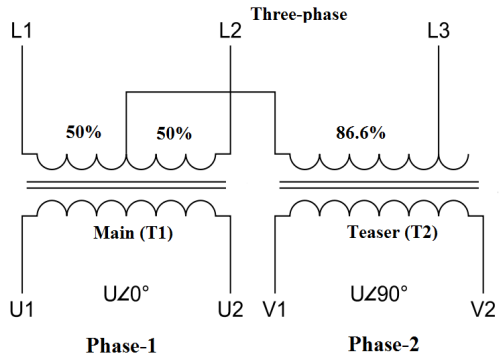


Fig. 4 Schematic diagram of Scott-T connection.

The complete system is then connected for operation and data collection. Modified induction motor, Scott-T transformer and electromagnetic dynamometer brake are assembled for final purpose of measurement in the picture of Fig. 5.

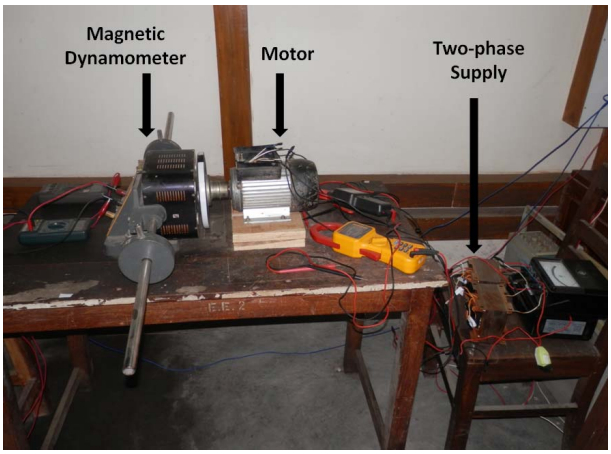


Fig. 5 Complete system with electromagnetic dynamometer.

IV. DATA COLLECTION PHASE

The modified motor is used to analyze the performance of the induction motor under single-phase and two-phase modes of operation. An electromagnetic dynamometer brake is used

to measure generated torque and power by the motor at various loads. With increasing the DC excitation voltage of the dynamometer the motor shaft is loaded. For various loads input power, output power, output torque and rotor speed are measured for both single-phase and two-phase operations.

A. Single-phase operation

In single-phase mode of operation a sinusoidal voltage of 220V is applied in one phase terminal of the modified motor. Varying the load of the shaft various data are collected.

TABLE I
DATA COLLECTION WITH SINGLE-PHASE

Input power (Watt)	Output power (Watt)	Developed torque (Nm)	Rotor speed (rpm)
38	0	0	1498
40	6.1	0.04	1495
46	9.1	0.06	1488
56	15	0.1	1477
70	26.8	0.17	1464
93	43.6	0.3	1428
123	59	0.4	1383
146	68.6	0.49	1350

B. Two-phase operation

In two-phase mode of operation two sinusoidal voltages, each being 220V, from the main and teaser windings of the Scott-T transformer are applied in both the terminals of each phase of the same motor. Here corresponding data are also collected varying the electromagnetic load.

TABLE II
DATA COLLECTION WITH TWO-PHASE

Input power (Watt)	Output power (watt)	Developed torque (Nm)	Rotor speed (rpm)
30	0	0	1500
33	4.6	0.03	1498
40	15.2	0.1	1494
52	19.7	0.13	1490
63	30	0.2	1472
83	47.7	0.31	1465
100	59	0.39	1452
134	89.5	0.6	1420
162	119.7	0.82	1401
205	137.5	0.97	1352
242	161.3	1.17	1322

V. COMPARATIVE ANALYSIS

With no load, both single-phase and two-phase modes have similar speed. In both cases, rotational losses are also similar when there is no load. When load is increased, developed torque and output power of the two-phase mode go higher than that of the single-phase mode. These two attributes of two-phase mode are compared with the single-phase mode in Fig. 6 and Fig. 7 respectively.

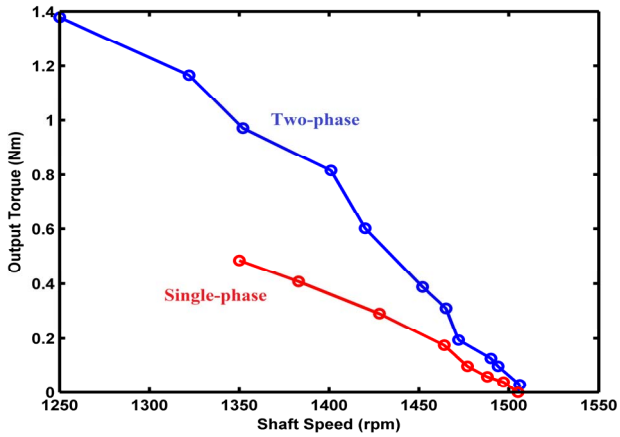


Fig. 6 Developed torque comparison.

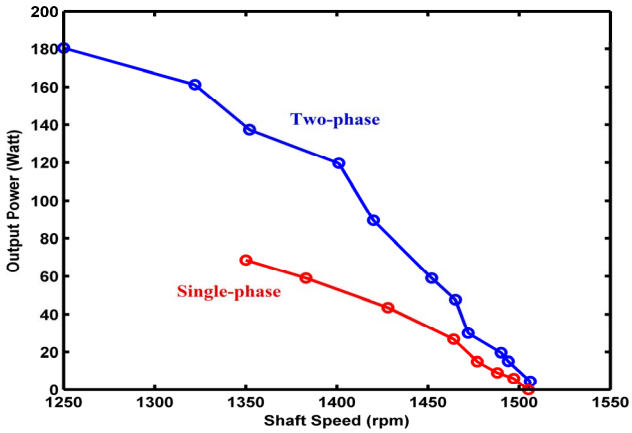


Fig. 7 Output power comparison.

It is prominent from the above figures that two-phase motor produces higher torque, which means more power for the same frame size of the motor. Table III shows that a 83% increase in torque is obtained at a 5% slip. Therefore, higher rated torque is achieved in two-phase operation.

TABLE III
POWER GAIN IN TWO-PHASE

	Single-phase		Two-phase	
Slip	2.7%	5%	2.7%	5%
Developed Torque (Nm)	0.17	0.3	0.31	0.55
Torque Gain	-	-	82.35%	83.33%

Besides, stator copper loss is also calculated for both modes of operation and their comparative analysis with increasing the load is shown in Fig. 8. Stator copper loss is increased for both modes with increasing load as it draws more currents through the stator winding. Copper loss is calculated using the input stator currents and stator resistance measured from the DC test of induction motor.

If we assume a motor slip of 5%, then copper losses in single-phase and two-phase are as follows:

$$P_{copper,\sin gle-phase} = 20.78\ \mathrm{W}$$

$$P_{copper,two-phase} = 16.5\ \mathrm{W}$$

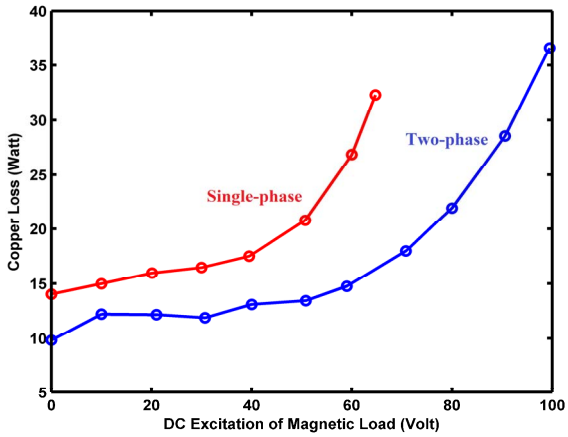


Fig. 8 Stator copper loss comparison.

Greater efficiency is also observed while motor is running in two-phase mode. Efficiency comparison between these two modes is illustrated in Fig. 9 with respect to shaft speed.

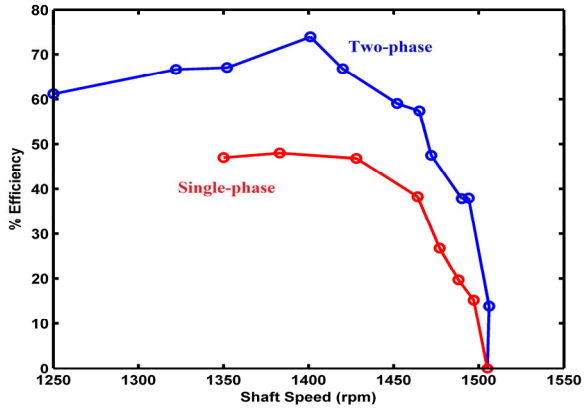


Fig. 9 Efficiency comparison with respect to shaft speed.

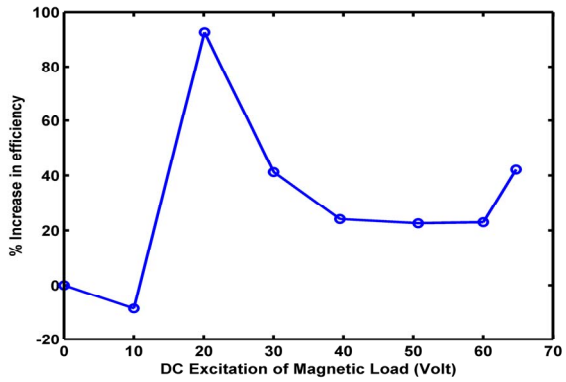


Fig. 10 Efficiency increment in two-phase mode.

It can be seen that with smaller load both single and two-phase modes gives similar efficiency. But with the increase of load, efficiency goes higher for two-phase operation the increment of which is in the range of 22.6% to at most 92.4% with an average of 29.6%. This increase in efficiency is shown in Fig. 10.

These results show that the two-phase operation increases the output power and reduces the copper losses, which means higher efficiency. This is due to the fact that, with the contribution of the two windings, a higher output can be obtained even with less current in each individual winding.

In this experiment, a single-phase induction motor is used to convert it into a two-phase induction motor driven with a two-phase supply. If we have a look at total the initial cost of both single-phase and two-phase motor, then it can be clearly seen that total conversion from single-phase to two-phase induction motor requires only less than 30 $ including cost of Scott-T transformer for getting two-phase supply.

TABLE IV
CONVERSION COST ANALYSIS

Single-phase		Two-phase	
Motor core, frame and rotor	17 $	**Motor core, frame and rotor**	17 $
Winding cost	13 $	**Winding cost**	15.5 $
Starting capacitor	1.5 $	**Scott-T transformer**	32 $
Centrifugal switch	4 $	-	-
Total cost	35.5 $	**Total cost**	64.5 $

Apart from these advantages, two-phase induction motor has benefits such as higher starting torque that is required to operate with high-starting-torque loads. Moreover, unlike single-phase motor, no humming noise and less vibration are associated with two-phase motor which is also essential for silent operation of motors.

VI. CONCLUSION

Single-phase induction motors are widely used for pumping purpose in households, irrigations and industries as well. A big portion of electricity bill is carried out by these motor pumps in a developing country like Bangladesh where energy crisis is a significant issue. If we replace them by more efficient two-phase motors with a minimal cost proposed here, then we can achieve starting torque with a higher rated power which means supplying water at a higher level is possible in a more efficient way. Moreover if we drive this motor by solar powered DC voltage through a two-phase inverter then it can reduce the load of our national grid power system to a great extent.

The results obtained, has confirmed that two-phase mode of operation gives better performance. Increased rated torque indicates more power in the same frame size thus less initial cost. Besides, improved efficiency translates into less running cost of motor. These savings justify the additional cost for motor conversion and two-phase supply circuits.

REFERENCES

[1] Stephen J. Chapman, *Electric machinery fundamentals*, 4th ed., New York: McGraw-Hill, 2005, pp. 131,132

[2] Jignesh Parmar. (2012, May 6). Scott-T Connection of Transformer [Online].Available: http://electricalnotes.wordpress.com/2012/05/06/scott-t-connection-of-transformer/

[3] Alexa, D., "Static frequency converter for supplying an asynchronous two-phase motor", in *Electric Power Applications, IEE Proceedings B*, vol. 134, issue: 1, 1987, pp. 57-60

[4] Holmes, D.G., and Kotsopoulos, A., "Variable speed control of single and two phase induction motors using a three phase voltage source inverter", in *IEEE Industrial Applications Society annual meeting conference records*, vol. 1, Toronto, Ont., 1993, pp. 613-620

[5] Collins, E.R., Jr., "Single-phase induction motor adjustable speed drive: direct phase angle control of the auxiliary winding supply", in *Industry Applications Society Annual Meeting, IEEE,* vol. 1, 1988, pp. 246-252

[6] M. Asgar, S. Mansourpour, E. Afjei, "Matrix converter for two-phase induction Motor Application: Analysis, Modeling and Simulation", in *Power Electronics, Drive Systems and Technologies Conference (PEDSTC)*, Tehran, Iran, 2011, pp. 246-250

[7] Naser Abdel-Rahim, Adel Shaltout, "Operation of Single-Phase Induction Motor as Two-Phase Motor", in *IECON 02 [Industrial Electronics Society, IEEE],* vol. 2, Spain, 2002, pp. 967-972

Design and Analysis of Isolated Integrated Charger for PHEV

Rajib Baran Roy[1*], Enamul Basher[2] and Joshua Mithil Biswas[3]
[1, 3] Department of Electrical and Electronic Engineering,
University of Information Technology and Sciences (UITS), Dhaka, Bangladesh
[2] Department of Electrical and Electronic Engineering,
Bangladesh University of Engineering and Technology, Dhaka, Bangladesh
[*]rajib.baran@uits.edu.bd

***Abstract*—The operation and efficiency of HEV (hybrid electrical vehicle) greatly depends on its charging system. The HEV uses grid power to charge the battery. The traction circuit components are not normally engaged during the charging time, so there is a possibility to use them in the charger circuit to have an on-board integrated charger. Most charges are non isolated type having low charging power capacity. Therefore on board isolated high power charger is required for PHEV (plug in hybrid electrical vehicle) which may improve the charging efficiency of the battery. The size and price are important considerations for such integrated charger. In this paper, an integrated on board charger is proposed which provides high charging power and unity power factor operation. In the proposed design, asynchronous machine and bi directional converter are used so that the machine can act as motor during traction and isolated transformer during charging. The boost converter topology is used for the bidirectional converter which can be used as inverter during traction. The proper operation of integrated charger greatly depends on grid synchronization and controlling mechanism of converter and asynchronous machine. The Park transformation is used for mathematical electromechanical model of the asynchronous machine. A controller scheme is designed by using the concepts of conventional controlling scheme of IPM (induction permanent magnet) motor/generator set and decoupled current control method of boost converter. The whole system is simulated in Matlab/Simulink based model in order to verify the system operation of the proposed charger. The simulation results show good performance of the charger leading to machine speed stability and smooth grid synchronization.**

***Index Terms*—Bidirectional converter, boost converter topology, integrated charger, PHEV.**

I. Introduction

The performance of battery modules in HEV depends on the charger. Generally there are two types of battery chargers which are on-board type and stand-alone (off-board) type [1], [2]. The on-board type of charger always should be carried by the vehicle which increases the weight of the system. With respect to weight, space and cost, it is difficult to install high power capacity on-board chargers in HEV. The assurance of galvanic isolation for battery charger is also an import issue [3].The integrated charger has electric machine, converter/inverter and charger circuit which reduces the system weight, space and cost [4]. It can be a suitable option for both onboard and off-board charger. Several different types of integrated chargers are proposed for vehicle application where the electric machine is used as inductor or isolated transformer and the vehicle remains standstill during charging[2],[5].For non-isolated integrated chargers, extensive protection and shielding are used [5]. However for isolated charger, a wound rotor induction machine is used as an isolated transformer which causes low system efficiency due to high magnetization current. So far no integrated chargers which are proposed by manufactures and academicians combine both galvanic isolation and high efficiency features. Though the isolated integrated chargers provide isolation but most of them are of low power capacity [3].In this paper, an approach is made for design and analysis of high power isolated integrated charger. In this design, asynchronous motor is used as isolated transformer during charging of the battery from grid whereas in traction mode, it is used as electric motor for the vehicle. The boost converter topology is incorporated in the proposed design so that the asynchronous machine remains stationary during charging [4]. In order to implement boost converter topology, the conventional squirrel cage or phase wound configuration of the asynchronous machine is modified to split wound configuration. The bidirectional converter is used for this design which acts as converter during charging and acts as inverter during traction. The proposed on board integrated charger provides both galvanic isolation and high efficiency. The determination of controlling topology for the integrated charger is the major concern of this design.

II. Proposed Isolated Integrated Charger

In the proposed design, the windings of the asynchronous electric machine are reconfigured which can be used as a grid isolated transformer. In traction mode the machine acts as motor whereas in charging mode its windings are arranged by switching device so that it can act as an isolated transformer [6]. The rotor of the three phase asynchronous machine is made of PM (permanent magnet) rather than electromagnet for simplification of machine construction so that it can start from initial stationary state without any extra starting arrangement [7]. This rotary three-phase machine constitutes a three-phase bidirectional converter which can be used as inverter as well. The figure 1 shows the schematic diagram of the proposed integrated charger for a PHEV. The charging power of the battery is controlled by bidirectional converter and therefore high power charging is possible in this configuration. During charging, the inverter is synchronized with the grid voltage through stator windings of the electric

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

machine. In order to keep synchronism, the gate pluses of switching devices of the bidirectional converter are synchronized with the grid voltage by using PWM technique [5], [8]. The stator windings act as an isolated three-phase transformer and the converter can control the dc voltage and current at the battery side [7]. The torque of the electric machine is kept zero during synchronization with the grid. For proper operation of the bidirectional converter, the dc bus voltage needs to be more than the peak ac line voltage which also ensures unity power factor operation [9]. This is attained by using Y −Δ connection of the stator windings of the machine.

The arrangement of stator and rotor windings of the two pole asynchronous machine is an important consideration for bidirectional operation of the machine [10]. The three phase stator winding is divided into six windings which are shifted 30^0 electrical degrees around the stator periphery [4]. These six windings can be considered as two sets of three-phase windings. The figure 2 shows the cross section of the machine where the direct axis components of the stator windings are a_1, b_1, c_1 for the first set and a_2, b_2, c_2 for the second set which are shifted 30^0 electrical degrees on the stator periphery. Similarly the quadrature axis components are a'_1, b'_1, c'_1 for the first set and a'_2, b'_2, c'_2 for the second set.

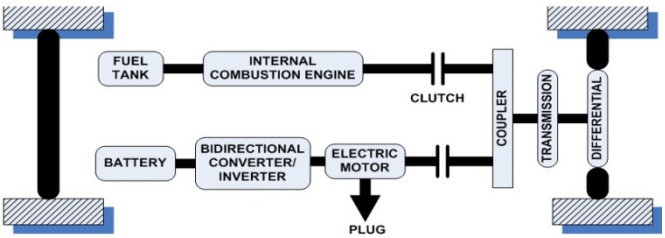


Fig. 1 Asynchronous motor based proposed integrated charger

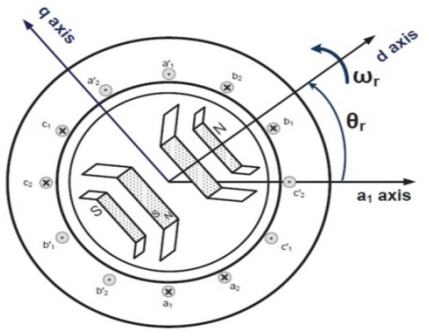


Fig. 2 Cross-sectional view of split stator windings

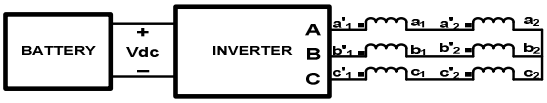


Fig. 3 System operation in traction mode

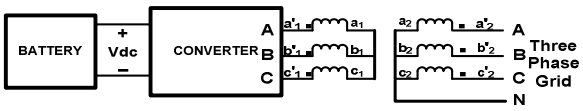


Fig. 4 System operation in charging mode

A. Mathematical Model of Six Stator Windings

A mathematical model of the machine is developed based on the assumption that all stator windings and magnetomotive forces of rotor magnets are sinusoidally distributed. This machine has six stator windings. A special abc to dq transformation is used which is based on the extended version of Park transformation for six variables instead of three [11], [12]. The system equations in the dq frame are reduced to a six order system while all sin and cosine terms are eliminated from the equations [12]. The voltage equations for six windings can be derived as follows [10], [12],

$$v_{a1} = r_s i_{a1} + \frac{d}{dx}\varphi_{a1} \quad \text{(1)}$$

$$v_{b1} = r_s i_{b1} + \frac{d}{dx}\varphi_{b1} \quad \text{(2)}$$

$$v_{c1} = r_s i_{c1} + \frac{d}{dx}\varphi_{c1} \quad \text{(3)}$$

$$v_{a2} = r_s i_{a2} + \frac{d}{dx}\varphi_{a2} \quad \text{(4)}$$

$$v_{b2} = r_s i_{b2} + \frac{d}{dx}\varphi_{b2} \quad \text{(5)}$$

$$v_{c2} = r_s i_{c2} + \frac{d}{dx}\varphi_{c2} \quad \text{(6)}$$

where v_{a1}, v_{b1}, v_{c1}, v_{a21} , v_{b2} , v_{c2} , i_{a1} , i_{b1} , i_{c1} , i_{a2}, i_{b2} , i_{c2}, φ_{a1}, φ_{b1}, φ_{c1}, φ_{a2}, φ_{b2}, φ_{c2} and r_s are windings voltages, currents, flux linkages and resistance respectively. The stator current i_s and voltage v_s can be derived by the following equations [9],

$$i_s = [\, i_{a1} \quad i_{b1} \quad i_{c1} \quad i_{a2} \quad i_{b2} \quad i_{c2} \,]^T \quad \text{(7)}$$

$$v_s = [\, v_{a1} \quad v_{b1} \quad v_{c1} \quad v_{a2} \quad v_{b2} \quad v_{c2} \,]^T \quad \text{(8)}$$

The developed electromagnetic torque in the machine can be calculated as follows [8],

$$T_e = \frac{P}{2}\left(\frac{1}{2}\, i_s T \frac{\partial L_s}{\partial \theta_r} i_s + i_s\, T \frac{\partial \varphi_{pm}}{\partial \theta_r}\right) \quad \text{(9)}$$

where, P is the number of poles, Ψ_{pm} is the stator winding fluxes due to rotor magnets and θ_r is the angular displacement of rotor.

B. Operation during Traction and Charging

The proposed integrated charging system has two modes of operation which are traction and charging. In traction mode, two set of stator windings are connected to each other in series to constitute a three-phase winding set. These three windings can be connected in Δ or Y to form a classical three phase machine. In traction mode, the machine acts as a motor which is powered by the battery through the inverter. The figure 3 shows the system diagram in traction mode. For charging mode, the two set of stator windings are separated from each other and windings a_2, b_2 and c_2 are connected to the grid for battery charging via converter. The system is reconfigured according to the scheme shown in the figure 4. In charging mode, the machine is kept standstill and it acts as an isolated transformer for transferring power from grid to the converter. The converter uses this isolated voltage source to charge the battery by means of machine leakage inductances (yielding a three phase boost converter).

C. Grid Synchronizing and Controlling Topology

In order to keep synchronization with the grid, it is necessary to determine the controlling topology of the proposed integrated charger [13]. When the battery is connected to the grid during charging, the dc voltage, stator current and rotor speed of the asynchronous machine need to be measured for controlling of the system [11]. For proper synchronization, both the magnitude and angle of grid side stator winding voltage and grid voltage need to be same [14]. Since the magnitude of voltage of the asynchronous machine

is a function of the speed and flux so by controlling the flux the voltage can be controlled. In a classical electrical machine usually the reference value for the direct component of the machine voltage is considered zero but in this integrated charger this value is used as a control parameter to change the induced voltage magnitude in machine [12],[15]. In order to match the voltage angles of the grid side stator windings and grid voltage, a PI controller is used to adjust the speed reference of the machine with the angle error signal [16]. The direct component of the voltage is close to zero so the angle error is replaced by the voltage direct components error in the PI controller for phase synchronization [15]. The figure 5 shows the schematic diagram of the charge control system of the proposed charger. When both voltage magnitude and angle error signals are within the predefined bands, the asynchronous machine is synchronized with the grid and the system is ready for the charge operation [17].

The direct and quadrature current control and feed forward compensation are used for this decoupled control scheme. The controller output sets the reference value for the direct component of the current that controls the power [15]. Two independent PI controllers are used to generate reference values u_α and u_β for the converter. The feed forward terms are added to reference values of the decoupled system in direct and quadrature axis to improve the system performance [13]. In the control system of the proposed design, currents are replaced by the difference of the primary and secondary winding currents. The control scheme of figure 5 is an extension of the classical boost converter control with some modifications. Two sets of three-phase machine currents and rotor position are measured and used in the controller. The grid voltage is also measured and used in the controller for proper operation (feed-forward compensation) [18]. The rotor speed error (the difference between the reference speed and true speed) is added to the controller by the means of a proportional controller to reduce these oscillations [2], [18]. By changing the set point of the direct and quadrature component of the current, the bidirectional flow of real and reactive power within the system can be controlled [19].

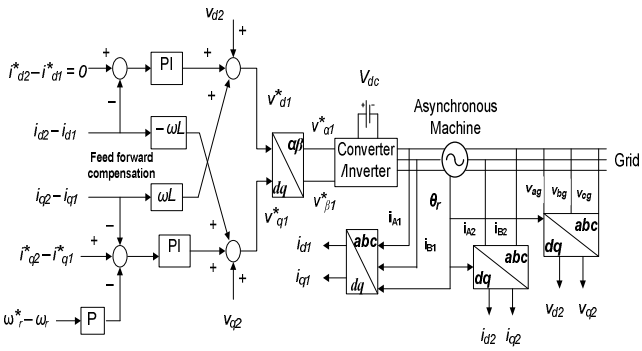


Fig. 5 Block diagram for charge control of proposed integrated charger

III. Simulated Model and Simulation Results

A 4 pole permanent magnet 3 phase asynchronous machine of 20 kW is considered for PEV vehicle. The base speed and maximum speed of the machine are 1500 and 6500 rpm respectively. During traction mode two set of stator windings are connected in delta whereas in charging mode the stator windings are disconnected from each other so that the machine can act as a transformer. The battery is considered of capacity of 200 Vdc. The charge power is limited to 12.5 kW due to the machine thermal limit. The bidirectional converter is synchronized to grid by using pulse width modulation technique. During charging, the voltage level of the grid side stator winding is adjusted by controlling the motor flux (the direct component of the current). The voltage angle is adjusted by speed control of the motor by means of the quadrature component of current [19]. The simulated model of the integrated charger for traction and charging mode is designed by using Matlab/Simulink.

A. Traction Mode

The figure 6 shows the simulated model of the proposed integrated charger during traction mode. A three-phase asynchronous machine of rated 2.24 kW, 220 V, 1725 rpm is fed by a sinusoidal PWM inverter. The base frequencies of the sinusoidal reference wave and triangular carrier wave are 60 Hz and 1980 Hz respectively for pulse width modulation. The inverter takes input from the battery and stator windings of the asynchronous machine are connected to the output side of the inverter. The load torque applied to the machine's shaft is constant which is set to its nominal value of 11.9 N.m. The machine attains its maximum speed which is considered 1 p.u. within 10 microseconds. Due to high switching frequency (1980 Hz) for PWM, the maximum time step has been limited to 10 microseconds and noise is introduced to rotor and stator currents which can be observed from the simulated results. The noise introduced by the PWM inverter is also observed in the torque waveform. However, the motor's inertia prevents this noise from appearing in the motor's speed waveform. The presence of ripple in torque waveform is about 1% of the rated torque. The figure 7 represents the waveforms for rotor speed and torque during traction mode.

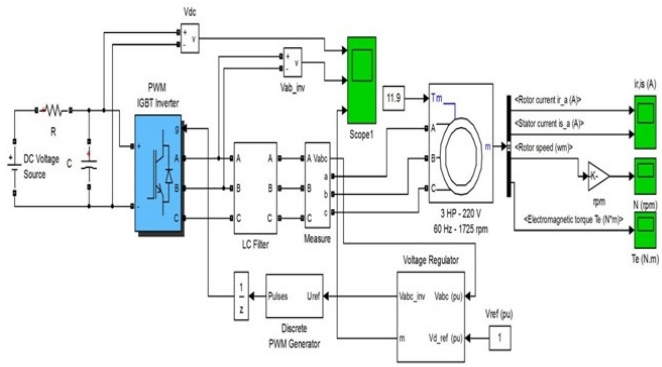


Fig. 6 Simulated model of the integrated charger during traction

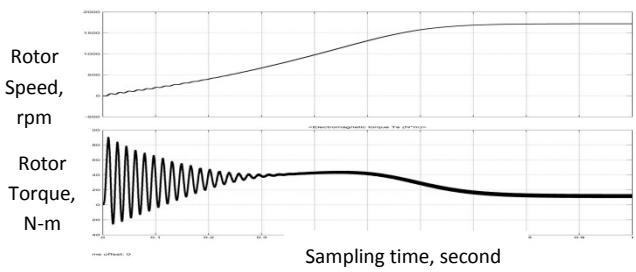


Fig. 7 Rotor speed and torque during traction

B. Charging Mode

The battery is discharged for rotation of asynchronous machine. When the SOC (state of charge) drops below 40%, the battery is charged by the bidirectional converter from the

grid when the asynchronous machine acts as a transformer. The SOC is calculated as,

$$SOC = 100(1 - \frac{1}{Q} \int_0^t i(t)dt\ ) \text{ ..... } (16)$$

The simulated model of the proposed charger during charging is shown in figure 8. The waveforms for charging voltage and SOC of battery are shown in the figure 9. It can be inferred from the waveforms of battery voltage and SOC that step charging occurs because of smooth synchronization of the charger with grid.

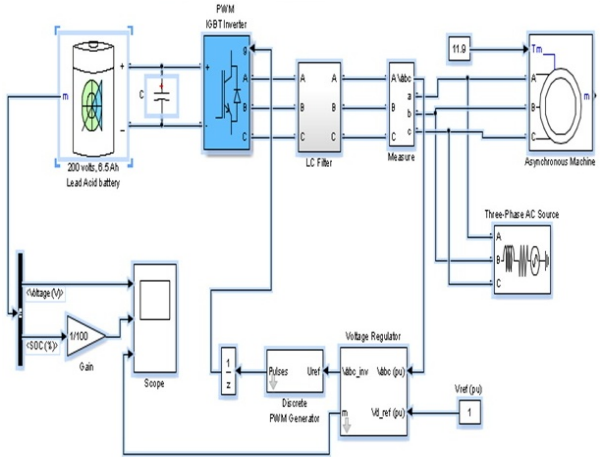


Fig. 8 Simulated model of the integrated charger during charging

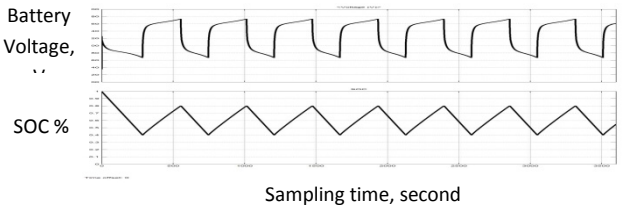


Fig. 9 Battery voltage and SOC during charging

IV. CONCLUSIONS

The proposed integrated battery charger for PHEV is cost effective and compact which has unit power factor capability. The controlling topology of the bi-directional converter is determined carefully so that it can act as converter during charging and inverter during traction. Though the use of high frequency carrier for PWM of bidirectional converter, the stator and rotor currents of the asynchronous machine become asymmetrical and distorted due to addition of noise signals from converter but the speed and torque wave shapes are less affected by noise signals due to proper synchronization of the machine with grid. The ripples in speed and torque wave shapes are less than 1% of rated speed and torque of the machine. The waveforms of battery voltage and SOC also represent smooth grid synchronization and better performance during charging. Further development in the proposed design is possible by using different winding configuration and other types of machine. The lower system efficiency during charging can be improved by using new techniques and control parameters in control scheme of the proposed design. Microcontroller based controlling of bidirectional converter can be used for better operation of the overall system. The losses and effects of magnetizing current for machine are not considered in this design. The controlling of the reactive power during the charging can affect the charging performance and therefore reactive power compensation needs to be considered in improved design of the integrated charger.

REFERENCES

[1] A. Emadi, Y. J. Lee, and K. Rajashekara, "Power electronics and motor drives in electric, hybrid electric, and plug-in hybrid electric vehicles,", *IEEE Transactions on Industrial Electronics*, vol. 55, no. 6, pp. 2237–2245, 2008.

[2] S. Y. Kim, I. Jeong, K. Nam, and H.-S. Song, "Three-port full bridge converter application as a combined charger for phevs," in *Vehicle Power and Propulsion Conference, 2009. VPPC '09. IEEE*, pp. 461 – 465, *2009*.

[3] M. Dumenjo, S. Monge, J. Yus, J. Delgado, M. Espadaler, J. Apruzzese, and J. Pereira, "A novel pulse width modulation to operate a bidirectional two-stage dc-ac converter with high-frequency isolation in discontinuous conduction mode for photovoltaic applications," *Power Electronics and Applications*, pp. 1–8, 2009.

[4] S. Haghbin, K. Khan, S. Lundmark, M. Alakula, O. Carlson, M. Leksell, and O. Wallmark, "Integrated chargers for ev's and phev's: examples and new solutions," *Electrical Machines (ICEM), 2010 XIX International Conference*, pp. 1 –6, 2010.

[5] S. Haghbin, M. Alakula, K. Khan, S. Lundmark, M. Leksell, O.Wallmark, and O. Carlson, "An integrated charger for plug-in hybrid electric vehicles based on a special interior permanent magnet motor," *Vehicle Power and Propulsion Conference (VPPC) Proceedings*, Lille, France, 2010.

[6] L. De Sousa, B. Silvestre, and B. Bouchez, "A combined multiphase electric drive and fast battery charger for electric vehicles," *IEEE Vehicle Power and Propulsion Conference VPPC) Proceedings*, 2010, France, Oct. 2010.

[7] A. Bruyre, L. De Sousa, B. Bouchez, P. Sandulescu, X. Kestelyn, and E. Semail, "A multiphase traction/fast-battery-charger drive for electric or plug-in hybrid vehicles," *IEEE Vehicle Power and Propulsion Conference (VPPC) Proceedings*, 2010, France, Oct. 2010.

[8] S. Lacroix, E. Laboure, and M. Hilairet, "An integrated fast battery charger for electric vehicle," *IEEE Vehicle Power and Propulsion Conference (VPPC) Proceedings*, 2010, France, Oct. 2010.

[9] L. D. Sousa and B. Bouchez, "Combined electric device for powering and charging," International Patent WO 2010/057892 A1, 2010.

[10] L. Solero, "Nonconventional on-board charger for electric vehicle propulsion batteries,", *IEEE Transactions on Vehicular Technology*, vol. 50, no. 1, pp. 144 –149, Jan. 2001.

[11] W. K. Thong and C. Pollock, "Low-cost battery-powered switched reluctance drives with integral battery-charging capability," *IEEE Transactions on Industry Applications*, vol. 36, no. 6, pp. 1676 – 1681, 2000.

[12] G. Pellegrino, E. Armando, and P. Guglielmi, "An integral battery charger with power factor correction for electric scooter," *IEEE Transactions on Power Electronics*, vol. 25, no. 3, pp. 751 –759, 2010.

[13] C. Stancu, S. Hiti, and E. Mundt, "Mobile electric power for medium and heavy duty hybrid electric vehicles," *Power Electronics Specialists Conference, 2004. PESC 04. 2004 IEEE 35th Annual*, vol. 1, pp.228 – 234, 2004.

[14] L. Tang and G.-J. Su, "Control scheme optimization for a low-cost, digitally-controlled charger for plug-in hybrid electric vehicles," *Energy Conversion Congress and Exposition (ECCE), 2010 IEEE,* pp.3604 –3610, 2010.

[15] G.-J. Su and L. Tang, "Control of plug-in hybrid electric vehicles for mobile power generation and grid support applications," *Applied Power Electronics Conference and Exposition (APEC), 2010 Twenty-Fifth Annual IEEE*, pp. 1152 –1157. 2010.

[16] C. Liaw and H. Chang, "An integrated driving/charging switched reluctance motor drive using three phase power module," *IEEE Transactions on Industrial Electronics,* 2010.

[17] H. Plesko, J. Biela, J. Luomi, and J. Kolar, "Novel concepts for integrating the electric drive and auxiliary dc converter for hybrid vehicles," *IEEE Transactions on Power Electronics*,vol. 23, no. 6, pp. 3025–3034, 2008.

[18] Y.-J. Lee, A. Khaligh, and A. Emadi, "Advanced integrated bidirectional ac/dc and dc/dc converter for plug-in hybrid electric vehicles," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 8, pp. 3970 -3980, 2009.

[19] K. Khan, S. Haghbin, M. Leksell, and O. Wallmark, "Design and performance analysis of a permanent-magnet assisted synchronous reluctance machine for an integrated charger application," *XIX International Conference on Electrical Machines (ICEM),* pp. 1 –6, 2010.

Design of Resonant Controller for Hard Disk Drive Servo System using a Mixed Passivity and Small-Gain Approach

*Md. Arifur Rahman[1], Abdullah Al Mamun[1], Kui Yao[2]
[1]National University of Singapore, 4 Engineering Drive 3, Singapore 117576
[2]A*STAR Institute of Materials Research and Engineering, 3 Research Link, Singapore 117602
Email: *a0092557@nus.edu.sg

***Abstract*—Effective compensation of mechanical resonant modes is one of the design challenges in read/write head-positioning servomechanism of hard disk drive (HDD) servo. Resonant modes create oscillation in the system response and may lead to instability if not compensated properly. This paper presents a resonant controller designed in a simple and yet efficient way to suppress the effects of resonant modes of Voice Coil Motor (VCM) actuator used in HDD. The design is based on a mixed passivity and small gain approach to an unstable plant model. Since the original theorem is for stable systems, therefore a stabilizable controller is used for the design. Experimental results substantiate the effectiveness of the controller.**

***Keywords*—VCM, Lead-lag, Passivity, Small-gain**

I. Introduction

In hard disk drives (HDD), information bits are recorded in concentric data tracks on rotating disks. The write-head used for recording and the read-head for retrieving recorded bits are fabricated on a single piece of slider which is positioned on the tracks by a head-positioning servomechanism. In single-stage servo, a voice-coil-motor (VCM) actuator is used to move several sliders over several magnetic media coated disks spun by a spindle motor. Nowadays, HDD servomechanism uses dual-stage actuation with VCM providing coarse motion and piezoelectric secondary actuator for fine positioning. But the work presented here addresses the resonance issues in a single-stage actuator. The HDD servo-system has two modes of operation - track seeking and track following. In seeking mode, the head is moved from one track to another in minimum possible of time. In track-following mode, the head is maintained as close as possible to the center of the track.

The HDD industry has experienced tremendous growth in recording density over the years and presently the density is around 1 Tbits/in^2. It is projected to exceed the 10 Tbits/in^2 barrier in the near future. This will require increase in both density of tracks (*track density*) and density of bits on a track (*bit density*). Increased track density means smaller track width and thus, improved accuracy in positioning the read/write head in spite of various disturbances. The high track density needs a servo controller with high bandwidth for eliminating the effects external disturbances. But resonant modes of the actuator prohibit bandwidth to be pushed to higher frequencies and may even lead to closed loop instability.

A popular method of suppressing the mechanical resonances in high precision micro and nano positioning systems follows the open loop design approach by cascading notch filters with the plant [1]-[4]. These modes, however, have a wide range of perturbation in their characteristic parameters, e.g., resonant frequency and damping coefficient, especially in mass produced systems like HDD, which are mainly caused by batch-to-batch variation. Even in the same batch, resonant modes may vary due to ambient temperature changes, mechanical wear, excessive shock and disk rotation. This requires the magnitude plot of the notch filters to be wide and deep enough to accommodate all possible variations. Moreover, in a system with constrained sampling frequency, if a resonant mode exits near the Nyquist frequency, the digital realization of the notch filter fails to effectively suppress that mode. Multi-rate realization of the notch filter offers a possible solution [5], but it requires an accurate model of the plant and again robustness is compromised. Many feedback controllers have been proposed to damp the resonant mode of the actuator. Positive position feedback (PPF) control is a popular method which is designed using root locus technique where unknown parameters have to be determined for a second order controller and PPF suffers from the problem of low gain and phase margins [6]. An adaptive notch filter is proposed in [7], comprising of a tunable finite-duration impulse response (FIR) filter. The tuner is designed to identify the resonance frequency of the actuator rapidly and automatically adjust the center frequency. A similar method is used in [8], in which the center frequency follows the change in resonant frequency based on the balance of two types of frequency weighted variances of the control input signal. But adaptive methods are often complex and computationally expensive.

A resonant controller to damp the first resonant mode of the VCM actuator is proposed in this paper. The method is simple and easy to implement. A mixed passivity and small gain approach is followed to achieve large damping of the resonant mode with robust finite gain stability. Since the theorem is applicable for interconnection of two stable systems and VCM actuator is essentially a double-integrator, the plant is first stabilized by a stabilizing controller.

Small-gain approach is a possible solution to guarantee finite gain (FG) stability of the interconnection between the stabilized VCM and resonant controller. The small-gain theorem states that the feedback interconnection between two stable linear systems is stable if the loop-gain of the two systems is strictly less than one [9]-[10]. As the gain of the VCM actuator is large near resonance frequency, the small-gain approach requires that the gain of controller be low to make the loop-gain less than one.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

The rest of the paper is organized as follows. Mixed passivity and small gain theorem is discussed in section 2. Experimental setup used for this work is described in Section 3. The frequency response of the VCM actuator and response of the identified model are also given in the same section. Design of the controller is explained in Section 4 and experimental results are shown in Section 5. Conclusion of our findings and scope of future works are given in Section 7.

II. Mixed Passivity and Small Gain

Let $\mathcal{R}$ be the set of all real rational transfer function matrices and $\mathcal{RH}_\infty$ be the set of all real rational stable transfer functions. The largest and smallest singular values of a matrix A will be denoted by $\overline{\sigma}(A)$ and $\underline{\sigma}(A)$.

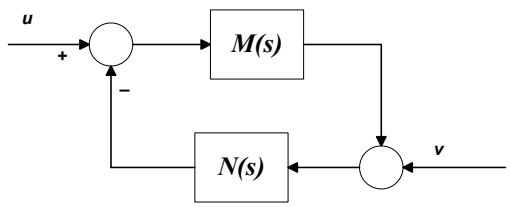


Fig. 1: Interconnection of systems with negative feedback

Consider the negative feedback interconnection between $M(s)$and $N(s)$ as shown in Fig. 1, where $M(s)$ and $N(s)$ are systems with "mixed" passivity and finite-gain properties. $M(s)$ and $N(s)$ are respectively bounded by gain k_1 and k_2 on different frequency intervals. According to [11] and [12], the theorem for mixed passivity and small gain properties of two interconnected system is as follows:

Theorem 1: Let $M(s) \in \mathcal{RH}_\infty$ and $N(s) \in \mathcal{RH}_\infty$ be two strictly proper, causal and linear time invariant systems which are connected as shown in Figure 1 where $M(s)$ is mixed passive and finite-gain system bounded by a gain $k_1 > \overline{\sigma}(M(0))$ and $N(s)$ is mixed passive and finite gain system bounded by a gain $k_2 > \overline{\sigma}(N(0))$. The interconnection of Figure 1 is finite-gain stable if (i) $k_1 k_2 < 1$, (ii) there exists two sets frequency bands: (a) a set of frequency intervals over which $M(s)$ and $N(s)$ both are passive or negative passive (b) a set of frequency intervals over which both $M(s)$ and $N(s)$ are respectively bounded by gain k_1 and k_2.

III. Experimental Setup

The setup for identification experiment (Fig. 2(b)) consists of a dynamic signal analyzer (DSA), a driver for the VCM actuator and the Laser Doppler Vibrometer (LDV) OFV 5001. The sweep sine excitation signal generated by the DSA is applied to the driver which provides required voltage-to-current conversion to drive the VCM actuator. The displacement of the actuator tip is measured using the LDV. The resolution of the LDV is set to 50 nm/V. The VCM actuator used for this work is obtained from a commercially available HDD (Western Digital's Caviar Black). With the help of the setup described above, the frequency response of the actuator is measured using sweep sine excitation signal generated by frequency response analyzer coded in LabVIEW. The analyzer produces the frequency response data and the result is saved for further processing. A linear model of the VCM actuator is identified from the frequency response. Experimental frequency response and the Bode plot of the identified model are shown in Fig. 3. Only the two most prominent resonances are considered in the identification.

For real-time implementation of controller, the setup is modified by replacing the DSA with a real-time controller card (dSPACE SD1104) (Fig. 2(c)). The sampling interval is set to 25 μs.

Remark: The two signals used in frequency response measurement are in volts. So the identified transfer function is unit-free. The output signal can be converted into displacement simply by multiplying by the scaling factor of 50 nm/V.

IV. Controller Design

A. Stabilizing Controller

Since the rigid body dynamics of the actuator is a double integrator, a lead compensator is used to stabilize it. We also design a lag compensator to increase the low-frequency slope of the Bode (magnitude) plot which is required for better tracking in low frequency range. The compensator is designed according to the bode-stability criteria [1] and is expressed by,

$$C(s) = K_d \frac{\left(\frac{s}{\omega_2}+1\right)\left(\frac{s}{\omega_3}+1\right)}{\left(\frac{s}{\omega_1}+1\right)\left(\frac{s}{\omega_4}+1\right)}, \qquad (1)$$

where $\omega_1, \omega_2, \omega_3, \omega_4$ are the corner frequencies of the lag and lead sections. The gain K_d is chosen to satisfy the condition

$$\left|\frac{1}{C(j\omega)P(j\omega)}\right|_{\omega=j2\pi f_v} = 1 .$$

f_v is the gain crossover frequency.

B. Resonant Controller

Now, the transfer function $M(s)$ in Figure 1 is the closed loop transfer function of VCM with lag-lead controller. Next we design a resonant controller $N(s)$ and the general transfer function for the resonant controller is as follows:

$$N(s) = -K_r \frac{s^2 + 2\xi_r \omega_r s}{s^2 + 2\xi_r \omega_r s + {\omega_r}^2}, \qquad (2)$$

where $K_r > 0$ is the gain of the controller, $\xi_r > 0$ and $\omega_r > 0$ are the damping constant and the frequency at which resonant mode needs to be damped. In our design we choose ω_r= frequency of the first resonant mode, desired damping ratio ξ_r=0.6 and then choose $K_r > 0$ in such a way that theorem 1 is satisfied.

V. Simulation Results

With the identified plant model and designed controller, some simulation analysis is done. Fig. 4(a) shows the frequency responses of the stabilized VCM model and frequency response of the interconnected system with resonant controller where sufficient damping can be observed for the first major resonant

mode. Simulated step responses (Fig. 4(b)) also proves the effectiveness of the proposed controller where less oscillation is observed for designed interconnected system with resonant controller.

V. Experimental Results

The performance of the proposed controller is evaluated experimentally using three types of reference signals. Firstly, a step input of 2 Volt is applied. The displacements measured with and without resonant controller are shown in Fig. 5(a) and Fig. 5(b), respectively. Secondly, triangular waveform of 50 Hz is used as reference input and corresponding outputs is shown in Fig. 5(c), Fig. 5(d). Finally, a sum of sinusoids $r(t) = \sin(2\pi \times 100t) + 0.8\sin(2\pi \times 200t) + 0.5\sin(2\pi \times 300t)$ is used as the reference input. Fig. 5(e) and Fig. 5(f) show the corresponding output with and without resonant controller, respectively. It is clearly evident from the results that oscillation in output becomes smaller for all cases when the resonant controller is used. Although the controller is designed to increase the damping of the first resonant mode, it is found that damping is increased for other higher frequency modes as well. Therefore it can be concluded that the proposed controller is able to suppress the effects of resonant modes.

VI. Conclusion

A simple and yet effective method of designing resonant controller is presented. The unstable plant is first stabilized and then a resonant controller is designed using a mixed passivity and small gain approach. Experimental results prove that the proposed method is effective. Although a VCM actuator of HDD is used as the experimental platform, this method can be useful in a wide range of applications. We are further investigating the applicability of this approach for designing a multi-mode resonant controller and also apply it in a dual-stage HDD servo system.

Acknowledgement

This work is supported by the Singapore National Research Foundation (NRF) under CRP Award No. NRF-CRP-4-2008-06.

References

[1] Al Mamun A, Guo G, Bi C. *Hard disk drive: mechatronics and control*, Boca Raton, USA: Taylor and Francis Publications, CRC Press; 2006.

[2] Thum CK, Du C, Chen BM, Ong EH, Tan KP, "A unified control scheme for track seeking and following of a hard disk drive servo system", *IEEE Trans Control Syst Technol*, vol.18, no.2, pp. 294–306, March 2010.

[3] Lan W, Thum CK, Chen BM, "A hard disk drive servo system using composite nonlinear feedback control with optimal nonlinear gain tuning method", *IEEE Trans Ind Electron*, vol.57, no.5, pp.1735-1745, May 2010.

[4] Suthasan T, Mareels I, Al-Mamun A., "System identification and controller design for dual actuated hard disk drive", *Control Eng Pract*, vol.12, no.6, pp.665–76, June 2004.

[5] Hyde JM, Seering WP., "Using input command pre-shaping to suppress multiple mode vibration", *in Proc. IEEE international conference on robotics and automation*, Sacramento, CA, 1991; p. 2604–09.

[6] D. Magee, "Optimal filtering to improve performance in hard disk drives: Simulation results," *in Proc. American Control Conference*, San Diego, California, June 1999; pp. 71–75.

[7] C. I. Kang and C. H. Kim, "An adaptive notch filter for suppressing mechanical resonance in high track density disk drives," *Microsystem Technology*, vol. 11, pp. 638–652, Aug. 2005.

[8] K. Ohno and T. Hara, "Adaptive resonant mode compensation for hard disk drives," *IEEE Trans. Industrial Electronics*, vol. 53, no. 2, pp. 624–630, Apr. 2006.

[9] C. A. Desoer and M. Vidyasagar, *Feedback Systems: Input-Output Properties*. San Diego, CA, USA: Academic Press, 1975.

[10] M. Green and D. J. N. Limebeer, *Linear Robust Control*, E. Cliffs, Ed.Upper Saddle River, NJ, USA: Prentice-Hall, 1996.

[11] S.K. Das, H.R. Pota and I.R. Petersen, "Stability analysis for interconnected systems with "mixed" passivity, negative-imaginary and small-gain properties," *in Proc. 3rd Australian Control Conference (AUCC), Nov. 2013;* pp. 201-206.

[12] Wynita M. Griggs, Brian D.O. Anderson and Alexander Lanzon, A "mixed" small gain and passivity theorem in the frequency domain, *Systems & Control Letters*, vol. 56, no. 9–10, pp. 596-602, October 2007.

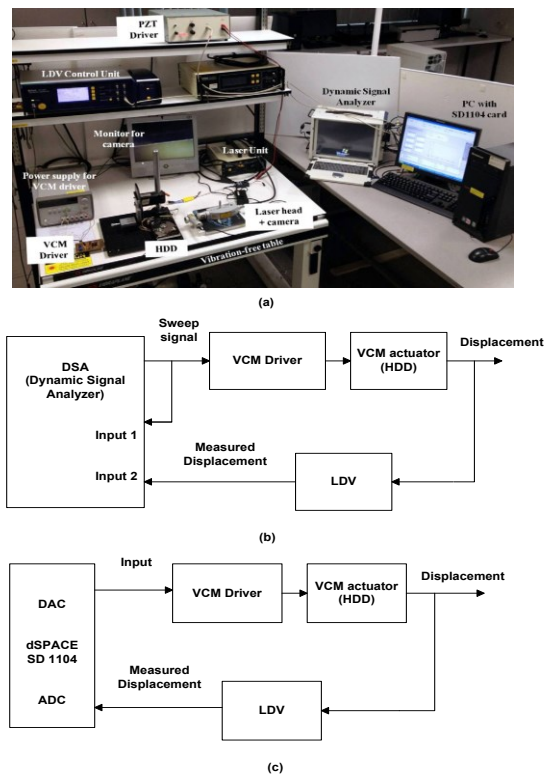


Fig. 2: (a) Experimental setup (b) Frequency response experiment (c) Controller implementation

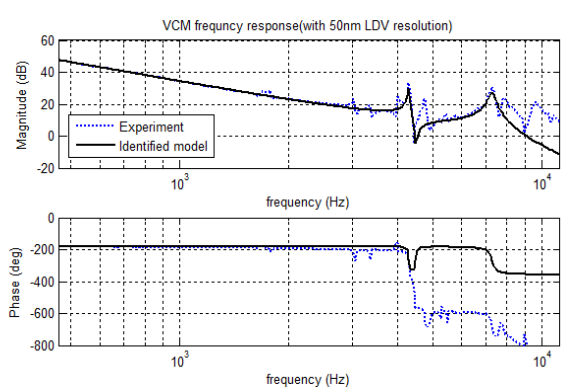


Fig. 3: Frequency response of VCM actuator

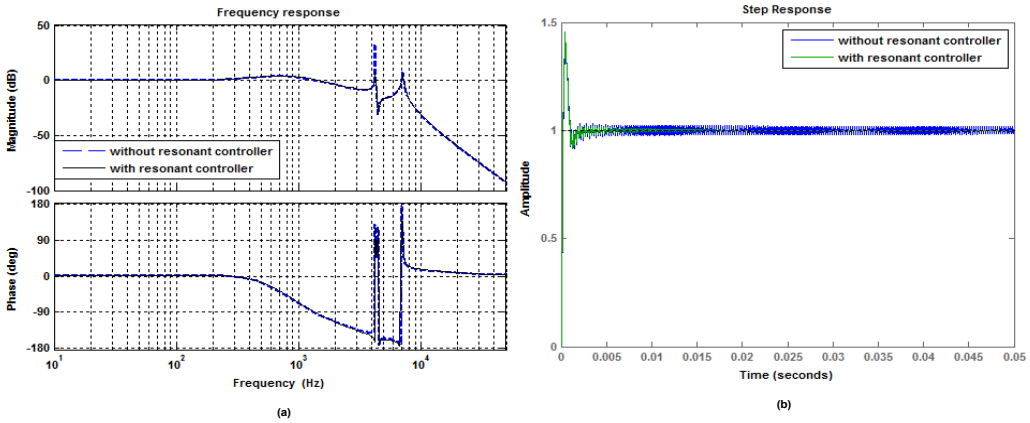


Fig. 4: Simulated (a) Frequency response (b) Step response

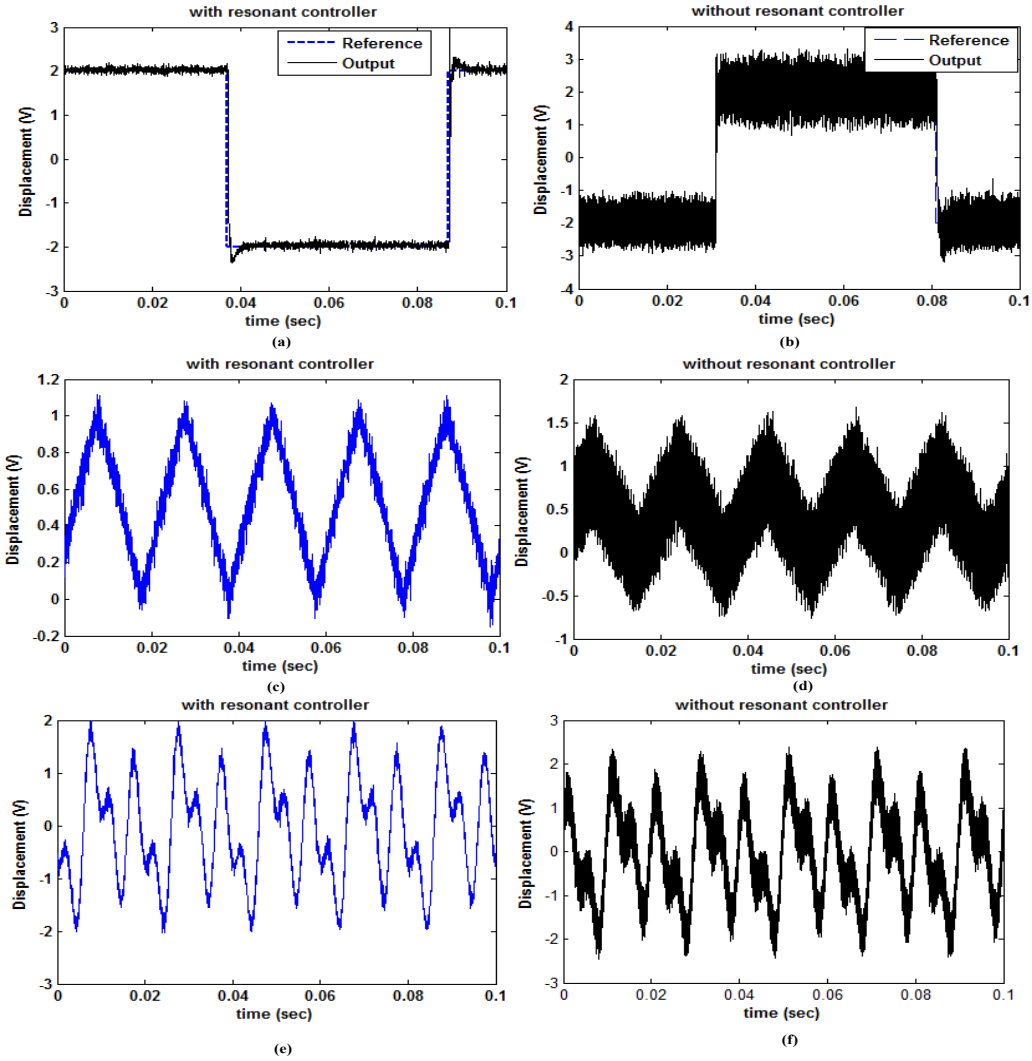


Fig. 5: Step response (a) with resonant controller (b) without resonant controller; Output for Triangular wave of 50 Hz frequency (c) with resonant controller (d) without resonant controller; Output for multiple frequency reference (e) with resonant controller (f) without resonant controller

Analysis of Bending Effects on Hexagonal and Rectangular Photonic Crystal Fiber

Tamal Sarkar,[1,*] Mofijur Rahman[1] and Anamika Dey[2]

[1]Department of Electrical & Electronic Engineering, Rajshahi University of Engineering & Technology
Rajshahi-6204, Bangladesh

[2]Department of Computer Science & Engineering, University of Asia Pacific,
Dhaka, Bangladesh

*tamalsarkarad@gmail.com

Abstract—**This paper presents photonic crystal fibers (PCF)based on hexagonal and rectangular structures designed to be operated in a range of wavelength 1.3 to 1.7μm and effects of various parameters on bending loss of the different structured index guiding PCFs have been investigated. According to simulation, bending loss is reduced down to 10^{-12}dB/km from 10^{-5}dB/km by changing bending radius and core diameters.Due to having low bending loss, these PCF structures can be used in applicationsof optical technologies including telecommunication and sensing applications.**

Index Terms—**Photonic crystal fiber (PCF), Bending loss, Bending radius, Core diameter, Wavelength.**

I. Introduction

Photonic crystal fiber has drawn emerging attention due to its attractive features such as wideband dispersion flattened characteristics, high birefringence, and low confinement loss and bending loss in transmission. Bend losses are a frequently encountered problem in fiber optics: optical fibers exhibit additional propagation losses when they are bent. Typically, these losses rise very quickly once a certain critical bend radius is reached. This critical radius can be very small (a few millimeters) for fibers with robust guiding characteristics (high numerical aperture), whereas it is much larger (often tens of centimeters) for single-mode fibers with large mode areas.Generally, bend losses increase strongly for longer wavelengths, although the wavelength dependence is often strongly oscillatory due tointerference with light reflected at the cladding/coating boundary, and/or at the outer coating surface. The increasing bend losses at longer wavelengths often limit the usable wavelength range of a single-mode fiber.Photonic crystal fibers can have very low bend losses even far beyond the single-mode cut-off wavelength. For estimating the magnitude of bend loss, the equivalent index method [1] can be used. The basic idea behind this technique is to calculate the mode distributions for an effective index which contains a term accounting the modified path lengths at different transverse positions. This is convenient and usually a good approximation, provided that there is no light reflected e.g. from the outer cladding surface back to the fiber core.The magnitude of bend losses has some dependence on the polarization[2-3].Bending not only introduces losses, but can also reduce the effective mode area. This is particularly true for large mode area step-index fibers. Also bending induces birefringence [4-5].

We, in this paper, propose two PCF structures with low bend loss ranges between in the order of 10^{-12} dB/km to the order of 10^{-5} dB/km. A simple five-ring Hexagonal PCF (H-PCF) is investigated and found to have comparatively higher bend-sensitivity in terms of bend loss dependent on wavelength(1.3-1.7μm), bend radius(1-7mm) and core diameter(1-8μm) firstly. Moreover, The Rectangular PCF (R-PCF) with five rings are developed and its propagation loss properties due to curvature are assessed using Finite Element Method (FEM) and a less sensitive structure is found in terms of bending effects on different guiding properties.The comparison job among the proposed structures was carried out in terms of bend loss and the effects on different parameters on it to find a bend-insensitive, low loss photonic crystal fiber.

II. Design Methodology

A. Hexagonal Structure

Fig.1 shows the simple geometry of the H-PCF with air-hole diameter, *d* and pitch, Λ between the adjacent air-hole rings. Pitch, Λ is the spacing between air hole centers on the adjacent rings and the spacing between air holes centers on the same ring is Λ_1. Pitch Λ_1 is related to Λ by the relation $\Lambda_1 \approx 0.866\Lambda$. Hexagonal PCF structure has isosceles triangular unit lattices [1] with vertex angle of 60^0. Owing to isosceles triangular unit lattices, it contains more air-holes in the cladding region for the same number of rings compared to conventional PCFs. The number of air-holes of an octagonal structure for rings 1, 2, 3, 4, 5 is, respectively, 6, 12, 18, 24, and 30. This results in a higher air-filling ratio and a lower refractive index around the core, thereby providing strong light confinement ability.

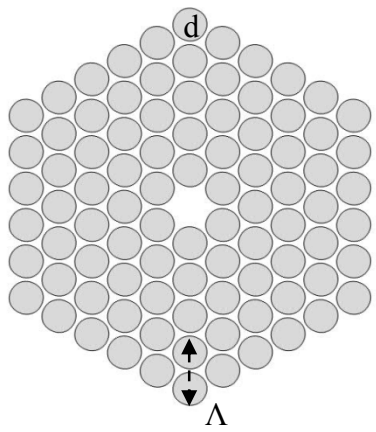


Fig. 1. Geometry of the proposed five-ring H-PCF with air-hole diameter *d*,, air-hole pitch Λ and number of ring, $N_r = 5$

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

The core diameter is A, where 'A' equals 2(Λ-*d*/2). The air-holes of diameter d are located at each corner of the isosceles triangle resulting lower refractive index due to higher air-filling fractions (AFFs) around the core.

B. *Rectangular Structure*

Fig. 2 shows the simple geometry of the Rectangular-PCF with air-hole diameter, *d* and pitch, Λ between the adjacent air-hole rings. Pitch, Λ is the spacing between air hole centers on the adjacent rings and the spacing between air holes centers on the same ring is Λ_1. Pitch Λ_1 is related to Λ by the relation $\Lambda_1 \approx 0.866\Lambda$.The number of air-holes of an rectangular structure for rings 1, 2, 3, 4, 5 is, respectively, 12, 20, 28,36 and 44. This results in a higher air-filling ratio and a lower refractive index around the core, thereby providing strong light confinement ability. The core diameter is 2*a*, where '*a*' equals Λ-*d*/2. The air-holes of diameter d are located at each corner of the isosceles triangle resulting lower refractive index due to higher air-filling fractions (AFFs) around the core.

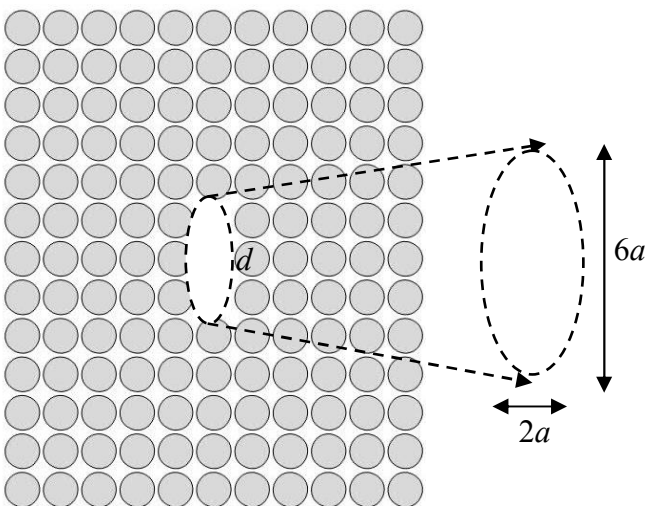


Fig. 2. Geometry of the proposed five-ring Rectangular-PCF with air-hole diameter *d*, air-hole pitch Λ .

III. Numerical Methodology

The COMSOL software 4.2 version is used as a simulationtool.Finite element method (FEM) with perfectly matched layers (PML) boundary condition is used to calculate the chromatic dispersion, effective area and confinement loss of the proposed PCFs.The FEM directly solves the Maxwell equations to best approximate the value of the effective refractive index. Various theoretical methods exist to predict curvature (bending) loss in optical fibers.The usual approach is to use a simplified formula introduced by Marcuse [1], applicable to weakly guided fibers,for sufficiently large radii of curvature. This formula agrees well with experiment of single mode fiber, after adjustments are made for bend-induced stress. Bent fiber simulations were performed using a finite difference approach, the beam propagtion method (BPM) , in conjuction with conformal mapping technique. The first step in the process was to transform the circularly curved fiber to an equivalent, straight fiber by the process of conformal mapping [6]. This process is schematically in Fig. 3. Co-ordinate transformation allowed the bent fiber to be represented by an equivalent, straight fiber, with modified refractive index disribution, $n'(x, y)$

$$n'(x, y) = n(x, y) \exp\left(\frac{x}{R}\right) \quad (1)$$

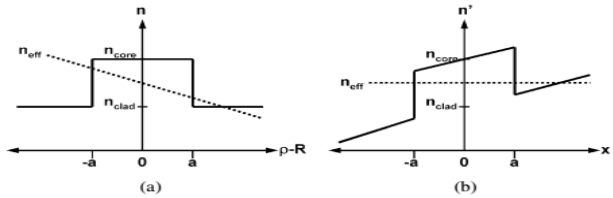


Fig. 3. Refractive index distribution of an unstressed,bent fiber and (b)that of its equivalent, straight fiber after conformal mapping.

Once the modal effective refractive index, n_{eff} is obtained by solving an eigen value problem drawn from Maxwell equations using the COMSOL software 4.2,The bend loss L_c is obtained from the imaginary part ofn_{eff} as follows [7]

$$Lc = 8.686 \times k_0 \mathrm{Im}[n'(x, y)] \quad (2)$$

Where, $\mathrm{Im}[n_{eff}]$ is the imaginary part of the refractive index, $k_0 = 2\pi/\lambda$is the wave number in the free space.

IV. Simulation Results and Discussion

A. Bending Radius Effects on Bend Loss

When fiber is bent, a propagation loss takes place due to the bend in optical fiber. The smaller the bending radius, the greater is the bending loss [8]. Fig. 4 shows the bending loss characteristics of x polarization for optimum design parameters with d/Λ=0.86 and pitch Λ=1μm. In the figure ,the bending loss is calculated by varying the bending radius from 1mm to 7mm.The amount of loss significantly decreases for the radius change from 1mm to 2mm and then the rate of fall slows down for further increase in bending radius [9].

B. Core Diameter Influence on Bending Loss

The bending loss is seen to stabilize with the increasing core diameter. As the core diameter hence Mode Field Diameter (MFD) increases, the loss due to bending seems to be decreased to a lower value and gets a stable value onwards. Fig. 5 illustrates the bending loss characteristics of x polarization for optimum design parameters with d/Λ=0.86 and bending radius 2mm and wavelength 1.55 μm. The above figure describes the bend induced loss change with the change in pitch Λ from 1μm to 4 μm.

C. Wavelength Dependence of Bending Loss

The bending loss increases considerably with the increase in wavelength of the propagating light through the fiber. In the telecom window, the losses are small, but increasing wavelength causes greater losses in H-PCF. The above figure depicts the bending loss characteristics of x polarization for optimum design parameters with d/Λ=0.86 and bending radius 2mm and pitch Λ=1μm. The bending loss or curvature loss rises with the rise in wavelength from 1.3μm to 1.7μm.

D.The Effects of Various Parameters on Bend Lossin H-PCF:

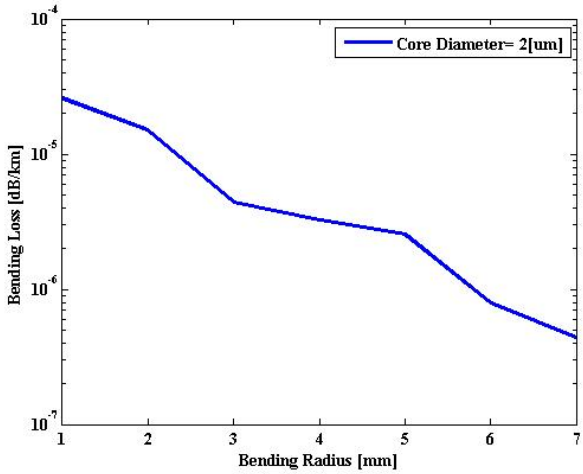


Fig. 4. Bending radius dependent bending loss curve in H-PCF.

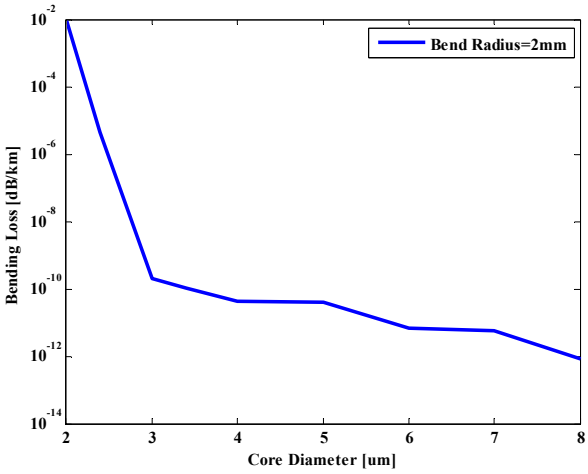


Fig. 5. Effect of Core diameter extension on Bending Loss in H-PCF.

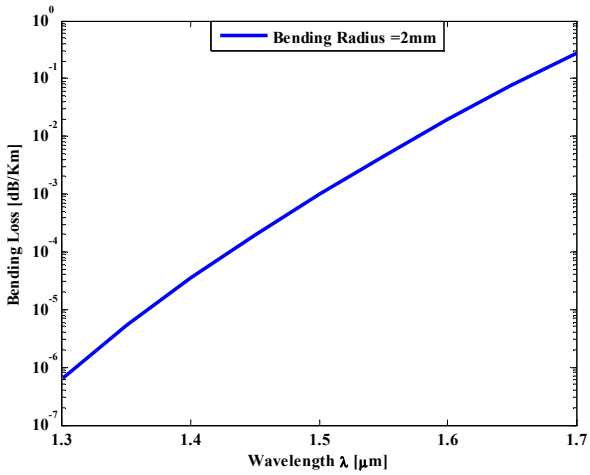


Fig. 6. Wavelength dependence of Bending loss in H-PCF.

E. The Effects of Various Parameters on Bend Loss in R-PCF:

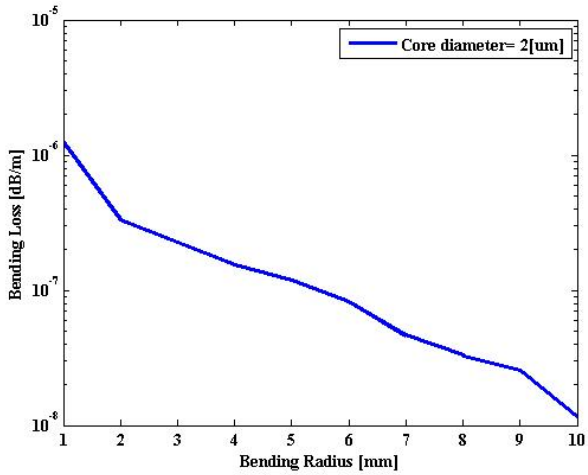


Fig. 7. Bending radius dependent bending loss curve in R-PCF.

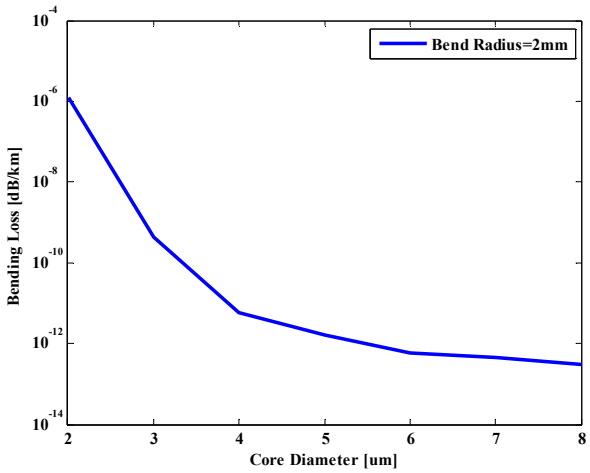


Fig. 8. Effect of Core diameter extension on Bending Loss in H-PCF.

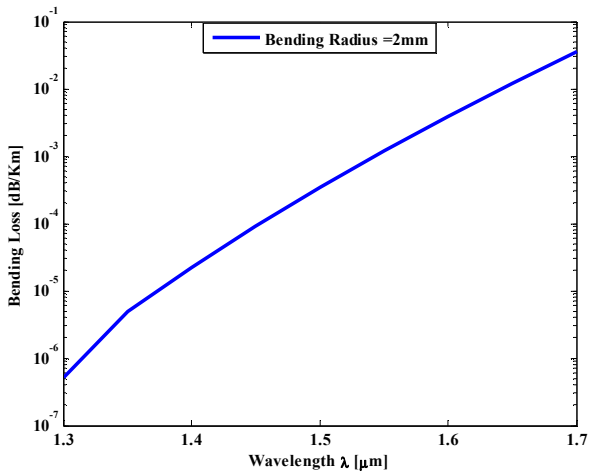


Fig. 9. Wavelength dependence of Bending loss in R-PCF.

Fig. 10 illustrates the curvatureloss characteristics of proposed photonic crystal fibers with design parameters as d/Λ=0.86 and pitch Λ=1μm and wavelength 1.55 μm. The bending radius is varied from 1mm to 7mm and corresponding responses of different proposed PCF structures were plotted on the curve.The H-PCF and the Rectangular PCF shows almost a flat loss response through the entire range of bending radius under consideration. The bend loss in H-PCF is higher than that of the Rectangular PCF all through.Fig. 11 presents the core diameter effects on bending loss of various photonic crystal fibers having design parameters d/Λ=0.86 and bending radius R=2mm.The bend loss in H-PCF is the largest and the lowest loss is obtained from the Rectangular structure shows a consistent inverse exponential loss curve with the increase in core diameter. It is observed that bend loss associated with the lowest core diameter is the highest and more than 1dB/km for H-PCF and it decreases to the order of 10^{-6} in H-PCF. Fig. 12 depicts the bending loss behaviour of different photonic crystal fibers with optimum design parameters with d/Λ=0.86 and pitch Λ=1μm and bending radius R=2mm.The bend loss in H-PCF is the largest among the proposed structures and the lowest loss is obtained from the Rectangular structure.. It is observed that bend loss associated with the lowest wavelength is the lowest and of 10^{-6} order and it increases up to 0.1dB/km for H-PCF which is the highest at wavelength 1.7 μm while lowest loss at this point is associated with Rectangular PCF and is of 10^{-2} order.

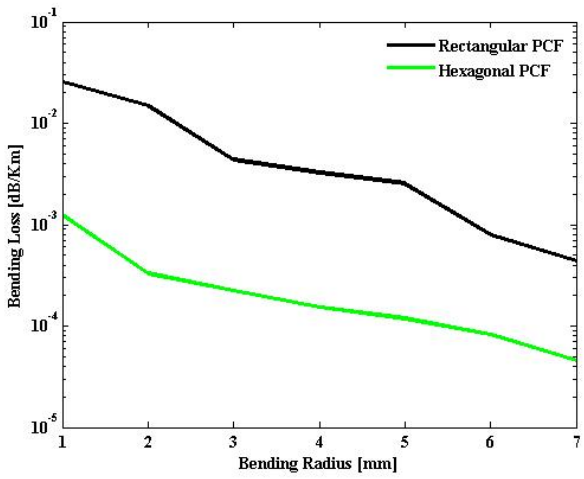


Fig. 10. Comparison of different PCF structures in terms of bending loss induced by a certain range of bend radius of the fibers.

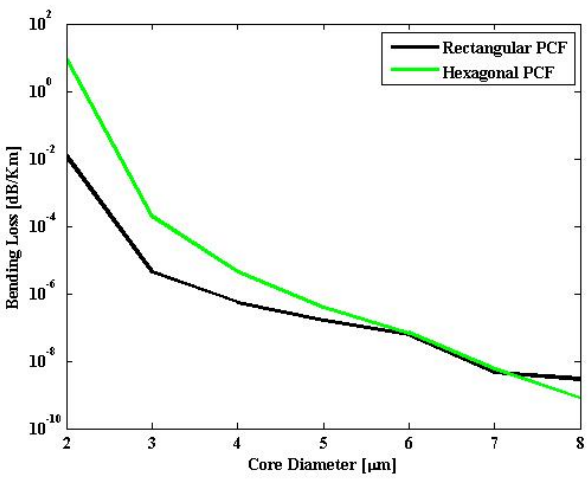


Fig. 11. Comparison of bending loss of proposed PCF structures influenced by core diameter effects on fibers.

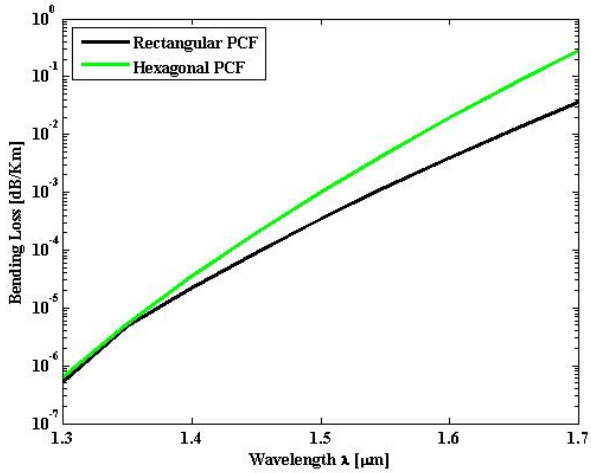


Fig. 12. Comparison of various PCF structures considering the effects of wavelength on bend loss .

V. Conclusion

In summary, we have proposed that afivering Hexagonal PCF (H-PCF) is found to have comparatively higher bend-sensitivity in terms of bend loss of the order 10^{-5} dB/km which is dependent on wavelength, bend radius and core diameter. The Rectangular PCF (R-PCF) with five rings are developed and its propagation loss properties due to curvature are assessed and observed to be a less sensitive structure in terms of bending effects on different guiding features with bend losses of the order 10^{-12} dB/km.

References

[1] D. Marcuse, "*Curvature loss formula for optical fibers*", J. Opt. Soc. Am. 66 (3), 216 (1976).

[2] D. Marcuse, "*Field deformation and loss caused by curvature of optical fibers*", J. Opt. Soc. Am. 66 (4), 311 (1976).

[3] D. Marcuse, "*Influence of curvature on the losses of doubly clad fibers*", Appl. Opt. 21 (23), 4208 (1982).

[4] R. Ulrich et al., "*Bending-induced birefringence in single-mode fibers*", Opt. Lett.5 (6), 273 (1980).

[5] S. J. Garth, "Birefringence in bent single-mode fibers", J. Lightwave Technology. 6, 445 (1988).

[6] R. T. Schermer and James H Cole, "*Improved bend loss formula verified for optical fiber by simulation and experiment",* Quantum Electronics, IEEE Journal of (Volume:43 , Issue: 10)

[7] K. Saitoh, M. Koshiba, T. Hasegawa, E. Sasaoka, "*Chromatic dispersion control in photonic crystal fibers: application to ultra-flattened dispersion,"* Opt. Express, vol. 11, no. 8, pp.843-852, April 2003.

[8] HuseyinAdemgil, ShyqyriHaxha, Fathi Abdel Malek , "*Highly nonlinear Bending-Insensitive Birefringent Photonic Crystal Fibres*", Scientific Research Engineering, 2010, 2,page 608-616.

[9] Ngoc Hai Vu, In-Kag Hwang and Yong-Hee Lee, "*Bending loss analyses of photonic crystal fibers based on finite difference time-domain method*", Optics letters, January 15,2008,vol. 33, no.2.

Highly Birefringent Modified Circular Photonic Crystal Fiber for Dispersion Management over E+S+C+L Wavelength Bands

M. Mejbaul Haque, [1,*] M. Shaifur Rahman,[1] M. Selim Habib,[2] and M. Samiul Habib[2]

[1]Department of Electrical and Electronic Engineering, Khulna University of Engineering & Technology
Khulna 9203, Bangladesh

[2] Department of Electrical and Electronic Engineering, Rajshahi University of Engineering & Technology
Rajshahi 6204, Bangladesh

[*]mmhaque.eee@gmail.com

Abstract— **In this paper, we propose a highly birefringent modified circular photonic crystal fiber (M-CPCF). The proposed structure is found extremely suitable for chromatic dispersion compensation of standard single mode fiber (SMF) over E+S+C+L wavelength bands. Finite element method (FEM) with perfectly matched layer boundary condition has been applied to explore its guiding properties. Simulation results confirm the possibility of large negative dispersion coefficient and relative dispersion slope (RDS) of -650 ps/(nm.km) and 0.0036 nm^{-1} respectively at 1550 nm wavelength and effective dispersion coefficient of about ± 0.5 ps/(nm.km) from 1360 to 1640 nm wavelength. The proposed fiber also demonstrates a high birefringence of order 2.1×10-2 that allows the fiber to maintain a single polarization. The proposed fiber can be a promising candidate in high speed transmission system for broadband dispersion compensation and sensing applications as well.**

Index Terms—**Chromatic dispersion, Birefringence, Effective dispersion, Dispersion compensation, Photonic crystal fiber, Finite element method.**

I. INTRODUCTION

Photonic crystal fiber (PCF) consists of microscopic array of air channels running down their length that make a low index cladding around the undoped silica core [1]. It is possible to control the optical properties of PCF with this type of structure which is not achievable with conventional fiber structure [2]. Thus, PCF offers great flexibility in tuning dispersion [3]-[4] which is significant in designing dispersion compensating fiber. Recently wavelength division multiplexing (WDM) technique is being adopted to handle higher bit rate transmission system of more than 40 Gbps. However, the performance of optical WDM system is affected by the transmission impairment caused by dispersion in standard transmission fiber. This necessitates dispersion compensation to improve the quality of WDM system. Presently, dispersion compensating fibers (DCFs) is widely used and commercially available for dispersion compensation [5]-[7]. DCFs should have large negative dispersion to compensate the accumulated dispersion encountered into the single mode fiber (SMF) used for WDM transmission systems. At the same time, the RDS of DCF must be equal to SMF which is 0.0036 nm^{-1} at the operating frequency of 1550 nm. Hence, compensation of accumulated dispersion and dispersion slope of conventional fibers are simultaneously required for broadband applications. To reduce the cost and minimize the confinement loss, the DCF should be as short as possible. However, it is quite difficult to obtain high negative dispersion over a wide band of wavelength using conventional DCFs. In addition, conventional DCFs have some limitations related to their structure, for instance, materials of different thermal expansion coefficients (i.e., germanium-doped core and silica cladding regions) [8]. So, the key issue is to design a DCF which simultaneously ensures large negative dispersion over a wide band. Therefore, in designing DCFs, it is also important to consider other properties like dispersion slope, RDS, confinement loss, and birefringence [8]. On the other hand, high negative dispersion and high birefringence is very essential for a PCF to be applicable in broadband dispersion compensation and sensing application as well. PCF is first proposed in [9] for broadband dispersion compensation. However, the design experiences a small effective area and raises complexity in coupling light into the fiber. A dual core PCF with highly doped internal core is proposed in [10] for dispersion compensation which achieves a large dispersion peak of about □59,000 ps/(nm.km) and hence the available bandwidth for dispersion compensation becomes narrow. In addition, doping the internal core increases confinement loss and makes the fabrication process more difficult. M-OPCF designed with six air hole rings is proposed in [11] which obtains a negative dispersion of −239.5 ps/(nm.km) at 1550 nm but the RDS does not match accurately with the RDS of single mode fiber. In recent times, several PCF designs are proposed for addressing the issue of broadband dispersion compensation and sensing simultaneously. For example, square lattice photonic crystal fiber is proposed in [12] which provides unmatched RDS of 0.0034 nm^{-1} and insufficient negative dispersion of only −204.4 ps/(nm.km) requiring longer fiber to compensate the accumulated dispersion. A highly birefringent broadband dispersion compensating octagonal PCF is proposed in [8] shows a better negative dispersion coefficient of −400 to −725 ps/(nm km) over S to L-bands with perfect slope matching and a birefringence of 1.8×10^{-2}. Furthermore, a spiral microstructure optical fiber is proposed in [13] which successfully achieves negative dispersion coefficient of about −327 ps/(nm.km) at 1550 nm wavelength with high birefringence of about 1.79×10^{-2}.

In this work, we propose highly birefringent M-CPCF for dispersion management over E+S+C+L wavelength bands. The proposed M-CPCF structure shows a high negative

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

dispersion over the wavelength ranging from 1360 to 1640 nm and particularly −650 ps/(nm.km) with RDS of 0.0036 nm^{-1} at 1550 nm wavelength. The effective dispersion is obtained within ± 0.5 ps/(nm.km) over E+S+C+L wavelength bands. Moreover, the proposed M-CPCF shows a high birefringence of about 2.1×10^{-2} which is also suitable for sensing applications. Confinement loss is also found 0.012315 dB/m at 1550 nm wavelength. It is highly expected that the proposed M-CPCF should be greatly applicable in WDM communication system and polarization maintaining application like sensing applications.

II. Design Guidelines of the Proposed M-CPCF

Fig. 1 depicts the transverse cross section of the proposed HyC-CPCF that has five air-hole rings. The cladding is formed in combination with the circular and elliptical air holes. The symmetry of the core is intentionally broken by removing two circular air-holes while replacing those with elliptical air-holes. However, some extra elliptical air-holes are added into the cladding region to have better dispersion properties and fiber nonlinearity. This design is said to be modified circular in the sense that it consists of air-holes those comes from different shape i.e. circular and elliptical in this design. The material of the proposed structure is pure silica. The total number of air-holes for 1st, 2nd, 3rd, 4th, and 5th rings are respectively 6, 8, 12, 24 and 48. To achieve this number of air-holes, the air-holes on the first ring are rotated at an angle 60^0 while air-holes on the 2nd to 5th rings are rotated at an angle 45^0, 30^0, 15^0, and 7.5^0 respectively. The air-hole diameter of 1st, 3rd and 4th ring is d_1 while air-hole diameter of the 2nd and 5th ring is selected as d_2 and d_3 respectively. d_s and d_f are the minor axis and major axis lengths of the elliptical air-holes respectively. The proposed M-CPCF structure produces a higher degree of freedom regarding the total number of geometrical design parameters namely, pitch, □, air-hole diameters ($d_1$to d_3) and ellipticity, η. With the precise control of these design parameters, optimal fiber transmission properties such as high negative dispersion coefficient, low confinement loss, and high birefringence is possible to achieve simultaneously in particular frequency regimes.

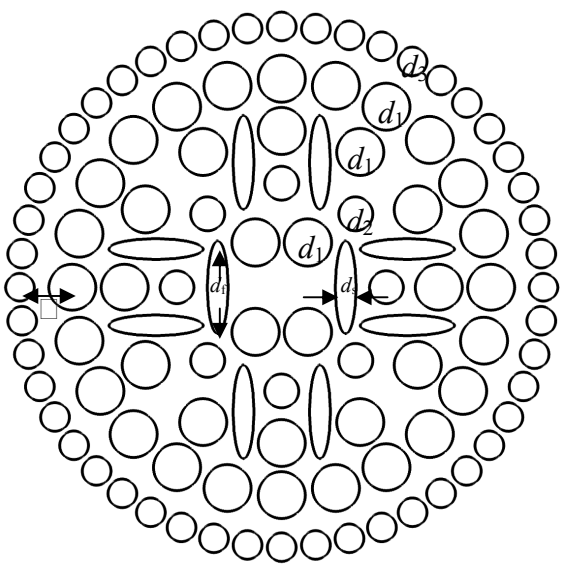


Fig. 1 Transverse cross section of the proposed HyC-CFCF structure showing pitch, □, air-hole channel with diameters, d_1, d_2 and d_3 and ellipse with dimension d_s and d_f.

III. Numerical Method

The finite element method (FEM) perfectly matched layer boundary condition is applied for numerical simulation to explore the guiding properties of proposed HyC-CPCF structure. Commercial full vector finite element software (COMSOL) with first order triangular vector edge elements is used to evaluate the value of modal effective refractive index n_{eff} . The wavelength dependent refractive index of optical material (silica) is calculated using three-term Sellmeier equation and directly included in the simulation. Once the modal effective index n_{eff} is obtained, dispersion $D(\lambda)$, effective dispersion D_e , confinement losses L_c , birefringence B and effective area, A_{eff} can be determined by the using the following equations (1) − (4) [13]-[15].

$$D(\lambda) = -\frac{\lambda}{c}\frac{d^2\,\mathrm{Re}[n_{eff}]}{d\lambda^2} \quad \text{ps/(nm.km)} \tag{1}$$

$$L_c = 8.686\times\frac{2\pi}{\lambda}\times\mathrm{Im}[n_{eff}] \quad \text{dB/m} \tag{2}$$

$$B = \left|n_x - n_y\right| \tag{3}$$

$$A_{eff} = \frac{\left(\iint |E|^2 dxdy\right)^2}{\iint |E|^4 dxdy} \quad \mu\text{m}^2 \tag{4}$$

Eq. (1) represents the total chromatic dispersion since material dispersion is directly included in the FEM calculation process with the help of Sellmeier equation.

IV. Numerical Results and Discussion

The designed M-CPCF offers design flexibility in tailoring the chromatic dispersion, birefringence, nonlinearity and confinement loss which is required to achieve application specific properties from the fiber. At first, the wavelength dependence nature of chromatic dispersion for optimized parameters □=1.0 µm, d_1=0.95 µm, d_2=70 µm, d_3=60 µm, η=0.22 µm is investigated and shown for both x and y polarized optical fields in Fig. 2. A negative dispersion coefficient of -224 and -650 ps/(nm.km) at operating wavelength of 1550 nm are observed for x and y polarization mode respectively. The y-polarized mode is about three times higher in magnitude than x polarized mode which represents fast axis mode.

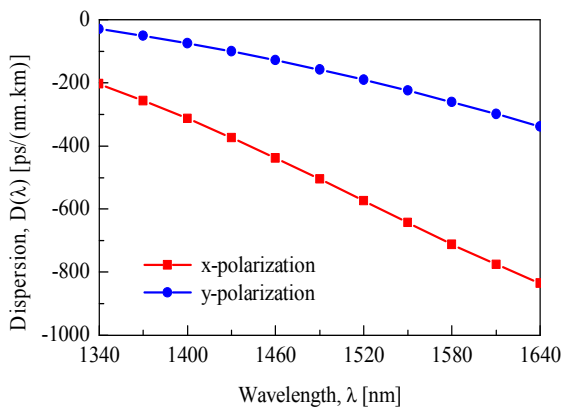


Fig. 2 Chromatic dispersion of the proposed HyC-CPCF as a function of wavelength, λ for x- and y-polarization.

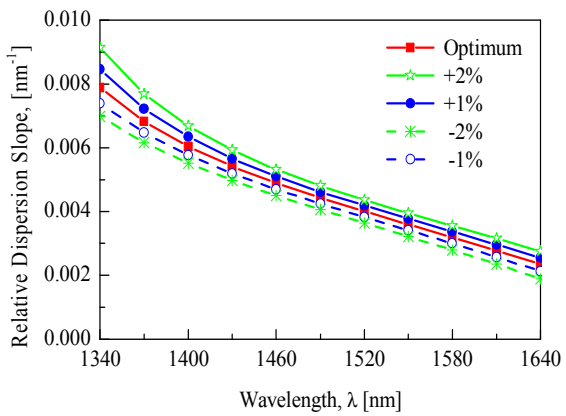


Fig. 3 RDS of the proposed HyC-CPCF for optimum design parameters and variation of pitch, □.

Hence y-polarized mode is considered to calculate other wavelength dependent parameters that include residual dispersion, residual dispersion slope (RDS), and confinement loss. However, the designed fiber shows a very high negative dispersion coefficient of about -204 to -836 ps/(nm.km) over 1340 to 1640 nm wavelength band for optimal design parameters. The high negative dispersion coefficient for the proposed M-CPCF particularly at 1550 nm wavelength is about -650 ps/(nm.km) which is higher than [8]. High negative dispersion is a key parameter in designing the fiber that should be as large as possible especially to reduce the length of dispersion compensation fiber. At the same time, the RDS profile of the designed fiber should be properly matched with the SMF for dispersion compensation application that requires a RDS of 0.0036 nm^{-1} at 1550 nm wavelength. Therefore, it is important to take into consideration the dispersion and dispersion slope simultaneously in designing PCF for dispersion compensation and also for parameter optimization [16].

RDS is shown in Fig. 3 for optimum design parameters as a function of wavelength and also by tuning the pitch, □. The RDS is exactly found 0.0036 nm-1 at 1550 nm for optimum design parameters. However, no sudden change in RDS profile is noticed over the entire band of interest.

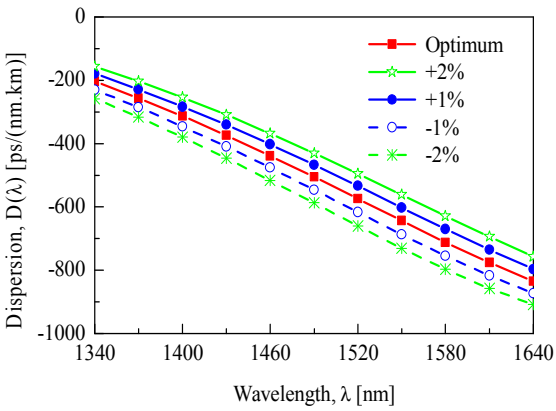


Fig. 4 Dispersion curve of the proposed HyC-CPCF for optimum design parameters and fiber's global diameter variation of order ±2%.

In a standard fiber draw, 2% variation in fiber's global diameter may occur during the fabrication process which may affect the dispersion accuracy [16]. After shaping the optimum dispersion curve and RDS profile, global diameter is varied up to 2% to check the dispersion accuracy about the optimum parameters. Fig. 4 illustrates the dispersion curve for optimum design parameters and due to fiber's global diameter variation. The optimum dispersion figure at operating wavelength of 1550 nm is -650 ps/(nm.km). The dispersion accuracy of the fiber up to ±2% variation in fiber's global diameter is found to be ±86 ps/(nm.km) which is less than [16]. So our proposed design can ensure the better dispersion accuracy.

Fig. 5 illustrates the effective dispersion curve of an optimized 1.04 km long M-CPCF as a function of wavelength after compensating the accumulated dispersion of 40 km long standard SMF used as a transmission fiber. To compensate dispersion for a high bit rate transmission system, for example 40Gbps system, the effective dispersion should not be higher than ±0.8 ps/nm/km [8]. Fig. 5 clearly shows that the effective dispersion is lower than ±0.8 ps/(nm.km) over E +S+C+L wavelength bands which gives an indication that our proposed design is enough capable to compensate accumulated dispersion of transmission fiber over the E to L wavelength bands. The most important feature of the proposed M-CPCF is that it provides two zero dispersion at 1430 and 1550 nm wavelength respectively. It also offers more bandwidth for dispersion compensation than [8] and the required length is smaller than [16]. The design is thus suitable for high bit rates transmission system.

We have then carefully investigated the birefringence property for optimum parameters in the wavelength range 1340-1640 nm. According to simulation, the birefringence of the proposed fiber seems to follow almost linear shape that increases with increase in wavelength as shown in Fig. 6.

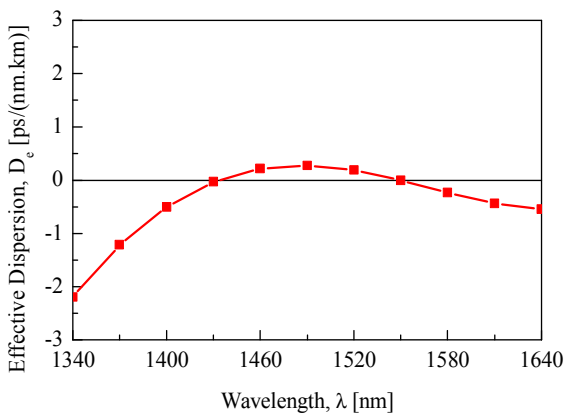


Fig. 5 Effective dispersion curve of the proposed HyC-CPCF for optimum design parameters against wavelength.

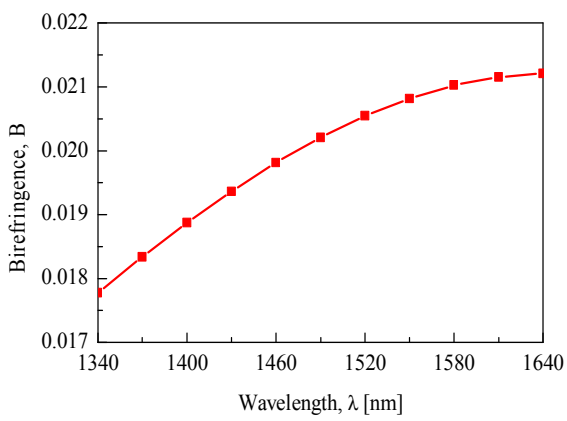


Fig. 6 Birefringence property of the proposed M-CPCF for the optimum design parameters.

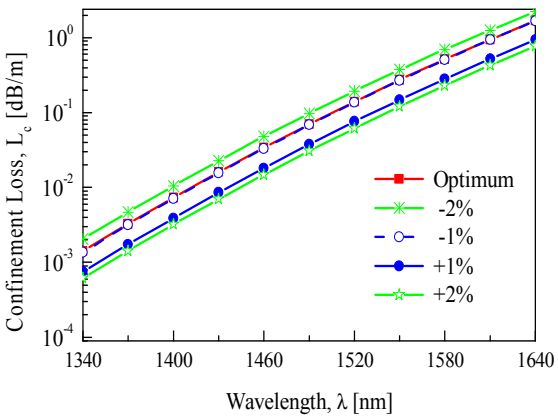


Fig. 7 Confinement loss of the proposed M-CPCF for the optimum design parameters.

Particularly, the birefringence is 0.021 at operating wavelength of 1550 nm which is 42 times higher than that of the conventional PM optical fiber. This indicates that the proposed M-CPCF can provide high birefringence which is very important and highly essential to eliminate the effect of PMD in optical transmission system. In addition to this, high birefringence is desirable in many other applications such as sensing application. For example, highly birefringent fibers are extensively used in fiber loop mirrors as a major component for optical fiber sensing applications. The high birefringence with high negative dispersion of the proposed M-CPCF would be attractive for optical fiber sensor design, optical amplification system and long distance data transmission systems as well.

Finally, we have studied the confinement loss against wavelength for optimum design parameters and also for fiber's global diameter variation. Fig. 7 shows the confinement loss of the proposed M-CPCF. It is clearly seen that the confinement loss is increased with wavelength and particularly found 0.012315 dB/m at 1550 nm wavelength.

V. Conclusion

This paper has presented M-CPCF based on FEM to concurrently ensure high negative dispersion, high birefringence those are crucial for PCF to be applicable in broadband dispersion compensation and sensing applications. The proposed M-CPCF gives a high negative dispersion of -650 ps/(nm.km) with a perfect RDS of 0.0036 nm-1 at 1550 nm wavelength and also provides an effective dispersion of about ± 0.5 ps/(nm.km) from 1360 to 1640 nm wavelength. This feature makes the designed M-CPCF appropriate for broadband dispersion compensation applications. In addition, the designed M-CPCF has shown high birefringence in the order of 2.1×10-2 at 1550 nm wavelength and thus has the potential to be suitable for PMD compensation and sensing applications.

References

[1] J. C. knight, "Photonic crystal fibers," *Nature* , vol. 424, pp. 847–851, 2003.

[2] J. C. Knight, T. A. Birks, P. S. J. Russell, and D. M. Atkin, "All-silica single-mode optical fiber with photonic crystal cladding," *Optics Letters*, vol. 21, pp. 1547–1549, 1996.

[3] S. M. A. Razzak, and Y. Namihira, "Proposal for highly nonlinear dispersion-flattened octagonal photonic crystal fibers," *IEEE Photonics Technology Letters*, vol. 20, pp. 249–251, 2008.

[4] K. Saitoh, M. Koshiba, T. Hasegawa, and E. Sasaoka, "Chromatic dispersion control in photonic crystal fibers: application to ultra-flattened dispersion," *Optics Express*, vol. 11, pp. 843–852, 2003.

[5] L. G. Nielsen, S. N. Knudsen, B. Edvold, T. Veng, D. Magnussen, C. C. Larsen, and H. Damsgaard, "Dispersion compensating fibers," *Optical Fiber Technology*, vol. 6, pp. 164–180, 2000.

[6] L. P. Shen, W. P. Huang, and S. S. Jian, "Design and optimization of photonic crystal fibers for broadband dispersion compensation," *IEEE Photonics Technology Letters*, vol. 15, pp. 540, 2003.

[7] M. Koshiba, and K. Saitoh, "Structural dependence of effective area and mode field diameter for holey fibers," *Optics Express*, vol. 11, pp. 1746–1756, 2003.

[8] Md. Selim Habib, Md. Samiul Habib, S. M. Abdur Razzak, and Md. Anwar Hossain, "Proposal for highly birefringent broadband dispersion compensating octagonal photonic crystal fiber," *Optical Fiber Technology*, vol. 19, pp. 461–467, 2013.

[9] T. A. Birks, D. Mogilevtsev, J.C. Knight, and P. St. J Russell, "Dispersion compensation using single-material fibers," *IEEE Photonics Technology Letters*, vol. 11, pp. 674–676, 1999.

[10] A. Huttunen, and P. Torma, "Optimization of dual-core and microstructure fiber geometries for dispersion compensation and large mode area," *Optics Express*, vol. 13, pp. 627–635, 2005.

[11] S. F. Kaijage, Y. Namihira, N. H. Hai, F. Begum, S. M. A. Razzak, T. Kinjo, K. Miyagi, and N. Zou, "Broadband dispersion compensating octagonal photonic crystal fiber for optical communication applications," *Japanese Journal of Applied Physics*, vol. 48, pp. 052401–052408, 2009.

[12] N. Ehteshami, and V. Sathi, "A novel broadband dispersion compensating square-lattice photonic crystal fiber," *Optical and Quantum Electronics*, vol. 44 ,pp. 323–335, 2012.

[13] Md. Samiul Habib, K. M. Nasim, Md. Selim Habib, Md. Imran Hasan, and Redwan Ahmad, "Relative dispersion slope matched dispersion compensating highly birefringent spiral microstructure optical fibers using defected core," *Optical Engineering*, vol. 52, pp. 096110, 2013.

[14] K. Saitoh, and M. Koshiba, "Full-vectorial imaginary-distance beam propagation method based on a finite element scheme: application to photonic crystal fibers," *IEEE Journal of Quantum Electronics*, vol. 38, pp. 927–933, 2002.

[15] M. Selim Habib, M. Samiul Habib, S. M. A. Razzak, Y. Namihira, M. A. Hossain, and M. A. Goffar Khan, "Broadband dispersion compensation of conventional single mode fibers using microstructure optical fibers," *Optik- International Journal for Light and Electron Optics*, vol. 124, pp. 3851–3855, 2013.

[16] Md. Imran Hasan, M. Samiul Habib, M. Selim Habib, and S. M. A. Razzak, "Design of hybrid photonic crystal fiber: Polarization and dispersion properties," *Photonics and Nanostructure- Fundamentals and Applications*, vol.12, pp. 205-211, 2014.

Formation of Micro/Nano-metric Holes/voids in Al_2O_3-coated Boro-aluminosilicate Glass by Single Pulse Femtosecond Laser Irradiation

Md. Shamim Ahsan,[1,*] Man Seop Lee,[2] and Ragib Shakil Rafi[1]

[1]Electronics and Communication Engineering Discipline, Khulna University
Khulna 9208, Bangladesh

[2]Department of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST)
373-1 Guseong-dong, Yuseong-gu, Daejeon 305-701, Republic of Korea.
*shamim@kaist.ac.kr

Abstract—**In this paper, we report the formation of periodic array of micro/nano-scale holes/voids on the surface of and inside Al_2O_3-coated boro-aluminosilicate glass by means of single pulse femtosecond laser irradiation. The lowest diameter of the nano-holes printed on the Al_2O_3 coating reaches down to 700 nm, whereas the lowest diameter of the voids inside the glass was 1.1 μm. The fabricated micro/nano-metric holes and voids show uniformity in size and shape throughout the sample area. We also investigate the surface property of the nanovoids' evolved glass surface. We don't observe any structure on the sample surface, when the voids have been evolved inside the glass sample. In addition, we investigate the formation mechanism of the micro/nano-holes or voids in our double-layer samples. We strongly believe that, the periodic micro/nano-voids array will be useful in fabricating various photonic/optical devices, especially, in spreading light.**

Index Terms—**Femtosecond Laser, Al_2O_3-coated boro-aluminosilicate glass, micro-holes/voids, nano-holes/voids.**

I. Introduction

During the last several decades, the demand of forming micro/nano-metric structures in a large variety of materials has increased significantly due to their applications in a diverse field of science and technology. High precision, easy processing steps, and processing time are the important parameters for selecting the patterning technology. Currently, there are several established patterning technologies for micro/nano-structuring of materials: nanoimprint lithography [1], electron beam lithography [2], ion beam lithography [3], photo lithography [4], X-ray lithography [5], physical vapour deposition [6], chemical vapour deposition [7], and laser processing technologies [8–18]. Compared to other patterning technologies, laser sources have shown greater promise for micro/nano-patterning of materials due to their simplicity, high precision, and high throughput. Among various laser sources, femtosecond lasers are widely accepted by the researchers as well as the micro/nano-machining industry because of their non-linear properties and debris free patterning that influence the precise formation of micro/nano-scale structures in various materials [10–19].

Depending on the application area, researchers have proposed the formation of a large variety of micro/nano-structures on the surface of and inside both the transparent and non-transparent materials [11–19]. Formation of micro/nano-holes or voids in transparent materials including glasses has attracted many researchers due to their wide range application area [14–18]. However, it is still challenging to fabricate uniform holes/voids' array with size down to several micrometer to several hundred nanometers precisely in transparent materials. Although it is easy to pattern micro/nano-scale holes/voids on the surface of a single transparent material, it is quite difficult to pattern holes/voids inside double-layer materials without damaging the top layer. Besides, the threshold energy and holes/voids characteristics are solely dependent on the material properties, which vary significantly for different materials. Thus, it is significant to investigate the formation of micro/nano-holes or voids in various transparent materials.

In this paper, we demonstrate for the first time the fabrication of periodic micro/nano-holes or voids' array on the surface of and inside the Al_2O_3-coated boro-aluminosilicate glass samples. Nano-holes' array with diameter down to 700 nm have been observed on Al_2O_3 surface, whereas micro-voids with diameter down to 1.1 μm has been reported inside the glass. More importantly, these micro/nano-holes or voids show excellent uniformity throughout the sample area. Furthermore, we analyze the characteristics of the micro/nano-metric holes/voids, printed on the surface of and inside the Al_2O_3 coated boro-aluminosilicate glass samples. Our experimental result shows that, it is possible to fabricate voids array inside the boro-aluminosilicate glass without damaging the top Al_2O_3 layer. In addition, we examine the influence of numerical aperture and magnification of objective lens on the formation of holes/voids in our samples. Furthermore, we analyzed the formation mechanism of the holes/voids in Al_2O_3-coated boro-aluminosilicate glass samples. We believe that, this holes' array in the double-layer structure will be valuable to design various optical devices in spreading light such as LEDs.

II. Experimental Details

We carried out our experiments using a Ti:sapphire femtosecond laser beam of Gaussian shape that operates at the central wavelength of 786 nm. The pulse width and the pulse repetition rate of the laser pulses were 183 fs and 1 kHz. We applied p-polarized laser beam to pattern the samples. The femtosecond laser beam was guided by means of several mirrors and an achromatic objective lens (50× or 100×) onto the samples, placed on the translation stage. The numerical aperture of the 50× and 100× lenses are 0.42 and 0.7. Fig. 1(a) illustrates the schematic diagram of the experimental setup.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

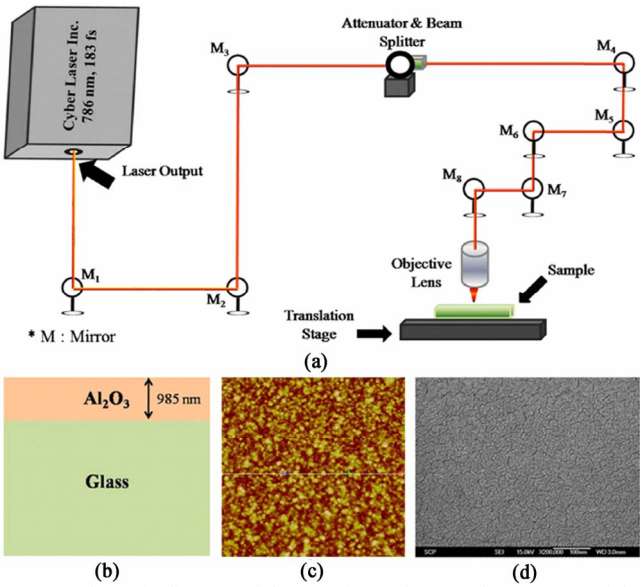


Fig. 1 (a) Schematic diagram of the experimental setup; (b) Structure of the Al_2O_3-coated boro-aluminosilicate glass sample; (c) AFM image of the Al_2O_3 bare surface; (d) SEM image of the boro-aluminosilicate glass surface before Al_2O_3 coating.

During our experiments, we used specially designed Al_2O_3-coated boro-aluminosilicate glass samples (Al_2O_3 thickness: ~ 985 nm; glass thickness: ~ 650 µm). The sample consists of two layers, where the Al_2O_3 layer was coated on the boroaluminosilicate glass surface, which is shown in Fig. 1(b). The atomic force microscope (AFM) image of the Al_2O_3 bare surface is depicted in Fig. 1(c). Using AFM image, we obtain the roughness (R_a) information of the Al_2O_3 bare surface, which is ~ 1.3 nm. To pattern the Al_2O_3-coated glass samples, we mounted them on top of a 3-axis translation stage, where the resolution of the translation stage in the x, y, and z directions were 100 nm. To examine the surface morphology of the material surface, we used a field emission scanning electron microscope (FE-SEM). Before SEM examination, the samples were cleaned in ultrasonic bath in de-ionized water for 30 minutes. The samples were sputter coated with a thin layer (~ 30 nm) of gold coating to provide conductivity for SEM examination. The SEM image of the boro-aluminosilicaate glass surface without the Al_2O_3 coating is shown in Fig. 1(d).

The laser beam was passed through an integrated circular attenuator & beam splitter, where the function of the attenuator was to control the power of the laser beam. The splitter splits the beam into s-polarized and p-polarized laser beams, passes the p-polarized laser beam and reflects the s-polarized laser beam in the perpendicular direction (this beam is blocked). We utilized the p-polarized laser beam to pattern periodic micro/nano-holes or voids' array on the surface of Al_2O_3 or inside boro-aluminosilicate glass. The micro/nano-patterned samples were investigated under an optical microscope. Before microscope examination, the samples were cleaned in ultrasonic bath in de-ionized water for 30 minutes. By controlling the irradiation conditions of the laser beam, we formed periodic holes/voids in our samples.

III. Results and Discussion

A. *Formation of Micro/nano-metric Holes' Array Using 50× Objective Lens*

By illuminating single laser pulse in each spot, we fabricated periodic micro-holes' array on Al_2O_3 surface. The laser beam was focused on the sample surface and scanned at a scanning speed (v) of 2 mm/s with vertical scanning step (a) of 2 µm. At a pulse repetition rate of 1 kHz and a scanning speed of 2 mm/s, the horizontal distance among the holes was 2 µm. On the other hand, the vertical distance among the holes was also 2 µm. For fabricating micro-holes with diameter of 1.7 µm, 1.3 µm, and 900 nm, the energy of the laser beam (E) was 500 nJ, 305 nJ, and 152 nJ, respectively. Figs. 2(a–c) represent the micro/nano-holes fabricated on the Al_2O_3 surfaces using 50× objective lens.

In order to produce nano-voids 3 µm inside the Al_2O_3-coated glass sample, we focused the laser beam inside the glass sample. Nano-voids with diameter down to 800 nm have been evolved inside the glass sample at a laser energy of 250 nJ, which is represented in Fig. 2(d). Because of the energy loss in the Al_2O_3 coating and the boro-aluminosilicate glass, more energy is required when the voids are formed inside the glass sample. If we closely look at the optical microscope images of Fig. 2, we find that the fabricated micro/nano-holes and voids show great consistency in size and shape throughout the sample area.

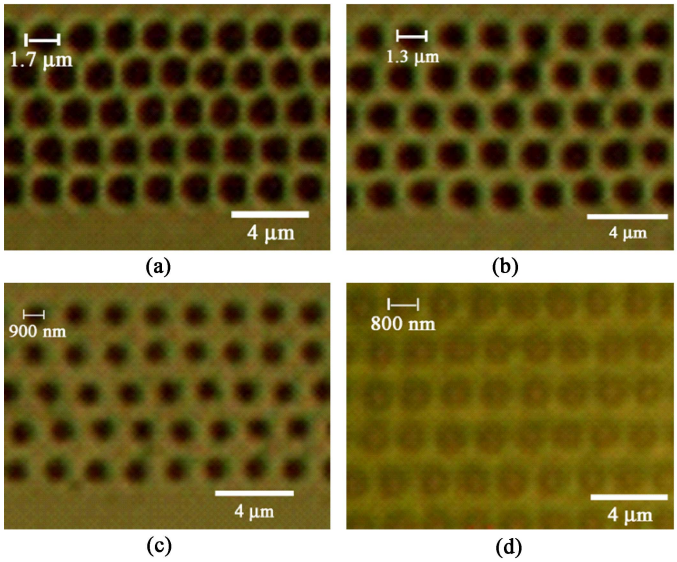
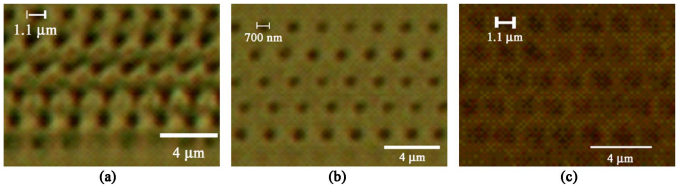


Fig. 2 Micro/nano-holes' array on the Al_2O_3 surfaces using 50× achromatic objective lens at a scanning speed of 2 mm/s and vertical scanning step of 2 µm. (a) E = 500 nJ; (b) E = 305 nJ; (c) E = 152 nJ; and (d) E = 250 nJ.

B. *Formation of Micro/nano-metric Holes/voids' Array Using 100× Objective Lens*

We also printed periodic micro/nano-holes or voids' array on the surface of Al_2O_3 and inside Al_2O_3-coated boro-aluminosilicate glass using 100× achromatic objective lens. To fabricate periodic micro-holes array on the Al_2O_3 surfaces, we focused the laser beam on the Al_2O_3 surface with laser energy of 607 nJ for 1.1 µm diameter holes and 430 nJ for 700 nm diameter holes. Figs. 3(a) and (b) illustrate the optical microscope images of the micro-holes' array on the Al_2O_3 surfaces.

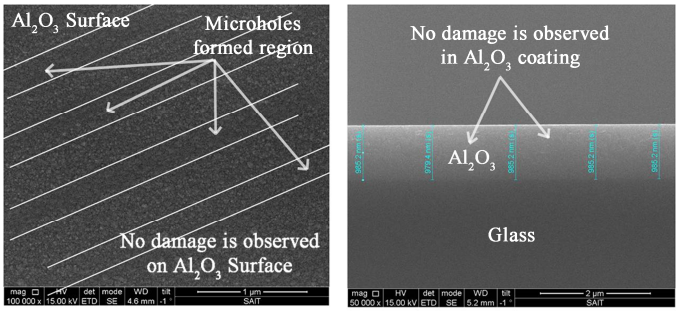
In order to fabricate micro-voids inside Al_2O_3-coated glass, we focused the laser beam 3 µm inside the glass by decreasing the laser energy down to 280 nJ. As a result, micro-voids' array with diameter of 1.1 µm are evolved inside the glass, which is shown in Fig. 3(c). The micro-voids developed inside the glass samples without damaging the top Al_2O_3 layer.



(a) (b) (c)

Fig. 3 Micro/nano-scale holes/voids' array on Al_2O_3 surface and inside Al_2O_3-coated boro-aluminosilicate glass using 100× achromatic objective lens at a scanning speed of 2 mm/s and vertical scanning step of 2 µm. (a) Micro-holes on the Al_2O_3 surface at E = 607 nJ; (b) nano-holes on the Al_2O_3 surface at E = 430 nJ; (c) micro-holes inside the boro-aluminosilicate glass at E = 280 nJ.

We also investigated the micro-voids evolved Al_2O_3 surface under SEM. Fig. 4(a) illustrates the SEM image (top view) of the micro-voids induced sample surface of Fig. 3(c), whereas Fig. 4(b) depicts the SEM image (side view) of the Al_2O_3-coated glass sample. If we closely look at the Al_2O_3 surface of Fig. 4(a), we don't observe any holes on the Al_2O_3 surface. The side view of Fig. 4(b) also confirms the absence of any holes on the top or inside the top layer (Al_2O_3 layer) of the sample. These results indicate that, femtosecond laser can fabricate voids inside the bottom layer without damaging the top layer, which is achieved due to the non-linear properties of the femtosecond laser beam.



(a) (b)

Fig. 4 SEM image of the micro-voids evolved Al_2O_3-coated boro-aluminosilicate glass sample. (a) Top view of Al_2O_3 surface; (b) side view of the Al_2O_3-coated glass sample.

C. Formation Mechanism of the Micro/nano-scale Holes/voids in Transparent Materials

In transparent materials, when we apply single pulse in a spot, holes/voids are evolved in the material depending on the focus point of the laser beam. We can expect filamentation effect of the laser beam in transparent materials. High intensity femtosecond laser beam can create electron plasma filaments in transparent materials that cause the extension of the focal length of the laser beam. Alternate self-focusing and defocusing of electron plasma guides a femtosecond laser beam through a prolonged narrow channel, known as filamentation. The critical power (P_c) of self-focusing is obtained by the following equation [18]:

$$P_c = \frac{3.77\lambda^2}{8\pi n n_2} \tag{1}$$

where λ is the wavelength of the incident laser beam, n is the linear and n_2 is the non-linear terms of refractive index of the medium.

As mentioned before, we utilized Gaussian shape laser beam during our experiments. For a Gaussian laser beam, we can determine the radius of the focal spot size using the following equation [19].

$$w \approx \frac{\lambda f}{\pi w_0} \tag{2}$$

where λ is the wavelength of the incident light, f is the focal length and w_0 is the 1/e Gaussian beam radius. With the decrease in laser energy, the intensity of the Gaussian laser beam decreases in every part of the focal spot. As a result, only the central part of the Gaussian laser beam can cross the threshold energy of a material. As a result, by decreasing the laser energy we can reduce the diameter of the micro/nano-holes or voids. However, when the laser energy of the central part of the Gaussian laser beam is below the threshold energy of a material, we cannot produce holes or voids in that material. Thus, the lowest diameter of the nano-holes/voids depends on the threshold energy of that material and the lowest possible energy of the central part of the laser beam. These parameters are dependent on the material and the optical equipments. In order to fabricate holes/voids in any material, the flatness of the translation stage is very important. Slight variation in the stage flatness may vary the position of the holes/voids in the material.

IV. Conclusions

In summary, we reported the fabrication of periodic micro-nano-metric holes/voids on the surface of and inside Al_2O_3-coated boro-aluminosilicate glass. The lowest diameter of the fabricated nano-holes, printed on Al_2O_3 coating, was measured as ~ 700 nm. The nano-voids of 1.1 µm diameter have been developed inside the glass sample without damaging the top layer of Al_2O_3. Furthermore, we investigated the formation mechanism of the nano-holes and nano-voids in the Al_2O_3-coated glass samples. It is well known that, micro/nano-holes and micro/nano-voids can cause scattering of incident light. Thus, the proposed micro/nano-metric holes/voids' array will be suitable for spreading light, which property can be utilized to fabricate various optical/photonic devices.

Acknowledgment

We are very much thankful to R&D Centre, Samsung Corning Precision Materials, Republic of Korea for their Al_2O_3-coated boro-aluminosilicate glass samples and technical & financial support.

References

[1] S. H. Ko, I. Park, H. Pan, C. P. Grigoropoulos, A. P. Pisano, C. K. Luscombe, and M. J. Fréchet, "Direct nanoimprinting of metal nanoparticles for nanoscale electronics fabrication," *Nano Lett.*, vol. 7, pp. 1869–1877 (2007).

[2] J. Grunes, J. Zhu, E. A. Anderson, and G. A. Somorjai, "Ethylene hydrogenation over platinum nanoparticle array model catalysts fabricated by electron beam lithography: determination of active metal surface area," *J. Phys. Chem. B*, vol. 106, pp. 11463–11468 (2002).

[3] K. Arshak, M. Mihov, A. Arshak, D. McDonagh, and D. Sutton, "Focused ion beam lithography-overview and new approaches," *Proc. International conference on Microelectronics*, 2004, p. 459.

[4] R. C. Jaeger, *Introduction to Microelectronic Fabrication*, 1st ed., USA, Prentice Hall, 2002.

[5] F. Cerrina, *Microlithography, Micromachining and Microfabrication*, 1st ed., USA, SPIE Press, 1997.

[6] D. M. Mattox, *Handbook of Physical Vapor Deposition (PVD) Processing: Film Formation, Adhesion, Surface Preparation and Communication Control*, 1st ed., USA, Noyes Publications, 1998.

[7] A. R. Barron, "Chemical Vapor Deposition," Connexions, Rice University, USA [Online]. Available: http://cnx.org/content/m25495/latest/.

[8] S. Calxto, M. R. Agullar, F. J. S. Marin, and L. C. Escobar, "Rod and spherical silica microlenses fabricated by CO_2 laser melting," *Appl. Opt.*, vol. 44, pp. 4547–4556 (2005).

[9] C. Vass and K. Osvay, and B. Hopp, "Fabrication of 150 nm period grating in fused silica by two-beam interferometric laser induced

backside wet etching method," *Opt. Express*, vol. 14, pp. 8354–8359 (2006).

[10] L. Qi, K. Nishi, and Y. Namba, "Regular subwavelength surface structures induced by femtosecond laser pulses on stainless steel," *Opt. Lett.*, vol. 34, pp. 1846–1848 (2009).

[11] M. S. Ahsan and M. S. Lee, "Femtosecond laser induced nanostructures in soda-lime glass," *J. Laser Micro/Nanoeng.*, vol. 7, pp. 202–207 (2012).

[12] M. S. Ahsan, F. Deqanda, M. S. Lee, H. Sekita, and T. Sumiyoshi, "Formation of superhydrophobic soda-lime glass surface using femtosecond laser pulses," *Appl. Surf. Sci.*, vol. 265, pp. 784–789 (2013).

[13] F. Ahmed, M. S. Ahsan, and M. S. Lee, "Near-field modification of femtosecond laser beam to enhance single-shot pulse filamentation in glass medium," *Appl. Phys. A*, vol. 114, pp. 1161–1165 (2014).

[14] E. N. Glezer, M. Milosavljevic, L. Huang, R. J. Finlay, T. –H. Her, J. P. Callan, and E. Mazur, "Three-dimensional storage inside transparent materials," *Opt. Lett.*, vol. 21, pp. 2023–2025 (1996).

[15] A. P. Joglekar, H. Liu, G. J. Spooner, E. Meyh□fer, G. Mourou, and A. J. Hunt, "A study of the deterministic character of optical damage by femtosecond laser pulses and applications to nanomachining," *Appl. Phys. B*, vol. 77, pp. 25–30 (2003).

[16] M. S. Ahsan, Y. G. Kim, and M. S. Lee, "Formation mechanism of nanostructures in soda-lime glass using femtosecond laser," *J. Non-Cryst. Solids*, vol. 357, pp. 851–857 (2011).

[17] Q. Zhang, H. Lin, B. Jia, L. Xu, and M. Gu, "Nanogratings and nanoholes fabricated by direct femtosecond laser writing in chalcogenide glasses," *Opt. Express*, vol. 18, pp. 6885–6890 (2010).

[18] M. S. Ahsan, Y. Y. Kwon, I. B. Sohn, Y. C. Noh, and M. S. Lee, "Formation of periodic micro/nano-holes array in boro-aluminosilicate glass by single-pulse femtosecond laser machining," *J. Laser Micro/Nanoeng.*, vol. 9, pp. 19–24 (2014).

[19] D. Grobnic, S. J. Mihailov, C. W. Smelser, and H. Ding, "Sapphire fiber bragg grating sensor using femtosecond laser radiation for ultrahigh temperature applications," *IEEE Photon. Technol. Lett.*, vol. 16. pp. 2505–2507 (2004)

A Novel Approach to Design Photonic Crystal Based Optical Power Splitter with a Flow Control Mechanism.

Arif Ahmed[1*], Md. Rokibul Hasan Robin[1], Md. Osman Goni[1]
[1]Dept. of Electronics and Communication Engineering,
Khulna University of Engineering and Technology
Khulna - 9203, Bangladesh.
*kuet0909026@gmail.com

Abstract— **We proposed and demonstrated an efficient method to split electromagnetic wave in two dimensional photonic crystal by using a T-shaped splitter. By measuring normalized electromagnetic field, it is shown that the equal outputs of splitter can be limited and controlled by using a control rod which position can be varied. It is also showed that the output of two ports can be unequal and it can be controlled by varying the position of control rod.**

Index Terms— **photonic crystal, splitter, control rod.**

I. Introduction

"Photonic Crystals" (PC) are dielectric structures with periodic modulation of the refractive index on the scale of the wavelength of the light [1], [2]. Due to this periodic modulation of refractive index, a photonic bandgap (PBG) is generated and propagation of electromagnetic wave is forbidden for wave vectors within this bandgap. Over the past several years, various important scientific and engineering applications of the control of light emission and propagation and the trapping of photons have been realized such as line defect waveguide [3],[4],planar waveguide [5],[6], coupled cavity waveguide [7],[8]. The optical power splitter or divider is an indispensable component in photonic integrated circuits which is used in fiber optic networks. Basic structure of the splitter is divided into two parts, an input and an output port. Ideally, input power is equally divided into the output ports without any significant reflection and radiation loss.

In a PC, a simple Line defect waveguide is created by removing a row of rods along one of its main crystalline directions. Sharp bend of light wave was done in 90º by Attila Mekis, J. C. Chen, I. Kurland, Shanhui Fan, Pierre R. Villeneuve, and J. D. Joannopoulos [9]. We can enhance the model by removing the rods in both directions which can work as a T-shaped power splitter [10].

We have shown the control mechanism of the T-shaped power splitter by placing a control rod in between the paths. We can control the output power at the ports by changing the placement of the control rod.

II. Structure and Design

In experimental environment we used a 2D photonic crystal. The crystal consists of silicone rods having a refractive index of 3.48. We took a square block having the sides of 5.5µm. The center to center distance of the rods is chosen to be 500nm. This is a frequency tunable model. These parameters can be changed to select a different frequency. Our model is tuned to 200 THz for the simulation. To create the line defects silicone rods are removed in a style that it creates a path of 11µm and have the periodicity to the next rod in line of □ = 500nm.

For photonic conduction there are two requirements (1) strong localization of the individual defect modes, (2) weak interaction between these localized modes. Under these conditions the transmission of photons in a crystal can be achieved regardless of changing the direction. We have experimented with a T-shaped beam splitter with different configurations.

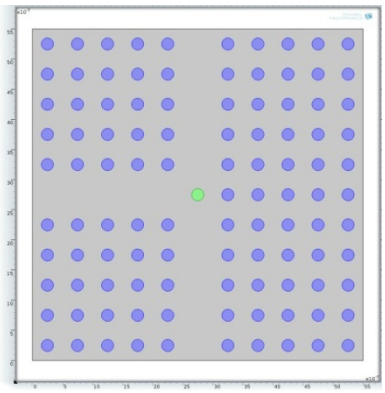


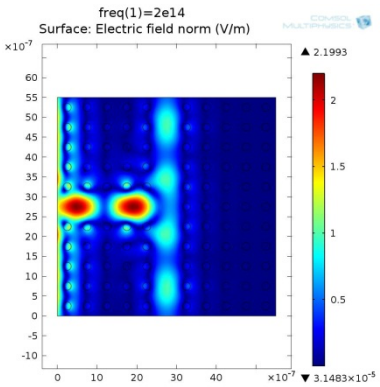
Figure - 1: Silicone rod based T-shaped photonic crystal beam splitter with □=500nm f=200THz. The control rod is marked in green. The output power at each port can be controlled by varying the position of the control rod.

III. Experiments

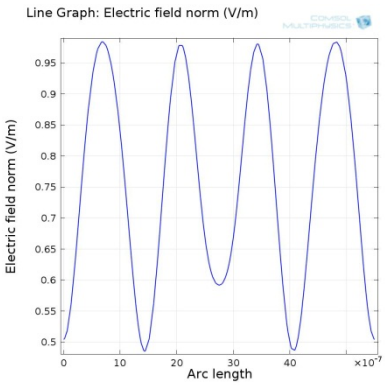
We first experimented with incident electromagnetic wave to the line defect. This experiment shows the result that the wave power is spilt into the top and bottom channel equally. The system is frequency selective. So the comparison of response at different frequency is not a concern in this case. We measured the distribution of field in the channel. This is quite a sinusoidal shape and shows a loss that tends to zero. The total power is conserved at the outputs.

Electric field in the output channels can be controlled with introduction of a control rod at the center. The position of the center rod controls the field (Fig-3). We can either introduce an amount of attenuation to the flow of the wave or we can drive the flow to a specific channel if we want.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

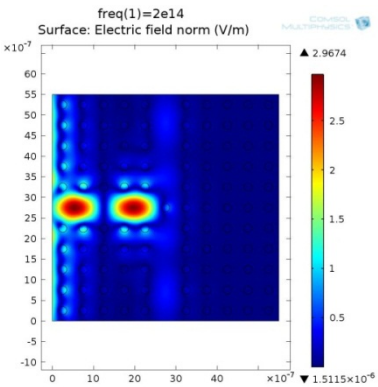


(a)

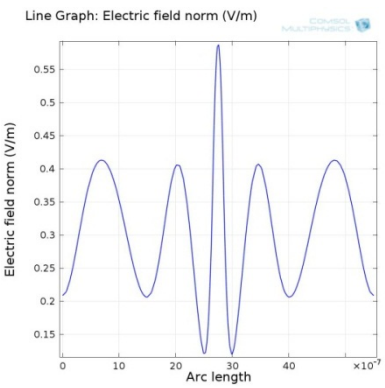


(b)

Fig-2: (a) The electric field flow in the splitter (without the control rod). The red line shows the point we measured the distribution. (b) The plot shows the measurement of electric field on the line. The center indicates the point of split at the crystal center.



(a)



(b)

Fig-3: (a) splitter with a control rod at the center. (b) The field distribution over the channels. The center peak indicates the absorption by the control rod.

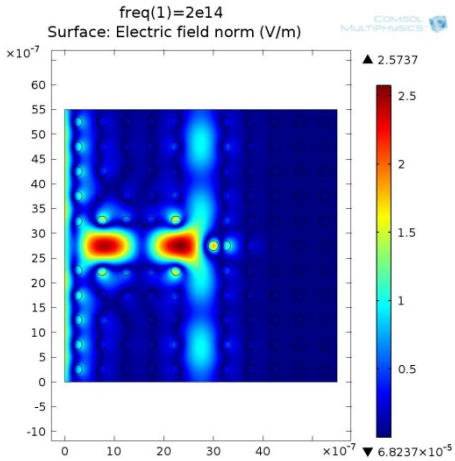


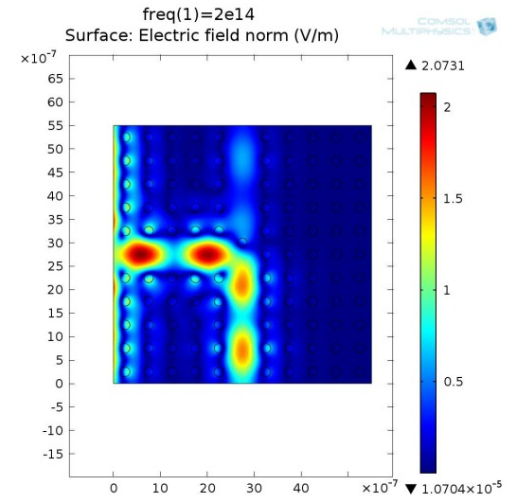
Fig-4: Output power with control rod moved towards positive x direction

Fig-5: control mechanism of the control rod. (a) More power at the bottom channel, (b) More power at the top channel.

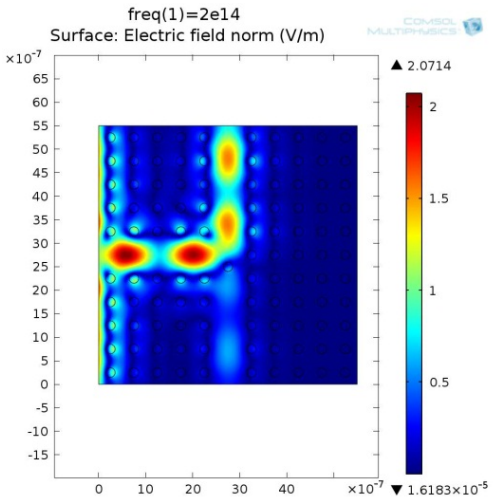
IV. Result analysis & discussion:

In the first experiment, we did not use the control rod. This gives us equal power split into both channels with a very low attenuation. Fig-2 shows the output power at the channels. We used 2v/m and got 0.99V/m at each channels. Then we introduced the control rod between the channels at central position. This gives also gives us an equal distribution of power at both channels but has a large amount of attenuation (Fig-3). Fig-3(a) shows the attenuated flow of the wave in the output channels. Fig-3(b) is a line graph for the field distribution in the output channels. The peak at the centre is for the absorption property of the control rod. We can see a large amount of power is absorbed by the centre rod. As earlier we used 2 V/m (peak) as the input and the graph is showing about 0.45 V/m at the output channels.

If we move the control rod horizontally towards the wall of silicone rods, more waves can pass through the gap between the input channel and the output channels. In fig-4 there is more power flowing in the output channels in comparison with the previous setup in fig-3(a). The opposite can be done by moving the control rod towards the input channels. This will decrease the output power at both channels. In all the cases the field distribution in both the output channels are equal. The change in the position of the control rod in horizontal fashion does provide same attenuation for both channels.



(a)



(b)

Fig-5: control mechanism of the control rod. (a) More power at the bottom channel, (b) More power at the top channel.

The field applied to the input can be driven to a specific channel out of the two by moving the control rod vertically. When the control rod is moved in the vertical direction, it provides more attenuation to one of the paths and thus driving the power flow to the other channel. The distribution of the power in the channels can be controlled by the position by the control rod position. If it is placed in line with the side wall of any one output channel, the other channel will carry most of the power. In fig-5(a), the control rod is placed in a fashion that 70% of power can flow in the bottom channel and the rest 30% in the top channel. As we are using 2V/m as input, we get peak of 1.42v/m in the bottom channel and 0.58v/m at the top channel (fig-5(a)). The same setup is repeated with the control rod placed at the mirror position of the previous setup (fig-5(b)). This gives the exact same result except that the maximum power is flowing in the top channel.

This model is a frequency tuneable system. The pitch and the radius of the rods are tuned to a specific frequency. In our experiments we have chosen 200 THz as the resonant frequency. In other frequencies the model does not provide a satisfactory result as a waveguide or a conducting path for the wave. At 200 THz the model works fine.

V. Conclusion

This paper has discussed about power splitter in optical frequency region based on photonic crystal. A T-shaped power splitter has been proposed to achieve an equal distribution over the channels. It is also scrutinized to establish a control mechanism for the splitter. A novel controlling method is also proposed by using a control rod. A flexible control is achieved with the control rod technique. The technique is studied extensively and investigated with different length and other conditions.

The results presented here indicated that this type of optical power splitter can be used in optical circuits. This can also be used to feed two or more output channels with the same signal.

VI. References

[1] J.D.Joannopoulos, R.D.Meade and J.N.Winn, *Photonic Crystals: Molding the flow of light*, (Princeton University Press, New York, 1995).

[2] C.M.Soukoulis, *Photonic Crystals and light Localization in the 21st century*, (Kluwer Academic Publishers, 2000).

[3] M. M. Sigalas, R. Biswas, K. M. Ho, C. M. Soukoulis, D. Turner, B.Vasiliu, S. C. Kothari, and S. Lin, Microwave Opt. Technol. Lett. 23, 56 (1999).

[4] A. Chutinan and S. Noda, Appl. Phys. Lett. 75, 3739(1999); S. Noda, K. Tomoda, N. Yamamoto, and A. Chutinan, Science 289, 604 (2000)

[5] A. Mekis, J. C. Chen, I. Kurland, Shanhui Fan, Pierre R. Villeneuve, and J. D. Joannopoulos, Phys. Rev. Lett. 77, 3787 (1996).

[6] T. Baba, N. Fukaya, and J. Yonekura, Electron.Lett. 35, 654 (1999).

[7] M. Bayindir, B. Temelkuran, and E. Ozbay, Phys. Rev. Lett. 84, 2140 (2000).

[8] M. Bayindir and E. Ozbay, Phys. Rev. B 62, R2247 (2000)

[9] Attila Mekis, J. C. Chen, I. Kurland, Shanhui Fan, Pierre R. Villeneuve, and J. D. Joannopoulos, *"High Transmission through Sharp Bends in Photonic Crystal Waveguides"*, Phys. Rev. Lett. 77, 18(1996).

[10] Mehmet Bayindir, B. Temelkuran, and E. Ozbay, *"Photonic-crystal-based beam splitters"*, Appl. Phys. Lett. 77, 24(2000)

Characteristics of Stress-Birefringence in Polymer Optical Strip Waveguides with Air Cladding

Md. Faruque Hossain,[1,*] Md. Mamunur Rahman,[1] and Hau Ping Chan[2]

[1]Department of Electronics and Communication Engineering, Khulna University of Engineering and Technology
Khulna-9203, Bangladesh

[2]Department of Electronic Engineering, City University of Hong Kong
83 Tat Chee Avenue, Kowloon Tong, Hong Kong

[*]fhossain97@yahoo.com

Abstract— **Thermal stress which is unavoidably generated in the fabrication process alters the optical property of the waveguide materials and thus hinders the realization of polarization insensitive devices. This work reports the detailed numerical investigation of stress-induced birefringence in polymer strip waveguide. A more realistic model is used in the finite element analysis to estimate the stresses induced over the entire sequential fabrication process. It is observed that the birefringence in the strip channel is non-uniform which is maximum near the lower cladding and decreases to zero towards the top of the channel. The variation in birefringence is also influenced by the waveguide geometries. Such birefringence characteristics are obtained and discussed in a generalized form, allowing for their applicability with various polymer materials. Thus, the generalized characteristics can be used to determine the stress-birefringence of a polymer strip waveguide with air cladding only by knowing the stress-birefringence in its planar film. This will eliminate the necessity of extensive numerical analysis for the prediction of thermal stress in polymer waveguides and will be more accurate than analytical solutions.**

Index Terms—**Optical waveguide, Thermal stress, Polymer waveguides, Birefringence.**

I. Introduction

Stress, which is unavoidably generated in the fabrication process, has long been a key issue for the yield improvement of many optical waveguide devices [1]–[4]. The stress is generated in the waveguide materials mainly due to the thermal expansion mismatch among the constituent materials. In polymer based optical waveguides, the stress can also be generated due to the polymerization shrinkage even at room temperature when cured under UV light [5]. Due to the photo-elastic effect, this stress may cause anisotropic changes in the material refractive index (i.e., material birefringence), which then makes the device polarization dependent [4]. The problem is more severe in case of polymer based optical waveguides, since the polymer film has large thermal expansion coefficient and hence experience large thermal mismatch with inorganic substrate. To reduce such stress-induced effects, therefore, an accurate estimation of stresses and their influence on the optical properties of the waveguide materials are important. Polymer materials are very attractive in the fabrication of optical waveguide devices due to their low cost, ease of processing, simplicity of integration onto desired surfaces, and energy-saving property (e.g., large thermo-optic coefficient) [3], [6].

A large number of works have been devoted to study the stress-induced material birefringence in optical waveguides [3], [4], [7]. Most of these works used either analytical [3] or numerical methods [4] for the stress analysis, since the measurement of stress is difficult in tiny waveguides or inside the components. Available analytical methods are simple but insufficient to capture the complex behavior of polymer materials and thus to estimate the stresses accurately in polymer waveguides [7]. Numerical analysis based on finite element (FE) method is, therefore, often employed for the analysis of stress-birefringence in optical waveguides. Recently, we have proposed a more realistic FE model to accurately estimate the stresses in polymer waveguides [7]. A successful application of this model has also been demonstrated by the efficient design of polarization insensitive devices considering the stress-induced effects [8]. However, such accurate prediction of stresses requires extensive material properties (e.g., elasto-plastic and viscoelastic behaviour, chemical kinetics for cure reaction, etc) which are usually unavailable in literature for many optical polymers and also difficult to characterize in thin films.

This work presents a generalized characteristic of stress-birefringence in polymer strip waveguide based on the detail numerical investigation of thermal stress in various waveguide structures using the realistic FE model. The obtained characteristic is expressed by an empirical formula which then can be used to determine the stress-birefringence in an air-clad polymer strip waveguide avoiding the requirements of extensive numerical analysis. Since the strip waveguide has been widely used in fabricating various optical devices [2], the presented results will be useful for the treatment of stress-induced problems in many of those optical devices.

II. FE Modeling and Photo-elastic Theory

A. Finite Element Modeling

Fig. 1 shows the cross section of a strip waveguide (width, w; and height, h). As the material chosen, the waveguide is comprised of benzocyclobutene (BCB) for the core layer and epoxy (OPTOCAST 3507) for the lower cladding layer. In the fabrication process [2], epoxy was spin-coated on silicon (Si) wafer first, cured under UV light at room temperature (24 °C) and post-cured at 120 °C. Then the benzocyclobutene (BCB) was spin coated on the epoxy and cured at 250 °C. Finally, the strip was patterned by conventional photolithography and dry etching process. Although the quantitative values are derived for the BCB waveguide, the stress-birefringence characteristics would be applicable for other polymers as well.

To accurately predict the stresses, we employed our developed FE model [7] in MSC Marc software. Comparing to the conventional models, this model is more realistic for the

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

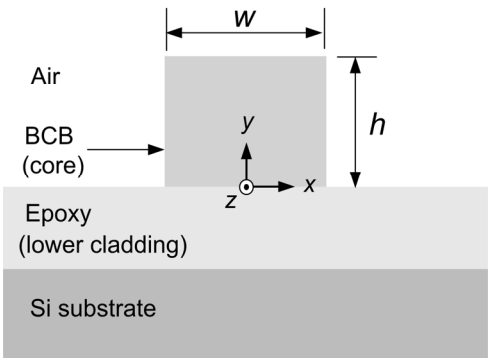


Fig. 1 Cross-section of a strip channel waveguide. The strip width and height are denoted by w and h respectively. The aspect ratio R is defined as w/h.

thermo-mechanical analysis of polymer optical waveguides over the entire sequential fabrication process. In the simulation, the geometry of the final structure was modeled at the very beginning and only the substrate was considered stress free at room temperature. The BCB layer was deactivated (i.e. exhibits no stiffness) till the end of the curing of epoxy. Thereafter, it was activated in a "strain-free" manner. Stress induced by the curing shrinkage was considered in our model by using coupled cure-thermal-mechanical approach. Creep model was used to simulate the stress relaxation process. The etching process was simulated by removal of the selected material elements which demonstrate zero stresses and strains on the post elements.

The strip waveguide analyzed in this work is of various heights around 3.0 µm (2.25 µm – 4.5 µm) and has a lower cladding thickness of 10 µm on 500 µm thick Si substrate. For the stress calculation, the waveguide geometry having dimension of 5 mm × 513 µm × 5 mm is used in the FE model which consists of over 8000 hex elements and 10000 nodes with required thermal and mechanical boundary conditions.

B. Photo-elastic Theory

In the elastic range of material, the stress dependence of the material refractive index is expressed by the Neumann-Maxwell equation [9]:

$$n_x = n_0 - C_1\sigma_x - C_2(\sigma_y + \sigma_z),$$
$$n_y = n_0 - C_1\sigma_y - C_2(\sigma_x + \sigma_z), \quad (1)$$

where n_x and n_y are the principal refractive indices in the in-plane (parallel to x-axis in Fig. 1) and out-of-plane (parallel to y-axis) direction, respectively, n_0 is the refractive index without stress, σ_x, σ_y and σ_z are components of principal stresses in the x, y and z direction, and C_1 and C_2 are the stress-optic coefficients. If the waveguide is invariant along the direction of light propagation (z-direction), the material birefringence can be calculated as $n_x - n_y = (C_2 - C_1)(\sigma_x - \sigma_y)$.

III. Results and Discussions

A. Stress Distributions

From the simulation model, the principal stresses (σ_x, σ_y, and σ_z) are noted at room temperature for two process steps: before etching and after etching. Before etching, the stresses in the planar BCB film (core layer) are obtained as $\sigma_{x,film} = \sigma_{z,film} = 44.1$ MPa and $\sigma_{y,film} = 0.657$ MPa for the in-plane and out-of-plane directions, respectively. Being a planar structure before etching, these stresses are uniform all over the BCB layer. A summary of the properties of fully cured materials used in the FE simulation are shown in Table I. Although not explicitly given in the table, a cure and temperature dependence for the CTE and elastic modulus is considered. The film stress can be verified easily in this stage by measuring the warpage of the structure [8]. In this work, the warpage and thus the stress was measured using Ambios XP-2 stylus profiler. The corresponding measured in-plane stress (45 ± 3 MPa) is in good agreement with the simulated result (44.1 MPa).

TABLE I
Material properties used in the FE analysis

Materials	Temperature (°C)	Elastic Modulus (GPa)	CTE (ppm/°C)	Poisson's ratio
Silicon	-	130	2.6	0.27
Epoxy T_g = 145 °C	Below T_g	2.5	65	0.34
	Above T_g	0.25	200	
BCB T_g = 350 °C	Below T_g	2.9	56.7	0.34

When the slab structure was etched to pattern strip waveguide, the stresses were relaxed and became non-uniform. However, it is very difficult, sometimes impractical, to measure that stress in the tiny channel waveguides. Fig. 2(a) and (b) shows the simulated in-plane (σ_x) and out-of-plane (σ_y) stress distributions in the strip channel (after etching) of width w = 3.0 µm and height h = 2.0 µm. The stress values are normalized by the in-plane stress retained in the film ($\sigma_{x,film} = \sigma_{z,film} = \sigma_{film}$) before etching. Unlike the planar (slab) waveguide, the stress distributions in the channel waveguide are highly non-uniform.

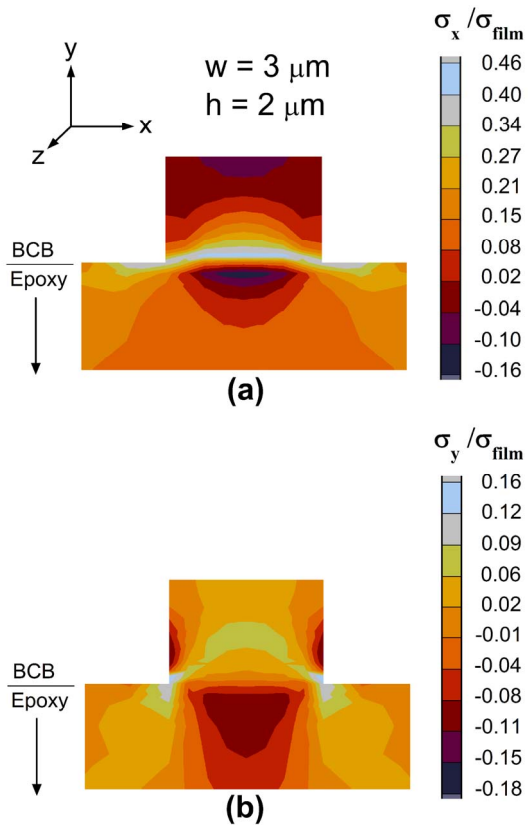


Fig. 2 a) In-plane, σ_x and b) out-of-plane, σ_y stress distributions in the strip waveguide consisting width, w = 3.0 µm and height, h = 3.0 µm. The stresses are normalized by the corresponding film (i.e., before etching) stress.

Fig. 2(a) reveals that the in-plane stresses (σ_x) in the strip channel are greatly relieved. The remaining maximum in-plane stress is about 50% of the film stress, but only in an extremely narrow area adjacent to the lower cladding or substrate. This is due to the absence of constraining force in the free standing part of strip structure compare to the bottom [10]. As shown in Fig. 2(b), a significant amount of out-of-plane stress (σ_y) also exist in the channel which is now comparable to the in-plane stress. Herein, positive stress is defined as tensile, whereas the negative stress is compressive.

To examine the influence of strip geometries on the evolution of stresses, numerical analysis is performed for a broad range of strip dimensions (2.25–4.5 μm in width and 2.0–4.0 μm in height). A more detail of such geometric influence will be discussed in the next section through its influence on the material birefringence. As mentioned earlier, the prediction of the stress-birefringence characteristics is of our primary interest.

B. Stress-Birefringence Characteristics

Once the stress distributions are obtained, combining them with photo-elastic coefficients using (1), the stress-induced birefringence (Δn) are determined for the waveguide material as $\Delta n = n_x - n_y = (C_1 - C_2)(\sigma_x - \sigma_y)$. For a BCB thin film, C_1 and C_2 are 99×10^{-12} Pa^{-1} and 31×10^{-12} Pa^{-1}, respectively at an optical wavelength of 1550 nm, whereas these are zero for epoxy [11]. Before etching, the birefringence calculated for the BCB film is 0.003, which is uniform all over the film layer. In this case, it was a slab waveguide and the material index of BCB can be measured easily by prism coupler. The measured birefringence was 0.0034 for BCB at 1550 nm wavelength [2], which is comparable to the calculated result of 0.003. Due to the zero values of photo-elastic coefficients, no birefringence is calculated here for the epoxy layer.

The birefringence retained in the BCB waveguide after etching is shown in Fig. 3 as a function of the position (y) along the height for various waveguide geometries. Similar to the waveguide stress, birefringence is also normalized by the value that retained in the planar (before etching) film. Such normalization will help us understand the general feature of the birefringence characteristics in channel waveguides. Fig. 3 shows that the birefringence is higher near the lower cladding or substrate, with its maximum value of about 50% of the birefringence remained in the film before etching. The birefringence decreases exponentially to the zero value as it approaches the top of the channel. As explained earlier, the material birefringence in the planar waveguide is induced mainly by the in-plane stress. In the etching process, however, the in-plane stress relaxes remarkably, while more of the out-of-plane stress is induced in the strip region. This is the reason why the material birefringence decreases sharply as it approaches the top of the channel.

Fig. 3 reveals that the significant birefringence remains only up to the height of about 2 μm regardless of the dimensions of the strip. For a given width, the birefringence retaining in the strip waveguide is constant, irrespective of the channel height. This is appeared in Fig. 3(a) with the essentially identical lines for w = 3 μm, and h = 2, 3 and 4 μm. However, Fig. 3(b) shows that the birefringence at a particular height slightly increases with the increase in strip width. The birefringence does not vary significantly along the width of the strip channel.

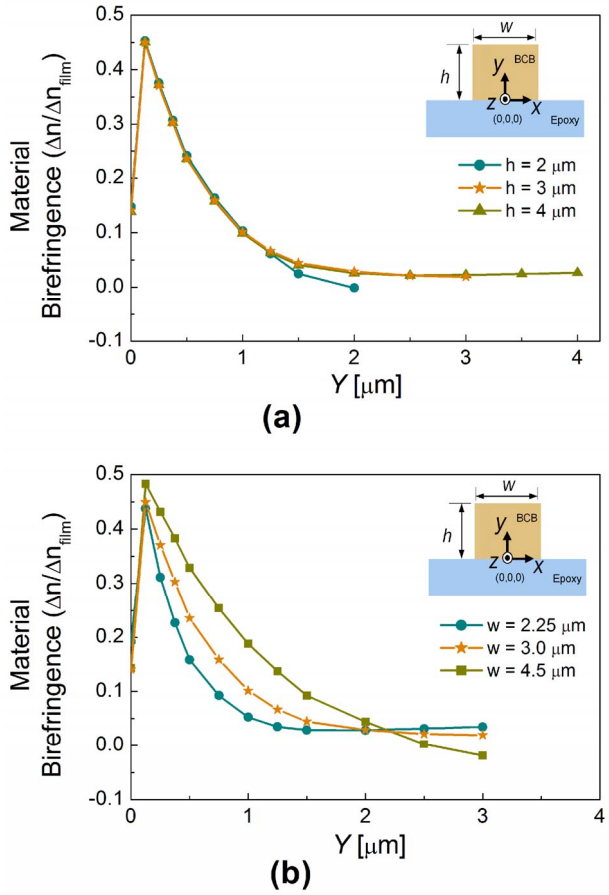


Fig. 3 Calculated normal birefringence ($\Delta n / \Delta n_{film}$), normalized by the corresponding value before etching, along the height of a strip channel (a) for different height (h = 2– 4μm), width w = 3.0 μm and (b) for different width (w = 2.25– 4.5 μm), height h = 3 μm.

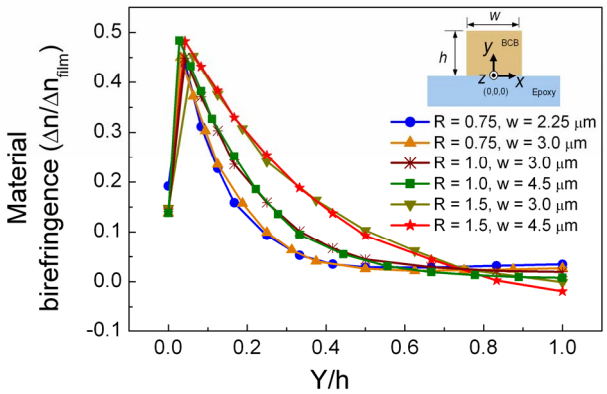


Fig. 4 Generalized characteristics of the stress-birefringence in strip channel waveguide. The birefringence is normalized by the value of that remains in the film before etching and the position along height is normalized by the maximum height of the corresponding strip channel.

Based on the aforementioned observations, a generalized stress-birefringence characteristic can be obtained by plotting the normal birefringence against the position (y/h) normalized by the corresponding maximum height. Fig. 4 shows the re-illustrated curves of Fig. 3, in which the x-axis is replaced by the normal position (normalized by the maximum height of the corresponding waveguide) instead of the actual position along the height. As can be observed, the stress-induced birefringence remaining in the strip waveguide is now simply related to the birefringence in planar film (Δn_{film}) and the aspect ratio ($R = w/h$) of the strip channel.

Using the curve fitting technique, the obtained generalized characteristics can be expressed by the following empirical relation:

$$\Delta n / \Delta n_{film} = 0.56\exp(-k_1 y_n) + k_2 (y_n / h)^2 \qquad (2)$$

where y_n is the normalized position (y/h), and k_1 and k_2 are the fitting constants. The values of these constants depend on the aspect ratio R of the waveguide and can be obtained as:

$$k_1 = 5.55 + 0.45(1-R), \quad k_2 = 0.16 + 0.7(1-R). \qquad (3)$$

Thus, by knowing the stress-induced birefringence before etching (Δn_{film}), which can be easily measured, the birefringence remaining in a polymer strip waveguide can be determined using (2) and (3).

Fig. 5 shows the estimated material birefringence in strip channel using the FE method and the developed empirical relation (2). The formulated birefringence for two different waveguides having same aspect ratio (R = 1.5) agree very well with those obtained using the FE method.

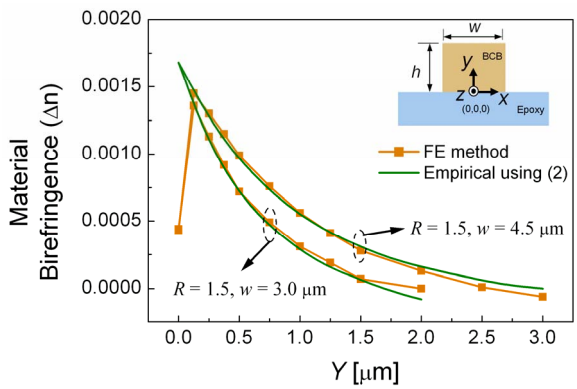


Fig. 5 Material birefringence (Δn) along the height of a strip channel obtained through the FE method (line with markers) and the developed empirical relation (2) (smoothed line). The birefringence is derived for the BCB film with birefringence (Δn_{film}) of 0.003.

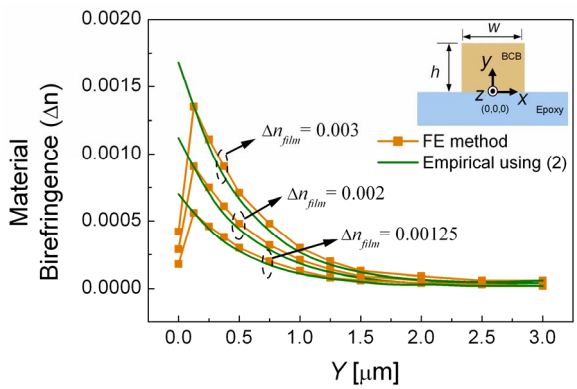


Fig. 6 Comparison between the material birefringence (Δn) estimated using the FE method (line with markers) and the empirical relation (smoothed line) in a strip channel for different film birefringence (Δn_{film}) of 0.003, 0.002 and 0.00125. The birefringence is derived for the strip waveguide with aspect ratio of unity (R = 1).

For different polymers, the value of Δn_{film} can be different due to different levels of induced stresses or different values of stress-optic coefficients, or both. Fig. 6 illustrates the birefringence (Δn) in a strip channel for different film birefringence of (Δn_{film}) of 0.003, 0.002 and 0.00125. A good agreement is found between the birefringence estimated using the FE method and the empirical relation given in (2) and (3). This indicates that the normalized channel birefringence for a given aspect ratio will get the same characteristics irrespective of the different levels of film birefringence (Δn_{film}). This is because the stress relaxes in the etching process mostly due to the elastic properties of the material. On the other hand, the birefringence varies linearly with the stress-optic coefficient. Thus, the obtained stress-birefringence characteristics for the strip channel will remain valid for the typical dimensions of optical waveguide comprising any other isotropic polymers. Such estimation of non-uniform birefringence will also be more accurate than the analytical methods which result only a uniform value [8].

IV. Conclusions

This work presents a detailed numerical investigation of stress-induced material birefringence in an optical strip waveguide composed of polymer materials. The stresses in the waveguide material are determined first through numerical solutions of the thermo-mechanical problem of different waveguide structures. Then, by combining them with the stress-optic coefficients, the material birefringence is obtained. It is found to be non-uniform in the strip channel. In the core layer, the maximum birefringence is observed near the lower cladding, which decreases to zero towards the top of the channel. This birefringence slightly increases with the increase in strip width. Such stress-induced characteristics are presented and discussed in a generalized empirical form, allowing for their applicability with other polymer materials. The obtained stress-induced characteristics can now be included in the optical design of waveguide devices in order to achieve more accurate design results.

References

[1] K. Wörhoff, B. J. Offrein, P. V. Lambeck, G. L. Bona, and A. Driessen, "Birefringence compensation applying double-core waveguiding structures," *IEEE Photon. Tech. Lett.*, vol. 11, no. 2, pp. 206-208, Feb. 1999.

[2] S. Y. Cheng, K. S. Chiang, and H. P. Chan, "Polarization-insensitive polymer waveguide Bragg gratings," *Micro. Opt. Technol. Lett.*, vol. 48, no. 2, pp. 334-338, Feb. 2006.

[3] M. Huang, "Thermal Stresses in optical waveguides," *Optics Lett.*, vol. 28, no. 23, pp. 2327-2329, Dec. 2003.

[4] W. N. Ye, D.-X. Xu, S. Janz, P.Cheben, M.-J. Picard, B. Lamontagne, and N. G. Tarr, "Birefringence control using stress engineering in silicon-on-insulator (SOI) waveguides," *J. Lightw. Technol.*, vol. 23, no. 3, pp. 1308-1318, Mar. 2005.

[5] Y. Lin, W. Liu, and F. G. Shi, "Adhesive joint design for minimizing fiber alignment shift during UV curing," *IEEE Trans. Advanced Packaging*, vol. 29, no. 3, pp. 520-524, Aug. 2006.

[6] H. Ma, A. K.-Y. Jen, and L. R. Dalton, "Polymer-based optical waveguides: materials, processing and devices," *Adv. Mater.*, vol. 14, no. 19, pp. 1339-1365, Oct. 2002.

[7] M. F. Hossain, H. P. Chan, M. A. Uddin, and R. K. Y. Li, "Stress-induced birefringence characteristics of polymer optical rib waveguides," *J. Lightw. Technol.*, vol. 27, no. 21, pp. 4678-4685, Nov. 2009.

[8] M. F. Hossain, H. P. Chan, and A. Z. Kouzani, "Efficient design of polarization insensitive polymer optical waveguide devices considering stress-induced effects," *Opt. Express,* vol. 22, no. 8, pp. 9334-9343, Apr. 2014.

[9] K. Okamoto, *Fundamentals of Optical Waveguides.* San Diego, CA: Academic, 2000, pp. 265.

[10] Y.-L. Shen, S. Suresh and I. A. Blech, "Stresses, curvatures, and shape changes arising from patterned lines on silicon wafers," *J. Appl. Phys.*, vol. 80, no. 3, pp. 1388-1398, Aug. 1996.

[11] M. F. Hossain, H. P. Chan and M. A. Uddin, "Simultaneous measurement of thermo-optic and stress-optic coefficients of polymer thin films using prism coupler technique," *Appl. Opt.*, vol. 49, no. 3, pp. 403-408, Jan. 2010.

Dispersion and Nonlinear Characteristics of a Photonic Crystal Fiber (PCF) with Defected Core and Various Doping Concentration

Md. Sajjad Hossain[1,*], Kishan Neupane[1], Md. Shihab Bin Hafiz[1], and Satya Prasad Majumder[1]

[1]Department of Electrical and Electronic Engineering,
Bangladesh University of Engineering and Technology, Dhaka-1205, Bangladesh
[*]sajjadhossain034@gmail.com

Abstract— **A novel Nearly Zero Dispersion Fiber (NZDF) at operating wavelength is designed having GeO_2 doped SiO_2 square defected core and investigated with Finite Element Method (FEM) using COMSOL. The proposed PCF shows promising dispersion characteristics (0.0001787ps/km-nm) with very low effective area (3.03μm^2) and high non-linear parameter (67.23 W^{-1} km^{-1}) at 1.55μm, making it suitable candidate for chromatic dispersion controller and nonlinear optical application.**

Index Terms—**FEM, NZDF, DCF, DFF**

I. Introduction

Photonic crystal fibers[1], consisting of a central defected region surrounded by multiple air holes running parallel to the fiber length, have been one of the most interesting developments in fiber optics. To achieve nearly-zero ultra-flattened chromatic dispersion in PCFs, several fascinating designs have been proposed so far. Among of them, the conventional PCF designs consist of uniform optimized air-holes [2]-[4]. Photonic crystal fibers guide light by arranging it within a periodic array of microscopic air holes that run along the entire fiber length [5]. In this case, light is guided by the total internal reflection, and the air holes are used to reduce the index of the cladding region. Different arrangement of air holes gives different dispersion characteristics. The task of controlling the chromatic dispersion is a very important problem in dispersion controllers [6], nonlinear systems [7] or designing practical optical communication systems [8]. A nonlinear PCF with several kinds of air-hole diameters [9] and a PCF with two–defected air-hole rings [10] have also been presented so far. Inserting a defected air hole [11] in the core shows ultra-flattened dispersion with low confinement loss. Zero dispersion is also seen at several irregular structures [12]-[13].

In this paper, a PCF containing GeO_2 doped defected SiO_2 square in the core with four rings of hexagonal air hole rings is proposed. Nearly zero dispersion is achieved at operating wavelength 1.55μm. A high nonlinear gain with low effective area has also been obtained at this wavelength. A full-vectorial finite-element method [14] with anisotropic perfectly matched layer has been used to compute the effective mode index. The RF physics feature of COMSOL MULTIPHYSICS 4.3b is used for simulation purpose.

II. Fiber Structure

The cross section of the proposed PCF structure is shown in Fig 1(a). It is composed of circular air holes in the cladding arranged in a triangular array with lattice constant Λ and diameter d. The central core region is perturbed by a GeO_2 doped SiO_2 square with side a. The host material is regular silica. The total number of air-hole rings was chosen to be four in order to simplify as much as possible the structural composition of the PCF. The insertion of the extra GeO_2 doped SiO_2 square in the core region is the novel concept of the present design and according to our extensive literature research is a novel idea hasn't been reported yet. The existence of the central defected air-hole has the function to control the waveguide dispersion properties of the fiber as will be demonstrated later on. The NZDF can be fabricated by the sol-gel method [15] which gives well-arranged air holes.

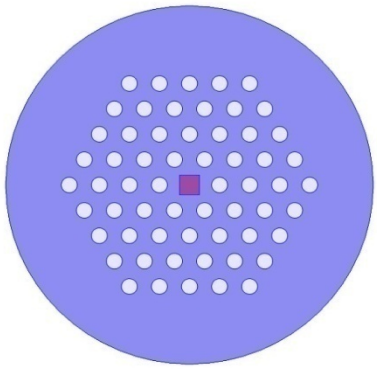


Fig. 1(a) Schematic cross section of proposed PCF structure.

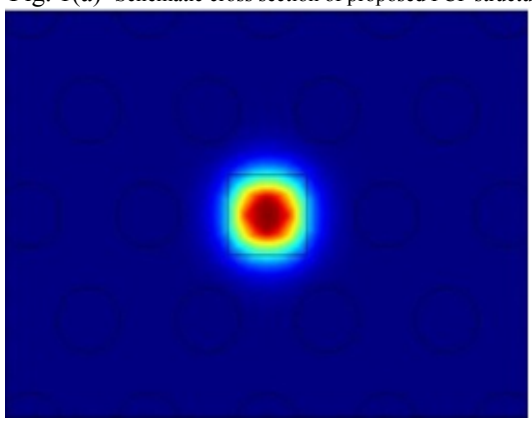


Fig. 1(b) Electric field distribution of proposed PCF structure at wavelength 1.55μm.(The red portion indicates the highest electric field)

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

III. Theoretical Background

We carry out simulation of a GeO_2 doped SiO_2 square defected core PCF structure using COMSOL based on the following theoretical background. Results obtained are exported to MATLAB for further computation. We calculate the total dispersion by the exact definition of the total chromatic dispersion as [11]

$$D = -\frac{\lambda}{c}\frac{\partial^2 n_{eff}}{\partial \lambda^2} \tag{1}$$

where n_{eff} is the effective index of the fundamental mode. Refractive index of SiO_2 and GeO_2 can be found from Sellmeier's equation [16]

$$n^2 = 1 + \sum_{i=1}^{3} \frac{A_i \lambda^2}{\lambda^2 - l_i} \tag{2}$$

where A_i and l_i are the oscillator strength and oscillator wavelength respectively. The refractive index of GeO_2 doped SiO_2 can be computed from the following equation [16]

$$n^2 = 1 + \sum_{i=1}^{3} \frac{[SA_i + X(\mathrm{GA}_i - SA_i)]\lambda^2}{\lambda^2 - [Sl_i + \mathrm{X}(\mathrm{Gl}_i - Sl_i)]^2} \tag{3}$$

where SA, Sl, GA, Gl are the Sellmeier coefficients for the SiO_2 and GeO_2 glasses and X is the doping concentration of GeO_2 and the material dispersion based on Sellmeier's equation has been taken into account explicitly in the effective index of the PCF. We also calculate Effective Area by [17]

$$A_{eff} = \frac{\left(\iint_s |E_t|^2 \, dxdy\right)^2}{\iint_s |E_t|^4 \, dxdy} \tag{4}$$

Then using effective area we calculate the Nonlinearity parameter by [18]

$$\gamma = \frac{2\pi n_2}{\lambda A_{eff}} \tag{5}$$

where nonlinear refractive index, n_2 for GeO_2 doped SiO_2 can be calculated as [18]

$$n_2 \cong 2.507 + .505\Delta \tag{6}$$

IV. Results and Discussions

The wavelength dependence of Silica glass, GeO_2 and GeO_2 doped SiO_2 based on sellmeier's equation is shown in Fig 2. In Fig. 3 we calculate the chromatic dispersion for different doping concentration of GeO_2 in central core. We can clearly see that the continuous increment of the doping concentration of GeO_2 in the defected core downshift the chromatic dispersion to negative values. In Fig 4-6 we show the variation of the design parameters Λ, *d*, *a* to the total dispersion curve of the PCF to investigate the parameter dependence of the dispersion and nonlinear properties of proposed PCF keeping the doping concentration of GeO_2 doped SiO_2 defected core constant at 53.23%.

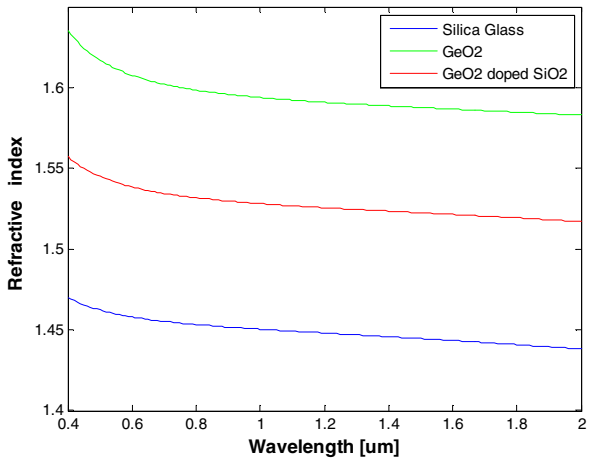


Fig. 2 Plots of refractive index as function of wavelength for Silica Glass, GeO_2 and GeO_2 doped SiO_2 square defected core with GeO_2 doping conc. X=53.23%.

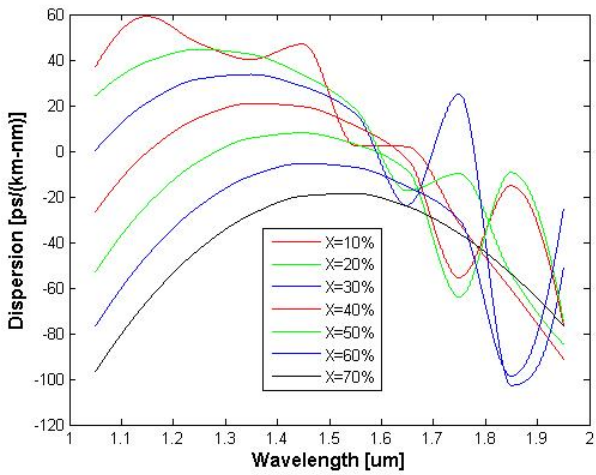


Fig. 3 Plots of dispersion coefficient D versus wavelength for a GeO_2 doped SiO_2 square defected core PCF for different doping concentration of GeO_2, X.

Specifically Fig. 4 shows the variation of the lattice constant Λ to the total dispersion obtained using Eq. (1), for Λ=1.50μm (blue line), Λ=1.55μm (red line), and Λ=1.60μm (green line). Figure 5 shows the variation of the air hole diameter d in the cladding of the PCF, for d/Λ=0.52 (blue line), d/Λ=0.53 (red line), d/Λ =0.54 (green line), while Fig. 6 shows the variation to variation of the defected air-hole diameter in the core of the PCF, for a/Λ=1.312 (blue line), a/Λ=1.325 (red line), and a/Λ=1.338 (green line). According to these numerical results, there exist an optimized set of design parameters that is Λ, d, a and X that lead to ultra-flattened total chromatic dispersion. These optimized values have been obtained through our numerical simulations to be Λ=1.55μm, d/Λ=0.53, a/Λ=1.325 and X=53.23%. Thus the optimized chromatic dispersion responses of the PCF under consideration correspond to the red lines in Fig 4-6. To quantify the performance of our proposed PCF we have achieved two zero dispersion wavelength in O-band and C-band.

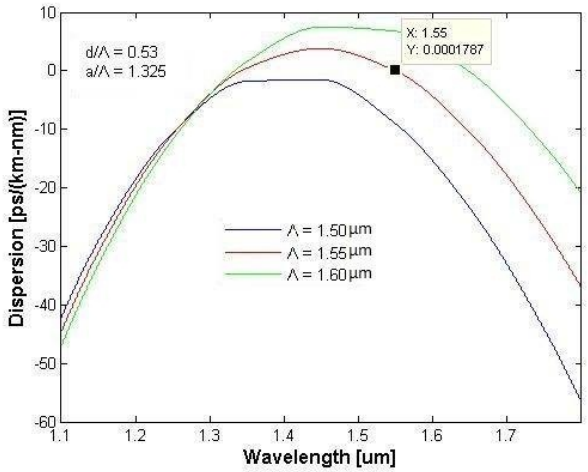


Fig.4 Variation of dispersion versus wavelength with lattice constant Λ as varying parameter for GeO_2 doped SiO_2 square defected core PCF with d/Λ =.53 and a/Λ =1.325 where Λ=1.50µm (blue line), Λ=1.55µm (red line), Λ=1.60µm (green line) with GeO_2 doping conc. X=53.23%.

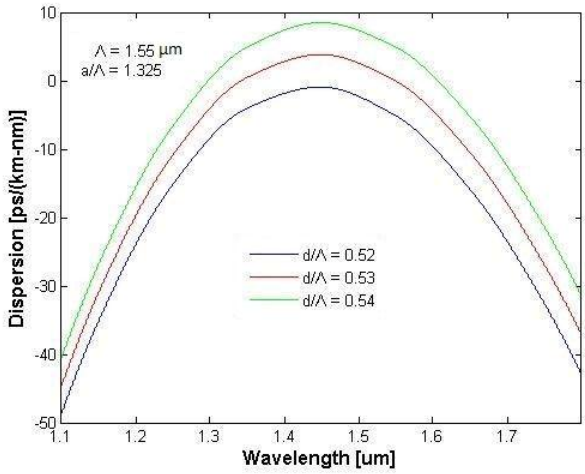


Fig. 5 Variation of dispersion versus wavelength with air hole diameter d as varying parameter for GeO_2 doped SiO_2 square defected core PCF with Λ =1.55 µm and a/Λ =1.325 where d/Λ=0.52 (blue line), d/Λ= 0.53(red line), d/Λ=0.54 (green line) with GeO_2 doping conc. X=53.23%.

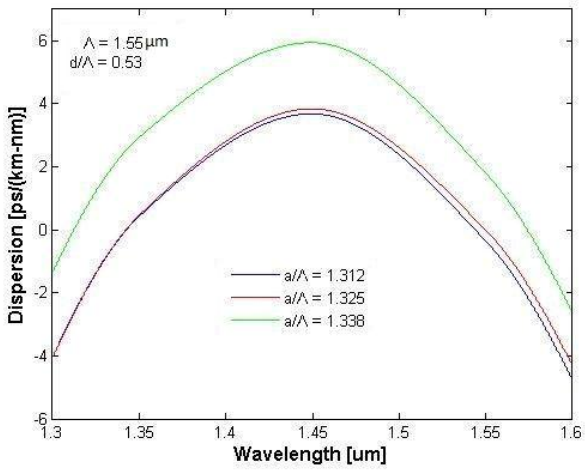


Fig. 6 Variation of dispersion versus wavelength with defected square side length 'a' as varying parameter for GeO_2 doped SiO_2 square defected core PCF with Λ =1.55 µm and d/Λ =1.53 where a/Λ=1.312 (blue line), a/Λ=1.325 (red line), a/Λ=1.338 (green line) with GeO_2 doping conc. X=53.23%.

These results Fig 3-6 show that the newly proposed structure is indeed sensitive to the design parameter X, Λ, d and a. The advantage of our structure is the fact that is much simpler from the design point of view, in comparison to previous designs because it contains less number of design parameters. We have shown already that the proposed PCF structure can exhibit ultra-flattened dispersion characteristics. In practical applications however only the near-zero dispersion characteristics may not be enough for justifying the usefulness of the fiber. Our proposed PCF structure has small effective mode area and high nonlinear gain for nonlinear applications. Fig 7 shows the effective area as a function of wavelength.

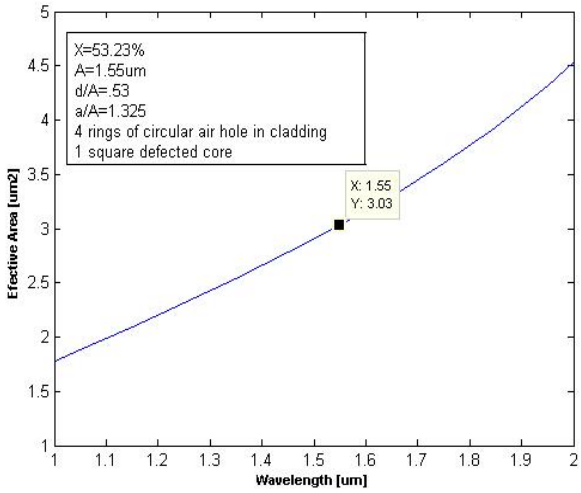


Fig. 7 Variation of Effective area in different wavelength difference for the proposed NZDF for optimized design parameters Λ=1.55 µm, d/Λ=0.53, and a/Λ=1.325.

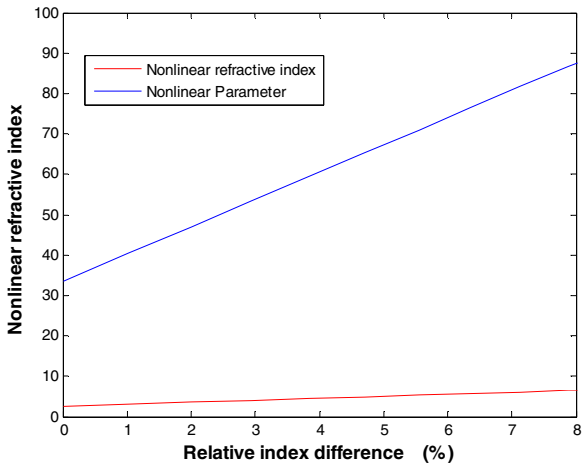


Fig. 8 Plots of Nonlinear refractive index and Nonlinear parameter as a function of relative index difference for the proposed NZDF for optimized design parameters Λ=1.55 µm, d/Λ=0.53, and a/Λ=1.325 d/Λ=0.53(red line), d/Λ=0.54 (green line) with GeO_2 doping conc. X=53.23%.

The nonlinear refractive index, n_2 was plotted using the Eq. 6 as a function of relative index difference as in Fig 8. The nonlinear gain increases while n_2 increases.As the effective area increase, so the nonlinear gain decreases as in Fig 9. At operating wavelength we get small effective area with high nonlinear gain which can be used for Super Continuum Generation, Four Wave Mixing and other nonlinear applications.

In comparison with the ultra-flattened structure, we clearly see in Fig 10 that at operating wavelength the chromatic dispersion is 0.304ps/(km-nm) [11] and proposed structure

gives 0.0001787ps/ (km-nm). The proposed structure also has low effective area than the ultra-flattened structure.

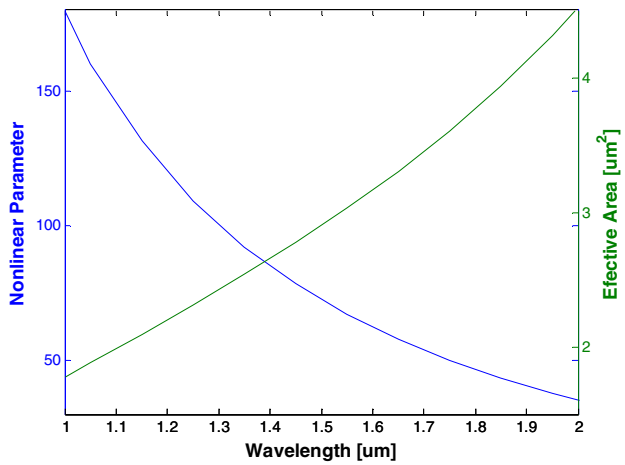


Fig. 9 Variation of effective area and nonlinear parameter as a function of wavelength for the proposed NZDF for optimized design parameters Λ=1.55 µm, d/Λ=0.53, and a/Λ=1.325 d/Λ=0.53(red line), d/Λ=0.54 (green line) with GeO_2 doping conc. X=53.23%.

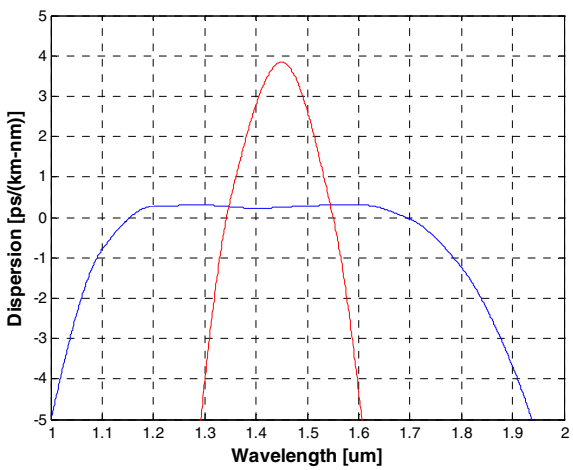


Fig. 10 Comparison of dispersion characteristics between proposed NZDF structure(red line) and ultra-flattened structure (blue line)[11].

V. Conclusions

Dispersion and nonlinear characteristics of GeO_2 doped SiO_2 square defected core PCF is modelled. It is found that at GeO_2 doping concentration 53.23% in defected core with lattice constant Λ=1.55µm, d/Λ=0.53, and a/Λ=1.325,the NZDF shows dispersion coefficient of 0.0001787ps/ (km-nm) with small effective area 3.03µm^2 and high nonlinear gain 67.23W^{-1}km^{-1} at operating wavelength. So this fiber is a suitable candidate for telecommunication application.

References

[1] J. C. Knight, T. A. Birks, P. St. J. Russell, and D. M. Atkin, "All-silica single-mode optical fiber with photonic crystal cladding," Opt. Lett. 21, 1547-1549 (1996)

[2] A. Ferrando, E. Silvestre, J. J. Miret, and P. Andres, "Nearly zero ultraflattened dispersion in photonic crystal fibers," Opt. Lett. 25, 790-792 (2000)

[3] A. Ferrando, E. Silvestre, P. Andres, J. J. Miret, and M. V. Andres, "Designing the properties of dispersion- flattened photonic crystal fibers Opt. Express 9, 687-697 (2001),

[4] W. H. Reeves, J. C. Knight, P. St. J. Russell, and P. J. Roberts, "Demonstration of ultra-flattened dispersion in photonic crystal fibers," Opt. Express 10, 609-613 (2002),

[5] J. C. Knight, "Photonic crystal fibers," Nature 424, 847-851 (2003).

[6] K. Saitoh, M. Koshiba, T. Hasegawa, and E. Sasaoka, "Chromatic dispersion control in photonic crystal fibers: application to ultra-flattened dispersion," Opt. Express 11, 843-852(2003),

[7] V. Finazzi, T. M. Monro, and D. J. Richardson, "Small core silica holey fibers: Nonlinearity and confinement loss trade-offs," J. Opt. Soc. Am. B 20, 1427-1436 (2003)

[8] M. D. Nielsen, C. Jacobsen, N. A. Mortensen, J. R. Folkenberg, and H. R. Simonsen, "Low-loss photonic crystal fibers for transmission system and their dispersion properties," Opt. Express 12, 1372-1376(2004),

[9] K. Saitoh and M. Koshiba, "Highly nonlinear dispersion-flattened photonic crystal fibers for supercontinuum generation in a telecommunication window," Opt. Express 12, 2027-2032 (2004),

[10] T. L. Wu and C. H. Chao, "A novel ultraflattened dispersion photonic crystal fiber," IEEE. Photon. Technol. Lett. 17, 67-69 (2005).

[11] K. Saitoh and M. Koshiba,"Ultra-flattened chromatic dispersion controllability using a defected core photonic crystal fiber with low confinement loss . *OPTICS EXPRESS 8365* , Vol. 13, No. 21.(2005)

[12] Arti Agrawal, N. Kejalakshmy, J. Chen, B. M. A. Rahman, and K. T. V. Grattan. Golden spiral photonic crystal fiber: polarization and dispersion properties. *OPTICS LETTERS* , Vol. 33, No. 22 . (2008).

[13] Muhammad Nazmul Hossain, M. Shah Alam,Dihan Md. Nuruddin Hasan, and K. M. Mohsin. A Highly Nonlinear Spiral Photonic Crystal Fiber for Tailoring Two Zero Dispersion Wavelengths in the Visible Region. PHOTONICS LETTERS OF POLAND, VOL. 2 (3), 143-145 (2010)

[14] K. Saitoh and M. Koshiba, "Full-vectorial imaginary-distance beam propagation method based on a finite element scheme: application to photonic crystal fibers," IEEE J. Quantum Elencron. 38, 927-93 3 (2002).

[15] R.T. Bise,and D.J.Trevor "Sol-gel derived microstructured fiber Fabrication and characterization,"in Optical Fiber Comm Conf.(OFC), Washington, DC, Mar.2005,3, Optical Society of America

[16] Ge-SiO2 dispersion Appl Optics 1984

[17] M. Pourmahyabadi* and Sh. Mohammad Nejad," Numerical Analysis of Index-Guiding Photonic Crystal Fibers with Low Confinement Loss and Ultra-Flattened Dispersion by FDFD Method", Iranian Journal of Electrical & Electronic Engineering, Vol. 5, No. 3, Sep. 2009

[18] G.P.Agrawal,"Nonlinear Fiber Optics",Fourth Edition

Prospects of Zinc Sulphide as an Alternative Buffer Layer for CZTS Solar Cells from Numerical Analysis

F. Haque,[1] N.A. Khan,[1] K.S. Rahman,[1] M.A. Islam,[1] M. M. Alam,[3] K. Sopian[1] and N.Amin[1,2,3*]

[1]Solar Energy Research Institute (SERI), The National University of Malaysia, 43600 Bangi, Selangor, Malaysia

[2]Department of Electrical, Electronic and Systems Engineering, The National University of Malaysia, 43600 Bangi, Selangor, Malaysia

[3]Advanced Materials Research Chair, Chemistry Department, College of Sciences, King Saud University, Riyadh 11451, Kingdom of Saudi Arabia

[*]nowshad@eng.ukm.my

Abstract—**Zinc Sulphide (ZnS) is a promising candidate to be an alternative buffer layer to the commonly used cadmium sulphide (CdS) in CZTS solar cells. In this study, buffer layer parameters like layer thickness and buffer layer bandgap have been investigated by Analysis of Microelectronic and Photonic Structures (AMPS-1D) to find out the higher conversion efficiency. A promising result has been achieved with an efficiency of 14.49% (with Voc = 0.81 V, Jsc = 28.85 mA/cm^2 and Fill factor = 67.5) by using ZnS as a buffer layer. It is also found that the high efficiency of CZTS absorber layer thickness is between 2 µm and 4 µm. From the simulation results, it is revealed that higher efficiency can be achieved for the buffer layer bandgap around 3.10 eV - 3.25 eV. This result can be explained by the practical work as the bandgap of ZnS is largely dependent on the preparation conditions and stoichiometry. In conclusion, numerous influences of buffer layer are investigated in CZTS solar cell that can lead to the fabrication of high efficiency devices.**

Index Terms—**CZTS absorber layer, ZnS buffer layer, AMPS 1-D, Thin film solar cells.**

I. Introduction

Thin film photovoltaics (PV) has achieved greater interest during the last few decades for its low material usage, low temperature processing method, variety of deposition process, compatibility with low cost substrates, and also low production costs [1], [2]. Thin film solar cells based on Cadmium Telluride (CdTe), Copper Indium Diselenide (CIS) and Copper Indium Gallium Diselenide (CIGS) has already reached commercialization stage. However, factors like the limitations in the supply of Indium (In) and Tellurium (Te) due to scarcity issue and the restrictions in the usage of heavy metal Cadmium (Cd) because of the toxicity issue have raised concerns about limitations in PV production capacity [3]. In this context, Cu_2ZnSnS_4 (CZTS) as a p-type compound semiconductor having an optimal bandgap of 1.4-1.6 eV in the visible range and a high optical absorption co-efficient (>10^4 cm^{-1}) has attracted much attention [4], [5]. It consists of abundant, inexpensive and non-toxic materials which makes it a strong and promising candidate as a photovoltaic absorber layer to replace CIGS [6], [7]. CZTS based thin film solar cells are still in the R&D stage and are being examined by extensive research activities in order to enhance the overall efficiency for using as a potential absorber for thin film solar cells [8]. To date, the highest efficiency obtained from CZTS solar cells are found as 8.4% and 11.1% for pure sulphur (CZTS) and mixed selenosulpfur (CZTSSe), respectively [9], [10].

To achieve well-behaving PV devices, the selection of an appropriate buffer layer material is very important. In thin film solar cells, buffer layer is an intermediate layer which lies between the absorber and window layer [11]. A buffer layer is expected to make a stable and good p-n junction with the p-type absorber to facilitate electrical conduction. An n-type material with higher bandgap is desired as a buffer layer in order to ensure maximum transmission of photons to reach the absorber layer [12]. Till now, the n-type compound semiconductor Cadmium Sulphide (CdS) having a bandgap of 2.45 eV is the most common and widely used material as a buffer layer [13], [14]. But researchers are in the quest for alternative non-toxic buffer layer materials due to the presence of toxic Cd in CdS [15]. Among different potential alternative materials Zinc Sulphide (ZnS) an n-type compound semiconductor having a wide direct bandgap of 3.5-3.8 eV is considered as one of the most promising materials [16], [17]. The higher bandgap energy of ZnS, compared to CdS is capable of enhancing the blue response of the photovoltaic cells by reducing photon absorption loss. Also the study of the lattice parameters of ZnS and CZTS reveal that, it is an excellent choice for obtaining lattice matched buffer-absorber interface which is another crucial factor for obtaining a high efficiency solar cell [18].

The main theme of this simulation work is to investigate the performance of a CZTS solar cell containing an alternative Cd-free non-toxic buffer layer of ZnS.

II. Methodology

Numerical simulation has now become an essential part of thin film photovoltaic cell designing. It is an important way to predict cell performance and to test the viability of the proposed structure. In this work, Analysis of Microelectronic and Photonic Structures (AMPS-1D) was used to investigate the CZTS solar cell performance with ZnS as the buffer layer. The AMPS-1D program has been developed to realistically simulate the electrical characteristics of the thin film heterojunction solar cells. It has already proven its role as a powerful tool in understanding the device operation and physics for single crystal, poly-crystal and amorphous structures. The schematic view of the structure used in this study is shown in Fig. 1. In the simulation, n-type ZnO:Al, n-type ZnS, p-type CZTS layer parameters are emphasized to investigate the benefits of using ZnS as an alternative buffer

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

layer in CZTS solar cells. By incorporating the various material parameters into AMPS-1D, for all the analysis aspects, changes in the values for efficiency, open circuit voltage, short circuit current and fill factor as well as the effect of different operating temperatures are observed. CZTS is a polycrystalline material and some physical parameters, especially bandgaps depend on doping conditions and compositional ratio as well [20]. The same goes true for other poly crystalline compound semiconductors as well. The role of CZTS absorber layer thickness on the cell performance is investigated by varying the thickness from 1 µm to 4 µm. For the ZnS buffer layer, the bandgap is varied from 3.10 eV to 3.40 eV. The thickness of the same layer is varied from 30 nm to 80 nm to observe the role of buffer layer thickness. Finally, the effect of operating temperature on the cell performance is explored for temperatures ranging from 300 K to 400 K. Table I shows the description of the entire parameters used in this analysis. These values were chosen on the basis of theoretical considerations, experimental data and existing literature or in some cases by reasonable estimations [19], [20].

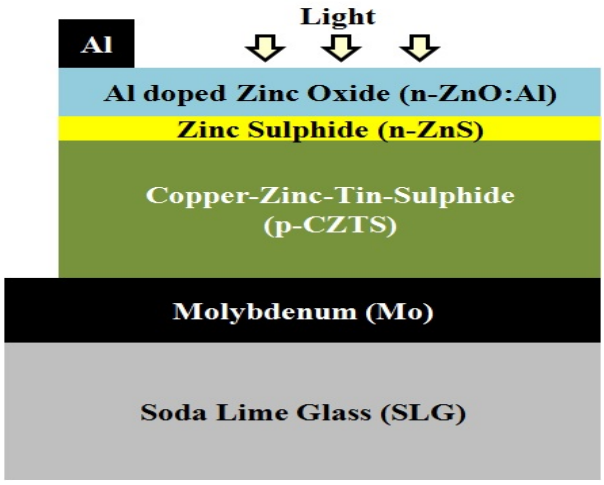


Fig. 1 CZTS solar cell structure used in the simulation.

TABLE I
SEMICONDUCTOR PARAMETERS USED IN THE SIMULATION

Parameters	n-ZnO:Al	n-ZnS	p-CZTS
Thickness, W (µm)	0.2	0.03-0.08	1-4
Dielectric ratio, $\varepsilon/\varepsilon_o$	9.0	10	10
Electron mobility, μ_n (cm^2/Vs)	100	100	100
Hole mobility, μ_p (cm^2/Vs)	25	25	25
Carrier concentration, n/p (cm^{-3})	10^{18}	9.0×10^{17}	2×10^{15}
Bandgap, Eg (eV)	3.30	3.10-3.40	1.40
CB Density of states, N_C (cm^{-3})	2.2×10^{18}	1.5×10^{18}	2.2×10^{18}
VB Density of states, N_V (cm^{-3})	1.8×10^{19}	1.8×10^{18}	1.8×10^{19}
Electron affinity, χ (eV)	4.60	4.50	4.10

III. RESULTS AND DISCUSSION

A. Variation of CZTS absorber layer thickness

In order to explore the effects of absorber layer thickness in the overall cell performance, the CZTS absorber layer is varied from 1 µm to 4 µm. During the simulation, the ZnS buffer layer thickness was considered as 40 nm. It has been found that, the efficiency of the solar cell increases with the increase in absorber layer thickness as shown in Fig. 2. Actually thicker absorber layer increases the minority carrier diffusion length. This means the diffusion length for the electrons in the p-type absorber increases. The thicker absorber layer allows more photons with longer wavelengths to be absorbed resulting in more electron-hole pair (EHP) generation. The generated electrons are subsequently collected resulting improved cell efficiency. Moreover, the depletion region also increases with the increased absorber layer thickness. Hence, electrons can drift towards the n-type buffer due to the electric field produced by the depletion region easily.

From Fig. 2, it is observed that, the solar cell output parameters are almost constant for the absorber layer thickness over 3 µm. This phenomenon indicates that, there lies an optimum value of thickness beyond which the output parameters do not show significant improvement. For the proposed cell in this work, the optimum value of absorber layer thickness is considered as 3 um. A conversion efficiency of 14.49% (V_{oc}= 0.81 V, J_{sc}= 28.85 mA/cm^2, FF = 67.5) has been achieved with the optimum thickness of 3 µm CZTS absorber layer.

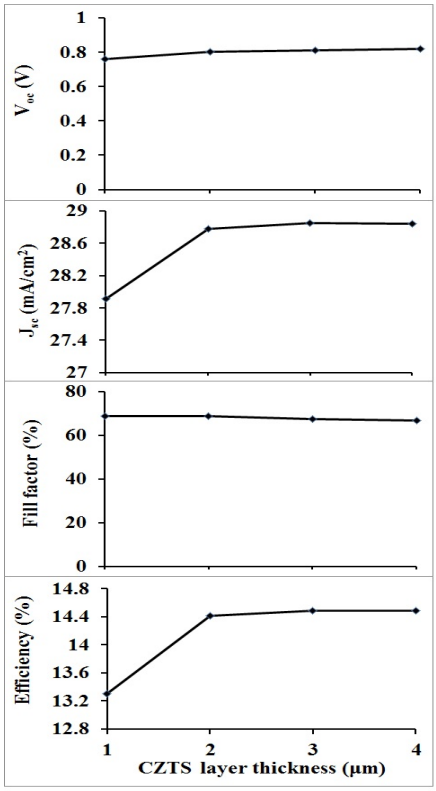


Fig. 2 Cell performance with various CZTS thickness.

B. Variation of ZnS buffer layer thickness

The proposed alternative ZnS buffer layer is free from toxic elements. Both zinc and sulphur are cheap and abundant in the nature. In this simulation, the ZnS buffer layer thickness has been varied between 30 nm to 80 nm to observe the effect of buffer layer thickness. It is found that, the more thinner the buffer, the higher is the efficiency of the overall cell. Basically

thinner buffer layers allow the majority of the photons to be transmitted towards the absorber layer reducing undesired electron hole pair generation in the buffer region. As the thickness of the buffer layer is raised, the efficiency tends to drop as the photons are being absorbed in the buffer layer portion of the thin film solar cell.

Solar cell with 40 nm of ZnS buffer layer recorded the highest efficiency for the proposed thin film solar cell. Although from the simulation results, shown in Fig. 3 depicts the possibility of raising the efficiency value with thinner buffer layers, but from the practical point of view a question will be raised about their feasibility. A trade off or an optimum value of buffer layer thickness must be maintained, as the buffer layer serves as a heterojunction partner for the CZTS absorber layer. Moreover, it serves as a protection layer during TCO sputtering and insulates pin-holes to avoid shorting between the TCO and absorber layers. Fabricating a thin film solar cell with minimum buffer layer thickness may damage the whole cell by shorting the absorber layer and TCO layer (used for the collection of carriers).

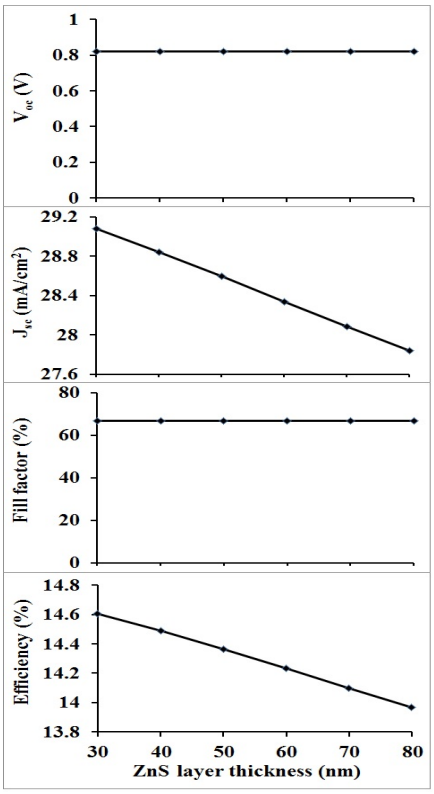


Fig. 3 Overall performance in respect to various ZnS thickness.

C. Effect of the buffer layer bandgap

The bandgap of the buffer layer plays an important role for heterojunction thin film solar cells. The bandgap value for ZnS is typically 3.54 eV for cubic phase. Whereas, the bandgap for conventional CdS buffer layer is 2.45 eV. The higher bandgap of ZnS compared to the conventional CdS can enhance the blue response of the photovoltaic cell by reducing absorption towards the lower wavelength region. Although theoretically, the buffer layer of ZnS lies between 3.5-3.7 eV but it can be tuned to match with the absorber and window layer by varying different deposition parameters like substrate temperature, pressure, stoichiometry, deposition techniques, etc. In our previous work [13], smaller values of ZnS bandgap than the theoretical range stated earlier were found by practical deposition. In this study, the bandgap of ZnS is varied between 3.10 eV to 3.40 eV. The reason for choosing slightly lower bandgap value than the typical for ZnS is because of the band bending in the TCO/ZnS interface. Since the TCO material ZnO:Al has a bandgap of 3.30 eV, therefore, ZnS having bandgaps higher than the ZnO:Al will result in lower efficiencies. Fig. 4 shows the variation of V_{oc}, J_{sc}, Fill factor and efficiency with the change of bandgap for the ZnS buffer layer.

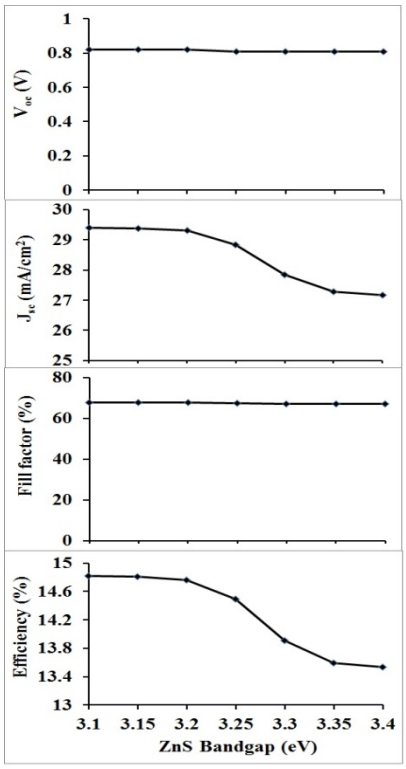


Fig. 4 Cell performance with various ZnS bandgap.

D. Operating temperature effects

The operating temperature of the solar cell plays a crucial role in determining the performance of the solar cell. It affects the parameters such as: effective density of states, absorption coefficients, electron and hole mobility, bandgaps and carrier concentration of the materials. The effect of operating temperature has been investigated in this work for the temperatures ranging from 300 K to 400 K. From Fig. 5 it is observed that, the efficiency and other output parameters degrade drastically with the increase in the operating temperature. With higher temperature, an additional thermal energy is gained by the electrons. This extensive energy gained by the electrons create an unstable situation and causes to recombine with the holes before they could reach the depletion region and get collected.

The reason for the decreasing trend of V_{oc} and J_{sc} with higher operating temperatures can be understood from this phenomenon. The optimum temperature for solar cell operation is 300 K, which is room temperature. But in practical scenarios, the solar panels are installed outside with direct exposure of sunlight in an open environment. The operating temperature of a solar panel increases as the sun light causes heating of the solar panels. Hence the solar cell usually operates at temperatures higher than 300 K.

Considering the baseline parameters, the final J-V curve is shown in Fig. 6.

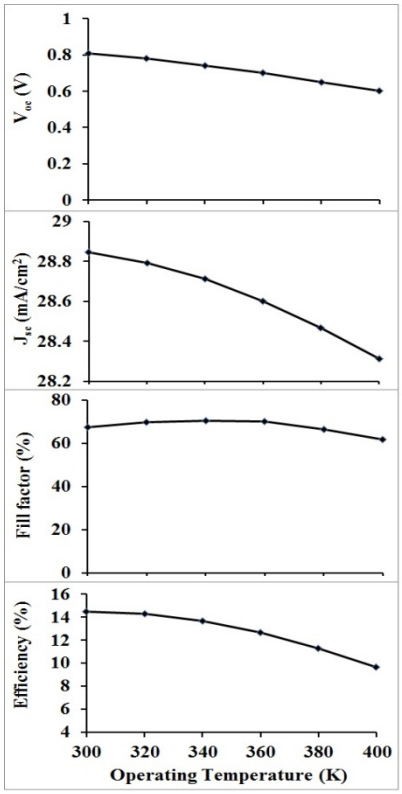


Fig. 5 Cell performance with various operating temperatures.

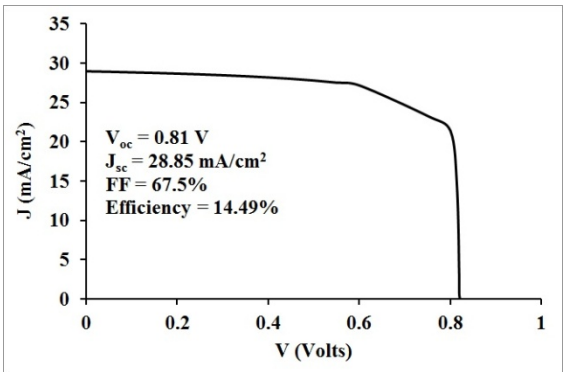


Fig.6 J-V characteristics of the best CZTS cell.

IV. Conclusions

From this numerical analysis, the electrical performances of the CZTS solar cells having ZnS as an alternative to conventional CdS buffer layer have been investigated in terms of buffer layer bandgap and thickness. Increasing absorber layer thickness results in higher efficiency upto a certain extent. Changing the buffer layer thickness also results in higher efficiency as well as improved J_{sc}. There is also room for improvement of the overall efficiency of the proposed solar cell by tuning the bandgap of the buffer layer material. The highest efficiency of 14.49% (with Voc = 0.81 V, Jsc = 28.85 mA/cm2 and Fill factor = 67.5) has been achieved after optimization of the structure. A new challenge to fabricate large-area CZTS thin film solar cell with ZnS as a non toxic and Cd-free buffer layer has been opened up by transferring the new device structure case developed in AMPS-1D.

Acknowledgment

This work is the Ministry of Higher Education research grant with codes FRGS/1/2013/TK07/UKM/01/3 and ERGS/1/2012/TK07/UKM/02/4. Appreciations are also extended to the Solar Energy Research Institute (SERI), The National University of Malaysia and the VP unit of King Saud University (KSU) of Saudi Arabia for invaluable contributions.

References

[1] V. Kheraj, K.K. Patel, S.J. Patel, D.V. Shah, "Synthesis and characterization of Copper Zinc Tin Sulphide (CZTS) compound for absorber material in solar-cells," *Journal of Crystal Growth*, vol. 362, pp. 174-177, 2013.

[2] K. L. Chopra, P. D. Paulson, V. Dutta, "Thin-Film Solar Cells: An Overview," *Progress in Photovoltaic Research and Applications*, vol. 12, pp. 62-69, 2004.

[3] H. Wang, "Progress in thin film solar cells based on Cu_2ZnSnS_4," *International Journal of Photoenergy*, Article ID 801292, 2011.

[4] A.I. Inamdar, S. Lee, K. Jeon, C. H. Lee, S.M. Pawar, R. S. Kalubarme, C.J. Park, H. Im, W. Jung, H. Kim, "Optimized fabrication of sputter deposited Cu_2ZnSnS_4 (CZTS) thin films," *Solar Energy*, vol. 91, pp. 196-203, 2013.

[5] L. Li, B.L. Zhang, M. Cao, Y. Sun, J.C. Jiang, P.F. Hu, Y. Shen, L.J. Wang, "Facile synthesis of Cu_2ZnSnS_4 nanocrystals and its use for dye-sensitized solar cells applications," *Journal of Alloys and Compounds*, vol.551, pp. 24-29, 2013.

[6] J. Wang, S. Li, J. Cai, B. Shen, Y. Ren, G. Qin, " Cu_2ZnSnS_4 thin films: Facile and cost-efffective preparation by RF-magnetron sputtering and texture control," *Journal of Alloys and Compounds*, vol. 552, pp. 418-422, 2013.

[7] M.A. Islam, K.S. Rahman, F. Haque, M. Akhtaruzzaman, M.M. Alam, Z.A. Alothman, K. Sopian, N. Amin, "Properties of low temperature vacuum annealed CZTS thin films deposited on polymer substrate," *Chalcogenide Letters*, Vol. 11, No. 5, pp. 233-239, 2014.

[8] H. Katagiri, " Cu_2ZnSnS_4 thin film solar cells," *Thin Solid Films*, vol. 480-481, pp. 426-432, 2005.

[9] Y. Zhao, W. Zhou, J. Jiao, Z. Zhou, S. Wu, "Aqueous synthesis and characterization of hydrophilic Cu_2ZnSnS_4 nanocrystals," *Materials Letters*, vol. 96, pp. 174-176, 2013.

[10] C.M. Fella, Y.E. Romanyuk, A.N. Tiwari, "Technological status of Cu2ZnSn(S,Se)4 thin film solar cells," *Solar Energy Materials and Solar Cells*, vol. 119, pp. 276-277, 2013.

[11] M. I. Hossain, P. Chelvanathan, M. Zaman, M.R. Karim, M.A. Alghoul, N. Amin, "Prospects of Indium Sulphide as an alternative to Cadmium Sulphide buffer layer in CIS based solar cells from numerical analysis," *Chalcogenide Letters*, Vol. 8, No. 5, pp. 315-324, 2011.

[12] S. Abermann, "Non-vacuum processed next generation thin film photovoltaics: Towards marketable efficiency and production of CZTS based solar cells," *Solar Energy*, vol. 94, pp. 37-70, 2013.

[13] F. Haque, K.S. Rahman, M.A. Islam, M.J. Rashid, M. Akhtaruzzaman, M.M. Alam, Z.A. Alothman, K. Sopian, N. Amin, "Growth Optimization of ZnS thin films by RF magnetron sputtering as prospective buffer layer in thin film solar cells," *Chalcogenide Letters*, Vol. 11, No. 4, pp. 189-197, 2014.

[14] Y. Bouznitt, Y. Beggah, A. Boukerika, A. Lahreche, F. Ynineb, "New co-spray way to synthesize high quality ZnS films," *Applied Surface Science*, vol. 284, pp. 936-941, 2013.

[15] A. Wei, J. Liu, M. Zhuang, Y. Zhao, "Preparation and characterization of ZnS thin films prepared by chemical bath deposition," *Materials Science in Semiconductor Processing*, vol. 16, pp. 1478-1484, 2013.

[16] D. H. Hwang, J. H Ahn, K. N. Hui, K. S. Hui, Y. G. Son, "Structural and optical properties of ZnS thin films deposited by RF magnetron sputtering," *Nanoscale Research Letters*, vol. 7:26, 2012.

[17] P. K. Nair, M. T. S. Nair, V. M. Garcia, O.L. Arenas, Y. Pena, A. Castello, I.T. Ayala, O. Gomezdaza, A. Sanchez, J. Campos, H. Hu, R. Suarez, M. E. Rincon, "Semiconductor thin films by chemical bath deposition for solar energy related applications," *Solar Energy Materials & Solar Cells*, vol. 52, pp. 313-344, 1998.

[18] Le-Xi Shao, Kuen-Huei Chang, Huey-Liang Hwang, "Zinc sulfide thin films deposited by RF reactive sputtering for photovoltaic applications," *Applied Surface Science*, vol. 212-213, pp. 305-310, 2003.

[19] N. Amin, M.I. Hossain, P. Chelvanathan, M. Zaman, K. Sopian, "Prospects of Cu_2ZnSnS_4 (CZTS) solar cells from numerical analysis," in 6th International Conference on Electrical and Computer Engineering, *ICECE 2010*, Dhaka, Bangladesh, 2012.

[20] M.A. Olopade, O.O. Oyebola and B.S. Adeleke, "Investigation of some materials as buffer layer in copper zinc tin sulphide (Cu_2ZnSnS_4) solar cells by SCAPS-1D," Pelagia Redearch Library, vol. 3, No. 6, pp. 3396-3400, 2012.

Design Optimization of CdTe Thin Film Solar Cells from Numerical Analysis

N. A. Khan[1], K. S. Rahman[1], F. Haque[1], N. Dhar[1], M. A. Islam[1], M. Akhtaruzzaman[1], K. Sopian[1] and N. Amin[1,2,3*]

[1]Solar Energy Research Institute, The National University of Malaysia, 43600 Bangi, Selangor, Malaysia
[2]Department of Electrical, Electronic and System Engineering, FKAB,
The National University of Malaysia, 43600 Bangi, Selangor, Malaysia
[3]Advanced Materials Research Chair, Chemistry Department, College of Sciences,
King Saud University, Riyadh 11451, Kingdom of Saudi Arabia

[*]E-Mail: nowshad@eng.ukm.my

Abstract— **In this paper, a modified structure for CdTe thin film solar cell was proposed by numerical analysis with an addition of a novel ZnO buffer to improve the conversion efficiency. The CdS window layer was reduced to 50 nm together with the insertion of zinc oxide (ZnO) as the buffer layer to prevent forward leakage current. The thickness of CdTe absorber layer was varied from 1000 nm to 5000 nm and as well as the operating temperature was also varied from 25°C to 165°C. The numerical simulation was done by Analysis of Microelectronic and Photonic Structures (AMPS-1D) simulator. The highest conversion efficiency obtained was 20.27% (Voc = 1.08 V, Jsc = 24.93 mA/cm^2, FF = 0.83) with 5000 nm CdTe absorber layer and 50 nm CdS window layer. Moreover it was observed that with the increase in operating temperature, the normalized efficiency decreased linearly at a gradient of 0.2%/°C, which reveals that the CdTe thin film solar cell has higher stability.**

Index Terms—**Thin film solar cell, CdTe, ZnO, AMPS-1D.**

I. Introduction

The solar energy is the most reliable and abundant form of renewable energy today that may serve as a major alternative option for the ever growing need of energy where energy from non-renewable sources are becoming scarce with time. The environmental friendly feature and its abundant nature made renewable energy as one of the most promising topic in present times. The solar cells which work on the principle of photovoltaic effect could be a possible alternative energy supplier to meet the crisis. Thin film based solar cells are now becoming a popular topic among the researchers and commercials because of its higher efficiency, low manufacturing cost and better stability [1]-[4]. Cadmium Telluride (CdTe) based solar cells are ideal because of its higher absorption coefficient (>5×10^5/cm) and direct band gap energy (Eg) of 1.45eV which corresponds quite significantly with the sunlight spectra [5]-[7]. CdTe is a very promising absorber layer because of its long term performance stability in thin film heterojunction solar cells [8]. Furthermore, various deposition techniques could be used to deposit CdTe thin films which include close space sublimation (CSS), vapour transport deposition (VTD), physical vapour deposition (PVD), metal organic chemical vapour deposition (MOCVD), electro-deposition and screen printing [9]. The objective of this paper is to obtain a highly efficient and stable CdTe solar cell which had been studied using AMPS-1D simulator. The effect of varying absorber layer thickness and variation in cell operating temperature were analysed in this work. The highest efficiency (□) obtained for the variation in CdTe absorber layer thickness and variation in operating temperature were 20.27% and 18.89% respectively.

II. Modelling and Simulation

The reason behind numerical analysis for a solar cell is a vital way to predict the cell performance and to check the feasibility of the proposed structure. The parameters that can be varied during modelling of a solar cell are more than 50 and these parameter values depend on deposition methods, fabrication techniques, ambient condition, etc. The simulation for this work was done by Analysis of Microelectronic and Photonic Structures (AMPS-1D) to study the performance of CdTe thin film solar cell. The AMPS-1D simulation program has been tested for CdTe and CIGS thin film solar cells and the program has been developed to realistically simulate the electrical characteristics of the thin film heterojunction solar cells.

Fig.1 below shows the conventional baseline structure for CdTe thin film solar cell as well as the modified superstrate structure (Glass/SnO_2/ZnO/Cds/CdTe/Ag) to obtain better conversion efficiency. The modified structure has a new or an extra layer (ZnO) in between TCO and CdS to obtain better performance and thus it can be said that the front contact consists of TCO and a buffer layer (ZnO). In the modified structure, silver (Ag) is used as the back contact instead of gold (Au) keeping the low cost feature into consideration.

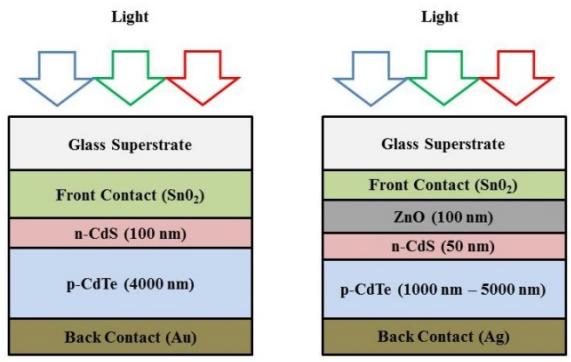


Fig.1 (a) Conventional CdTe Solar Cell Structure (b) Modified CdTe Solar Cell Structure

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

In this analysis after introducing a buffer layer (ZnO), at first the thickness of the CdTe absorber layer were varied from 1000 nm to 5000 nm and secondly the operating temperature were varied from 25°C to 165°C . It was ensured that only one parameter varied at a time during the simulation process whereas the other parameters were kept fixed.

The table below illustrates the details of all the parameters used in this analysis which were selected on the basis from literature, theory and experimental data [7], [10]-[11].

TABLE I
PARAMETERS USED FOR SIMULATION

Parameters	SnO_2	ZnO	n-Cds	p-CdTe
Thickness, W (μm)	0.5	0.1	0.05	1.00 - 5.00
Dielctric ratio, ε/ε□	9.0	9.0	10.0	9.40
Electron mobility, μ_C (cm^2/Vs)	100	100	100	320
Hole mobility, μ_P (cm^2/Vs)	25	25	25	40
Carrier concentration, n, p, (cm^{-3})	1.0×10^{20}	1.5×10^{18}	1.1×10^{18}	1.0×10^{18}
Band gap, E_g (eV)	3.30	3.30	2.40	1.50
Density of states, N_C (cm^{-3})	2.2×10^{18}	2.2×10^{18}	2.2×10^{18}	7.5×10^{17}
Density of states, N_V (cm^{-3})	1.8×10^{19}	1.8×10^{19}	1.8×10^{19}	1.8×10^{19}
Electron affinity, χ (eV)	4.20	4.50	4.50	4.28

III. RESULTS AND DISCUSSION

This simulation work is done by insertion of a new layer of ZnO to the baseline structure to improve the performance of the conventional CdTe solar cells. The Cds layer was reduced to 50 nm in order to achieve better performance of CdTe thin film solar cells but this reduced thickness might allow forward leakage current to flow to the front contact through possible pinholes in CdS layer. To prevent this undesired forward leakage current a high resistive buffer layer of ZnO was inserted in between the front contact TCO and CdS window layer. The starting point of this work was by modelling a device which comprises of four layers SnO_2/ZnO/CdS/CdTe respectively. The structure for the device was kept fixed during the full simulation process. The performance of the cell and its dependency on the CdTe layer thickness from 1000 nm to 5000 nm and on the operating temperature from 25 °C to 165 °C had been simulated using the AMPS-1D simulator.

A. Effect of CdTe Layer Thickness

The theoretical document shows that the minimum thickness required for CdTe thin films to absorb 99% of the incident photons with energy greater than the band gap (Eg) is approximately 1 to 2 μm [12]. In this work, the CdTe layer thickness was varied from 1000 nm to 5000 nm and the cell output parameters such as Jsc, Voc, FF and η were measured using AMPS-1D simulation. Then the J-V curve and graph for Q.E was plotted for the parameters which showed the highest efficiency during thickness variation. The cell output parameters which are Jsc, Voc, FF and η had been shown in Fig: 2 with the variation in CdTe layer thickness.

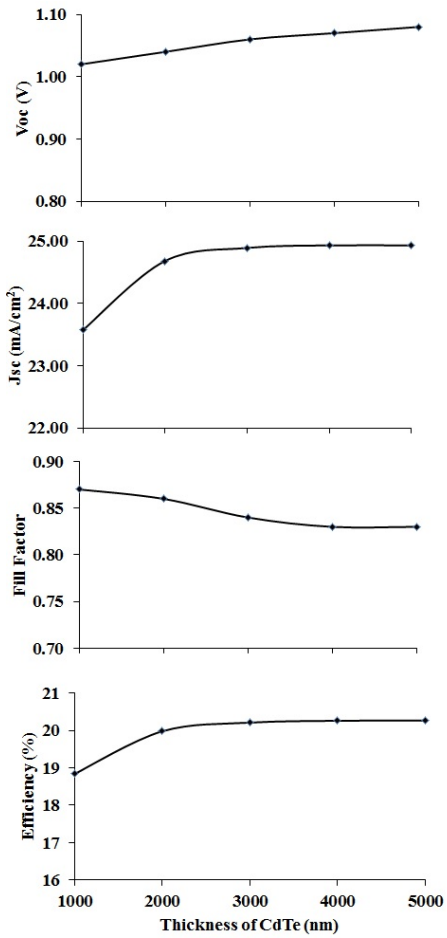


Fig.2 Effect of CdTe thickness on cell output parameters

The variation in CdTe absorber layer thickness caused the output parameters to change gradually. The value of Voc improved at a slow rate with the increase in thickness but the change in Jsc was significant when the thickness of CdTe absorber layer increased from 1000 nm to 2000 nm and after that the Jsc value increased at a slow rate and had a value of 24.93 mA/cm^2 when the thickness went up to 5000 nm. On the other hand, the FF reduced slightly and then almost remained constant from 3000 nm to 5000 nm thickness of CdTe absorber layer. However the efficiency improved significantly with the increase in CdTe absorber layer thickness from 1000 nm to 3000 nm and then the rate of increase became less and the highest conversion efficiency attained was 20.27% (Voc = 1.08 V, Jsc = 24.93 mA/cm^2, FF = 0.83) at 5000 nm thickness of CdTe absorber layer. The J-V curve and the QE graph were illustrated in Fig: 3. Improvement in cell efficiency with the increase in thickness of CdTe absorber layer was evident from the J-V curve and the figure also showed that the final cell output characteristics were having acceptable low series resistance and high shunt resistance, confirming higher Jsc, Voc, FF and which eventually contributed to higher conversion efficiency. The QE graph showed that the light absorption was good from wavelengths of 450 nm to 750 nm.

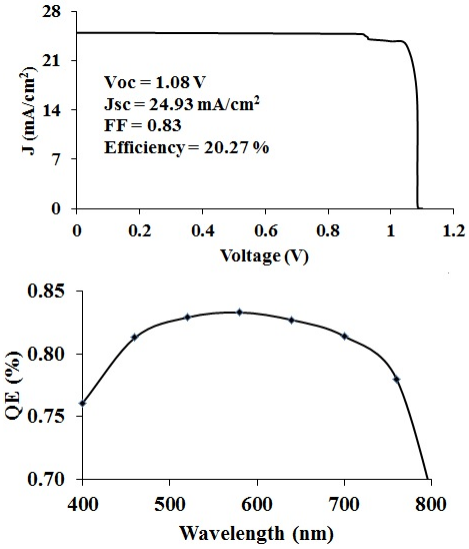


Fig.3 J-V and QE characteristics for thickness variation of CdTe

B. *Effect of Operating Temperature*

The operating temperature plays a very important part in real situations that affects the performance of the overall solar cell. The parameters like the effective density of states, absorption coefficients, electron and hole mobility, carrier concentrations and band gaps of the materials are affected at higher operating temperatures [13]. In this part of the simulation work, the operating temperature of the solar cell was varied from 25 °C to 165 °C. The parameters such as Voc, Jsc, FF and η were measured and illustrated in Fig.4.

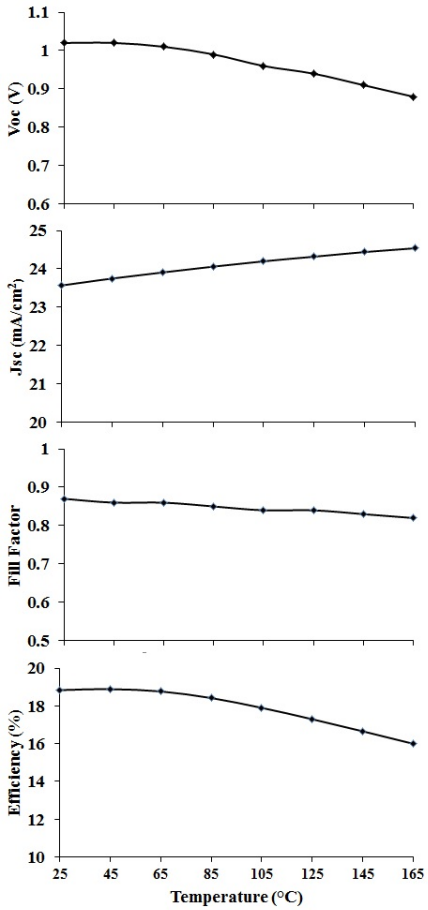


Fig.4 Effect of temperature on cell output parameters

The value of Voc decreased gradually with the increase in operating temperature and similar trend was observed for the FF graph. However the value of Jsc somewhat increased steadily with the rise in operating temperature. The highest conversion efficiency obtained was 18.89% (Voc = 1.02 V, Jsc = 24.54 mA/cm^2, FF = 0.82) at 45 °C and after that the efficiency (η) decreased with the increase in temperature up to 165 °C. Then the graph for J-V and QE was plotted for the CdTe thin film solar cell that obtained the highest efficiency during the variation in operating temperature. From the QE plot in Fig.5, it is observed that the maximum light absorption were from between wavelengths of 450 nm to 650 nm.

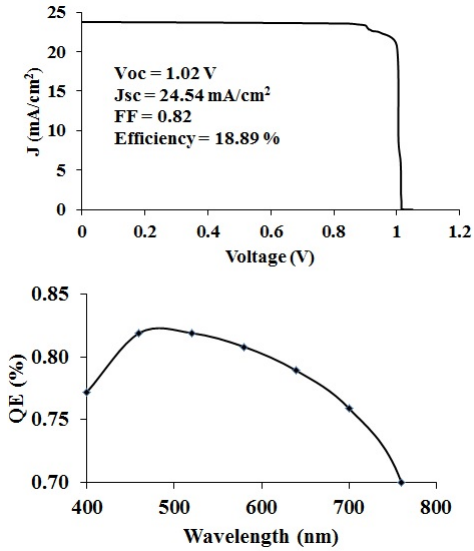


Fig.5 J-V and QE graph for operating temperature variation of CdTe

The effect of temperature could be measured by another parameter named as the temperature coefficient, which means the change of power of solar cell for increase or decrease of operating temperature per degree centigrade. From Fig.6, it is observed that the normalised efficiency (η) decreased linearly at a gradient of 0.2% / °C with the increase in operating temperature which also showed the higher degree of stability of the CdTe thin film solar cell at higher operating temperature or in stressed conditions. The above mentioned statements are in good agreement with the related works [14] – [15].

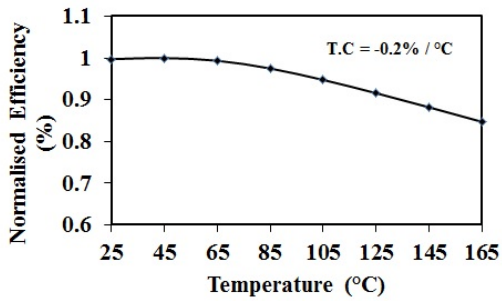


Fig.6 Effect of operating temperature on normalized efficiency

IV. CONCLUSIONS

The numerical analysis in this work showed design optimization of the conventional CdTe based solar cells with minimal changes keeping a cautious concern over minimal production cost. A highly efficient CdTe thin film solar cell with an absorber layer thickness of 5000 nm and CdS thickness of 50 nm has been achieved. Furthermore, by

reducing the thickness of CdS window layer and insertion of ZnO as a buffer layer helped to raise the conversion efficiency up to 20.27% (Voc = 1.08 V, Jsc = 24.93 mA/cm^2, FF = 0.83). Finally, from the temperature coefficient (T.C = -0.2%/°C), it showed that the stability of this CdTe thin film solar cell was good at higher operating temperatures. Therefore, it is hoped that with this modified structure with all realistic approaches or modification, an efficient and cost effective CdTe thin film solar cell can be fabricated using standard fabrication technique for practical implementation.

ACKNOWLEDGMENT

The authors would like to acknowledge and appreciate the contribution of the Ministry of Education of Malaysia through its research grant with code FRGS/1/2013/TK07/UKM/01/3. Appreciations are also extended to the VP unit of the King Saud University (KSU) of Saudi Arabia.

REFERENCES

[1] MD. Sharafat Hossain, Nowshad Amin, M. A. Matin and M. Mannir Aliyu, A Numerical Study on the prospects of high efficiency ultra-thin ZnxCd1-xS/CdTe Solar Cell, *Chalcogenide Letters, Vol. 8, No. 3, March 2011, p. 263 – 272*

[2] X. Wu, J.C. Keane, R.G. Dhere, C. DeHart, D.S. Albin, A. Duda, T.A. Gessert, S. Asher, D.H. Levi, P. Sheldon, 16.5% Efficient CdS/CdTe Polycrystalline *Thin-Film Solar Cell, Proc. of 17th European Photovolt. Sol. Energy Conf., 2001, Munich, Germany*

[3] Nowshad Amin, Akira Yamada and Makoto Konagai, Effect of ZnTe and CdZnTe Alloys at the Back Contact of 1-µm-Thick CdTe Thin Film Solar Cells, *Japanese Journal of Applied Physics 41(5A), Part 1 (2002)*

[4] M. A. Matin, M. Mannir Aliyu, Abrar H. Quadery and Nowshad Amin, Prospects of novel front and back contacts for high efficiency cadmium telluride thin film solar cells from numerical analysis, *SOLMAT, 94, 1496 (2010)*

[5] Aminul Islam, M.A. Matin, Yusuf Sulaiman and Nowshad Amin, A highly efficient CdS:O/CdTe Solar cell by Numerical Analysis, *Regional Engineering Postgraduate Conference (EPC) 2010*

[6] M. A. Matin, Nowshad Amin, Azami Zaharim and Kamaruzzaman Sopian, A Study towards the Possibility of Ultra-Thin CdS/CdTe High Efficiency Solar Cells from Numerical Analysis, *WSEAS Transactions on Environment and Development*

[7] M. A. Islam, Yusuf Sulaiman, Nowshad Amin, A Comparative Study of BSF Layers for Ultra-Thin CdS:O/CdTe Solar Cells, *Chalcogenide Letters, Vol. 8, No. 2, February 2011, p. 65 – 75*

[8] D.I Batzner, A. Romeo, H. Zogg, R. Wendt, A.N. Tiwari, Development of efficient and stable back contacts on CdTe/CdS solar cells, *Thin Solid Films 387 (2001) 151*

[9] Brian E. McCandless and James R. Sites, *Handbook of Photovoltaic Science and Engineering, ISBN: 0-471-49196-9*

[10] M A Islam, Shrarafat Hossain, Nowshad Amin, M. Mannir Aliyu, Yusuf Sulaiman and Kamaruzzaman Sopian, Investigation for Optimum Structure of CdS:O/CdTe Solar Cell from Numerical Analysis, *Advanced Materials Research Vols. 622-623 (2013) pp 1194-1198*

[11] Aminul Islam, M. A. Matin, Mohammed Mannir Aliyu, Yusuf Sulaiman and Nowshad Amin, A Numerical Analysis on CdS:O Window Layer for Higher Efficiency CdTe Solar Cells, *Developments in Renewable Energy Technology (ICDRET), 2009*

[12] Nowshad Amin, Kamaruzzaman Sopian, Makoto Konagai, *Solar Energy Materials and Solar Cells (SOLMAT) 91, 1202 (2007).*

[13] J. Fan, Theoretical temperature dependence of solar cell parameters, *Solar Cells 17, 309 (1986)*

[14] M. A. Matin, Nowshad Amin and Kamaruzzaman Sopian, Ultra-Thin High Efficiency CdS/CdTe Thin Film Solar Cells from Numerical Analysis, *The 3rd International Conference on RENEWABLE ENERGY SOURCES (RES'09), Canary Islands, Spain, July 1-3, 2009*

[15] Nowshad Amin, Mahmud A. Matin and Kamaruzzaman Sopian, Prospects of Novel Front and Back Contacts for High Efficiency Cadmium Telluride Thin Film Solar Cells from Numerical Analysis,

Modeling Combined Effect of Temperature and Irradiance on Solar Cell Parameters by MATLAB/ Simulink

Partha Sarathi Paul*, Sudipta Mondal, Nasrin Akter and Sharif Mohammad Mominuzzaman

Department of Electrical and Electronic Engineering,
Bangladesh University of Engineering and Technology,
Dhaka-1000, Bangladesh
Email: *partha_psp149@yahoo.com

Abstract— **The main objective of this paper is to make a detailed modelling of a single solar cell incorporating basic operating equations and include the dependence of cell temperature on irradiance as a combined effect in the model. For this purpose at first the change of open circuit voltage and short circuit current at varying operating conditions like irradiance and temperature is theoretically explored. After that, a simulink model is made using MATLAB to simulate the separate and combined effect of irradiance and temperature on single solar cell and the output curves are analyzed to compare the model with used module's datasheet. Then fall of open circuit voltage with irradiance due to temperature effect is illustrated and explained in detail .**

Index Terms— **NOCT, PV module, sun simulator, SRE.**

I. Introduction

Due to diminishing fossil fuel energy resources and increasing demand for power, the world is gradually moving towards sustainable renewable energy resources. Sustainability and abundance of solar radiant energy are important factors characterize the energy through the PV (photovoltaic) effect among the renewable resources. Despite the intermittent nature of sunlight solar energy is widely available and completely free of cost. Solar irradiance and temperature are the essential factors that affect the output characteristics of a solar cell. The open circuit voltage increases in logarithmic way with solar irradiation whereas the short circuit current increases linearly. The dominant effect with increasing cell's temperature is the linear decrease of open circuit voltage, the cell is then subjected to less efficiency. The short circuit current slightly increases with temperature.

In brief temperature and irradiance are the key factors which have a substantial effect on the performance of solar cell. Prior to our thesis work, we have come across theoretical and experimental data where the effect of temperature on solar cell is shown but the irradiance is kept constant and other data are found where the temperature is kept constant and irradiance is kept constant. However, it is quite preposterous to assume that one of these variables is kept constant and other is varied indefinitely in environmental operating condition. In reality, both irradiance and temperature can vary simultaneously in environmental operating condition. In our thesis work we have constructed a simulink model of solar cell using MATLAB which can simulate the effect of simultaneous change of irradiance and temperature on solar cell and approximate the environmental operating condition where both the irradiance and temperature of solar cell is liable to change.

II. General Model Representation

A general theoretical model of PV module is studied from the reference [1-3]. As shown in Fig. 1, the equivalent circuit of the general model is composed of photo current source, diode, parallel resistor expressing the leakage current, and series resistor describing the internal resistance to the current flow. The *I-V* characteristic equation of a PV cell is given as:

$$I = I_{PH} - I_S\left(\exp\left[\frac{q(V+IR_S)}{kTA}\right]-1\right) - \frac{(V+IR_S)}{R_{SH}} \quad (1)$$

where, I_{PH} is a light-generated current or photo current, I_S is the cell saturation f dark current, q (=1.6×10^{-19}C) is the electron charge, k (=1.38×10^{-23}J/K) is Boltzmann constant, T is the cell working temperature, A is the ideality factor, R_{SH} is the shunt resistance, and R_S is the series resistance.

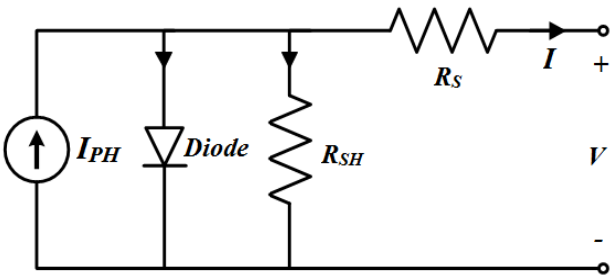


Fig. 1 Equivalent circuit of a solar cell

The photocurrent mainly depends on the solar irradiance and cell working temperature, which is given as:

$$I_{PH} = \lambda\left(I_{SC} + K_I\ (T - T_r)\right) \quad (2)$$

where, I_{SC} is the cell short-circuit current at a 25 ^{0}C and 1 kW/m^2, K_I is the cell short-circuit current temperature coefficient, T_r is the cell reference temperature, and λ is the ratio of ambient solar irradiance to 1sun irradiance (1000 W/m^2). On the other hand, the cell saturation current varies with cell temperature, which is described as:

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

$$I_S = I_{RS} \left(\frac{T}{T_r}\right)^3 \exp\left[q\, E_G \frac{\left(\frac{1}{T_r} - \frac{1}{T}\right)}{kA}\right] \quad (3)$$

where, I_{RS} is the cell reverse saturation current at a reference temperature and a solar radiation, E_G is the band gap energy of the semiconductor used in the cell which can is related to temperature as [4]:

$$E_G(T) = E_G(0) - \frac{\alpha T^2}{T + \beta} \quad (4)$$

where, α and β are constants specific to each semiconductor. For silicon $E_G(0) = 1.169\ eV$, $\alpha = 4.9 \times 10^{-4}\ eV/K$ and $\beta = 655\ K$ [5].

Finally, the reverse saturation current at reference temperature (T_r) can be approximately obtained as:

$$I_{RS} = \frac{I_{SC}}{\exp\left[\frac{qV_{OC}}{N_s kAT}\right] - 1} \quad (5)$$

where, I_{SC} is the short circuit current, N_s is the number of series connected cell and V_{OC} is the open circuit voltage of the module at reference temperature.

III. Temperature Effect

In this section, change of open circuit voltage (V_{OC}) and short circuit current (I_{SC}) with temperature is theoretically explored keeping irradiance level constant.

A. Change of V_{OC}

Temperature has a dominant effect on V_{OC} which can be described by the following equation [6]:

$$V_{OC2} = V_{OC1} \left(\frac{T_2}{T_1}\right) + \frac{E_G}{e}\left(1 - \frac{T_2}{T_1}\right) \quad (6)$$

where, V_{OC1} is the open circuit voltage at temperature $T_1{}^0K$ and V_{OC2} at $T_2{}^0K$. If we take the differential form of equation (5) then we get:

$$\frac{dV_{OC}}{dT} = \frac{1}{q}\frac{dE_G}{dT} - \frac{1}{T}\left(\frac{E_G}{q} - V_{OC}\right) \quad (7)$$

After analyzing these equations we can conclude that for silicon solar cell, for every 1^0C increase in temperature V_{OC} drops by about 0.002/0.55≈0.4% from its room temperature value [7].

B. Change of I_{SC}

Change of short circuit current with temperature is negligible. Actually, there is a very small increase of short circuit current with temperature.

IV. Irradiance Effect

In this section, change of open circuit voltage (V_{OC}) and short circuit current (I_{SC}) with irradiance is theoretically studied keeping temperature constant.

A. Change of V_{OC}

Open circuit voltage changes logarithmically with irradiance (indicated by I_1 and I_2) as follows:

$$V_{OC2} = V_{OC1} + \frac{nk_B T}{e} \ln\left(\frac{I_2}{I_1}\right) \quad (8)$$

B. Change of I_{SC}

Short circuit current is linearly dependent on irradiance which is given by the following equation:

$$I_{SC2} = I_{SC1} \left(\frac{I_2}{I_1}\right) \quad (9)$$

V. Combined Effect of Temperature and Irradiance

Cell temperature is a parameter of great importance as it has immense effect on cell efficiency and other output parameters. It depends on module encapsulating material, it's thermal dissipation, absorption properties, working point and also on atmospheric parameters such as irradiance level, ambient temperature, wind speed etc.

It is very common to use a parameter named NOCT (Nominal Operating Cell Temperature) as an indication of module temperature. NOCT is defined as the mean solar junction temperature with an open rack mounted nodule in SRE (Standard Reference Environment) which is defined as a condition when the tilt angle is at normal incidence to the direct solar beam at local solar noon, total irradiance is 800 W/m^2, ambient temperature is 20 ^{0}C, wind speed is 1m/s, with no electrical load.

To relate the cell temperature with the change of irradiance Ross, R.G. (1980) approximation is used in the mode. The approximation is described as [8]:

$$T_{Cell} = T_{air} + \frac{NOCT - 20}{80} S \times \frac{1000}{10000} \quad (10)$$

where, T_{Cell} is the cell temperature T_{air} is the air temperature and S is the irradiance level.

VI. Simulink Model and Output Curves

The characteristics of MSX 60 PV module as shown in Table I is used in the MATLAB model of single solar cell [9].

Table I
Electrical characteristics of MSX 60 PV module

Description	MSX 60
Maximum power (P_m)	60 W
Voltage at P_{max} (V_m)	17.1 V
Current at P_{max} (I_m)	3.5 A
Guaranteed minimum power	58 W
Short circuit current (I_{SC})	3.8 A
Open circuit voltage(V_{OC})	21.1 A
Temperature co-efficient of V_{OC}	-(80±10) mV/ ^{0}C
Temperature co-efficient of I_{SC}	(0.065±0.15) % / ^{0}C
Temperature co-efficient of power	-(0.5±0.05) % / ^{0}C
NOCT	47±2 ^{0}C

As this module consists of 36 polycrystalline silicon solar cells electrically configured as two series strings of 18 cells each so we divided the voltage parameters by 36 in our model. The used values of parameters in the model are shown in TABLE II.

TABLE II Model parameters	
Parameters	Value
Voltage at P_{max} (V_m)	17.1/36 V
Current at P_{max} (I_m)	3.5 A
Open circuit voltage(V_{OC})	21.1/36 V
Short circuit current (I_{SC})	3.8A
Temperature co-efficient of I_{SC} (K_I)	3 mA/°C
Air temperature (T_{air})	25°C
Reference temperature (T_r)	25°C
NOCT	47°C

Using the parameter values from TABLE II and incorporating the equations from (1) to (5) and (10), a simulink model for single silicon solar cell is procured. By this model, *I-V* and *P-V* curve can be obtained by varying temperature with fixed irradiance as shown in Fig.2 and Fig.3 or varying irradiance with fixed temperature as shown in Fig.4 and Fig.5.

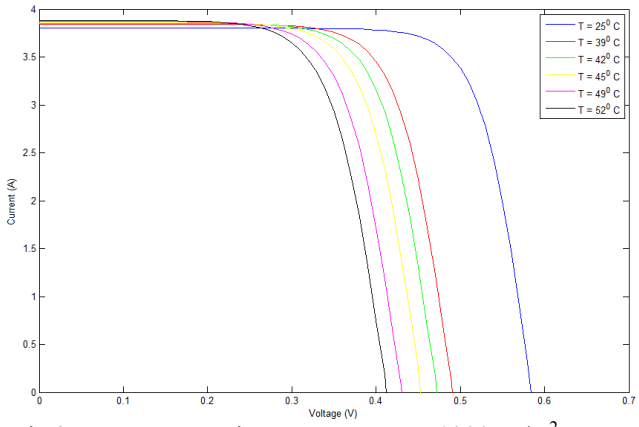


Fig.2 *I-V* curve varying temperature at 1000 W/m^2 constant irradiance level

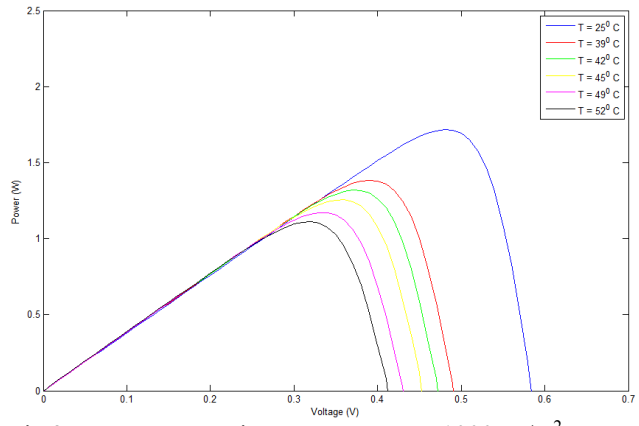


Fig.3 *P-V* curve varying temperature at 1000 W/m^2 constant irradiance level

Again when the relation in equation (10) is used then the combined effect of irradiance and temperature on *I-V* and *P-V* characteristics can be obtained as shown in Fig.6 and Fig.7.

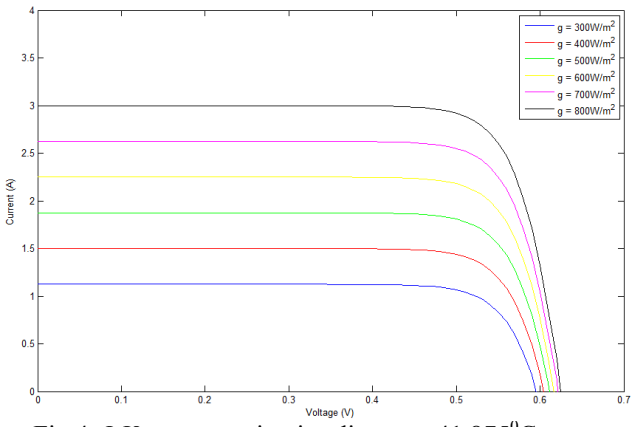


Fig.4 *I-V* curve varying irradiance at 41.875^{0}C constant temperature

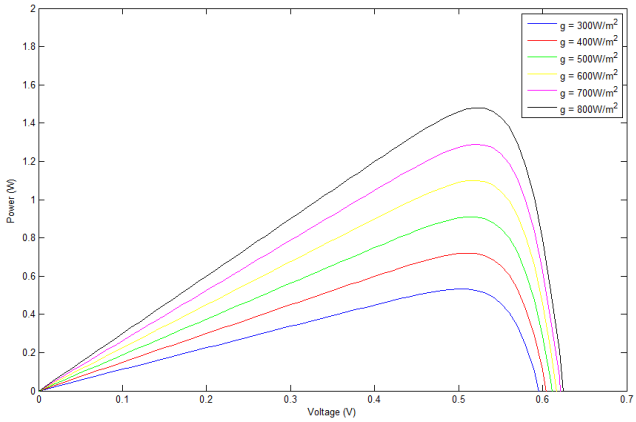


Fig.5 *P-V* curve varying irradiance at 41.875^{0}C constant temperature

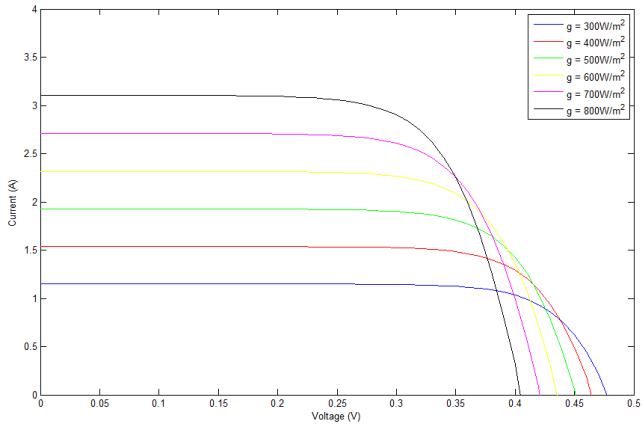


Fig.6 *I-V* curve of the combined effect

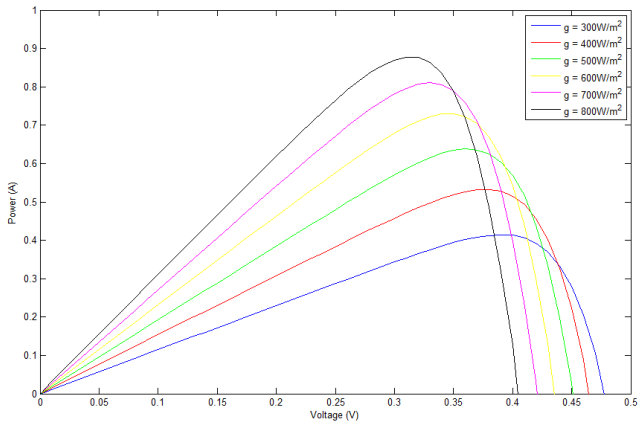


Fig.7 *P-V* curve of the combined effect

In the last case, irradiance is varied as result the cell temperature does not remain constant as in normally practiced modelling cases, rather it varies inherently according to the approximation in equation (10). As a result we get a practical operating scenario of varying cell temperature with irradiance.

VII. Analysis and Discussion

A. Conformity with practical module

In Fig.2 the blue curve is for 1000 W/m^2 irradiance and 25^{0}C operating cell temperature which is the reference operating condition of MSX 60 PV module. From the data sheet of MSX 60 PV module it can be verified that output curve (i.e. V_{OC}= 0.586V=21.1/ 36 and I_{SC}= 3.8A) is almost same at the reference condition as the data sheet value; which implies the accuracy of the simulink model [10].

B. Temperature and Irradiance effect separately

From Fig.2 it is observed that, with the rise of temperature there is a small increase in short circuit current and a remarkable decrease in open circuit voltage as expected theoretically from equation (7). One additional thing from Fig.3 is the fall of power output with the rise of temperature as expected.

From Fig.4, it can be noticed that, with the rise of irradiance the short circuit current rises remarkably as a result of linear relation as shown in equation (9) but the rise open circuit voltage is not so remarkable because of having a logarithmic relation as discussed by equation (8). Fig.5 can be explained as a manifestation of the rise of power with irradiance level.

C. Temperature and Irradiance effect combindly

From Fig.6 it is noticed that, as with the rise of irradiance the cell temperature also rises; both the open circuit voltage and short circuit current changes simultaneously by their dominant issues called temperature and irradiance respectively. Short circuit current rises with irradiance and open circuit voltage falls with temperature remarkably. As a result we get a rarely cited but practically obvious scenario. That is the fall of open circuit voltage with the rise of irradiance due to temperature effect. This is because temperature has a dominant influence on open circuit voltage as the temperature level increases the open circuit voltage drops steeply. Though the open circuit voltage increases logarithmically with irradiance the decrease due to increase of temperature has more far reaching effect on open circuit voltage which not only overcomes this increased amount but also decreases open circuit voltage to some extent. So, we observe an overall decrease in open circuit voltage.

Fig.7 also demonstrates the same concept as in the case of Fig.6. This fall of open circuit voltage with irradiance can be verified experimentally by using sun simulator. As a result of unavailability of this equipment, an experimental data in almost same scenario taken from reference [11] is shown in Fig.8. In this experiment, a polycrystalline silicon solar cell is attached to a metal plate and its temperature is stabilized at different levels like 50^0, 40^0, 30^0 and 20^0 C for various irradiance levels regulated by sun simulator. As shown in Fig.8 it is obvious that, open circuit voltage at high irradiance value (for example 100 mw/cm^2) may fall below of its lower irradiance value (for example 20 mw/cm^2 at 20^{0}C) at an elevated temperature (50^{0}C).

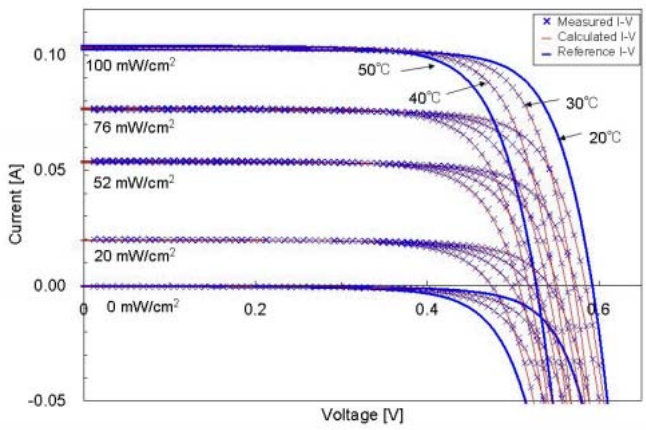


Fig.8 Measured (circles) and calculated (lines) *I-V* curves of a polycrystalline Si solar cell. *I-V* curves measured at irradiance 0 mw/cm^2 and 100 mw/cm^2 and T=20^{0}C and 50^{0}Cwere used for reference *I-V* curves [11].

VIII. Conclusions

In most of the previous modelling practices only the separate effect of irradiance and temperature are focused. So according to them, with the rise of irradiance the open circuit voltage invariably rises. Temperature effect is not considered so much. From the analysis of this paper, finally it is obvious that, with the rise of irradiance at a very higher value when the temperature of the cell is also increased remarkably then open circuit voltage must fall than its lower irradiance value due to the temperature effect if the cell temperature is not controlled appropriately.

References

[1] S.W. Angrist, Direct Energy Conversion, Vol. 4, Allyn and Bacon Inc., Boston, 1982, pp. 177-227.

[2] O. Wasynczuk, Dynamic behavior of a class of photovoltaic power system, IEEE Transactions on Power Apparatus and Systems 102 (1983) 3031-3037.

[3] J.C.H. Phang, D.S.H. Chan, J.R. Philips, Accurate analytical method for the extraction of solar cell model parameters, Electronics Letters 20 (1984) 406-408.

[4] Antonio Luque, Steven Hegedus, Handbook of Photovoltaic Science and Engineering, 2nd edition, chapter 3(Jeffery L. Gray, The Physics of Solar Cell),1 March, 2011.

[5] S.M. Sze & Kwok K. Ng, hysics of Semiconductor Devices, 3rd edition, 2007.

[6] Safa. O. Kasap, Optoelectronics and Photonics; Principles and Practices, 1st edition Prentie Hall, 2001.

[7] C.Hu and R.M. White, Solar cells, McGraw-Hill Inc, New York, 1983.

[8] Ben Richard Hughes, Ng Ping Sze Cherisa, and Osman Beg, Computational Study of Improving the Efficiency of Photovoltaic Panels in the UAE, World Academy of Science, Engineering and Technology, Vol: 5, 2011-01-22.

[9] S. Sheik Mohammed, Modeling and Simulation of Photovoltaic Module Using MATLAB/Simulink, International Journal of Chemical and Environmental Engineering, Vol. 2, No. 5, October 2011.

[10] MSX 60 Datasheet, BP solar Global Marketing, 2002.

[11] Yuki Tsuno, Yoshihiro Hishikawa and Kosuke Kurokawa, Translation Equations for Temperature and Irradiance of the I-V Curves of Various PV Cells and Modules,Vol.2, IEEE, 2006.

Design and Analysis the Characteristics of A Cost-Effective Polymer Based Bulk Heterojunction Tandem Solar Cell

Md. Jubaer Hossain Pantho*, Noor A Junnat and Mohammad Jahangir Alam
Department of Electrical and Electronic Engineering (Bangladesh University of Engineering and Technology, Bangladesh)
Email: *hossain_jubaer@yahoo.com

***Abstract*—In this paper, a device model that describes the current-voltage characteristics of polymer:fullerene bulk heterojunction solar cell has been developed. Two organic PV cells of two different materials are cascaded together to form a tandem organic solar cell. In the design, PPV is implemented as the top cell and P3HT:PCBM blend is used as bottom cell. The simulation has been carried out using Silvaco ATLAS software. The power curve is also derived. Furthermore, the ideal efficiency and the Fill Factor of the tandem cell have been predicted from the curve. The stability of the model is ensured by introducing a thin buffer layer of ZnO coated on top of the solar cell.**

***Index Terms*—Fill Factor, exciton, tandem, heterojunction, solar cell**

I. Introduction

In today's world, the increasing rate of fossil-fuel consumption and its accompanying greenhouse effect have resulted in research efforts that focus on renewable energy. Among various other renewable energy sources photovoltaic cells (PV) comprises a large portion. Photovoltaic cells allow for device fabrication of large areas of lightweight and flexible substrates using low-cost solution-processing methods [1]. Thin film PV cells are promising approach and it has tremendous potential. After the introduction of silicon as the prime semiconductor material, silicon PV diodes became widely available [2]. Photoconductivity and electroluminescence were reported in organic crystals for the first time in the 1950s [3]. The first highly conductive polymer was chemically doped polyacetylene, reported in 1977 [4]. Since then, organic molecules as semiconducting polymers are being increasingly used in optoelectronic devices. Organic materials were introduced to photovoltaic devices because they can be chemically synthesized and low cost. They can be easily manufactured compared to silicon based cells, and this is due to the molecular nature of the materials used. Organic materials are easier to work with and can be used with thin film substrates that are 1,000 times thinner than silicon cells. They are becoming established as one of the future photovoltaic technologies for lowest-cost power production.

Solution-processed tandem bulk-heterojunction solar cells have gained serious attention during the last few years and are becoming established as one of the future photovoltaic technologies for lowest-cost power production. The organic photovoltaic(OPV) cells are normally comprised of p-doped and n-doped materials that are combined to form an active layer, as called bulk heterojunction (BHJ). In this type of photovoltaic cell, the electron donor and acceptor are mixed together, forming a polymer blend. One of the most challenging parts to further increase the efficiency potential of organic solar cells undoubtedly is the design of organic semiconductors that offer better photon harvesting across the entire solar spectrum, better charge-transport properties and better environmental stability. Here, an attempt is made to design an organic solar cell with high stability, high hole mobility, high crystallinity and able to absorb wavelength up to about 650 nm. P3HT:PCBM is used as the organic material to fulfill these criterion. A higher stability is achieved by using a thin coating layer of ZnO on top of the solar cell. Recently, Engineering researchers at the University of Arkansas have achieved the highest efficiency ever in a solar cell, after coating the solar cells with a thin layer of zinc oxide [5]. The purpose of this oxide layer is practiced here too.

The purpose of this contribution is to simulate a tandem organic solar cell structure using Silvaco ATLAS [6] software and give a state of art report on that plastic tandem solar cell based on conjugated polymers and analysis the operating characteristics of the tandem OPV consisting of two sub-cells. In the design of the tandem cell, PPV is implemented as the top cell and P3HT:PCBM blend is used as the bottom cell. A ZnO layer is implemented on top of the solar cell to provide better stability [7]. Afterwards simulated device data are used to extract physical parameters that are subsequently used to model tandem device operation under a variety of conditions.

II. Device Principle

Solar cell is the device that converts the light energy of the sun to electricity. The conversion steps and loss mechanism of light power into electric power are shown in Fig. 1.

In solar cells, the photovoltaic process of converting energy is composed of four steps [8].

$$\eta_{eff} = \eta_{abs} \times \eta_{diss} \times \eta_{trans} \times \eta_{col} \quad (1)$$

Where,

η_{abs} = light absorption efficiency
η_{diss} = exciton dissociation efficiency
η_{trans} = charge transport efficiency
η_{col} = charge collection efficiency

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

An organic semiconductor is an organic material with semiconductor properties [9]. Single molecules, short chain (oligomers) and organic polymers can be semiconductive. These materials are based on the unconventional properties of the carbon atom. The principle of operation of Organic photovoltaic cells utilizes conjugated π and π^* electron orbitals for light absorption, emission, charge generation and transportation. Light absorption leads to generation of excitons. When sunlight is directed onto photosensitive semiconducting organic materials, electrons in HOMO are excited to LUMO. After light absorption, inorganic semiconductors produce free carriers at once, while organic semiconductors require additional processes to produce free carriers. First, organic semiconductors form excitons, which are strongly bound electron hole pairs. These pairs then diffuse inside of the organic semiconductor, which is independent on an applied electric field. The typical exciton diffusion length in an organic semiconductor is around 10-20nm, and exciton dissociation occurs only at the interface between the donor and the acceptor. This is due to the offset of HOMO states of the donor and the acceptor. If excitons do not reach the interface, they recombine and the absorbed energy is dissipated without generating photocurrent. Therefore, to generate power efficiently, the excitons have to be dissociated and collected at electrodes before recombination.

If a solar cell is short circuited with illuminating by sunlight there will have a short circuit current. This current is known as photo current I_{ph}. The greater is the light intensity, the higher is the photo generation rate and larger is I_{ph}.

$$I_{sc} = -I_{ph} = -KI \tag{2}$$

Here K is a constant that depends on the particular device. When a resistance is connected across the solar cell the total current flow through the resistance is given by,

$$I = I_d - I_{ph} \tag{3}$$

I_d is due to the normal pn junction behavior. It is given by the diode characteristics,

$$I_d = I_o[\exp\left(\frac{eV}{nkT}\right) - 1] \tag{4}$$

Here I_o is the reverse saturation current and n is the identity factor that depends on the semicondutor material. In an open circuit the net current is zero. This means $I_d = I_{ph}$. So the total current through the resistor is,

$$I = -I_{ph} + I_o[\exp\left(\frac{eV}{nkT}\right) - 1] \tag{5}$$

III. SIMULATED TANDEM OPV CELL STRUCTURE

Realization of a photovoltaic product requires excellent materials for outstanding good cells. The tandem organic solar cell simulated here consists of two organic PV subcells of two different materials cascaded together to form a solar cell. In the top cell PPV polymer is used and in the bottom Cell P3HT:PCBM polymer layer is used. Low Work Function Polymer Electrode for anode and Aluminum Cathode are used as contacts [10]. The simulation is done using Silvaco ATLAS. ATLAS is a device simulation framework which enables device technology engineers to simulate the electrical, optical, and thermal behavior of semiconductor devices.

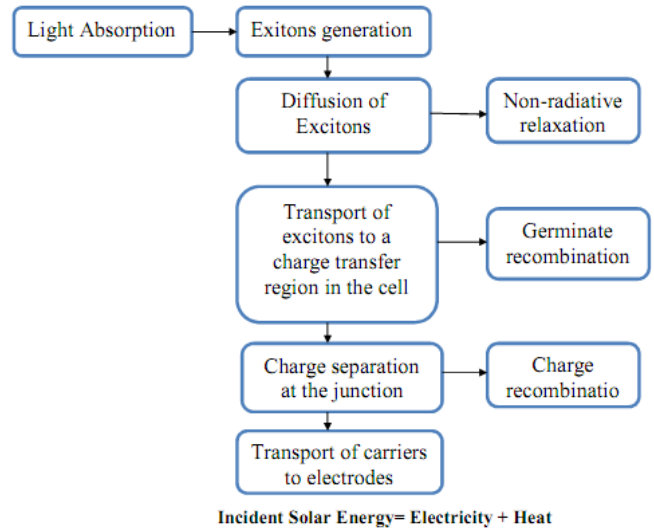


Fig. 1. Conversion steps and loss mechanism of light power into electric power

PPV (polyphenylene vinylene) is a conducting polymer of the rigid-rod polymer host family [11]. They can be easily synthesized in good purity and high molecular weight. P3HT: PCBM or poly(3-hexylthiophene):1-(3 methoxycarbonyl) propyl-1-phenyl[6,6]-C61 is the most efficient OPV material so far [12]. This is because the characteristics of P3HT:PCBM favors high stability, high hole mobility, high crystallinity and able to absorb wavelength up to about 650 nm. In addition, C60 in PCBM is strong electron acceptor, enables fast charge separation and is the least expensive and most energy saving fullerene. The molecular structures of PCBM and P3HT are shown in Fig. 2. Organic polymers have wider gap than semiconductors. Thus they give an efficient absorption at near UV part. So is P3HT:PCBM blends. The gap of the blend is approximately 1.8eV. So the longest absorption wavelength should be around 650 nm. The Absorptance spectrum of P3HT film, PCBM film, and P3HT:PCBM film is shown in Fig. 3.

One of the greatest obstacles that must be overcome before organic photovoltaic cells become commercially viable is the inability to produce stable solar cells with long lifetimes. Semiconducting polymers are generally unstable in the presence of light and oxygen. The diffusion of oxygen and water into the cell is recognized as the greatest source of device degradation when OPVs are tested in ambient air. To overcome this problem a thin buffer layer of ZnO is used on top of the OPV cell which scavenges for oxygen and humidity, improving the stability of the organic Layer and increasing

Fig. 2. Structures of P3HT and PCBM

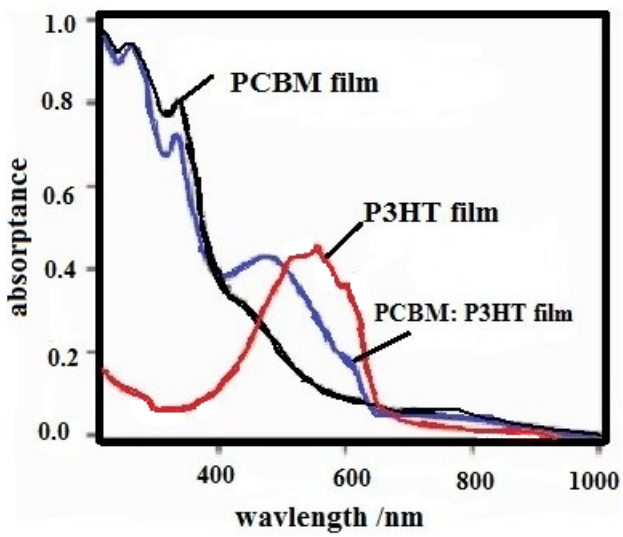


Fig. 3. Absorptance spectrum of P3HT film, PCBM film, and P3HT:PCBM film

the efficiency as well.

For simulation purposes, AM1.5 solar spectrum condition has been used. Which means input optical power is approximately 1000KW/m2 or 100mW/cm2 [13].

The simulated structure of Tandem Organic Solar Cell is shown in Fig. 4.

IV. Simulation and Analysis

J-V Characteristics and output power curve of the tandem cell are predicted using the Silvaco ATLAS software. They are shown in Fig. 5 and Fig. 6.

Using these plots simulation outputs are analyzed and following output parameters are obtained:

1. J-V Characteristics
2. Power Curve
3. Short-circuit Current Density (J_{sc})
4. Open Circuit Voltage (V_{oc})
5. Fill Factor (FF)
6. Efciency (η)

Calculation of J_{sc}, V_{oc} are done by the following equations:

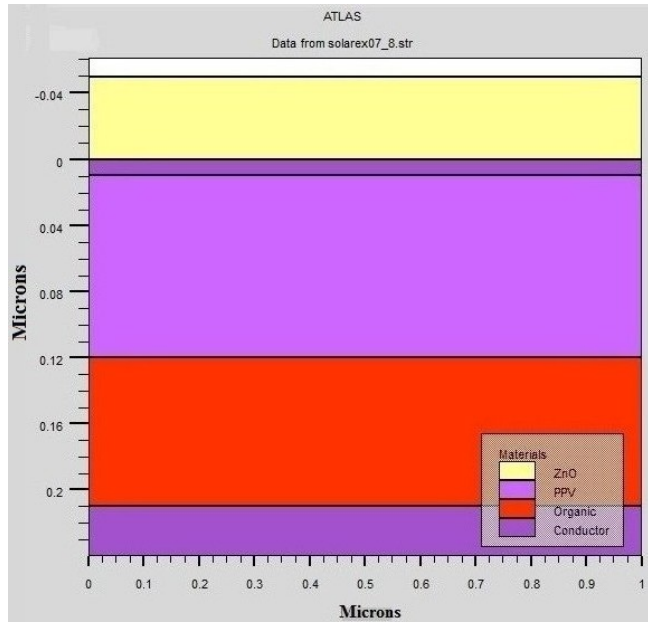


Fig. 4. Structure of the simulated tandem organic solar cell

J_{sc}=value of current when V=0 in J-V curve
V_{oc}=value of voltage when J=0 in J-V curve
P_m = peak value of power in power curve

$$P_m = V_m J_m \tag{6}$$

The Fill Factor (FF) and efficiency (η) of the solar cell is calculated from:

$$FF = \frac{P_m}{V_{oc} \times J_{sc}} \tag{7}$$

$$\eta = \frac{FF \times V_{oc} \times J_{sc}}{P_{in}} \tag{8}$$

The Simulated results are given in tabulated form in table I

TABLE I
Extracted Data from Simulation of Tandem Organic Solar Cell

Parameters	Results Under Illumination
J_{sc}(mA/cm2)	36.70
V_{oc}(V)	0.68
P_m (mW/cm2)	10.73
V_m(V)	0.43
J_m (mA/cm2)	23.94
Fill Factor (FF)	0.43
Efficiency η	10.73%

It is found that an Ideal efficiency of 10.73% can be obtained from this solar cell with a Fill Factor of 0.43. The open circuit voltage of the proposed cell is 0.68V and the short circuit current is 36.70 mA/cm2, which is quite good.

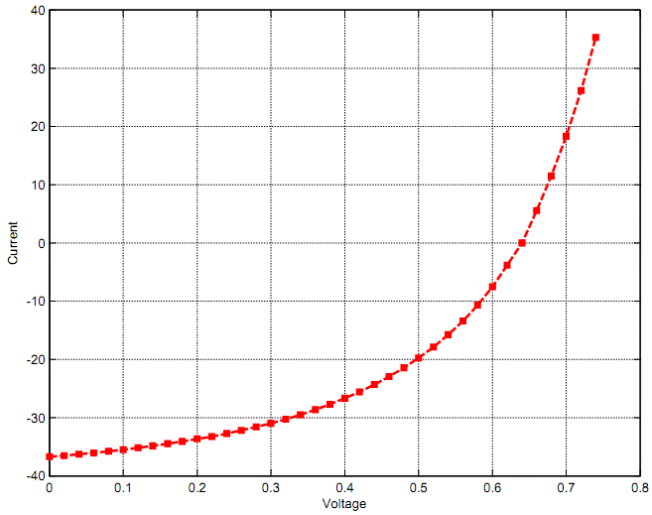


Fig. 5. J-V Characteristics of Tandem Organic Solar Cell

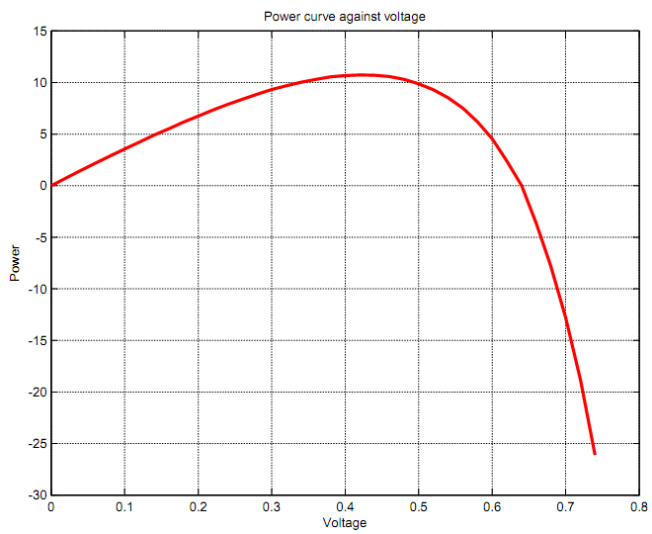


Fig. 6. Output power curve of Tandem Organic Solar Cell

V. Conclusion

In this paper, an efficient tandem organic solar cell based on polymer bulk heterojunction has been designed and its J-V characteristics are studied. The top cell comprises of PPV and the in bottom cell a P3HT:PCBM fullerene polymer layer is used. A thin layer of ZnO is coated on top of the solar cell to ensure the stability of the solar cell. Furthermore, this buffer layer increases the efficiency of the solar cell as well. An overall ideal efficiency of 10.73% is attained which is quite compatible. The efficiency of this photovoltaic cell can be further increased by introducing advance light trapping management [14].

References

[1] L.J.A. Korster, E.C.P. Smits, V.D. Mihailetchi, and P.W.M. Blom, Device model for the operation of polymer/fullerene bulk heterojunction solar cell, *Physical Review B*, Vol. 72, (2005) pp. 085205-1, 085205-9

[2] A. Shah, "Photovoltaic Technology-The Case for Thin-Film Solar Cells", in *Science Volume 285* issue 5428, 1999

[3] M. Pope, H. P. Kallmann, P. Magnante," J. Chem. Phys.", 38, 2042, 1963.

[4] C. K. Chiang, C. R. Fincher, Jr., Y. W. Park, A. J. Heeger, H. Shirakawa, E. J. Louis, S. C. Gau, A. G. MacDiarmid, "Phys. Rev. Lett", 39, 1098, 1977.

[5] Y.F. Makableh, R. Vasan, J.C. Sarker, A.I. Nusir, S. Seal, M.O. Manasreh. "Enhancement of GaAs solar cell performance by using a ZnO solgel anti-reflection coating." *Solar Energy Materials and Solar Cells*, 2014.

[6] "SILVACO, ATLAS Users Manual", Device Simulation Software, chap-1, p-1-2

[7] S.M. AbdulAmohsin, J.B. Cui, M.Z. Mohammed, "Study on ZnO/P3HT:PCBM nanowire solar cells," *Photovoltaic Specialists Conference (PVSC)*, 2013 IEEE 39th , vol., no., pp.2685,2689, 16-21 June 2013

[8] Myung-Su Kim, "Understanding Organic Photovoltaic Cells: Electrode, Nanostructure, Reliability, and Performance", Art 2.3.1.

[9] What_are_organic_semiconductors: http://www.iapp.de/orgworld/Basics

[10] Kipyo Hong, Se Hyun Kim, Chanwoo Yang, Tae Kyu An, Hyojung Cha, Chanjun Park, Chan Eon Park,"Photopatternable, highly conductive and low work function polymer electrodes for high-performance n-type bottom contact organic transistors",*ScienceDirect*,31 December 2010.

[11] Chen, W.B.; Deng, M. Z.; Zou, H. J.; Li, C. Y.; Deng, L. F.; Lai, P.T., "Improved efficiency of organic photovoltaic MEH-PPV/PCBM device with CuPc-doping," *Electron Devices and Solid-State Circuits (EDSSC), 2010 IEEE International Conference of* , vol., no., pp.1,4, 15-17 Dec. 2010

[12] "An overview on P3HT:PCBM, the most efficient organic solar cell material so far". Ge, Weihao Email: wge1@utk.edu *Solid State Physics II*, Spring 2009

[13] S.O.Kasap,"Optoelectronics and Photonics: Principles and Practices" Chap-6 p-255

[14] F. J. Beck, A. Polman, K. R. Catchpole. "Tunable light trapping for solar cells using localized surface plasmons", *J. Appl. Phys*. 105, 114310, 2009.

Modeling and Analysis of Inorganic, Organic and Hybrid Solar Cells

Nasrin Akter*, Partha Sarathi Paul, Sudipta Mondal and Sharif Mohammad Mominuzzaman
Department of Electrical and Electronic Engineering,
Bangladesh University of Engineering and Technology,
Dhaka-1000, Bangladesh.
*nasrin.eee.08@gmail.com

***Abstract* —The main objective of this paper is the modeling and analysis of inorganic homo-junction and inorganic-organic hybrid solar cells. For this, an independent model is developed that is capable of simulating the I-V characteristics of any homo-junction solar cell. The efficiency from the model is 14.897% at standard conditions and from PC1D, a commercial software the efficiency is 14.37% under the same conditions. Also hybrid series and hybrid parallel solar cells are modeled. The efficiency of the single inorganic cell is increased from 15% to 19.5% by the hybrid parallel connection.**

***Index Terms*—Inorganic cell, Hybrid cell, Efficiency**.

I. INTRODUCTION

The research in renewable energy has become an increasingly important topic in the 21st century with the problem of energy crisis becoming more and more aggravated, resulting in increased exploitation and search for new energy resources such as wind, water, geothermal and solar energy around the world. Photovoltaic solar energy, which is inexhaustible, environment friendly and widely available, has been increasingly used to generate electric power from sunshine. A solar cell (also called a photovoltaic cell) is an electrical device that converts the energy of light directly into electricity by the photovoltaic effect.

In this paper, an independent model is developed that is capable of simulating the I-V characteristics of any homo-junction inorganic solar cell. This paper also includes the modeling and study of inorganic-organic hybrid solar cells.

II. SOLAR CELL MODEL

A general mathematical description of I-V output characteristics for a PV cell is studied from [1]-[2]. The equivalent circuit of solar cell comprises of a current source connected in anti-parallel with a diode, series resistance of the cell R_s and the shunt resistance of the cell R_{sh} as shown in Fig. 1. The voltage-current characteristic equation of a solar cell is given as

$$I = I_{ph} - I_s[e^{\frac{q(V+IR_s)}{nKT_c}} - 1] - \frac{V + IR_s}{R_{sh}} \quad (1)$$

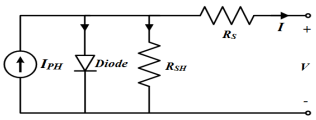


Fig. 1 Equivalent circuit of a solar cell.

where I_{ph} is a light-generated current or photocurrent, I_s is the cell saturation of dark current, q (= 1.6×10^{-19} C) is an electron charge, K (= 1.38×10^{-23} J/K) is Boltzmann's constant, T_c is the cell's working temperature, n is the ideality factor, R_{sh} is the shunt resistance, and R_s is the series resistance. The photocurrent mainly depends on the solar insolation and cell's working temperature, which is described as

$$I_{ph} = [I_{sc} + K_I(T_C - T_{Ref})]\lambda \quad (2)$$

where I_{sc} is the cell's short-circuit current at 25°C and 1kW/m^2, K_I is the cell's short-circuit current temperature coefficient, T_{Ref} is the cell's reference temperature, and λ is the solar insolation in kW/m^2. The cell's saturation current varies with the cell temperature, which is described as

$$I_s = I_{RS}\left(\frac{T_C}{T_{Ref}}\right)^3 e^{\frac{qE_g}{nK}\left(\frac{1}{T_{Ref}} - \frac{1}{T_C}\right)} \quad (3)$$

where I_{RS} is the cell's reverse saturation current at a reference temperature and a solar radiation, E_g is the bang-gap energy of the semiconductor used in the cell and varies with temperature as (e.g., $E_g(T) \approx E_g(0) - \frac{\alpha T^2}{T+\beta}$). The reverse saturation current at reference temperature can be approximately obtained as

$$I_{RS} = \frac{I_{sc}}{e^{\frac{qV_{oc}}{nKT_{Ref}}} - 1} \quad (4)$$

The short-circuit current at standard conditions, I_{sc}, is analysed by dividing the solar spectrum into many segments, each having a narrow range of wavelength and finding the current due to each spectral segment. Since each spectral segment is essentially monochromatic, it may be assigned a single wavelength λ and absorption coefficient $\alpha(\lambda)$. The generation rate of electron-hole pairs at a distance x from the semiconductor surface as shown in Fig. 2 is given by

$$G(\lambda, x) = \alpha(\lambda)F(\lambda)[1 - R(\lambda)]e^{-\alpha(\lambda)x} \quad (5)$$

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

where $\alpha(\lambda)$ is the absorption coefficient, $F(\lambda)$ the number of incident photons/cm^2/s per unit bandwidth in this spectral segment, and $R(\lambda)$ is the fraction of these photons reflected from the surface. α depends on k_e, the extinction coefficient (e.g., $\alpha = \frac{4\pi k_e}{\lambda}$) and R depends on n_r, the refractive index and k_e (e.g., $R = \frac{(1-n_r)^2 + k_e^2}{(1+n_r)^2 + k_e^2}$) where

$$\overline{n} = n_r - ik_e \tag{6}$$

is the complex index of refraction.

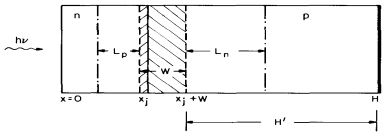


Fig. 2 Solar-cell dimensions and minority-carrier diffusion lengths.

The total short-circuit photocurrent density at a given wavelength is the sum of electron, hole and depletion-region short-circuit currents as taken from [3]-[4]. The total cell short-circuit current density is the sum of the contributions from all the spectral segments

$$J_{sc} = \sum_i qF(\lambda_i)[1 - R(\lambda_i)]SR(\lambda_i) \tag{7}$$

where

$$SR(\lambda) = \frac{J_n(\lambda) + J_p(\lambda) + J_{dr}(\lambda)}{qF(1-R)} \tag{8}$$

The open-circuit voltage, V_{oc} at standard conditions is determined using the following equation

$$V_{oc} = \frac{nKT}{q}\ln(\frac{I_{sc}}{I_s} + 1) \tag{9}$$

III. HYBRID SOLAR CELL

A hybrid solar cell is basically a combination of an inorganic and an organic solar cell. These two cells can be connected in either series or parallel connection. The I-V relationship of the organic cell is studied from [5].

$$J = J_{ph} - J_{01}[e^{B_1(V+JR_s)} - 1] - J_{02}[e^{B_2(V+JR_s)} - 1] - \frac{V + JR_s}{R_{sh}} \tag{10}$$

J_{01} represents the electronic conduction phenomena in the quasi-neutral region of the junction, such as diffusion, recombination and the drift effects, J_{02} corresponds to the carrier recombination via deep levels in the space charge region of the junction.

1) *Series Connection:* The I-V characteristics of a hybrid series cell is given by

$$V = \frac{n_1KT}{q}\ln(\frac{I_{ph1} - I}{I_{s1}} + 1) + \frac{n_2KT}{q}\ln(\frac{I_{ph2} - I}{I_{s2}} + 1) - IR_{s1} - IR_{s2} \tag{11}$$

where I_{ph1} and I_{ph2} are the photocurrents, R_{s1} and R_{s2} are the series resistances and I_{s1} and I_{s2} are the diode reverse saturation currents of the inorganic and organic cells respectively as shown in Fig. 3.

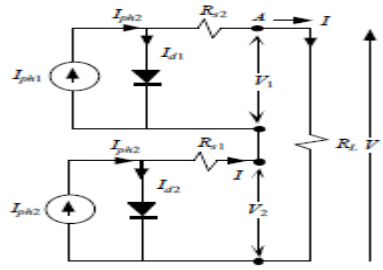


Fig. 3 Two mismatched solar cells in series under a load R_L

2) *Parallel Connection:* When an inorganic cell and an organic cell with photocurrents I_{ph1} and I_{ph2}, series resistances R_{s1} and R_{s2} and parallel resistances R_{sh1} and R_{sh2} respectively are connected in parallel driving a load R_L, the I-V characteristics of the hybrid cell is given by

$$I = (I_{ph1} + I_{ph2}) - I_{s1}[e^{\frac{q(V+IR_{se})}{n_1KT}} - 1] - I_{s2}[e^{\frac{q(V+IR_{se})}{KT}} - 1] - I_{s3}[e^{\frac{q(V+IR_{se})}{2KT}} - 1] - \frac{V + IR_{se}}{R_{sh1}} - \frac{V + IR_{se}}{R_{sh2}} \tag{12}$$

where

$$R_{se} = \frac{R_{s1}R_{s2}}{R_{s1} + R_{s2}} \tag{13}$$

IV. MODELING AND ANALYSIS

1) *Inorganic Cell Modeling:* A typical Si solar cell is first simulated using PC1D. Next, the same solar cell structure is modeled using MATLAB coding. For this, First I_{sc} and V_{oc} are determined at standard conditions and compared with I_{sc} and V_{oc} from PC1D simulation. The resulting I_{sc} and V_{oc} values at standard conditions are then used for modeling the solar cell at different irradiance and temperature conditions. The parameters used in the simulation are summarized in Table I.

Table I
Si solar cell parameters

Parameter	Value
Cell area, A	100 cm^2
Width of entire cell, H	100 µm
Width of n-layer, T_n	0.5 µm
Doping of n-layer, N_D	1×10^{18} cm^{-3}
Doping of p-layer, N_A	1×10^{16} cm^{-3}
Surface recombination velocity, S_p, S_n	1×10^{5} cm/s
Recombination lifetime, τn, τp	1000 µs
Irradiance, G	1000 W/m^2

Parameter	Value
Temperature, T	$25°C$
Effective density of states in conduction band, N_c	$2.8\times10^{19}\ cm^{-3}$
Effective density of states in valence band, N_v	$1.04\times10^{19}\ cm^{-3}$
α	4.9×10^{-4} eV/K
β	655 K
$E_g(0)$	1.169 eV
Ideality factor, n	1.2
Electron mobility, μn	$1110\ cm^2V^{-1}s^{-1}$
Hole mobility, μp	$196\ cm^2V^{-1}s^{-1}$

The complex index of refraction at each wavelength from 0.2066 μm to 0.835 μm with a wavelength separation of 0.025 μm is taken from [6] and from this α and R at each wavelength are determined. Table II shows the I_{sc} and V_{oc} values obtained from simulation for Si cell and Table III shows the resulting parameters of the solar cell using these values.

Table II

I_{sc} and V_{oc} values obtained from simulation for Si

From PC1D simulation		From MATLAB code	
Isc	3.122 A	Isc	3.1239 A
Voc	0.5789 V	Voc	0.6094 V

Table III

Si solar cell simulation results at 1000 W/ m^2 and $25°C$

From PC1D simulation		From modeling	
Isc	3.122 A	Isc	3.124 A
Voc	0.5789 V	Voc	0.6089 V
Pm	1.437 W	Pm	1.4897 W
FF	0.7951	FF	0.7831
Efficiency (%)	14.37	Efficiency (%)	14.897

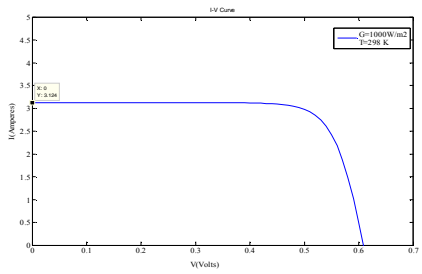


Fig. 4 I-V curve of a Si solar cell at 1000 W/ m^2 and $25°C$

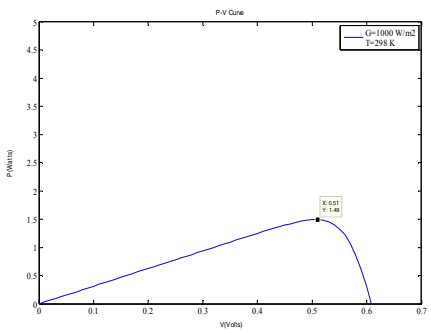


Fig. 5 P-V curve of a Si solar cell at 1000 W/ m^2 and $25°C$

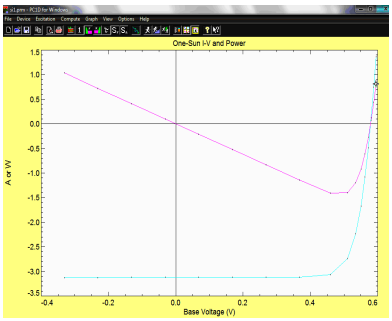


Fig. 6 I-V and P-V curves of a Si solar cell at 1000W/ m^2 from PC1D

2) *Hybrid cell modeling:*

Table IV
Hybrid solar cell main parameters

Area, A= 1 cm^2			
G=1000 W/ m^2, T= $25°C$			
Inorganic cell		Organic cell	
Photo-current, I_{ph1}	31.2 mA	Photo-current, I_{ph2}	8 mA
R_{s1}	0.05 Ω	R_{s2}	97.7 Ω
V_{oc1}	0.6094 V	Diode reverse saturation current, I_{s2}	2.3×10^{-14} A
α	5.4×10^{-4} eV/K		
β	204 K	Diode reverse saturation current due to recombination, I_{s3}	5×10^{-10} A
$E_g(0)$	1.519 eV		

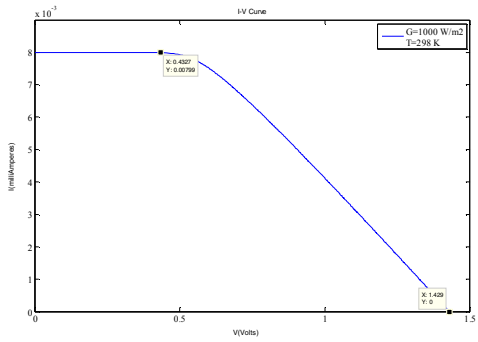


Fig. 7 I-V characteristics of a hybrid series solar cell

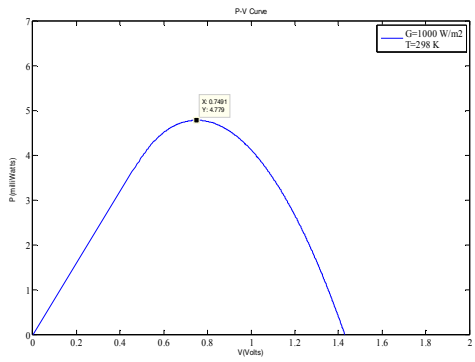


Fig. 8 P-V characteristics of a hybrid series solar cell

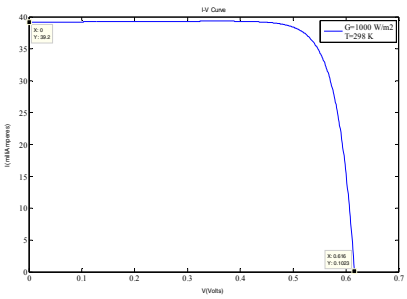


Fig. 9 I-V characteristics of a hybrid parallel solar cell

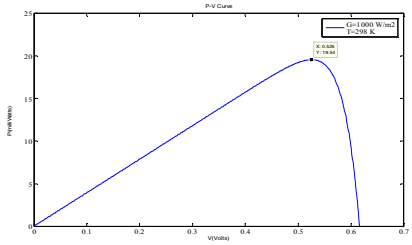


Fig. 10 P-V characteristics of a hybrid parallel solar cell

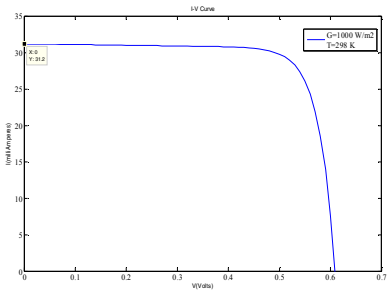


Fig. 11 I-V curve of a Si solar cell at 1000 W/ m^2 and $25°C$

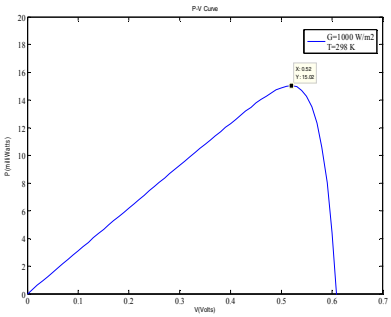


Fig. 12 P-V curve of a Si solar cell at 1000 W/ m^2 and $25°C$

Table V
Overall performances of Si and hybrid solar cells

Irradiance, G=1000 W/ m^2 , Temperature, T= $25°C$					
Area, A=1 cm^2					
Single Si		Hybrid Series		Hybrid Parallel	
Isc	31.2 mA	Isc	8 mA	Isc	39.2 mA
Voc	0.6086 V	Voc	1.429 V	Voc	0.616 V
Pm	15.0159 mW	Pm	4.7792 mW	Pm	19.5354 mW
FF	0.7908	FF	0.4181	FF	0.809
Efficiency (%)	15.0159	Efficiency (%)	4.7792	Efficiency (%)	19.5354

Comparing the characteristics of the hybrid series cell with the single Si solar cell it is found that the performance of the hybrid cell is extremely poor. The main reason for this is the limiting of the cell short-circuit current by the short-circuit current of the organic cell. The short-circuit current of the organic cell is low because of its very high series resistance. Although the open-circuit voltage of the hybrid cell is improved much on account of the series connection, the lower short-circuit current reduces the overall efficiency of the cell to a much lower value than that of the Si solar cell. Conversely, the performance of the hybrid parallel cell is better than that of the single Si cell. This is also mainly because of the short-circuit current but in an opposite sense. The short-circuit current of the organic cell increases that of the Si cell due to the parallel connection. The open-circuit voltage of the hybrid cell is also high. As a result, the overall cell efficiency is improved.

A probable way of improving the performance of the series cell is paralleling multiple organic cells and connecting that combination in parallel with the Si cell. This would improve the short-circuit current of the total cell and hence enhance the efficiency.

V. CONCLUSIONS

In this paper the I-V and P-V characteristics of inorganic and organic solar cells are modeled. From this modeling the short-circuit current, open-circuit voltage, maximum power, fill factor and efficiency of homo-junction Si is obtained. The results are compared with that of a commercial software and it is found that the model gives nearly the same results as a software. This model can be applied to any homo-junction solar cell given the parameters like doping concentrations, layer widths, area, complex indices of refraction, minority carrier lifetimes etc.

Two hybrid solar cells are also modeled. It is observed that the parallel connection increases the efficiency of the single Si solar cell by an appreciable amount but the series combination degrades the efficiency.

REFERENCES

[1] S. Sheik Mohammed, "*Modeling and Simulation of Photovoltaic module using MATLAB/Simulink*," International Journal of Chemical and Environmental Engineering, Volume 2, No.5, October 2011.

[2] Huan-Liang Tsai, Ci-Siang Tu, and Yi-Jie Su, Member, IAENG, *"Development of Generalized Photovoltaic Model Using MATLAB/SIMULINK,"* Proceedings of the World Congress on Engineering and Computer Science 2008, WCECS 2008, October 22 - 24, 2008, San Francisco, USA.

[3] S.M.Sze and Kwok K.Ng, *"Physics of Semiconductor Devices,"* 3rd ed., John Wiley & Sons, Inc., Hoboken, New Jersey.

[4] Chenming Hu, Richard M.White, *"SOLAR CELLS From Basics to Advanced Systems,"* McGraw-Hill Book Company.

[5] Ali Cheknane, Hikmat S. Hilal, Faycal Djeffal, Boumedie`ne Benyoucef, Jean-PierreCharles, *"An equivalent circuit approach to organic solar cell modeling,"* Microelectronics Journal 39 (2008) 1173–1180; doi:10.1016/j.mejo.2008.01.053.

[6] http://refractiveindex.info/legacy/?group=CRYSTALS&material=Si&option=Aspnes&wavelength=2.480

Thickness Optimization and Composition Grading Effect in Heterojunction CIGS Solar Cell

Ayed Al Sayem, Yeasir Arafat*, and Md. Mushfiqur Rahman
Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology (BUET)
Dhaka, Bangladesh
*arafat@eee.buet.ac.bd

Abstract—**Thin Film Solar Cell is a technology which has a great potential as it has low material consumption and relatively high efficiency to others. Among thin film solar cells CIGS is a promising one. But due to scarcity of gallium and indium its industrial production can be affected. Here in this paper via simulation (using device simulator) we have tried to achieve good acceptable efficiency using very thin layer of CIGS absorber layer with the help of grading and appropriate gallium composition and by optimizing emitter efficiency which is close to the efficiency of CIGS cells having thick absorber layer.**

Keywords—Thin Film; CIGS Solar Cell; Composition Grading.

I. Introduction

Solar cell is usually made of semiconducting materials (including compound materials) which converts radiant optical energy into useful electrical energy. At present most of the solar devices or cells manufactured from wafers of mono crystalline or polycrystalline silicon or amorphous silicon which are the dominant technology in the commercial market. Solar cells based on the thin film of direct band gap semiconducting materials are other technologies with great potential in harvesting green energy. One such thin film technology is based on a compound of Copper, Indium, Gallium and Selenium abbreviated as $Cu(In,Ga)Se_2$ or CIGS [1]. As CIGS is a direct band gap material the absorber thickness for CIGS solar cell is much less than Silicon solar cells usually in between 1.5 to 2.5 micro meter. But materials used in CIGS solar cell are not abundant in nature. The critical material for CIGS solar cell is Indium and gallium. The production capability of gallium is 167 metric tons per year where the production capability of indium is at 600 metric tons per year [2]. With the increasing demand of electricity if CIGS solar cells have to keep impact on solar power supply then material used for CIGS solar cell device has to be reduced by not degrading its efficiency. In this work by device simulation we have tried to find a way to lessen the absorber thickness by not affecting the efficiency.

II. Grading Effect

One of the very common ways used to increase the efficiency of a CIGS solar cell is composition grading in the CIGS absorber [3]-[5] Grading can be of various types:

- Front Grading
- Back Grading
- Double Grading
- Full reverse grading

Grading in CIGS is of course a complex manufacturing process. Here in this work we have tried only the full reverse grading.

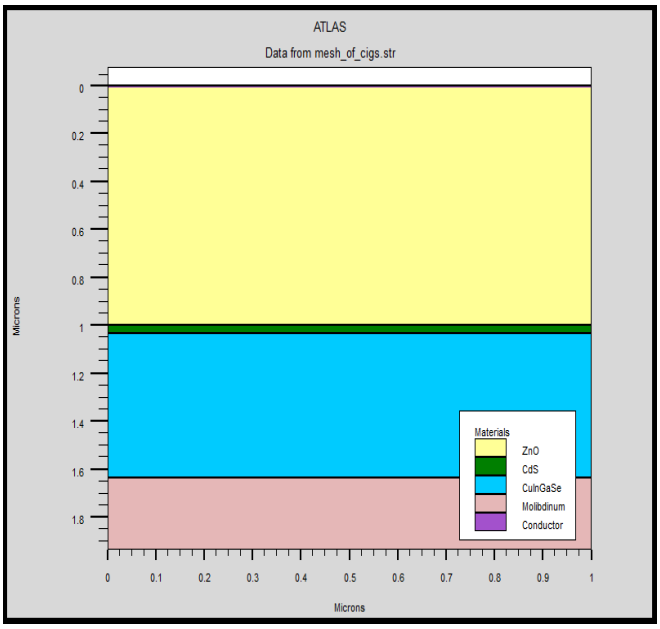


Fig. 1 Structure of CIGS solar cell.

The optimum thickness for a CIGS absorber is usually 1.5 to 2 micro meters. Here we first simulated the cell under AM1.5 for different thickness from 0.4 micron to 2 micron. The gallium composition was fixed for 0.3 throughout the absorber. We did not use higher gallium composition because it is already shown in literature that the optimized gallium composition for CIGS absorber is in the 0.3 to 0.4 range. We considered the possible defects in CdS and CIGS. We vary the thickness of CIGS absorber from 0.4 micro meters to 1.5 micro meters. For both uniform and reverse graded gallium composition the important findings open circuit voltage, short circuit current, fill factor and finally efficiency are shown below:

III. Simulation And Results

The simulation was done under AM1.5 spectrum.

A. Thickness Variation:

1) Open Circuit Voltage: Open circuit voltage change for different thickness level is shown in the following figure.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

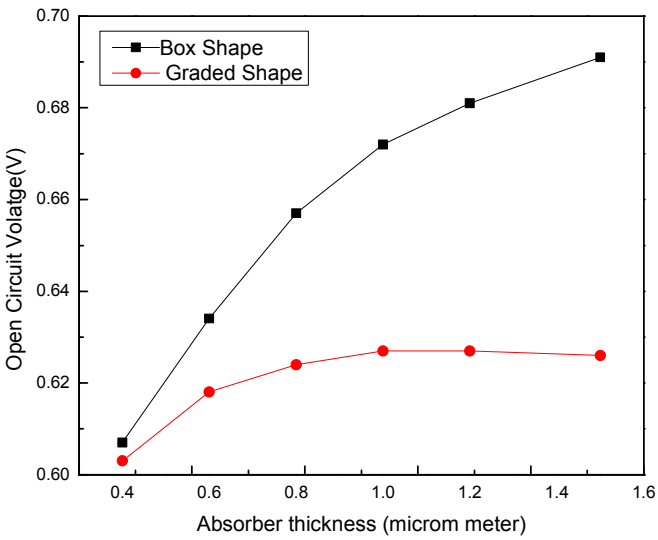


Fig. 2 Open circuit voltage vs. thickness variation (both for uniform and graded composition).

Open circuit voltage was lower in graded than uniform gallium composition for thickness level from .4 micro meters to 1.5 micro meters.

2) Short Circuit Current: Short circuit current density variation for different thickness level is shown in the following figure.

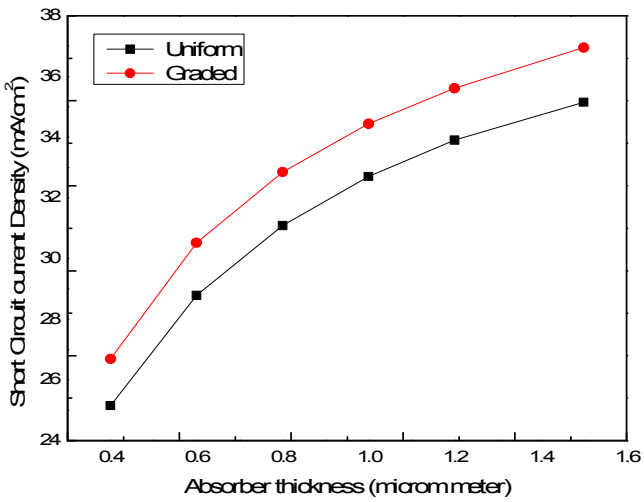


Fig. 3 Short circuit current density vs. thickness variation (both for uniform and graded composition).

Short circuit current density increased for both uniform and graded gallium composition as the thickness of the absorbing layer is increased which is very much expected as the photo generation gets better for thick absorbing layers. But for graded composition short circuit current density was higher than uniform composition for all level of thickness.

3) Fill Factor: Fill factor is defined as-

$$Fill\ Factor = \frac{V_m I_m}{V_{OC} I_{SC}}$$

where,

V_{oc} = Open circuit voltage

I_{sc} = Short circuit current

V_m = maximum obtainable voltage

I_m = maximum obtainable current

Fill factor variation for different thickness level is shown in the following figure.

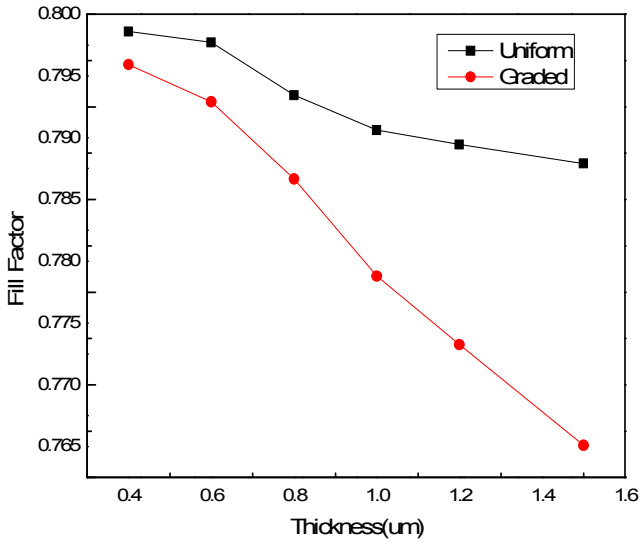


Fig. 4 Fill factor vs. thickness variation (both for uniform and graded composition).

In case of fill factor we have found that at low thickness fill factor obtains much better value and with increase of thickness it decreases for both uniform and graded gallium composition. In case of graded composition we see that fill factor is lower than uniform composition for all level of thickness and for higher thickness (especially after thickness of .8 micro meter) difference of fill factor between graded and uniform gallium composition is greater which can be good reason for low efficiency at high thickness level for graded composition than uniform composition.

4) Efficiency: Efficiency of the solar cell did not increase for graded gallium composition for higher thickness level. Efficiency of the solar cell increased with a nonlinear shape for uniform gallium composition for thickness level of .4 micro meters to 1.5 micro meters. The increase of efficiency can be explained simply that with more absorber the current generation gets enriched and also the diode characteristic gets better and so the open circuit voltage. But efficiency is higher for graded than uniform gallium shape for lower absorber thickness level. For thickness level up to 0.8 micro meter graded composition of gallium gives higher efficiency than uniform. But above 0.8 micro meters efficiency is lower than uniform composition of gallium for graded.

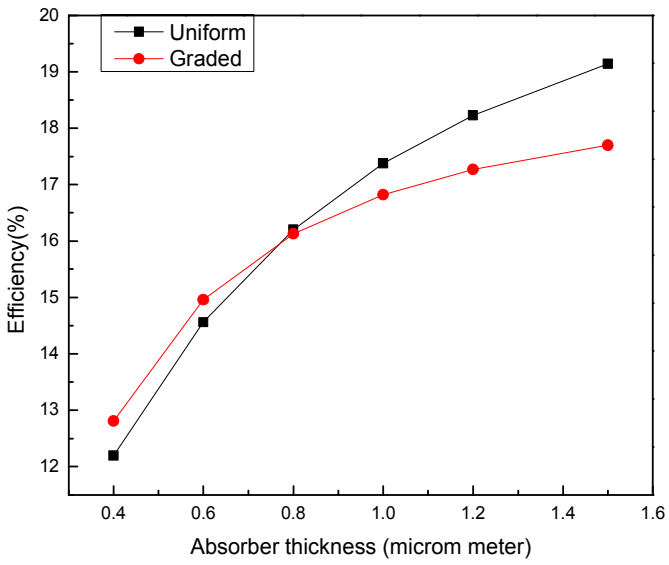


Fig. 5 Efficiency vs. thickness variation (both for uniform and graded composition).

B. *Grading Effect Optimization*

Our goal is to obtain better efficiency having an absorbing layer of low thickness so we kept the absorber thickness at 0.6 micro meter and as we found grading is useful in obtaining better efficiency when the absorber thickness is low, we increase the grading by increasing the peak value of gallium composition at the back contact. The grading combinations used here were 0 to 0.4 (Type A), 0 to 0.5 (Type B), 0 to 0.6 (Type C) and 0.1 to 0.6 (Type D).

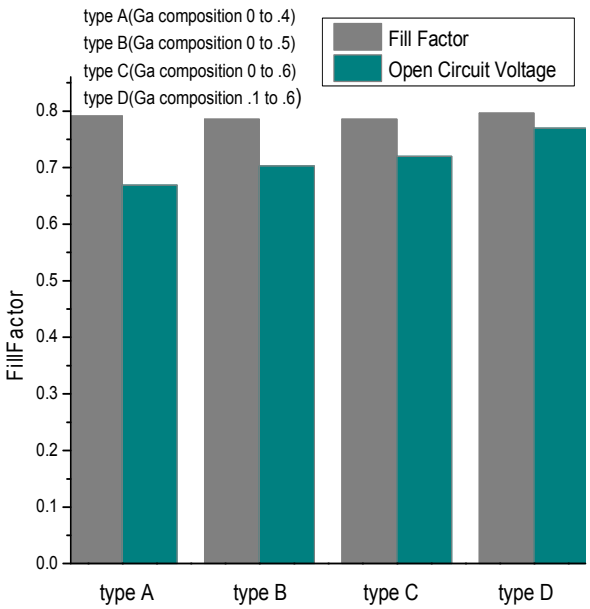


Fig. 6 Fill factor and open circuit voltage for different grading levels.

The Open circuit voltage of the cell was mostly affected by higher grading and it increased with the increase of grading. In case of type D where the Gallium composition was increased from 0.1 to 0.6 the open circuit voltage was further increased whereas the fill factor remained almost constant for all the types.

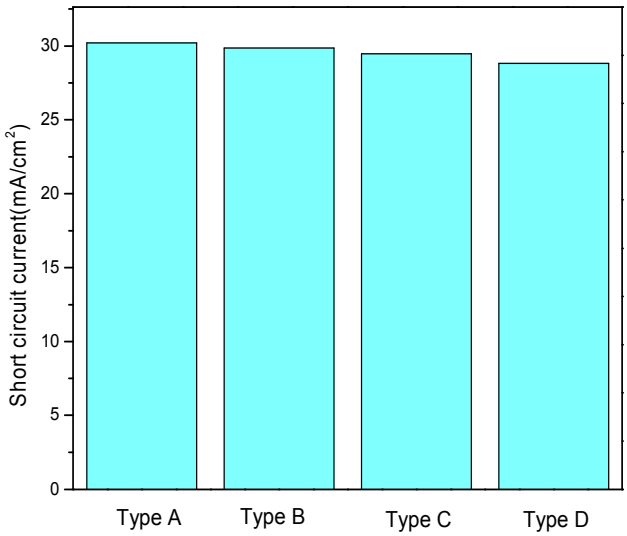


Fig.7 Short circuit current for different grading levels.

The short circuit current of the solar was adversely affected d with the increase of grading though it is previously shown that for graded composition short circuit current density is higher than uniform composition. But increasing the grading didn't help increasing the short circuit efficiency. But the key factor is that which is dominant factor affecting efficiency open circuit voltage or short circuit current. Here comes the necessity of optimization and by perfect optimization accurate profile can be give better efficiency.

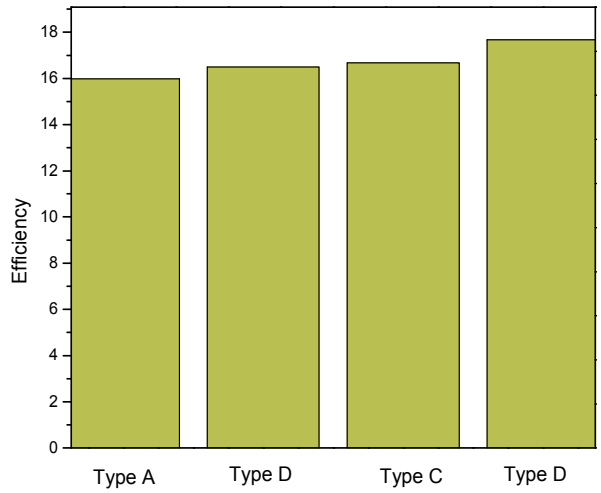


Fig. 8 Efficiency for different grading levels.

Short circuit current density variation for different graded gallium composition profile is shown in the following figure. It is already shown in literature that the efficiency of a solar cell gets increased with the increase in the gallium composition (not more that 30-40%) our goal was to achieve the higher efficiency (which is possible in the greater thickness level) in the lower thickness level by using graded composition of gallium. The efficiency of the solar cell got increased with the higher graded composition of gallium for type A, Type B and Type C. The efficiency was further increased at type D where more gallium was introduced and composition was changed from 0.1 to 0.6 and it obtained highest efficiency at 0.6 micron thickness level and the reason of this can be the more induced gallium and also the high grading effect.

C. Emitter Optimization: Lowering Thickness of Cds:

Our goal is to achieve efficiency as close as the highest efficiency using thinner absorber layer of CIGS. Hence so far we have done reverse grading and chosen optimized gallium composition to gain efficiency of 17% which is 2% lower than the highest efficiency. We varied the thickness of the Cds emitter thickness from 10 nm to 40 nm and for l0 nm.

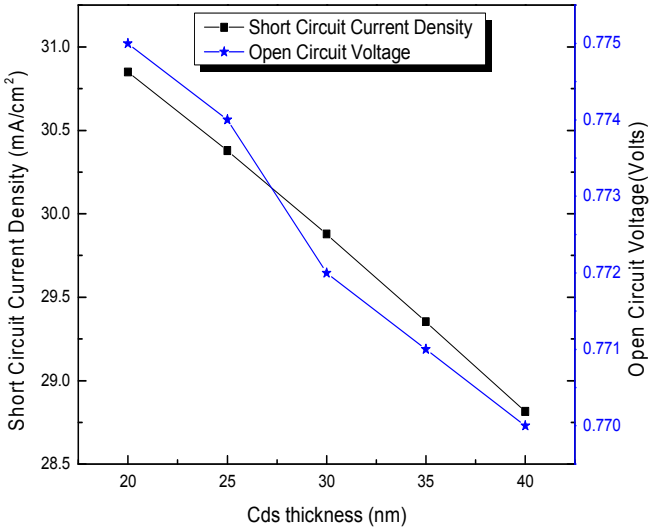


Fig. 9 Short circuit current density and open circuit voltage variation for different emitter thickness.

We achieved the highest efficiency of 19.1% at 0.6 micron with gallium composition 0.1 to 0.6 using a very thin emitter of 10 nm. The following figures: Fig.9 shows the CdS thickness effect on open circuit voltage and short circuit current, Fig.10 shows the effect on fill factor and Fig.11shows the effect on overall efficiency.

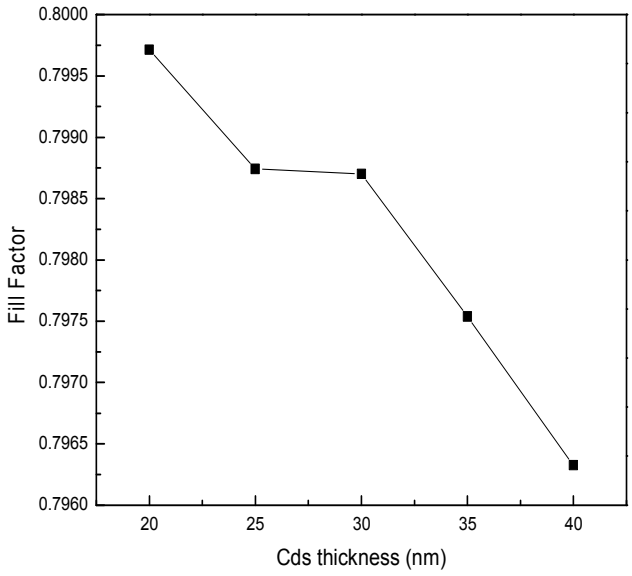


Fig.10 Fill factor variation for different emitter thickness.

Fill Factor is not that much affected by lowering Cds thickness though it gets better for lower Cds thickness. Efficiency is greatly included by emitter thickness. Efficiency reaches up to 19% for by lowering emitter thickness from 40 nm to 10 nm.

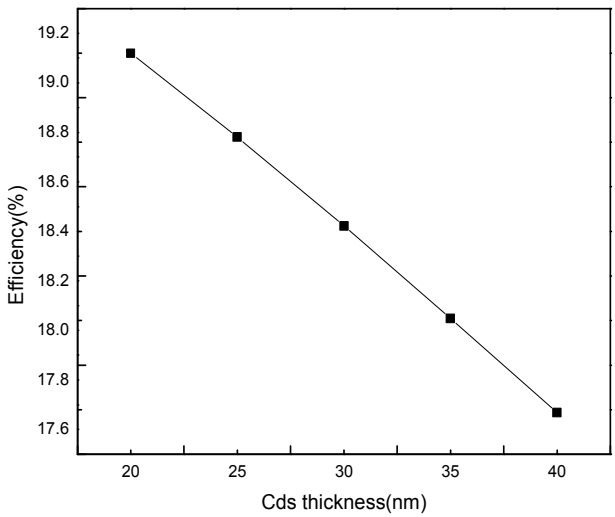


Fig. 11 Efficiency variation for different emitter thickness.

Our goal was to achieve near highest efficiency of the thick film CIGS solar by optimizing all possible ways. Here by reverse grading and Cds thickness optimization we show that it is possible to achieve high efficiency though using very thin absorber layer (0.6um).

IV. Conclusions

In this paper by simulation we have tried to optimize the thickness of CIGS solar cell. For mass production the thickness of cigs absorber must be reduced to a marginable limit by keeping good conversion efficiency as Indium and Gallium are not abandoned in nature. Grading which can be helpful in achieving high efficiency has been used here and using a lower thickness and higher band gap grading and also by increasing gallium component in the absorber the efficiency achieved is near the highest efficiency. Here we want to clarify that though the gallium composition has been increased for achieving high efficiency but as total quantity it is still a lot lower than conventional 1.5 to 2.5 micron thick CIGS solar cell. Again for achieving high efficiency having a thinner absorber layer it is quiet important to increase the emitter efficiency by using an appropriate emitter thickness. So by appropriate grading and emitter thickness it possible to have good high efficiency using thinner absorber layer.

References

[1] Joar Johansson, "Modelling and Optimization of CIGS Solar Cell Module," Master's Thesis, Lundsuniversitet, Dec., 2007.

[2] US Geological Survey, Mineral Commodity Summaries (2009) /https:// minerals.usgs. gov/minerals/pubs/mcs/index.htmlS.

[3] Jiyon Song , Sheng S. Li , C.H. Huang , O.D. Crisalle , T.J. Anderson. Device modeling and simulation of the performance of $CuIn_{1-x}Ga_xSe_2$solar cells. Solid-State Electronics 48 (2004) 73–79.

[4] O. Lundberg, M. Edoff, L. Stolt. The effect of Ga-grading in CIGS thin film solar cells. Thin Solid Films. Volumes 480–481, 1 June 2005, Pages 520–525.

[5] T. Dullweber, G. Hanna, W. Shams-Kolahi, A. Schwartzlander, M.A. Contreras, R. Noufi, H.W. Schock. Study of the effect of gallium grading in Cu(In,Ga)Se2 Thin Solid FilmsVolumes 361–362, 21 February 2000, Pages 478–481

The effect of the variation of the concentration factor and interdot distance of $In_xGa_{1-x}N$/GaN QD–IBSC

Nezam Uddin,[1,*] Kazi Nymul Haque,[2] Atiqul Islam,[3] Niloy Chandra Saha,[4] and Sushanta Paul[5]

Department of Electrical and Electronic Engineering, Khulna University of Engineering and Technology
Khulna–9203, Bangladesh
*nezam104@yahoo.com

Abstract— **Power conversion efficiency of conventional solar devices are low because low energy photons cannot excite carriers to the conduction band, therefore do not contribute to the device's current and high energy photons are not efficiently used due to a poor match to the energy gap. Efficiency of solar cell may be increased in various ways; now-a-days quantum dot intermediate band solar cell (QDIBSC) is most promising approach among them. Introducing intermediate levels into the energy gap of a conventional solar cell; low energy photons can be used to promote charge carriers in a stepwise manner to the conduction band thereby enhancing the current while maintaining a large open-circuit voltage. In this thesis $In_xGa_{1-x}N$/InN quantum-dot intermediate band solar cell is calculated by means of solving schr□dinger equation according to the kronig-penney model. On the basis of particular assumptions, the power conversion efficiency is calculated. The results reveal that the $In_xGa_{1-x}N$/InN quantum dot intermediate-band solar cell gives much larger power conversion efficiency than that of conventional solar cells and the power conversion efficiency strongly depends on the size of the quantum dot and the inter dot distance.**

Index Terms—**Quantum dot, intermediate band, concentration factor, interdot distance, $In_xGa_{1-x}N$.**

I. Introduction

The efficiency limit for a single junction solar cell is 33.5%. As for example the efficiency of non-concentrator solar cell has now reached 28.8% in single junction thin film GaAs solar cells under the global Air Mass 1.5 spectrum at 25°C [1]. Various researches have been performed for increasing the efficiency of solar cell with the composition of different semiconductor material.

Multi-junction solar cells (MJSCs) and intermediate band solar cells (IBSCs) are the two major designs for concentrator photovoltaic using inorganic III–V semiconductors. In case of MJSC; several PN junctions are connected in series. In this design each PN junction contributes to the absorption which produces broader solar spectrum. But for efficient absorption the photon energy should be nearest to the bandgap energy because only those photons will be absorbed with energy larger than the band gap energy of the material. Other parameter such as lattice constant or strain balance or thermal expansion co-efficient of each material should be same. For these criteria to be maintained the choice of the material combination become limited. The highest efficiencies of 43.5% have been demonstrated for MJSCs. Neglecting losses a modelled efficiency of 52.4% has been found for a triple junction photovoltaic cell with Ge, GaInP and GaAs [2]. Again the efficiency for non concentrator 3 junction solar cell composed of InGaP/GaAs/InGaAs is found 37.5%. From these results it has been shown that the efficiency of 3 junction solar cells exceeds the limit of single junction solar cell with the increased complexity [1].

For achieving higher efficiency limit; the intermediate band solar cell (IBSC) can be a novel approach. Assuming maximum sunlight concentration; the efficiency limit for IBSC solar cell is 63.2 % whereas 40.7% for single gap solar cell [3].The concept of IBSC was proposed by A. Luque and A. Marti in 1997. It has high conversion efficiency in comparison to the MJSC at a low level of complexity. IBSC is a single junction solar cell. Without affecting the open circuit voltage (Voc) it increases the photocurrent keeping the advantages of the single junction solar cell present [1].

Various types of material combinations are being used for IBSC such as the intermediate band solar cell concept can be implemented using InAs quantum dots (QDs) in a GaAs barrier material. The non-concentrator efficiency of 30.2% has been predicted for this material system. But under concentration, the efficiency of intermediate band solar cell in this material system increases to 51.6%, as compared to 36.7% for bulk GaAs solar cells, under a 1000 sun concentration [1].

In this paper we investigated the $In_xGa_{1-x}N$ /InN Quantum-Dot Intermediate-Band Solar Cell. The III–nitride material system is promising for integration into high efficiency PV systems [4]. One of the advantages of using the Indium Gallium Nitride ($In_xGa_{1-x}N$) material is the range of the bandgap energy from 0.64 eV to 3.4 eV by varying the indium composition (x), which covers most of the solar spectrum from ultraviolet to infrared region with characteristics of direct bandgap, high absorption coefficients, and excellent irradiance resistance. In low–dimensional semiconductor devices the $In_xGa_{1-x}N$ material system presents excellent properties[5]. By tuning the size of QDs and the interdot distance; Deng et al. theoretically investigated the InxGa1–xN/GaN QD–IBSC and $In_xGa_{1-x}N$ /InN QD–IBSC [6]. These works do not clarify about the effects of QD size, interdot distance and indium content on IB position and width which are very important factors for the realization of the performance of a QD–IBSC [7].

In our work, we have investigated the effect of variation of Concentration factor, interdot distance and indium content of InxGa1-xN /InN QD-IBSC and also on solar cell performance which may open the door for future experiment. The $In_xGa_{1-x}N$/InN QDIB solar cells are computed according to the Schrödinger equation and the Kronig–Penney model. The results reveal that the optimum indium content in InxGa1-xN is about 55% and the maximum efficiency under the maximal concentrated sunlight is about 60.3% with the optimum size of the QD (about 3.3nm) and optimum ID (about 2nm) based on specific assumptions.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

II. THEORY

A quantum dot is confined in three dimensions and has a size limited by the de Broglie wavelength in all directions. For simplicity, the orthorhombic symmetry for QDs is shown in Fig. 1. The QDs are embedded in periodic arrays. A single electron in such a system can be described by the Schrödinger equation [8]

$$[-\frac{\hbar^2}{2}\nabla_r \frac{1}{m^*}\nabla_r + V(r)]\varphi(r) = E\varphi(r) \quad (1)$$

Where m^* is the effective mass of electron, the potential V(r) corresponds to an infinite sequence of QDs of sizes Wx, Wy and Wz separated by the barriers of the thickness bx by and bz. Here Lx Ly Lz are periods of co-ordinates x,y and z, respectively $L\varsigma = W\varsigma + B\varsigma$

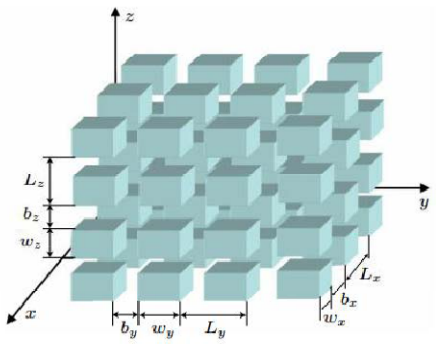


Fig. 1 Schematic Diagram of the Orthorhombic QDs.

where m^* is the effective mass of electron. The potential *V(r)* corresponds to an infinite sequence of QDs of the sizes *wx, wy* and *wz* separated by the barriers of thicknesses *bx*, *by* and *bz* as shown in Fig. 1. Here *Lx*, *Ly* and *Lz* represent the periods of QDs in the coordinates *x*, *y* and *z* respectively. The size of QDs and interdot distance are assumed to be equal (*wx* = *wy* = *wz* and *bx* = *by* = *bz*).

According to the Kronig–Penney model, solution of the Schrödinger equation has already been accomplished by following the Lazarenkova and Balandin envelope function approximation, which has the following form [8]

$$\mathrm{Cos}(L\varsigma q\varsigma) = \cos(K_\varsigma^w W_\varsigma)\cos(K_\varsigma^b b_\varsigma) - .5\left(\frac{K_\varsigma^w}{K_\varsigma^b}\frac{m_b^*}{m_w^*} - K\varsigma b K\varsigma w\, mw{*}mb{*} \times \sin(K\varsigma w W\varsigma)\sin(K\varsigma b b\varsigma)\right. \quad (2)$$

If $E_\varsigma \geq V_0$

$$\mathrm{Cos}(L\varsigma q\varsigma) = \cos(K_\varsigma^w W_\varsigma)\cosh(K_\varsigma^b b_\varsigma) - .5\left(\frac{K_\varsigma^w}{K_\varsigma^b}\frac{m_b^*}{m_w^*} - K\varsigma b K\varsigma w\, mw{*}mb{*} \times \sin(K\varsigma w W\varsigma)\sinh(K\varsigma b b\varsigma)\right. \quad (3)$$

If $0 < E_\varsigma < V_0$

Where

$$K_\varsigma^w = \frac{1}{h}\left(\sqrt{2m_w^* \times mod(E_\varsigma)}\right),\ K_\varsigma^b = \frac{1}{h}\left(\sqrt{2m_b^* \times mod(E_\varsigma - V_0)}\right)$$

where m_b^* and m_w^* are the effective masses of barrier region and quantum dot (QD) region, respectively, $\hbar$ is Planck's constant, $E\varsigma$ is the single electron energy along the particular coordinate axis direction and V_0 is the potential barrier.

III. PERFORMANCE ANALYSIS OF $IN_XGA_{1-X}N$/INN QD-IBSC

The effects of the variation of the concentration factor and interdot distance of $In_XGa_{1-X}N$ QD-IBSC will be shown in this section.

A. Effect of Variation of Concentration Factor x of $In_XGa_{1-X}N$

The maximum efficiency and short circuit current density, open circuit voltage has been calculated for the devices with the various concentration factor x. The other parameter including width of QD and inter dot distance of the cell kept constant.

TABLE I
SIMULATION RESULTS OF VARYING THE CONCENTRATION OF $IN_XGA_{1-X}N$ (FOR QD WIDTH W=2.5NM B=1-4NM)

Device No	Width of QD w (nm)	Width of Interdot Distance b(nm)	Open Circuit Voltage Voc (volt)	Short Circuit Current Jsc (A/cm2)	Maximum Efficiency □ (%)
1	2.5	2	1.5578	3169.4	53.8206
2	2.5	3	1.5578	3141.4	53.3453
3	2.5	4	1.5578	3137.1	53.2730

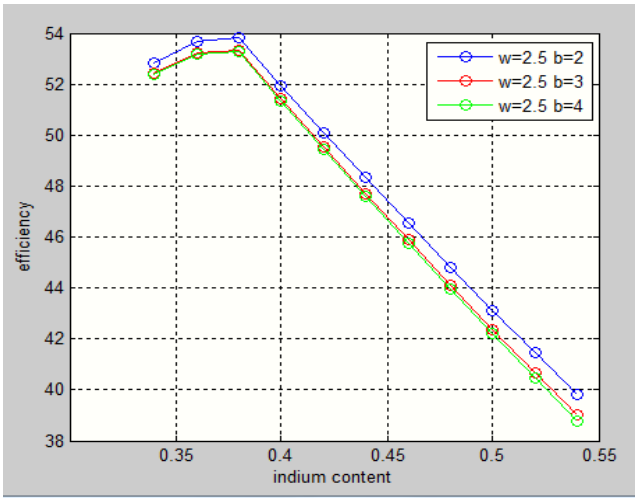


Fig. 1 Indium Content and ID-Dependent Efficiency in the InxGa1-xN/InN Device (for w=2.5nm b=1-4nm)

The short circuit current, open circuit voltage, intermediate band energy is shown in fig. 2 for b=2nm and w=2.5

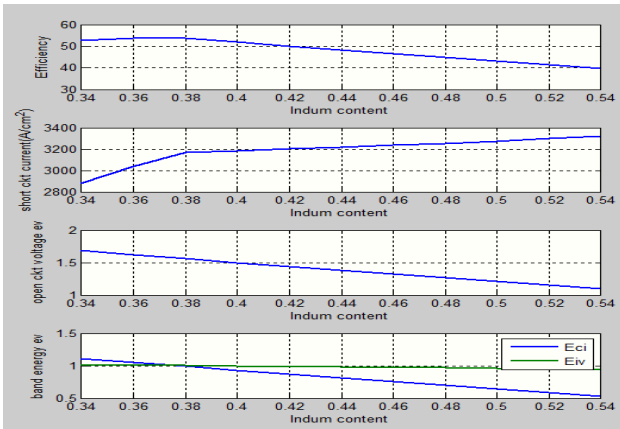


Fig. 2 Efficiency, Short Circuit Current, Open Circuit Voltage, Intermediate Band Energy for (w=2.5nm, b=2nm)

TABLE II
SIMULATION RESULTS OF VARYING THE CONCENTRATION OF $In_xGa_{1-x}N$
(FOR QD WIDTH W=3NMB=1-4NM)

Device No	W idth of QD w (nm)	Width of Interdot Distance b(nm)	Open Circuit Voltage Voc (volt)	Short Circuit Current Jsc (A/cm2)	Maximum Efficiency (%)
1	3	1	1.4363	3495.1	54.72
2	3	2	1.4363	3304.8	51.74
3	3	3	1.4363	3278.2	51.33
4	3	4	1.4363	3274.2	51.26

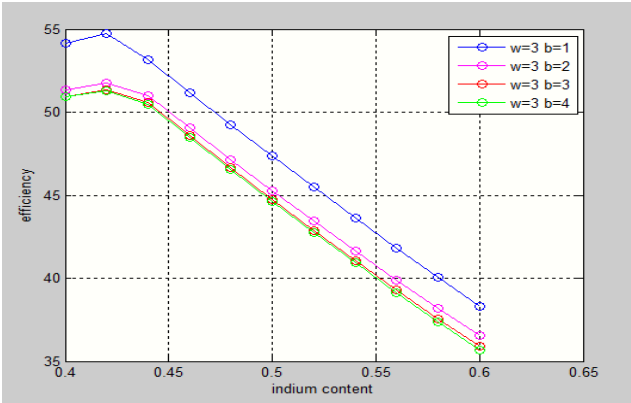


Fig. 3 Indium Content and ID-Dependent Efficiency in the InxGa1-xN/InN Device (for w=3nm b=1-4nm)

The short circuit current, open circuit voltage, intermediate band energy are shown in fig. 4 for b=1nm and w= 3nm.

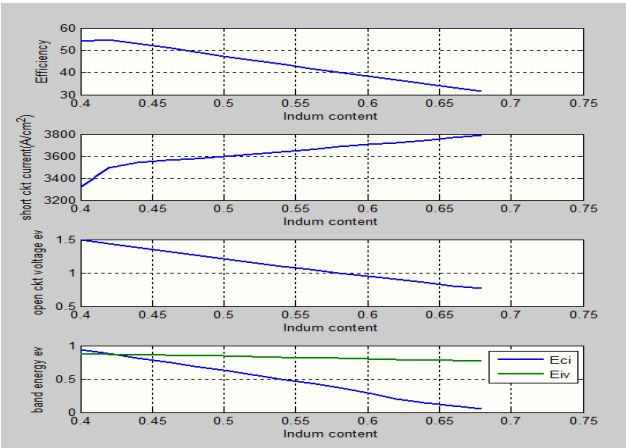


Fig.4 Efficiency, Short Circuit Current, Open Circuit Voltage, Intermediate Band Energy for (w=3.0nm, b=1nm)

TABLE III
SIMULATION RESULTS OF VARYING THE CONCENTRATION OF INXGA1-XN
(FOR QD WIDTH W=3.5NM B=1-4NM)

Device No	Width of QD w (nm)	Width of Interdot Distance b(nm)	Open Circuit Voltage Voc (volt)	Short Circuit Current Jsc (A/cm2)	Maximum Efficiency □ (%)
1	3.5	1	1.3195	3627.6	52.1780
2	3.5	2	1.3195	3467.9	49.8817
3	3.5	3	1.3195	3443.0	49.5228
4	3.5	4	1.3195	3438.6	49.4603

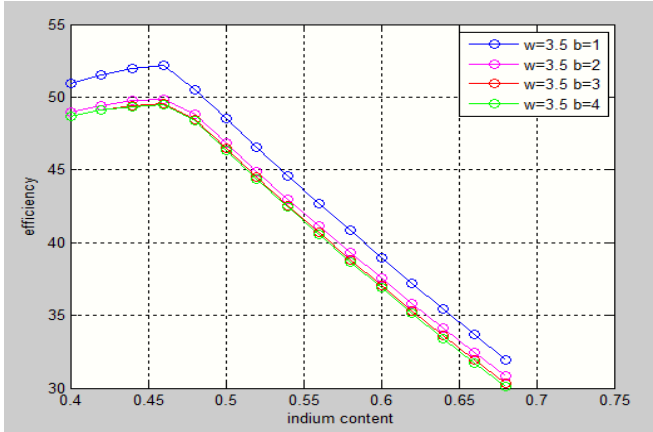


Fig. 5 Indium Content and ID-Dependent Efficiency in The InxGa1-xN/InN Solar Cells (for w=3.5nm b=1-4nm)

The short circuit current, open circuit voltage, intermediate band energy are shown in fig. 6 for b=1nm and w=-3.5nm.

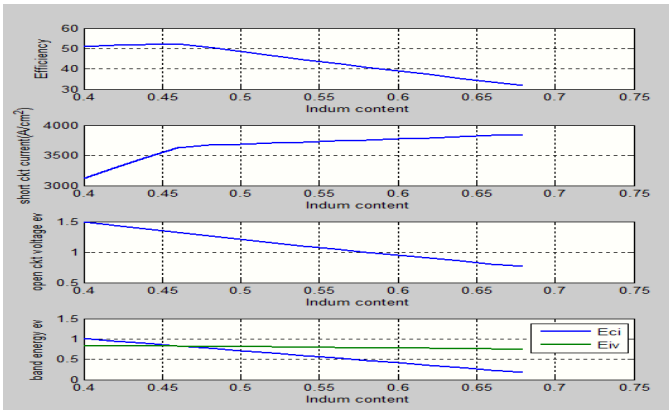


Fig. 6 Efficiency, Short Circuit Current, Open Circuit Voltage, Intermediate Band Energy for (w=3.5nm, b=1nm)

TABLE IV
SIMULATION RESULTS OF VARYING THE CONCENTRATION OF $In_xGa_{1-x}N$
(FOR QD WIDTH W=4NM B=1-4NM)

Device No	Width of QD w (nm)	Width of Interdot Distance b(nm)	Open Circuit Voltage Voc (volt)	Short Circuit Current Jsc (A/cm2)	Maximum Efficiency □ (%)
1	4	1	1.3195	3639.1	50.0942
2	4	2	1.3195	3511.7	48.3406
3	4	3	1.3195	3340.3	48.0464
4	4	4	1.3195	3486.5	47.9938

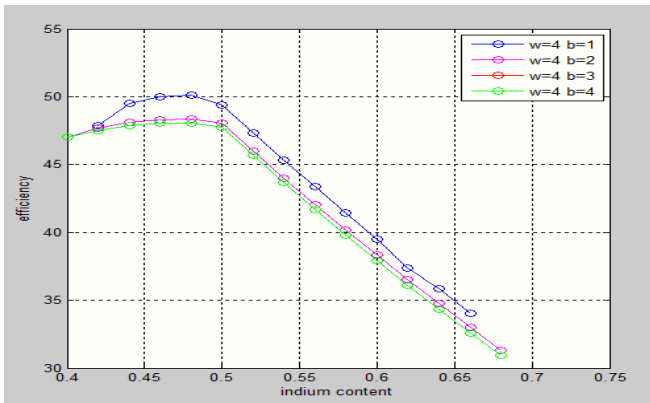


Fig. 7 Indium Content and ID-Dependent Efficiency in the InxGa1-xN/InN Solar Cells (for w=4nm b=1-4nm)

The short circuit current, open circuit voltage, intermediate band energy are shown in fig. 8 for b=1nm and w=4nm.

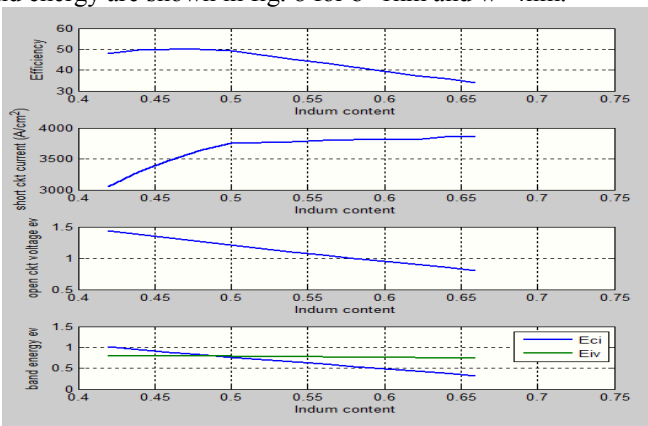


Fig. 8 Efficiency, Short Circuit Current, Open Circuit Voltage, Intermediate Band Energy for (w=4nm, b=1nm)

B. Effect of variation of Interdot Distance of InxGa1-xN/InN QD-IBSC

The maximum efficiency and short circuit current density, open circuit voltage has been calculated for the devices by varying interdot distance of QD-IBSC. The other parameter including width of QD and concentration factor x of InxGa1-xN is kept constant.

TABLE V
SIMULATION RESULTS OF VARYING THE INTERDOT DISTANCE OF INXGA1-XN (FOR X=0.48 AND W=3.3-4NM)

Device No	Width of QD w (nm)	Concentration Factor x of InxGa1-xN	Open Circuit Voltage Voc (volt)	Short Circuit Current Jsc (A/cm2)	Maximum Efficiency □ (%)
1	3.3	.48	1.26282	3636.32	50.0942
2	3.5	.48	1.26282	3668.96	50.5065
3	3.8	.48	1.26282	3695.36	50.87
4	4	.48	1.26282	3640.8	50.1173

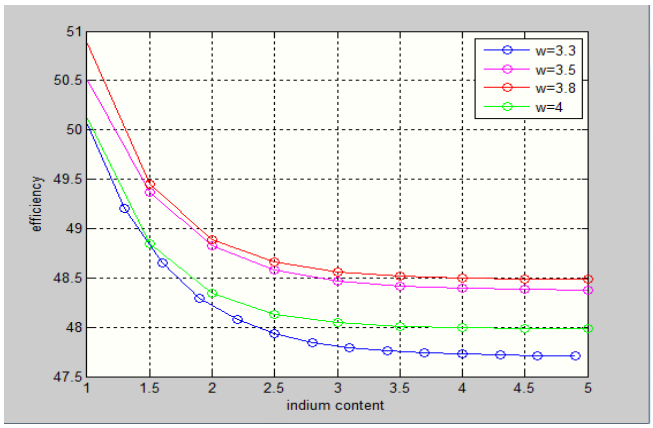


Fig. 9 ID and the Size of Each QD-Dependent Efficiency in the InxGa1-xN/InN QDIB Solar Cell.

IV. CONCLUSION

The results have been achieved under maximum concentrated sunlight. The size of each QD and the ID are not selected arbitrarily, but are chosen after calculation. Fig. 3 reveals that the optimal indium content depending on ID is about 0.42 and maximum efficiency is about 54.72%, which is obviously lower than the limit efficiency of 63.1%. This is mainly because the ID and size of each QD are not optimum, and the restriction of the existence of only one IB in the potential well ; which results in the different ranges of indium content to the differences of ID in fig. 3. The intermediate band makes the electron transition easier from valence band to conduction band and improves the absorption of sunlight. The intermediate band divides the band gap of InxGa1-xN into three parts, the photons whose energies between the band gap of intermediate band to conduction band and the valence band to conduction band are absorbed for the existence of intermediate band, while not be absorbed in bulk materials solar cells. Therefore, the power conversion efficiencies of QDIB solar cells are higher than those of bulk material solar cells. The smaller interdot distance causes the easier coupling of electrons wave functions from adjacent QDs and the easier formation of intermediate band.

The size of each QD and the ID are tuned in fig. 9 but they are not changed arbitrarily. The different ranges of the ID depend on the differences of the size of one QD so as to fulfill assumption 1. Fig. 9 reveals that the optimal size of each QD is about 3.8 nm and the optimal ID is about 1nm. The maximum efficiency is about 50.87% under the maximum concentrated sunlight which is close to limit efficiency of 63.1%. The smaller size of QD makes the quantum effect more obvious and the formation of intermediate band easier. The size of one QD and the ID vary the width and the location in potential well of IB, and the power conversion efficiency ultimately depends on the width and location of IB. Therefore, the power conversion efficiency dependent on the size of each QD and ID is depicted as fig. 9.

In summary, the InxGa1-xN/InN QDIB solar cells are computed according to the Schrödinger equation and the Kronig–Penney model. The results reveal that the optimum indium content in InxGa1-xN is about 55% and the maximum efficiency under the maximal concentrated sunlight is about 60.3% with the optimum size of the QD (about 3.3nm) and optimum ID (about 2nm) based on specific assumptions.

REFERENCES

[1] Tian Li, Robert E. Bartolo, and Mario Dagenais, “Challenges to the concept of an intermediate band in InAs/GaAs quantum dot solar cells,” *APPLIED PHYSICS LETTERS 103, 141113*, vol. 103, pp. 141113 - 141113-4, 2013.

[2] Qiao-Yi Wang , Rorison, J., “Modelling of quantum dot intermediate band solar cells: effect of intermediate band linewidth broadening,” *Optoelectronics, IET* , vol. 8, pp. 81-87, April 2014.

[3] Antolin, E., Marti, A., Farmer, C.D., Linares, P.G., “Reducing carrier escape in the InAs/GaAs quantum dot intermediate bandsolar cell,” *Journal of Applied Physics* , vol. 108, pp. 064513 - 064513-7, September 2010

[4] Na Lu and Ian Ferguson, “III–nitrides for energy production: photovoltaic and thermoelectric applications,” Semiconductor Science and Technology, vol. 28, no. 7, pp. 4023–4033, 2013.

[5] Ashraful Ghani Bhuiyan, Kenichi Sugita, Akihiro Hashimoto, and Akio Yamamoto, “InGaN Solar Cells: Present State of the Art and Important Challenges,” IEEE Journal of Photovoltaics, vol. 2, no. 3, pp. 276–293, July 2012.

[6] Qingwen Deng, Xiaoliang Wang, Cuibai Yang, Hongling Xiao, Cuimei Wang, Haibo Yin, Qifeng Hou, Jinmin Li, Zhanguo Wang, Xun Hou “Computational Investigation of InxGa1–xN/InN Quantum–Dot intermediate–Band Solar Cell,” Chin. Phys. Lett., vol. 28, no. 1, pp. 8401–8404, 2011.

[7] Islam, M.M.; Bhuiyan, M.T.I. ; Rahman, M.T. ; Bhuiyan, A.G., “The effect of quantum dot size, interdot distance and indium content on InxGa1–xN/GaN QD–IBSC,” International Conference on Electrical Information and Communication Technology (EICT), 2013 , pp. 1-4, February 2014

[8] O. L. Lazarenkova, A. A. Balandin, “Miniband formation in a quantum dot crystal,” J. Appl. Phys., vol. 89, no. 10, pp. 5509–5515, 2001.

Voltage Stability Management With the Help of Optimization Techniques

Md. Imran Azim,[1*] Md. Masud Rana,[2] Jawad Tahmeed Bin Taufik,[3]

[1] School of Engineering and Information Technology, University of New South Wales
Australian Defence Force Academy, Canberra BC 2610, Australia

[2,3] Department of Electrical and Electronic Engineering, Rajshahi University of Engineering and Technology (RUET)
Kazla, Rajshahi-6204, Bangladesh

*imran.azim89@gmail.com

Abstract—**This paper deals with the injection of reactive power, which is vitally important in case of too much load variations, by means of Static VAR Compensator (SVC) so that both, the voltage profile and voltage stability of the power system can be improved. The reactive power management is formulated as an optimization problem that can be solved by the evolutionary advanced optimization algorithms. Two well known algorithms, namely linear algorithm and genetic algorithm are applied in this case and it has been found that the genetic algorithm gives more efficient outputs compared to the linear algorithm as it is based on the biological metaphor, in which best individuals are selected among parents and offspring generation. As far as this paper is concerned, 0.2 MVAR reactive power would be enough for voltage compensation and maintaining stability throughout the system.**

Index Terms—**System Stability, SVC, Linear Algorithm, Genetic Algorithm.**

I. Introduction

One important fact is that power system operators and planners are still facing challenges of integrating renewable energy sources into power system grids with ensuring security, reliability and stability. Out of this context, Reactive Power Management (RPM) has been arisen as a critical issue that entails the requested operation and planning actions to be implemented so that not just the voltage profile but the voltage stability can be improved [1]. Moreover, the reactive power planning of VAR sources are defined by RPM and the precise implementation of it has been permitted by the active power networks, in which active power flow is influenced by Flexible AC Transmission Systems (FACTS) devices [2-3]. Conventionally, reactive power planning has been formulated as an optimization problem and Optimal Power Flow (OPF) solutions can be applied to solve this [4]. In this situation, the utilization of optimization techniques, including Linear Algorithm (LA) and Genetic Algorithm (GA), a type of evolutionary optimization algorithm developed by Holland in 1975 [5], are fruitful.

This paper gives a clear idea on linear algorithm and genetic algorithm and also proofs the advantages of genetic algorithm over linear algorithm in solving linear objective function, involving load factor, fluctuated voltage and amount of reactive power. It goes without saying that the proposed methodology will aid the power system operators to investigate reactive power management in smart grid system integrated with renewable energy for the improvement of system stability.

II. Reactive Power Management

Active network requires a dynamic reactive power support so that voltage can be regulated but the exceeding voltage limit variation is one of the technical barriers for the integration of renewable energy source. The implementation of FACTS devices, such as SVC solves this problem as they monitor the network. Moreover, the local generation of reactive power reduces its import from the feeder, thus reduces the associated losses, and improves the voltage profile and voltage security [6].

Usually, the load of a system is never constant rather it varies from time to time. Slight variation of load does not affect the system but large load variation is problematic as it changes constant supply voltage and urges the reactive power management [7]. Load variations with respect to time can be realized from load curve.

The ratio of average load to maximum demand is called load factor which is always less than 1 because average load is smaller than maximum demand. The load factor plays a key role in determining the overall cost per unit generated as higher value of it makes the cost per unit generation lesser.

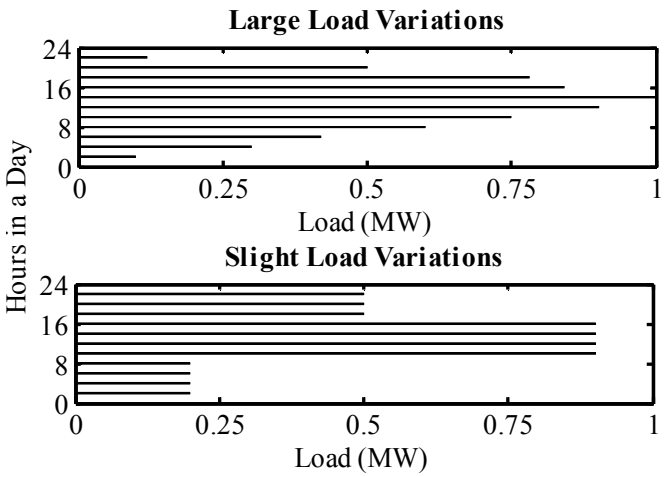


Fig. 1. Daily load curve.

However, too much reduction in load causes overvoltage. The effects of it are large magnetizing current, overheating and low power factor, which urges the large KVA rating of the equipments.

On the other hand, large enhancement in load causes the voltage to drop and induction motors connected to the system may not start due to the lack of sufficient torque.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

Therefore, voltage of the system should be kept constant [8] and to make sure of it, the provision should be made for the injection of reactive power [9]. In this paper, SVC, which is capable of exchanging reactive current to control certain parameters used in power system, is taken into account for this purpose. The dynamic nature of the SVC lies in the use of thyristor devices such as GTO, IGCT.

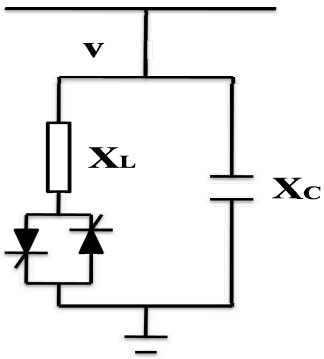


Fig.2. Static VAR compensator diagram.

As is observed that v is the voltage at Thyristor based SVC connection point that is being controlled. X_L is the total inductance and X_C is the capacitance.

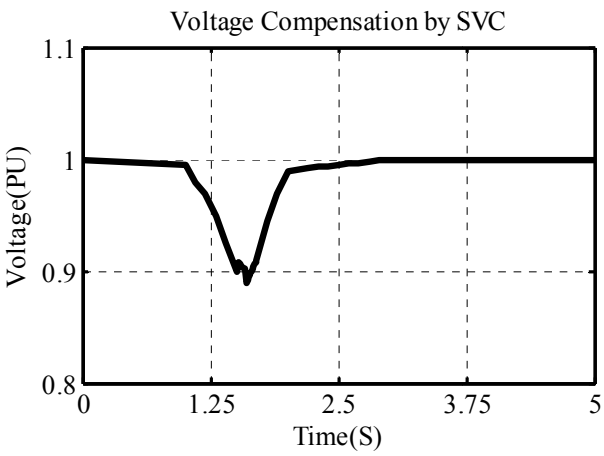


Fig. 3. Reactive power injections by SVC.

III. Advanced Optimization Techniques

Optimization techniques are used to find a set of design parameters, x = {x_1, x_2, ...,x_n}, that can in some way be defined as optimal. In a simple case this might be the minimization or maximization of some system characteristic that is dependent on x. In a more advanced formulation the objective function, f(x), to be minimized or maximized, might be subject to constraints in the form of equality constraints, inequality constraints, and/or parameter bounds ; both upper and lower. When both the objective function and the constraints are linear functions of the design variable, the problem is known as a Linear Programming (LP) problem. Quadratic Programming (QP) concerns the minimization or maximization of a quadratic objective function that is linearly constrained.

The objective of utilizing optimization algorithms in this paper is to find the optimal values of load factor, λ, voltage fluctuation, V caused by the undesired load variations and also the optimum amount of reactive power, Q that is needed to be injected for voltage compensation by SVC.

To begin with, two assumptions are made. Firstly, voltage should be decreased by high load factor and secondly, higher amount of reactive power should be injected if voltage decreases.

TABLE I
INITIAL CONSIDERED VALUES IN PER UNIT (PU)

Load factor, λ (PU)	Fluctuated voltage, V (PU)	Reactive power injection, Q (PU)	Objective function, f(λ,V,Q)
0.12	0.98	0.80	0.18
0.17	0.95	0.89	0.21
0.45	0.87	0.97	0.33

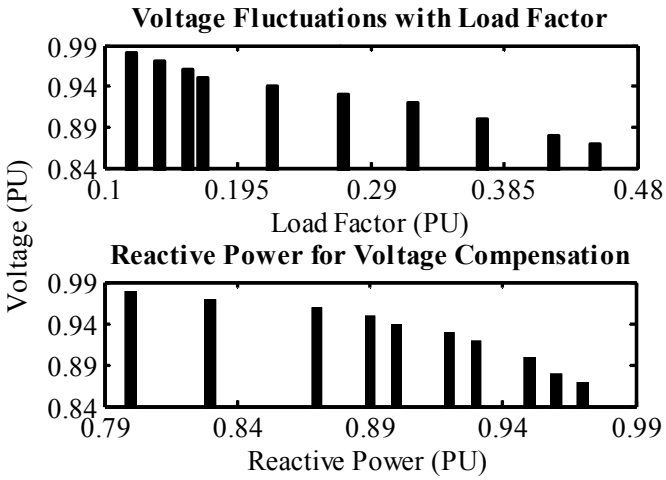


Fig. 4. Assumed conditions to start optimization algorithms.

From table, solved linear equations,

$$0.12x+0.98y+0.8z=0.18 \quad (1)$$

$$0.17x+0.95y+0.89z=0.21 \quad (2)$$

$$0.45x+0.87y+0.97z=0.33 \quad (3)$$

Solutions: x=0.404; y=0.0357; z=0.1206

Derived objective function,

$$f(\lambda,V,Q)=0.404\lambda+0.0357V+0.1206Q \quad (4)$$

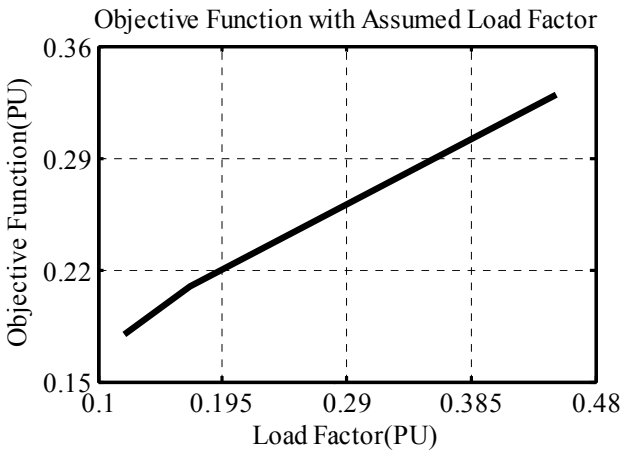


Fig. 5. Relation of load factor to the estimated objective function.

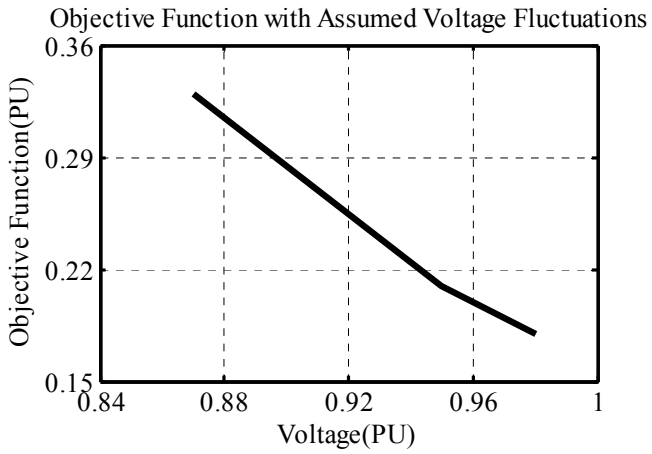


Fig. 6. Relation of voltage fluctuations to the estimated objective function.

A. Linear Algorithm

Since the objective function is linear, linear algorithm has been applied so that the optimal values of the required parameters can be determined. In linear algorithm, the constraints are also linear but solutions derived from it are minuscule values and are not satisfactory as seen from Fig. 7, and to overcome these, the implementation of an advanced algorithm is necessary.

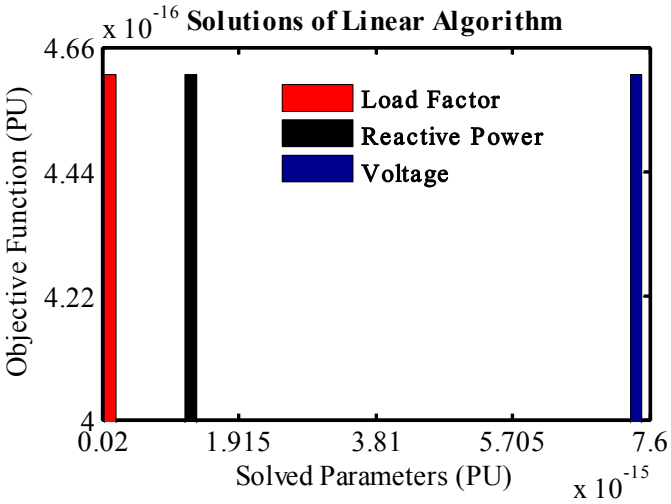


Fig. 7. Optimum values solved by linear algorithm.

B. Genetic Algorithm

Genetic algorithm is the most up-to-date optimization technique that is operated according to the biological concept. It is a meta-heuristic optimization method, in which a problem is solved by improving the candidate solution based on certain criteria [10]. It is begun with the initialization of population. Each solution of population is called chromosomes which simply represent the values of the parameters to be optimized. Then the objective function is evaluated and the chromosomes are selected based on higher values of objective function. Cross is made between two parent values. Mutation brings variety in the values. This process continues until the best possible optimum values of the parameters are achieved and finally after providing results, the algorithm procedure is terminated [11-12].

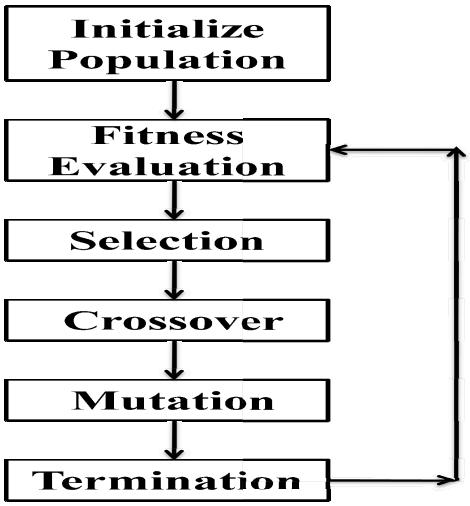


Fig.8. Basic structure of Genetic Algorithm (GA).

TABLE II
DATA FOUND AFTER 20 GENERATIONS

Generations	Best f(λ,V,Q)	Mean f(λ,V,Q)
1	0.12320	0.23490
2	0.09140	0.20250
3	0.06830	0.15330
4	0.06830	0.12700
5	0.06830	0.09560
6	0.06830	0.09057
7	0.06830	0.07959
8	0.06830	0.07412
9	0.06830	0.06872
10	0.06830	0.06870
11	0.06830	0.06853
12	0.06830	0.07097
13	0.06830	0.07353
14	0.06830	0.06830
15	0.06830	0.06830
16	0.06830	0.06830
17	0.06830	0.06830
18	0.06830	0.06830
19	0.06830	0.06830
20	0.05932	0.06785

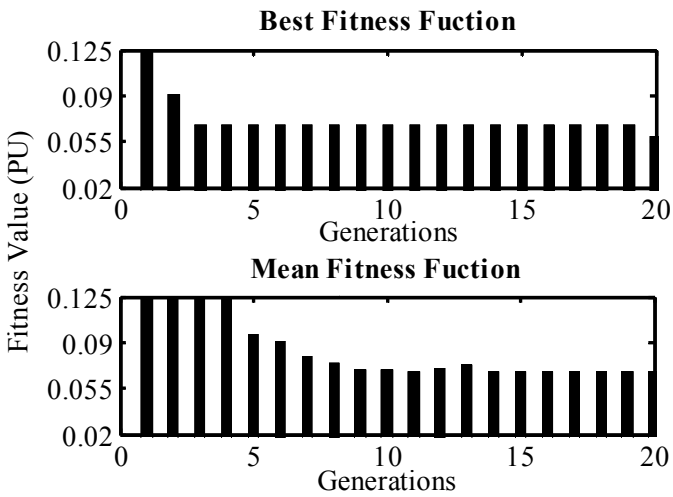


Fig. 9. Best and mean fitness functions after 20 generations.

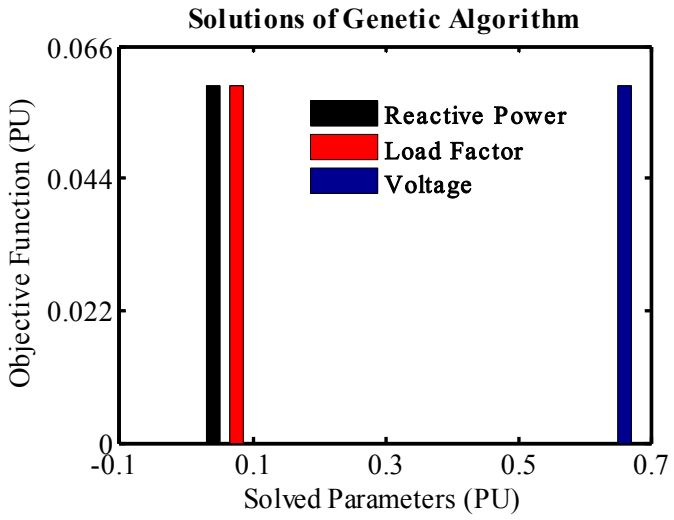


Fig. 10. Optimum values solved by genetic algorithm.

TABLE III
OPTIMIZED SOLUTIONS of GENETIC ALGORITHM

Load Factor, λ (PU)	Fluctuated voltage, V (PU)	Reactive power injection, Q (PU)	Objective Function, $f(\lambda, V, Q)$
0.07639978387976154	0.6617988711062396	0.04006849638182075	0.059323993049564

TABLE IV
ASSUMED BASE VALUE of the PARAMETERS

Maximum demand	Supply voltage	SVC rating
1 MW	230 V	5 MVAR

$$\text{Average load} = \text{Load factor} \times \text{Maximum demand} \quad (5)$$

The actual values of the parameters can be found from the following equation

$$\text{Actual value} = \text{Base value} \times \text{Per unit value} \quad (6)$$

TABLE V
EVALUATED ACTUAL VALUES of the PARAMETERS

Average load	Supply voltage	SVC rating
0.07639978387976154 MW	152.2137265 V	0.200342481 MVAR

Therefore, if it is considered that the maximum demand of the system is 1 MW, while the maximum values of supply voltage and the maximum rating of available SVC are 230 V and 5 MVAR respectively, then the solution of genetic algorithm in terms of actual value suggests that 76.4 kW average load causes the voltage to fluctuate approximately to 152 V and 0.2 MVAR reactive power is needed to supply to the system for voltage compensation.

IV. CONCLUSIONS

The procedure of attaining system stability with the help of Static VAR Compensator (SVC) has been presented in this paper as it has the capability of providing reactive power. The utilization of optimization algorithms is fruitful in this case as they give optimum values. Two algorithms are applied and the comparative solutions show that genetic algorithm solves the linear objective function and gives the optimum values of load factor, voltage fluctuation and reactive power in a more precise way than that of linear algorithm.

In future, these optimized methodologies would find its extensive utility when it comes to dealing with reactive power management in a large power system.

REFERENCES

[1] Miller, T. J. E, *Reactive Power Control in Electric Systems*, Wiley Inter-science.

[2] Xiong H., Cheng H., Li H., "*Optimal Reactive Power Flow Incorporating Static Voltage Stability Based on Multi-objective Adaptive Immune Algorithm,*" Journal of Energy Conversion and Management, 49, pp. 1175–1181, 2008.

[3] Lahatani N., Aouzellag D., Mendil B, "*Contribution to the Improvement on Voltage Profile in Electrical Network with Wind Generator Using SVC Device*," Renewable Energy, 35 (1), pp. 243–248, 2010.

[4] Raoufi H., Kalantar M, "*Reactive Power Rescheduling with Generator Ranking for Voltage Stability Improvement*," Journal of Energy Conversion and Management, 50, pp. 1129–1135, 2009.

[5] J. H. Holland, Genetic Algorithms. Scientific American, 1992.

[6] Hedayati, H., Nabaviniaki, S. A., Akbarimajd, A., "*A Method for Placement of DG Units in Distribution Networks*", *IEEE Transactions on Power Delivery,* pp. 1620–1628, 2008

[7] Ajjarapu V., *Computational Techniques for Voltage Stability Assessment and Control*, Springer, 2006.

[8] V. K. Mehta, *Principles of Power System*, 4th Revised Edition, S. Chand & Company Ltd., 2011.

[9] S. Deshmukh, B. Natarajan, A. Pahwa, "*State Estimation and Voltage/VAR Control in Distribution Network With Intermittent Measurements*," IEEE Transactions on Smart Grid, pp.1-10, 2013.

[10] Holland. J, "*Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence,*" Univ. of Michigan Press, 1975.

[11] P. C. Ramaswamy, G. Deconinck, "*Smart Grid Reconfiguration Using Simple Genetic Algorithm and NSGA-II,*" 3rd IEEE PES Smart Grid Technologies Europe (ISGT Europe), Berlin, 2012.

[12] R. Joshua Samuel Raj, Vasudevan, "*Grid Scheduling with smart Genetic Algorithm,*" International Journal of Grid Computing and Multi Agent Systems, vol. 2, no. 1, January-June, pp. 1-10, 2011.

Power System Stabilizer Tuning Based on Frequency Response Method

Asif Ferdoush,* M. G. Rabbani

Department of Electrical & Electronic Engineering, Eastern University
Dhaka-1205, Bangladesh
*asifferdoush@gmail.com

Abstract— **Stability phenomena of the power system are largely studied by the researchers over the last few decades. In order to ensure a stable power system, power system stabilizer (PSS) is adopted which can increase the stability margin by providing supplementary feedback signal. For an efficient stabilization, the proper tuning of PSS parameters is very important. Now a days, the tuning appears a big challenge as the changes in power requirement take place continually. This paper deals with the tuning process of PSS connected to the single machine infinite bus (SMIB) system. The tuning method is decided based on the transfer function from SMIB model and the system frequency response. Simulation results are presented with and without PSS to demonstrate the effectiveness of proposed tuning method.**

Index Terms— **Stability, low frequency oscillation, damping torque, PSS, transfer function, frequency response.**

I. Introduction

To ensure power system security and protection, an extensive research has been carrying out worldwide on the stability of power system. It is very important and a major requirement to secure system operation for an efficient power system [1]. Continuous network extension and ongoing interconnection due to increased load demand raises complexity of the power system [2]. After being perturbed by any kind of disturbances (i.e. continuous load change, line to line or line to ground fault, lightening etc.) from power grid, the rotor oscillation of the synchronous generator takes place. This low frequency oscillation must be damped out quickly before it gets stronger with increasing amplitude [3]. Otherwise, the synchronism may be deteriorated and synchronous generator will be tripped out from the grid by means of protective relays. Therefore, it is imperative that special control methods are used to perform the stabilization and adequately damping these low frequency oscillations [1]. In order to do that, PSS is employed with the system which is equipped with generator control system to provide adequate damping torque to the rotor during oscillations. As a result, power system stability gets improved and thereby extends the power transfer capability [4]. Now a days, population based methods like Particle Swarm Optimization (PSO), Genetic Algorithm (GA) etc. are very commonly used for PSS tuning problems [4]. In this tuning approach, conventional power system stabilizer (CPSS) is studied and its parameters are tuned by using frequency response.

II. Background of Rotor Oscillation

When power system is subjected to any kind of disturbances, the rotor angle of synchronous generator oscillates [5]. The oscillation mainly comes from the change in power requirements. Normally under steady state operation, there maintains a balance between mechanical power given by the prime mover and electrical power output of the generator [6]. But with the increase in load, the output of electrical power also increases relative to the input of mechanical power and in turn, both powers appear to be unbalanced. As a result, the stator revolving field of generator increases with constant rotor field. Therefore, the resulting electromagnetic torque, which opposes the rotor rotation, also changes [7]. So, rotor speed tends to decrease for increasing load. But input mechanical torque given by the prime mover opposes this speed to be changed. Hence, an electromechanical oscillation occurs according to the laws of motion of a rotating body [6]. When load decreases, the same interaction as mentioned above between rotor and stator field occurs but in reverse way and rotor speed tends to increase. When rotor oscillates, its speed deviates slightly from the rated speed in positive and negative manner. So field must be increased when speed deviation is negative and vice versa. In general, the excitation system of generator is equipped with PSS which can sense the speed change and provides appropriate damping torque based on the proper tuning technique [3].

III. System Model

The following is the feedback power control model where GS is generator and E_g is generated voltage to the grid which is fed back to the summing point to regulate itself by modulating the rotor field, E_{FD}. Turbine provides mechanical power P_m to the generator shaft. In this model, PSS is added to provide more damping by controlling the field current I_f. It senses the change of rotor speed and gives supplementary feedback signal E_{PSS} following a disturbance. So, with the change of rotor speed ω, the rotor field also changes by PSS. For a proper stabilization, the change of field current must be in phase with the change in rotor speed [3].

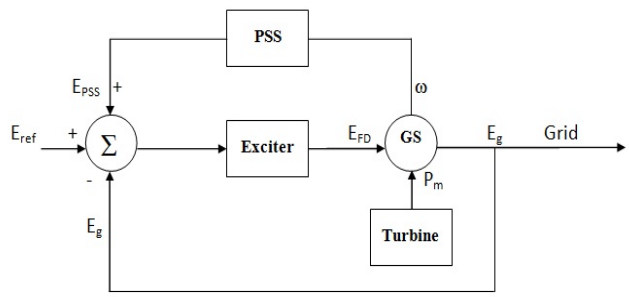


Fig. 1. Feedback power control model

A. Linear Model

The real power control model is a very highly nonlinear model and sophisticated to use. Hence the following

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

simplified linear model is used for synchronous generator system [8].

$$\Delta E'_q = \frac{k_3}{1+sk_3T'_{d0}}\Delta E_{FD} - \frac{k_3k_4}{1+sk_3T'_{d0}}\Delta\delta \quad (1)$$

$$\Delta T_e = k_1\Delta\delta + k_2\Delta E'_q \quad (2)$$

$$\Delta V_t = k_5\Delta\delta + k_6\Delta E' \quad (3)$$

$$2H\Delta\dot{\omega} = \Delta T_m - \Delta T_e - \Delta T_D \quad (4)$$

$$\Delta T_D = D\Delta\omega \quad (5)$$

$$\Delta\delta = \frac{\omega_0}{s}\Delta\omega\ ;\quad H \overset{\Delta}{=} \frac{\frac{1}{2}J(\omega_B\frac{2}{P})^2}{S_B} \quad (6)$$

The parameters of the above equation are summarized below:

TABLE I
SYSTEM PARAMETER

Parameter	Description
E_{FD}	Field voltage
δ	Rotor angle
V_t	Terminal voltage of generator
E'_q	Excited voltage
T'_{d0}	d-axis transient time constant
H	Inertia
T_D	Damping torque
T_e	Electrical torque
T_m	Mechanical torque

B. PSS

A commonly used approach is based on the conventional PSS structure which is composed of a gain, a wash out circuit and a cascade of two-phase lead network [9]-[11]. The output of PSS is then limited by the limiter to its required value. The PSS output is normally a signal applied to the reference input of the excitation system. The generator speed is assumed as the input of the PSS structure [12].

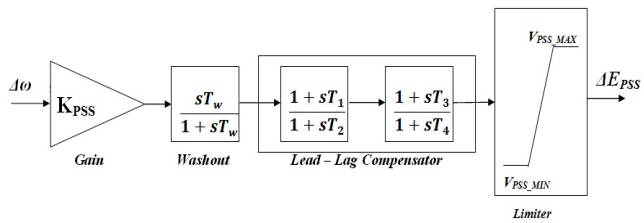


Fig. 2. Conventional PSS structure.

The transfer function of the PSS block is given below:

$$PSS = K_{PSS}\left(\frac{sT_w}{1+sT_w}\right)\left(\frac{1+sT_1}{1+sT_2}\right)\left(\frac{1+sT_3}{1+sT_4}\right) \quad (7)$$

TABLE II
PSS PARAMETERS

Parameter	Description
K_{PSS}	PSS gain compensator
T_w	Washout time constant
T_1, T_2, T_3, T_4	PSS phase compensator parameters

C. Single Machine Infinite Bus (SMIB) System

From the above linear equation a model is found which is also called Heffron-Phillips SMIB model and given next [13]. Equations of the constants k_1-k_6 are given in the appendix [14].

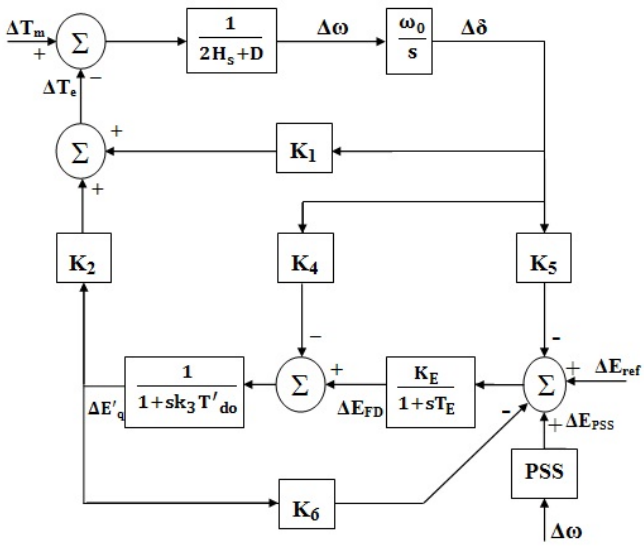


Fig. 3. Linearized synchronous generator control model.

From the above model, the stabilization of the power system means the stabilization of the power angle δ. By controlling the power angle, speed deviation can be damped out. But from the equation (4), δ depends on electromagnetic torque T_e and damping torque T_D. In the short period of time, the input mechanical torque T_m from turbine has no effect on δ, following a disturbance, because of its high inertia. And from the equation (5) damping torque always changes with the speed change Δω. Then, to make the system stable, the controlling parameter left is T_e in order to create more damping. And maximum damping can be obtained by keeping ΔT_e in phase with Δω [10, 13]. In that case, the frequency response method is applied to adjust the phase. So the transfer function between ΔT_e and Δω with and with PSS is derived from the SMIB model. In the above, torque and power have been used as substitute to each other; since, in per unit system, they are very nearly equal [7].

1) *System Transfer Function in Absence of PSS*: When PSS is not equipped with the model the transfer function between ΔT_e and Δω can be found as:

$$\Delta E_{FD} = -\left(\frac{K_E}{1+sT_E}\right)\left(k_5\Delta\delta + k_6\Delta E'_q\right) \quad (8)$$

$$\Delta E'_q = \frac{\left(\frac{k_3}{1+sk_3T'_{d0}}\right)\left(-\frac{K_Ek_5}{1+sT_E} - k_4\right)}{\left(1+\frac{k_3k_6K_E}{(1+sk_3T'_{d0})(1+sT_E)}\right)}\Delta\delta \quad (9)$$

So the close loop transfer function from Δω to ΔT_e in absence of PSS is:

$$G_0(s) = \frac{\Delta T_e}{\Delta\omega};$$

$$G_0(s) = \frac{\omega_0}{s}\left(k_1 - \frac{\left(\frac{k_2k_3}{1+sk_3T'_{d0}}\right)\left(\frac{K_Ek_5}{1+sT_E} + k_4\right)}{\left(1+\frac{k_3k_6K_E}{(1+sk_3T'_{d0})(1+sT_E)}\right)}\right) \quad (10)$$

2) *System Transfer Function in presence of PSS:* The relation between ΔT_e and $\Delta\omega$ in presence of PSS is:

$$\Delta E_{FD} = \left(\frac{K_E}{1+sT_E}\right)\left(\Delta E_{PSS} - k_5\Delta\delta - k_6\Delta E'_q\right) \quad (11)$$

$$\Delta E'_q = \frac{\left(\frac{k_3}{1+sk_3T'_{d0}}\right)\left(\frac{K_E}{1+sT_E}\right)PSS}{\left(1+\frac{k_3k_6K_E}{(1+sk_3T'_{d0})(1+sT_E)}\right)}\Delta\omega - \frac{\left(\frac{k_3}{1+sk_3T'_{d0}}\right)\left(\frac{k_5K_E}{1+sT_E}+k_4\right)\frac{\omega_0}{s}}{\left(1+\frac{k_3k_6K_E}{(1+sk_3T'_{d0})(1+sT_E)}\right)}\Delta\omega \quad (12)$$

So, in presence of PSS the transfer function becomes

$$G_{PSS}(s) = k_1\frac{\omega_0}{s} + \frac{\left(\frac{k_3}{1+sk_3T'_{d0}}\right)\left(\frac{K_E}{1+sT_E}\right)PSS}{\left(1+\frac{k_3k_6K_E}{(1+sk_3T'_{d0})(1+sT_E)}\right)} - \frac{\left(\frac{k_3}{1+sk_3T'_{d0}}\right)\left(\frac{k_5K_E}{1+sT_E}+k_4\right)\frac{\omega_0}{s}}{\left(1+\frac{k_3k_6K_E}{(1+sk_3T'_{d0})(1+sT_E)}\right)} \quad (13)$$

The phase of the Bode diagram will be considered based on the above transfer function between ΔT_e and $\Delta\omega$ with or without PSS. The proposed tuning method will be described as follows keeping ΔT_e in phase with $\Delta\omega$.

IV. Tuning Methodology and Results

The tuning method applied for PSS parameters in this study is based on the frequency response. As described earlier that the change in electrical torque must be in phase with speed change. Hence from the transfer function above between ΔT_e and $\Delta\omega$, either with or without PSS, we get Bode phase plot.

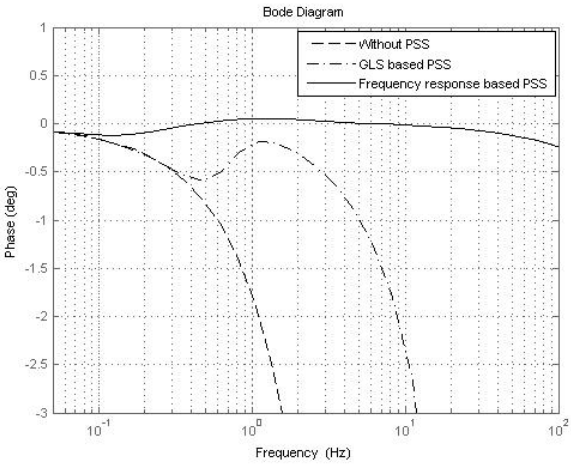


Fig. 4. Frequency response of the system.

As it is not possible to arrange for the phase to be zero for all frequencies, so concentration must give on the frequency range where most of the power oscillations exist, between 0.1 and 3 Hz [13]. For the proper stabilization, the phase lag between ΔT_e and $\Delta\omega$ in this frequency range must be less than 90 degree [15]. Although in Fig. 4, without PSS the phase lag between them is less than 90 degree in this frequency range, the phase lag, for optimal tuning, is further studied to reduce it nearly to zero. So the value of gain and lead-lag compensator parameter by using frequency response is optimized. Initially, during tuning process, the value of $T_1 = T_3 = 0.01$ sec, $T_2 = T_4 = 0.05$ sec and $T_w = 10$ sec were selected. The value of K_{PSS} was increased slowly starting from 1.0 with the change of T_1 and T_2. It is found that the lower value of K_{pss} provides negligible effect on the tuning. K_{PSS} was increased until leading phase is appeared in the above frequency range. Then T_1 was increased and T_2 was decreased simultaneously to make the lagging phase between ΔT_e and $\Delta\omega$ near to zero. The parameters of the system used in this study for simulation has been taken from elsewhere where the approach was to tune the PSS parameter using genetic local search algorithm [16]. From the Bode plot in the frequency range from 0.1 to 3 Hz, the GLS based PSS shows the phase lag very close to that for without PSS. But after tuning by the present approach described above, the phase lag in 0.5 Hz is reduced near to zero, where the phase was - 0.75 degree without PSS and -0.6 degree with GLS based PSS.

The simulation results of the proposed tuning, for step disturbance of 3.5% and 5% in mechanical input, are shown in Fig. 5 and Fig. 6 for SMIB system as discussed above.

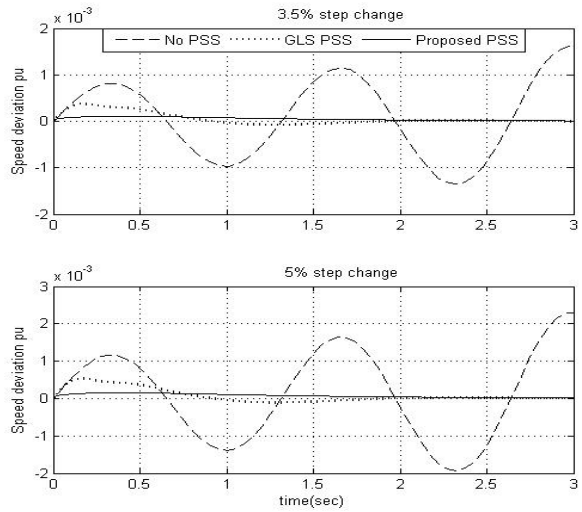


Fig. 5. Speed deviation $\Delta\omega$ vs time for disturbance ΔT_m=3.5% & 5%.

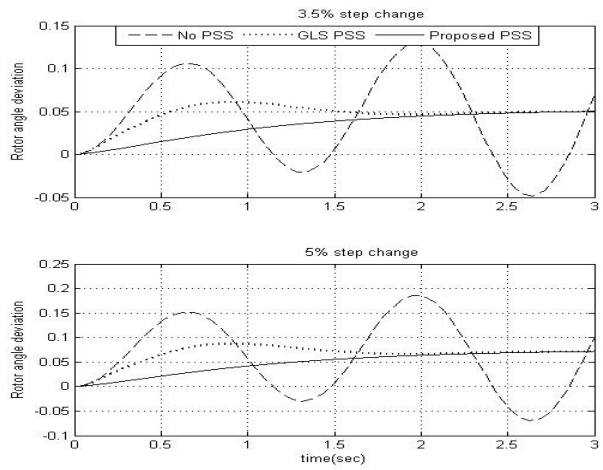


Fig. 6. Rotor angle deviation $\Delta\delta$ vs time for disturbance ΔT_m=3.5% & 5%.

If no PSS is equipped with the excitation system, the speed deviation of the rotor following a step disturbance will continue with increasing amplitude as depicted in Fig. 5. The GLS based PSS shows adequate damping but this is not optimum for perfect stabilization. The more damping, as shown by solid line in Fig. 6, is created after tuning the PSS parameters by frequency response and the deviation is reduced significantly. After the tuning, the rotor angle deviation is also damped out more quickly than GLS based PSS. In Fig. 6, it is shown that rotor angle starts to be stable after 2.5 seconds without any overshoot which justifies the effectiveness of the proposed tuning method.

V. Conclusions

In this paper, a design scheme of power system stabilizer for a single machine infinite bus system has been developed. The PSS parameters are tuned as per requirement while other system parameters remain the same. The proposed tuning method is based on the frequency response of the system. The transfer function between the electrical output torque change and corresponding speed deviation is derived. Thus, by decreasing the phase lag between them, the rotor angle oscillation is undermined. The system performance is observed under MATLAB environment. The result shows the effectiveness of the proposed tuning scheme.

Acknowledgment

We would like to thank Dr. A K M Alamgir and Mr. Mahmoad Hasan for their valuable suggestion during the preparation of this paper.

References

[1] S. Sheetekela, and K. A. Folly, "Multimachine power system stabilizer design based on evolutionary algorithm," in *44th International Universities' Power Engineering Conference*, Glasgow, Scotland, 1-4 Sept. 2009.

[2] Z. Bao-Hui, L. Cheng, Z. Hao, A. Klimek, B. Zhiqian, Y. Deng, G. Yao, and J. Shu, "Study on security protection of transmission section to prevent cascading tripping and its key technologies," in *IEEE/PES Transmission and Distribution Conference and Exposition: Latin America,* Bogota, pp. 1-13, 13-15 Aug. 2008.

[3] B. Tang, M. L. McAnelly, and G. Guoxiang, "Experimental parameter tuning of power system stabilizer," in *24th Chinese Control and Decision Conference (CCDC),* Taiyuan, China, 23-25 May 2012.

[4] P. Mitra, C. Yan, L. Grant, G. K. Venayagamoorthy, and K. A. Folly, "Comparative study of population based techniques for power system stabilizer design," in *15th International Conference on Intelligent System Applications to Power Systems (ISAP)*, Curitiba, Brazil, 8-12 Nov. 2009.

[5] D. Menniti, A. Burgio, A. Pinnarelli, and N. Sorrentino. "Synchronizing fuzzy power system stabilizer and fuzzy FACTS device stabilizer to damp electromechanical oscillations in a multi-machine power system," in *IEEE Power Tech Conference Proceeding,* Bologna, 23-26 Jun. 2003.

[6] P. Kundur, et al. "Definition and classification of power system stability IEEE/CIGRE joint task force on stability terms and definitions," *IEEE Transaction on Power System*, vol. 19, pp. 1387-1401, Aug. 2004.

[7] P. Kundur, *Power System Stability and Control*, Eds. N. J. Balu, and M. G. Lauby. Vol. 7, New York: McGraw-hill, 1994.

[8] P.M.Anderson and A.A.Fouad, *Power System Control and Stability*, Institute of Electrical and Electronics Engineers, Inc. 2003.

[9] R. Gupta, B. Bandyopadhyay, and A. M. Kulkarni. "Design of decentralized power system stabilizers for multimachine power system using model reduction and fast output sampling techniques," in *5th Asian Control Conference*, Melbourne, Victoria, Australia, 20-23 Jul. 2004.

[10] P. Kundur, et al. "Application of power system stabilizers for enhancement of overall system stability," *IEEE Transactions on Power Systems*, vol. 4, pp. 614-626, May 1989.

[11] P. P. Kumar, M. R. Babu, and Saraswathi, "Dynamic analysis of single machine infinite bus system using single input and dual input PSS," *International Electrical Engineering Journal*, vol. 3, pp. 632-641, 2012.

[12] O. M. Awed-Badeeb, "Damping of electromechanical modes using power system stabilizers (PSS) Case: Electrical Yemeni network," *Journal of Electrical Engineering*, vol. 57, pp. 291-295, 2006.

[13] A. Murdoch, S. Venkataraman, R. A. Lawson, and W. R. Pearson "Integral of accelerating power type PSS. I. theory, design, and tuning methodology," *IEEE Transactions on Energy Conversion*, vol. 14, pp. 1658-1663, Dec. 1999.

[14] F. P. Demello, and C. Concordia. "Concepts of synchronous machine stability as affected by excitation control," *IEEE Transactions on Power Apparatus and Systems*, vol. 88, pp. 316-329, Apr. 1969.

[15] E. V. Larsen, and D. A. Swann. "Applying power system stabilizers part III: practical considerations," *IEEE Transactions on Power Apparatus and Systems,* vol. 100, pp. 3034-3046, Jun. 1981.

[16] M. Rashidi, and F. Rashidi. "Power system stabilizer tuning using multi-objective genetic algorithm," *International Review of Electrical Engineering*, vol. 6, pp. 356-363, Jul.-Aug. 2011.

Appendix

Generator parameters:

$x_d = 0.973$ p.u.; $x'_d = 0.19$ p.u.; $x_q = 0.55$ p.u.; $v_{t0} = 1.00$ p.u.; $\omega_0 = 314$ rad/sec; $T'_{d0} = 7.765$ sec;

Exciter parameters:

$K_e = 50$, $T_e = 0.05$ sec;

Transmission line parameters:

$r_e = 0$ p.u.; $x_e = 0.997$ p.u.;

PSS parameters:

$T_w = 10$ sec;

PSS	K_{PSS}	$T_1=T_3$	$T_2=T_4$
GLS based PSS	35.1469	0.2078	0.05
Frequency Response based PSS	200	0.15	0.002

Calculation of Heffron-Phillips constants:

$$k_1 = \frac{x_q - x'_d}{x_e + x'_d} i_{q0} E_0 \sin\delta_0 + \frac{E_{q0} E_0 \cos\delta_0}{x_e + x_q};$$

$$k_2 = \frac{E_0 \sin\delta_0}{x_e + x'_q}; \; k_3 = \frac{x_e + x'_d}{x_e + x_d}; k_4 = \frac{x_d - x'_d}{x_e + x'_d} E_0 \sin\delta_0;$$

$$k_5 = \frac{x_q}{x_e + x_q}\frac{e_{d0}}{e_{t0}} E_0 \cos\delta_0 + \frac{x'_d}{x_e + x'_d}\frac{e_{q0}}{e_{t0}} E_0 \sin\delta_0;$$

$$k_6 = \frac{x_e}{x_e + x'_d}\frac{e_{q0}}{e_{t0}};$$

The parameters i_{q0}, e_{d0}, e_{q0}, i_{d0}, E_{q0}, E_0 and δ_0 can be calculated from the following equations.

$$i_{q0} = \frac{P_0 e_{t0}}{\sqrt{(P_0 x_{q0})^2 + ({e_{t0}}^2 + Q_0 x_q)^2}};$$

$$e_{d0} = i_{q0} x_q; \quad e_{q0} = \sqrt{{e_{t0}}^2 - {e_{d0}}^2}; \quad i_{d0} = \frac{Q_0 + {i_{q0}}^2 x_q}{e_{q0}};$$

$$E_0 = \sqrt{(e_{d0} + i_{q0} x_e)^2 + (e_{q0} - i_{d0} x_e)^2};$$

$$E_{q0} = e_{q0} + i_{d0} x_q; \quad \delta_0 = \tan^{-1}\left(\frac{e_{d0} + i_{q0} x_e}{e_{q0} - i_{d0} x_e}\right).$$

Transient Performance Improvement of Power System by Optimal Design of SVC Controller Employing Genetic Algorithm

Md. Shafiullah[1*], Md. Shafiul Alam[2], Md. Ismail Hossain[3], Md Nazmul Hasan[4]

[1,2]Dept. of EE, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia
[3,4]Dept. of EEE, International Islamic University Chittagong, Dhaka, Bangladesh
[*]shafiullah@kfupm.edu.sa

Abstract- **With the increase of loading of long transmission line, transient stability of power system on the occurrence of a fault become one of the major concern for power engineers. Instability problem of power system on occurrence of the fault can be solved with the use of Static VAR Compensator (SVC) controller. Assessing the transient stability of a large inter connected power network very much nonlinear and tremendously complicated. To mitigate this complication and nonlinearity, a substantial research with SVC controller has been done in the power industry. To improve the system stability for dynamic and transient disturbances SVC plays a very important role, which is a shunt connected FACTS devices. In this work, genetic algorithm based optimal design of SVC controller to improve the transient performance of power system is proposed. The parameters of the damping controller by optimal selection of damping ratio of the critical eigenvalue employing genetic algorithm are tuned. The effectiveness of the developed optimal controller is investigated by simulating different abnormal conditions of a single machine infinite bus (SMIB) system using MATLAB based PST.**

Index Terms: **Controller, FACTS, GA, PST, SMIB, SVC.**

NOMENCLATURE

δ = rotor angle, ω= angular frequency, e'_q= generator internal voltage, ΔE_{fd} = field excitation, P_m= Input mechanical power, P_e =Output electrical power, P_D = Damping Coefficient, M= System inertia coefficient, ω_b = base angular frequency, x_d, x_q = Direct and quadrature axes reactance, x'_d, x'_q = Direct and quadrature axes transient reactance, i_d , i_q = Direct and quadrature axes currents, v_d , v_q = Direct and quadrature axes voltages, v_t, v_{tr} = Terminal and reference voltages, u_E = Control input, T_w = Time constant of the washout block, T_F = Time constant of the limiter block, T_E = Time constant of the excitation system, T'_{do} = Open circuit field time constant, K_1-K_6 = Fourth order model constants, K_C= Gain of the compensator, K_E = Gain of the excitation system

I. INTRODUCTION

Power system network is the largest and the most complex system ever devised by man, which consist of thousands of buses and hundreds transformers, generators, capacitor banks, cables and lines. The power export import arrangements are done through transmission networks but due to the growth of demand on such large system the capacity of the transmission networks have to be enhanced by adding new parallel lines. The construction of new transmission line requires significant amount of time as well as it requires huge amount of investment, consequently the utility planners look forward to utilize full capacity of the existing resources and transmission lines but getting maximum power transfer through the existing transmission networks with minimum price has become one of the utmost challenges for today's transmission system operators. In meeting this challenge, early researchers explored the possibility of using series compensation of transmission line. In [1], R.E. Marbury et. al. introduced series capacitor as an element connected in series with the line to improve the voltage by interleaving negative reactance whereas in [2] several methods of applying series capacitor are discussed. The concepts of thyristor assisted FACTS was first introduced by N. G. Hingorani [3] to reduce transmission cost. To facilitate initialization of the steady state parameters a new linear continuous TCSC model in state-space form was derived [4]. A hybrid power system having both TCSC and series capacitors, a SDC is proposed by X. Zheng, Z. Xu, and J. Zhang [5] to mitigate sub synchronous resonance (SSR) caused by the series capacitors. The proposed damping controller is composed of a band pass filter, a gain, two or more lead-lag blocks and a limiter. The authors used the deviation of the generator speed as the input signal than that of bus voltage magnitude, line current, active power etc. as reported in [6, 7]. A. A. Edris proposed VAR compensator to enhance load-ability of low capacity power system in [8].

Power transmission capacity and system stability improvement by Static VAR Compensator (SVC) have been presented in [9] where with the help of voltage control loop the bus voltage has been controlled. The voltage control loop is comprised of thyristor controlled reactor (TCR), voltage regulator, voltage magnitude transducer and thyristor switched capacitor (TSC) firing circuit. In real time operation there are uncertainties in power system reactance of the transmission path and SVC regulator time constant so the retuning the SVC controller is necessary when these uncertainties are large enough besides those uncertainties power system parameters and quantities are changed due to the change in system structure such as load addition, generation outage over time. In order to minimize above difficulties, a new nonlinear robust adaptive SVC controller which can deal with both constant and time varying systems is proposed in [10]. R. Griinbaum in [11] presented several benefits of FACTS devices in power transmission systems, TCSC and SVC are shown to be very vigorous and effective means to enhancing damping for long distance transmission system.

As the demand of energy is increasing every day and to

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

meet the new demands some portion of the transmission network becomes overloaded as well as because some disturbances the power system become unstable, the transient stability refers the ability of a power system to maintain synchronism through finding a new equilibrium point (operating point) subjected to a large disturbance [12] has become a matter of fact at the power system industry. The improvement of transient stability using UPFC and SVC are reported in [13] whereas Abedin and Malik reported in [14], with the help of dynamic action the SVC can offer damping and can enhance overall power system stability. Low frequency oscillation of power system can be well damped employing SVC which has been discussed [15] whereas in [16] a novel genetic algorithm based PI controller has been designed to improve system stability. PSO algorithm is effectively utilized, here by synchronized tuning of synergetic excitation damping controller and SVC based damping controllers the power system stability has been improved which has been addressed in [17] where as a supplementary controller a lead - lag compensator is employed for the SVC. A novel optimization technique for simultaneous coordination between PSS and SVC for multimachine power system as a damping controller to damp the oscillation is reported in [18] where chaos theory and particle swarm optimization technique are hybridized to form a chaotic particle swarm optimization (CPSO). In [19] a popular meta-heuristic optimization technique usually known as Firefly Algorithm (FA) is addressed to design SVC controller in multi machine power system. With the coordination PSS and SVC the enhancement of power system stability is investigated in [20].

This work proposes a genetic algorithm based SVC controller to damp system oscillation to improve transient stability upon occurrence of fault and the developed optimal controller is applied to a SIMB to validate by simulating abnormal conditions using MATLAB based PST.

II. POWER SYSTEM MODELING

A. Synchronous Generator

In this work a SMIB system has been taken and an SVC has been set up in transmission network as shown in fig. 1. A local load with admittance $Y_L = g + jb$ is connected to the bus of generator and the transmission line has two impedances for two sections named $Z_1 = R_1+jX_1$ and $Z_2 = R_2+jX_2$ which has been shown in the following figure.

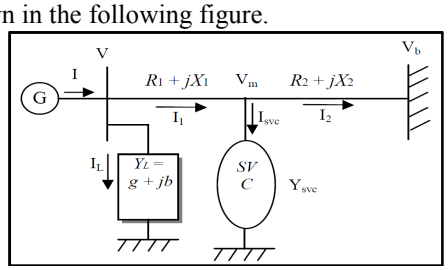


Fig.1 SVC connected Single Machine Infinite Bus System

The fourth order model of the generator has been considered which comprising of the generator internal voltage (produced by generated EMF) equation, IEEE standard type excitation system as well as the electromechanical swing equation. Equations (1) and (2) show the electromechanical swing equation. Internal voltage produced by generated EMF can be represented by equation (3) and the regulator exciter system is represented as in IEEE type ST in equation (4):

$$\dot{\delta} = \omega_b(\omega - 1) \quad (1)$$

$$\dot{\omega} = \frac{1}{M}(P_m - P_e - P_D) \quad (2)$$

$$\dot{e'_q} = \frac{1}{T'_{d0}}\left[E_{fd} - e'_q - (x_d - x'_d)i_d\right] \quad (3)$$

$$\Delta E_{fd} = \frac{k_E}{1+sT_E}\left[v_{tr} - v_t + u_E\right] \quad (4)$$

The terminal voltage can be expressed in terms of direct axis and quadrature axis voltage which is shown in equation (5). Equations (6) and (7) show the expressions for direct axis and quadrature axis voltages respectively.

$$v_t = (v_d + jv_q) = \sqrt{(v_d^2 + v_q^2)} \quad (5)$$

$$v_d = x'_q i_q \quad (6)$$

$$v_q = -x'_d i_d + e'_q \quad (7)$$

Output power of the generator is expressed in terms of quadrature and direct axes quantities as shown in equation (8)

$$P_e = v_d i_d + v_q i_q \quad (8)$$

System data for the proposed model of fig.1 is given below in table I

Table I: System Data

Line Data (pu)	Generator Data (pu)
$X_1 = 0.4985$	$x_q = 0.55$
$X_2 = 0.4985$	$x_d' = 0.19$
$R_1 = 0.017$	$x_d = 0.973$
$R_2 = 0.017$	D=0
b= 0.262	M=9.26
g= 0.249	$T_{do}' = 7.76$

B. SVC based Stabilizer

To control the reactive power injection of a chosen bus a special type of shunt compensation device is used in transmission network which is commonly known as static VAR compensator (SVC). By regulating the reactive power injection SVC enhances steady state power transmission limit, hence it improves the stability (transient) of the power system [21]. The power swing damping controller [9] is shown in Fig. 2. To provide the compensation two or more lead-lag blocks are used and to produce modulation signal the output of the compensator is passed through first order lag which is commonly known as a limiter. To deactivate the controller in steady state operation a washout/rest block is connected. Here the input signal is the change in generator angular frequency and the output is the state x_8.

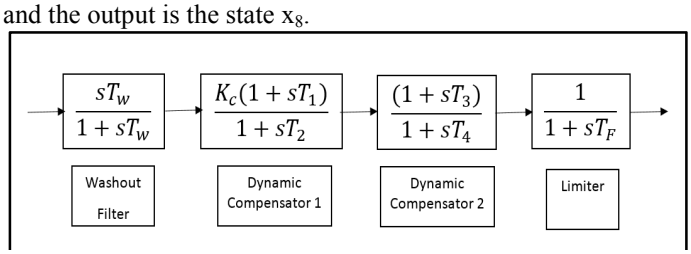


Fig. 2 SVC power swing damping controller

C. System Linearized Model

The incremental linearized model around an operating point is usually employed in the design of damping controllers but the fourth order generator model represented by equations (1)-(4) are nonlinear, so we have to linearize the equations to find

$$\dot{x} = f(x, u_E) \quad (9)$$

$$x \text{ are the states and } u_E \text{ is the input}$$

For the purpose of linearization first of all, we have to linearize the direct and quadrature axes currents [22] then substituting in equations (1)-(7) to have the following linearized equations.

$$\Delta\dot{\omega} = a_{11}\Delta\omega + a_{12}\,\Delta\delta + a_{13}\,\Delta e_q' \quad (10)$$

$$\Delta\dot{\delta} = a_{21}\Delta\omega \quad (11)$$

$$\Delta\dot{e_q'} = a_{32}\,\Delta\delta + a_{33}\Delta e_q' + a_{34}\,\Delta E_{fd} \quad (12)$$

$$\Delta\dot{E}_{fd} = a_{42}\,\Delta\delta + a_{43}\,\Delta e_q' + a_{44}\,\Delta E_{fd} + a_{48}\,x_8 \quad (13)$$

$$\dot{x_5} = a_{51}\,\Delta\omega + a_{52}\,\Delta\delta + a_{53}\,\Delta e_q' + a_{55}\,x_5 \quad (14)$$

$$\dot{x_6} = a_{61}\Delta\omega + a_{62}\Delta\delta + a_{63}\,\Delta e_q' + + a_{65}x_5 + a_{66}\,x_6 \quad (15)$$

$$\dot{x_7} = a_{71}\,\Delta\omega + a_{72}\,\Delta\delta + a_{73}\,\Delta e_q' + a_{75}x_5 + a_{76}\,x_6 + a_{77}\,x_7 \quad (16)$$

$$\dot{x_8} = a_{87}\,x_7 + a_{88}\,x_8 \quad (17)$$

III. SVC PARAMETERS TUNING USING GA

Genetic algorithms are probabilistic search methods and it is based on the ideas of evolutionary processes. It is a global search algorithm that is founded on the concept of natural genetics and the Darwinian survival-of-the-fittest code [23]. The GA based SVC controller parameters found using GA are shown in Table II. The Eigen values of the disturbed system and the system with damped controller are presented in Table III. The Eigen values in the left half plane shows the system stability with SVC.

Table II: Optimal parameters found GA based SVC

Gain (K_c)	Lead time (T_1=T_3)	Lag time (T_2=T_4)
90.4278	2.0686	0.5175

Table III: Eigen Values associated with fourth order linear model and with SVC based controller

Fourth Order Linear Model	With SVC Controller
-204	-204.01
-.36 + 5.70i	-3.01 + 7.12i
-0.36 - 5.70i	-3.01 - 7.12i
4.52	-6.74
--	-0.42 + 1.23i
--	-0.42 - 1.23i
--	-0.45 +.01i
--	-0.45 -.01i

IV. SIMULATION TOOL

For the simulation purpose MATLAB and MATLAB based Power System Toolbox (PST) has been used. To find the optimal values SVC parameters genetic algorithm has been coded in MATLAB environment. These optimal values are used to simulate a three phase fault in PST. The details information about the tool are available in [24].

V. SIMULATION RESULTS

Transient analysis of a SIMB system has been performed without and with GA based SVC controller to show positive effect of damping controller on damping power system oscillation. A three phase disturbance has been applied at t=0.3 second at infinite bus with subsequent tripping of line. Improvement in transient performance of the system is observed in time domain simulation with GA based SVC controller than that of without controller. The maximum level of compensation is depicted in fig. 3 provided by the SVC in the after fault scenario and then it decreases gradually. Bus voltage behavior with and without GA based SVC is presented in fig. 4. The figure shows after experiencing an oscillation of few seconds the bus voltage become flattened with a GA based optimal controller whereas for uncontrolled system oscillation stays live.

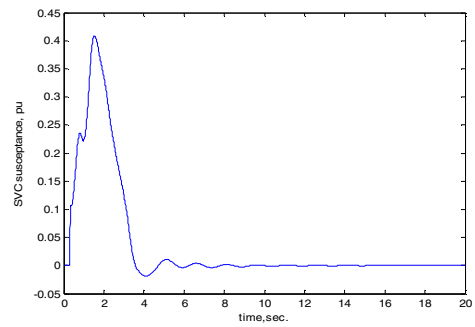


Fig. 3Variation of SVC susceptance with respect to time

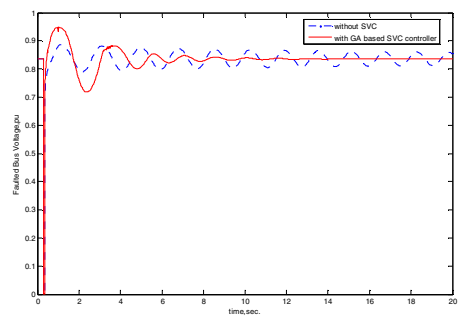


Fig. 4 Comparison of voltage profile with and without GA based SVC controller

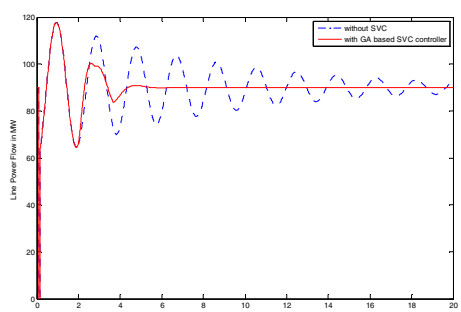


Fig. 5 The line power flow

Line power flow settles to its steady state value after a few seconds of oscillations with a GA based SVC damping controller; however, without SVC controller line power oscillates throughout the entire simulation period, which is not desired in power system operation at all. Line power flow behavior is shown in fig. 5. Oscillation in machine angle is well damped with GA based SVC damping controller parameters is depicted in fig. 6. The figure clearly depicts that system is stable with optimal controller. Finally, the fig. 7 shows the effectiveness of GA based SVC damping controller to damp machine speed deviation. Oscillation in machine speed deviation is observed during the entire simulation period without designed SVC. However, GA based SVC damping controller damps machine speed deviation and shows stable operation.

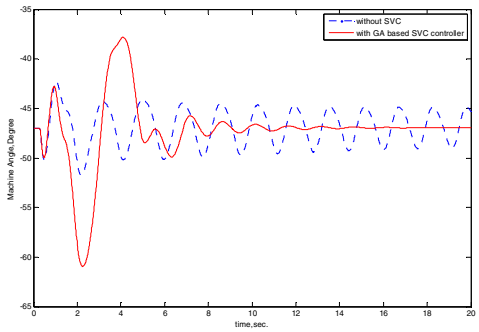


Fig. 6 Comparison of oscillation in machine angle

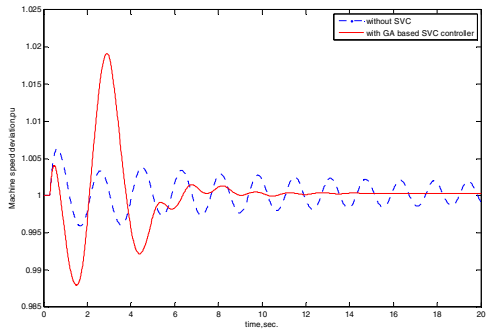


Fig. 7 Comparison of angular speed deviation

VI. CONCLUSIONS

In this work, a GA based SVC controller is designed for damping oscillations in power system. The proposed controller has been applied to SMIB system. Simulation has been performed in MATLAB based PST software and simulation results show the effectiveness of the designed controller. As at severe contingencies and stressed operating conditions the mathematical model of a power system is usually involved with a huge set of nonlinear equations, to ensure improved dynamic performance of the system the controller could be designed to reflect the nonlinearity.

VII. APPENDIX

The parameters used in equations (10)-(17) are given below:

$$a_{11} = -\frac{D}{M}, a_{12} = -\frac{K_1}{M}, \quad a_{13} = -\frac{K_2}{M}, \qquad a_{21} = \omega_0,$$

$$a_{32} = -\frac{K_4}{T'_{d0}}, \; a_{33} = -\frac{1}{K_3 T'_{d0}}, \; a_{34} = -\frac{1}{T'_{d0}}, a_{42} = -\frac{K_E K_5}{T_E},$$

$$a_{43} = -\frac{K_E K_6}{T_E}, \; a_{44} = -\frac{1}{T_E}, \; a_{48} = \frac{K_E}{T_E}, a_{51} = -\frac{D}{M},$$

$$a_{52} = -\frac{K_1}{M}, \; a_{53} = -\frac{K_2}{M}, a_{55} = -\frac{1}{T_w}, a_{87} = \frac{1}{T_F}$$

$$a_{61} = -\frac{DT_1K_c}{MT_2}, \; a_{62} = -\frac{T_1K_1K_c}{MT_2}, \; a_{63} = -\frac{T_1K_2K_c}{MT_2}$$

$$a_{65} = \left(\frac{K_c}{T_2} - \frac{T_1K_c}{T_2T_w}\right), \; a_{66} = -\frac{1}{T_2}, \qquad a_{71} = = -\frac{DT_1T_3K_c}{MT_2T_4}$$

$$a_{72} = -\frac{T_1T_3K_1K_c}{MT_2T_4}, \; a_{73} = -\frac{T_1T_3K_2K_c}{MT_2T_4}, a_{77} = -\frac{1}{T_4},$$

$$a_{75} = \frac{T_3}{T_4}\left(\frac{K_c}{T_2} - \frac{T_1K_c}{T_2T_w}\right), \; a_{76} = \left(\frac{1}{T_4} - \frac{T_3}{T_2T_4}\right), a_{88} = -\frac{1}{T_F}$$

REFERENCES

[1] R. E. Murbury, J. B. Owens, "*New Series Capacitor Protective Devices*," AIEE Transactions, March 1946, Vol. 65, pp. 142-145.

[2] S. B. Crary, L. E. Saline, "*Location of Series Capacitors in High Voltage Transmission System*," AIEE transactions, December 1953.

[3] N. G. Hingorani, "*High Power Electronics and Flexible AC Transmission System*," IEEE Power Engineering Review, July 1998.

[4] D. Jovcic, and G. N. Pillai, "*Analytical Modeling of TCSC Dynamics*," IEEE Transactions on Power Delivery, April 2005, Vol. 20, No. 2, pp. 1097-1104.

[5] X. Zheng, Z. Xu, J. Zhang, "*A Supplementary Damping Controller of TCSC for Mitigating SSR*," IEEE Power and Energy Society General Meeting, 2009, pp. 1-5 .

[6] J. X. Ling, Z. J. Guo, W. H. Feng, L. K. Jun, N. Lin, P. H. Chao, "*Design of Adaptive TCSC Damping Controller Based on Oscillation Energy Analysis*," IEEE/PES Transmission and Distribution Conference and Exhibition: Asia and Pacific Dalian, China, 2005, pp. 1-5.

[7] Q. Liu, V. Vittal, N. Elia, "*LPV Supplementary Damping Controller Design for a Thyristor Controlled Series Capacitor (TCSC) Device*," IEEE Transactions on Power Systems, August 2006, Vol. 21, No. 3, pp. 1242-1249.

[8] A. A. Edris, "*Controllable Var Compensator: A Potential Solution to Loadability Problem of Low Capacity Power Systems*," IEEE Transactions on Power Systems, August 1987, Vol. 2, No. 3, pp. 561-567.

[9] E. V. Larsen, J. H. Chow, "*SVC Control Concepts for System Dynamic Performance*," in IEEE Tutorial Course: Application of Static Var Systems for System Dynamic Performance, 87TH0187-5-PWR, 1987, pp. 36-53.

[10] R. Yan, Z. Y. Dong, T. K. Saha, J. Ma, *Nonlinear Robust Adaptive SVC Controller Design For Power Systems*, IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century, 2008, pp. 1-7.

[11] R. Griinbaum, M. D. Grijp, V. Moshi, "*Enabling Long Distance AC Power Transmission by Means of FACTS*," IEEE AFRICON, 23-25 September 2009, Nairobi, Kenya.

[12] K. R. Padiyar, *"Power System Dynamics: Stability and Control,"* Second Edition, BS Publications, Hyderabad, 2002.

[13] S. N. Siva, S. Ravik, *"Transient Stability Improvement Using UPFC and SVC"* ARPN Journal of Engineering and Applied Sciences Vol. 2, No. 3, June 2007

[14] T. Abedin, O. P. Malik, *"Intelligent SVC Control for Transient Stability Enhancement,"* IEEE Power Engineering Society General Meeting, August 2005, Vol. 2, pp. 1701-1707

[15] T. Abedin, O. P. Malik, *"Genetic Algorithm Based Simultaneous Coordination of PSS and FACTS Controllers for Power Oscillations Damping"*, IEEE ICSET 2012, pp. 85-90

[16] Y.P. WANG, D.R. HUR, N.R.WATSON, J. ARRILLAGA, S.S. MATAIR"A Genetic Algoritms Approach To Design An Optimal Pi Controller For Static Var Compensator" IEEE 2000, International Conference on Volume3, Issue , 2000 Page(s):1557-1562 vol.3 .

[17] T. Ademoye, A. Feliachi, A. Karimi, *"Coordination of Synergetic Excitation Controller and SVC Damping Controller Using Particle Swarm Optimization."* Proceedings of 2011 IEEE Power and Energy Society General Meeting, 24-29 July 2011, pp. 1 – 8, Detroit, MI, USA.

[18] M. Eslami, H. Shareef, A. Mohamed and M. Khajehzadeh, *"Coordinated Design of PSS and SVC Damping Controller Using CPSO."* 5th International Power Engineering and Optimization Conference (PEOCO), 6-7 June 2011, pp. 11 – 16, Shah Alam, Malaysia.

[19] N. N. Islam, M. A. Hannan, H. Shareef and A. Mohamed, *"SVC Damping Controller Design Based on Firefly Optimization Algorithm in Multi Machine Power System."* In 2013 IEEE Conference on Clean Energy and Technology (CEAT), 18-20 Nov. 2013, pp. 66 – 70, Lankgkawi Malaysia.

[20] H. Fallahi, M. R. Aghamohammadi, A. Parizad, A. Mohamadi, *"Enhancing Power System Oscillation Damping Using Coordination between PSS and SVC"* proc. IEEE EPECS, pp. 1-7, 2009.

[21] L. Gyugyi, "*Power Electronics in Electric Utilities: Static Var Compensators*," Proc. Of IEEE, April 1988, Vol. 76, No. 4, pp. 383-394.

[22] M. A. Abido, *"Design of PSS and STATCOM – Based Damping Stabilizers Using Genetic Algorithms"* Power Engineering Society General Meeting, IEEE 18-22 June 2006, pp.8.

[23] S. Rajasekaran, G. A. Vijayalakshmi Pai *"Neural Networks, Fuzzy Logic and Genetic Algorithms: Synthesis and applications"* PHI Learning Pvt. Ltd., 2006.

[24] J. H. Chow, G. Rogers, *Power System Toolbox Version 3.0, available at http://www.ecse.rpi.edu/pst/PST.html*

Analytical Approach to Investigate the Stability of a System with Simultaneous AC-DC Power Flow through Existing AC Transmission Line

Mohammad Tawhidul Alam, [1,*] and Q. Ahsan[2]

[1]Department of Electrical and Electronic Engineering
Ahsanullah University of Science and Technology, Dhaka, Bangladesh
[2]Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology, Dhaka, Bangladesh
[*]*Email: tawhidul_bclt@yahoo.com*

***Abstract*—The simultaneous ac-dc power transmission system can improve the transient stability of a power system. A mathematical model for ac-dc composite system, suitable for stability analysis, is developed in this paper. To compare the ac transmission system with the composite system a comprehensive analysis in terms of stability is made using the proposed model. A numerical analysis is also carried out to compare the stability of the composite system with that of a system with ac transmission only.**

***Index Terms*—AC-DC composite system, critical clearing angle, critical clearing time, stability improvement.**

NOMENCLATURE

P_l	Steady state power flow through unconverted ac line in p.u
E_g	Internal generated voltage of the generator in p.u
E_r	Receiving end voltage in p.u
X	Transfer reactance of unconverted ac system in p.u
I_{th}	Thermal limit of the conductor
I_L	Steady state current of unconverted line
δ	Load angle of the generator for unconverted system
H	Inertia constant
ω_s	Angular frequency
δ_0	Prefault steady state load angle for unconverted system
δ_{cr}	Critical clearing angle for unconverted system
T_{cr}	Critical clearing time for unconverted system
P_{comb}	Combined steady state ac-dc power flow
I_{dc}	DC component of the composite system line current
I_{AC}	AC component of the composite system line current
δ_{ac}	Rotor angle due to ac power transfer
δ_{dc}	Rotor angle due to dc power transfer
δ_{CR}	Critical clearing angle in converted system
P_{dcmax}	Maximum transmissible dc power
P_{acmax}	Maximum ac power in ac-dc composite system
P_{emrg}	Maximum combined ac-dc power that can be transfer at any emergency period

I. INTRODUCTION

Long EHV ac transmission line has a limitation of being loaded heavily because of transient stability problem. Flexible AC Transmission System (FACTS) can be used in existing EHV ac transmission line to improve stability [1-2]. A combined system like EHV ac and HVDC system can be operated in parallel for stability improvement [3]. A generic concept combining both transient angle and voltage stabilities of parallel ac-dc transmission is introduced in [4]. The authors of this paper clearly exposed that the utilization of inherent short term overload capacity of the dc system converters can increase the stability margin of ac system.

Another new approach, which is very much close to ac-dc parallel system, is proposed to increase stability of existing transmission line in [5]. In this proposed approach same transmission line carries ac power along with dc power simultaneously. The main benefit of simultaneous ac-dc power transmission is that it does not require separate HVDC transmission line with its right-of-way. The authors of [5] shows stability improvement through numerical simulation, considering transient fault at different location of transmission line but did not considered the impact of dc portion of power on the rotor angle of generating unit. K.P. Basu [6] proposed a methodology to improve the stability of simultaneous ac-dc system where the system is no more ac-dc system after clearing the fault until the system enters into the normal state. In this proposed methodology ac-dc composite system would be converted into dc system just after clearing the fault by tripping the ac circuit breaker and when the system becomes healthy ac breakers would be turned on and its need special control mechanism for circuit breaker operation. To date, all the works those have been done on stability analysis for simultaneous ac-dc transmission in the literature are based on numerical approach but no work has been found on the basis of analytical approach.

To investigate the behavior of a system comprehensively an appropriate mathematical model is required. Moreover, the

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

change of which parameter of the system impact how much on its performance by using the mathematical model someone can evaluate easily.

In this paper, a comprehensive analysis is presented on simultaneous ac-dc system by developing appropriate mathematical model for improving the transient stability of an existing long ac transmission line. A numerical analysis also performed for the validation of developed mathematical model.

II. Simultaneous AC-DC Power Flow

In simultaneous ac-dc power transmission system same conductor is used to transmit ac power along with dc power, the schematic diagram of simultaneous ac-dc flow is shown in fig. 1. The main concern of this paper is to convert existing long EHV ac line into ac-dc composite line. According to the figure, a remote generating unit is feeding to an infinite bus through a long transmission line. A portion of ac power, taking from the generator bus, converted into dc form and feeding to the existing ac transmission line through the neutral point of zigzag transformer.

At the receiving end, dc portion of power is taking away from the neutral point of zigzag transformer and converted into ac before feeding to the infinite bus. Zigzag connection is used at both ends of the transmission line to avoid magnetic saturation in the transformer core due to dc current.

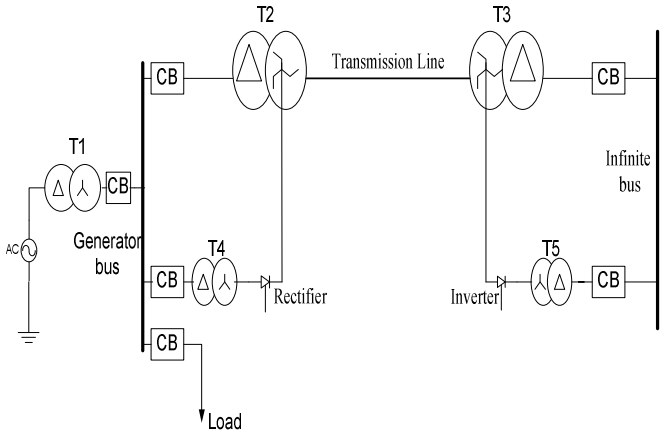


Fig. 1 AC-DC composite power transmission system

In converting ac line into ac-dc composite line the insulator remain unchanged and thus the maximum value of the phase voltages must be kept same. The output dc current of rectifier is equally divided among all the three phases and each conductor carries ac-dc modulated current wave. A communication link is needed between rectifier and inverter for better control of dc power flow.

III. Stability Model

In case of ac transmission line the power flow equation is given as follows.

$$P_l = \frac{E_g E_r}{X} sin\delta \quad (1)$$

$$X = X_{te} + X_{Tr} \quad (2)$$

Here,

X_{te} = Reactance that includes the reactance of generator and transformer in p.u.

X_{Tr} = Reactance of the transmission line in p.u.

For long ac transmission line minimum steady state stability margin is recommended as 30% and steady state power transfer angle would be 40^0 – 44^0 [7]. Although the thermal limit is qualified as line design and it depends on conductor selection but according to the loadability characteristics of a long transmission line the thermal limit is much higher than the steady state stability limit.

According to the equal area criteria of transient stability analysis of an ac system, the critical fault clearing angle and time for ac system are presented in (3) and (4) respectively.

$$\delta_{cr} = cos^{-1}[(\pi - 2\delta_0)sin\delta_0 - cos\delta_0] \quad (3)$$

$$T_{cr} = \sqrt{\frac{4H(\delta_{cr}-\delta_0)}{\omega_s P_l}} = \sqrt{(\delta_{cr} - \delta_0)}\sqrt{\frac{4H}{\omega_s P_l}} = \lambda_{ac}\sqrt{\frac{4H}{\omega_s P_l}} \quad (4)$$

Here

$\lambda_{ac} = \sqrt{(\delta_{cr} - \delta_0)}$

A generalize mathematical model for steady state power flow in ac-dc composite system is presented in (5)

$$P_{comb} = k\beta P_l + \sqrt{2}(1-k)\gamma P_l \quad (5)$$

Here

k = Percentage of ac voltage mix

1- k = Percentage of dc voltage mix

β = Ratio between I_{AC} and I_L

γ = Ratio between I_{dc} and I_L

The main advantage of ac-dc composite system is that in case of emergency period after clearing a transient fault huge energy can be transferred from sending end to the receiving end by increasing the dc power using converter control within a shorter period of time. In this case emergency short term over loading capacity of the converters can be utilized. The power transfer at the emergency period considering 50% short term over loading of the converters is as follows[4].

$$P_{emrg} = P_{acmax} \sin \delta_{ac} + P_{dcmax} \quad (6)$$

$$P_{dcmax} = \frac{3}{\sqrt{2}}(1-k)\gamma P_l \quad (7)$$

$$P_{emrg} = k\beta P_l + \frac{3}{\sqrt{2}}(1-k)\gamma P_l \quad (8)$$

After being subjected to a transient fault the critical clearing angle for the composite system can be derived as,

$$P_{dcmax}\delta_{CR} - P_{acmax}\cos(\delta_{CR} - \delta_{dcm}) = P_{dcmax}\delta_{max} + P_{acmax}\cos(\delta_{max} - \delta_{dcm}) - P_{comb}(\delta_{max} - \delta_0) \quad (9)$$

Here
δ_{CR} = Critical fault clearing angle for composite system.
δ_{dcm} = Rotor angle due to emergency dc power
δ_0 = Rotor angle due to steady state combined power
$\delta_{max} = \pi + \delta_{dc} - \delta_{ac}$

For a pure ac system it is easier to establish the simple expression for critical clearing angle that is presented in (3) but it is not feasible in composite ac-dc system by solving the equation (9). In this case several numerical techniques can be applied to evaluate the critical clearing angle. Nevertheless, a simple expression for Critical fault clearing angle is presented in (10) considering some approximation for evaluating the ac part of the combined emergency power.

$$\delta_{CR} = \delta_{max} - \frac{P_{comb}}{P_{emrg}}(\delta_{max} - \delta_0) \quad (10)$$

$$P_{emrg} = \frac{1}{2}(\beta_m + \beta_s)kP_l + \frac{3}{\sqrt{2}}(1-k)\gamma P_l \quad (11)$$

Here
B_m = The value of β for maximum ac power transmission
β_S = The value of β for steady state power transmission

The critical fault clearing time of composite system has similar type of expression of unconverted ac system that is presented below.

$$T_{CR} = \sqrt{\frac{4H(\delta_{CR}-\delta_0)}{\omega_s P_{comb}}} \quad (12)$$

$$T_{CR} = \lambda_{comb}\sqrt{\frac{4H}{\omega_s P_l}} \quad (13)$$

Combining the equation (4) and (13) the expression for critical clearing time can be further modified as,

$$T_{CR} = \frac{\lambda_{comb}}{\lambda_{ac}} T_{cr} \quad (14)$$

$$T_{CR} = C\, T_{cr} \quad (15)$$

Here, $C = \frac{\lambda_{comb}}{\lambda_{ac}}$

For the improvement in transient stability for ac-dc combined system in comparison with the unconverted ac system the value C must be greater than one, i.e. $C > 1$. For example, if the value of k is 0.5 that gives 50% ac voltage and 50% dc voltage mix and if the value of α is considered 0.5 the value of C would be 1.55. This represents 55% improvement in transient stability when ac-dc combined system is considered. In this case same power transfer angle is considered for ac power transmission in composite ac-dc system and for unconverted ac system which is 44^0 [7]. Considering some reasonable values of k and α the Stability Improvement(SI) is presented in Table I.

Table I clearly show that for a particular value of α if the value of k is increased the transient stability of the composite system will decrease. The increase in k means increase in ac voltage mix and decrease in dc voltage mix. The value of α has also a significant impact on SI. For a particular value of k smaller decrease in α gives a greater increase in SI.

TABLE I
TRANSIENT STABILITY OF AC-DC COMPOSITE SYSTEM IN COMPARISON WITH THE UNCONVERTED AC SYSTEM

k	$\alpha = \frac{I_L}{I_{th}}$	C	Stability Improvement(SI) (%)
0.5	0.50	1.55	55
	0.45	1.60	60
	0.40	1.65	65
0.52	0.50	1.54	54
	0.45	1.59	59
	0.40	1.63	63
0.54	0.50	1.52	52
	0.45	1.57	57
	0.40	1.62	62
0.56	0.50	1.51	51
	0.45	1.56	56
	0.40	1.61	61
0.58	0.50	1.49	49
	0.45	1.54	54
	0.40	1.59	59

IV. NUMERICAL ANALYSIS WITH SIMULATION RESULT

For the numerical simulation, a power system is considered that presented in appendix. In case of unconverted pure ac line, considering 30% steady state stability margin [7], following results are found after simulation.

Steady state loading = 436MW
Steady state power transmission angle = 44.44^0
Line current = 0.7574kA
Value of α = 0.4208
Critical clearing time T_{cr} = 210ms.

Now, ac-dc composite system is analyzed through developed mathematical model and obtained results are compared with those obtained from the numerical simulation using MATLAB. Table II shows that the improvement in stability found from developed mathematical model are very much close to those found from numerical simulation.

TABLE II
COMPARISON OF STABILITY IMPROVEMENT BETWEEN DEVELOPED MODEL AND NUMERICAL SIMULATION

$\alpha = \frac{I_L}{I_{th}}$	k	Improvement in critical clearing time (%)		% of error
		with developed model	with numerical simulation	
0.4208	0.50	63.63	65.50	2.85
	0.52	62.44	64.40	3.04
	0.54	61.20	63.29	3.30
	0.56	59.89	62.04	3.47
	0.58	58.52	60.77	3.70
	0.60	57.08	59.50	4.06

The developed model gives a small percentage of errors due to the fact that some logical assumptions are considered in the expression of β and δ_{CR} for bringing the simplicity in the mathematical model of ac-dc composite system.

Again the ac-dc composite system is analysed, using developed mathematical model, to investigate the impact of different dc voltage mix on stability improvement. In this case steady state loading is considered as much as the steady state loading of unconverted ac system. The change in dc voltage in composite system has huge impact on SI that is shown in fig.2. It is clearly observed from the figure that the stability improvement increases with the increase in the value of dc voltage mix and obtained improvements for this particular system varies from 22% to 64% with the dc voltage mix variation from 10% to 50% respectively.

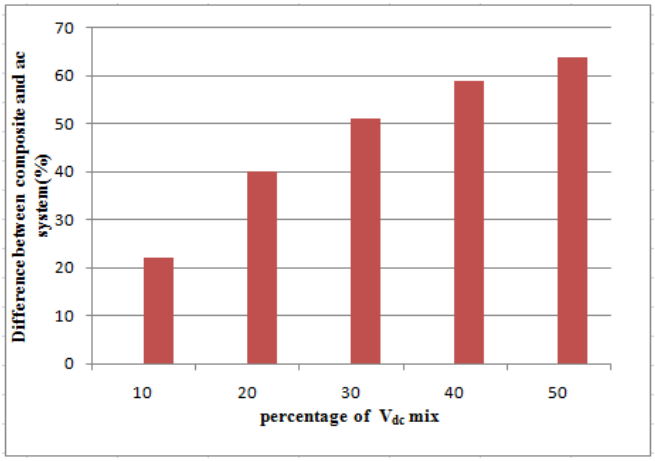


Fig. 2 Stability Improvement(SI) of ac-dc composite system.

V. Conclusion

Simultaneous ac-dc power transmission through the existing ac line improves the transient stability. The improvement in stability for the composite system (ac-dc simultaneous power flow) is shown in the paper both analytically and through numerical simulation. Moreover, ac-dc composite power transfer system provides the benefit of ac-dc parallel transmission system without claiming separate line and right-of- way. The mathematical model developed in this paper will be helpful and analytical tool for power transmission planner to take right decision whether they will go for ac-dc composite system.

Appendix

A sample power system is considered for the simulation study with numerical data

A. Unconverted AC line configuration

Base power=550MVA

Line voltages are 24kV, 132kV, 345kV for the generator, generator bus and transmission line respectively.

Ratings of different components are as follows:

Two identical generators each has the following rating 550MVA, 24kV, 50Hz, H=3.5sec, 30% transient reactance.

Transformer T_1&T_2(Between generator and generator bus) each of 550MVA, 24/132 kV, 50Hz, 10% reactance.

Transformer T_3 : 1100MVA,132/345 kV,50Hz, 8% reactance.(At the sending end of the line)

Transformer T_4 : 1100MVA,345/132 kV,50Hz, 8% reactance.(At the receiving end of the line)

Transmission line: ACSR(Moose; commercial name) twin bundle conductor, 345kV, 50Hz, 400km single circuit line has following parameters

X_{Tr} = 0.33086 Ω/km/phase, R = 0.03252 Ω/km/phase

$Y = 3.33797*10^{-6}$ S/km/phase

current carrying capacity of each sub conductor is 0.9kA.

I_{th}(thermal limit)=1.8kA,

B. Composite AC-DC configuration

Transformer T_3 = 550MVA, 132/172.5kV, 50Hz, 8% reactance.

Transformer T_4 = 550MVA,172.5/132kV, 50Hz,8% reactance.

References

[1] L. Wang, and C.T. Hsiung "Dynamic Stability Improvement of an Integrated Grid-Connected Offshore Wind Farm and Marine-Current Farm Using a STATCOM" *IEEE Transaction on Power System, Vol. 26 ,No. 2, May 2011.*

[2] N. Johansson, L. Ängquist, and H. P. Nee "An Adaptive Controller for Power System Stability Improvement and Power Flow Control by Means of a Thyristor Switched Series Capacitor (TSSC)" *IEEE Transaction on Power System, Vol. 25 ,No. 1, November 2010.*

[3] P. K. Dash, B. Puthal, o. P. Malik and G. S. Hope "Transient Stability and Optimal Control of Parallel A.C.-D.C. Power System" *IEEE Transaction on Power Apparatus and Systems, Vol. PAS-95, No. 3 May/June 1976.*

[4] A.E. Hammad "Stability and Control of HVDC and AC Transmissions in Parallel" *IEEE Transaction on Power Delivery, Vol. 14, No.4, October 1999.*

[5] H. Rahman and B. H. Khan, "Stability Improvement of Power System by Simultaneous AC–DC Power Transmission" *ELSEVIER, Electric Power System Research 78(2008) 756-764.*

[6] K.P. Basu "Stability Enhancement of Power System by Controlling HVDC Power Flow through the Same AC Transmission Line" *2009 IEEE Symposium on Industrial Electronics and Application (ISIEA 2009), October 4-6, 2009, Kuala Lumpur, Malaysia.*

[7] R.D. Dunlop, R. Gutman, P.P. Marchenko "Analytical Development of Loadability Characteristics for EHV and UHV Transmission Line" *IEEE Transaction on Power Apparatus and Systems, Vol. PAS-98, No. 2 March/April 1979.*

Load Frequency Control of Multi-Microgrid Using Energy Storage System

A. Hasib Chowdhury, Md. Asaduz-Zaman*
Dept. of EEE, Bangladesh University of Engineering & Technology
Dhaka-1205, Bangladesh
*masad.eee06@gmail.com

Abstract—**Control of active power and frequency in a power system is referred to as load frequency control (LFC). A linear combination of tie line power between interconnected system and frequency deviations, defined as area control error (ACE), is regulated to zero by LFC such that frequency and tie-line power errors are forced to zeros and a balance is achieved between generation and load. This paper presents an investigation on the LFC of multi-microgrid (MMG) using energy storage system (ESS). The tie-line connected MMG system consists of two microgrids, each made up of a synchronous generator, a PV source, a wind power system, ESS and load. Control strategies are developed based on both, frequency deviation and ACE, to regulate the ESS output. A performance index is defined for comparative assessment of control strategies. Effectiveness of the frequency control strategies is studied through simulation.**

***Index Terms*—Load frequency control, Multi-microgrid, Energy storage system, Area control error, Integral square error**

I. Introduction

Conventional power generation from fossil fuels face the issues of climate change, depletion of fossil fuel resources, and rising energy demand. These challenges can be addressed by integrating distributed generators with the distribution system to supply electricity from renewable energy sources such as wind, photovoltaic and biomass [11].

Microgrids (MG) are low voltage (LV) grids that interconnect distributed generation (DG) microsources and storage devices via feeders with small loads. They can either operate grid-connected to the medium voltage (MV) distribution system or in autonomous mode. However, an MG can only supply a maximum load capacity of 10 MVA [11].

The new concept of multi-microgrids (MMGs) is related to a higher level structure, formed at the medium voltage level, consisting of several LV MGs and DG units connected on adjacent MV feeders. The possibility of having a large number of controllable MGs, DG units and MV loads under demand side management (DSM) control requires the use of a hierarchical control scheme that enables an efficient control and management of this kind of system. These concepts are being developed under the EU financed More-Microgrids Project [10].

Control of active power and frequency in a power system is referred to as load frequency control (LFC). Since energy storage system (ESS) is an integral part of MGs and it can provide fast active power compensation, it also can be used to improve the performance of LFC. ESSs show much promise for mitigating the dynamics introduced by non-dispatchable variable generation [13]. Storage facilities possess additional dynamic benefits such as load levelling, spinning reserve, area regulation, long line stabilization, power factor correction and black start capability [12].

MMG is a new concept and load frequency control of MGG using ESS yet is not studied. This paper presents an investigation on the load frequency control of MMG using ESS. The MMG consists with two MGs, each made up of a diesel generator, a PV sources, a wind power generator, ESS and load. The MGs are interconnected by tie-lines. Schematic diagram of an MG and an MMG are shown in Figs. 1(a) and 1(b). Control strategies are developed based on both, frequency deviation and ACE, to regulate the ESS output. A performance index is defined for comparative assessment of control strategies. Effectiveness of the frequency control strategies is studied through simulation.

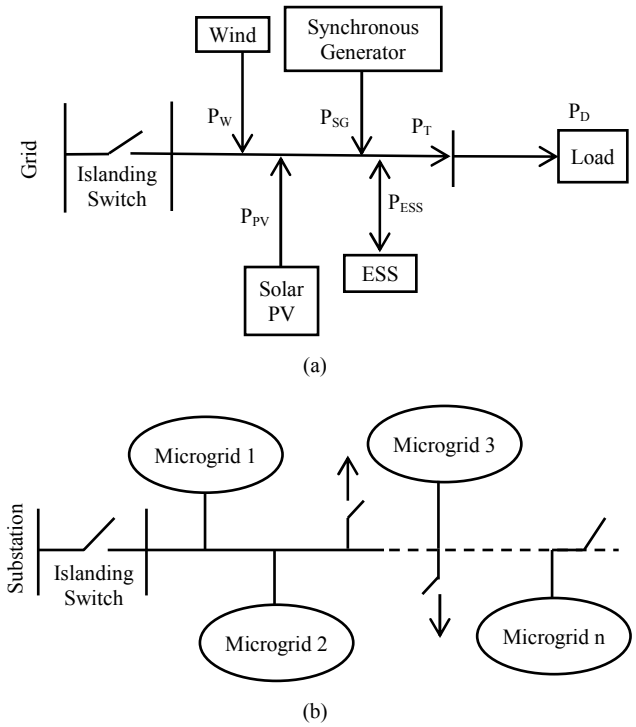


Fig.1 Schematic diagram of (a) microgrid system, (b) multi-microgrid system

II. System Modelling

In this section the modelling of different component of microgrid system is presented.

A. Synchronous Generator System

Standby synchronous generator (SG) work autonomously to supply the deficit power to the microgrid system to meet the supply-load demand balance condition. Smaller MGs usually use diesel generator but larger MGs may use turbine driven generator. The synchronous generator system considered here consists of speed-governor and turbine system. The transfer function of governor and the turbine can be expressed by [2]:

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

$$G_G(s) = \frac{\Delta P_V(s)}{\Delta P_g(s)} = \frac{K_g}{1+sT_g} \quad (1)$$

$$G_T(s) = \frac{\Delta P_m(s)}{\Delta P_V(s)} = \frac{K_T}{1+sT_T} \quad (2)$$

where T_g and T_T are the time constant of governor and turbine, respectively. The value of governor gain constant (K_g) and turbine gain constant (K_T) is set to 1.0. The details of modelling is given in [1,2]. The governors typically have a speed regulation (R) of 5-6 percent from zero to full load [2].

B. Wind Power Generation System

In this paper, the standard deviation for wind power output has been simulated by the following function.

$$dP_{WP} = 0.8\sqrt{P_{WP}} \quad (3)$$

Random output fluctuation is derived from white noise block with some manipulation in MATLAB/SIMULINK is shown in Fig. 3. It is multiplied by the standard deviation to calculate output fluctuation on the model [7,9].

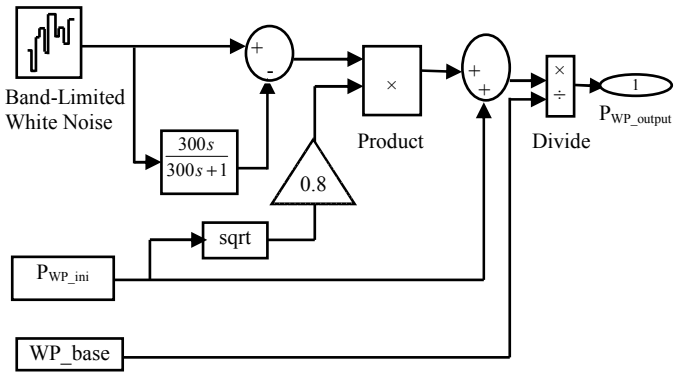


Fig.3 MATLAB/SIMULINK model of wind power system

C. Solar Photovoltaic Model

A PV system consists of many cells connected in series and parallel to provide the desired voltage and current. The voltage and current relationship is non-linear in nature. The maximum power output of the PV array varies according to solar radiation or load current. Therefore, control strategy is required to use solar radiation effectively in order to obtain maximum power. The output power of the PV system can be expressed as,

$$P_{PV} = \eta S\varphi[1 - 0.005(T_a + 25)] \quad (4)$$

Where η is conversion efficiency of the PV array, S is measured area of PV array (m^2), φ is solar irradiation (kW/m^2) and T_a is ambient temperature (^0C). The transfer function of the solar PV system can be expressed by a simple linear first order lag [3,4],

$$G_{PV}(s) = \frac{\Delta P_{PV}}{\Delta\varphi} = \frac{K_{PV}}{1+sT_{PV}} \quad (5)$$

where K_{PV} is gain constant and T_{PV} is time constant.

D. Energy Storage System

Energy storage systems play an important role to effectively supply deficit energy of power generation subsystems of the hybrid power system very fast to maintain system stability [3]. With the rapid advances in power electronics technology, the battery energy storage system has recently been shown to have the capability to quickly control both its active and reactive power output at switching frequency well above the kHz range. While the main applications of such a device are usually for load levelling, harmonic cancellation and voltage control, once installed, it can also be utilized to provide additional damping to power system swings to improve both transient and dynamic stability [4].

The transfer functions of the energy storage systems (ESS) can be taken as first order lag,

$$G_{ESS}(s) = \frac{\Delta P_{ESS}}{\Delta\omega} = \frac{K_{ESS}}{1+sT_{ESS}} \quad (6)$$

where K_{ESS} is gain constant and T_{ESS} is time constant.

E. Power Deviation and System Frequency Variation

To maintain a stable operation of an autonomous system, the total power generation must be effectively controlled and properly dispatched to meet the total power demand of the connected loads [3].

The total power generation (P_T) by the microgrid system shown in Fig.1 is the algebraic summation of synchronous generator power (P_{SG}), wind power (P_W), solar photovoltaic power (P_{PV}) and power of energy storage system (P_{ESS}).

$$P_T = P_{SG} + P_W + P_{PV} \pm P_{ESS} \quad (7)$$

The difference between total power generation P_T and power demand reference P_D is given by Eq. 8.

$$\Delta P_e = P_T - P_D \quad (8)$$

Because system frequency is changed with net power variation, the system frequency variation $\Delta\omega$ is calculated by

$$\Delta\omega = \frac{\Delta P_e}{K_{sys}} \quad (9)$$

where K_{sys} is system frequency characteristic constant of the microgrid system. Since an inherent time delay exists between system frequency variation and power deviation, the transfer function for system frequency variation to per unit power deviation can be expressed by

$$G_{sys}(s) = \frac{\Delta\omega}{\Delta P_e} = \frac{1}{K_{sys}(1+sT_{sys})} = \frac{1}{D+Ms} \quad (10)$$

where M and D are, the equivalent inertia constant and damping constant of the microgrid system, respectively.

III. Interconnection Of Microgrids With Tie-Line

Reliable power supply to the connected load may be achieved by the interconnection of neighbouring isolated MGs through tie-line. Each isolated MG is assumed as control area. Tie-line bias is a control characteristic that determines how much a particular area of an interconnected system should respond to system frequency variation. Tie-line is used to exchange energy between control areas to provide inter-area support in case of any abnormal condition. If there is a mismatch between the power generation and load demand, the deviation in frequency for a particular area occurs [4]. Interconnected microgrid with two-area interconnection is represented as shown in Fig.4.

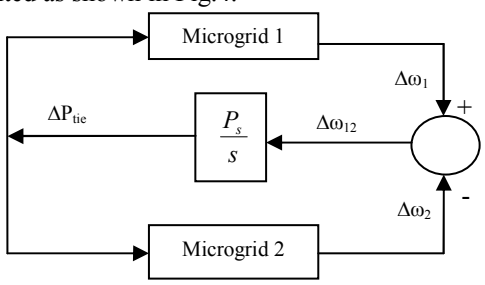


Fig.4 Modelling of tie-line connection

Expressing the tie-line power deviation (ΔP_{tie}) in terms of frequency deviation ($\Delta\omega$) and synchronizing power co-efficient (P_s) is represented as

$$\Delta P_{tie} = P_s(\int \Delta\omega_1 dt - \int \Delta\omega_2 dt) \tag{11}$$

Taking the laplace transform we have

$$G_{tie} = \frac{\Delta P_{tie}(s)}{\Delta\omega(s)} = \frac{P_s}{s} \tag{12}$$

The area control error (ACE) of the i^{th}microgrid is given by,

$$ACE_i = \sum_{j=1}^{n} \Delta P_{ij} + B_i \Delta\omega_i \tag{13}$$

where B is the frequency bias factor for i^{th}microgrid. Fig. 5 shows the block diagram of two interconnected MG system. It shows that the ESS may be controlled either by frequency deviation ($\Delta\omega$) or ACE.

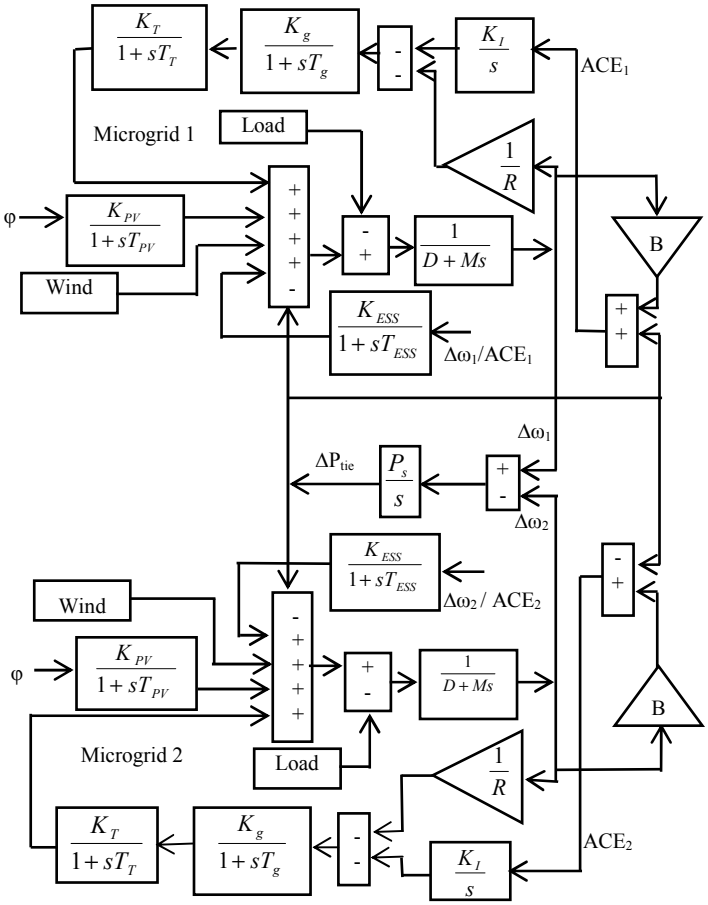


Fig.5 Block diagram of two interconnected microgrid system

IV. Simulation Results and Analysis

The block diagram of interconnected MMG systems shown in Fig. 5 are simulated and analysed. Parameter values of the configurations are taken from [1-4,14-16] and listed in Table I. The synchronizing power co-efficient (P_s) is set to 1.5pu. The average value of wind power and solar power is kept about 0.4pu and 0.2pu for MG1 and 0.3pu and 0.2pu for MG2, respectively. The load at both MG is 0.7pu. During all simulation process a load change of 0.1pu at MG1 from 0.7pu to 0.8pu at t=1.0s is considered.

Figures 6 and 7 present simulation results with $\Delta\omega$ and ACE feedback to ESS, respectively. Frequency deviation ($\Delta\omega$), output power of energy storage system (P_{ESS}) and tie-line power flow deviation (ΔP_{tie}) is presented in both the figures. All quantities in the plots are in per unit (pu) values. The suffix 1 and 2 used for MG1 and MG2 quantities, respectively.

TABLE I
PARAMETER VALUES OF STUDIED MG SYSTEMS

Parameter	Definition	MG1	MG2
R(pu)	Speed regulation	0.05	0.04
$T_g(s)$	Governor time constant	0.1	0.1
$T_T(s)$	Turbine time constant	0.4	0.4
$T_{PV}(s)$	Solar PV time constant	1.5	1.4
$T_{ESS}(s)$	ESS time constant	0.1	0.1
K_{ESS}	ESS gain constant	-10	-8
K_I	Integral gain	5	7
B(pu)	Frequency biasing factor	10	12.5
M(pu)	Inertia constant	0.8	0.7
D (pu)	Damping constant	0.02	0.03

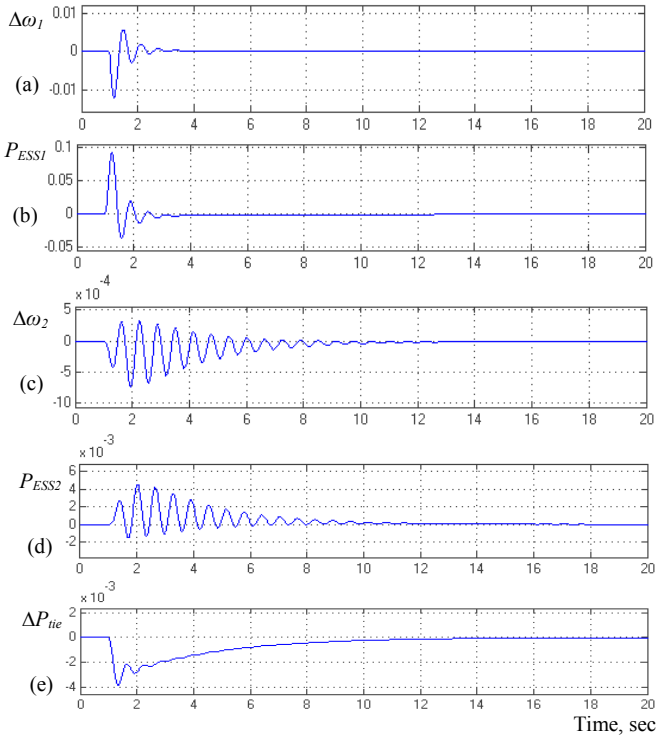


Fig.6 Simulation output for $\Delta\omega$ feedback to ESS control

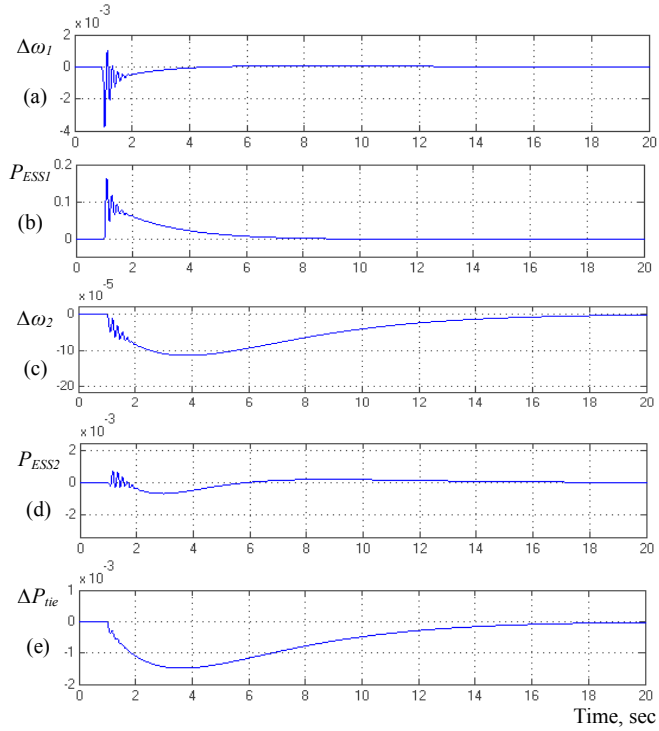


Fig.7 Simulation output for ACE feedback to ESS control

From Figs. 6 and 7, it can be seen that the frequency deviation and tie-line power flow deviation is lower when the ACE is used as the feedback for load frequency control using ESS compare to frequency deviation feedback. An integral square error (ISE) is considered as performance index (PI) to quantify the result analysis. ISE is given by [4,15],

$$ISE = \int e(t)^2 dt \tag{14}$$

where $e(t)$ is the error signal which may be frequency deviation or tie-line power flow deviation. The ISE values for such deviation are tabulated in Table II.

TABLE II
ISE FOR FREQUENCY DEVIATION AND TIE-LINE POWER FLOW DEVIATION

Error Signal	ISE (pu) for	
	$\Delta\omega$ feedback	ACE feedback
Frequency deviation at MG1	3.177×10^{-5}	1.445×10^{-6}
Frequency deviation at MG2	4.198×10^{-7}	6.946×10^{-8}
Tie-line power flow deviation	1.827×10^{-5}	1.120×10^{-5}

It can be seen from Table II that the ISE for frequency deviation and tie-line power flow deviation is low when ACE is feedback to ESS control compared to that when frequency deviation is feedback. From qualitative and quantitative comparison it can concluded that the ACE controller has a better performance for load frequency control using the ESS.

V. CONCLUSION

This paper presented load frequency control (LFC) of two interconnected microgrid systems where both frequency deviation and area control error (ACE) are tested as feedback signals for controlling the output power of energy storage system. The employed mathematical models for the various component of the system are represented by first-order transfer functions to simplify the tasks of system simulation. During simulation process a step load change is applied. The frequency deviations and tie-line power flow deviation profile for interconnected multi-microgrid system were studied, both qualitatively and quantitatively. Integral square error (ISE) is considered as performance index for comparative assessment of control strategies. It is observed that control strategy that regulates the ACE to zero is most suitable for load frequency control in multi-microgrid system.

REFERENCES

[1] P. Kundur, "Power System Stability and Control," New York, USA: McGraw-Hill, 1994.

[2] H. Saadat, " Power System Analysis," New York, USA:McGraw-Hill, 1999.

[3] D. J. Lee, and L. Wang,"Small-Signal Stability Analysis of an Autonomous Hybrid Renewable Energy Power Generation/Energy Storage System Part I:Time-Domain Simulations," *IEEE Transactions on Energy Conversion*, vol. 23, no. 1, pp. 311-320, March 2008.

[4] P. K. Ray, S. R. Mohanty, and N. Kishor,"Small-Signal Analysis of Autonomous Hybrid Distributed Generation Systems in Presence of Ultracapacitor and Tie–Line Operation," *Journal of Electrical Engineering*, vol.61, no.4, pp.205-214, 2010.

[5] T. Senjyu, T. Nakaji, K. Uezato, and T. Funabashi,"A Hybrid Power System Using Alternative Energy Facilities in Isolated Island," *IEEE Transactions on Energy Conversion*, vol. 20, no. 2, pp. 406-414, June 2005.

[6] M. A. Sofla, and R. King,"Control Method for Multi-Microgrid Systems in Smart Grid Environment—Stability, Optimization and Smart Demand Participation," *IEEE PES, Innovative Smart Grid Technologies* (ISGT), pp. 1-5, Date 16-20 January, 2012.

[7] M. Matsubara, G. Fujita, T. Shinji, T. Sekine, A. Akisawa, T. Kashiwagi, and R. Yokoyama,"Supply and Demand Control of Dispersed Type Power Sources in Micro Grid," Proceedings of the 13th International Conference, Intelligent Systems Application to Power Systems, pp.67 – 72, 2005.

[8] X. Li, Y. Li, X. Hanb, and D.Hui,"Application of Fuzzy Wavelet Transform to Smooth Wind/PV Hybrid Power System Output with Battery Energy Storage System," *The Proceedings of International Conference on Smart Grid and Clean Energy Technologies*, vol. 12, pp .994–1001, 2011.

[9] X. Li, D. Hui, L. Wu, and X. Lai,"Control Strategy of Battery State of Charge for Wind/Battery Hybrid Power System,"IEEE International Symposium on Industrial Electronics (ISIE), pp. 2723 – 2726, 2010.

[10] N. J. Gil, and J. A. P. Lopes, "Hierarchical Frequency Control Scheme for Islanded Multi-Microgrids Operation," in IEEE Lausanne Power Tech Lausanne, pp.473-478, 2007.

[11] S. Chowdhury, S. P. Chowdhury, and P. Crossley, "Microgrids and Active Distribution Networks," The Institution of Engineering and Technology, 2009.

[12] S.K. Aditya, and D.Das,"Load-Frequency Control of an Interconnected Hydro-Thermal Power System with New Area Control Error Considering Battery Energy Storage Facility," *International Journal of Energy Research*, vol.24, pp.525-538, 2000.

[13] M. U. Usama, D. Kelle, and T. Baldwin, "Utilizing Spinning Reserves as Energy Storage for Renewable Energy Integration," Power Systems Conference, Clemson University, pp.1-5, March 2014.

[14] D. Kottick, M. Blau, and D. Edelstein,"Battery Energy Storage for frequency regulation in an Island Power System", *IEEE Transactions on Energy Conversion*, vol. 8, no.3, pp.455-459, September 1993.

[15] Prakash K. Ray, Soumya R. Mohanty, Nand Kishor,"Proportional–integral controller based small-signal analysis of hybrid distributed generation systems," Energy Conversion and Management, vol. 52, no.4, pp.1943–1954, April 2011.

[16] H. Bevrani, F. Habibi, P. Babahajyani, M. Watanabe, and Y. Mitani, "Intelligent Frequency Control in an AC Microgrid: Online PSO-Based Fuzzy Tuning Approach," *IEEE Transactions on Smart Grid*, vol. 3, no. 4, pp.1935 – 1944, December 2012.

Parameterization of Active Disturbance Rejection Controller for Load Frequency Control

Md. Mijanur Rahman[1*], A Hasib Chowdhury[2]
[1]Department of EEE, Eastern University
Dhaka-1205, Bangladesh
[2]Department of EEE, Bangladesh University of Engineering and Technology
Dhaka-1205, Bangladesh
*mijan@easternuni.edu.bd

Abstract—**Active disturbance rejection control (ADRC) has gained significant attraction due to its simple tuning method and robustness against process parameter variations, whether the plant is linear or non-linear, time-varying or time-invariant. In load frequency control, the use of ADRC is proposed to regulate the area control error, frequency error and tie-line error in the presence of internal dynamics and external disturbance. This paper presents a study on the effects of state observer bandwidth, controller bandwidth and sampling period on the performance of the ADRC based load frequency control of single- and multi- area power systems. The study gives good insight in choosing controller parameters for ADRC based LFC.**

Index Terms — **Load frequency control, Active disturbance rejection control, Parameterization**

I. Introduction

The control of active power and frequency in a power system is referred to as load frequency control (LFC). Stable operation of interconnected power system requires both constant frequency and constant tile-line power exchange [1]. An area control error (ACE), defined as a linear combination of tie line power and frequency deviations, is taken as the controlled output of LFC. LFC regulates ACE to zero such that frequency and tie-line power errors are forced to zeros.

PID controller is widely used in LFC. It is simple to implement but usually gives long settling time and produces large frequency deviation [2]. Fuzzy logic based PID and linear matrix inequality (LMI) tuned PID are also implemented for LFC. In recent years, active disturbance rejection control (ADRC) based LFC has been proposed [3-6]. The ADRC technique was first proposed by J. Han [7] and modified by Z. Gao [8, 9]. It does not require precise model information of the physical system, is very robust against parameter uncertainties and external disturbance, and shows fast dynamic response with no overshoot [10].

In practical applications, there are multiple parameters that require tuning in an ADRC controller. A concept of bandwidth parameterization is proposed in [8]. Using this concept, ADRC has only two tuning parameters, the observer bandwidth (ω_0) and feedback controller bandwidth (ω_c). Researchers have considered different values for ω_0 and ω_c. A range of ω_0 is presented in [8] but no procedure for selecting the observer bandwidth is offered. A thumb rule has been illustrated between the relationship of ω_0 and ω_c in [8,11]. For LFC of a multi-area power system, ω_0 is considered as four times to ω_c [9]. In other works on LFC [5,6] ω_0 is considered as five times of ω_c. In [12] a third controlling parameter, the sampling frequency (T), has been considered, which improves the response time.

This paper presents a comprehensive study on the effects of change in the three parameters (ω_c, ω_0 and T) on the performance of an ADRC based LFC in single- and multi-area power systems. The study gives good insight in choosing the controller parameter for an ADRC based LFC.

II. ADRC based LFC model

The block diagram of an ADRC based LFC is presented in Fig. 1. The 'Plant' block represents a single area of a power system, outputs of this block are frequency and tie flow deviations that are combined to form ACE. The ADRC block takes in the ACE and produces the input u for the plant. Frequency deviation is also feedback to the plant through the speed droop coefficient. What follows in this section is the detailed model of the controller.

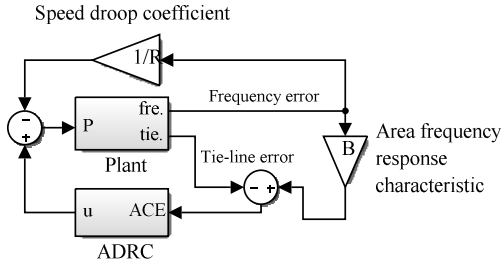


Fig. 1 Block diagram of ADRC based LFC

A. Transfer Function for an n-th Order Plant

A plant with disturbance can be represented as

$$Y(s) = G_p(s).U(s) + W(s) \quad (1)$$

Where $U(s)$ is the input, $Y(s)$ is the output, and $W(s)$ is the generalized disturbance. The transfer function of the physical plant $G_p(s)$ can be represented as

$$\frac{Y(s)}{R(s)} = G_p(s) = \frac{b_m s^m + b_{m-1}\ s^{m-1} + \cdots + b_1\ s + b_0}{a_n\ s^n + a_{n-1}\ s^{n-1} + \cdots + a_1\ s + a_0} \quad (2)$$

Where $R(s)$ is the reference input and a_i (i =1,….,n) and b_j (j =1,…, m) are coefficient of the transfer function.

The basic idea of ADRC design is based on the transfer function of the plant without zeros. Thus, in order to implement ADRC for the system represented by (1), it is necessary to develop an equivalent model of (2) so that the transfer function only has poles. Any error between the two models can be included into the generalized disturbance term. In order to develop the non-zero equivalent model of (2), the following polynomial long division is conducted on $1/G_p(s)$.

$$\frac{1}{G_p(s)} = \frac{a_n\ s^n + a_{n-1}\ s^{n-1} + \cdots + a_1\ s + a_0}{b_m s^m + b_{m-1}\ s^{m-1} + \cdots + b_1\ s + b_0}$$
$$= c_{n-m} s^{n-m} + bc_{n-m-1} s^{n-m-1} + \ldots + c_1 s + c_0 + G_{left} \quad (3)$$

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

In (3), c_i (i=0,…,n-m) are coefficients of polynomial division result, and the $G_{left}(s)$ is a reminder, which can be represented by

$$G_{left}(s) = \frac{d_{m-1}\ s^{m-1} + d_{m-2}\ s^{m-2} + \cdots + d_1\ s + d_0}{b_m s^m + b_{m-1}\ s^{m-1}\ + \cdots + b_1\ s + b_0} \quad (4)$$

In (4), d_j (j=0,…,m-1) are coefficient of the numerator of the remainder. Substituting (3) into (1) gives,

$$[c_{n-m}s^{n-m} + bc_{n-m-1}s^{n-m-1} + \cdots + c_1 s + c_0 + G_{left}(s)].Y(s) = U(s) + W'(s) \quad (5)$$

Where, $W'(s) = W(s)/G_p$

Equation (5) can be rewritten as

$$[c_{n-m}\ s^{n-m}]\,Y(s) = U(s) - [bc_{n-m-1}\ s^{n-m-1} + \ldots + c_1 s + c_0 + G_{left}(s)].Y(s) + W'(s) \quad (6)$$

Finally we have,

$$s^{n-m}\,Y(s) = \frac{1}{c_{n-m}}U(s) + D(s) \quad (7)$$

Where,

$$D(s) = -\frac{1}{c_{n-m}}[bc_{n-m-1}\,s^{n-m-1} + \ldots + c_1 s + c_0 + G_{left}(s)].Y(s) + \frac{1}{c_{n-m}}W'(s) \quad (8)$$

From (3), it can be seen that

$$c_{n-m} = \frac{a_n}{b_m} \quad (9)$$

However, it is difficult to get the expression of the other coefficients in (3) and (4). But, for the development process of ADRC, $D(s)$ is treated as the generalized disturbance and estimated in time domain so that the exact expression for the c_i and d_j are not needed.

From (7), it is seen that the two characteristics (relative order between input and output and controller gain) have been extracted from the plant by modifying the Laplace transform. Instead of using the order of plant n, the relative order n-m may be utilized as the order of the controller system and equation (7) can be written as

$$s^{n-m}\,Y(s) = bU(s) + D(s) \quad (10)$$

Where,

$$b = \frac{1}{c_{n-m}}$$

B. Design of ADRC

ADRC can be understood as a combination of an extended state observer (ESO) and a state feedback controller. The ESO is utilized to observe the generalized disturbance, which is also taken as an extended state, and the state feedback controller is used to regulate the tracking error between the real output and a reference signal for the physical plant [7]

Extended state observer

The effectiveness of ADRC is dependent on the accurate estimation of the generalized disturbance $D(s)$. Consequently an ESO needs to be developed to estimate $D(s)$ in real time.

Now the state space model of (7) is,

$$sX(s) = AX(s) + BU(s) + E(s)\,D(s) \quad (11)$$

$$Y(s) = CX(s)$$

Where, $X(s) = \begin{bmatrix} X_1(s) \\ X_2(s) \\ \vdots \\ X_{n-m}(s) \end{bmatrix}$

$$A = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & \ddots & 1 & \ddots & \vdots \\ \vdots & \ddots & 0 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 1 \\ 0 & 0 & \ldots & \ldots & 0 \end{bmatrix}_{(n-m+1)*(n-m+1)} \qquad B = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ b \\ 0 \end{bmatrix}_{(n-m+1)*1}$$

$$C = [1 \quad 0 \quad \ldots \quad 0]_{1*(n-m+1)} \qquad E = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}_{(n-m+1)*1}$$

The ESO of the plant is

$$SZ(s) = AZ(s) + BU(s) + L(Y(s) - \hat{Y}(s)) \quad (12)$$

$$\hat{Y}(s) = CZ(s)$$

Where Z(s) is the estimated state vector given by

$$Z(s) = [z_1(s) \quad z_2(s) \ldots \quad z_{n-m}(s)]^T$$

L is the observer gain vector given by

$$L = [\beta_1 \quad \beta_2 \ldots \quad \beta_{n-m}]^T$$

In order to locate all the Eigen values of the ESO to $-\omega_0$, the observer gains are chosen as

$$\beta_i = \binom{n-m}{i}.\,\omega_0^i, \qquad i = 1, \ldots, n\text{-}m$$

With a well tuned ESO, $Z_i(s)$ will be able to estimate the value of $X_i(s)$ closely (i= 1 ,…, n-m). Then we have

$$Z_{n-m}(s) = \hat{D}(s) \approx D(s) \quad (13)$$

Feedback Controller

If the control input is designed as

$$U(s) = (U_0(s) - Z_{n-m}(s))\,/\,b \quad (14)$$

Equation (10) will reduce to a pure integral part, i.e.,

$$s^{n-m}\ Y(s) = b.\frac{u_0(s) - Z_{n-m}(s)}{b} + D(s) = U_0(s) - \hat{D}(s) + D(s) \approx U_0(s) \quad (15)$$

The control law for the pure integral plant is

$$U_0(s) = k_1\,(R(s) - Z_1(s)) - k_2\,Z_2(s) - \ldots - k_{n-m-1}\,Z_{n-m-1}(s) \quad (16)$$

Where R(s) is a reference input. With the control law in (16), Y(s) will be drive to R(s) which is zero for LFC. To further simplify the process, all the closed loop poles of the PD controller are set to $-\omega_c$. Then the controller gain in (16) has to be selected as

$$K_i = \binom{n-m-1}{n-m-i}.\,\omega_c^{n-m-1} \quad (17)$$

where i = 1,…., n-m-1

III. Simulation Results

The efficacy of the ADRC based LFC with change in observer bandwidth (ω_0), controller bandwidth (ω_c) and controller sampling time (T) is studied for single area and multi-area power system using MATLAB Simulink. Controller bandwidth is selected based on the time required to settle down the response. Feedback controller will reject the internal dynamics and external disturbance which will be estimated by the ESO. So the observer bandwidth must be higher than the feedback controller bandwidth [8]. Effectiveness of observer's disturbance estimation depends on the observer gain vector and the gain vectors are represented as the function of observer bandwidth (ω_0).The parameters of the single- and multi- area power systems are taken from [3] and [4], respectively.

A. Single Area Power System

The block diagram model of a single area power system with ADRC based LFC is shown in Fig. 2. Assuming coherency generators are represented by a single equivalent generator. A non-reheat turbine system is considered. Effects of parameter variation and for a step load change on frequency and tie flow deviations and, ACE are presented in Figs. 3 and 4 and in Table I. It can be seen that the peak amplitudes do not change much. This is because for all the cases the disturbance considered is the same.

The controller bandwidth (ω_c) is tuned according to the requirement of how fast and steady the output is wanted to track the set point. Higher bandwidth corresponds to better command following, disturbance rejection and sensitivity to parameter variations. However, bandwidth is limited by the presence of sensor noise and due to the higher value of controller bandwidth (ω_c= 5 rad /sec), it may push the system to its limit, leading to oscillations or even instability as shown in Figs. 3(a) and (c).

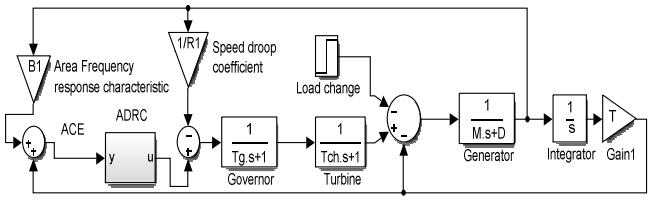


Fig. 2 Model of a single area power system with ADRC based LFC

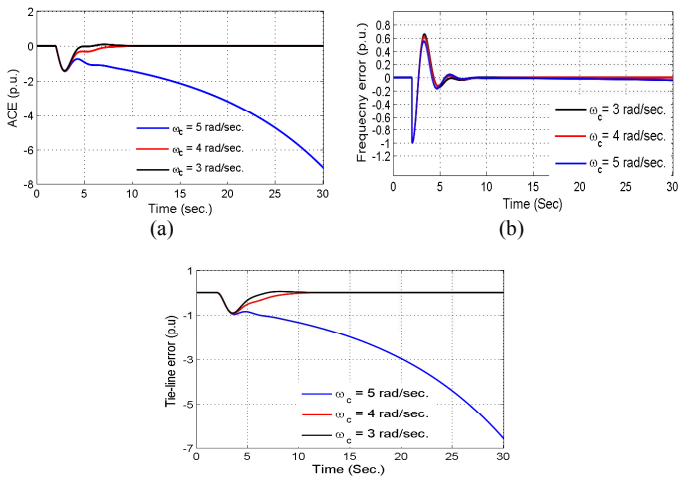


Fig. 3 Effect of controller bandwidth (ω_c) on (a) ACE, (b) frequency error, and (c) tie-line error

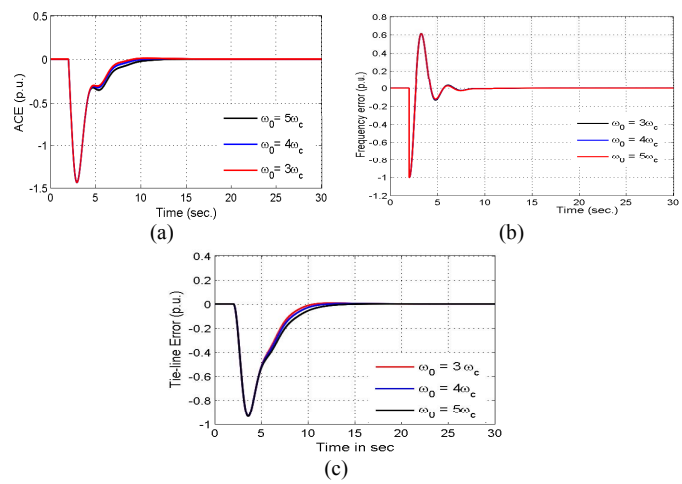


Fig. 4 Effect of observer bandwidth (ω_0) on (a) ACE, (b) frequency error, and (c) tie-line error

TABLE I

PERFORMANCE OF SINGLE AREA POWER SYSTEM

Error type	Parameters variation	Peak Amplitude (p.u.)	Time at peak overshoot (sec.)	Settling time (sec.)
Effects of change in controller bandwidth (ω_c) (rad/sec)				
ACE	ω_c=3	-1.7	3.0	10.0
	ω_c=4	-1.7	3.0	10.0
	ω_c=5	-1.7	3.0	unstable
Frequency error	ω_c=3	-0.65	2.0	9.0
	ω_c=4	-0.6	2.0	8.0
	ω_c=5	-0.55	2.0	7.5
Tie-line error	ω_c=3	-0.9	3.5	10.0
	ω_c=4	-0.9	3.5	12.0
	ω_c=5	-0.9	3.5	unstable
Effects of change in observer bandwidth (ω_0) (rad/sec)				
ACE	$\omega_0 = 3\omega_c$	-1.4	3.0	11.0
	$\omega_0 = 4\omega_c$	-1.4	3.0	9.0
	$\omega_0 = 5\omega_c$	-1.4	3.0	10.0
Frequency error	$\omega_0 = 3\omega_c$	-0.6	2.5	8.0
	$\omega_0 = 4\omega_c$	-0.6	2.5	8.0
	$\omega_0 = 5\omega_c$	-0.6	2.5	8.0
Tie-line error	$\omega_0 = 3\omega_c$	-0.9	4.0	11.0
	$\omega_0 = 4\omega_c$	-0.9	4.0	12.0
	$\omega_0 = 5\omega_c$	-0.9	4.0	12.0

As ω_c and ω_0 are increased, the noise in control signal u also increases. Any notable change does not found in magnitude of ACE due to change in observer bandwidth in Fig. 4(a) but for higher bandwidth, the response take longer time to settle down in case of tie-line error in Fig.4(c). It is very difficult to establish any relationship between observer bandwidth and controller bandwidth but simulation results indicates that observer bandwidth should be chosen four times of controller bandwidth to get the shortest settling time as can be seen from Figs. 4(a), (b) and (c) and Table I.

In general, the shorter the sampling period the higher the control degree of accuracy for the same plant with increasing observer bandwidth [12]. However, for load frequency control, larger observer bandwidth increases the settling time as can be seen in Figs. 4(a) and (b) and Table I. However, it increases the noise sensitivity of the system.

B. Multi-Area Power System

An ADRC based LFC for a three area interconnected power system is also modelled and simulated. The model consists of three generation units, each of which is composed of three major parts: governor, turbine and generator. All generators in one area respond coherently, so they are represented by an equivalent generator. All turbine units are to be considered non-reheat type. Fig. 5 presents the dynamic model of one-area in the multi-area power system.

A 0.01 p.u. load change is added to each area. The simulation results are presented in Table II for area-1 by varying the three controlling parameters. Peak amplitude is still same for most of the cases due to same load change consideration. The settling time is calculated by considering the tolerance limit of the response 0.01 p.u. It is seen that the time required reaching the peak overshoot and settling time increases with increasing controller bandwidth (ω_c) and controller sampling period, respectively. However, a small change in settling time is noted when the multiplying factor

between ω_0 and ω_c is increased. This is because a large observer bandwidth increases noise sensitivity.

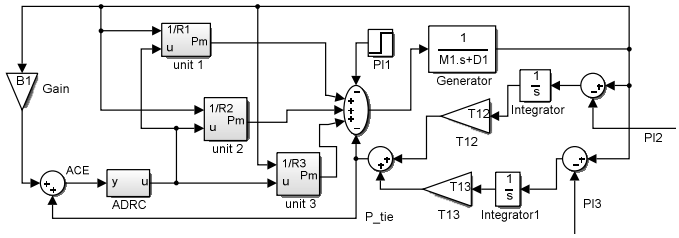


Fig. 5 Dynamic model of one-area in a multi-area power system

TABLE II

PERFORMANCE OF MULTI-AREA POWER SYSTEM

Error type	Parameters variation	Peak amplitude (p.u.)	Time at peak overshoot (sec.)	Settling time (sec.)
Effects of controller bandwidth variation (ω_c) (rad/sec)				
ACE	ω_c=3	-0.02	2.0	25
	ω_c=4	-0.02	2.5	25
	ω_c=5	-0.02	3.5	27
Frequency error	ω_c=3	-0.0075	1.5	26
	ω_c=4	-0.005	2.0	27
	ω_c=5	-0.008	3.5	27
Tie-line error	ω_c=3	-0.002	3.5	23
	ω_c=4	-0.008	3.0	23
	ω_c=5	-0.023	3.0	22
Effects of observer bandwidth variation (ω_0) (rad/sec)				
ACE	$\omega_0 = 3\omega_c$	-0.02	2.5	22
	$\omega_0 = 4\omega_c$	-0.02	2.5	24
	$\omega_0 = 5\omega_c$	-0.02	2.5	24
Frequency error	$\omega_0 = 3\omega_c$	-0.008	2.0	22
	$\omega_0 = 4\omega_c$	-.0008	2.0	23
	$\omega_0 = 5\omega_c$	-0.008	2.0	22
Tie-line error	$\omega_0 = 3\omega_c$	-0.002	2.5	22
	$\omega_0 = 4\omega_c$	-0.008	2.0	21
	$\omega_0 = 5\omega_c$	-0.022	2.5	20

IV. CONCLUSION

This paper presents a comprehensive study on the effects of change in the ADRC parameters in load frequency controller to regulate the area control error, frequency deviations and tie-line error in single- and multi- area power systems. It is observed that for ADRC based load frequency control, a controller bandwidth of 4 rad/sec for both power system and observer bandwidth of four times of controller bandwidth gives good result for single area power system in Table I. However, observer bandwidth should five times of controller bandwidth for multi-area power system to get the lower settling time in Table II. In the future, the parameters for LFC of two- and three- area power system with re-heat turbines and hydro-electric turbine should be studied. It will help to enhance the range of applications of higher order ADRC to the power system.

REFERENCES

[1] K. Tomsovic, D. E. Bakken, V. Venkatasubramanian, and A. Bose, "Designing the Next Generation of Real-time Control, Communication, and Computations for Large Power Systems", in Proceedings of the IEEE, vol. 93, no. 5, pp. 965-979, May 2005

[2] G. Panda, S. Panda, and C. Ardil, "Automatic Generation Control of Interconnected Power system with Generation Rate Constraints by Hybrid Neuro Fuzzy Approach", International Journal of Electrical Power and Energy Systems Engineering, vol.2, no.1, pp. 13-18, 2009

[3] A.H.M. Sayem, "Active disturbance rejection control for load frequency control of Bangladesh power system," M. Eng. thesis, Bangladesh University of Engineering and Technology, Bangladesh, Nov. 2012

[4] Y. Zhang, L.Dong and Z. Gao, "Load frequency control of multi-area power system," American control conference, June, 2009.

[5] Y. Zhang, L.Dong, "On design of a robust load frequency controller for interconnected power system," American control conference, July, 2010

[6] Y. Zhang, L.Dong and Z. Gao, "A robust decentralized load frequency controller for interconnected power system," ISA Transactions vol. 51, pp. 410-419, 2012

[7] Z. Gao, Y. Huang and J. Han, "An alternative paradigm for control system design," Proc. of IEEE Conference on decision and control, vol.5, no.4-7, pp.4578-4585, Dec. 2001

[8] Z.Gao, "Scaling and Bandwidth-Parameterization Based Controller Tuning," Proc. of the American Control Conference, pp. 4989-4996, 2003

[9] Z. Gao, "Active disturbance rejection control: A paradigm Shift in feedback control system design," Proc. of American control conference, pp. 16-18, June 2006

[10] R. Miklosovic, Z. Gao, "A robust two-degree-of-freedom control design technique and its practical application," Proc. of IEEE Industry Applications Conference, vol. 3, 2004

[11] Dan Wu, K. Chen, and X. Wang," Tracking control and active disturbance rejection with application to nonlinear machining," International Journal of Machine Tools and Manufacture, vol. 47, pp. 2207-2217, 2007

[12] S. Shi, J. Lu and S Zhao, "On design analysis of linear active disturbance rejection control for uncertain system, "International Journal of Control and Automation, vol.7. No.3, pp. 225-236, 2014

A Fully Integrated Low Power UWB LNA Using Self-Body-Bias Technique for 6.2−10.6 GHz Applications

Mohammad Nahidul Karim, Shahadat Hossain*[#], Simul Barua[#], Koushik Barua[#], Kazi Raisul Islam[#]

[#]Department of Electrical and Electronic Engineering, American International University-Bangladesh, Dhaka-1213, Bangladesh
*hossainshahadat.eee@gmail.com

***Abstract*— In this paper, a low power CMOS UWB low noise amplifier (LNA) suitable for 6.2−10.6 GHz application is presented. Using simultaneous noise matching (SNIM) and self-body-bias technique, the proposed LNA is designed to operate with low supply voltage in order to reduce power consumption. The simulated results showed that at 0.62V supply voltage, the LNA consumed 3.84mW with a maximum gain of 10 dB and average noise figure of 4.55 dB over the band of interest. Minimum port reflection parameters, S_{11} and S_{22} for the proposed amplifier were found <-10 dB whereas 1−dB compression point was found at -16.2 dBm.**

Keywords*— *Ultra Wideband; FCC; CMOS; RF; LNA

I. Introduction

Due to broad application scopes for ultra wideband (UWB) systems, radio frequency circuit designers are motivated to design unique, power efficient circuits that can facilitate such applications. According to FCC regulation released in 2002, 3.1−10.6 GHz band has been allocated for unlicensed wideband wireless communication. Moreover, recent IEEE 802.15.3 standard is focused on using two sub bands of the unlicensed band, 3.1−4.85 GHz and 6.2−9.7 GHz for high speed data transmission [1−2]. To support high speed data communication, the frontend of the receiver must be efficient enough to suppress the noise, minimize the return losses and provide sufficient gain within a low power budget.

Low Noise Amplifier (LNA) is the most crucial part of receiver frontend as the sensitivity of overall receiver end is dependent on it. Because of higher level of integrability, lower feature size, higher power efficiency and lower cost, CMOS technology has become a more obvious choice for on-chip radio transceiver implementation. But, designing of an LNA involves consideration of many trade-offs among gain, noise, linearity, return losses and power dissipation [3]. Various topologies for wideband and ultra wideband CMOS LNAs have been developed over the past few decades [1−2][4−11]. Among them, simultaneous input noise matching (SNIM) technique offers simultaneous noise and input matching [4]. Recently reported self-body-biasing has demonstrated ultra low power operation at low voltage supply [5]. Integration of SNIM technique with self-body-bias technique can be used to design an ultra low power UWB CMOS LNA that is capable of operating at low supply voltage.

In this paper, a low power; fully integrated UWB LNA using SNIM and self-body-bias technique for 6.2−10.6 GHz application in 0.18μm CMOS technology is proposed. In section II, the circuit analysis is introduced and the design equations are discussed. Section III provides the simulated results, demonstrates the effects of layout parasites on overall performances and draws the summary as well as comparisons of the result with several published works. The conclusion is presented in section IV.

II. Circuit Analysis

The equation for threshold voltage of an n−channel MOSFET can be written as [12]:

$$V_{th} = V_{th0} + \gamma\left(\sqrt{\left|2\varphi_f - V_{bs}\right|} - \sqrt{\left|2\varphi_f\right|}\right) \qquad (1)$$

Where, V_{th} is the threshold voltage, V_{th0} is the threshold voltage for body-source voltage, V_{bs}=0, γ is a process dependent parameter, φ_f is dependent on semiconductor. For a standard CMOS technology, V_{bs} adds an extra parameter to change V_{th}. Thus, a CMOS amplifier can be operated at a lower supply voltage by lowering V_{th} which can be achieved changing V_{bs}. But, this technique raises the need for extra biasing (body biasing) in the circuitry. If the body biasing is done within the circuit, it transforms self-body-bias technique. This technique reduces the need for extra bond pads, thus reducing the effective die area.

The proposed two-stage 0.18μm LNA circuit is design based on a self-body-biased degenerated common source common gate cascode structure using SNIM technique. The first stage, which is degenerated common source amplifier, performs the main amplification and cascoded common gate amplifier creates effective isolation between input and output. The second stage is a common source amplifier, which performs further amplification and designed to lower the return loss at output. Fig. 1 shows the proposed circuit schematic.

A. Input matching

In order to ensure minimum reactive part of the input impedance throughout the band of interest to achieve flat gain and optimum noise figure, a chebyshev filter is used at the input of the LNA shown in fig. 1. The gate voltage of degenerated common source MOS is provided within the circuit by connecting the drain terminal of cascaded common gate MOS by a resistance, R_{bias}. C_{bias} is provided for an ac

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

ground. So, the input impedance remains unaffected by inclusion of R_{bias} and C_{bias}. Considering the small signal model of an inductively degenerated LNA, the input impedance of the MOS transistor from node x [6]:

$$Z_{in}(s) = 1/sC_{gs} + s(L_s + L_g) + g_{m_1} L_s / C_{gs}$$
$$= 1/sC_{gs} + s(L_s + L_g) + \omega_T L_s \qquad (2)$$

where, g_{m1} is the transconductance of the M_1 transistor, C_{gs} is gate-source capacitance the real part of $Z_{in}(s)$, $\omega_T L_s$ is equal to the real part of source impedance R_s, where $\omega_T = g_{m1}/C_{gs}$. Size of M_1 was chosen for proper input impedance matching.

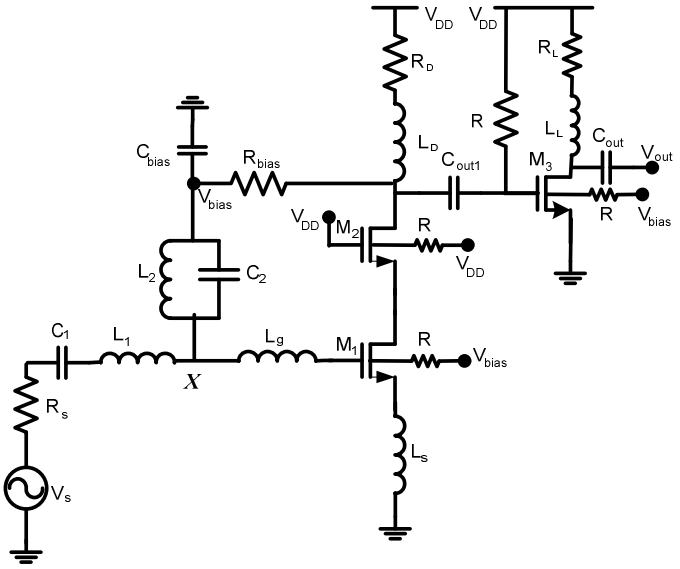


Fig. 1 Proposed circuit for ultra wideband LNA

B. Noise Figure Analysis

In SNIM technique [4], the optimum noise impedance, Z_{opt} can be shifted by varying the width of the MOS transistor so that, $Z_{opt}=Z_{in}^*=Z_s$ is achieved. This ensures both input and noise matching at the same time. Equation (2) defines the parameters on which Z_{in} is dependent. The width of the transistor, gate voltage and transconductance was chosen to match the aforementioned requirements. The noise figure is mostly affected by the noise contribution from the first amplifying stage whereas the noise originated from other stages is reduced by gain of previous stages.

C. Self-Body-Biasing and Self-Biasing

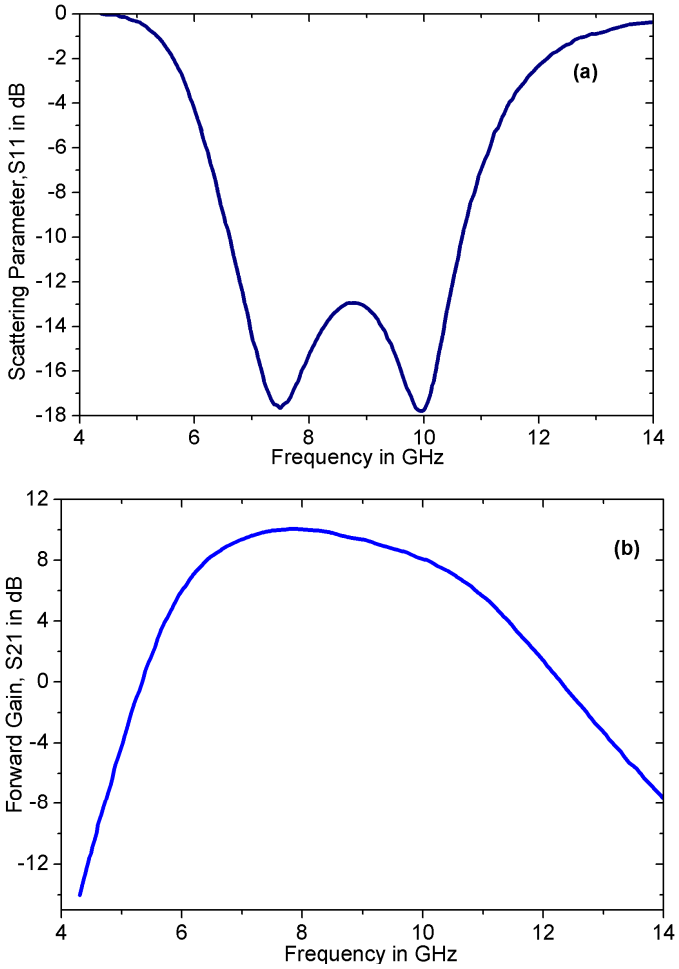
The proposed LNA architecture includes three NMOS transistors and all the transistors are forward body biased. Body biasing introduces two problems to be addressed; body biasing may introduce additional bondpads in the die and forward body bias which turns on the source to body junction that results in additional power consumption. That is why the proposed architecture has adopted self-body-bias technique, where all the forward body biasing is done through a resistance (R≈10KΩ). For most of design topologies proposed on LNA design [2][9-11], at least one of the MOS transistors is needed to be gate biased from external supply. The proposed design has biased all the MOS transistors inside the main circuit design. The degenerated common source amplifier's MOS transistor is biased through R_{bias}.

D. Output Impedance Matching

The filtered and amplified signal from cascade stage is fed to a second amplifying stage for further amplification. In fig. 1, the output is seen through capacitor, C_{out}(~1pF) . To reduce the quality factor of the inductor L_L, R_L in series with load inductor is used and widen the bandwidth of the signal received at output. The common source amplifier used in the second stage is biased through resistance, R. Capacitor C_{out1} is used to filter out any dc component from incoming signal to M_3 transistor.

III. Simulated Results and Comparison

The proposed UWB LNA has been designed in 0.18µm CMOS model parameters and simulated in Hspice RF. V_{bs1}, V_{bs2} and V_{bs3} were self forward biased to reduce the threshold voltage of M_1, M_2 and M_3 respectively. For a fixed bias current, lowering V_{th} results in lower supply voltage requirement thus lowering overall power consumption. Simulated results have shown that the proposed LNA consumes 3.84mW from a 0.62V supply voltage. The drain current drawn by cascode amplifying stage and the second amplifying stage have been found 2.06mA and 4.16mA respectively. The chebyshev filter used at input has ensured a better input impedance matching. Fig. 2 illustrates the input



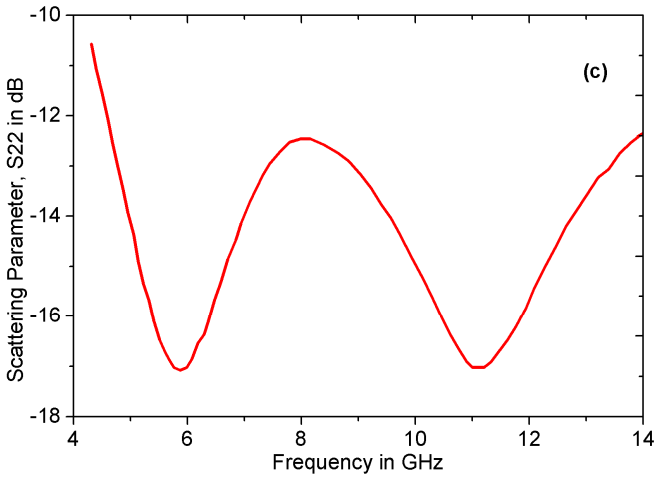


Fig. 2 Simulated results of scattering parameters (a)S_{11}, (b)S_{21} and (c)S_{22} in dB

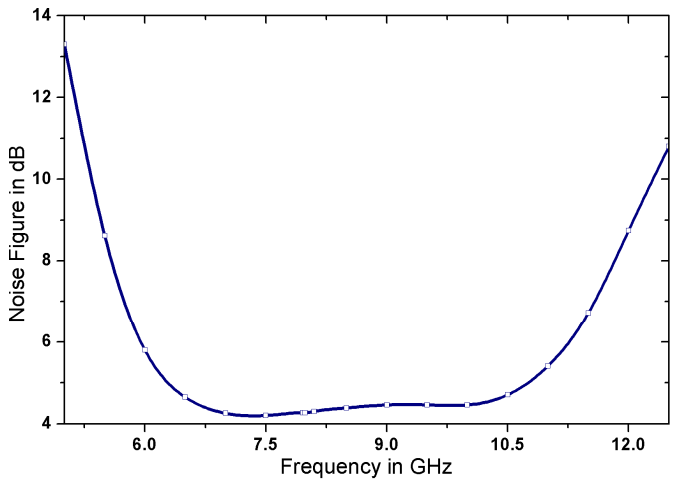


Fig. 3 Simulated results of Noise Figure

reflection coefficient (in Fig. 2(a)) in dB to demonstrate the input matching of the LNA in which S_{11}<-10 dB from 6.6−10.6 GHz. The maximum forward gain, S_{21} is achieved 10 dB at 7.98 GHz (in Fig. 2(b)). Output matching parameter, S_{22} is found <-13 dB throughout the operational band(in Fig. 2(c)). The circuit has shown better reverse isolation (<-35 dB over the whole frequency range). The simulated noise figure of the LNA is shown in fig. 3. The noise figure is minimum (4.20 dB) at 7.5 GHz and average noise figure is ~4.55 dB.

Fig. 4 depicts the range of linearity for the LNA. The input-port referred 1-dB compression point for the LNA is found -16.2 dBm. According to the harmonic balance analysis (simulation was performed in HSpice RF), the 3rd –order input referred inter-modulation product was found -6.77 dBm at 8 GHz. All the capacitor used in this design is metal-insulated-metal type capacitor (0.2p-1pF) where model parasites were considered. The inductors used in this design ranges between 0.4n-1.7nH. Parasitic resistances and capacitances of the spiral inductor model were also considered in the design. The simulated unconditional stability factor, μ for the proposed design was found greater than 1 from 5.5−12.5 GHz.

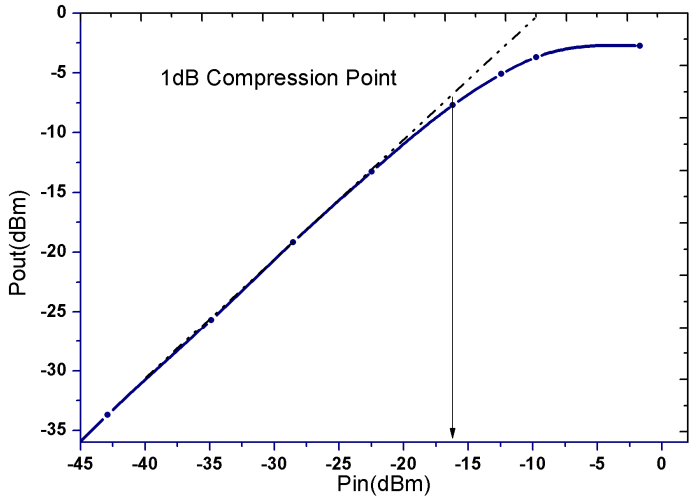


Fig. 4 Input referred 1-dB Compression Point

Table I summarizes the results and compares the design performances with several published UWB LNA designs.

TABLE I
SUMMARY OF THE PROPOSED LNA AND COMPARISON WITH OTHER PUBLISHED WORKS

Specification	This Work	[7]	[8]	[9]	[10]	[11]
Data	*Simulated*	*Measured*	*Simulated*	*Simulated*	*Measured*	*Measured*
Year	2014	2010	2007	2014	2012	2013
CMOS Process(μm)	0.18	0.18	0.18	0.09	0.18	0.18
Supply (V)	0.62	1.8	0.6	0.75	1.1	1.1
BW (GHz)	4.4	7.5	7.5	7	7.5	7.5
S_{21}(dB)	10	13.2	10	10-32	10.8	14.8
S_{11}(dB)	-10~-18.5	-20	-10	-10	-8.1	<-11.5
S_{22}(dB)	<-13	N/A	<-10	-15	-9.5	<-10.7
NF_{min}(dB)	4.20	4.5	3.8	.2	5.5	3.5
NF_{ave} (dB)	4.55	<5	4.65	1.1	>5.5	>3.5
ICP (dBm)	-16.2	-11	N/A	N/A	<-16	N/A
IIP3 (dBm)	-6.77	-1.4	1	1	-6.4	-13
P_{diss}(mW)	3.84	23	7	19.3	6.4	7.8
Operating Range (GHz)	6.2−10.6	3.1−10.6	2.8−9.1	1−11	3.1−10.6	3.1-10.6
FOM	7.9	0.53	5.02	5*	2.1	3.7

*calculation was done performance parameters seen at 7 GHz

The figure of merit (FOM) is defined here to evaluate the circuit performance as,

$$FOM = \frac{S_{21}(dB).Operating\ Freq(GHz)}{Power(mW).V_{dd}(Supply)NF_{\min}(dB)}$$

These results imply that the proposed 10dB, 3.84mW, UWB CMOS LNA is suitable for low voltage, high-gain and low power application.

IV. Conclusion

This paper presents an optimized low power self body biased UWB CMOS LNA. Simulated results illustrate the circuit performs better in terms of gain, input output matching and linearity. These results indicate that the proposed architecture is suitable for low power 6.2−10.6 GHz UWB applications.

Acknowledgment

Authors would like to thank the Department of Electrical and Electronic Engineering of American International University−Bangladesh for their co-operation and support.

References

[1] Mirvakili, A., Yavari, M.," A linear wideband CMOS LNA for 3-5 GHZ UWB systems," in International SoC Design Conference 2008 , pp. 150-153

[2] Dehqan A., Kargaran, E., Mafinezhad, K., Nabovati, H., "An Ultra Low Voltage Ultra Low Power CMOS UWB LNA Using Forward Body Biasing," IEEE 55th International Midwest Symposium on Circuits and Systems (MWSCAS), 2012, pp. 266-269

[3] T. H. Lee, The Design of CMOS Radio Frequency Integrated Circuits, Cambridge Univ., 1998

[4] Nguyen, T., Kim, N., Ihm, G., Yang, M., and Lee, S.,'' CMOS Low Noise Amplifier Design Optimization Techniques," in Transactions on microwave theory and techniques, 2012, pp. 1-10

[5] Chang J.-F., and Lin Y.-S.," 0.99 mW 3–10 GHz common-gate CMOS UWB LNA using T-match input network and self-body-bias technique," in Electronics letters, vol. 47, no. 11, May 2011.

[6] Bevilacqua, A., Niknejad, A., "An Ultrawideband CMOS Low Noise Amplifier for 3.1–10.6 GHz Wireless Receivers," in IEEE Journal of Solid State circuits, vol. 39, no. 12, pp. 2259-2268, December 2004.

[7] B. Park, S. Choi, S. Hong, "A low-noise amplifier with tunable interference rejection for 3.1 to 10.6-GHz UWB systems," IEEE Microwave and Wireless Components Letters, vol. 20, no. 1, pp. 40-42, Jan 2010.

[8] Yueh-Hua Yu, Yi-Jan Emery Chen, and Deukhyoun Heo, "A 0.6-V Low Power UWB CMOS LNA," in IEEE Microwave and Wireless Components Letters, , vol. 17, no. 3, March 2007

[9] Pandey, S.K, Agrawal, S., Singh, J., and Parihar, M.S.," A wideband LNA design based on body bias suitable for low voltage and low power application" in IEEE International Conference on Electronics, Computing and Communication Technologies (IEEE CONECCT), pp. 1-6, 2014

[10] Meng-Ting Hsu, Yu-Chang Hsieh ; An-Cheng Ou," Design of low power UWB CMOS LNA using RC feedback and body-bias technology" in IEEE 10th International Conference on ASIC (ASICON), China, pp. 1-4, 2013

[11] Meng-Ting Hsu, Yu-Hua Lin, and Jing-Cheng Yang," Design of UWB CMOS LNA based on current-reused topology and forward body-bias for high figure of merit" in Asia-Pacific Microwave Conference Proceedings (APMC), pp. 772-774, 2013

[12] B. Razavi, Design of Analog CMOS Integrated Circuits, Delhi: Tata McGraw-Hill Edition, 2002.

An Ultra Wide Band Low Noise Amplifier for 10.8 to 20.1 GHz Applications

Md. Saiful Arefin Mojumder*, Md. Sariful Islam, and M. Ziaur Rahman Khan
Bangladesh University of Engineering & Technology
Dhaka-1000, Bangladesh
Email: ujjalbuet@gmail.com

***Abstract*—Design and simulation of an Ultra Wide Band Low Noise Amplifier for 10.8 to 20.1 GHz applications is presented. A CG-CS topology is adopted in this work. CG is used for input matching in the whole band of operation. Output of the CG is fed to the CS amplifier input to increase the gain of the overall amplifier. The circuit is designed in IBM 130 nm CMOS Technology and simulated in HSPICE RF. The simulated results of the circuit shows 10.2 dB of forward gain (S_{21}) with a bandwidth of 9.3 GHz, -18 dB input return loss (S_{11}), reverse isolation (S_{12}) of -44 dB and -32 dB of output matching parameter (S_{22}) at center frequency, 14 GHz. This circuit exhibits a Nf of 5.03 dB at center frequency and 5.44 dB of average Nf. ICP is -17.2 dBm at the center frequency. It consumes only 7.38 mW when driven from 1.2 V source.**

Keywords*—Ultra-wide-band (UWB), low noise amplifier (LNA), noise figure (Nf), S-parameters, common gate (CG), common source (CS).

I. Introduction

With an increasing trend towards utilizing higher frequencies for wideband communication applications, Ku-Band (12 to 18 GHz) is of special interest as there are many important commercial and research applications in this frequency band such as: satellite communication for entertainment, high speed internet access [1], global snow-cover monitoring, weather prediction, air-borne scatterometers and DVB-S applications [2].

As the demand increases, the operating frequency of the transceiver system increases because the frequency band below 10 GHz is becoming more and more crowded [3]. GaAs, AlGaN materials are used in high frequency wireless system design for lower noise figure and higher frequency applications. But, the major drawback of those systems is cost. CMOS wireless system is used for lower cost, flexibility, and available technology.

Integral part of wireless system is the RF front end LNA (Low Noise Amplifier). As this is the first component, it should have higher gain, lower noise figure and lower area [4]. Moreover, higher is the gain provided by the LNA, lower will be the overall noise figure of the RF receiver. Bandwidth is also major concern for ultra-wide-band communication. In order for the receivers in a single chip to support multiple frequencies, the bandwidth of the receivers has to be very high. Apart from that, short distance data communication uses short pulse for communication. To support this type of communication, ultra wide-band receiver is the first requirement. Ultra-Wide-Band (UWB) communication is also a possible solution for wireless interconnects.

High frequency UWB systems are a new wireless technology capable of transmitting data over a wide frequency spectrum with very high data rates and very low power consumption [5]. For a wireless UWB systems, UWB LNA design is vital. This amplifier should have high gain, low noise figure and low power consumption. The gain and bandwidth of an amplifier cannot be simultaneously increased beyond a limit [6]. Hence, designing an amplifier circuit providing high gain and large bandwidth is quite a challenging task.

Several works have been reported in the frequency range of 12 to 18 GHz [1, 7–9]. Gil has reported a CMOS LNA operating in the 13 GHz region with a very low forward gain and relatively high power consumption [7]. A 14 GHz LNA reported by Kuo-Liang has moderate forward gain, but the power consumption is very high [9]. Meanwhile, Cheng-Chi Yu has reported a low voltage and low power LNA having a wide bandwidth in the 12 to 18 GHz frequency range at the expense of large number of stages [10]. All the above works are done using TSMC 180 nm technology. Very few works have been done using IBM 130 nm CMOS technology. However, use of 130 nm technology allows smaller chip area with low power consumption.

Here the design and simulation of a UWB LNA using IBM 130 nm CMOS technology is reported. A CG-CS structure is used to implement this design. This amplifier is designed especially for the applications utilizing frequencies in between 12 to 18 GHz.

II. Circuit Analysis

The proposed low noise amplifier has been shown in Fig. 1. The common gate stage provides an excellent and a relatively easy way to match input impedance for a wide range of frequency. For maximum power transfer from antenna to amplifier, the input impedance of the amplifier has to be 50 Ω. The input impedance for the common gate (CG) stage is nearly equal to $1/g_m$. Hence, in this stage, the values of width (w) and length (L) of the n-channel MOSFET, M_1 have been chosen so that the input impedance is very close to 50 Ω. The values of capacitance C_i (=500 pF) and inductor L_i (=160 fH) have ensured the input reactance to be resonated out at a

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

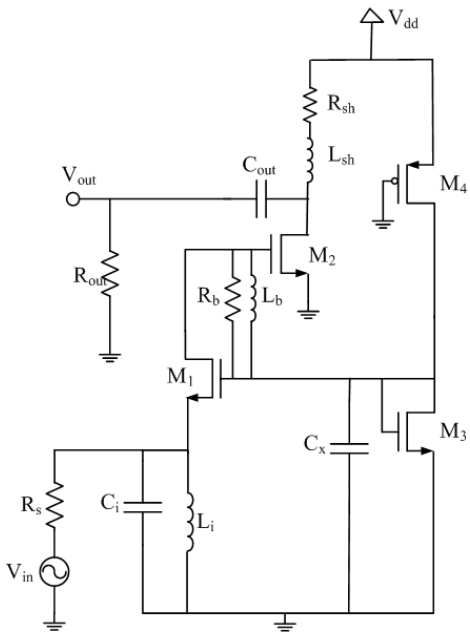


Fig. 1: Proposed LNA Circuit.

frequency not greater than $1/(2\pi\sqrt{L_i C_i})$ (=17.8 GHz) because of the presence of C_{gs}. Besides, the reverse isolation is better for this stage because of the absence of the C_{gd} path.

Since the gain of a CG amplifier is not sufficiently high compared with its inherent noise figure, a CS stage has been added at the output of the CG stage for providing a better amplification while adding very little to the noise figure. The value of inductance L_b (=1.7 nH) has been chosen so that it resonates with the parasitic capacitances ($C_{gs} = 51$ fF and $C_{gd} = 15.87$ fF) of the n-channel MOSFET, M_2 at a frequency which is equal to the resonant frequency obtained at the CG input stage. Here the resonant frequency is approximately 14.9 GHz.

The n-channel MOSFET, M_3 and the p-channel MOSFET, M_4 have been used to ensure the proper biasing of M_1 and M_2. The capacitance, C_x reduces the effect of the gate to source parasitic capacitance of M_1.

The load is designed to achieve considerably flat and large gain. The load is a shunt peaking load. To increase the gain, load impedance has to be high. L_{sh} is used to reduce the capacitive effect at the output node. Large inductor means higher quality factor, which will decrease the bandwidth of the amplifier. So, L_{sh} is a compromise in between large gain and bandwidth. R_{sh} is introduced to control the quality factor of the inductor. High value of R_{sh} ensures lower quality factor for higher bandwidth. Higher value of R_{sh} reduces the load impedance and finally reduces the gain of the amplifier.

III. Simulation and Results

This proposed circuit is designed in IBM 130 nm CMOS 8RF process. The circuit is simulated in HSpice RF. Variation of forward gain, S_{21} is shown in Fig. 2. S_{21} should be as high

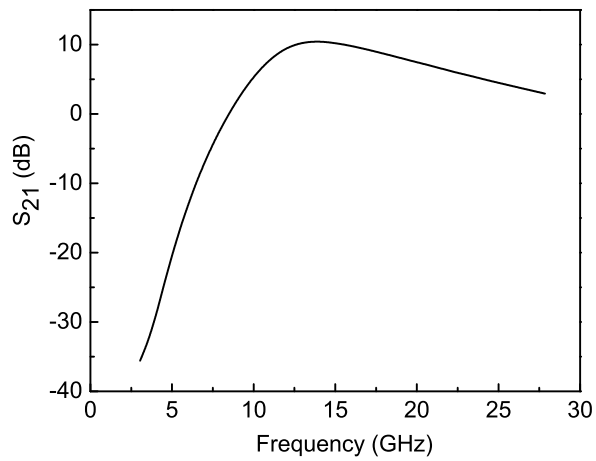


Fig. 2: Forward Gain vs. Frequency

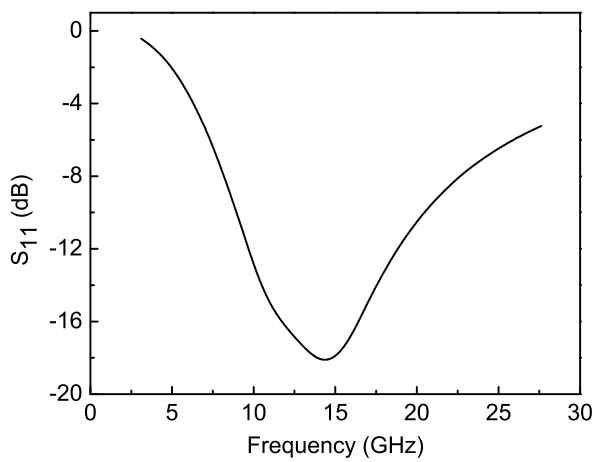


Fig. 3: Input Impedance Matching (S_{11}) vs. Frequency

as possible for an amplifier. The peak forward gain of the LNA is 10.2 dB at a frequency of 14 GHz. It is obvious from the Fig. 2 that 3-dB bandwidth of the amplifier is 9.3 GHz. The wide band impedance matching achieved by using the CG portion at the input followed by a high gain provided by the CS portion is responsible for such a wide bandwidth. The input return loss, S_{11} should be as low as possible. Generally, S_{11}, below -10 dB, ensures sufficient input impedance matching.

As seen from the Fig. 3, S_{11} is -18 dB at 14 GHz. Apart from that, S_{11} stays below -11 dB in the entire bandwidth. The reverse isolation, S_{12} has to be very low. Reverse isolation characterizes the leakage of power from output node to input node. A very low value of S_{12} indicates that the amplifier is stable in the entire bandwidth. S_{12} is below -40 dB in the entire BW for the present work as seen from in Fig. 4. Any value of output impedance matching parameter, S_{22} less than -10 dB is preferable. S_{22} is -28.1 dB at 14 GHz as indicated by Fig. 5. Moreover, S_{22} remains below -10 dB in the entire bandwidth.

The noise figure (Nf) is 5.03 dB at 14 GHz as seen from

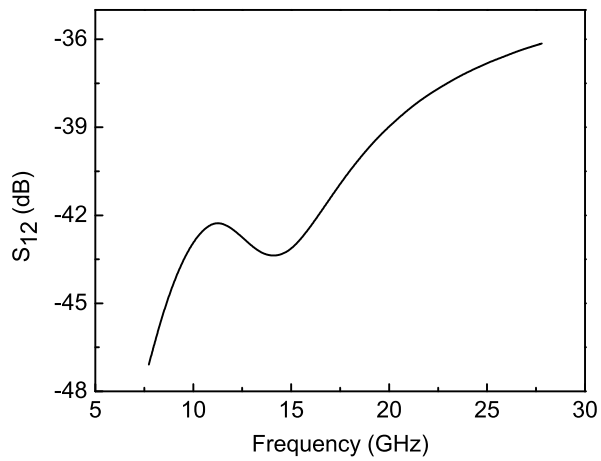


Fig. 4: Reverse isolation vs. Frequency

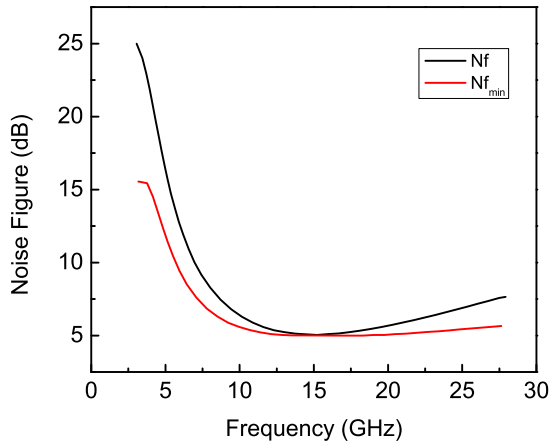


Fig. 6: Noise Figure vs. Frequency

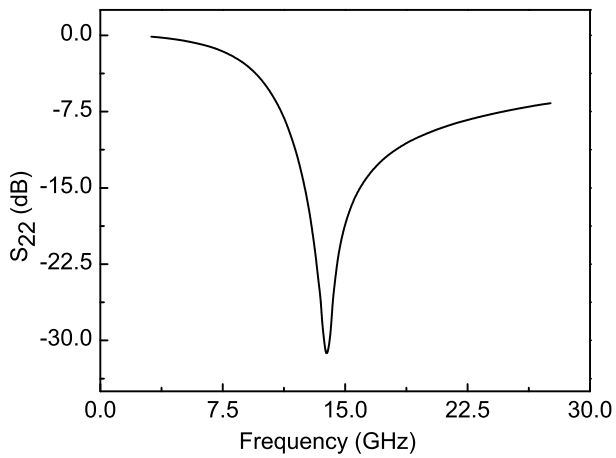


Fig. 5: Output Impedance Matching vs. Frequency

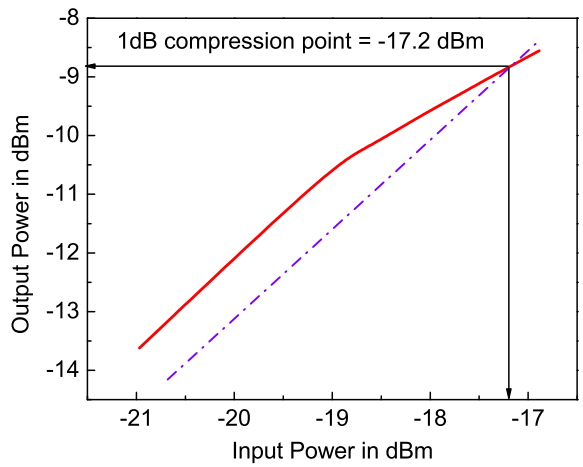


Fig. 7: Output Power vs. Input Power

Fig. 6. However, for such ultra wide band applications, an average noise figure, 5.44 dB is better figure of merit than spot noise figure. Fig. 7 depicts that the 1-dB compression point is -17.2 dBm.

IV. Comparison with Previously Reported LNAs

Table I presents a thorough comparison of this work with different reported state-of-the-art LNAs. All but the proposed LNA in this work have been designed using 180 nm CMOS technology. Apart from ref. [11], all other previously reported amplifiers have more than one stage. Infact, ref. [12] (cascade) and [10] have used three stages to provide sufficient gain as well as input and output matching. The bandwidth that can be utilized by the LNA proposed here, far outnumbers all reported LNAs. The reverse isolation of this LNA is also very impressive when compared with the rest. The power consumption is less when compared with the most of the designs. The output return loss (S_{22}) of the designed LNA is also less than the reported LNAs.

TABLE I: Comparison of Reported CMOS LNAs

Ref.	Freq. (GHz)	BW (GHz)	S_{21} (dB)	S_{11} (dB)	S_{12} (dB)	S_{22} (dB)	P_{DC} (mW)
[11]	10	4.5	14	-10	-25	-13	15
[12] (cascade)	13	<4	14.48	-18.15	na	-16.52	24.61
[12] (cascode)	13	<3	14.73	-18.68	na	-21.56	5.82
[4]	15	6	9.72	-11.05	-22.56	-10.2	10
[10]	15	6	13.83	< -10	< -30	> -10	7.6
This work	14	9.3	10.2	-18	-44	-32	7.38

Nonetheless, all those boons have been achieved at the expense of slightly higher noise level. Apart from that, the 1-dB compression point of this amplifier is slightly less than that of the other reported LNAs. Nonetheless, the 1-dB compression point evaluated for the proposed LNA closely matches with that of ref. [10]. However, this LNA can be used up to a reasonable power level of the input signal.

V. Conclusion

The design and performance of an UWB Low Noise Amplifier for the applications ranging for 10.8 to 20.1 GHz has been analyzed in this work. Comparing with the other state-of-the-art LNAs, this amplifier has higher bandwidth with superior input and output matching. Due to the lack of available data, this design cannot be compared with 130nm technology. The presented design has center frequency of 14 GHz with a bandwidth of 9.1 GHz ranging from 10.8 to 20.1 GHz. The use of CG-CS configuration permits wideband input impedance matching along with sufficient gain. The forward gain is 10.2 dB and the reverse isolation is less than -35 dB at the center frequency of 14 GHz. This LNA takes the advantage of less power consumption (7.38 mW) and higher bandwidth simultaneously. Hence, LNA is suitable for low power applications of 10.8 to 20.1 GHz.

Acknowledgment

The authors would like to thank IBM for providing the 130 nm CMOS library files and Professor A.B.M. Harun-ur-Rashid of the Department of EEE, BUET for his support.

References

[1] B. Afshar, Ali M. Niknejad, "X/Ku Band CMOS LNA design techniques", *IEEE Custom Integrated Circuits Conference*, pp. 389 - 392, 2006.

[2] Wan-Rone Liou, Siddarth Rai Mahendra, and Tsung-Hsing Chen, "A Wideband LNA Design for Ku-Band Applications", *International Conference on Communications, Circuits and Systems*, Chengdu, China, pp. 680 - 684, 2010.

[3] Yiqun Cao, Vadim Issakov, Marc Tiebout, "A 2kV ESD-Protected 18GHz LNA with 4dB NF in 0.13m CMOS ", *IEEE International Solid-State Circuits Conference*, San Francisco, CA, USA, pp. 194 - 606, 2008.

[4] S. M. Shahriar Rashid, Apratim Roy, Sheikh Nijam Ali & A. B. M. H. Rashid, "A 23.5 GHz Double Stage Low Noise Amplifier Using .13m CMOS Process with an Innovative Inter-Stage Matching Technique", *5h International Conference on Wireless Communications, Networking and Mobile Computing, WiCom '09*, Beijing, China, pp. 1 - 4, 2009.

[5] Andrea Bevilacqua and Ali M. Niknejad, "An Ultrawideband CMOS Low-Noise Amplifier for 3.110.6-GHz Wireless Receivers, *IEEE J. Solid-state Circuits*, Vol. 39, No. 12, pp. 2259 - 2268, 2004.

[6] H. W. Bode, Network Analysis and Feedback Amplifier Design. NewYork: Van Nostrand, 1945.

[7] Gil J., Kwangseok Han, Hyungcheol Shin, "13 GHz 4.67 dB NF CMOS low-noise amplifier", *Electronics Letters* , vo1. 39, no. 14, pp. 1056- 1058, 2003.

[8] Chung-Yu Wu, Shahroury, F.R., "A Low-Voltage CMOS LNA Design Utilizing the Technique of Capacitive Feedback Matching Network", *Electronics, Circuits and Systems, 2006. ICECS '06. 13th IEEE International Conference on* , pp. 78-81, 2006.

[9] Kuo-Liang Deng et al., "A Ku-band CMOS low-noise amplifier", *Proc.RFIT*, pp. 183 - 186, 2005.

[10] Cheng-Chi Yu, Jiin-Hwa Yang, Lien-Chi Su, Heng-Yi Lee, and Chang-Chih Chen, "A Low-voltage and Low-power Low Noise Amplifier for Ku-band Application", *PIERS Proc.*, Suzhou, China, pp. 538 - 540, 2011.

[11] Sen Wang, "Design of a 9 GHz CMOS Low Noise Amplifier Using Gain-Enhanced Technique", *Microwave and Optical Technology Letters*, vol. 53, issue 3, pp. 479 481, 2011.

[12] Yu-Lin Wang & Man-Long Her, "Design and Implementation of the Cascade and Cascade Low Noise Amplifiers for 13 GHz", *Cross Strait Quad-Regional Radio Science and Wireless Technology Conference*, Harbin, China, pp. 625 - 629, 2011.

ELC Resonator Based Narrow Band Metamaterial Absorber

Md.Washim Zayed, Md.Shafuzzaman and Md.Osman Goni
Department of Electronics and Communication Engineering
Khulna University of Engineering and Technology, (KUET)
Khulna-9203, Bangladesh
washimzayed@gmail.com, shafuzzamankuet@gmail.com and osmanhku@gmail.com

Abstract— **The aim of this paper is to construct an absorber using metamaterial structure which operates at narrow band. The proposed structure is simulated using Microwave studio by CST. Simulations show that S11 of our first structure is -15.20 dB and the absorbance is 97.74% in the resonant frequency at 5.962 GHz and for second structure absorption occurs at 5.296 GHz with absorbance of 99.859% and S11 of -28.243 dB. The frequency of the proposed structures where absorption occurs is reduced significantly as compared to existing structure, but can be more efficient in absorbance.**

Index Terms—**Metamaterial absorber; ELC resonator; absorbance**

I. Introduction

Electromagnetic (EM) metamaterials (MTMs) are usually artificial composite materials consisting of periodic structural units much smaller than the wavelength of the incident radiation are designed to usually gain their properties from their structure and that are not found in nature [1].

A metamaterial absorber manipulates the loss components of the complex effective parameters, electric permittivity and magnetic permeability of metamaterials, to create a high electromagnetic absorber.

To obtain high absorption through resistive and dielectric loss, MTMs can be characterized by electric permittivity and magnetic permeability which can be minimized both transmittance and reflectance simultaneously at resonance frequency [6]. Initially metamaterials absorber was structured by Split ring resonator (SRR) [3], but Electric field driven (ELC) resonating structures [4] are now extensively used as absorbers since EM waves travel lesser distance in ELC compared to the conventional SRR based structure [8].

The coupling between metamaterial structures and electromagnetic radiation via capacitance and inductance suggests that it is to change the resonance frequency also refraction index, and electromagnetic energy may also be stored and to yield a large absorption. Since permeability and permittivity are characterized by a real part (which describes propagation in a medium) and an imaginary part (which describes absorption and loss in a medium), perhaps it is possible to manipulate the imaginary component of ε and μ to be simultaneously large over a given frequency range and thus absorb radiation well. The approach is similar to the one used for negative index metamaterials, to optimize the imaginary portion of the permittivity and permeability. In 2008, Landy et. al introduced a metamaterial structure with high effective absorption over a narrow band in the microwave range (10-12.5 GHz) and near unity absorption at 11.5 GHz [7]. This was accomplished with a metamaterial structure consisting of a silver double split ring resonator (similar to the electrically resonant structures described in [8]). The electrically resonant split rings couple to the electric field without any magnetic interaction, while the magnetic response was a product of the induction current driven between the cut wire and the center conductor of the split ring resonator. The resonance that is established between the inductive and capacitive portions of the "circuit" allows energy to be stored and subsequently dissipated via ohmic and dielectric loss.

The ELC resonator can be described qualitatively in terms of its equivalent circuit [e.g. Fig. 2(b)]. A capacitor-like structure couples to the electric field and is connected in parallel. Two capacitive loops are connected in parallel, so that the resonant frequency of the circuit model is $\omega=\sqrt{(2/(LC))}$ [10]. This allows the electric field to drive the LC resonance providing both positive and negative electric polarization at different frequencies along the resonance curve.

To get a high absorption metamaterial structure need to be much smaller dimension than the wavelength. The performance of the metamaterial absorber is considered by bandwidth and the sensitivity to the incident electric polarization. Previously split ring resonator (SRR) based metamaterirals structure is used for the absorber which is travel at the long distance, Landy et al. has preferred metamaterial absorbers using the Electric Field Driven LC (ELC) resonator structure where the wave has to travel much shorter distance [7].

In here compact ELC driven metamateral absorber structures has been proposed, where the resonance frequency reduces to 5.296 GHz and 5.962 GHz from 6.46GHz [2]. In this paper, we consider absorbance (e.g. sensitivity) of the material in a particular bandwidth where the 2nd proposed structure shows 99.859% absorbance which is more efficient than the existing structure [2].

II. Design

The design of these metamaterial structures is based on simultaneous resonant excitation of an electric dipole and magnetic dipole at a particular bandwidth. The structures consist of a tri layer. Top layer, on which radiation is incident is a structured metallic layer created by copper sheet, separated from the bottom by an intermediate dielectric layer of FR4 substrate. The structure is backed by copper sheet. When EM wave is incident on the structure, the electric field is coupled with the top ELC structure [3], which can control the value of electric permittivity $\varepsilon(\omega)$ and the anti-parallel current between the top layer and the metal ground plane can

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

be coupled with the magnetic field which controls the magnetic permeability $\mu(\omega)$ of the resonating structure [9].

The frequency dependent reflectance [R(ω)] and transmittance [T(ω)] were obtained from the S-parameters in the simulation package and the absorbance was calculated as, $A(\omega)= 1-R(\omega)-T(\omega)$.where $A(\omega)$, $|S11(\omega)|^2$ and $|S21(\omega)^2|$ are the absorbance, reflected power and transmitted power respectively at an angular frequency ω. We can see that, if we want to get the perfect absorption, (e.g. $A(\omega) = 1$). We should make both $T(\omega)$ and $R(\omega)$ equal to 0. However, it is much difficult to get the two parameters to be 0 simultaneously. To simplify the design, here we adopt a ground plane which filled with copper to make sure that the transmission is 0 during the simulation. Therefore, we only need to minimize the reflection in the absorber design.

The absorbance equation can also be defined as,

$$A(\omega) = 1 - |S11|^2 - |S21|^2 \quad (1)$$

Here, S21 is zero because of the copper ground plane, so equation (1) can be written as,

$$A(\omega) = 1 - |S11|^2 \quad (2)$$

So the absorbance of our structure is only dependent on |S11|.

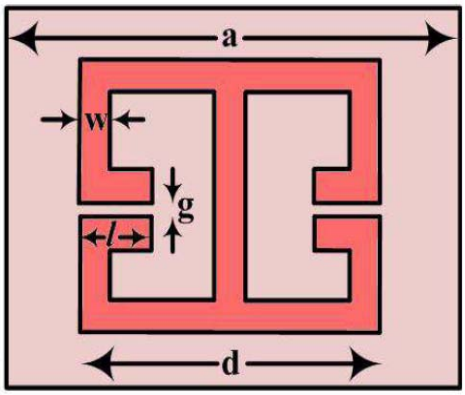


Fig. 1 Front side of the Reference structure.

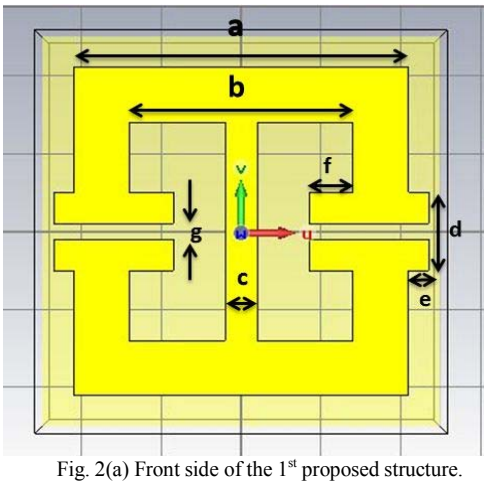


Fig. 2(a) Front side of the 1st proposed structure.

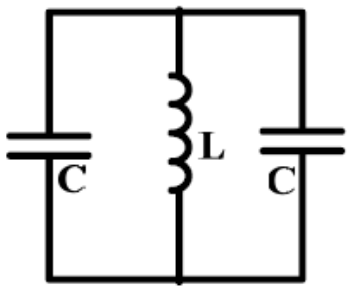


Fig. 2(b) Equivalent L-C circuit of 2(a)

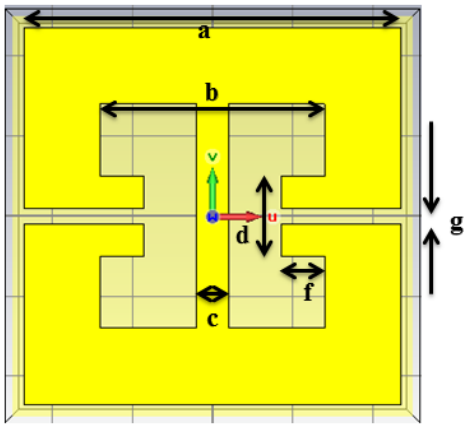


Fig. 3 Front side of the 2nd proposed structure

TABLE I

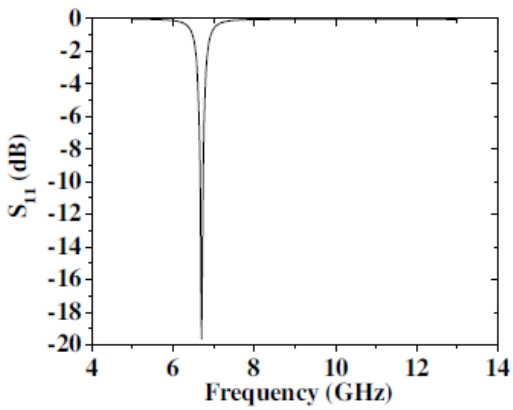
DIMENSIONS OF THESE STRUCTURES

Symbol	Dimension (Fig.2)	Dimension (Fig.3)
a	4.2(mm)	4.7(mm)
b	2.8(mm)	2.8(mm)
c	0.4(mm)	0.4(mm)
d	1.0(mm)	1.0(mm)
e	0.25(mm)	NA
f	0.55(mm)	0.55(mm)
g	0.2(mm)	0.2(mm)

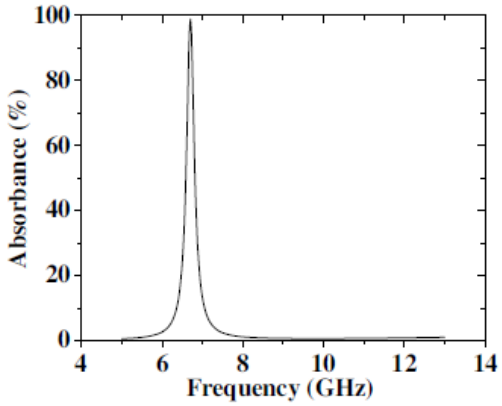
III. Simulated Results

We performed computer simulations of proposed metamaterial absorber structures using the commercial transient solver Microwave Studio by CST. A unit cell was simulated using boundary conditions H(t)=0 in Xmin and Xmax direction and E(t)=0 in Ymin and Ymax direction. We used FR4 substrate with square of 5mm edge and the copper layers have the thickness 0.035 mm.

Somak Bhattacharyya and Kumar Vaibhav Srivastava [2], showed for their structure an absorbance of 99.2% at 6.70 GHz using HFSS software.

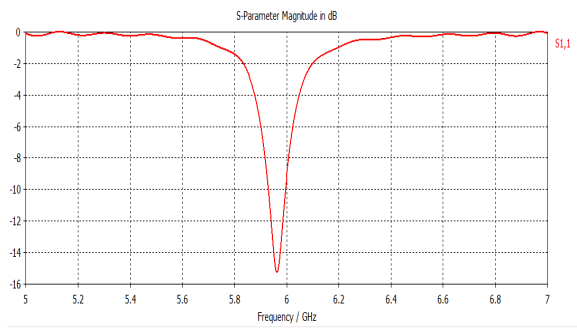


4(a)

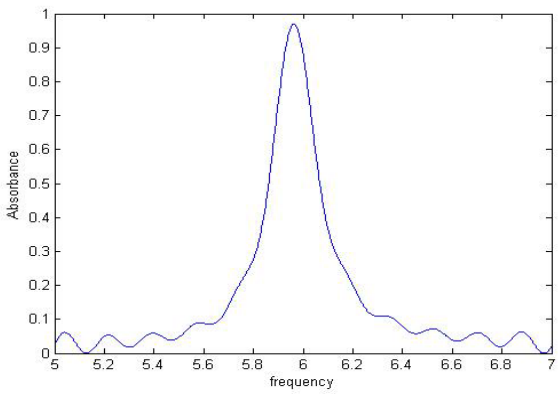


4(b)

Fig.4 (a) s11 plot and (b) absorbance at 6.46 GHz is about 99.2% of the reference structure in [2].



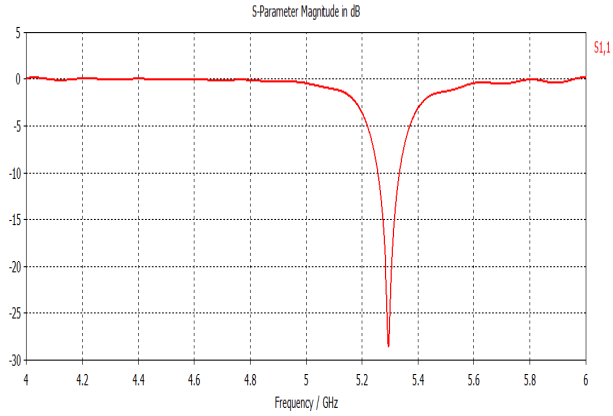
5(a)



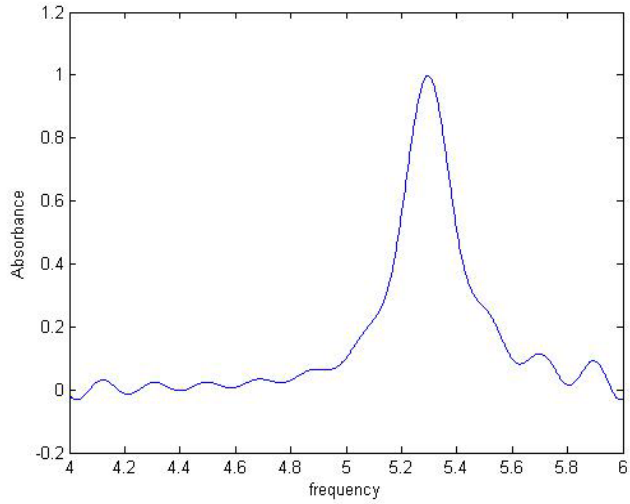
5(b)

Fig.5 (a) s11 plot and (b) absorbance at 5.962GHz is about 97.74% of the 1st proposed structure.

The simulation of the 1st proposed structure shows S11 of -15.20dB and the absorbance is 97.74% at the resonant frequency of 5.962GHz. The absorbance curve (fig. 5(b)) is plotted using MATLAB.



6(a)



6(b)

Fig. 6 (a) s11 plot and (b) absorbance at 5.296GHz is about 99.859% of 2nd proposed structure.

The simulation of the 2nd proposed structure shows a spike in the plot of S11 to -28.243 dB and the corresponding absorbance value A(ω) of 99.859% at the resonant frequency 5.296 GHz as shown in fig. 6(a). The absorbance value is calculated from Eq.(2) and the MATLAB plot is shown in fig 6(b).

The resonant frequency of our structure changed due to the change in capacitance as a result of change in the structure.

IV. Conclusions

We present two metamaterial absorber structures that works at narrow frequency band and the first structure shows absorbance of 97.74% in the resonant frequency 5.962 GHz and for second structure absorption occurs at 5.296 GHz with absorbance of 99.859% which is more efficient than the existing structure in [2]. The structure can be used for battle field and airborne radar applications [11].

REFERENCES

[1] Veselago, V. G., *"The electrodynamics of substances with simultaneously negative value of ε and μ"*, Sov. Phys.Usp., Vol. 10, No. 4, 509-514, 1968.

[2] Somak Bhattacharyya and Kumar Vaibhav Srivastava, *"Ultra-Thin Metamaterial Absorbers Using Electric Field Driven LC(ELC) Resonator Structure. "* PIERS Proceedings, Kuala Lumpur, MALAYSIA, March 27-30, 2012.

[3] Saptarshi Ghosh#1, Debdeep Sarkar#2, Somak Bhattacharyya#3, Kumar Vaibhav Sriavstava,*"Design of an Ultra-Thin Dual Band Microwave Metamaterial Absorber. "* ATMS INDIA 2013.

[4] Landy, N. I., S. Sajuyigbe, J. J. Mock, D. R. Smith, and W. J. Padilla, *"Perfect metamaterial absorber."* Physical Review Letters, Vol. 23, 207-402, 2008.

[5] Li, M. H., L. H. Yang, B. Zhou, X. P. Shen, Q. Cheng, and T. J. Cui, *"Ultrathin multiband gigahertz metamaterial absorbers."* Journal of Applied Physics, Vol. 110, 014909, 2011.

[6] S. Gu, J.P. Barrett, T.H. Hand, B.I. Popa, S.A. Cummer, *"A Broadband low-reflection metamaterial absorber,"* Journal of Applied Physics *108*,064913 (2010).

[7] Landy, N. I., et al. *"Perfect Metamaterial Absorber."* Physical Review Letters. Vol. 100, No. 207402. 23 May 2008.

[8] Padilla, W. J., et al. *"Electrically Resonant Terahertz Metamaterials: Theoretical and Experimental Investigations."* Physical Review B. Vol. 75, No. 041102(R). 15 January 2007.

[9] F. Sepideh, F. Keyvan and A. Zahra, *"A Polarization-Insensitive Metamaterial Absorber with a Broad Angular Band," 20th Iranian Conference on Electrical Engineering, (ICEE2012)*, May 15-17, 2012,Tehran, Iran.

[10] D. Schurig,a_ J. J. Mock, and D. R. Smith, American Institute of Physics. "*Electric-field-coupled resonators for negative permittivity metamaterials*" 24 January 2006.

[11] *Federal Radar Spectrum Requirement*, U.S. Department of Commerce, National Telecommunications and Information Administration, May2000.

A New Approach to Design Microstrip Low Pass Filter Using Novel Defected Ground Structure

G. M. Rafiquzzaman, * Abu Talha Sadi, and Md. Nurunnabi Mollah
Department of Electrical and Electronic Engineering,
Khulna University of Engineering and Technology
Khulna, Bangladesh
*kuet0703017@live.com

***Abstract* - A microstrip transmission line on a Binomial distributed hybrid defected ground structures (DGS) with interleaving photonic bandgap structures (PBGS) are proposed. Four different approaches are utilized to realize the better performance of the low pass filter (LPF). In all these cases, the length, width or both is made non-uniform in accordance with the co-efficient of Binomial distribution. It can be seen that, by applying Binomial distribution on DGS and interleaving PBGS the performance of LPFs are significantly improved.**

Index Terms* - *Defected ground structure (DGS), photonic bandgap structure (PBGS), low pass filter (LPF), Binomial distribution.

I. Introduction

The modern microwave communication system requires high performance, compact size and low cost devices. To fulfil the requirement and to enhance performance some new structures like Photonic Bandgap Structure (PBGS), Defected Ground Structure (DGS), Substrate integrates waveguide (SIW) etc. are existing on the communication system. In 1987, Yablonovitch and John proposed PBG [1, 2] which utilizes metallic ground plane properly and breaks the traditional microwave circuit fixed design to surface components and distributions of the medium circuit plane [3]. Recently many researchers have paid more attention to photonic bandgap structures (PBGSs) due to their unique properties. PBGSs are photonic analogous to semiconductor and the electromagnetic (EM) waves are impeded in PBG perturbed substrates as photons are impeded in the semiconductor. Due to their periodic nature, they are called electromagnetic bandgap structure (EBGS). The advantages of PBG structured components and devices have the ability of creating wider stopband which is very useful for suppressing the surface wave, spurious transmission and leakage in waveguides, striplines and filters [4]-[6].

Several DGS configurations have been described in this paper. MD. N. Mollah [7] proposed two 1-D hybrid DGSs for the low pass filter in the ground plane of a microstrip transmission line. He reported the optimum filling factor (FF) value as 0.46 to generate a very low ripple passband and wide rejection band. FF is defined as the volumetric ratio of one PBG element to one unit cell [8]. In this paper, we have reported that the performance of a microstrip transmission line having non uniform distribution (Binomially distributed) of DGS at a single frequency.

In this paper, we have focused on different binomially distributed DGSs at different FFs. In Binomial distributions, the amplitudes of PBG elements are non-uniform [9]. Therefore, FF is not a single valued parameter. In our study, we have defined the FF as the ratio of central DGS element and inter-element spacing. This research conveys the message of optimum FF at about 4 GHz cutoff frequency. Binomially distributed DGSs with optimum FF are applied to unwanted insertion loss in the band rejection region. Sufficient experimental investigations will be reported on the variation of number of DGSs and interleaved PBGSs in order to observe the effect on the low pass characteristic in a LPF.

Finally, comparing the insertion loss (IL) and return loss (RL) performances of the LPFs with Binomially distributed DGSs and hybrid DGSs, Binomially distributed DGSs and hybrid DGSs provide better performance than uniformly distributed DGSs in terms of suppression of the ripple in the passband and return and insertion-loss bandwidths.

II. Theory

Binomial Distribution

Binomial distribution is the determination of the probability distribution for the discrete number of successes in an independent sequence of experiments. The coefficients of the polynomial are determined by the expression (1).

$$(1+x)^{N-1} = 1 + (N-1)x + \frac{(N-1)(N-2)}{2!}x^2 + \frac{(N-1)(N-2)(N-3)}{3!}x^3 + \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

The coefficients of the series expansion for different values of N are expressed in terms of the Pascal's triangle [6]. N represents the number of EBG elements at the ground plane, then the amplitudes of the different EBG elements are calculated from the coefficients. We have taken seven (7) EBG elements in our designs by using the coefficient of the Pascal's triangle.

$N = 7$	1	6	15	20	15	6	1

Fig. 1 Binomial coefficient in Pascal's triangle

The proportional amplitude coefficients (P) are as follows:
$P_4=0.05$ $P_3=0.30$ $P_2=0.75$ $P_1=1$ $P_2=0.75$ $P_3= 0.30$ $P_4=0.10$

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

These values are calculated from Pascal's triangle for $N = 7$.

III. Proposed Designs

The conventional PBGSs are connected by narrow vertical slots to form dumbbell shape periodic structures. This is known as dumbbell shaped DGS [10] which is shown in Figure 2. This is constructed by etching defected areas on the ground plane and a connecting slot between defected rectangles. The equivalent circuit of DGS consists of the additional inductance and capacitance, and the equivalent inductance is dominant over the capacitance. It is much easier to establish the equivalent circuit because the transmission, reflection, and suppression characteristics of DGS microstrip line are very clear and definite [9]. The arm length of the larger rectangle is 'b' and width is 'a'. The vertical slot is rectangular with 'w' × 'g' dimension; 'w' is known as width and 'g' is known as gap.

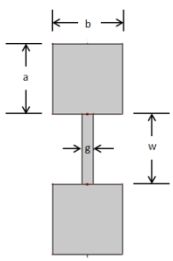


Fig. 2 Geometry of a unit cell of a dumbbell shape DGS.

Bragg's condition [11] is used to calculate the dimension of the uniform PBGS. By using dumbbell shaped DGS, we have proposed three (3) types of designs based on Bragg's condition to improve the performance of the LPF. The following designs are explained below one by one.

A. Design 1

In design 1, we have used Taconic substrate having dielectric constant 2.45 and height of 31 mils. The geometry of LPF designed by non-uniform dumbbell shaped DGS is shown in figure 3. The PBGS elements of the DGS are square in size and each PBGS has both arm's length is calculated to be 205 mils corresponding to FF=0.5 to analyse S-parameter performance. The non-uniform dumbbell shaped DGS elements have the period of 411 mils. The length and width of the narrow slot are 30 and 200 respectively. For the width of the narrow slot of non-uniform dumbbell shaped DGS, we have used the coefficient of the Binomial distribution for 7 element DGSs. We have applied Binomial distribution on the length (b) of every DGS. For N=7 the coefficients are 1, 0.75, 0.3, 0.05 and the length (b) of the PBGSs are taken 205, 153.75, 61.5, 6.15 mils. The FF of the central element is 0.5 for both PBGS and narrow slot of the DGS. The design is shown in Figure 3.

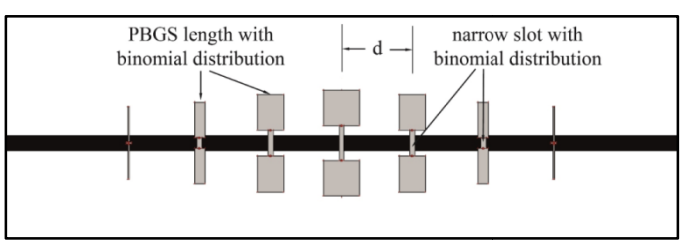


Fig. 3 Geometry of a Binomially distributed LPF designed by varying narrow slot width (w) and length (b) of PBGS of dumbbell shaped DGS etched in the ground plane. The period of DGS is 'd' and the substrate is Taconic with dielectric constant= 2.45 and height= 31 mils.

B. Design 2

In design 2 we have kept the area of narrow slot constant (30×100). The FF of the central element is 0.5 for PBGS of the DGS. The other design pattern is similar as design 1. The design is shown in Figure 4.

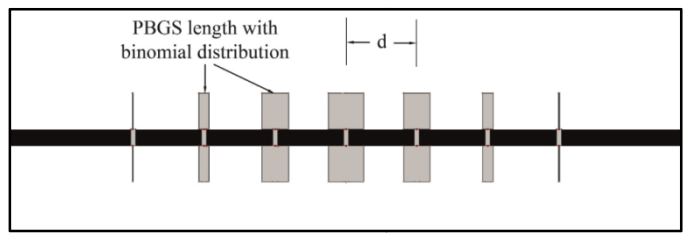


Fig. 4 Geometry of a Binomially distributed LPF designed by varying length (b) of PBGS of the dumbbell shaped DGS etched in the ground plane. The period of DGS is 'd' and the substrate is Taconic with dielectric constant= 2.45 and height= 31 mils.

C. Design 3

In design 3, the dumbbell shaped DGS pattern is followed by design 2. But the square EBGSs have been interleaved that is 1-D in nature and the square PBGSs are reduced by Binomial distribution. The arm lengths of square PBGS are taken as 200, 120, 40, 6 mils for the coefficients of 1, 0.6, 0.2, 0.03 for N=8 as of Binomial distributions. With a view to generating low pass filter (LPF), this interleaved PBGS will be useful. The combined design of the PBGS - DGS is shown in Fig. 5.

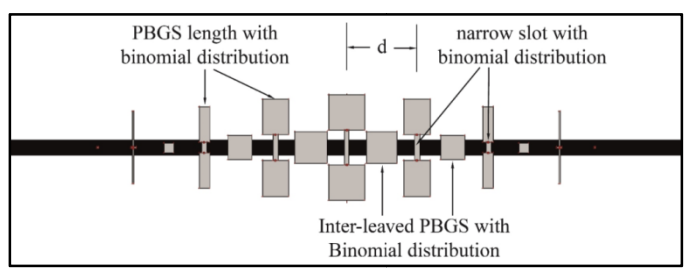


Fig. 5 Geometry of a Binomially distributed LPF designed by varying both length (b) of PBGS and width (b) of the narrow slot of the dumbbell shaped DGS and interleaved square PGBS etched in the ground plane. The period of dumbbell shaped DGS is 'd' for both DGS and interleaved PBGS and the substrate is Taconic with dielectric constant= 2.45 and height= 31 mils.

D. Design 4

The design 4 is different from previous designs. The widths (a) of PBGSs and the width of narrow slot of the dumbbell

shaped DBGs and lengths of the interleaved PBGSs are changed by Binomial distribution. The widths (a) of PBGSs of dumbbell shaped DGSs are 200, 150, 60, 10 mils with lengths (b) were kept constant. The lengths of those EBGS are 205, 153.75, 61.5, 6.15 mils. For PBGS of dumbbell shaped DGS the filling factor of the central element has kept 0.48 and for interleaved PBGS the filling factor has kept 0.5. The narrow slots have designed similar as Design 1. The combined design of the PBGS - DGS is shown in Fig. 6.

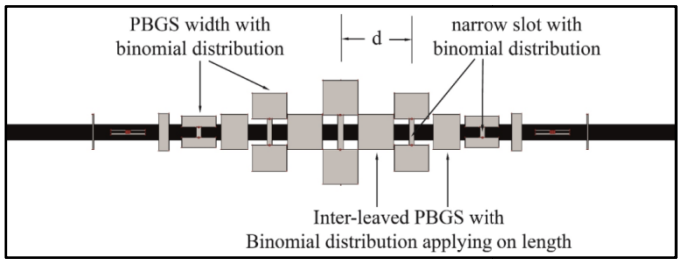


Fig. 6 Geometry of a Binomially distributed LPF designed by varying both width (a) of PBGS and width (w) of the narrow slot of the dumbbell shaped DGS and lengths of the interleaved PGBS etched in the ground plane. The period of dumbbell shaped DGS is 'd' for both DGS and interleaved PBGS and the substrate is Taconic with dielectric constant= 2.45 and height= 31 mils.

IV. Results and Discussions

We have simulated all designs by method of moment (MOM) based EM software Zealand IE 3D which is commercially available for researchers. We have investigated the performance of different designs in terms of cut-off frequency, ripple height and depth and width of the rejection band.

A. Design 1

The S-parameters vs. frequency of Design 1 is shown in Figure 7. It can be seen that the return loss curve is almost straight after 4GHz. But it is observable that there is a ripple on the passband in insertion loss curve. Maximum isolation of the rejection band is -69 dB and minimum isolation is -25 dB.

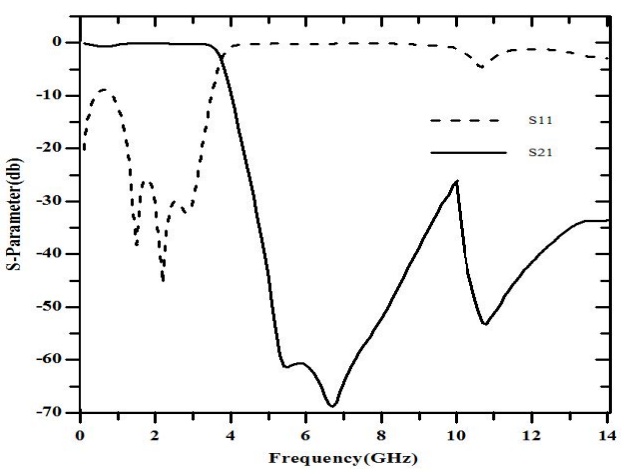


Fig. 7 S-parameters performance of a transmission line over dumbbell shaped DGSs according to design 2 with FF=0.5 for center element. The substrate is Taconic with dielectric constant= 2.45 and height= 31 mils.

B. Design 2

The S-parameters vs. frequency of Design 2 is shown in Figure 8. It can be seen that return loss curve is almost straight after 4GHz. But it is observable that there is a ripple on the passband in insertion loss curve. Maximum isolation of the rejection band is -55 dB and minimum isolation is -29 dB. For design 2 we observe that the improvement of the IL curve at 10GHz than design 1. The pick point is being reduced for design 2 in the IL curve.

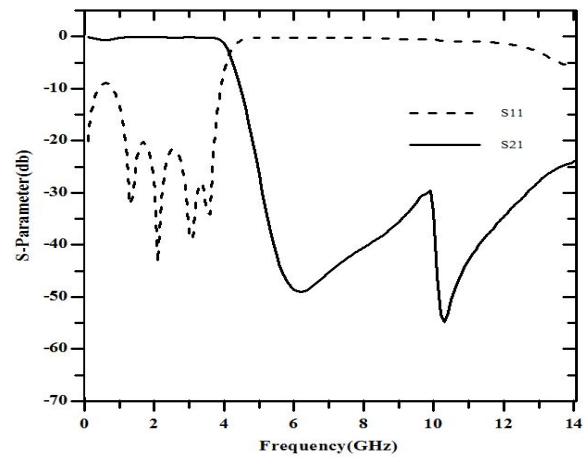


Fig. 8 S-parameters performance of a transmission line over dumbbell shaped DGSs according to design 3 with FF=0.5 for center element. The substrate is Taconic with dielectric constant= 2.45 and height= 31 mils.

C. Design 3

The S-parameters vs. frequency of a dumbbell shape DGS with inter-leaved PBGS of Design 3 is shown in Figure 9. There is a single ripple on the passband in IL curve. Maximum isolation of the rejection band is -65 dB and minimum isolation is -31 dB.

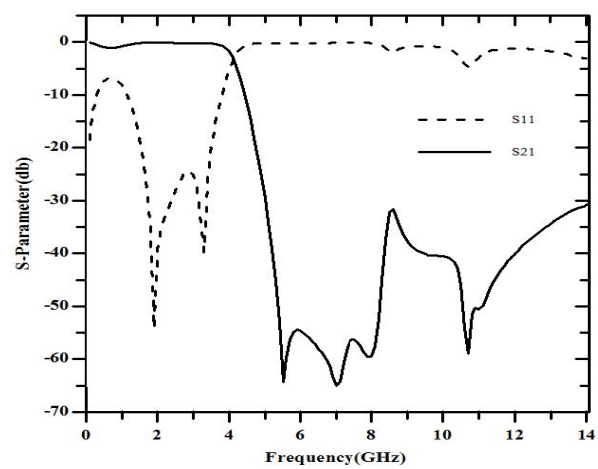


Fig. 9 S-parameters performance of a transmission line over hybrid dumbbell shaped DGSs according to design 4 with FF=0.5 for center element. The substrate is Taconic with dielectric constant= 2.45 and height= 31 mils.

D. Design 4

The S-parameters vs. frequency of a dumbbell shape DGS with inter-leaved PBGS of Design 4 is shown in Figure 10.

There is a single ripple on the passband in IL curve. Maximum isolation of the rejection band is -75 dB and minimum isolation is -28 dB. This design gives the best performance among all the designs that we have proposed. Passband of the insertion loss curve is smoother and the rejection band curves are below -20db and going down gradually. The changing of the width of PBGS of dumbbell shaped DGS and at the same time the changing of the length of interleaving PBGSs have caused a great impact on the results.

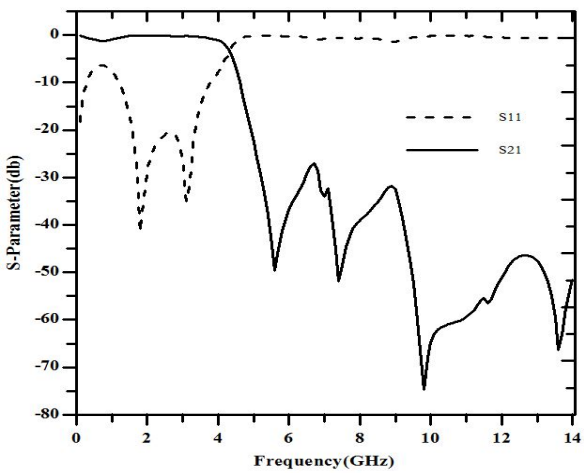


Fig. 10 S-parameters performance of a transmission line over hybrid dumbbell shaped DGSs according to design 5 with FF=0.5 for center element. The substrate is Taconic with dielectric constant= 2.45 and height= 31 mils.

V. Conclusions

We have investigated S-parameter performances 50Ω transmission line on binomially distributed dumbbell shaped DGS and hybrid DGS having FFs of 0.5 for the center element. For these S-parameter performance's investigations, we have used Taconic substrate with dielectric constant of 2.45 and height of 31 mils. It can be seen that Binomial distribution for dumbbell shaped DGS or hybrid dumbbell shaped DGS offers promising performance. Considering the improved performance in terms of RL, IL, ripples, passband and rejection band we have got the best performance for design 5. The depth of stopband and performance of the passband is different for different designs, but they provide a wider scope to the designer for the specific design purposes of a low pass filter. In these investigations conductor losses, substrate loss and loss due to conductors are not included. All the investigations have been done on an infinite ground plane. Additionally, fabrication may cause variation in the size and period of the lattice structures.

To evaluate the performance of binomially distributed dumbbell shaped DGSs and hybrid dumbbell shaped DGSs, we have tested many structures and found the best designs we have provided in this paper. Our main purpose of this design is to reduce ripple from the passband of the LPF and to get improved performance from the band rejection portion. The idea is a novel idea as far of the author's knowledge. Therefore, the design with DGSs and hybrid DGSs are more attractive and may find applications in microwave devices and components.

References

[1] Yablonovitch, E., "Inhibited spontaneous emission in solid-state physics and electronics," *Physical Review Letters*, vol. 58, no. 20, pp. 2059–2062, May 1987.

[2] John, S., "Strong localization of photons in certain disordered dielectric superlattices," *Physical Review Letters*, vol. 58, no. 23, pp. 2486–2489, June 1987.

[3] L. H. Weng, Y. C. Guo, X. W. Shi, and X. Q. Chen "An Overview On Defected Ground Structure" *Progress In Electromagnetics Research B*, vol. 7, pp. 173–189, 2008

[4] R Kim and H.Y.D.Yang. "Surface waves and leaky waves in integrated circuit structures with planar periodic dipole loading," *in Antennas and Propagation Society International Symposium*. *IEEE*, Salt Lake City, UT, USA, 2000, pp. 346-349

[5] K-P.Ma, J. Kim. F-R.Yang. Y.Qian and T. Itoh "Leakage suppression in stripline circuits using a 2-D photonic bandgap lattice," *Microwave symposium digests. IEEE MTT-S Int*., Anaheim, CA, USA, 1999, pp. 73-76

[6] F-R Yang, K-P Ma, Y.Qian and T.Itoh "A uniplanar compact photonic-bandgap (UC-PBG) structure and its applications for microwave circuits," IEEE Trans. "*Microwave Theory and Tech.*, vol 47, no.8 pp. 1509-1514, Aug.1999.

[7] Md. Nurunnabi Mollah, "Planar Electromagnetic Bandgap Structures and Applications", Ph.D. dissertation, Nanyang Technological University, Singapore, March 2005.

[8] Mohammad N Mollah, Nemai Chandra Karmakar, Jeffrey S. Fu, " A Binomially Distributed Photonic Bandgap Structure (PBGS) and Its Application to Bandpass Filter", *RF and Microwave Computer Aided Engineering*, vol. 16, pp. 355-366, July 2006.

[9] C. A. Balanis, *Antenna Theory Analysis and Design*, 2nd ed., New York: Wiley, 1997.

[10] Chul-Soo Kim, Jun-Seok Park, Dal Ahn, and Jae-Bong Lim, "A Novel 1-D Periodic Defected Ground Structure for Planar Circuits," *IEEE-MWCL*, vol.10, no. 4, pp.131-133, April 2000.

[11] K. O. Hill and G. Meltz, "Fiber Bragg Grating technology fundamentals and overview", *J Lightwave Tech*. vol. 15, no. 5, pp. 1263-1276, Aug. 1997.

Design of a Compact Dual-Band Band-Pass Filter for Global Positioning System and Fixed Satellite Applications

Md. Rokunuzzaman[1],* Mohammad Tariqul Islam,[2] Md. Samsuzzaman[2], Md. Rasheduzzaman[3] Rezaul Azim[1] and Afaz Uddin Ahmed[1]

[1]Center for Space Science (ANGKASA), Universiti Kebangsaan Malaysia
43600 Bangi, Selangor, Malaysia

[2]Department of Electrical, Electronic and Systems Engineering, Universiti Kebangsaan Malaysia
43600 Bangi, Selangor, Malaysia

[3]Department of Electrical Electronics and Communication Engineering, Military Institute of Science and Technology
Dhaka-1216, Bangladesh

*robelhk@yahoo.com

Abstract—**In this article, a dual-band band-pass filter is developed for Global Positioning System (GPS) and Fixed Satellite (FS) applications. The filter is a planar structure composed of parallel coupled lines with reduced length. The dual band filter is developed using the cascaded coupled lines. High isolation between two pass-bands is achieved. Compared to the previous works, the proposed filter has lower insertion loss at 1.5 GHz and 7.5 GHz. A compact size of 15.7 mm × 33.4 mm is achieved for the proposed filter using Arlon DiClad 880 substrate with a thickness of 0.508 mm.**

Index Terms—**Bandpass filter (BPF), Dual-band, Global Positioning System (GPS), Fixed Satellite (FS).**

I. Introduction

With the rapid growth of microwave applications and services of communication, the desire for multi-frequency equipment is increasing. Moreover, the frequency is divided into categories for different applications. To separate one frequency from another, filter is an essential element. Low cost, compactness and heavy duty products are in great need for the present microwave applications. High performance multi-band filters such as in [1, 2] are making their marks towards the use of multi-band. In recent years, a number of journals are published focusing on multi-band filters for microwave frequencies [3-6]. Currently, there are a large number of literatures promoting multi-band antennas such as [7-9]. To use multi-band antennas it is required for the backend of the antenna to be developed for multi-band accessibility.

To design band-pass (BPF) filter, a transmission line varying continuously is used to match the impedance in [10]. Non-uniform coupled filter design method is shown in [11] realizing arbitrary transmission line filters. While designing dual-band BPF filter, the challenge for microwave circuit designer is to design the filter in a compact area with high performance. Previously, microwave circuit designers used to develop two individual BPF filters to achieve dual-band BPF response from a filter [12]. To combine the performance of the two filters, diplexer type structures are introduced to match the impedance of both filters and hence resulting larger dimension for the filter design. In the later version of dual-band BPF filter design, open-stub cascaded structures are used to achieve the performance increasing the complexity of the circuit and size increment [13].

In this paper, a planar dual band filter is proposed with band pass characteristics. The two centre frequencies of the filter are 1.5 GHz and 7.5 GHz. High transmission characteristics is achieved at both bands by utilizing cascaded coupled lines. The proposed antenna is well matched for 50 Ω impedance microwave circuits. At both pass-bands low insertion loss is achieved. The presented dual-band BPF is designed at a compact dimension of 524.38 mm^2 with a substrate thickness limited to 0.508 mm.

II. Design of the Filter

Fig. 1 shows the layout of the dual-band BPF. Arlon DiClad 880 substrate with a thickness of h = 0.508 mm is used to design the filter. The dielectric constant and loss tangent of the selected substrate are ε_r = 2.2 and tan δ = 0.0009. the basic structure of the filter consists of two tapered I/O paths and a coupled structure in the middle. The coupled structures shown in the fig. 1 follows a pattern that can be described as stepped impedance resonators. By using the coupling mechanism between the microstrip lines, the higher order mode resonances of the filter can be omitted or shifted to another resonance point [14]. Two identical microstrip structures are coupled with each other at the middle of the BPF filter. The arms of the coupling structures are folded to decrease the total size of the filter which makes the filter design more compact. At the edge of each microstrip coupling structure, a cascaded structure is introduced to cope with the different impedance characteristic. It can be observed from the layout that two different thickness microstrip lines are coupling with each other. It is valid for both identical microstrip coupled structures. The irregularity of the thickness can be expressed as: for the long length ($2\theta_1$) in the middle of a cascaded structure the higher impedance is (Z_1) and at both end of the long length, two short length (θ_2) with an impedance (Z_2) lower than the impedance (Z_1), the resonance condition for both even and odd mode can be derived as followed [15]:

$$K * \cot\theta_2 = -\cot\theta_1 \quad \text{(Even-mode)} \tag{1}$$

$$K * \cot\theta_2 = \tan\theta_1 \quad \text{(Odd-mode)} \tag{2}$$

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

Where, $K = (Z_2/Z_1)$ the ratio between two impedances [14]. The length θ_2 is corresponding to the physical length of the stepped impedance resonator. The ratio 'τ' of the total length of the cascaded structure can be expressed as follows:

$$\tau = \frac{\theta_2}{(\theta_1 + \theta_2)} \tag{3}$$

By substituting (3) into the equation (1) and (2) where $\theta_1+\theta_2 = \theta_t$ we find:

$$K * \cot(\tau\theta_t) = -\cot(\theta_t(1-\tau)) \tag{4}$$

$$K * \cot(\tau\theta_t) = \tan(\theta_t(1-\tau)) \tag{5}$$

The resultant θ_t is dependent on the ratios τ and K where θ_t can have several types of solution. For design presented in this paper, the characteristics of this structure are changed repeatedly to optimize the design to resonate at the frequencies 1.5 GHz and 7.5 GHz respectively. The I/O paths of the proposed filter are tapered so that the impedance of the coupled structure can be matched with the 50 Ω impedance of the backend structure it is mounted on. The feed-line of the filter starts from two identical points at both sides of the filter. After an increment of '*a*' the feed-line is banded to reduce the dimension of the filter from both sides. The bending of the feed-line extensively decreases the total size of the filter, making it an ideal filter device for any compact system. At the end of the feeding, the line is tapered and the thickness is increased until the length '*d*' where, at the edge it touches the excitation point.

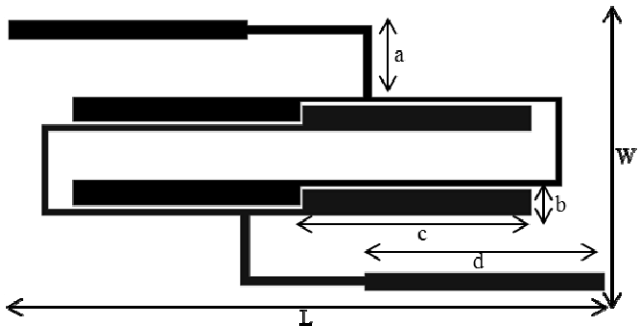


Fig. 1 Layout of the proposed BPF filter. The filter dimensions are: L = 33.4 mm, W = 15.7 mm, a = 4 mm, b = 1.4 mm, c = 12.8 mm, d = 13.4 mm.

III. Results

The proposed dual-band BPF is designed using the commercially available software High Frequency Structural Simulator (HFSS) version 15.0. A plane copper layer is assigned at the bottom of the substrate and assigned as the ground plane of the filter. The calculation is carried out using two lumped ports assigned 50 Ω impedance so that the filter can be assembled with 50 Ω microwave structures. Fig. 2 shows the S-parameter response of the proposed filter. It can be observed from the figure that an insertion loss of less than -39 dB is achieved at 1.5 GHz and -23.8 dB is achieved at 7.5 GHz. The S_{11} and S_{22} response are almost identical to each other and overlaps each other throughout the frequency span. The S_{21} response of the filter shows that at the first resonance of the BPF, the transmission loss is as low as -0.11 dB and the -3 dB bandwidth starts from 1.2 GHz till 1.8 GHz with a 40% fractional bandwidth (FBW) centred at 1.5 GHz frequency. At the second resonance the transmission loss is lower at 7.5 GHz with a value of -0.53 dB and the -3 dB bandwidth starts from 7.1 GHz till 7.8 GHz with the FBW of 10% centred at 7.5 GHz. The -10 dB bandwidth of S_{11} and S_{22} at the first resonance starts from 1.3 GHz till 1.7 GHz and at the second resonance, the bandwidth starts from 7.3 GHz until 7.7 GHz.

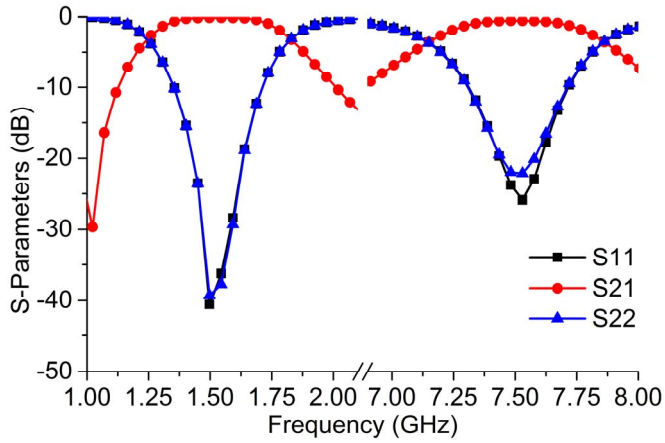


Fig. 2 S-parameter responses of the BPF with respect to the frequency.

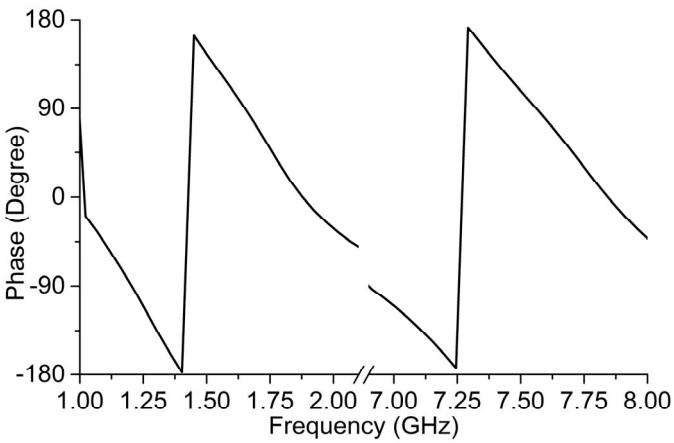


Fig. 3 Phase response of the BPF at resonance frequencies.

Fig. 3 depicts the phase response at the resonance bands of the BPF. It can be observed from the figure that the phase response change with the frequency is smooth and there is no irregularity of phase change within the pass-bands of the BPF.

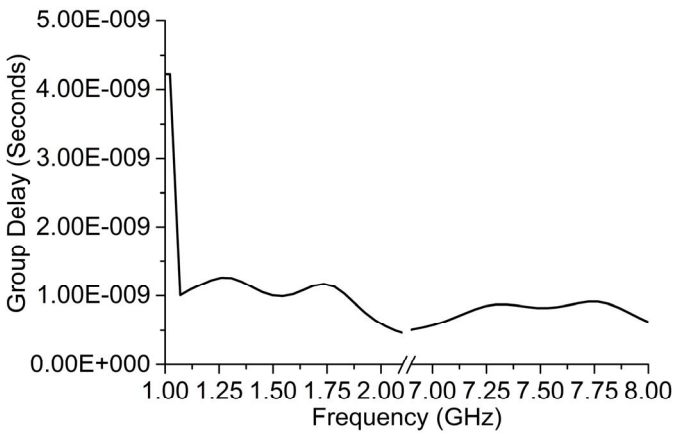


Fig. 4 Group delay within the I/O of the BPF.

Fig. 4 shows the group delay between two I/O ports of the filter. From the figure, it can be perceived that the group delay is not constant throughout the frequency. However, within the resonance bandwidths, the average group delay is below 1 ns. Although some ripples can be noticed, nonetheless, the fluctuation of time is very narrow which can be ignored.

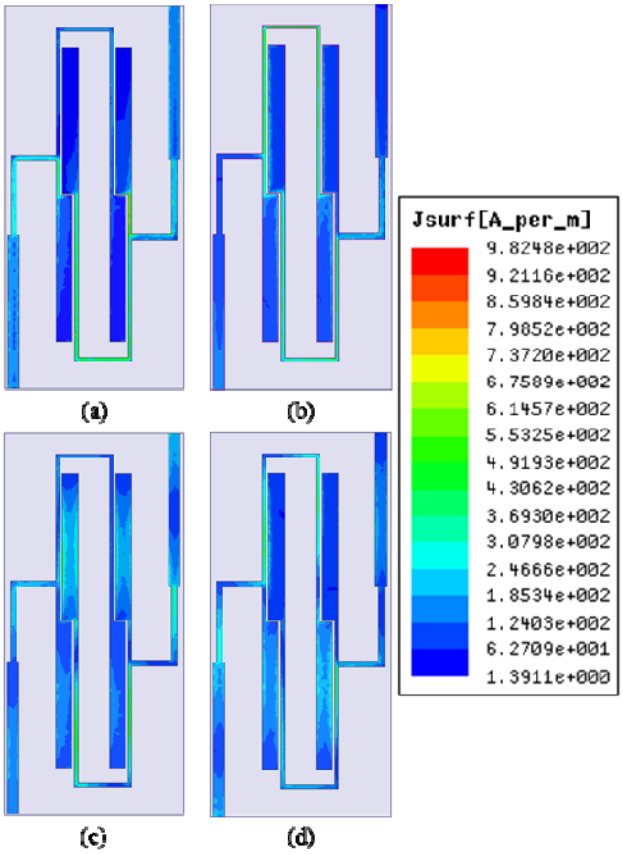


Fig. 5 Current distribution pattern of the BPF at (a) 1.5 GHz and 0° phase, (b) 1.5 GHz and 90° phase, (c) 7.5 GHz and 0° phase and (d) 7.5 GHz and 90° phase.

Fig. 5 shows the current distribution pattern of the filter at both resonance frequencies. The current distribution shows a peak of $9.8248\times e^2$ Amp/meter at all the current distribution patterns. In fig. 5 (a) the current distribution is shown for 1.5 GHz at a phase of 0°. It can be observed that the current density is at its peak over the feed-lines. The current distribution in Fig. 5 (b) depicts that the current density is increasing at the band lines connecting the coupling hands of the filter for 90° phase difference at 1.5 GHz. At 7.5 GHz 0° phase, current is dense mostly at the edge of the feeding point and the coupling hands shown in Fig. 5 (c). Fig. 5 (d) presents the current distribution pattern of the filter for 7.5 GHz 90˚ phase. One side of the coupled hands are excited with more current compared to the other side. The connecting line of the coupling hand also shows highly dense current distribution.

IV. Conclusions

A compact, low loss novel dual-band BPF is presented by means of coupled structures. The filter is designed for GPS and FS system for which the frequency mostly comes from satellites and filter is required for high cancellation from noise. The proposed filter shows good frequency responses at 1.5 GHz and 7.5 GHz with a bandwidth of 600 MHz and 700 MHz respectively. The filter has a very low transmission loss. Good isolation between two operating bands is achieved. The phase response gives a stable performance at both operating bands. An average group delay of less than 1 ns is achieved between two I/O ports. The proposed filter comprises a dimension of 15.7 mm × 33.4 mm and the Arlon Diclad 880 substrate thickness of 0.508 mm.

Acknowledgment

The authors would like to thank Centre for Space Science (ANGKASA) and Universiti Kebangsaan Malaysia for sponsoring this work.

References

[1] C.-W. Tang, S.-F. You, and I.-C. Liu, "Design of a dual-band bandpass filter with low-temperature co-fired ceramic technology," *Microwave Theory and Techniques, IEEE Transactions on,* vol. 54, pp. 3327-3332, 2006.

[2] C.-W. Tang and S.-F. You, "Using the technology of low temperature co-fired ceramic to design the dual-band bandpass filter," *Microwave and Wireless Components Letters, IEEE,* vol. 16, pp. 407-409, 2006.

[3] F. Sen, W. Bian, C. Jia, S. Shou-jia, and L. Chang-Hong, "Novel Second-Order Dual-Mode Dual-Band Filters Using Capacitance Loaded Square Loop Resonator," *Microwave Theory and Techniques, IEEE Transactions on,* vol. 60, pp. 477-483, 2012.

[4] H. Chen and Y. X. Zhang, "Design of a novel dual-band bandpass filter using microstrip/CPW structure," in *Communications, Circuits and Systems, 2008. ICCCAS 2008. International Conference on*, 2008, pp. 1216-1218.

[5] Z. Songbai and Z. Lei, "Synthesis Design of Dual-Band Bandpass Filters With <formula formulatype="inline"> <img src="/images/tex/946.gif" alt="\lambda /4"> </formula> Stepped-Impedance Resonators," *Microwave Theory and Techniques, IEEE Transactions on,* vol. 61, pp. 1812-1819, 2013.

[6] C. Wei-Shin and C. Chi-Yang, "Analytical Design of Microstrip Short-Circuit Terminated Stepped-Impedance Resonator Dual-Band Filters," *Microwave Theory and Techniques, IEEE Transactions on,* vol. 59, pp. 1730-1739, 2011.

[7] M. Zaman, M. Islam, N. Misran, and J. Mandeep, "Analysis of Resonance Response Performance of C-Band Antenna Using Parasitic Element," *The Scientific World Journal,* vol. 2014, 2014.

[8] M. Zaman, M. Islam, N. Misran, and B. Yatim, "Analysis of Glass-Reinforced Epoxy Material for Radio Frequency Resonator," *The Scientific World Journal,* vol. 2014, 2014.

[9] M. R. Zaman, R. Azim, N. Misran, M. F. Asillam, and T. Islam, "Development of a Semielliptical Partial Ground Plane Antenna for RFID and GSM-900," *International Journal of Antennas and Propagation,* vol. 2014, p. 7, 2014.

[10] M. Le Roy, A. Pérennec, S. Toutain, and L. C. Calvez, "The continuously varying transmission-line technique-application to filter design," *Microwave Theory and Techniques, IEEE Transactions on,* vol. 47, pp. 1680-1687, 1999.

[11] G. Xiao, K. i. Yashiro, N. Guan, and S. Ohkawa, "An effective method for designing nonuniformly coupled transmission-line filters," *Microwave Theory and Techniques, IEEE Transactions on,* vol. 49, pp. 1027-1031, 2001.

[12] J.-T. Kuo, T.-H. Yeh, and C.-C. Yeh, "Design of microstrip bandpass filters with a dual-passband response," *Microwave Theory and Techniques, IEEE Transactions on,* vol. 53, pp. 1331-1337, 2005.

[13] X. Guan, Z. Ma, P. Cai, Y. Kobayashi, T. Anada, and G. Hagiwara, "Synthesis of dual-band bandpass filters using successive frequency transformations and circuit conversions," *Microwave and Wireless Components Letters, IEEE,* vol. 16, pp. 110-112, 2006.

[14] M. Makimoto and S. Yamashita, "Bandpass filters using parallel coupled stripline stepped impedance resonators," *Microwave Theory and Techniques, IEEE Transactions on,* vol. 28, pp. 1413-1417, 1980.

[15] C.-F. Chen, T.-Y. Huang, C.-P. Chou, and R.-B. Wu, "Microstrip diplexers design with common resonator sections for compact size, but high isolation," *Microwave Theory and Techniques, IEEE Transactions on,* vol. 54, pp. 1945-1952, 2006.

Electromagnetic Wave Propagation Characteristics in Single Walled Metallic Carbon Nanotube

Md. Shantanu Islam[1], M A Matin[2] and M. Mofazzal Hossain[1]*

[1]Department of Electronics and Communications Engineering, East West University
Aftabnagar, Dhaka-1212, Bangladesh

[2]Department of Electrical and Electronic Engineering, Institut Teknologi Brunei, Jalan Tungku Link, Gadong BE1410,
Brunei Darussalam

*dmmh@ewubd.edu

Abstract—**The metallic single walled carbon nanotubes have been proposed as a promising inter connector in intra-chip and inter-chip packaging applications as well as passive devices for future generation terahertz IC (Integrated Circuit) technology, due to their superior electrical and thermal properties compared with those of copper. In this paper a theoretical investigation is carried out to predict phase and group velocities and the attenuation characteristics of single walled metallic carbon nanotubes (SWCNTs), and bundled SWCNTs in the terahertz regime. The expressions for attenuation constant, phase constant, phase velocity, and group velocity have been derived using transmission line theory. It is found that the predicted phase and group velocities in single SWCNT, bundled SWCNTs are strongly frequency dependent and increases with frequency. It is also noticed that the group velocity decreases after having a peak at around 100 GHz for both single SWCNT and bundled SWCNT.**

Index Terms—**Attenuation Constant, Phase Velocity, Group Velocity, Bundled SWCNT.**

I. Introduction

One of the latest trends in GHz to THz technology [1] is to use single-walled carbon nanotubes (SWNTs) as building blocks of novel high-frequency devices and interconnects. CNTs have been drawn great interest due to their outstanding electronic, thermal, electromagnetic and mechanical properties. CNTs are basically rolled-up graphene sheets and are of two types: metallic and semiconducting depending on the chirality. Recent trend of technology reveals that metallic CNTs are going to replace the conventional copper as on-chip interconnect in VLSI circuits, transmission lines and coplanar waveguides in microwave integrated circuits (MIC), and dipole antennas for terahertz radiation [2]. Conventional copper does not offer reliable performance when it comes to transmit or detect very high frequency signal due to their skin effect at nanoscale dimensions. CNTs outperform copper as interconnect in VLSI circuits. The superior electrical and thermal properties of CNTs over copper are: (i) high current density (SWCNT: 10^9 A/cm^2, Cu: 10^7 A/cm^2) [3]; (ii) Large mean free path (SWCNT: 2.8 µm, Cu: 40 nm) [4]; (iii) High thermal conductivity (SWCNT: 5800 W/m-K, Cu: 385 W/m-K) [5]; (iv) Negligible skin effect at THz regime compared with copper [6]; (v) smaller resistance at THz regime compared with copper [4, 6]; (vi) high melting point (SWCNT: 3800 K, Cu:1356K). The mechanical properties of CNTs are also worth noting over copper. CNTs have high tensile strength (CNT: 22 GPa, Cu:0.22 GPa) [7] due to its sp^2 hybridization; thus, stronger than diamond.

Based on the Luttinger liquid theory, Burke [8] proposed the *RLC* equivalent circuit model of SWCNT considering only the length independent quantum resistance, quantum and electrostatic capacitances and kinetic and magnetic inductances. Later on, that model was revised adding the length dependent scattering resistance and imperfect metal-CNT contact resistances [4, 9-10]. Koo *et al* [4] discussed the latency and power dissipation of CNT based on *RLC* circuit model. Naemee *et al* [11] also discussed the latency of CNT based on similar type of *RLC* circuit model, and they also found that the phase velocity in SWCNT is about 1.76×10^6 m/s. But their phase velocity calculation was over simplified as they neglected the resistance of SWCNT (assuming lossless transmission line). Therefore, that result is valid only for an SWCNT whose length is less than the mean free path of electron. But in reality, SWCNTs are synthesized having a length of several millimeters, which offer a high resistance and therefore, huge power dissipation when used as a transmission lines at GHz to THz frequencies. As the use of CNTs in the GHz and THz applications (interconnect in ultra high speed VLSI circuits, transmission line, waveguides etc) are emerging; thus it is crucial to carry on the comprehensive investigation of wave velocities (both phase and group velocities) in SWCNTs taking into account the resistance.

In this paper, novel simple expressions of attenuation constant, phase constant, phase and group velocities are derived analytically from the expressions of complex propagation constant taking into account the high resistance, capacitance and inductance of SWCNT. We shall also compare the phase and group velocities among single SWCNT, bundled SWCNT.

II. Electrical properties of SWCNT

The properties of SWCNT are very different than those of conventional metal conductors. Some of the electrical properties of SWCNT relevant to interconnect applications are described in this section.

A. Electrical Resistance

The electrical resistance of SWCNT consists of three components: (i) Quantum resistance, (R_Q) (ii) Scattering resistance (R_QL/l_0), and (iii) Contact resistance arises from metal-CNT imperfect contacts, (R_c). The quantum resistance is the intrinsic resistance of CNT, which is independent of the length of CNT. However the quantum resistance is diameter dependent [12]. The scattering resistance attributed from the scattering effects, is length dependent. Hence the total resistance of SWCNT can be expressed as:

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

$$R_{CNT} = R_c + R_Q\left(1 + \frac{L}{l_0}\right) \quad (1)$$

Here, L: length of the SWCNT, and l_0: electron's mean free path in SWCNT, R_Q: quantum resistance (6.5 kΩ), and R_c: metal-CNT contact resistance.

The high resistance of SWCNT may be reduced using bundle of SWCNTs. The resistance for a bundle of SWCNT consists of "n" number of SWCNTs can be expressed as:

$$R_{bundle} = \frac{R_{CNT}}{n} \quad (2)$$

B. Capacitance

In an SWCNT, there are two conducting channels, and due to the electron spins (spin-up and spin-down) each channel has two degeneracy. Therefore, in an SWCNT, there are four conducting channels. According to the circuit model of Burk [8], there are two types of capacitances: (i) electrostatic capacitance, C_E and (ii) Quantum capacitance, C_Q. The total capacitance of an SWCNT can be expressed as:

$$C_{CNT} = \frac{4C_Q C_E}{4C_Q + C_E} \quad (3)$$

The capacitance for a bundle of SWCNT consists of "n" number of SWCNTs can be expressed as:

$$\frac{1}{C_{bundle}} = \frac{1}{C_Q^{bundle}} + \frac{1}{C_E^{bundle}} \quad (4)$$

Where, $C_Q^{bundle} = 4 \times C_q \times n$ and

$$C_E^{bundle} = 2C_{En} + \frac{n_w - 2}{2}C_{Ef} + \frac{3(n_H - 2)}{5}C_{En}$$

Here, n_H is the number of rows and n_w is the number of columns in the interconnect bundle; here $n=(n_H \times n_W)-n_H/2$ if n_H is even and $n=(n_H \times n_W)-(n_H-1)/2$ if n_H is odd [13].

C. Inductance

The total inductance of SWCNT consists of kinetic inductance and mutual inductance. In a single SWCNT the kinetic inductance is dominating, thus we can ignore the effects of mutual inductance. The kinetic inductance can be expressed as:

$$L_{CNT} = \frac{h}{8e^2 v_F} \quad (5)$$

The inductance for a bundle of SWCNT consists of "n" number of SWCNTs may be expressed as:

$$L_{bundle} = \frac{L_{CNT}}{n} \quad (6)$$

III. Expression of Wave velocities

The velocity of electromagnetic waves travelling through a transmission line has been characterised by: (i) phase velocity and (ii) group velocity. Phase velocity is defined as a mathematical formulation but not a realistic velocity. Group velocity is the real velocity of a travelling wave. Here we shall derive the expressions of phase velocity and group velocity in SWCNT, considering the SWCNT as a transmission line. The well known telegrapher equations can be written as follows [14]:

$$\frac{dV(z)}{dz} = -(R + j\omega L)I(z) \quad (7)$$

$$\frac{dI(z)}{dz} = -(G + j\omega C)V(z) \quad (8)$$

Where, R, G, L and C are the series resistance, parallel conductance, series inductance and parallel capacitance respectively. $V(z)$ and $I(z)$ are the voltage and current respectively at any location of the transmission line, and z is the direction of propagation. From theses equations after manipulation, we can write the voltage and current waves as:

$$\frac{d^2V(z)}{dz^2} - \gamma^2 V(z) = 0 \quad (9)$$

$$\frac{d^2I(z)}{dz^2} - \gamma^2 I(z) = 0 \quad (10)$$

The solution of equations (9) and (10) can be written as follows:

$$V(z) = V_0^+ e^{-\gamma z} + V_0^- e^{\gamma z} \quad (11)$$

$$I(z) = I_0^+ e^{-\gamma z} + I_0^- e^{\gamma z} \quad (12)$$

Where, the propagation constant, γ can be expressed as follows.

$$\gamma = \alpha + j\beta = \sqrt{(R + j\omega L)(G + j\omega C)} \quad (13)$$

In SWCNT we may assume G=0. Hence, from equation (13), we derived the following equations for attenuation constant (α) and phase constant (β).

$$\alpha = \frac{1}{\sqrt{2}}\left[\sqrt{\omega^4 L^2 C^2 + \omega^2 R^2 C^2}\right]^{\frac{1}{2}} \quad (14)$$

$$\beta = \frac{1}{\sqrt{2}}\left[\sqrt{\omega^4 L^2 C^2 + \omega^2 R^2 C^2} + \omega^2 LC\right]^{\frac{1}{2}} \quad (15)$$

According to the definition of phase velocity, it can be expressed as:

$$v_p = \frac{\omega}{\beta} \quad (16)$$

From equations (14), (15) and (16), the expression of phase velocity can be written as:

$$v_p = \frac{\sqrt{2}}{\left[L^2C^2 + \left(\frac{RC}{\omega}\right)^2 + LC\right]^{\frac{1}{2}}} \quad (17)$$

The group velocity, v_g can be defined as:

$$v_g = \frac{1}{\frac{d\beta}{d\omega}} \quad (18)$$

From equations (15) and (18), we may derive the expression of group velocity as:

$$v_g = \frac{4\sqrt{2}\left(\sqrt{\omega^4 L^2 C^2 + \omega^2 R^2 C^2} + \omega^2 LC\right)\sqrt{\omega^4 L^2 C^2 + \omega^2 R^2 C^2}}{3\omega^3 L^2 C^2 + 2\omega R^2 C^2 + 4\omega LC\sqrt{\omega^4 L^2 C^2 + \omega^2 R^2 C^2}} \quad (19)$$

IV. Results and discussions

The performance of electromagnetic wave propagation through the SWCNT may be quantified using the phase and group velocities profiles. The velocities in single SWCNT and bundled SWCNT have been calculated using the derived expressions in the previous section. Fig.1 and Fig.2 describe the dependence of phase and group velocity for single SWCNT and bundled SWCNT. From these figures, it is found that the both the wave velocities (phase velocity and group velocity) are strong functions of operating frequency, especially at the lower frequency range. It is observed that at higher frequency regime, the phase velocity becomes saturated [Fig.1] and the group velocity has a peak at around 100 GHz and then decreases with frequency [Fig.2]. It is also noticed that in both cases the wave velocity is much higher in bundled SWCNT compared with that in single SWCNT. This is due the reduced values of resistance and inductance in a bundled SWCNT. In a bundle, SWCNTs are considered in parallel. The phase velocity in single SWCNT found in this work, is in good harmony with that of Ref. [11]. Fig.3 shows the evolution of attenuation constant with frequency. Attenuation constant increases with the increase of frequency due to the reduction of skin depth at higher frequency [6], and attenuation constant is inversely proportional to skin depth. It is also found that the attenuation is much lower in a bundle of SWCNT than that of in a single SWCNT.

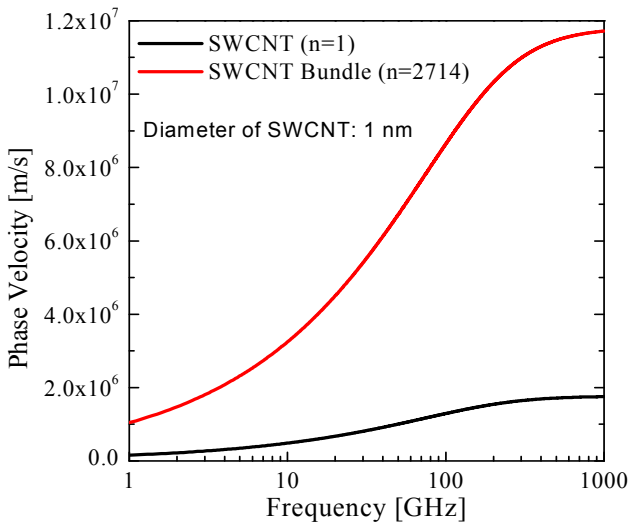


Fig.1 Frequency dependence of phase velocity for single SWCNT and bundled SWCNT

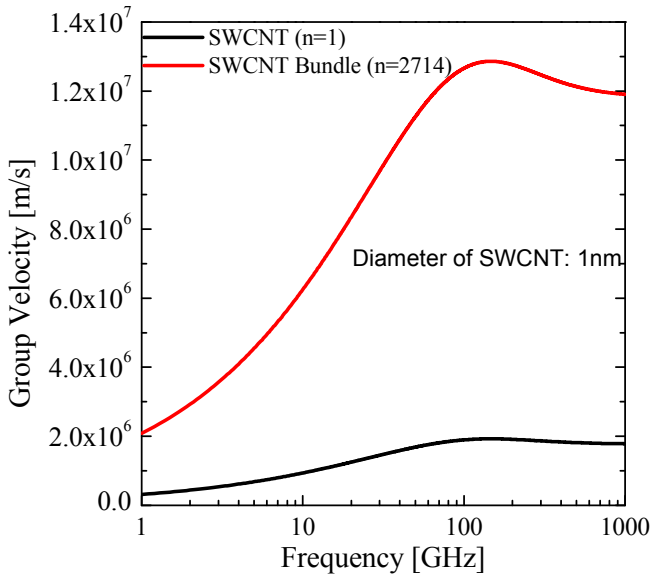


Fig.2 Frequency dependence of group velocity for single SWCNT and bundled SWCNT

We have also investigated the effects of CNT diameter on the group velocity and attenuation constant. The diameter (D) of SWCNT is related to the mean free path of electrons (l0) in the SWCNT by l0=1000D [6]. Fig.4, describes the effects of diameter of SWCNT on the group velocity of single and bundled SWCNTs at 100 GHz. It is noticed that the diameter has a very insignificant effect on the group velocity of single SWCNT; however, the group velocity of bundled SWCNT drastically falls with the increase of diameter of SWCNT. This is because, if we use larger diameter SWCNTs in a bundle (with a fixed dimension of bundle), the number of SWCNTs decreases. As a result the equivalent resistance and inductance of a bundle decrease, which in turn reduces the group velocity. Therefore, it is advantageous to use smaller diameter SWCNTs in a bundle to have higher group velocity. Fig.5 shows the effects of diameter of SWCNT on the attenuations constant of single and bundled SWCNTs at 100 GHz. It is found that in single SWCNT attenuation constant decreases with larger diameter, because larger diameter CNTs offer smaller resistance. On the other hand, the attenuation of bundled SWCNT slightly increases with diameter. The reason is similar as described in Fig. 4.

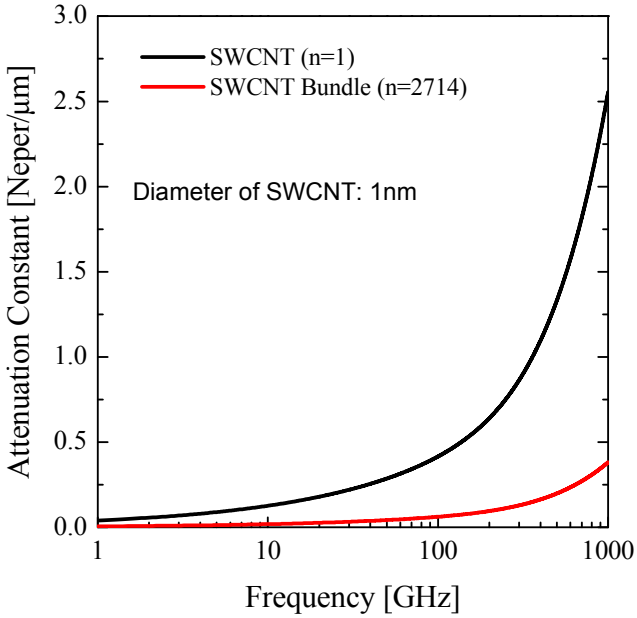


Fig.3 Frequency dependence of attenuation constant for single SWCNT and bundled SWCNT

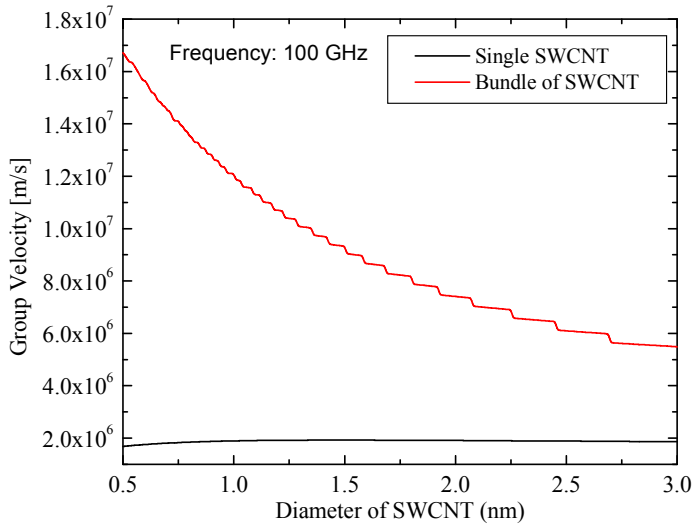


Fig.4 Effects of SWCNT's diameter on the group velocity at 100 GHz.

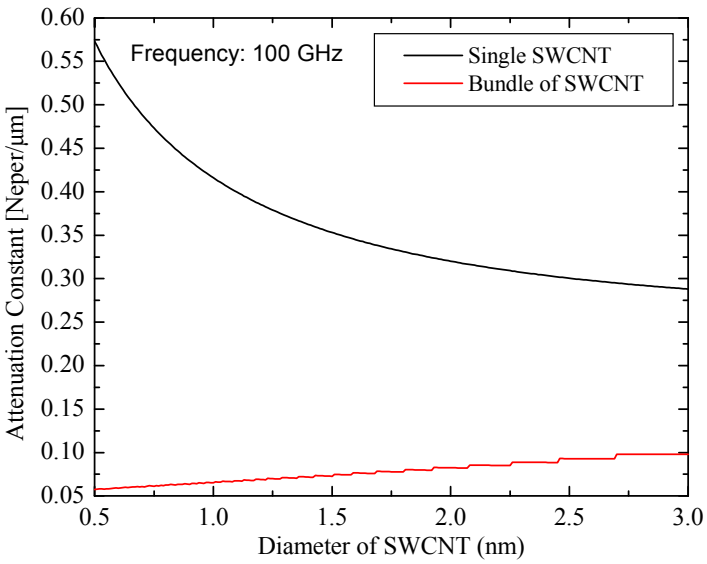


Fig.5 Effects of SWCNT's diameter on the attenuation constant at 100 GHz.

V. Conclusion

In this paper the novel expressions for phase and group velocity, and attenuation constant have been derived taking into account the losses in the SWCNT. We have discussed the electrical properties of single SWCNT and bundled SWCNT related to interconnect applications. It is found that the wave velocities and attenuation are strong functions of frequency. The wave velocities are much higher in an SWCNT bundle than that of in a single SWCNT. In contrast, the attenuation in a bundle of SWCNT is much lower than that of in a single SWCNT. Moreover, the diameter of SWCNT significantly affects the group velocity of bundled SWCNT and attenuation in single SWCNT. It may be recommended that to have higher group velocity and smaller attenuation, we should use bundle of smaller diameter SWCNTs.

References

[1] D. Dragoman, M. Dragoman, "Terahertz oscillations in semiconducting carbon nanotube resonant-tunneling diodes," *Physica E,* vol. 24, 282-289, 2004.

[2] J. Hao and G. Hanson, "Infrared and Optical Properties of Carbon Nanotube Dipole Antennas," *IEEE Trans. Nanotechnol.*vol.5, p.766-775, 2006.

[3] M. Radosavljevi´c, J. Lefebvre, and A. T. Johnson, "High-field electrical transport and breakdown in bundles of single-wall carbon nanotubes," *Phys. Rev. B, Condens. Matter*, vol. 64, no. 24, p. 241 307, Dec. 2001.

[4] Kyung-Hoae Koo, Hoyeol Cho, Pawan Kumar, and Krishna C. Saraswat, "Performance comparison between carbon nanotubes, optical and Cu for future high-performance on-chip interconnect applications," *IEEE Trans. Electron Devices*, vol.54, No. 12, pp 3206-3215, Dec 2007.

[5] J. Hone, M. Whitney, C. Piskoti, and A. Zettl, "Thermal conductivity of single-walled carbon nanotubes," *Phys. Rev. B, Condens. Matter*, vol. 59, no. 4, p. R2 514, 1999.

[6] Hong Li, and K. Banerjee, "High-frequency analysis of carbon nanotube interconnects and implications for on-chip inductor design," *IEEE Trans. Electron Devices*, vol. 56, No. 10, 2009.

[7] F. Li, H. M. Cheng, S. Bai, G. Su, and M. S. Dresselhaus, "Tensile strength of single-walled carbon nanotubes directly measured from their macroscopic ropes," *Appl. Phys. Lett.*, vol. 77, no. 20, pp. 3161–3163, 2000.

[8] P. J. Burke, "Luttinger Liquid Theory as a model of the gigahertz electrical properties of carbon nanotubes," *IEEE Trans. Nanotechnol.*, vol. 1, pp. 129–144, Sept. 2002.

[9] A. Naeemi and J. D. Meindl, "Design and performance modeling for single-walled carbon nanotubes as local, semiglobal, and global interconnects in gigascale integrated systems," *IEEE Trans. Electron Devices*, vol. 54, no. 1, pp. 26–37, Jan. 2007.

[10] N. Srivastava, H. Li, F. Kreupl, and K. Banerjee, "On the applicability of single-walled carbon nanotubes as VLSI interconnects," *IEEE Trans. Nanotechnol.*, vol. 8, no. 4, pp. 542–559, Jul. 2009.

[11] Azad Naeemi, Reza Sarvari and James D. Meindl, "Performance comparison between carbon nanotube and copper interconnects for gigascale integration (GSI)," *IEEE Trans. Electron Device Letters*, vol.26, No. 2, 2005

[12] F. Kreupl, A. P. Graham, G. S. Duesberg, W. Steinhoegl, M. Liebau, E. Unger, W. Hoenlein, "Carbon nanotubes in interconnect applications," *Microelectronic Eng.*, vol.64, pp.399-408, 2002.

[13] Navin Srivastava and Kaustav Banerjee, "Performance Analysis of Carbon Nanotube Interconnects for VLSI Applications," *IEEE/ACM International Conference on Computer-Aided Design*, ICCAD-2005, pp.383-390, 2005.

[14] D. M Pozar, *Microwave Engineering*; Wiley: New York, 1998.

Minimization of Output Fluctuation of a Fixed Speed IG Based Wind Farm without Capacitor by using PMSG

M.R.I. Sheikh,[1] and Zinat Tasneem[2,*]

[1,2]EEE Department, Rajshahi University of Engineering & Technology
Rajshahi-6204, Bangladesh
[*]tasneemzinat@gmail.com

Abstract—**Capacitors are considered as the prime element in a fixed speed Induction Generator (IG) based wind farm along with the other external controlling devices for providing the necessary reactive power compensation. But this paper proposes that the capacitors as well as the external controlling devices can be eliminated from the wind farm if Permanent Magnet Synchronous Generator (PMSG) is included in the system. The proposed model of PMSG works not only as a power producing generator but also as a controlling device for IG. A 50 MW wind farm has been considered for this study which consists of both IG and PMSG without any capacitors or controlling devices. Varying load condition has also been considered for a composite wind farm with Synchronous Generator (SG). Simulation results prove the effectiveness of the proposed system. Both the dynamic and transient analyses have been carried out by using the laboratory standard power system software package, PSCAD/EMTDC.**

Index Terms—**IG; PMSG; capacitor; variable load.**

I. Introduction

In recent years the world consumption of energy has increased enormously due to the massive industrialization which has been intensified rapidly in some geographical areas of the world remarkably in the countries of Asia. This consumption mostly depends on non-renewable energy sources, i.e. on fossil fuels and nuclear power resources. But these non-renewable energy sources are limited and they are the main reason of green house effect and environment pollution. Hence renewable energy sources are now rapidly taking their places as a clean and non-polluting energy source. Among the renewable energy sources, wind is a promising one. It is currently one of the most cost effective ways to produce electricity. Worldwide there are now over two hundred thousand wind turbines operating, with a total nameplate capacity of 282,482 MW as of end 2012 [1]. Since, the demand for wind energy is increasing day by day, it is important to assure that the power developed by the wind farm must be of good quality and the operation of the wind farm must be stable [2]. Since the wind speed is dynamic so the output parameters of a wind farm like the terminal voltage, output power, frequency etc fluctuate with the variable wind speed. To reduce this fluctuation, in conventional Induction Generator (IG) based wind farms, some external controlling FACTS (Flexible AC Transmission System) devices like STATCOM (Static Compensator), Unified Power Flow Controller (UPFC) etc are used. These are the non-producing element for a wind farm, they make the system more complicated and increase the system overall cost. Over recent years, a new scheme for the wind generators has become popular, which includes a Variable Speed Wind Turbine with Permanent Magnet Synchronous Generator (VSWT-PMSG). In this concept, PMSG can be directly driven by a wind turbine without gearbox and it is connected to the ac power grid through the power converters.

In our previous study [3-4], it is seen that PMSG can enhance the stability of the fixed speed wind farm if it is used in combination with IG. In [5] a PMSG based wind farm has been used to enhance the fault ride through capability of a fixed speed wind farm with a three-level IGBT converter-inverter set. In [6] a VSWT-PMSG has been used in HVDC connected offshore wind farm. Again a stabilization of a grid connected wind farm using PMSG is depicted in [2]. In these papers capacitor bank was used with IG to compensate for the reactive power. But capacitor banks are quite expensive and in case of varying wind speed, the reactive power demand from capacitor bank also become varying. Thus switching is needed, which reduces their lifetime. Therefore, this paper presents a unique technique to minimize the output fluctuation of a wind farm by using PMSG, which eliminates the need of external controlling devices and even the capacitors. The proposed model of PMSG is able to provide the necessary reactive power compensation and hence to maintain a constant voltage and less fluctuating power at the grid for which the capacitor bank is generally used along with IG. Variable load conditions have also been considered for a composite wind farm where IG and PMSG are used along with SG. It can be seen from the simulation results that PMSG itself works as a controlling device for the wind farm not only in dynamic and transient conditions but also in varying load conditions.

II. System Layout

Four cases have been considered in this study to show the effectiveness of the proposed wind farm model. These cases are explained in the following paragraphs.

A. Case-1, Wind Farm with IG

In the first case, a conventional wind farm with IG has been considered without any external controlling devices and capacitor bank. The total capacity of the farm is 50 MW. Fig. 1(a) shows the schematic diagram of case-1.

B. Case-2, Wind Farm with IG and Capacitor Bank

In the second case, conventional IG has been used along with a capacitor bank. This farm also has a capacity of 50 MW. But it doesn't include any controlling devices. Fig. 1(b) shows the schematic diagram of this farm.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

C. Case-3, Wind Farm with IG and PMSG

Case-3 includes a wind farm with both IG and PMSG. No capacitor bank or other controlling FACTS devices have been used. Here, a PMSG of 5 MW has been considered along with an IG having capacity of 45 MW. Thus the total capacity of the farm is 50 MW. Fig. 1(c) shows the schematic diagram of this farm.

D. Case-4, Wind Farm with IG, PMSG and SG

To observe the varying load conditions, a composite wind farm with IG, PMSG and SG have been considered in case-4. An IG of 5 MW along with a PMSG of 3 MW has been used with an SG of 100 MW capacity. Thermal governor is used to control the SG. SG depicted in [7] has been used. The schematic diagram is shown in Fig. 1(d).

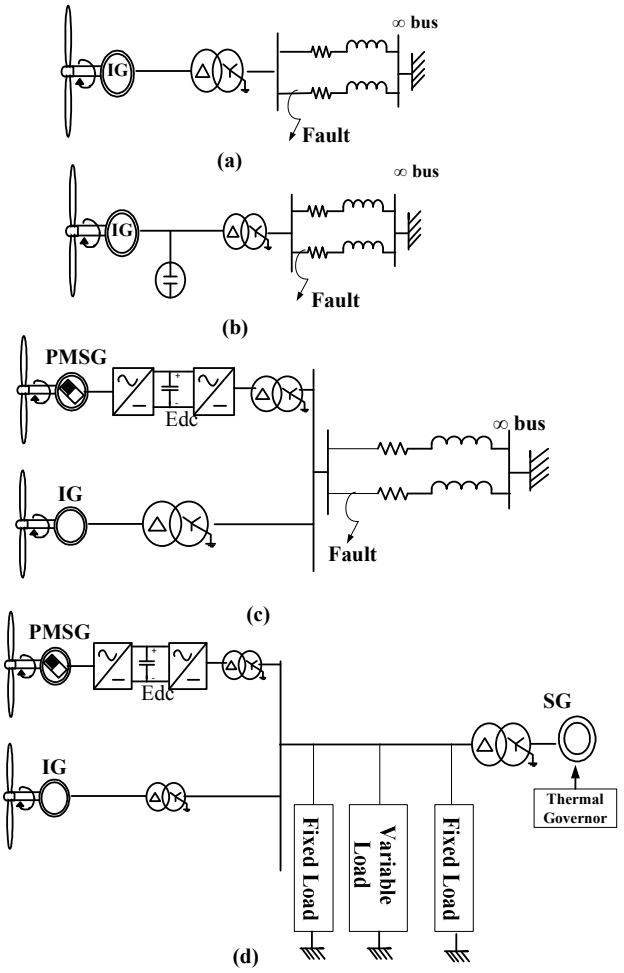


Fig. 1 Schematic diagram for (a) case-1, (b) case-2, (c) case-3 and (d) case-4.

III. PMSG Model

Fig. 2 shows a VSWT-PMSG system which is modelled with a fully controlled frequency converter. The PMSG model available in the package software PSCAD/EMTDC is used [8] in this study. The frequency converter consists of a generator side AC/DC converter, a DC link capacitor and a grid side DC/AC inverter. Each of the converter/inverter is a standard three phase two-level unit, composed of six insulated gate bipolar junction transistor (IGBTs) and anti parallel diodes. The full-rating power converter which is made up of two back- to-back IGBT bridges (the generator side converter and grid side inverter) are linked by a DC bus. The three phase AC output of the generator is rectified, and the DC output from the rectifier is fed to an IGBT-based grid side inverter. Its output is fed to a step-up transformer and then to the grid. The rating of the power converters corresponds to the rated power of generator plus losses. The power converters control the power flow to the grid.

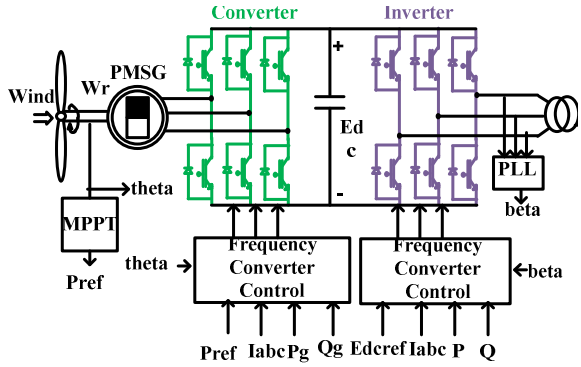


Fig. 2 Electrical scheme of VSWT-PMSG.

IV. Simulation Results

Simulations have been done by using Power System Computer Aided Design / Electromagnetic Transient including DC (PSCAD/EMTDC) program [8] for 350 seconds. The timing step of the simulation is chosen to be 0.01 sec. As mentioned earlier, four case studies have been considered in this study. The results have been shown for dynamic, transient analysis for the first three cases and variable load conditions for the last case.

A. Dynamic Stability Analysis

Fig. 3 shows the variable wind speed data used in the simulation. Fig. 4 shows the grid side voltage response for the first three cases. It can be seen that PMSG maintains almost a constant voltage without the capacitor bank of IG. The real power response is shown in Fig. 5. Here also the third case, that is IG and PMSG has the best response among the three cases, since this case provides the necessary reactive power compensation to the grid as can be seen from Fig. 6. And it also helps to maintain the frequency response stable as shown in Fig. 7.

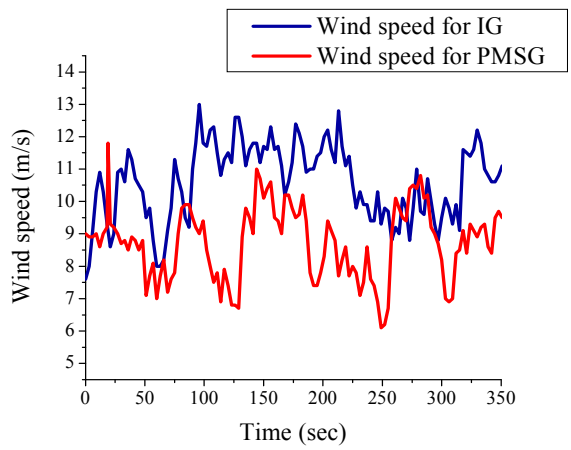


Fig. 3 Wind speed data.

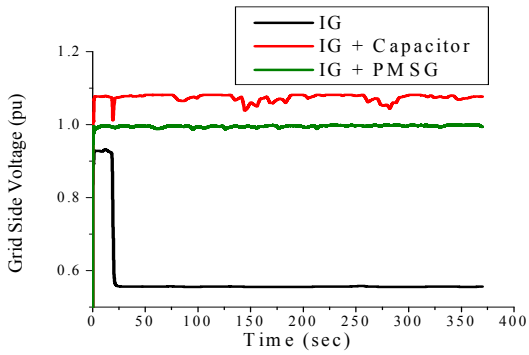


Fig. 4 Grid side voltage.

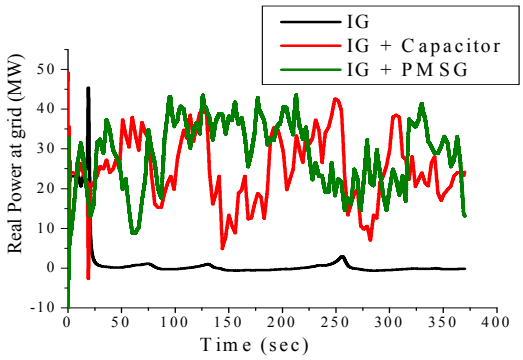


Fig. 5 Real power available at the grid.

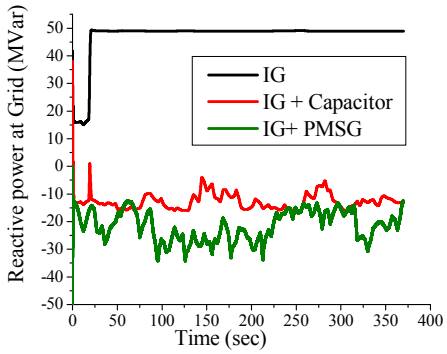


Fig. 6 Reactive power available at the grid.

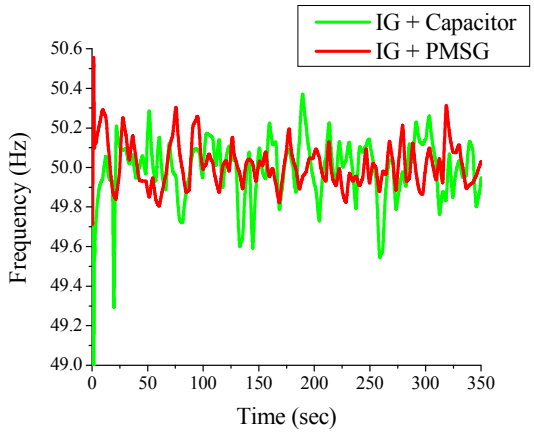


Fig.7 Frequency response.

B. *Transient Stability Analysis*

In this study, the dynamic wind speed data shown in Fig. 3 is used for transient stability analysis to make the response of the proposed system more practical. For this, a three line to ground (3LG) fault has been considered, which occurs at 27s, and the fault is cleared within 0.05s. Fig. 8 shows the grid side voltage response, where PMSG is able to recover the voltage to its normal state after the clearance of the fault just like the capacitors.

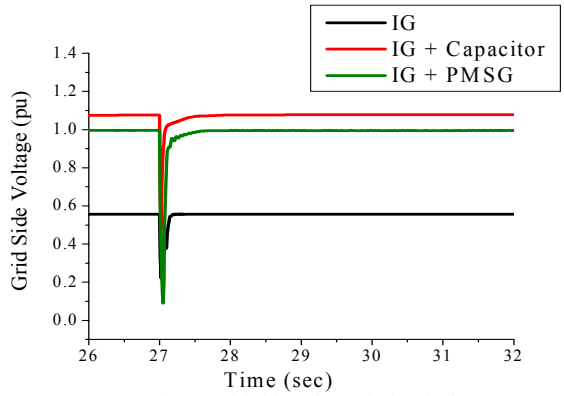


Fig. 8 Grid side voltage during fault.

Fig. 9 and Fig. 10 show the real and reactive power response at the grid respectively. As a results of the reactive power compensation during fault, in the third case (PMSG+IG), the real power attains the normal value after the fault as quickly as possible.

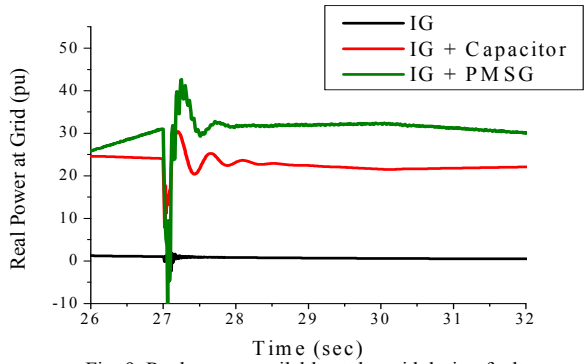


Fig. 9 Real power available at the grid during fault.

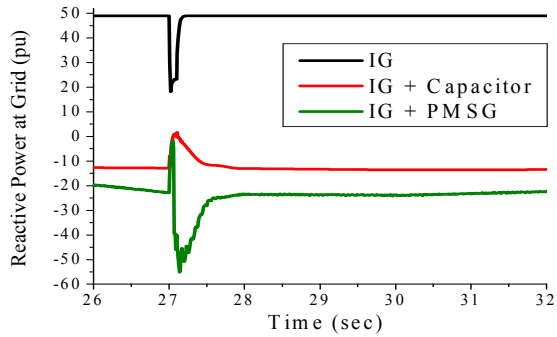


Fig. 10 Reactive power available at the grid during fault.

C. *Variable Load Conditions*

In case-4, variable load conditions have been considered. A 100 MW fixed load has been considered, where a 2MW load has been added at 50th second, a 3 MW load has been added at 60th second. At 70th second, 4 MW load has been reduced and at 80th second the 1 MW load has been removed. The variation of the power in the load side is shown in Fig. 11.

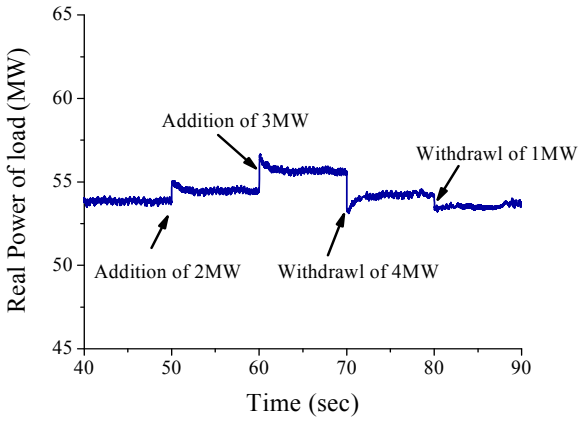


Fig. 11 Variation of real power with the variation of load.

The variation of the real power of SG and the frequency are shown in Fig. 12 and Fig. 13 respectively. As can be seen from these figures, there is an acceptable variation in frequency for the variation of load. This has been done by PMSG without any capacitor bank connected with IG. Moreover it doesn't need other controlling devices.

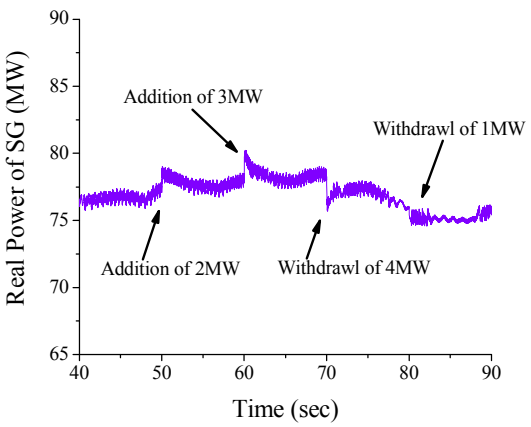


Fig. 12 Variation of SG's real power with the variation of load.

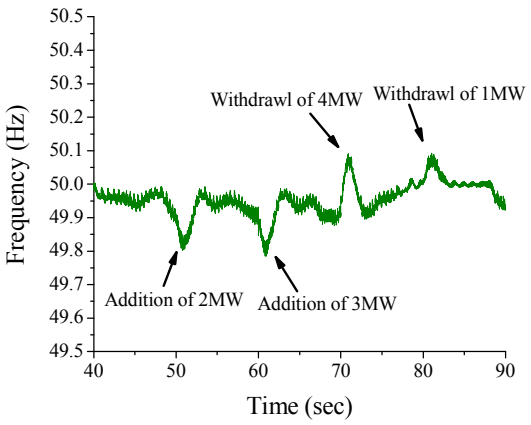


Fig. 13 Variation of frequency with the variation of load.

So, it is seen that PMSG can ensure the reliable operation of a fixed speed IG based wind farm. Therefore capacitor bank or other FACTS devices can be avoided. The simulation results show the effectiveness of the proposed wind farm during dynamic, transient and variable load conditions.

V. Conclusions

The proposed model of the wind farm consisting of both IG and PMSG has the ability to reduce the output fluctuations of a conventional fixed speed wind farm. To attain this, capacitor banks and other external controlling devices can be ignored. Simulation results prove that the PMSG itself can work as a generator as well as a controlling device and also it eliminates the need of the capacitor bank. Thus a large scale wind farm can be established by using the proposed technique. This method is equally effective for both new wind farms and for existing wind farms where expansion of installed capacity is required. Therefore, it is concluded that, the proposed system is a good tool for power system stability improvement of fixed speed wind generators by contributing to the electrical output and eliminating the costs of external controlling devices and capacitor banks.

References

[1] "Global Wind Report Annual market update 2012". *Global Wind Energy Council*. Retrieved 23 April 2013.

[2] M. Rosyadi, M.R.I. Sheikh, S.M. Muyeen, R. Takahashi, and J. Tamura, "A Control Strategy of Direct-Drive Permanent Magnet Wind Generator for Wind Farm Stabilization," *Annual Paper Meeting, Society of Rotatory Mission, Japan*, pp. 7-12, Aug. 2010.

[3] M.R.I. Sheikh, and Z. Tasneem, "Transient Stability Improvement of a Fixed Speed Wind Driven Power System Using Permanent Magnet Synchronous Generator," in *Proc. ICME* 2013, paper 224.

[4] Z. Tasneem, and M.R.I. Sheikh, "Improvement of Fixed Speed Wind Driven Power System Stability by Using Variable Speed PMSG Based Wind Generator," in *Proc. ICME* 2013, paper 184.

[5] S.M. Muyeen, R. Takahashi, T. Murata, and J. Tamura, "A Variable Speed Wind Turbine Control Strategy to Meet Wind Farm Grid Code Requirements" *IEEE Transactions on Power Systems,* Vol. 25, No. 1, February 2010.

[6] S.M. Muyeen, R. Takahashi, and J. Tamura, "Operation and Control of HVDC-Connected Offshore Wind Farm," *IEEE Transactions on Sustainable Energy,* Vol.1, No. 1, April 2010.

[7] M.R.I. Sheikh "Stabilization of a Grid Connected Wind Farm by Using SMES" a PhD thesis, Kitami Institute of Technology, Japan, September 2010.

[8] "PSCAD/EMTDC Manual," Manitoba HVDC Research Center, 1994.

A Probabilistic Index for Estimating Frequency Response of a Power System with High Wind Power Penetration

Nahid-Al-Masood[1, 2], Ruifeng Yan[1] and Tapan Kumar Saha[1]

[1]School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, Australia

[2]Department of Electrical and Electronic Engineering, BUET, Dhaka, Bangladesh

n.masood@uq.edu.au, ruifeng@itee.uq.edu.au, saha@itee.uq.edu.au

Abstract—**Frequency response is a key concern for power system security. Usually, it is estimated in a deterministic approach. But in most cases, power systems behave stochastically. Moreover, presence of wind power imposes further variability on generation profile of a power system. Hence, it appears more pragmatic to address these uncertainties during an evaluation of frequency response. This paper develops an index to assess frequency response of a power system with high wind power penetration. A probabilistic method is used for this purpose. K-means algorithm, a well-known data mining technique is employed. The developed index is called a 'probabilistic frequency response index'. A realistic interconnected power system is investigated for a case study. Impact of increased wind power penetration on probabilistic frequency response index is analysed.**

Index Terms—**Frequency response, Headroom, Inertia, K-means algorithm, Wind power.**

I. Introduction

Secure operation of an electric power system relies on appropriate balance between generation and load. One of the major objectives of a power system operator is to maintain system frequency within an acceptable range. However, maintaining system frequency within a desired limit is a challenge due to continuous variation of consumers' demand. Moreover, occasional events such as sudden loss of a large generating unit or load can cause an abrupt deviation in system frequency. Frequency response is a measure of the ability of a power system to arrest and then stabilize rapid change in frequency [1]. Frequency sensitivity index (β, in MW/0.1 Hz) is used as one of the measuring parameters of frequency response [1]. It is estimated by relating a mismatch between load and generation to a corresponding sudden change in system frequency.

Integration of renewable energy sources such as wind power is increasing rapidly in many countries. Addition of wind power replaces conventional synchronous generators. Unlike synchronous generators, modern wind turbine generators do not provide automatic frequency support when there is an imbalance between load and generation. Therefore, inclusion of a substantial amount of wind power plants complicates the frequency response issue of a power system.

A significant number of researches on power system frequency response have been reported in the current literature [2]-[8]. Computer models are developed in [2] to estimate an impact of Western Electricity Coordinating Council (WECC) utility's inertia on the system frequency response. Frequency transients are observed and recorded for this purpose. When there is a generator outage in a hydro based system, frequency response may exacerbate due to temporary decrease of power output by hydro generators. Hydro-fossil coordination can improve the situation [3]. A study from the utilisation of phasor measurement units is presented in [4] to estimate inertial frequency response of the Great Britain (GB) transmission network. Reduction of inertial response due to inclusion of renewable energy is observed. This method utilizes historical data of large generator losses.

Non-synchronous penetration level in the Ireland grid can be increased up to 50% in current network structure [5]. It is suggested that synthetic wind inertia and adjustment in protective relay settings can facilitate 75% non-synchronous generation. A study on frequency response of California and WECC due to high wind and solar penetration is presented in [6]. Here, influences of various generation mix and system headroom on frequency response are addressed. Electric vehicle (EV) loads are growing in GB power system. Different EV charging strategies such as "dumb charging" and "off-peak charging" may be useful to provide primary frequency response [7]. Inertial response of a variable speed wind turbine depends on local wind speed. Consequently, there exists an uncertainty in inertial response capability of a highly wind penetrated power system [8]. However in [8], there is no guideline about stochastic measurement of frequency response. Statistical analysis of an island power system in [9] demonstrates a reduction of overall system stiffness due to excessive wind power. Again, there is no clear advice on statistical determination of frequency response characteristics.

It seems that in most researches, frequency response is estimated at a particular load level. Historical data of generator trips are sometimes employed. Thus, frequency response characteristics reported in literature are deterministic in nature. However, in reality, generation and load pattern of a power system exhibit probabilistic behaviour. No such frequency response index, which reflects the probabilistic nature of a power system, is presented in literature yet.

This paper proposes a methodology to develop a probabilistic index for estimating frequency response of a power system. K-means data clustering algorithm is used for this purpose. The proposed methodology is applied on an interconnected power system. This paper then investigates an effect of high wind power on probabilistic frequency response index of the studied power system.

II. K-means Algorithm

In this paper, K-means algorithm [10] is used to form clustered load scenarios. Assume that historical load data of a power system are to be divided into *NL* load levels. That is,

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

NL represents the total number of clusters. Steps of K-means algorithm are shown below.

Step-1: Initial cluster means (M_i) are selected, where i denotes the i-th cluster and i = 1, 2, … …, *NL*.

Step-2: The distance (D_{ki}) from each load point (L_k) to each cluster mean (M_i) is calculated by

$$D_{ki} = \left| M_i - L_k \right| \tag{1}$$

where k = 1 to total number of load points in historical load data (*NT*).

Step-3: Load points are grouped by assigning them to the nearest cluster.

Step-4: A new mean of i-th cluster is recalculated using

$$M_i = \frac{\sum_{k \in IC} L_k}{NI_i} \tag{2}$$

where NI_i refers to the total number of load points in i-th cluster and *IC* denotes a set of the load points in i-th cluster.

Step-5: Steps 2 to 4 are repeated until all cluster means remain unchanged between successive iterations. At the end of iterations, each cluster is represented by its centre value. This centre is effectively the mean value of all load points inside a cluster.

Step-6: Probability of occurrence of i-th cluster (P_i) is computed by

$$P_i = \frac{NI_i}{NT} \tag{3}$$

K-means algorithm is a local optimal sorting approach. Therefore, it may lead to multi-solutions. To avoid it, number of clusters and initial cluster means should be specified properly.

III. Methodology

This section presents a methodology to develop a probabilistic frequency response index of a power system. Steps of the proposed methodology are described below.

Step-1: Historical load data of a power system are clustered in 24 hourly load scenarios using the K-means algorithm.

Step-2: For a clustered load scenario, different synchronous generators are tripped, one at a time, to create a disturbance. Dynamic simulation is performed in PSS®E simulation environment [11]. PSS®E software is widely used by transmission system planners and electricity market operators. In PSS®E dynamic study, 100% constant current load model is used for active power and 100% constant impedance load model is used for reactive power. For a generator, dynamic models of governor, exciter and stabilizer are used. A Python [12] script is written to automate the dynamic simulation process.

For a tripping of any synchronous generator, rotational speeds of other generators are recorded. Then the equivalent system frequency (f_{eq}) is calculated using (4) [6]

$$f_{eq} = \frac{\sum_{k=1}^{k=n} (S_k \times \omega_k)}{\sum_{k=1}^{k=n} S_k} \tag{4}$$

where S_k is the MVA rating of k-th generator, ω_k is the speed of k-th generator (in p.u.) and n is the total number of synchronous generators.

Step-3: Dynamic simulation after a disturbance is performed for a time frame t_0 to t_f. Here, t_0 refers to the moment when a disturbance is initiated and t_f denotes the final simulation time. Equivalent frequencies at different time samples (within t_0 to t_f) are computed using (4). The minimum equivalent frequency is then calculated, which represents the frequency nadir after a generator trip.

Step-4: System frequency nadirs (in Hz) for different synchronous generator trips in a specific clustered load scenario are recorded. Frequency nadir (*FN*) is modelled as a function of power outage due to a synchronous generator trip (P_o, in MW) using linear regression method. It can be expressed by (5) [13]

$$FN = (S_p \times P_o) + f_c \tag{5}$$

where S_p is the slope (in Hz/MW) and f_c is the constant part of linear regression model (in Hz).

The frequency sensitivity index (β, in MW/0.1 Hz) is calculated using (6) [13].

$$\beta = \frac{1}{\left| S_p \right|} \times 0.1 \tag{6}$$

Step-5: Frequency sensitivity indices for all 24 clustered load scenarios are evaluated using steps 2 to 4. Finally, a probabilistic frequency response index of a power system (β_{sys}) is estimated by

$$\beta_{sys} = \sum_{i=1}^{i=NL} (\beta_i \times P_i) \tag{7}$$

where β_i is the frequency sensitivity index of i-th clustered load scenario, P_i is the probability of occurrence of i-th cluster and *NL* = 24.

IV. Overview of the Studied Power System

A practical power system is studied in this paper. Installed capacity of the studied system is around 3,000 MW. In transmission level, 220 kV and 110 kV are used. The system has 250 buses, 65 synchronous generators and 4,000 circuit kilometres of transmission lines. The system is connected to its adjacent power network through a high voltage back to back DC interconnection. The maximum limit for both power export and import through this interconnection is 500 MW.

There are two large-scale wind farms in operation in the studied system. Capacity of the 1st wind farm is 140 MW and that of the 2nd is 170 MW. Another wind farm of capacity 300 MW is under construction and it is expected to be connected to the grid by the next few years. Total installed capacity of wind power will then be 610 MW. In this study, all wind farms are modelled as Type-3 (Doubly-fed induction generator) wind turbine generators. None of the wind farms has synthetic inertia and active power control (APC) features.

Wind power penetration level (*WPL*) of the studied system is estimated using (8) [13]

$$WPL = \frac{P_{wind}}{P_{sync} + P_{wind} + P_{import}} \tag{8}$$

where P_{wind} specifies the wind power in the studied system, P_{sync} is the power generation from synchronous generators and P_{import} denotes the power import to the studied system from the adjacent network through interconnection. In this paper, power export to the adjacent network from the studied system through interconnection (P_{export}) is modelled as a load supplied by the generators of the studied system. So, it is not shown as a separate quantity in (8).

V. Results and Discussion

Case studies are carried out for three different seasons, summer, winter and spring. The maximum, minimum and average load of the studied power system fluctuate from each other in various seasons. Historical hourly load data for last 8 years (2006-2013) are investigated. Table I shows a seasonal load profile of the system. It is observed that the system is of winter peak in nature.

TABLE I
SEASONAL LOAD PROFILE

Season	Maximum load (MW)	Minimum load (MW)	Average load (MW)
Summer	1460	690	1070
Winter	1760	820	1260
Spring	1640	685	1130

In this paper, a probabilistic frequency response index (β_{sys}) is separately estimated for three seasons. In each season, 24 hourly clustered load scenarios are formed. Estimation of β_{sys} and impact of high wind power generation on β_{sys} in different seasons are presented in the following subsections. Significant observations are also discussed.

A. Probabilistic Frequency Response Index

(i) Summer: 24 hourly clustered load scenarios are constructed using (1) and (2). Table II shows the clustered load, corresponding synchronous generation, wind power and interconnection power flow for summer.

Probability of occurrence of each cluster is calculated using (3). For each load level, dynamic simulation of the system is performed for different generator trips. Frequency sensitivity indices are computed using (4)-(6). Table III shows dynamic simulation results and probability of occurrence of different clusters during summer.

TABLE II
CLUSTERED LOAD PROFILE OF SUMMER

Cluster No.	Load (MW)	P_{sync} (MW)	P_{wind} (MW)	P_{import} (MW)	P_{export} (MW)
1	967	423	131	413	0
2	911	456	32	423	0
3	914	479	5	430	0
4	1035	510	78	447	0
5	1050	528	53	469	0
6	1080	1085	48	0	53
7	1202	965	63	174	0
8	1277	734	75	468	0
9	1267	1650	20	0	403
10	1374	1109	42	223	0
11	1310	1557	125	0	372
12	1252	1284	99	0	131
13	1279	1397	57	0	175
14	1362	1684	82	0	404
15	1208	1512	40	0	344
16	1202	1593	78	0	469
17	1169	1125	120	0	76
18	1235	1238	125	0	128
19	1141	1334	60	0	253
20	1158	1200	92	0	134
21	1132	1368	36	0	272
22	1093	1434	78	0	419
23	1102	911	124	67	0
24	1056	847	121	88	0

TABLE III
FREQUENCY SENSITIVITY INDICES AND CLUSTER PROBABILITIES

Cluster No.	β (MW/0.1 Hz)	Probability	Cluster No.	β (MW/0.1 Hz)	Probability
1	5.9948	0.0105	13	11.7980	0.0249
2	5.7097	0.0229	14	15.9160	0.0082
3	6.5236	0.0345	15	14.8500	0.0431
4	8.6207	0.0504	16	15.6495	0.0580
5	6.9784	0.0532	17	10.1225	0.0586
6	8.5172	0.0579	18	10.8284	0.0557
7	10.5241	0.0538	19	12.7178	0.0591
8	8.0978	0.0327	20	10.2923	0.0557
9	16.9434	0.0163	21	13.5722	0.0565
10	9.6787	0.0028	22	13.1475	0.0562
11	16.0359	0.0404	23	10.5031	0.0529
12	11.8821	0.0468	24	10.6338	0.0489

Now using (7), probabilistic frequency response index during summer ($\beta_{sys,sm}$) is found to be 11.15 MW/0.1 Hz. That is, on an average 11.15 MW imbalance between generation and load causes 0.1 Hz frequency deviation during summer.

(ii) Winter and spring: Following the similar procedure of summer, probabilistic frequency response index during winter and spring ($\beta_{sys,wt}$ and $\beta_{sys,sp}$ respectively) are computed. Table IV presents the results.

TABLE IV
PROBABILISTIC FREQUENCY RESPONSE INDICES

Season	Notation	Value (MW/0.1 Hz)
Summer	$\beta_{sys,sm}$	11.15
Winter	$\beta_{sys,wt}$	10.50
Spring	$\beta_{sys,sp}$	17.55

It is observed from Table IV that β_{sys} is minimum during the winter season. That is, this power system becomes most vulnerable in terms of frequency response during winter. During spring, β_{sys} is the maximum. So, the ability of the system to arrest frequency excursion after a disturbance is highest during spring, lowest during winter and in-between during summer.

Variation of β_{sys} during different seasons can be explained using the concepts of system inertia and headroom [5], [6]. When there is an imbalance between generation and load due to a generator trip, a portion of the stored kinetic energy from committed synchronous generators' rotor are released immediately. This is known as inertial response of a power system and it dominates initial frequency decline after a disturbance. Total inertia (IR) of a power system can be computed by

$$IR = \sum_{j=1}^{j=n} (S_j \times H_j) \tag{9}$$

where S_j is the rated MVA of j-th synchronous generator, H_j denotes the inertia constant (in s) of j-th synchronous generator and n is the total number of committed synchronous generators.

Headroom (HR) is an amount of generation that can be provided by committed synchronous generators (for arresting frequency decline) after a disturbance using governor control function. Frequency nadir after a generator trip is determined by the combination of inertia and headroom of a system. Headroom can be calculated using

$$HR = \sum_{j=1}^{j=ng} (P_{max,j} - P_{gen,j}) \tag{10}$$

where $P_{max,j}$ refers to the rated capacity of j-th synchronous generator (in MW), $P_{gen,j}$ is the power generation of j-th synchronous generator (in MW) and ng denotes the total number of synchronous generators with governor control enabled.

For the studied power system, the average inertia (IR_{sys}) and headroom (HR_{sys}) are estimated using

$$IR_{sys} = \sum_{i=1}^{i=NL} (IR_i \times P_i) \tag{11}$$

$$HR_{sys} = \sum_{i=1}^{i=NL} (HR_i \times P_i) \tag{12}$$

where IR_i and HR_i are the inertia and headroom of i-th clustered load scenario respectively.

Table V presents average inertia and headroom of the studied system for different seasons. It is observed that IR_{sys} and HR_{sys} are the maximum during spring. As a consequence, $\beta_{sys,sp}$ is the highest amongst three seasons. IR_{sys} of summer is less than that of winter. But HR_{sys} of summer is more than that of winter. However, $\beta_{sys,sm}$ is higher than $\beta_{sys,wt}$.

TABLE V
AVERAGE INERTIA AND HEADROOM

Season	IR_{sys} (MWs)	HR_{sys} (MW)
Summer	7900	332
Winter	8030	57
Spring	9770	600

Thus, it can be revealed that headroom has dominating impact on the probabilistic frequency response index.

B. Impact of Wind Power on Probabilistic Frequency Response Index

Impact of significant wind power penetration on the β_{sys} of the studied power system is investigated in this paper. Steps of the investigation are presented below.

Step-1: During a specific season, for each cluster, wind power penetration is increased incrementally. Penetration is expanded up to 610 MW, in a step of 30 MW at a time. To facilitate additional wind power, a certain amount of synchronous generators are replaced by wind machines. For each wind penetration case, average wind power penetration level (WPL_{sys}) is evaluated by

$$WPL_{sys} = \sum_{i=1}^{i=NL} (WPL_i \times P_i) \tag{13}$$

where WPL_i refers to the wind power penetration level at i-th clustered load scenario.

Step-2: In each wind penetration case, frequency sensitivity index (β) is determined using (4)-(6).

Step-3: Probabilistic frequency response indices (β_{sys}) are computed using (7) for different WPL_{sys}.

Step-4: β_{sys} is modelled as a function of WPL_{sys}. Model parameters are estimated using regression technique.

Fig.1 depicts a regression model between β_{sys} and WPL_{sys} during summer. Using extensive dynamic simulations, a mathematical relationship between β_{sys} and WPL_{sys} are estimated for each season. Relationships are given by (14), (15) and (16) for summer, winter and spring respectively.

$$\beta_{sys,sm} = -0.16(WPL_{sys,sm}) + 12.96 \tag{14}$$

$$\beta_{sys,wt} = -0.15(WPL_{sys,wt}) + 10.95 \tag{15}$$

$$\beta_{sys,sp} = -0.35(WPL_{sys,sp}) + 20.94 \tag{16}$$

Here, $WPL_{sys,sm}$, $WPL_{sys,wt}$ and $WPL_{sys,sp}$ denote average wind penetration level during summer, winter and spring respectively. It is observed that β_{sys} decreases with an increases of WPL_{sys} during different seasons. So, frequency response of the studied power system declines gradually due to high wind power penetration.

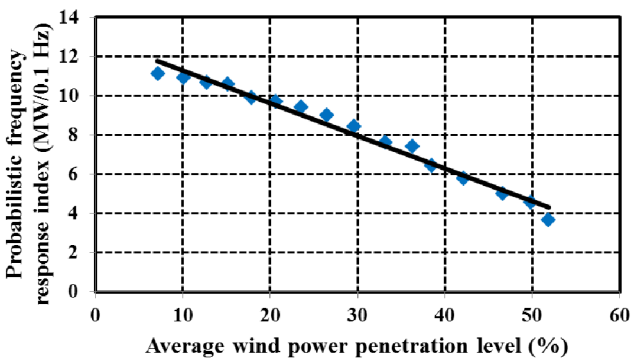


Fig. 1 Regression model between probabilistic frequency response index and average wind power penetration level during summer.

VI. CONCLUSIONS

This paper develops a probabilistic frequency response index (β_{sys}) for a power system using K-means clustering algorithm. Power export and import through interconnection at different clustered load scenarios affect the number of committed synchronous generators. As a consequence, system inertia and headroom are influenced. Frequency response index is thus found to be affected by both inertia and headroom.

It is observed that the studied power system shows more susceptible frequency response with increased wind power penetration. An estimated relationship between β_{sys} and average wind penetration level (WPL_{sys}) shall help a power system operator to assess the frequency behaviour in case of high wind power generation. In continuation to this study, further works will be done to find the implications of both wind and PV generation on probabilistic frequency response index, which can be useful for any power system.

ACKNOWLEDGMENT

The authors would like to thank the Australian Research Council and partner industries for their financial support through a Linkage Project.

REFERENCES

[1] "Review of the Recent Frequency Performance of the Eastern, Western and ERCOT Interconnections", North American Electric Reliability Corporation, Dec. 2010 [On line]. Available: http://www.nerc.com/FilingsOrders/us/FERCOrdersRules/Interconnection_Frequency_Performance.pdf [Accessed: 12-May-2014].

[2] D.P. Chassin, Z. Huang, M.K. Donnelly, C. Hassler, E. Ramirez and C. Ray, "Estimation of WECC System Inertia Using Observed Frequency Transients", *IEEE Trans. Power Systems*, vol.20, no.2, pp. 1190–1192, May 2005.

[3] L. Gao, J. Xia and Y. Dai, "Analysis of Power System Frequency Responses with Hydro Turbines Incorporating Load Shedding" *in Proc. 2010 The 5th IEEE Conference on Industrial Electronics and Applications*, pp. 893–897.

[4] P.M. Ashton, G.A. Taylor, A.M. Carter, M.E. Bradley and W. Hung, "Application of Phasor Measurement Units to Estimate Power System Inertial Frequency Response", *in Proc. 2013 IEEE Power and Energy Society General Meeting*, pp. 1–5.

[5] J. O'Sullivan, A. Rogers and A. Kennedy, "Determining and Implementing an Approach to System Frequency and Inertial Response in the Ireland and Northern Ireland Power System", *in Proc. 2011 IEEE Power and Energy Society General Meeting*, pp. 1–6.

[6] N.W. Miller, M. Shao, S. Venkataraman, C. Loutan and M. Rothleder, "Frequency Response of California and WECC under High Wind and Solar Conditions", *in Proc. 2012 IEEE Power and Energy Society General Meeting*, pp.1–8.

[7] M. Yunfei, W. Jianzhong, J. Ekanayake, N. Jenkins and J. Hongjie, "Primary Frequency Response from Electric Vehicles in the Great Britain Power System", *IEEE Trans. Smart Grid*, vol.4, no.2, pp. 1142–1150, June 2013.

[8] L. Ruttledge, N.W. Miller, J. O'Sullivan and D. Flynn, "Frequency Response of Power Systems with Variable Speed Wind Turbines", *IEEE Trans. Sustainable Energy*, vol.3, no.4, pp. 683–691, Oct. 2012.

[9] I.M. Dudurych, "Statistical Analysis of Frequency Response of Island Power System under Increasing Wind Penetration", *in Proc. 2010 IEEE Power and Energy Society General Meeting*, pp. 1–6.

[10] R. Billinton and W. Li, "Reliability Assessment of Electric Power Systems using Monte Carlo Methods". New York: Plenum Press, 1994, pp. 22–96.

[11] "Power System Simulator for Engineering" [On line]. Available: http://www.energy.siemens.com/hq/en/services/power-transmission-distribution/power-technologies-international/software-solutions [Accessed: 02-May-2014].

[12] "Python Programming Language" [On line]. Available: https://www.python.org [Accessed: 02-May-2014].

[13] Nahid-Al-Masood, R. Yan and T.K. Saha, "Frequency Response with Significant Wind Power Penetration: Case Study of a Realistic Power System", *accepted in 2014 IEEE Power and Energy Society General Meeting*.

Estimation of Maximum Wind Power Penetration Level to Maintain an Adequate Frequency Response in a Power System

Nahid-Al-Masood[1, 2], Ruifeng Yan[1] and Tapan Kumar Saha[1]

[1]School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, Australia

[2]Department of Electrical and Electronic Engineering, BUET, Dhaka, Bangladesh

n.masood@uq.edu.au, ruifeng@itee.uq.edu.au, saha@itee.uq.edu.au

Abstract— **Wind energy has an immense potential to play a vital role with conventional energy sources. Therefore, integration of wind power is increasing in many power systems. Unlike synchronous generators, modern wind farms have lack of automatic frequency support capability following a disturbance. Thus, an addition of wind power does impact the frequency response of a power system. Inadequate frequency response may occur due to presence of a large wind generation. This paper proposes a methodology for estimating the maximum wind power penetration level to ensure an adequate frequency response in a power system. Frequency nadir and rate of change of frequency after an outage of the largest generator of a power system are taken into consideration. Impact of two different wind power integration strategies (direct replacement of a synchronous generator with a wind machine and 2/3 de-commitment, 1/3 re-dispatch approach) on the maximum wind power penetration level are investigated. This paper also attempts to identify a system parameter that could most significantly influence a penetration of wind power.**

Index Terms—**Frequency response, Frequency sensitivity index, Headroom, Inertia, Wind power.**

I. Introduction

Conventional energy sources such as coal, oil and natural gas are fossil fuels and have limited availability. Moreover, they emit carbon dioxide (CO_2) and cause environmental pollution. On the contrary, renewable energy sources like wind and solar are amply available and they are pollution free. Among them, wind power has enormous possibility to become one of the major energy sources in many countries around the world. Global installed capacity of wind power at the end of 2013 is around 318 GW. It has increased by 35 GW than that of 2012 [1]. However, wind power penetration in a conventional system cannot be increased arbitrarily due to some technical issues. One major issue is the lack of sufficient generation reserve margin in a highly wind penetrated system. The other issue is associated with power system frequency response. Committed synchronous generators provide inertial and governor response after a generator in a system trips. In contrast, wind farms do not automatically provide these functions. Therefore, the maximum wind power, which can be integrated in a conventional system, is a major concern.

A number of investigations have been performed to identify the maximum allowable penetration level of wind power in a conventional system [2]-[9]. The Maximum wind penetration level is determined based on transient stability limit and frequency security constraints in [2]. Voltage stability margin and system capacity are employed for maximizing the wind penetration level in [3]-[5]. In [6], Particle Swarm Optimization (PSO) technique is used to develop an algorithm for attaining the maximum wind energy penetration level. A method incorporating modal analysis and Q-V curves for determining the maximum size of wind farms is presented in [7]. To identify the maximum wind penetration level, an integrated numerical algorithm using available information of the corresponding electricity generation system is proposed in [8]. A fast computational algorithm based on frequency response to calculate the maximum wind penetration level is presented in [9].

It is evident from the current literature that addition of wind power in a conventional system influences both frequency nadir (the minimum frequency following a disturbance) and rate of change of frequency (*ROCOF*) of a power system. However, none of the researches has considered these two criteria simultaneously in estimating the maximum wind power penetration.

Up to now, two different wind integration strategies have been reported in literature. In the first strategy, a certain quantity of synchronous generation is replaced with an equal amount of wind generation [2]. In the second strategy, wind power is added using 2/3 de-commitment, 1/3 re-dispatch approach (2/3-1/3 rule) [10]. This means that for every 3 MW of additional wind production, there is a 2 MW de-commitment of synchronous generation and a 1 MW reduction in dispatch of synchronous generation. In both strategies, system demand is assumed to be unchanged. Impact of these two approaches on the maximum wind power penetration is yet not found to be reported in literature.

This paper develops a methodology to estimate the maximum allowable wind power penetration level in a power system to maintain an acceptable frequency response. Both frequency nadir and *ROCOF* are taken into account to assess the frequency response after a disturbance. The proposed methodology is applied for a realistic power system. Six typical load scenarios based on different seasons are studied. Effect of different wind integration strategies on the maximum wind power penetration is investigated in this paper. A system parameter that could highly influence the maximum wind power penetration is also determined.

II. Key Terminologies

The important terminologies used in this paper are described below.

(i) Wind power penetration level (*WPL*): It is a ratio of total wind power generation to total load and losses. Alternatively, it can be computed using (1) [11].

$$WPL = \frac{P_{wind}}{P_{sync} + P_{wind} + P_{import}} \tag{1}$$

where P_{wind} is the total wind power in a system, P_{sync} is the total power generation from synchronous generators and

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

P_{import} denotes the power import through an interconnection. In this paper, power export through an interconnection is modelled as a load supplied by the synchronous generators of an exporting power system. Hence, it is not shown as a separate quantity in (1).

(ii) Maximum wind power penetration level (WPL_{max}): It is the highest limit of wind power that can be added in a power system maintaining system frequency or rate of change of frequency (after a disturbance) within an acceptable margin.

(iii) Frequency sensitivity index (β): It describes the responsiveness of system frequency when there is an imbalance between generation and load. It is measured in MW/0.1 Hz.

(iv) System inertia (IR): It is a measure of total stored kinetic energy in a power system. It can be computed by

$$IR = \sum_{i=1}^{i=n} (S_i \times H_i) \tag{2}$$

where S_i is the rated MVA of i-th synchronous generator, H_i denotes the inertia constant (in s) of i-th synchronous generator and n is the total number of committed synchronous generators.

(v) Headroom: It is an amount of generation that can be provided by committed synchronous generators for arresting frequency decline after a disturbance. Governor control function is used for this purpose. Headroom can be calculated using

$$HR = \sum_{i=1}^{i=ng} (P_{max,i} - P_{gen,i}) \tag{3}$$

where $P_{max,i}$ refers to the rated capacity of i-th synchronous generator (in MW), $P_{gen,i}$ is the power generation of i-th synchronous generator (in MW) and ng denotes the total number of synchronous generators with governor control enabled.

(vi) Proportion of governor responsive generation: It is a ratio between governor responsive power generation capacity and total power generation capacity of a power system [12]. It is denoted by K_t and computed using

$$K_t = (\sum_{i=1}^{i=ng} P_{max,i}) / (\sum_{i=1}^{i=n} P_{max,i}) \tag{4}$$

III. Methodology

This section presents the proposed methodology and essential computation techniques regarding this paper.

A. Estimation of the Maximum Wind Power Penetration Level

To estimate the maximum wind power penetration level, dynamic simulation of a power network is performed in PSS®E simulation environment [13]. PSS®E software is widely used in transmission system planning to provide stochastic and dynamic modelling features. In PSS®E dynamic case, 100% constant current load is used for active power and 100% constant impedance load is used for reactive power modelling. Steps of the proposed methodology are presented below.

Step-1: For a power system, the largest synchronous generator is tripped to create a disturbance. The equivalent system frequency (f) is then calculated using (5) [12]

$$f = [\sum_{i=1}^{i=n} (S_i \times \omega_i)] / [\sum_{i=1}^{i=n} S_i] \tag{5}$$

where S_i is the MVA rating of i-th synchronous generator, ω_i is the speed of i-th generator (in p.u.) and n is the total number of synchronous generators.

Step-2: Dynamic simulation is run for a duration of time (t) given by

$$t = t_{initial} : t_{final} \tag{6}$$

where $t_{initial}$ refers to the simulation starting time and t_{final} denotes the simulation ending time.

Frequency nadir due to a trip of the largest generator (f_{min}) is then computed using

$$f_{min} = min[f]_{t=t_{initial}}^{t=t_{final}} \tag{7}$$

Step-3: Rate of change of frequency after a trip of the largest unit is computed using (8) [14]

$$ROCOF = \frac{1}{2} \times \frac{\Delta P}{IR} f_0 \tag{8}$$

where ΔP refers to the power mismatch between generation and load (in MW), IR is the system inertia (in MWs) and f_0 denotes the nominal system frequency (in Hz).

Step-4: Wind power generation in a power network is increased incrementally. To facilitate additional wind power, first wind integration strategy (direct replacement of a synchronous generator with a wind machine) is followed. In each wind penetration level, f_{min} and $ROCOF$ are calculated using (5)-(8) after a trip of the largest synchronous generator.

Step-5: Wind penetration level (WPL) is increased until either f_{min} or $ROCOF$ reaches under frequency load shedding (UFLS) margin. That is, WPL is increased up to

$$f_{min} \le f_{LL} \text{ or } ROCOF \ge ROCOF_{max} \tag{9}$$

where f_{LL} is the lower limit of allowable system frequency band and $ROCOF_{max}$ is the upper limit of acceptable $ROCOF$ to avoid UFLS.

Step-6: A mathematical relationship between f_{min} and WPL or $ROCOF$ and WPL is modelled. It can be expressed as

$$f_{min} = g(WPL) \text{ if } f_{min} \text{ reaches its limit prior to } ROCOF \tag{10}$$

$$ROCOF = g(WPL) \text{ if } ROCOF \text{ reaches its limit prior to } f_{min} \tag{11}$$

where g (.) stands for a functional relationship. Unknown parameters of (10) and (11) are estimated using least square regression technique.

Step-7: Now, setting $f_{min} = f_{LL}$ in (10) or $ROCOF = ROCOF_{max}$ in (11), corresponding WPL is estimated. This WPL indicates the maximum wind power penetration level (WPL_{max}) in a power system.

Step-8: The above steps are repeated for the second wind power integration strategy (2/3-1/3 rule).

B. Computation of Frequency Sensitivity Index

Frequency sensitivity index (β) is determined by relating an imbalance between generation and load to the corresponding abrupt change in system frequency. To estimate β of a power system, following steps are executed.

Step-1: Different synchronous generators are tripped, one at a time, to create a single credible contingent event. After a contingency, dynamic simulation is performed in PSS®E software and frequency nadirs are recorded.

Step-2: Frequency nadir (FN) is modelled as a function of power outage due to a synchronous generator trip (P_o, in MW) using linear regression method. It can be expressed using (12) [11]

$$FN = (\alpha \times P_o) + f_c \tag{12}$$

where α is the slope (in Hz/MW) and f_c is the constant part of linear regression model.

Step-3: Finally, β (in MW/0.1 Hz) is computed by (13) [11].

$$\beta = \frac{1}{|\alpha|} \times 0.1 \tag{13}$$

IV. Case Study

A 250 bus practical power system is investigated in this paper. The studied power system is connected to its adjacent network through a DC interconnection, which facilitates both power import and export. The installed capacity of the system is around 3,000 MW including 310 MW of wind power installation. There are some wind farms under construction, which are expected to be connected to the grid by the next few years. Moreover, there are several suitable locations where further wind farms could be installed. Considering all prospective wind farms, it is observed that wind penetration level in the studied system might significantly increase.

Average demand of the system is around 1,500 MW. The maximum and minimum demand occurs during winter and summer season respectively. Six different load scenarios are studied in this paper. These scenarios are based on some typical load conditions depending on different seasons. The load scenarios are low and high loads of summer, winter and spring seasons.

V. Results and Discussion

This section presents the maximum wind power penetration level in the studied power system to provide a sufficient frequency response. Six different load scenarios are separately analyzed with two different wind power integration strategies. Simulation results and relevant discussion are presented below.

A. Maximum Wind Power Penetration Level

(i) First wind power integration strategy: To clarify the process of estimating WPL_{max}, summer high load scenario is considered as an example. Wind power penetration during summer high is increased incrementally using first strategy. In the studied system, power generation from the largest synchronous generator is 135 MW. For each *WPL*, 135 MW of synchronous generation is tripped. Corresponding f_{min} and *ROCOF* are computed using (5)-(8). As *WPL* increases, f_{min} starts to decrease and *ROCOF* starts to increase. It is observed that f_{min} reaches its UFLS margin prior to *ROCOF*. Hence, a mathematical relation between f_{min} and *WPL* is modelled. Fig.1 illustrates a regression model between f_{min} and *WPL* during summer high load scenario. Mathematical relationship between them is estimated by (14).

$$f_{min} = -0.036(WPL) + 49.43 \tag{14}$$

The lower limit of allowable frequency of the studied power system is assumed at 48 Hz. It is based on similar practices for small size power systems [15]. This lower limit of allowable frequency varies from system to system. Now setting $f_{min} = f_{LL} = 48$ Hz in (14), WPL_{max} during summer high load scenario is around 40%.

Similar procedures are followed for all other load scenarios. It is found that f_{min} reaches its limit prior to *ROCOF* for all load cases. Table I shows the maximum wind power penetration level at different load conditions.

(ii) Second wind power integration strategy: The maximum wind power penetration level is also estimated for six typical load scenarios using the second strategy. Like first integration strategy, f_{min} reaches its margin before *ROCOF* for all load cases. Table II shows the results and a comparison between the two strategies.

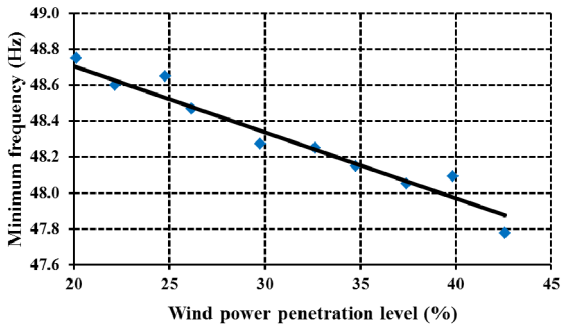


Fig. 1 Regression model between minimum frequency and wind power penetration level during summer high load scenario.

TABLE I
MAXIMUM WIND POWER PENETRATION LEVEL USING FIRST STRATEGY

Load scenario	Mathematical model between f_{min} and *WPL*	WPL_{max} (%)
Summer high	$f_{min} = -0.036(WPL) + 49.43$	40
Summer low	$f_{min} = -0.042(WPL) + 48.95$	23
Winter high	$f_{min} = -0.035(WPL) + 49.14$	33
Winter low	$f_{min} = -0.030(WPL) + 48.89$	30
Spring high	$f_{min} = -0.034(WPL) + 49.54$	45
Spring low	$f_{min} = -0.032(WPL) + 48.88$	28

TABLE II
MAXIMUM WIND POWER PENETRATION LEVEL USING SECOND STRATEGY AND COMPARISON TO THE FIRST STRATEGY

Load scenario	WPL_{max} (%)	
	Second strategy	First strategy
Summer high (SH)	43	40
Summer low (SL)	27	23
Winter high (WH)	36	33
Winter low (WL)	34	30
Spring high (SPH)	49	45
Spring low (SPL)	31	28

It is observed from Table II that, use of second wind integration strategy facilitates more wind power to be penetrated. In second strategy, 2/3 de-commitment and 1/3 re-dispatch approach is followed. For 1/3 re-dispatch, total system headroom may increase sometimes [12]. It can improve frequency nadir than that obtained using the first strategy. As a result, WPL_{max} is increased in all load scenarios when second wind integration strategy is implemented.

B. Computation of Correlation Coefficient

To detect a system parameter that could have the major impact on maximum wind power penetration level, correlation analysis is performed. Frequency sensitivity index of the investigated power system is estimated using (12) and (13). System inertia, headroom and proportion of governor responsive generation are calculated employing (2), (3) and (4) respectively. Table III shows the values at different load conditions.

Correlation coefficient between β and WPL_{max} is denoted by R_1 and expressed using (15).

$$R_1 = \frac{E(\beta \times WPL_{max}) - E(\beta) \times E(WPL_{max})}{\sigma_{\beta} \times \sigma_{WPL_{max}}} \tag{15}$$

where $E(.)$ refers to the expected value or mean value and $\sigma_{(.)}$ is the standard deviation. Similarly, the other correlation coefficients are computed using (16)-(18).

$$R_2 = \frac{E(IR \times WPL_{max}) - E(IR) \times E(WPL_{max})}{\sigma_{IR} \times \sigma_{WPL_{max}}} \tag{16}$$

TABLE III
FREQUENCY SENSITIVITY INDEX, INERTIA, HEADROOM AND PROPORTION OF GOVERNOR RESPONSIVE GENERATION

Load scenario	β (MW/ 0.1 Hz)	IR (MWs)	HR (MW)	K_t (%)
Summer high	14.30	9210	390	93
Summer low	6.00	5075	535	92
Winter high	12.00	9190	53	92
Winter low	9.60	6450	128	89
Spring high	16.70	9820	630	90
Spring low	9.25	6990	404	89

$$R_3 = \frac{E(HR \times WPL_{max}) - E(HR) \times E(WPL_{max})}{\sigma_{HR} \times \sigma_{WPL_{max}}} \quad (17)$$

$$R_4 = \frac{E(K_t \times WPL_{max}) - E(K_t) \times E(WPL_{max})}{\sigma_{K_t} \times \sigma_{WPL_{max}}} \quad (18)$$

where R_2 denotes the correlation coefficient between IR and WPL_{max}, R_3 is the correlation coefficient between HR and WPL_{max} and R_4 refers to the correlation coefficient between K_t and WPL_{max}.

Using the values from Table I, II and III, R_1 to R_4 are calculated for two wind integration strategies. Table IV presents the correlation coefficients.

To identify the most associated parameter with maximum wind power penetration level, the maximum value of correlation coefficient is computed by

$$R_{max} = max(R_1, R_2, R_3, R_4) \quad (19)$$

Using the values from Table IV, R_{max} is 0.99 and 0.98 in first and second wind power integration strategy respectively. The values correspond to the correlation coefficient between β and WPL_{max}. So, the frequency sensitivity index is mostly correlated to the maximum wind power penetration level. Fig.2 depicts WPL_{max} vs. β plot for both wind integration strategies. It is observed that for both cases, WPL_{max} exhibits a trend similar to that of β.

System inertia, headroom and proportion of governor responsive generation jointly contribute to the frequency sensitivity index. It is a composite term, which reflects the overall system characteristics. So, it seems rational to assume that frequency sensitivity index has the highest impact on maximum wind power penetration level.

TABLE IV
CORRELATION COEFFICIENTS

Correlation coefficient	First strategy	Second strategy
R_1	0.99	0.98
R_2	0.90	0.88
R_3	0.24	0.26
R_4	0.12	0.10

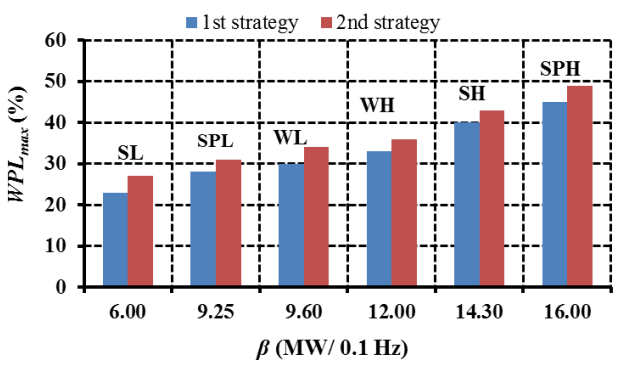


Fig. 2 Maximum wind power penetration level vs. frequency sensitivity index.

VI. CONCLUSIONS

This paper presents a methodology to estimate the maximum wind power penetration level to preserve a satisfactory frequency response in a given power system. If wind power penetration exceeds its maximum level, under frequency load shedding is likely to occur after a trip of the largest generating unit. As a consequence, service reliability of a power system might be affected. In this paper, a power system is studied in six different load scenarios with two wind power integration strategies. It is observed that amongst six load conditions, maximum wind power penetration level becomes the highest during spring high load condition. It is the lowest during summer low load scenario.

It reveals from this study that the key factors affecting the maximum wind power penetration can be system inertia, headroom, proportion of governor responsive generation and frequency sensitivity index. Amongst them, frequency sensitivity index may have the highest impact. Further works will be done to find a mathematical model, which will relate frequency sensitivity index with the maximum wind power penetration level.

ACKNOWLEDGMENT

The authors would like to thank the Australian Research Council and industry partners for their financial support through a Linkage Project.

REFERENCES

[1] "Global Wind Statistics-2013" [Online]. Available: http://www.gwec.net/wp-content/uploads/2014/02/GWEC-PRstats-2013_EN.pdf [Accessed: 15-May-2014].

[2] H. Ahmadi and H. Ghasemi, "Maximum Penetration Level of Wind Generation Considering Power System Security Limits", *IET Gener. Transm. Distrib.*, vol. 6, no. 11, pp. 1164–1170, Nov. 2012.

[3] A.A. Tamimi, A. Pahwa, S. Starrett, and N. Williams, "Maximizing Wind Penetration Using Voltage Stability Based Methods for Sizing and Locating New Wind Farms in Power System", *in Proc. 2010 IEEE Power and Energy Society General Meeting,* pp. 1–7.

[4] G. Coath and M. Dabbagh, "Effect of Steady State Wind Turbine Generator Models on Power Flow Convergence and Voltage Stability Limit", in *proc. 2005 Australasian Universities Power Engineering Conference.*

[5] N.T. Linh, "Voltage Stability Analysis of Grids Connected Wind Generators", *in proc. 4th IEEE Conference on Industrial Electronics and Applications, 2009*, pp. 2657–2660.

[6] S. Sreedharan, W. Ongsakul, J.G. Singh, I.M. Wartana and K. Buayai, "Development of PSO Based Control Algorithms for Maximizing Wind Energy Penetration", *in Proc. 2011 IEEE Power and Energy Society General Meeting*, pp. 1–6.

[7] A.A. Tamimi, A. Pahwa and S. Starrett, "Effective Wind Farm Sizing Method for Weak Power Systems Using Critical Modes of Voltage Instability", *IEEE Trans. Power Systems*, vol.27, no.3, pp. 1610–1617, Aug. 2012.

[8] J.K. Kaldellis, K.A. Kavadias and A.E. Filios, "A New Computational Algorithm for the Calculation of Maximum Wind Energy Penetration in Autonomous Electrical Generation Systems", *Applied Energy*, vol. 86, no. 7-8, pp. 1011–1023, Aug. 2009.

[9] H.Y. Yu, R.C. Bansal and Z.Y. Dong, "Fast Computation of the Maximum Wind Penetration Based on Frequency Response in Small Isolated Power Systems" *Applied Energy*, vol. 113, pp. 648–659, Jan. 2014.

[10] GE Energy, "Western Wind and Solar Integration Study: February 2008- February 2010," NREL report no. SR-550-47434, May 2010.

[11] Nahid-Al-Masood, R. Yan and T.K. Saha, "Frequency Response with Significant Wind Power Penetration: Case Study of a Realistic Power System", *accepted in 2014 IEEE Power and Energy Society General Meeting*.

[12] N.W. Miller, M. Shao, S. Venkataraman, C. Loutan and M. Rothleder, "Frequency Response of California and WECC under High Wind and Solar Conditions", *in Proc. 2012 IEEE Power and Energy Society General Meeting*, pp. 1–8.

[13] "Power System Simulator for Engineering" [On line]. Available: http://www.energy.siemens.com/hq/en/services/power-transmission-distribution/power-technologies-international/software-solutions [Accessed: 02-May-2014].

[14] H.R. Chamorro, M. Ghandhari and R. Eriksson, "Wind Power Impact on Power System Frequency Response", *in proc. 2013 North American Power Symposium*, pp. 1–6.

[15] "Wind Integration: International Experience, WP2: Review of Grid Codes", [Online]. Available: http://www.aemo.com.au [Accessed: 12-May-2014].

Voltage Stability Augmentation of IG Based Wind Farm Using Capacitor with Exciter Control System

D. Datta,[1,*] and M.R.I. Sheikh,[2]

[1,2]EEE Department, Rajshahi University of Engineering & Technology
Rajshahi-6204, Bangladesh
[*]dristidatta@gmail.com

Abstract—**Due to the recent expansion of renewable energy applications, wind energy conversion is receiving much interest all over the world. This is in no small part due to the rapid advances in the size of wind generators as well as the development of power electronics and their applicability in wind energy extraction, however, it has a great effect on the stability on power system. As wind turbine output is proportional to the cube of wind speed, the wind turbine generator output fluctuates due to wind speed variation. Thus this paper presents a new method to control the voltage stability by using reactive power compensation. It is observed that wind farm output voltage as well as grid voltage fluctuations can be controlled easily by applying this simple technique. Besides this, output voltage fluctuations can be smoothed by effectively controlling the reference voltage of the exciter of synchronous generator (SG). IEEE AC1A exciter is used in this study. Real wind speed data have been considered to obtained realistic responses. PSCAD/EMTDC software is used for simulation analyses.**

Index Terms—**Wind turbine; reactive power compensation; wind farm voltage; grid voltage; SG's exciter control.**

I. Introduction

Wind energy is the fastest growing and most widely utilized of the emerging renewable technologies in electrical systems [1]. The global energy challenge is to satisfy the growing demand to secure the energy sources and at the same time to face the threat equally urgent of climate change. In this regard renewable energies promise to be an important source for mankind. The rapid expenditure of fossil fuel has increased by leaps and bounds to improve our living standards. But it is a matter of great regret that, the world has limited amount of fossil fuel and nuclear power resources ([2], [3]).

One of the simplest methods of operating a wind turbine generation system (WTGS) is to use an induction generator (IG) connected directly to the power grid, because an induction generator is the most cost effective and robust machine for wind energy conversion. However, during start up, the induction generators need reactive power. As the reactive power drained by the induction generators is coupled to the active power generated by them, the variation of wind speed causes the variations of IGs real and reactive powers. These active and reactive power variations interact with the network and thus initiate voltage and power fluctuations.

Thus, reactive power compensation is a major issue; especially for fixed-speed WTGS for both steady state and transient conditions ([4] [5]). Usually, a fixed value capacitor bank is placed at wind farm terminal to provide the necessary reactive power demand at steady state. But it cannot maintain the wind generator terminal voltage constant under the randomly fluctuating wind speed. Moreover, IG needs a large reactive power during short circuit faults. In bulk power transmission systems, power electronics based controllers are frequently called Flexible AC Transmission Systems (FACTS). Presently, these devices are a viable alternative as they allow controlling voltages and currents of electric power systems. Voltage or current source inverter based FACTS devices such as static var compensator (SVC), static synchronous compensator (STATCOM), dynamic voltage restorer (DVR) etc. have been used for flexible power flow control, secure loading and damping of power system oscillation ([6], [7]). SVC is applied in [8] and [9] for reactive power compensation. Recently FACTS/ESS, i.e., FACTS with energy storage system (ESS), have emerged as more promising devices for power system applications [10]. But all the systems are very expensive. Some authors ([11], [12]) use new pitch control system to maintain output fluctuations, though this system can not control grid voltage with the fluctuating wind speed.

Therefore, this study investigate a new control system to maintain voltage fluctuation by correcting and controlling fixed capacitor bank as well as reference voltage of the exciter. Thus the system voltage performance improves effectively. Consequently, this study may be a good tool for power system application.

II. Model System For Simulation Analysis

The model system used in the simulation analysis is shown in Fig. 1. It consists of single wind generator (IG)

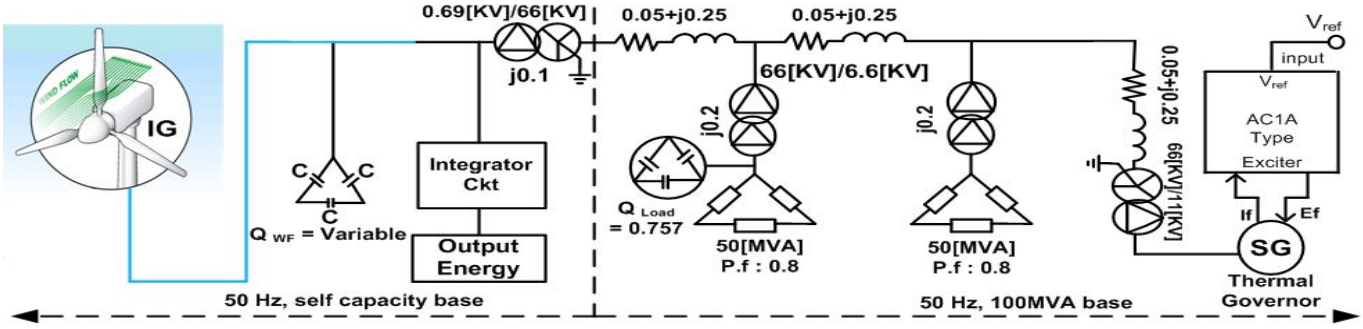


Fig. 1 Model circuit.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

and a thermal power generator (SG) with AC1A type exciter model. Here two loads have been considered and the total capacity of these loads is around 80MW. Core saturations of IG and SG are not considered for simplicity. The initial power flow and initial conditions are taken from [13]. In this model a capacitor bank has been considered and the values of the capacitor bank changes with respect to wind speed variations so that the p.f. becomes unity. The capacitor bank Q_{WF} performs the reactive power compensation in both steady state and dynamic wind variation. Similarly Q_{Load} is used at the terminal of load to compensate the voltage drop by the impedance of transmission lines.

III. Synchronous Generator Model

A. Governor [13]

The governor is such a device which automatically adjusts the rotational speed of the turbine and the generator output. The turbine is operated at a constant rotational speed, when the generator load remains constant. However, when the load changes, balance between the generator output and the load is not maintained and the rotational speed changes. When the load is removed, the governor detects the increase of the rotational speed and then the valve is closed rapidly so that an abnormal speed increase of the generators is prevented.

B. Governor for Thermal Generators [14, 15]

The governor model which is used in the simulation analysis is shown in Fig. 2 in which the values of 65M and 77M for thermal generator are shown in Table I. Where S_g is the revolution speed deviation [pu], 65M is the initial output, 77 M is the load limit (65M + rated MW output × PLM (%)), PLM is the spare governor operation (%) and Pm is the turbine output [pu].

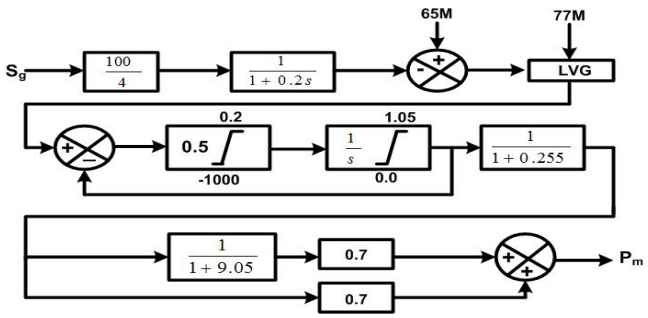


Fig. 2 Model of thermal governor.

TABLE I
Values of 65M and 77M for thermal generator

	IG: 20[MVA]
65M[pu]	0.653
77M[pu]	0.767
PLM[%]	15

IV. Pitch Controller

Conventional pitch controller as shown in Fig. 3 is used in the simulation analysis. The purpose of using the pitch controller is to maintain the output power of the wind generator at rated level by controlling the blade pitch angle of turbine blade when the wind speed is over the rated speed. Generally, the blade pitch operation system is complicated, but this paper simulates the pitch operation system by using a first order time delay system with time constant Tw = 5 sec. In addition, the pitch angle can not be changed instantly due to the rotational inertia of blade and mechanical limitations. Therefore, the rate of change of pitch angle is limited to 10 degrees per second in the simulations.

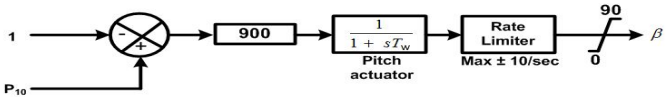


Fig. 3 Conventional pitch control system.

V. Proposed Capacitor Bank Correction System And SG's Exciter Control Technique

Generally, the value of capacitor bank for conventional model is kept constant. So when wind speed is larger than the rated value then output voltage becomes less than unity. On the other hand, when wind speed is less than rated value then output voltage fluctuates over the unity [16, 17]. This is the main problem for conventional model as it can not make the output voltage smooth. For this reason, a new technique has been proposed to control the output voltage fluctuations by making the capacitor value variable and the adjustment for different values of capacitor with respect to wind speed is shown in Table II. For the further smoothen of grid voltage, reference voltage of synchronous generator is also controlled to make the grid voltage approximately unity.

TABLE II
Capacitor Bank Correction

Wind Speed, V_w (m/s)	Capacitor Value, C (F) as shown in Fig. 1
Below 6	6600
7	7000
8	7600
9	9000
10	12000
11	16500
12	22200
13	24000
Above 13	26000

For this a decision circuit is used as shown in Fig. 4, where input voltage, Va is sensed by the control circuit and corresponding reference voltage (as shown in Fig. 1) is generated according to the Table III.

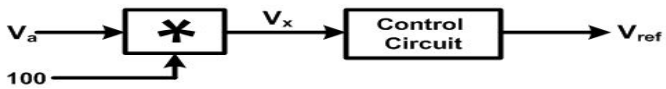


Fig. 4 Decision circuit.

TABLE III
Reference Voltage Correction

Input Voltage, V_x(Volt)	Reference Voltage, V_{ref} (Volt)
Below 99.4	1.08
99.5 to 99.7	1.07
99.8 to 99.9	1.06
100	1.05
100.1 to 100.2	1.04
Above 100.3	1.03

VI. Simulation Analyses

In this paper, simulation analyses have been carried out for three different cases as shown in Table IV. In order to investigate the performance of 3 cases with high wind power penetration real wind speed data have been considered which are measured in Hokkaido Island, Japan shown in Fig. 5. The simulation analyses have been performed by using PSCAD/EMTDC [18].

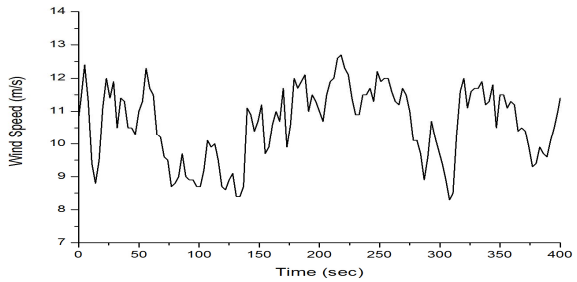


Fig. 5 Wind speed data.

TABLE IV
SIMULATION PATTERN

No. of Cases	Comparison Pattern
Case 1	Conventional model with capacitor bank control only
Case 2	Conventional model with reference voltage control only
Case 3	Conventional model with capacitor bank and reference voltage control

Simulation results for case 1 are shown from Fig. 6 to Fig. 9. Figure 6 shows the grid voltage, where the fluctuation is comparatively less when capacitor values are controlled with respect to wind speed variation instead of using fixed value of capacitor. From Fig. 7 and Fig. 8 it is seen that less reactive power compensation for both IG and SG is required than conventional model when capacitor bank is controlled. Figure 9 represents that the different values of capacitor changes proportionally with wind speed variation.

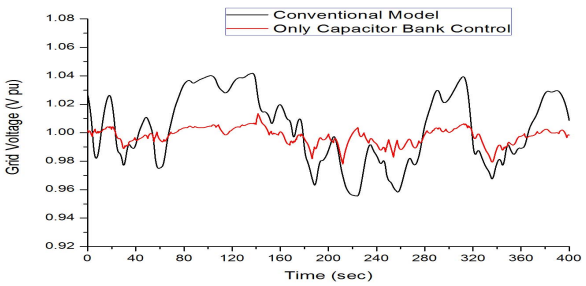


Fig. 6 Response of grid voltage [case 1].

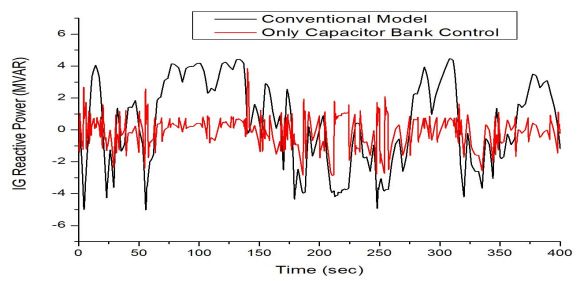


Fig. 7 Response of reactive power [case 1].

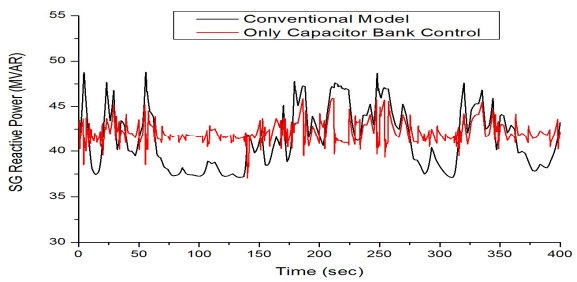


Fig. 8 Response of SG reactive power [case 1].

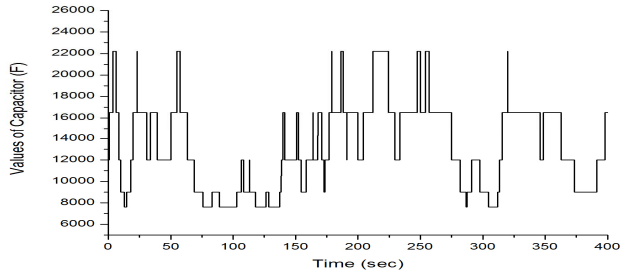


Fig. 9 Response of capacitor bank [case 1].

For case 2, only reference voltage is controlled with respect to grid voltage and fluctuation of grid voltage is comparatively less when reference voltage is controlled as shown in Fig. 10. For this case also less reactive power compensation is needed as shown in Fig. 11 and 12. The change of reference voltage with respect to grid voltage is shown in Fig. 13.

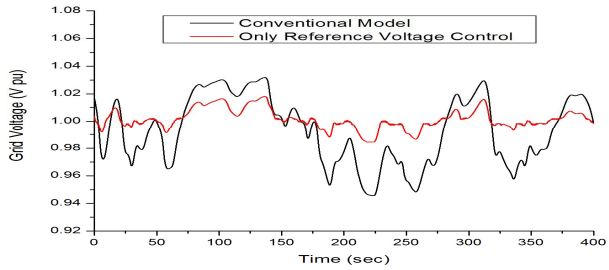


Fig. 10 Response of grid voltage [case 2].

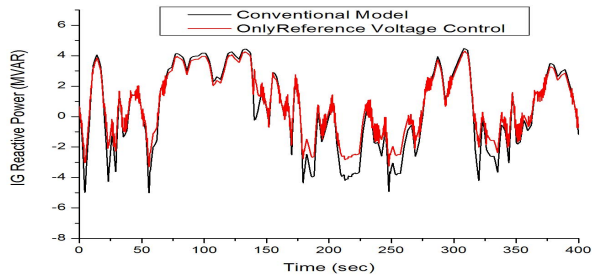


Fig. 11 Response of reactive power [case 2].

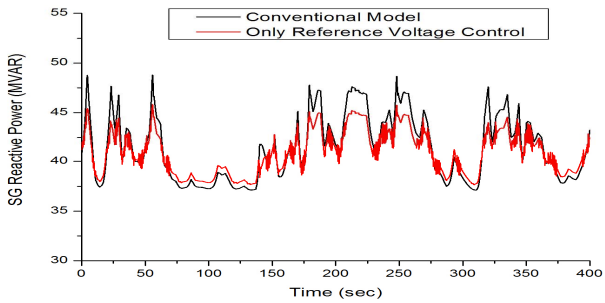


Fig. 12 Response of SG reactive power [case 2].

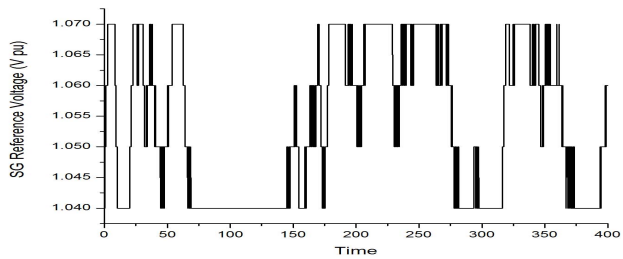


Fig. 13 Response of SG's reference voltage [case 2].

Both the values of capacitor bank along with reference voltage is controlled for case 3 and the grid voltage is almost near about unity which is shown in Fig. 14. Similarly, in this case less reactive power compensation is also required for IG and SG that is seen from Fig. 15 and 16 respectively.

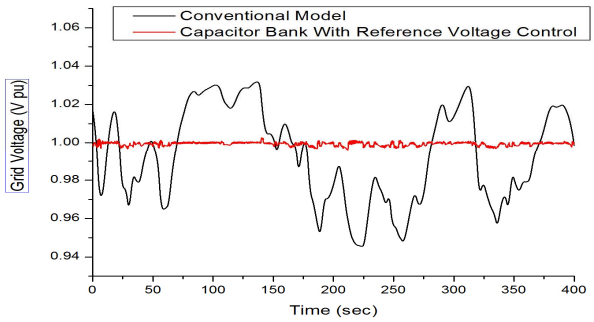


Fig. 14 Response of grid voltage [case 3].

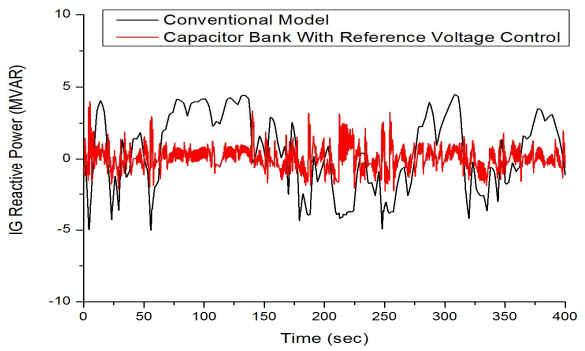


Fig. 15 Response of IG reactive power [case 3].

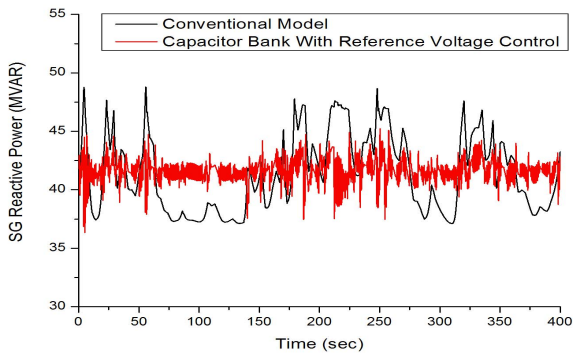


Fig. 16 Response of SG reactive power [case 3].

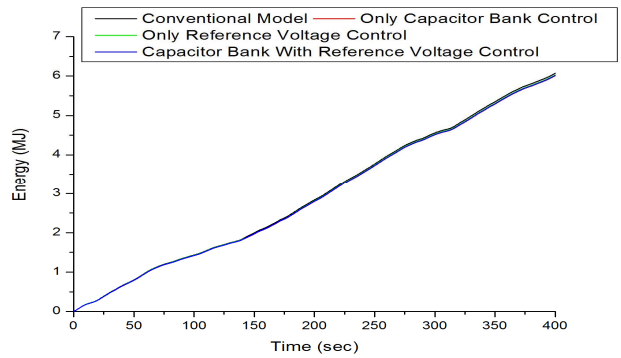


Fig. 17 Response of output energy [case 1, case 2, case 3].

An excellent result is observed from Fig. 17, which represents the comparison of output energy for 3 cases with conventional model and it is noticed that output energy is almost same to the conventional model. Hence there is no loss of energy.

VII. Conclusions

As grid voltage varies with the variation of wind speed, thus this study focuses on a new method to maintain the output voltage near about unity when wind speed remains in the acceptable level. It is confirmed from the above simulation results that by controlling capacitor bank and reference voltage along with conventional model, output grid voltage becomes smoother without any loss of energy, considering high wind penetration. Thus energy storage device can be neglected. As a result system may be more economic.

References

[1] T. Ackermann, Wind Power in Power Systems, John Wiley and Sons, Ltd., 2005.

[2] E. Haque and M.R.I Sheikh, “Grid frequency analysis with the issue of high wind power penetration”, *International Conference on Electrical Information and Communication Technology,* KUET, Khulna, Bangladesh, Dec. 2013.

[3] M.R.I. Sheikh, M.S.R. Ashrafi, and D. Datta, “Co-ordinated Frequency Control of Power System with High Wind Power Penetration”, *International conference on Electrical Engineering and Information & Communication Technology,* MIST, Dhaka, Bangladesh,12-14 Apr. 2014.

[4] M. R. I. Sheikh “Stabilization of a Grid Connected Wind Farm by Using SMES” a Ph.D thesis, Sep. 2010.

[5] S. M Muyeen, J. Tamura, and T. Murata, “Stability Augmentation of a Grig-connected Wind Farm," *Springer- Verlag, London,* ISBN 978-1-84800-315-6, Oct. 2008.

[6] Dr. L. Gyugyi "Unified Power-Flow Control Concept for Flexible AC Transmission Systems", *lEE Proceeding-C,* vo1.139, no.4, pp,323-33I, 1992.

[7] H.F.Wang, F.Li, and RG.Cameron, "Facts Control Design Nonparametric Models", *lEE Proceeding-Generation, Transmission and Distribution* vol. 146, no. 5, pp. 409-415, 1999.

[8] L.Cong, D.J.Hill, "Transient Stability and Voltage Regulation Enhancement via Coordinated Control of Generator Excitation and SVC", *Electrical Power and Energy system,* vol. 27, pp. 121-130, 2005.

[9] T. Ahmed, O. Noro, E. Hiraki, and M. Nakaoka, "Terminal Voltage Regulation Characteristics by Static Var Compensator for a three-phase Self-Excited Induction Generator”,*IEEE Transactions On Industry Applications,* vol. 40, no. 4, pp. 978-988, 2004.

[10] L. Zhang, C. Shen, M. L. Crow, L. Dong, S. Pekarek, and S. Atcitty, “Performance Indices for the Dynamic Performance of FACTS and FACTS with Energy Storage”, *Electrical Power Component and System* , vol. 33, no. 3, pp. 299-314, Mar. 2005.

[11] T. Yamazaki, R. T. Murata, T. Tamura, Fukushima, E. Ssano, K. Shinya, and T. Matstumoto, ‘‘Smoothing control of wind generator output fluctuations by new pitch controller’’, *IEEJ Transactions on Power and Energy*, vol. 129, no. 7, pp. 880-888,Jul. 2009.

[12] H. Umeki, R.Takahashi, T. Murata, and J. Tamura, ‘‘Smoothing control of wind generator output fluctuations by new pitch controller’’, *The Int. Conf. of Electrical Machines and Systems(ICEMS),*LS3C-1,2006.

[13] M.R.I. Sheikh, M.M. Haque, and A. Hossain, ‘‘Performance of governor system on minimizing frequency fluctuations with wind power generation’’, *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 2, no. 1, pp. 46--56, Feb. 2012.

[14] M.R.I. Sheikh, R. Takahashi, and J. Tamura, ‘‘Study on frequency fluctuations in power system with a large penetration of wind power generation’’, *International Energy Journal* , vol. 1, no. 1, Mar. 2011.

[15] Technical Reports, ‘‘Standard models of electrical power system,’’ *IEE of Japan*, vol. 754, pp. 40-43.

[16] C Luo, and B. T Ooi, ‘‘Frequency deviation of thermal power plants due to wind farms’’, *IEEE Transaction on Energy Conversion,*’’ vol. 21, no. 3, pp. 708-716, Sep. 2006.

[17] C. Carrillo, A.E. Feijoo, J. Cidras, and J. Gonzalez, ‘‘Power fluctuation in an isolated wind plant’’, *IEEJ Transactions on Power and Energy,* vol. 19, no. 1, pp. 217-221, Mar. 2004.

[18] Manitoba HVDC Research Center. 1994. PSCAD/EMTDC Manual.

A Competitive Framework to Regulate Day-Ahead Wind Power Predictions for Operational Planning in Electricity Markets

Lazaros Exizidis[1,*], Zacharie De Greve[1], Francois Vallee[2] and Jacques Lobry[1]

[1]Department of Electrical Engineering, Faculty of Engineering, University of Mons
Boulevard Dolez 31, 7000 Mons, Belgium

[2]Department of General Physics, Faculty of Engineering, University of Mons
Rue de Houdain 9, 7000 Mons, Belgium

[*]lazaros.exizidis@umons.ac.be

Abstract—**Wind energy, being an efficient and economically attractive Renewable Energy Source, attracts more and more investments as the electrical grids evolve into smarter grids. However, the stochastic nature of wind raises a lot of challenges for the system actors and planners. The different system actors (e.g. system operators, producers, consumers…), which are usually characterized by contradicting interests, use their own prediction models for taking into account the uncertainty of wind. The need for a framework that defines their interactions regarding prediction models is therefore mandatory, in order to allow a safe operation of the grid while ensuring a competitive framework for each of the stakeholders. In the present paper, a method for producing uncertainty forecasts is first presented. The uncertainty forecasts along with price scenarios for the day-ahead are used to investigate the potential conflicts of interests, with respect to the power forecast, between the major system actors (*i.e.* producers and the transmission/distribution system operator). The findings of this work indicate that the producer can maximize his income by inserting a bias in his prediction models which is in conflict with the System Operator's security constraints. In order to face this issue with a compromised solution, a competitive framework for the use of wind power predictions is suggested, according to which the interests of both actors will be satisfied.**

Index Terms—**Wind Power Forecasting, Scenarios, System Actors, Power Trading, Interactions Among System Actors.**

I. Introduction

Political decisions and the need for the use of more green energy sources, have led to the transformation of the classical distribution grids into smarter grids increasing integration of Renewable Energy Sources (RES), such as Wind Energy. However, the stochastic nature of wind, along with the lack of sufficient storage solutions, bring up the need for forecasting wind power in order to plan the future operation of the electric grid. Operational planning aims in taking preventive decisions to avoid, in day-ahead, critical situations such as over and under-voltages, congestions, etc. which are caused by the uncertain nature of the RES as well as the consumption patterns. The impact of the RES is further strengthened by the recent liberalization of the electricity market: when participating to the day-ahead market, wind power producers can only bid an expected value of power quantity, which will certainly deviate from the final delivered one. Therefore, a wind power forecast is needed in order to assist producers in estimating the quantity of power they are able to deliver.

Although wind power forecasting has been subject to an extensive research effort for the past years, there is no method that can provide a perfect forecast. The literature is very abundant on the topic, and concerns prediction horizons which range from very short-term [1]-[5] to day-ahead (DA) and long-term predictions [6]-[8]. Some of the most performant approaches regarding the prediction error make use of time series models [4], [7] and artificial neural networks [8]-[11].

As indicated before, a forecasting model can never provide a perfect forecast. Therefore, a wind power forecast is not that informative and helpful for operations regarding decision making. To cover this need an uncertainty forecast that provides information about the dispersion of observations around a forecasted value, should be achieved. However different system actors, are usually characterized by contradicting interests and use their own prediction models to cope with the uncertainty of wind. The need for a framework that defines their interactions regarding prediction models is mandatory, in order to allow a safe operation of the grid while ensuring a competitive framework for each of the stakeholders.

At subsection II.A of the current work we present a deterministic approach for wind power prediction based on AutoRegressive Moving Average models (ARMA) for the day-ahead. In order to use this point forecast for decision making, a method of constructing uncertainty forecasts consisting of wind speed scenarios is described at II.B. Wind speed scenarios are then transformed at II.C into wind power scenarios [12]. Finally, at section III, the wind power scenarios are employed along with the day-ahead price data to create realistic scenarios regarding the producer's participation to the energy market, the use of the wind power forecasts, and the conflicts of interests between the actors. The findings of this work indicate that the producer can maximize his income by inserting a bias in his prediction models which is in conflict with the System Operator's security constraints. A competitive framework is finally presented, which regulates the use of the forecast and proposes a compromised solution that aims to protect the interests of both actors.

II. Wind Power Forecasting For the Day-Ahead

For the purpose of this study, data from a wind site in the municipality of Wanze, Belgium were used. A wind speed prediction model is first proposed, followed by a transformation in the power domain.

A. Wind Speed Forecasting based on ARMA models

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

The available historical dataset consists of wind speed data for May 2013 in hourly measurements. The prediction horizon is one day-ahead, i.e. 01/06/2013. An ARMA model is described by (1), where Y_t is the investigated variable at time step t, φ_i and θ_i are the model's parameters, ε_t is white noise and p, q are the orders of the autoregressive and moving average part respectively:

$$Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \quad (1)$$

In order to fit a dataset in an ARMA model, it should respect the stationarity principle, thus we have to follow a standardization method as it is described in [11], which is based on extracting the average $\mu(t)$ of all data Y^*. The standardized value, Y_t, is then obtained by dividing the data with the standard deviation $\sigma(t)$, as given in (2),(3) and (4):

$$Y_t = \frac{Y_t^* - \mu(t)}{\sigma(t)} \quad (2)$$

$$\mu(t) = \frac{\sum_{i=0}^{d} Y^*_{m*i+t}}{d} \quad (3)$$

$$\sigma(t) = \sqrt{\frac{\sum_{i=0}^{d} (Y^*_{m*i+t} - \mu(t))^2}{d}} \quad (4)$$

where d is the number of considered days and m is the number of time intervals that comprise a day. Parameters p and q are calculated based on the Akaike Information Criterion (AIC) [13], which resulted in the set (p,q)=(1,3).

The Root Mean Square Error (RMSE) is compared in Fig. 1 with the RMSE of the persistence model. A forecast based on the persistence model is obtained by setting the forecasted value equal to the last measured one. This naïve predictor is usually used as a benchmark [7]. The selected ARMA model performs better than the persistence as seen in Fig. 1.

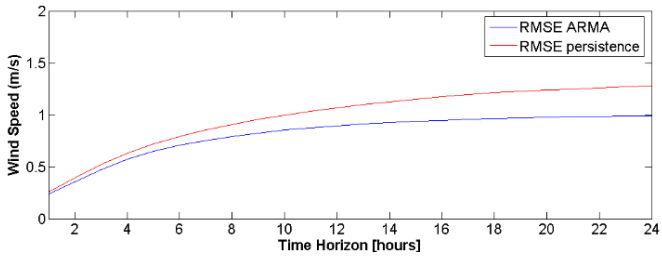


Figure 1: RMSE comparison between ARMA and Persistence models for the given dataset

B. Construction and Reduction of the Wind Speed Scenario Tree – Uncertainty Forecast

For the transformation of the point forecast into an uncertainty forecast, a method based on the simulation of the forecast error is used [14]. The absolute error of the prediction, i.e. the difference between observations and predicted values, is fitted into an ARMA(1,1) model as given in (5)-(6):

$$X_0 = \varepsilon_0 = 0 \quad (5)$$

$$X_t = \varphi_1 X_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1} \quad (6)$$

where X_t is the wind speed forecast error for the t-hour forecast and ε_t is a random Gaussian variable with standard deviation σ_z. The ARMA(1,1) parameters are defined by fitting the simulated RMSE to the available data, i.e. by minimizing (7):

$$Q(\varphi_1, \theta_1, \sigma_z) = \sum_{t=1}^{K} [RMSE_{measured}(t) - RMSE_{ARMA}(t)]^2 \quad (7)$$

$RMSE_{measured}$ is the real measured RMSE data and $RMSE_{ARMA}$ is the standard deviation of the forecast error $\sigma(X_t) = \sqrt{Var(t)}$, as given by (8)-(9):

$$Var(0) = 0, Var(1) = {\sigma_z}^2 \quad (8)$$

$$Var(t) = {\varphi_1}^2 Var(t-1) + \left(1 + {\theta_1}^2 + 2\varphi_1\theta_1\right) \sum_{i=1}^{t-1} {\varphi_1}^{i-1} \quad (9)$$

Once the model parameters are defined, the error of the prediction is simulated using Monte Carlo sampling, and a fan of error scenarios is constructed. For the current study, 1000 error scenarios with equal probabilities were generated. Each of these scenarios is added to the wind speed forecast resulting in a final fan of wind speed scenarios.

This big scenario tree, despite being well informative, is difficult to handle by the operational planning tool, mainly for computational issues. A reduction of the scenarios should be achieved, retaining as much as possible the information of the original tree. A reduction algorithm is used for this purpose, which clusters scenarios based on their mathematical distance and then deletes the one with the lowest probability. The probability of the remaining scenario will be equal to the sum of probabilities of the two clustered ones [15]:

$$d(\xi^i, \xi^j) = \left(\sum_{k=0}^{n_T} (v_k^i - v_k^j)^2\right)^{1/2} \quad (10)$$

where ξ are the investigated scenarios, n_T is the prediction time horizon and v^i_k is the wind speed at time k for the *i-th* scenario.

C. Transforming Wind Speed into Wind Power

A tool that will assist system actors in energy trading, should provide a prediction of the produced power. The power produced by a wind turbine is calculated by differencing the kinetic energy, E, of the turbine's wings under constant acceleration (11)-(12). The "power coefficient", C_p, that sets the limit according to Betz's law was set equal to 0.4.

$$E = m \cdot a \cdot d = m \frac{v^2 - u^2}{2} \quad (11)$$

$$P = \frac{dE}{dt} = \frac{1}{2} pAv(v^2 - u^2) C_p \quad (12)$$

where m is the mass of the moving object (kg), a is the acceleration (m/s^2), d is the distance covered (m), v is the wind speed (m/s), u is the initial wind speed (m/s), p=1.23 kg/m^3 is the air density and $A=\pi r^2$ is the swept area (m^2).

A commercial wind turbine with rotor diameter 116m and maximum capacity 3MW was used for the calculations. In Fig. 2 one can see the wind power scenarios after the reduction, followed by their probability of occurrence.

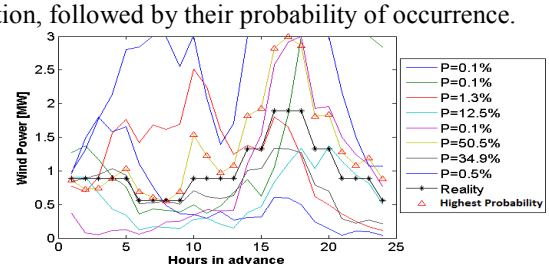


Figure 2: Reduced Wind Power Scenario Tree

III. Price Scenarios for the Day-ahead

Power trading at the DA market requires from the producer an evaluation of the real-time (RT) and day-ahead (DA) prices in order to optimize his market bid. At the current study Monte Carlo Sampling was used for the evaluation of the price scenarios.

A. Monte Carlo Sampling of Day-ahead and Real-time Prices

Price scenarios were constructed based on realistic hourly price data for March 2014, which are available online by the

New England Independent System Operator. Similarly subsection II.A, a standardization method is followed calculating the hourly average and the standard deviation. The desired scenarios are obtained according to (13)-(14), based on Monte-Carlo Simulations [16]:

$$Price_Scenarios_{DA}(j) = \varepsilon_{DA}(j) * \sigma_{DA}(t) + \mu_{DA}(t) \quad (13)$$

$$Price_Scenarios_{RT}(j) = \left(rho * \varepsilon_{DA}(j) + \sqrt{1 - rho^2}\right) * \varepsilon_{RT}(j) * \sigma_{RT}(t) + \mu_{RT}(t) \quad (14)$$

j=1,…,1000, rho is the correlation coefficient between RT and DA prices and ε are random samples in the range of [0,1]. The initial 1000 scenarios are reduced to 8, according to the same algorithm presented at subsection II.B.

B. *Conflicts of Interest at the Day-ahead Market with respect to Wind Power Predictions*

The producer's main interest in participating to the DA market is maximizing his income. Based on the tools presented we describe now how a producer can use an optimization algorithm to maximize his income, leading in many cases to high deviations between the bid and the delivered power, causing security and cost issues to the system operator (SO). On the other hand, the SO's main interest is to retain the stability and safety of the network and keep operational costs low. The procedure that is described is indicating the conflict of interests between the two major system actors with respect to the use of wind power predictions. A compromised scenario in the benefit of both actors is presented thereafter.

1) *Producer's profit maximization and conflicts of interest* [16]: The producer's profit at hour h can be calculated by the profit equation as given by (15):

$$profit_h = p_{DA,h} q_{del,h} + \left(p_{RT,h} - p_{DA,h}\right)\left(q_{del,h} - q_{bid,h}\right) \quad (15)$$

where $p_{DA,h}$, $p_{RT,h}$ are the day-ahead and real-time prices, $q_{del,h}$ is the delivered power and $q_{bid,h}$ is the bid placed by the producer for hour h of the DA. The first part of (15) describes the producer's profit if he settles all the produced power to the DA price whereas the latter part is the deviation settlement, which can be either negative (the producer pays the deviation) or positive (the producer is paid).

Considering this formula one can notice that the decision of the optimal bid value is based mostly on the difference between the RT and DA prices and not on the deviation between the delivered power and the bid. Decisions made by the producer in this direction can cause serious problems to the operation of the network as the deviation of power can be high and the SO will have to cover the missing power in real-time, and face at the same time stability issues. Therefore a deviation penalty is applied to the power producer, as given by (16):

$$profit_h = p_{DA,h} q_{del,h} + \left(p_{RT,h} - p_{DA,h}\right)\left(q_{del,h} - q_{bid,h}\right) - pen_{DA}\left|q_{del,h} - q_{bid,h}\right| \quad (16)$$

The maximization of the producer's expected profit, as stated in (17), is performed using a non-linear optimization algorithm defining the producer's point of view:

$$Q_h = \max_{q_{bid}} E[profit_h(q_{bid})] \quad (17)$$

The profit is maximized with respect to the placed bid, considering all scenarios for wind power production and RT/DA prices, along with their probability of occurrence. In Fig. 3, a bar plot with the suggested bid for each hour of the DA and for the two different cases (penalty/no penalty) is presented according to (17), along with the delivered power. As it is seen, if there is no penalty applied, the producer benefits from bidding maximum capacity (3MW) at hours 1AM, 3AM, 9AM, 10AM, 11AM and minimum at the rest hours. This is a strategy that obviously doesn't take into consideration the resulting deviation between bid and delivered power. In order to cope with this issue a penalty of \$5/MWh is applied to the deviation and the optimal bid from the producer's point of view is calculated. One notices that in this case (green bars the producer changes his strategy in most cases and bids based on the prediction, thus deviating less. However there are still cases where the producer benefits from bidding maximum capacity such as at hours 1PM, 10PM, 2AM, 4AM, 5AM. Those are the cases where the difference between RT and DA price is higher than the deviation penalty.

2) *Compromised solution for the best use of wind power predictions*: In order to face this conflict of interests that leads the producer to deliberately bid more aiming in higher profits, we suggest a framework based on an independent agent that will be responsible for the predictions according to which participating producers will place their bids. More particularly, the independent agent will be responsible for delivering wind power predictions for the day-ahead to the participating wind power producers who, in turn, will have to place a bid based on those predictions. A solution like this is obviously on the benefit of the SO as it will prevent from high deviations of power but is not that appealing for the producer since it reduces his profit. Therefore economic motivation should be provided to the producer to compensate his loss and motivate him for to participate. At the present work two different frameworks are suggested according to which producers are encouraged to participate to this service and both system actors are benefitted from their participation.

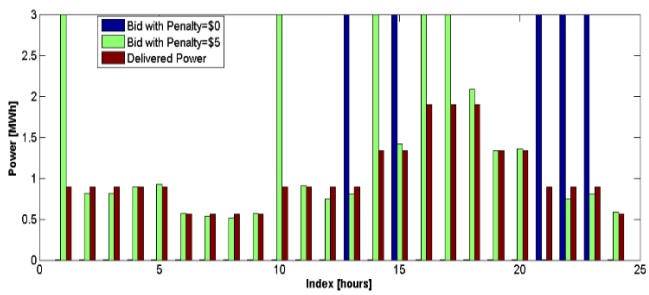


Figure 3: Bids based on the optimization of the producer's penalty

<u>Suggested Framework 1: Calculation of penalty costs based on the daily deviation</u>

Applying the penalty on the deviation of a longer time horizon (e.g. daily instead of hourly), offers the producer the possibility to counterbalance positive and negative hourly deviations, and thus get penalized on his daily deviation, which will be lower. Thus an accumulated daily penalty is applied under this framework instead of an hourly one. For the current case study we set the penalty at \$5/MWh as before, but instead of applying it on the hourly deviation we apply it on the daily one. The daily deviation is calculated by summing up positive and negative hourly deviations. For producers that do not bid under this framework (non-participating producers) the daily cost due to penalties will be \$50 for June 1st. In the contrary, a participating producer whose penalty is applied on his daily deviation, receives a daily penalty of \$37.8, which means \$12.2 less penalties at the end of the day. The same analysis for three more days, the

29th, 30th and 31st of May 2013, is presented in Fig. 4 (framework 1).

Suggested Framework 2: Establishment of a variable penalising factor based on the deviation at hour *h-1*

A variable penalty based on historical deviations is suggested as given at table 1, where *pen* is the penalty for non-participating producers. Considering the participation of the producer at this service we can calculate the changes in his decision making process, presenting the results for the current case study (June 1st, 2013). For a fixed penalty of \$5/MWh, the producer's bid results in high deviation at some hours, as described at III.B.1, gaining at the end of the day \$1007. On the other hand, for a variable profit as described at table 1, the producer follows the bidding strategy suggested by the independent agent which is based on the wind power prediction and his profit under this framework becomes \$1061. Therefore the producer will be recompensed, for compromising and deviating less, with \$54 higher profit at the end of the day. The result is presented also for the 29th, 30th and 31st of May 2013, as before, in Fig. 4 (framework 2).

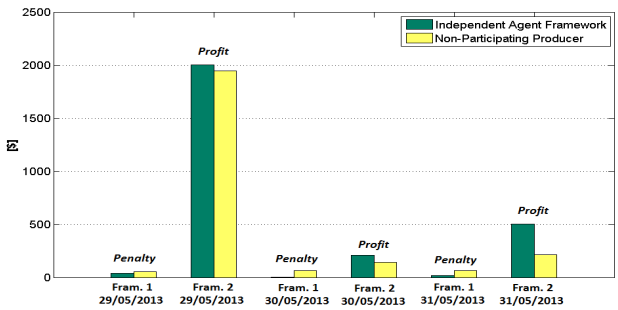


Figure 4: Producer's benefit for participating in the Independent Agent Service

Table 1: Adoption of the penalty according to the deviation history

Delivered Power and Bid deviation at hour *h-1*	**Corresponding penalty at hour *h***
$Dev \leq 20\%$	$50\% * pen$
$20\% < Dev \leq 30\%$	$80\% * pen$
$Dev > 30\%$	$100\% * pen$

IV. Conclusions

In this paper we have presented the construction of a complete tool that can assist the two major actors of an electricity network, DSO/TSO and wind power producers, in their decision making procedure regarding the use of wind power predictions. Wind power and DA/RT price scenarios were introduced in an optimization tool in order to maximize wind power producer's profit in the DA market.

This process has shown that the producer and the SO have conflicting interests and, in particular, even in the case that a deviation penalty is applied, the producer can maximize his income by deliberately bidding more energy than he is able to deliver. Therefore, we suggest a framework under which the predictions are given by an independent agent. In order to motivate the producer to participate to this regulated framework, a compromised solution can be reached by applying a variable deviation penalty and/or a penalty that is applied on a longer time horizon, increasing the producer's expected income motivating him to participate. Moreover this framework results in minimizing the power deviation, which is crucial for the stability of the network and thus for the SO.

Monte Carlo Sampling for the estimation of price scenarios will be replaced in the future by a multivariate ARMA-based prediction which will contribute to a more accurate input to the profit optimization algorithm. Furthermore, a combination of the two suggested frameworks while optimizing at the same time the penalty value with respect to the deviation history, will be further examined in order to approach an optimal framework considering also the economic cost and benefits from the producer's point of view.

Acknowledgment

This research is supported by the public service of Wallonia – Department of Energy and Sustainable Building within the framework of the GREDOR project.

References

[1] Kwon Kim, Young-Jun Seo, Kyong-Seob Moon and Young-Mi Lee, "Real-time Wind Power Prediction System Based on Smart-Grid in Jeju, Korea", *Journal of International Council on Electrical Engineering* Vol. 2, No. 2 (2012); pp. 194-200

[2] C.W. Poter and M. Negnevitsky, "Very short-term wind forecasting for Tasmanian power generation", *IEEE Transactions on Power Systems*, vol. 21, no. 2 (2006); pp. 965-972.

[3] Yu Jiang, Zhe Song, Andrew Kusiak, "Very short term wind speed forecasting with Bayesian structural break model", *Renewable Energy* 50 (2013); pp.637-647.

[4] G. Giebel, R. Brownsword and G. Kariniotakis. "The state of the art in short term prediction of wind power: A Literature Overview", *The European Commission under the 5th (EC) RTD Framework Programme* (1998-2002); pp.1-36.

[5] Xiaoqian,Bind Dong, Le Xie, Latanya Sweeney, "Adaptive Gaussian Process for short term wind speed forecasting", *Proceedings of the 2010 conference on ECAI 2010*; pp 661-666.

[6] T.M.H. El-Fouly, E.F. El-Saadany and M.M.A. Salama, "One day ahead prediction of wind speed using annual trends", *IEEE Power Engineering Society*, 18-22 June (2006); pp. 7-12.

[7] R.G. Kavasseri and K. Seetharaman, "Day Ahead Wind Speed Forecasting using f-ARIMA models", *Renewable Energy*, Vol. 35, No. 5, May 2009; pp.1388-1393.

[8] T.G. Barbounis, J.B. Theocharis, M.C. Alexiadis and P.D. Dokopoulos, "Long-Term Wind Speed and Power Forecasting using Local Recurrent Neural Network Models", *IEEE Transactions on Energy Conversion*, Vol. 21, No. 1, March 2006.

[9] H.G. Beyer, T. Degner, J. Hausmann, M. Hoffmann and P. Rujan", Short Term Prediction of Wind Speed and Power Output of a Wind Turbine with Neural Networks", *Proceedings of the EWEC '94 in Thessaloniki*, 10-14 Oct.; pp. 349-352.

[10] J.O. Tande and L. Landberg, "A 10 sec. Forecast of Wind Turbine Output with Neural Networks", *Proceedings of the EWEC '93* in Travemunde, 8-12 March; pp. 774-777.

[11] J.L. Torres, A. Garcia, M. De Blas and A. De Francisco, "Forecast of hourly average wind speed with ARMA models in Navarre (Spain)", *Solar Energy* 79 (2005); pp. 65-77.

[12] L. Exizidis, Z. De Greve, F. Vallee and J. Lobry, "Construction and Reduction of Aggregated Scenario Trees for the Day-Ahead Prediction of Wind Speed", Young Researcher Symposium IEEE Benelux Chapter, April 24-25 (2014), Ghent, Belgium.

[13] K. P. Burnham and D. R. Anderson, *Model selection and multimodel inference: a practical information-theoretic approach*, Springer: New York, 2002.

[14] N.Growe-Kuska, N. Heitsch and Werner Romisch, "Scenario Reduction and Scenario Tree Construction for Power Management Problems", *Contribution to the IEEE Power Tech Conference*, Bologne 2003.

[15] J. Dupacova, N.Growe-Kuska and Werner Romisch,"Scenario reduction in stochastic programming: An approach using probability metrics", *Mathematical Programming*, Vol. 95, No. 3 (2003); pp. 493-511.

[16] A. Botterud, Z. Zhou, J. Wang, R.J. Bessa, H. Keko, J. Sumaili, V. Miranda, "Wind Power Trading Under Uncertainty in LMP Markets", *IEEE Transactions on Power Systems*, Vol.27, No.2, May 2012

Lightning Surge Response of Wind Turbine Grounding at Low Resistivity Soil

Md. Raju Ahmed,* and Utpal Kumar Das**

Department of Electrical & Electronic Engineering
Dhaka University of Engineering & Technology (DUET)
Gazipur, Bangladesh
*mrahmed@duet.ac.bd, **apu_duet@yahoo.com

Abstract— **Accidents caused by natural phenomena such as lightning are serious problems for the diffusions of wind power generations. To exploit high wind conditions, wind turbine generators are often constructed along the seashore where few tall structures exist; therefore, they are often struck by lightning. Most of the breakdowns and malfunctions of the electrical and control systems inside wind turbine due to lightning are caused by the voltage rise of the grounding system. At a disposal site like seashore where the resistivity of soil is very low, the voltage rise at grounding at the wave front typically becomes larger than that of the steady state because of the inductive behaviour of the grounding system. Voltage rise at grounding of wind turbine due to lightning at a disposal site are computed using electromagnetic analysis approach, and designing of proper grounding system at low resistivity soil is discussed. Finite Difference Time Domain (FDTD) method is used for electromagnetic analysis. Designing of grounding system of wind turbine at a disposal site is discussed.**

Index Terms—**Lightning, wind turbine grounding, voltage rise, FDTD method.**

I. Introduction

With the rapid growth of wind power generation, accidents associated with wind turbine generator systems have increased, including those caused by natural phenomena such as lightning and typhoons. Lightning, especially, causes extensive and serious damage [1–3]. According to statistics, annually 4-8% of wind turbines in northern Europe get damages by lightning; this percentage is 14 in southern Germany [4]. Most of the breakdowns and malfunctions of the electrical and control systems inside wind turbines are caused by a rise in ground potential due to lightning [5, 6].

Wind turbine grounding system has to be effectively designed in order to prevent excessive over-voltages and potential gradients that may cause damage to equipment or threaten human life. To design the wind turbine grounding properly, many researchers have been studied the transient and steady state behavior of wind turbine grounding [7-12]. However, studies on transient behavior of grounding at very low resistivity soil are very rear [13].

In this article, the lightning surge response of a wind turbine grounding at very low resistivity soil is analyzed using electromagnetic analysis approach. In case of low resistivity soil the transient voltage rise at the wavefront of the lightning impulse current is very large as compared to the steady state value due to the inductive behavior of the grounding. IEC 61400-24 [14] suggest use of additional grounding rod/wire to reduce the grounding resistance, in this study it is shown that though the additional grounding rod/wire reduces steady state grounding resistance, it cannot reduces the maximum voltage rise of the grounding if its length is longer than certain value which is depending on the resistivity of soil and front time of injection current. Designing of proper grounding system at low resistivity soil to reduce the maximum voltage rise at the current injection point is discussed.

II. Analysis Method

Virtual Surge Test Lab (VSTL) [15], developed by Central Research Institute of Electric Power Industry (CRIEPI), Japan is used for electromagnetic analysis. VSTL is an electromagnetic analysis code based on FDTD method [16]. Fig. 1 shows the model of grounding electrode of wind turbine. The modelled grounding electrode is similar to the grounding of a small size wind turbine. Though many nations are moving towards modern large size wind turbine, due to lack of manufacturing and construction facility many locations are favourable for small and medium size wind turbine. The analytical setup for FDTD calculation is shown in Fig. 2.

A current lead wire is used to inject the lightning impulse current and a voltage measuring wire is used to measure the voltage at the current injection point. The voltage measuring wire and current lead wire are set orthogonally to reduce electromagnetic interference between them.

The dimension of the analysis space is 100m×125m×100m, and is divided into 150×200×150 cells. Liao's second-order absorbing boundary condition [17] is adopted at each face of the working volume.

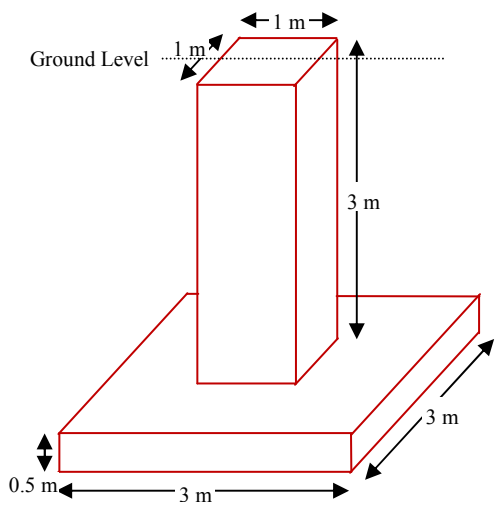


Fig. 1 Model grounding electrode of wind turbine. Steady-state grounding resistance is 30 Ω for ground resistivity of 500 Ωm.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

Non-uniform cell sizes are used to reduce the calculation time and memory required. The side lengths of a cell vary from 0.5m to 1.5m. The lower 40 m of the space represents the soil. The radius of all the wires including the electrodes of grounding system is 10 mm.

Soil is postulated homogeneous with relative permittivity of 10. Lightning current of 50 kA with 0.25/100µs waveform representing subsequent return-stroke current [18], as shown in Fig. 3, are injected into the grounding. The waveform of injected current is expressed by Heidler function.

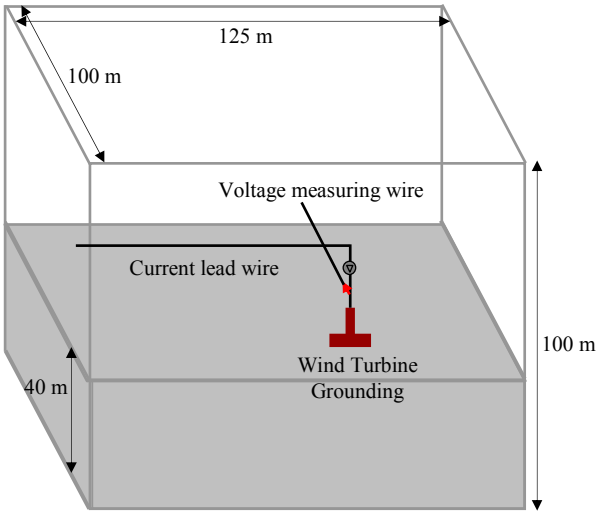


Fig. 2 Analytical setup for FDTD analysis.

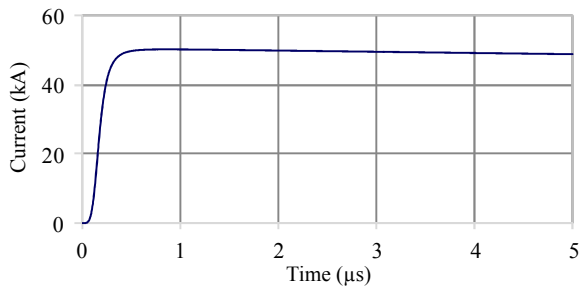


Fig. 3 Test current waveform, subsequent return-stroke current, rise time 0.25 µs.

III. Lightning Surge Response of Wind Turbine Grounding

The voltage rise at the current injection point of the modelled wind turbine grounding for injection of current waveform of Fig. 3 is shown in Fig. 4. When a wind turbine generator system is constructed at a place where resistivity of soil is very low, such as a coastal area, a significant potential rise occurs at wavefront due to the inductive behaviour of the grounding system. At high resistivity soil, the steady state voltage rise is very high, which is depending on the steady state grounding resistance. Steady state condition is generally emphasized during the design of any grounding system. Also the steady state response can be easily predicted using simple empirical formula. In this case the transient behaviour is not so prominent. However, at low resistivity soil the transient voltage rise is very high than the steady state voltage rise, also the transient characteristics cannot predicted without proper analysis. Therefore proper analysis of transient characteristics of grounding system at low resistivity soil is very important for their proper design.

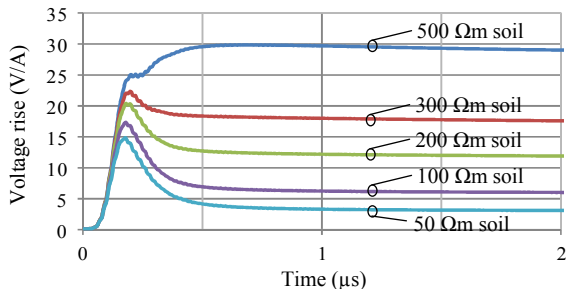


Fig. 4 Voltage rise at the current injection point of modelled wind turbine grounding for injection of subsequent return-stroke current, rise time 0.25 µs.

From Fig. 4, for 50Ωm soil the resistance of the wind turbine grounding is 3Ω, and the steady state voltage rise which is proportional to the grounding resistance is 3 volt per ampere of injection current. However, at wavefront the voltage rise is about 15 volt which is five times higher than steady state. This phenomenon is known as the inductive behaviour of grounding.

The inductive behaviour of grounding is dependent on the resistivity of soil and front time of the injection current, and increases with the decreases of the resistivity of soil and shorter the front time of injection current. As the transient voltage rise is higher at low resistivity soil than the steady state, therefore, transient characteristics should be consider during the design of grounding system at low resistivity soil.

IV. Design of Grounding System at Low Resistivity Soil

IEC- 61400-24 [14], recommend use of additional grounding rod or wire with the grounding system to reduce the maximum voltage rise at the current injection point. To investigate the influence of additional grounding rod/wire on the maximum voltage rise at the current injection point modification of grounding of Fig. 1 is considered as shown in Fig. 5.

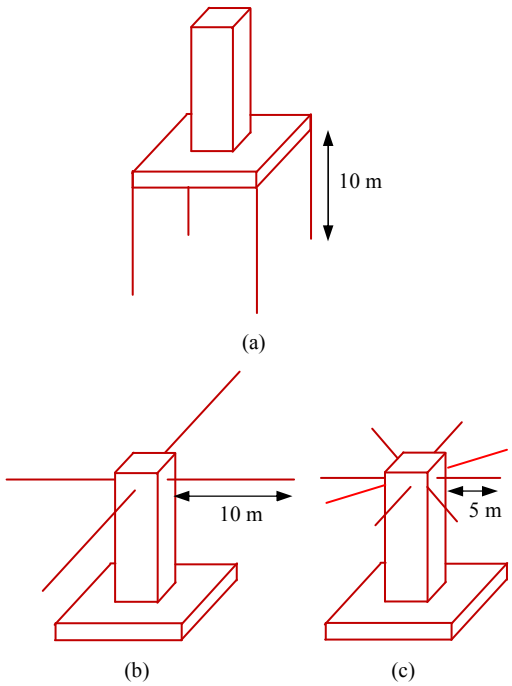


Fig. 5 Additional grounding rod/wire with the wind turbine grounding to reduce the maximum voltage rise, (a) 4 additional rods is at the bottom of the footing (b) 4 long grounding wires near the current injection point and (c) 8 short grounding wires near the current injection point.

Three modifications are considered for analysis, these are: Case a: 4 additional grounding rods is used at the bottom of footing; Case b: 4 additional long grounding wires is used near the current injection point; Case c: 8 additional short grounding wires are used near the current injection point.

The voltage rise at the current injection point for fundamental case (as Fig. 1) and case (a) of Fig. 5 varying the length of grounding rod is shown in Fig. 6. From Fig. 6, it is clear that if the length of the grounding rod at the bottom of the footing is longer than 3 m, the additional length have no effect on the maximum voltage rise at the current injection point, maximum voltage rise remain constant, though the effect of additional grounding rod is visible after the maximum voltage rise.

The voltage rise at the current injection point for fundamental case (as Fig. 1) and case (b) of Fig. 5 varying the length of grounding wire is shown in Fig. 7. From Fig. 7, it is clear that if the length of the grounding wire is longer than 10 m, the additional length have no effect on the maximum voltage rise. From previous section for case (a), it is found that if the length of grounding rod at the bottom of the footing of the grounding system is longer than 3m the maximum voltage rise at the current injection point remains the same. Therefore, the maximum length of the rod which effectively reduces the maximum voltage rise is 3m. However, if the grounding wire is used near the current injection point instead of the bottom of the footing of the grounding then the maximum length of the grounding wire which effectively reduce the maximum voltage rise at the current injection point is 10m.

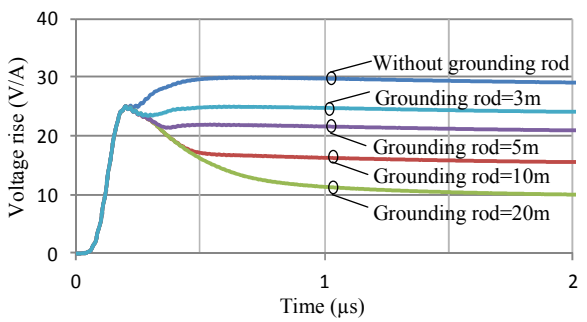


Fig. 6 Voltage rise at the current injection point for case (a) depending on the length of grounding rod, rise time of injection current is 0.25 µs. Resistivity of soil is 500Ωm.

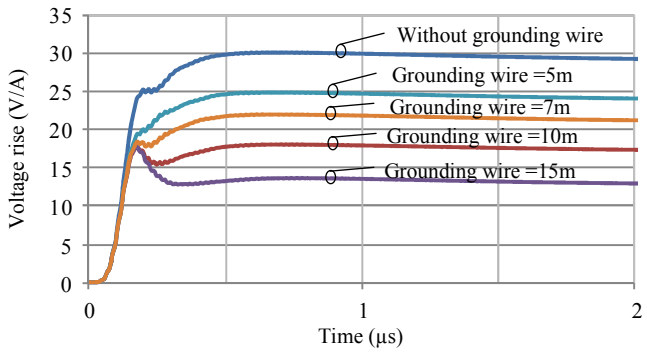


Fig. 7 Voltage rise at the current injection point for case (b) depending on the length of grounding rod, rise time of injection current is 0.25 µs. Resistivity of soil is 500Ωm.

The voltage rise at the current injection point for fundamental case (as Fig. 1) and case (c) of Fig. 5 varying the length of grounding wire is shown in Fig. 8. From Fig. 8, it is clear that if the length of the grounding rod is longer than 10 m, the additional length will have no effect on the maximum voltage rise.

The length of grounding wire effective to reduce the maximum voltage rise is same in this case as that of case (b), because of their same position. For case (b), if we use total 80m grounding wire, length of each grounding wire is 20m maximum voltage rise will be nearly 18 volt per ampere of injection current. For case (c) if we use total 80m grounding wire, length of each grounding wire is 10m maximum voltage rise will be reduced from 18 volt per ampere of injection current to nearly 15 volt per ampere of injection current. Therefore, at low resistivity soil, it is better to use several number of short grounding rod or wire instead of using few number of grounding rod or wire to reduce the maximum voltage rise at the current injection point. The additional grounding electrodes should be placed as close as the current injection point of the grounding system.

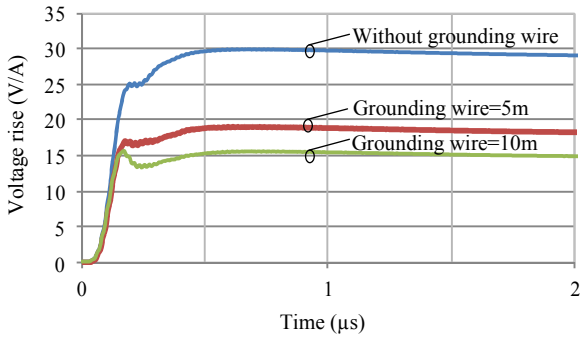


Fig. 8 Voltage rise at the current injection point for case (c) depending on the length of grounding rod, rise time of injection current is 0.25 µs. Resistivity of soil is 500Ωm.

V. Conclusions

This article has presented the results of analytical studies that investigated the grounding characteristics of a small or medium size wind turbine generation system at a low resistivity soil. Voltage rise at the grounding cause overvoltage of the power, communications, and control equipment placed in the tower and at the outermost insulating layer of any incoming cable, which could result in the breakdown or deterioration of the insulation. At a low resistivity soil, the grounding system shows inductive behaviour. Voltage rise at the wavefront becomes higher than the value calculated from the resistance of the grounding. This phenomenon is more significant at low resistivity soil.

IEC standard recommends using additional grounding rod or wire to reduce the grounding resistance, in this study, it is found that though the additional grounding rod/wire reduces steady state grounding resistance, it cannot reduces the maximum voltage rise of the grounding if its length is longer than certain value. Also the position of additional grounding wire/rod plays an important role to reduce the maximum voltage rise at the grounding. At low resistivity soil, it is better to use several number of short grounding rod or wire instead of using few number of long grounding rod or wire to reduce the maximum voltage rise at the current injection point. The additional grounding electrodes should be placed as close as the current injection point of the grounding system.

ACKNOWLEDGMENT

The author would thank CRIEPI, Japan for granting to use VSTL in the present study.

REFERENCES

[1] I. Cotton, B. Mcniff, T. Soerenson, W. Zischank, P. Christiansen, M. Hoppe-Kilpper, S. Ramakers, P. Pettersson and E. Muljadi, "Lightning protection for wind turbines," *in Proc. of the 25th International Conference on Lightning Protection*, Rhodes, Greece, 2000, pp. 848–853.

[2] *IEC. Wind Turbine Generator Systems—part 24: Lightning Protection.* TR 61400-24, 2002.

[3] *NEDO. Wind Turbine Failures and Troubles Investigating Committee Annual Report.* 2006; (in Japanese).

[4] J. Korsgaard, "Improved lightning protection system enhances the reliability of multi-MW blades" COWC presentation, Oct. 2005.

[5] K. Yamamoto, T. Noda, S. Yokoyama and A. Ametani, "An experimental study of lightning overvoltages in wind turbine generation systems using a reduced-size model," *Electrical Engineering in Japan, vol.* **158(4)**, pp. 22–30, 2007.

[6] K. Yamamoto, T. Noda, S. Yokoyama, A. Ametani, "Experimental and analytical studies of lightning overvoltages in wind turbine generator systems," *Electric Power Systems Research*, vol. **79(3)**, pp. 436–442, ISSN: 0378–7796, 2009

[7] Jenkins N and Vaudin A., "Earthing of wind farms," *Wind Engineering,* vol. **18(1)**, pp. 37–43, 1994.

[8] I. Cotton, "Windfarm earthing," *in Proc. of 11th International Symposium on High Voltage Engineering (ISH 99)*, London, UK, 1999, vol. 2, pp. 288–291.

[9] M. Lorentzou, N. Hatziargyriou and B. Papadias, "Analysis of wind turbine grounding systems," *in Proc. of 10th Mediterranean Electrotechnical Conference (MELECON 2000)*, 2000, pp. 936–939.

[10] Kontargyri VT, Gonos IF, Stathopulos IA," Frequency response of grounding system of wind turbine generators," *in Proc. of the 24th International Symposium on High Voltage Engineering (ISH2005)*, 2005.

[11] B. Hermoso, "Wind farm earthing installations: rated and lightning frequencies behavior," *in Proc. of International Conference on Grounding and earthing (GROUND'2006)*, 2006; pp. 411–414.

[12] O. Ukar, I. Zamora, R. Idiondo and A. Mugica' "Analysis for high frequencies of grounding systems, for wind turbines," *in Proc. of 27th International Conference on Lightning Protection*, Avignon, France, 2004, vol. II, pp. 1053–1057.

[13] K. Yamamoto, S. Yanagawa, S. Sekioka and S. Yokoyama, "Transient Grounding Characteristics of an Actual Wind Turbine Generator System at a Low-resistivity Site", *IEEJ Trans on Electrical and Electronic Engineering, vo. 5, pp.* 21–26, 2010.

[14] IEC TR 61400-24, Wind turbine generator systems, Part 24: Lightning Protection, 2002.

[15] T. Noda and S. Yokoyama, "Development of a general surge analysis program based on the FDTD method," *Trans. IEE of Japan*, vol. 121-B, no. 5, pp. 625-632, 2001 (in Japanese).

[16] K. S. Yee, "Numerical solution of initial boundary value problem involving Maxwell's equations in isotropic media," *IEEE Trans. on Antenna and Propagation,* vol. 14, no. 3, pp. 302-307, 1966.

[17] Z. P. Liao, H. L. Wong, B. P. Yang and Y. F. Yuan, "A transmitting boundary for transient wave analysis," *Science Sinica*, Vol. A27, No. 10, pp. 1063-1076, 1984.

[18] IEC 62305-1, *Protection against Lightning, Part 1: General Principle*, 2006.

Wide Speed Range Operation of PMSM Drives Using Nonlinear Flux Control Techniques

Garin M. Schoonhoven, *Student Member IEEE* and M. Nasir Uddin, *Senior Member IEEE*
Dept. of Electrical Engineering
Lakehead University
Thunder Bay, ON, Canada
E-mails: gmschoon@lakeheadu.ca, muddin@lakeheadu.ca

***Abstract*—This paper presents a nonlinear flux controller based wide speed operation of permanent magnet synchronous motor (PMSM) drive. An adaptive backstepping based nonlinear controller is developed to get the control laws while maintaining systems stability. For the nonlinear controller the flux is controlled based on maximum torque per ampere (MTPA) and flux weakening (FW) techniques for below and above the rates speed, respectively. The global asymptotic stability of the drive is demonstrated by Lyapunov stability criterion in conjuncture with Barbalat's lemma. The system nonlinearities are also accommodated through the online estimation of critical parameters. The proposed nonlinear flux controller based PMSM drive is investigated in simulation using MATLAB/Simulink software. The global stability and robustness of the driver are verified through simulation results.**

***Keywords*— Nonlinear control, IPMSM, adaptive backstepping, parametric estimation, flux control, and wide speed range operation.**

I. INTRODUCTION

Advances in the fields of semiconductor power electronics and magnetic materials have contributed to recent significant improvements in ac motor drives. In particular, permanent magnet synchronous machines (PMSM), have seen a rise in popularity; in part due to the increased availability of rare-earth alloys facilitating high density permanent magnets. Interior permanent magnet synchronous machines (IPMSM), a class of PMSM's, demonstrate excellent efficiency, nonlinear flux distribution, small effective air gap and high power density compared to other ac machines [1]. Further, reluctance torque can be realized as a function of the nonlinear flux distribution (i.e. non-saliency).

Despite advantageous features, control of IPMSM drives is a complex issue, resulting from nonlinear coupling of winding currents and magnetic saturation of the rotor core. Performance of the IPMSM drive varies with changing operating conditions due to magnetic saturation, affecting the d-q axis inductances. Thus to ensure robust machine performance, critical motor parameters need to be estimated adaptively. Significant research effort has been dedicated on various parameter estimation schemes [5]; and recently, adaptive backstepping has been gaining popularity in controller design for motor drives [6],[7]. These techniques allow for the accommodation of many machine nonlinearities and uncertainties.

Wide speed range operation of IPMS machines, principally above the rated speed remains a challenge [8]. When surpassing rated speed, phase voltage and current magnitudes can exceed the rated machine limits. Consequently, the magnetic saturation of the rotor core can cause the operating point of IPMS to shift, resulting in significant variation of d-q axis flux linkages [9]. Thus, performance of IPMSM drives may severely degrade above the rated speed. Performance drives in demanding applications require fast and accurate speed response and insensitivity to parametric variations. In particular, spindle drives, machine tools and traction drives require operation above rated speed (constant power operation). Motoring operation above the rated speed can be achieved by utilizing flux weakening (FW) techniques [9],[10], to ensure rated conditions are not exceeded. Furthermore, flux control below rated speed is necessary to exploit the reluctance torque, which is the principal advantage of IPMSM over other PM motors. In IPMSM drives, direct flux control is not possible; however, flux can be controlled indirectly through the demagnetizing effect of d-axis armature current. However, in the majority of reviewed papers, nonlinear control of IPMSM drives is established by first setting the d-axis current to zero [1]-[3]. This approach greatly simplifies mathematical design, however invariably limits the operating speed ranges and forfeits the contribution of reluctance torque due to the buried permanent magnets.

Consequently, this paper presents a nonlinear adaptive control technique for IPMSM drives, incorporating MTPA and FW based flux control algorithms to ensure wide speed range operation. The development of control laws is based on adaptive backstepping to provide robust system stability. Online estimation of key motor parameters ensures system uncertainties and nonlinearities are accommodated. The performance of the proposed nonlinear controller based IPMSM drive in Fig. 1 is tested in simulation at different operating conditions to demonstrate performance.

II. IPMSM MODEL

Through the use of the "*dq0*" transform, the machine model in stationary stator reference form is translated to the synchronously rotating d-q reference frame in as:

$$\dot{i}_d = \tfrac{1}{L_d} v_d + \tfrac{n_p L_q}{L_d} \omega_r i_q - \tfrac{R_s}{L_d} i_d \tag{1}$$

$$\dot{i}_q = \tfrac{1}{L_q} v_q - \tfrac{n_p L_d}{L_q} \omega_r i_d - \tfrac{n_p \phi_m}{L_q} \omega_r - \tfrac{R_s}{L_q} i_q \tag{2}$$

$$\dot{\omega}_r = \tfrac{1}{J}\left(\tau_e - \tau_m - B_m \omega_r\right) \tag{3}$$

$$\tau_e = \tfrac{3 n_p}{2}\left[(L_d - L_q) i_d i_q + \phi_m i_q\right] \tag{4}$$

where: v_d , v_q - *d-q* axis voltages; i_d , i_q - *d-q* axis currents; ω_r - angular velocity; τ_L - motor load torque; L_d , L_q - *d-q* axis inductances; B_m - viscous friction coefficient; J - moment of

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

inertia of motor and load; ϕ_m - rotor magnetic flux; R_s - stator resistance; " · " - derivative operation; n_p - motor pole pair.

III. Control Algorithm Design

This paper presents a high-performance adaptive speed controller that provides accurate motor control both above and below rated speed. Control is achieved by applying command control magnitudes v_d and v_q as per (15),(16). Most papers [1],[2] describing IPMSM control techniques set the *d*-axis current to zero, simplifying control, however restricting the motor performance characteristics. In this paper, the *d*-axis current is defined by flux control techniques above and below rated speed. Within rated speed range, i_d is defined by the maximum torque per ampere (MTPA) technique [2], and above rated speed i_d is defined by flux weakening [2],[4]. Thus, over the full speed range of the motor, i_d command current is defined as:

$$i_d^* = (1-n)\left[\frac{(L_d-L_q)}{\phi_m} i_q^2\right] - \frac{n}{L_d}\left[\phi_m - \frac{V_m^{'}}{n_p\omega_r} + \frac{L_q^2 n_p \omega_r}{2V_m^{'}} i_q^2\right] \tag{5}$$

where, $n(\omega_r) = \begin{cases} 0, & \omega_r \le \omega_{rated} \\ 1, & \omega_r > \omega_{rated} \end{cases}$

The principle objective is to achieve control of the IPMSM drive angular speed, thus by defining the speed error as the difference between set point and dynamic speed as:

$$e_\omega = \omega_r^* - \omega_r \tag{6}$$

An initial Lyapunov stability function can be defined as

$$V = \tfrac{1}{2} e_\omega^2 \tag{7}$$

By differentiating (7) and expressing in terms of the speed error dynamic

$$\dot{V} = e_\omega\left[\dot{\omega}_r^* + \frac{\tau_m}{J} + \frac{B_m}{J}\omega_r - \frac{3n_p(L_d-L_q)}{2J} i_d i_q - \frac{3n_p\phi_m}{2J} i_q\right] \tag{8}$$

In order to stabilize (8) the contribution of reluctance torque is regarded as minimal; if this assumption is not made, further control design becomes unfeasible. By choosing q-axis stabilizing current as

$$i_q^* = \frac{2J}{3n_p\phi_m}\left[\dot{\omega}_r^* + \frac{\tau_m}{J} + \frac{B_m}{J}\omega_r + k_\omega e_\omega\right] \tag{9}$$

then, $\dot{V} = -k_\omega e_\omega^2$

In order to force the d- and q-axis currents to track the reference trajectories error terms must be defined. Further, in light of the parametric uncertainty of the IPMSM and absence of direct torque measurement, parametric error dynamics are defined in terms of estimated parameters.

$e_d = i_d^* - i_d$, $e_q = i_q^* - i_q$, $e_D = \hat{L}_d - L_d$, $e_Q = \hat{L}_q - L_q$, $e_\lambda = \hat{\lambda} - \lambda$, $e_\beta = \hat{\beta} - \beta$ where, for convenience $\lambda = \frac{\tau_m}{J}$ and $\beta = \frac{B_m}{J}$

With the new error dynamics expressed, it becomes possible to define a comprehensive Lyapunov stability function as

$$V_2 = \tfrac{1}{2}(e_\omega^2 + e_d^2 + e_q^2 + \tfrac{1}{n_D}e_D^2 + \tfrac{1}{n_Q}e_Q^2 + \tfrac{1}{n_\lambda}e_\lambda^2 + \tfrac{1}{n_\beta}e_\beta^2) \tag{10}$$

Thus it is necessary to both redefine the speed error dynamics and express *d*- and *q*-axis errors in terms of the estimated command currents. By differentiating (10) the following is found

$$\dot{V}_2 = e_\omega\dot{e}_\omega + e_d\dot{e}_d + e_q\dot{e}_q + \tfrac{1}{n_D}e_D\dot{e}_D + \tfrac{1}{n_Q}e_Q\dot{e}_Q + \tfrac{1}{n_\lambda}e_\lambda\dot{e}_\lambda + \tfrac{1}{n_\mu}e_\beta\dot{e}_\beta \tag{11}$$

In order to express the error dynamic terms of (12)-(14), the initial speed error term in (6) must be expressed in terms of estimated functions of *d*-*q* axis currents, such that all error dynamics are functions of estimated parameters as:

$$\begin{aligned}\dot{e}_\omega = {} & \frac{3n_p(L_d-L_q)}{2J} e_d i_q - e_\lambda - e_\beta\omega_r - k_\omega e_\omega \\ & - \frac{3n_p(L_d-L_q)}{2J}\frac{(1-n)(\hat{L}_d-\hat{L}_q)i_q^3}{\phi_m} + \frac{3n_p\phi_m}{2J} e_q \\ & + \frac{3nn_p(L_d-L_q)i_q}{2J\hat{L}_d}\left[\phi_m - \frac{V_m^{'}}{n_p\omega_r} + \frac{\hat{L}_q^2 n_p\omega_r i_q^2}{2V_m^{'}}\right]\end{aligned} \tag{12}$$

$$\begin{aligned}\dot{e}_d = {} & \frac{2(1-n)(L_d-L_q)}{\phi_m} i_q\dot{i}_q + \frac{R_s}{\alpha L_d} i_d + k_d e_d \\ & - \frac{1}{L_d} v_d - \frac{n_p L_q}{L_d}\omega_r i_q - \frac{nV_m^{'}}{n_p L_d \omega_r^2} \\ & - \frac{nn_p L_q^2}{2L_d V_m^{'}}(\dot{\omega}_r i_q^2 + 2\omega_r i_q\dot{i}_q)\end{aligned} \tag{13}$$

$$\begin{aligned}\dot{e}_q = {} & \frac{2J\dot{\lambda}}{3n_p\phi_m} + \frac{2J\beta\dot{\omega}_r}{3n_p\phi_m} + \frac{3n_p(L_d-L_q)(B_m-k_\omega J)i_d i_q}{J\phi_m} + k_q e_q \\ & + \frac{(B_m-k_\omega J)i_q}{J} - \frac{2(B_m-k_\omega J)}{3n_p\phi_m}\lambda - \frac{2(B_m-k_\omega J)\omega_r}{3n_p\phi_m}\beta \\ & - \frac{1}{L_q} v_q + \frac{n_p L_d}{L_q}\omega_r i_d + \frac{n_p\phi_m}{L_q}\omega_r + \frac{R_s}{L_q} i_q\end{aligned} \tag{14}$$

With the error dynamics defined in terms of estimated parameters, the *d* and *q*-axis stabilizing command voltages are expressed as:

$$\begin{aligned}v_d = {} & R_s i_d - n_p\omega_r\hat{L}_q i_q + \frac{2(1-n)(\hat{L}_d-\hat{L}_q)\hat{L}_d i_q\dot{i}_q}{\phi_m} \\ & - \frac{nV_m^{'}}{n_p\omega_r^2} - \frac{nn_p\hat{L}_q^2}{2V_m^{'}}(\dot{\omega}_r i_q^2 + 2\omega_r i_q\dot{i}_q) + k_d e_d\hat{L}_d \\ & - \frac{3(1-n)n_p(\hat{L}_d-\hat{L}_q)^2\hat{L}_d i_q^3 e_\omega}{2J\phi_m e_d} + \frac{3n_p(\hat{L}_d-\hat{L}_q)\hat{L}_d e_\omega i_q}{2J} \\ & + \frac{3nn_p(\hat{L}_d-\hat{L}_q)\hat{L}_d i_q e_\omega}{2Je_d}\left[\phi_m - \frac{V_m^{'}}{n_p\omega_r} + \frac{\hat{L}_q^2 n_p\omega_r}{2V_m^{'}} i_q^2\right]\end{aligned} \tag{15}$$

$$\begin{aligned}v_q = {} & R_s i_q + k_q e_q\hat{L}_q + n_p\phi_m\omega_r + n_p\hat{L}_d\omega_r i_d \\ & + \frac{2\hat{L}_q(B_m-k_\omega J)}{3n_p\phi_m}\left(\frac{3n_p(\hat{L}_d-\hat{L}_q)i_d i_q}{2J} + \frac{3n_p\phi_m i_q}{2J} - \hat{\lambda} - \hat{\beta}\omega_r\right) \\ & + \frac{3n_p\phi_m\hat{L}_q e_\omega}{2J} + \frac{2J\hat{L}_q}{3n_p\phi_m}\hat{\beta}\dot{\omega}_r\end{aligned} \tag{16}$$

In order to find relations describing the parameters to be estimated, adaptive backstepping is employed to express the error dynamics e_D, e_Q, e_λ, e_β which are then differentiated to yield estimated values of d and q-axis inductance, load torque and mechanical coefficient respectively. Thus, having the Lyapunov derivative defined in (11), the next steps are to express the previously defined error dynamics (12)-(14) in terms of the *d* and *q*-axis command voltages defined in (15), (16). Grouping the resultant terms according to the respective error dynamics, cancelling opposing terms and simplifying terms containing higher order instances of error dynamics permits the expression of the parametric update laws. The update laws are as follows, consisting of d-q axis inductances, load torque and mechanical constant error dynamics respectively.

$$\dot{e}_D = -n_D \begin{bmatrix} -\frac{3nn_p e_\omega i_q}{2JL_d}\left(\phi_m - \frac{V_m^*}{n_p\omega_r} + \frac{L_q^2 n_p \omega_r i_q^2}{2V_m^*}\right) \\ -\frac{3n_p i_q e_d e_\omega}{2J} - \frac{2(1-n)i_q \dot{i}_q e_d}{\phi_m} - \frac{n_p \omega_r i_d e_q}{L_q} \\ +\frac{9(1-n)n_p e_\omega i_q^3 (L_d - L_q)}{2J\phi_m} - \frac{i_d i_q (B_m - k_\omega J) e_q}{J\phi_m} \end{bmatrix} \quad (17)$$

$$\dot{e}_Q = -n_Q \begin{bmatrix} \frac{n_p \omega_r i_q e_d}{L_d} + \frac{3nn_p e_\omega i_q}{2JL_d}\left(\phi_m - \frac{V_m^*}{n_p\omega_r} + \frac{L_q^2 n_p \omega_r i_q^2}{2V_m^*}\right) \\ +\frac{3n_p i_q e_d e_\omega}{2J} + \frac{nn_p(\dot{\omega}_r i_q^2 + 2\omega_r i_q \dot{i}_q) L_q e_d}{V_m^* L_d} \\ -\frac{9(1-n)n_p e_\omega i_q^3 (L_d - L_q)}{2J\phi_m} - \frac{2(1-n)i_q \dot{i}_q e_d}{\phi_m} + \frac{i_d i_q (B_m - k_\omega J) e_q}{J\phi_m} \end{bmatrix} \quad (18)$$

$$\dot{e}_\lambda = -n_\lambda \left[\frac{(B_m - k_\omega J)e_q}{n_p \phi_m} - e_\omega\right] \quad (19)$$

$$\dot{e}_\beta = -n_\beta \left[\frac{(B_m - k_\omega J)\omega_r e_q}{n_p \phi_m} - \omega_r e_\omega - \dot{\omega}_r\right] \quad (20)$$

By applying d and q-axis command voltages as defined in (15), (16) and with update laws as per (18)-(21), the resultant Lyapunov derivative can be shown to be negative semi-definite with bounded state variables as:

$$\dot{V}_2 = -k_\omega e_\omega^2 - k_d e_d^2 - k_q e_q^2 \le 0 \quad (21)$$

However, in order to demonstrate asymptotic stability of the non-autonomous controller, absolute stability criterion in the form of Barbalat's lemma is employed [11],[12]. Barbalat's lemma can be expressed in the following trio of conditions applied to the Lyapunov stability function [11].

I. As per (10) $V_2 \ge 0$, and is thus characterized by a lower bound

II. As per (22) $\dot{V}_2 \le 0$, thus the negative semi-definite condition is validated.

III. With $\ddot{V}_2 = -2k_\omega e_\omega \dot{e}_\omega - 2k_d e_d \dot{e}_d - 2k_q e_q \dot{e}_q$ verified to be bounded (error dynamics defined a priori; and with all design coefficients, command variables and error dynamics bounded), it can be validated that $\dot{V}_2$ is a continuous function of time.

Having satisfied Barbalat's lemma, it can be concluded that

$$\dot{V}_2 \to 0 \; as \{e_\omega, e_d, e_q\} \to 0$$

Thus, the system is shown to be convergent everywhere, demonstrating global asymptotic stability.

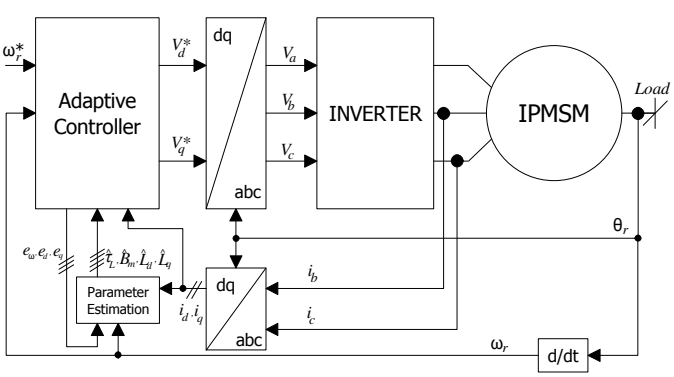


Fig. 1: Block diagram of the proposed controller

IV. Simulation Results

Simulation of the proposed nonlinear controller in Fig. 1 was carried out in a Matlab/Simulink software environment. Nominal machine parameters are as follows: R_s=0.242Ω, L_d=5.06mH, L_q=6.42mH, J=0.0133kgm^2, B_m=0.001Nm/rad/s, Φ_m=0.2449volts/rad/s, n_p=3, ω_r=183.3rad/s, V_m=183volts, P_{rated}=3.7kW.

To demonstrate the robustness of the adaptive laws, Fig. 2(a) compares the response of a scenario with an accidental incorrect mechanical coefficient to that using the adaptive estimation laws, at rated load condition. Fig. 2(b) shows reluctance torque contribution at rated conditons; comparatively this reluctance torque contribution is forfeit in control schemes using i_d=0. Fig. 3 shows the response of the proposed IPMSM drive for a step increase in speed, at full load from 183.3 rad/s at t=0s, to 350 rad/s at t=1s. It is seen from Fig. 3(a) that the drive can follow the command speed smoothly even with a command speed above rated. The increase in speed is also verified by the 'a' phase stator current in Fig. 3(b). The corresponding d-q axis currents are shown in Fig. 3(c). Flux weakening control can be seen from Fig. 3(c) as the d-axis current magnitude shifts downward as the command speed is increased above the rated speed. The corresponding dynamic error values of speed, i_d and i_q are shown in Fig. 3(d).

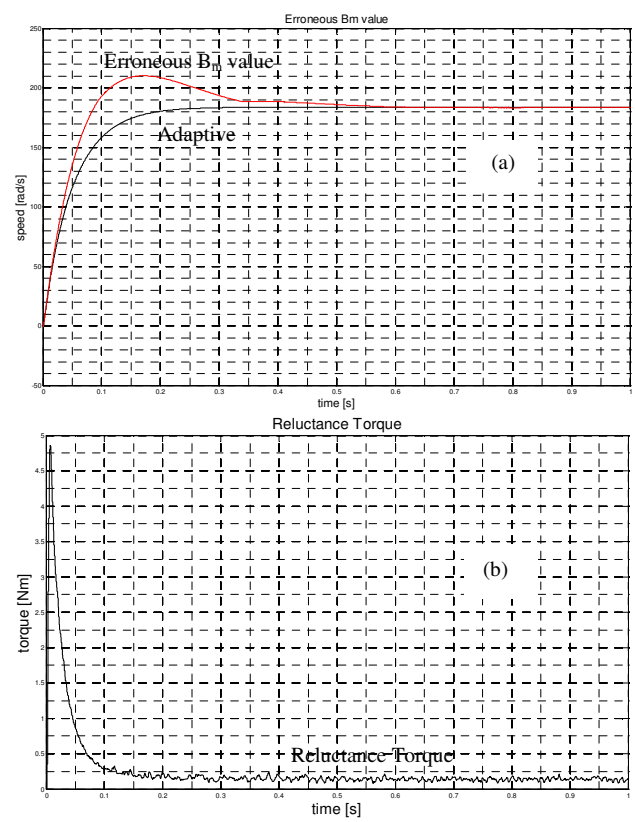


Fig 2: Robust operation and torque analysis of the proposed nonlinear controller based IPMSM drive: (a) speed response using erroneous B_m value at rated conditions (b) reluctance torque at rated conditions

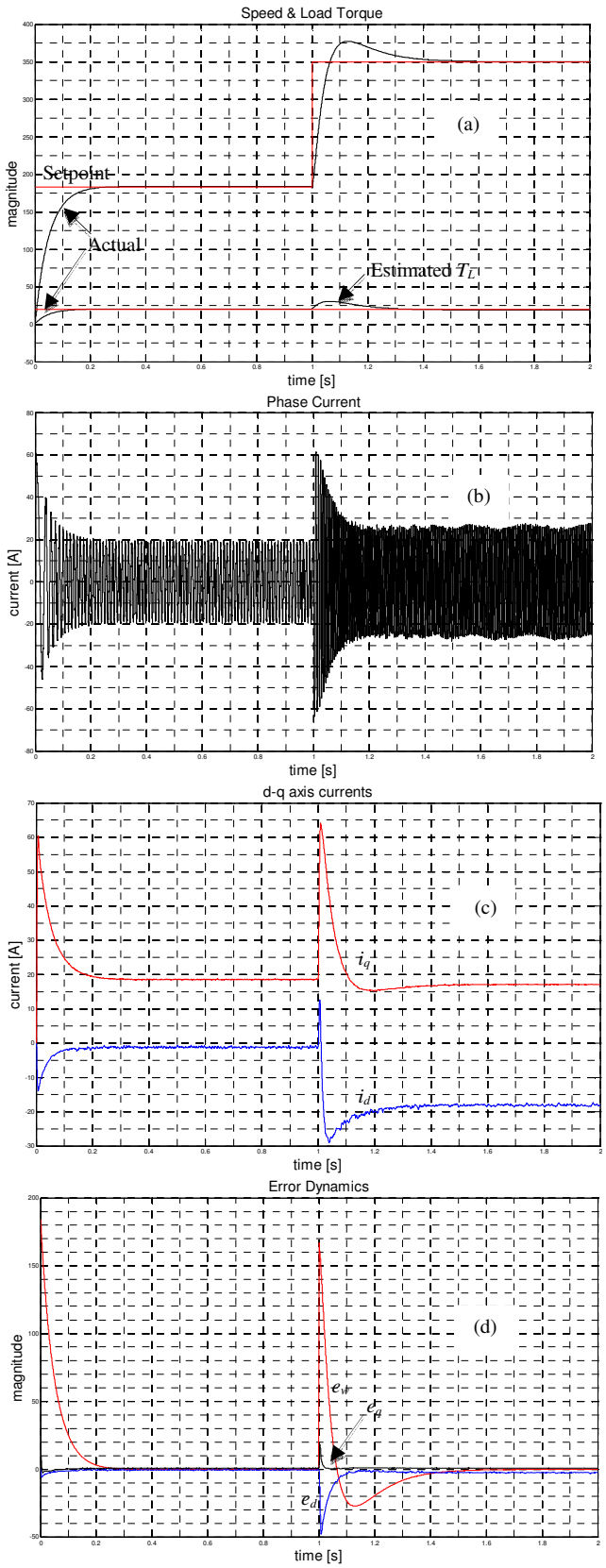


Fig 3: Simulated responses of the proposed nonlinear controller based IPMSM drive for step increase in speed (183.3 rad/s to 350 rad/s at 1s, with 20 Nm load): (a) speed and estimated torque (b) i_a (c) i_d , i_q (d) error dynamics

V. CONCLUSIONS

A nonlinear flux controller based wide speed operation of permanent magnet synchronous motor has been presented in this paper. The flux control based on MTPA and FW is considered in the design stage of the nonlinear controller. The MTPA and FW techniques are used to control the d-axis component of stator current, i_d, below and above the rated speed, respectively. To ensure system stability, control laws were developed based on adaptive backstepping technique; and system nonlinearities were considered through online estimation of motor parameters. Global asymptotic stability of the drive was validated through Lyapunov stability criterion and Barbalat's lemma. Effectiveness of the proposed nonlinear controller has been verified through simulation results at different operating conditions, demonstrating excellent dynamic performance.

REFERENCES

[1] M. A. Rahman, M. Vilathgamuwa, M. Uddin and K.-J. Tseng, "Nonlinear Control of Interior Permanent-Magnet," *IEEE Trans. Ind. App.,* vol. 39, no. 2, pp. 408-416, 2003.

[2] M. N. Uddin and M. M. I. Chy, "Online Parameter-Estimation-Based Speed Control of," *IEEE Trans. Industry Applications,* vol. 44, no. 5, pp. 1486-1494, 2008.

[3] J. Lau and M. Uddin, "Nonlinear Adaptive Position Control of an Interior," in *IEMDC*, San Antonio, 2005.

[4] Z. Fangyang, R. Feng, L. Jianjun and H. Peng, "Study on Flux–Weakening Control for PMSM," in *KAM Inter. Symp.*, Sanya, 2011.

[5] K. B. Lee, F. Blaabjerg, "Sensorless DTC-SVM for Induction Motor Driven by a Matrix Converter Using a Parameter Estimation Strategy", *IEEE Trans. on Ind. Electronics*, Vol. 55, No. 2, pp. 512-521, Feb. 2008.

[6] P. Vaclavek, P. Blaha, "Lyapunov-function-based flux and speed observer for AC induction motor sensorless control and parameters estimation", *IEEE Trans. on Industrial Electronics*, Vol. 53, No.1, pp. 138- 145, Feb. 2006.

[7] L. Harnefors, "Globally Stable Speed-Adaptive Observers for Sensorless Induction Motor Drives", *IEEE Trans. on Industrial Electronics,* Vol. 54, No. 2, pp. 1243-1245, April 2007.

[8] M. N. Uddin, T. S. Radwan and M. A. Rahman, "Performance of Interior Permanent Magnet Motor Drive over Wide Speed Range", IEEE Trans. on Energy Conversion, vol. 17, no. 1, March 2002, pp. 79-84.

[9] S. Morimoto, M. Sanada and Y. Takeda, "Effects and Compensation of Magnetic Saturation in Flux-Weakening Controlled Permanent Magnet Synchronous Motor Drives," *IEEE Transactions on Industry Applications*, vol. 30, no. 6, Nov./Dec. 1994, pp. 1632-1637.

[10] Jiunn-Jiang Chen and Kan-Ping Chen, "Automatic Flux-Weakening Control of Permanent Magnet Synchronous Motors Using a Reduced-Order Controller", *IEEE Transactions on Power Electronics*, vol.15, No.5, Sept. 2000, 881-890

[11] J. Mash, M. Pahlevaninezhad and P. Jain, "Nonlinear Control of Wind Energy Conversion System based on Control-Lyapunov Function", IEEE ECCE, 2013, 2955-2961

[12] H. K. Khalil, Nonlinear Systems, Upper Saddle River, Prentice Hall, 1996

Modeling and Transient Performance Analysis of a Hysteresis IPM Motor

S. F. Rabbi, Student Member, IEEE and M. A. Rahman, Life Fellow, IEEE
Department of Electrical and Computer Engineering
Memorial University of Newfoundland, St. John's, NL, Canada, A1B 3X5

sfr560@mun.ca, arahman@mun.ca

Abstract— **A hysteresis interior permanent magnet (IPM) motor is a solid-rotor self-starting hybrid synchronous motor combining both hysteresis phenomena and permanent magnet excitation inside the rotor. The rotor of a hysteresis IPM motor consists of a solid hysteresis ring made of composite materials of high degree of magnetic hysteresis energy per unit volume. High energy density rare earth permanent magnets are buried inside the rotor hysteresis ring. This paper presents the modeling and transient analysis of a hysteresis IPM motor using finite element methods. Electrical equivalent circuits of a hysteresis IPM motor have been presented considering the *d-q* axis saliency of the rotor. Transient finite element analysis of a 2.5kW 208V 3-phase 4-pole hysteresis IPM motor has been carried out using ANSYS Maxwell Electromagnetic field simulation software. Simulation results are presented and explained in this paper.**

Index Terms—**Magnetic hysteresis, permanent magnet motors, analysis, modeling, equivalent circuits, simulation.**

I. Introduction

A hysteresis interior permanent magnet (IPM) motor is a self-starting solid rotor hybrid synchronous motor. Its rotor has a cylindrical ring made of composite material like 17%/36% cobalt steel alloy, special Al-Ni-Co, Vicalloy, etc. with high degree of hysteresis energy per unit volume [1-5]. The rare earth permanent magnets are buried inside the hysteresis ring. The ring is supported by a sleeve made of non-magnetic materials like Aluminum which forces the flux to travel circumferentially inside the rotor ring [1-11]. Fig. 1 illustrates the rotor of a hysteresis IPM motor. The cross section of a hysteresis IPM motor depicting the position and orientation of permanent magnets is shown in Fig. 2. The inclusion of permanent magnets creates rotor saliency without changing the length of the physical airgap and provides an additional permanent source of excitation in the rotor. The induced magnetization of the hysteresis material inside the rotor ring always lags behind the time varying magnetic field. This time lag produces a torque, called hysteresis torque. Also, the induced eddy currents in the hysteresis ring generate some additional starting torque called the eddy current torque. The combination of the hysteresis torque and the eddy current torque results in a high starting torque during line starting of a hysteresis IPM motor which overcomes the magnet brake torque [1-7]. The eddy current torque becomes zero when the motor synchronizes but the hysteresis torque still exists, and the hysteresis ring behaves like a temporarily magnetized permanent magnet. The synchronous torque of a hysteresis IPM motor consists of the magnet alignment torque, reluctance torque and hysteresis torque. As a result, a hysteresis IPM motor has a larger synchronizing torque than that of a conventional line start IPM motor

This paper presents modeling and finite element analysis of a hysteresis interior permanent magnet motor. Electrical equivalent circuits of a hysteresis IPM motor in *d-q* axis are illustrated and explained in this paper. Transient finite element analysis of a 2.5kW 208V 3-phase 4-pole hysteresis IPM motor has been carried out using ANSYS Maxwell Electromagnetic field simulation software. The nonlinear hysteresis phenomenon of the rotor material is approximated using advanced vector hysteresis modeling [12]. Simulation results are presented and explained in this paper. Simulation results exhibit that the hysteresis IPM motor has a moderate starting current, high starting torque and good synchronization capability.

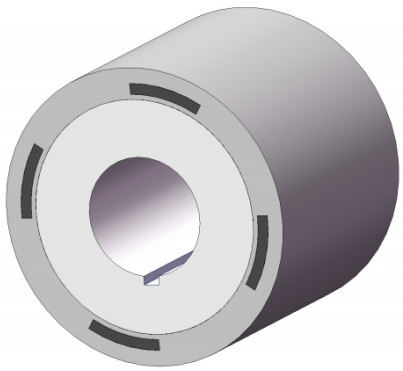


Figure 1. Rotor of a hysteresis IPM motor

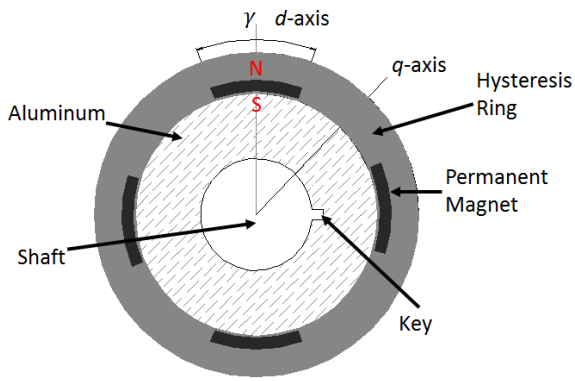


Figure 2. Cross section of hysteresis IPM rotor

II. Electrical Equivalent Circuits

The following assumptions are considered for developing the electrical equivalent circuit of a hysteresis IPM motor:

I. The stator is assumed to have a balanced 3φ sinusoidally distributed winding.

II. Only the fundamental component of the flux density *B* in response to the magnetizing field *H* is considered.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

III. The magnetic flux density is radial in the airgap, but circumferential inside the rotor hysteresis ring.
IV. The B-H curve of the hysteresis material is approximated using the elliptical modeling.
V. The permeability of the hysteresis material is assumed to be constant at its normalized saturated value.
VI. The maximum hysteresis lag angle is assumed to be between 30-70 degrees.
VII. The permanent magnet is assumed to be radially magnetized.

The *q-d* axis electrical equivalent circuits of a hysteresis IPM motor are depicted in Figs. 3a and 3b. The permanent magnet is modeled as a constant current source I_m. Eddy current effect is included in the circuit by modeling it as an equivalent resistance R_e. Hysteresis effect is modeled as an equivalent hysteresis resistance R_h, and an equivalent hysteresis inductance L_{hr}.

Hysteresis and eddy current resistances are divided into two components. One component reflects the hysteresis loss and the eddy current loss in the rotor, modeled by R_e and sR_h, respectively. The other component contributes to the output power of the motor, combining power derived from the effective eddy current resistance and the effective hysteresis resistance. The effective eddy current and hysteresis resistances are expressed as $((1-s)/s)R_e$ and $(1-s)R_h$, respectively, where s is the slip of the motor. In the sub-synchronous mode of operation, both the loss component and the effective load component of the eddy current resistance and the hysteresis resistance exist in the rotor, and results in high I^2R losses in the rotor. At synchronous speed, the eddy current resistance and the hysteresis loss component vanish and the torque produced by the motor is due to combinations of the permanent magnet and the hysteresis effects. The total developed electromagnetic torque at synchronous speed is the summation of the magnet alignment torque, reluctance torque and the hysteresis torque. The hysteresis IPM motor can synchronize larger load torques than the conventional squirrel cage line start IPM motor because of the additional hysteresis torque.

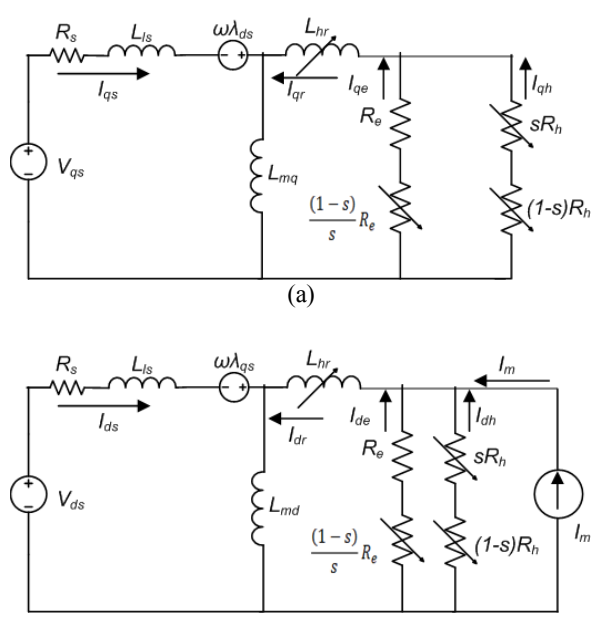


Figure 3. Electrical equivalent circuits of a hysteresis IPM motor: (a) *q*-axis and (b) *d*-axis

The *q-d* axis equations of the motor are derived from the equivalent circuits and are expressed as,

$$\begin{bmatrix} V_{qs} \\ V_{ds} \\ V_0 \end{bmatrix} = \frac{2}{3}\begin{bmatrix} \cos\theta_r & \cos(\theta_r - \frac{2\pi}{3}) & \cos(\theta_r + \frac{2\pi}{3}) \\ \sin\theta_r & \sin(\theta_r - \frac{2\pi}{3}) & \sin(\theta_r + \frac{2\pi}{3}) \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix}\begin{bmatrix} V_a \\ V_b \\ V_c \end{bmatrix} \quad (9)$$

$$\theta_r = \int_0^t \omega_{rm}(t)dt + \theta_r(0) \quad (10)$$

$$V_{qs} = R_s I_{qs} + \frac{d\lambda_{qs}}{dt} + \omega_r \lambda_{ds} \quad (10)$$

$$V_{ds} = R_s I_{ds} + \frac{d\lambda_{ds}}{dt} - \omega_r \lambda_{qs} \quad (11)$$

$$V_{qr} = 0 = R_r I_{qr} + \frac{d\lambda_{qr}}{dt} \quad (12)$$

$$V_{dr} = 0 = R_r I_{dr} + \frac{d\lambda_{dr}}{dt} \quad (13)$$

$$\lambda_{qs} = I_{qs}(L_{mq} + L_{ls}) + L_{mq}(I_{qe} + I_{qh}) \quad (14)$$

$$\lambda_{ds} = I_{ds}(L_{md} + L_{ls}) + L_{md}(I_{de} + I_{dh} + I_m) \quad (15)$$

$$\lambda_{qr} = (I_{qe} + I_{qh})(L_{hr} + L_{mq}) + L_{mq} I_{qs} \quad (16)$$

$$\lambda_{dr} = (I_{de} + I_{dh})(L_{hr} + L_{md}) + L_{md} I_{ds} + L_{md} I_m \quad (17)$$

$$R_r = \frac{\frac{R_e}{s} * R_h}{\frac{R_e}{s} + R_h} = \frac{1}{\frac{1}{R_h} + \frac{s}{R_e}} \quad (18)$$

$$T_e = \frac{3}{2} p(\lambda_{ds} I_{qs} - \lambda_{qs} I_{ds}) \quad (19)$$

where R_s is the stator phase resistance, L_{ls} is the per phase stator leakage inductance, L_{mq} and L_{md} are the *q-d* axis magnetizing inductances, respectively and ω_r is the electrical angular frequency of the rotor. V_{qs} and V_{ds} are the *q-d* axis stator voltages, V_{qr} and V_{dr} are the *q-d* axis rotor induced voltages, θ_r is the rotor position, λ_{qs} and λ_{ds} are the *q-d* axis stator flux linkages, λ_{qr} and λ_{dr} are the *q-d* axis rotor flux linkages, R_r is the total equivalent rotor resistance and T_e is the developed electromagnetic torque.

III. Finite Element Modeling of Hysteresis IPM Motors

Figure 1 illustrates one quarter of a 2D hysteresis interior permanent magnet motor finite element model which is developed in ANSYS Maxwell 2D, based on the geometric symmetry and electrical symmetry of the motor. The model is discretized into small triangular elements. All the triangles are assembled together by an adaptive mesh technique. In two dimensional magnetic field solutions, the vectors have only one component in the z-direction. The magnetic vector potential, A_z in each element is approximated using a second order quadratic polynomial [13]. Magnetic transient solver is

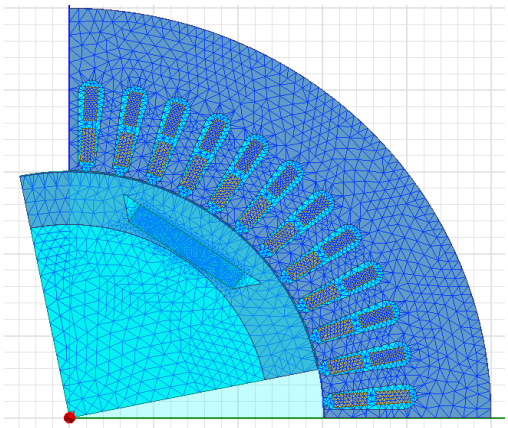


Figure 4. Finite element model of a hysteresis IPM motor

used to analyze the magnetic fields, energy, torque, losses, speed, currents and flux of the hysteresis IPM motor model using time stepping finite element analysis. In time stepping FEA, the desired vector field is computed at various time steps of a solution over a specified time period [13]. The time dependent magnetic equation which is used by the Maxwell magnetic transient solver is given by [13],

$$\nabla \times \upsilon \nabla \times A = J_s - \sigma \nabla V + \nabla \times H_c + \sigma v \times \nabla \times A \qquad (5)$$

where H_c is the coercivity of the permanent magnet, v is the velocity of the rotor, A is the magnetic vector potential, V is the electrical potential, υ is the reluctivity and J_s is the source current density. The excitation is set up as a balanced three phase ac voltage sources and the conductors are assumed to be stranded. The boundary condition of the model is chosen as the odd symmetry boundary which capitalizes the geometric and electrical symmetry of a model in order to reduce the computational time. In the odd symmetry boundary, the signs of the magnetic/ electric fields on one side of a symmetry plane are the opposite of those on the other side of the plane [13].The magnetic transient solver computes the magnetic vector potential in the z-direction A_z. The magnetic field intensity H and the magnetic flux density B are derived from A_z using the Maxwell's equations.

The magnetic flux distribution is displayed in Fig. 5. Fig. 6 illustrates the magnetic flux density. The flux is radial in the airgap and travels circumferentially inside the hysteresis ring.

IV. Transient Performance Analysis

The transient run-up responses of a 3-phase 4-pole 208V 2.5kW hysteresis IPM motor for different loads are shown in Fig. 7. The motor is run by a 3-phase Δ-connected 208V 60Hz balanced ac supply. The motor starts smoothly with some rotor oscillations at the low speed because of the magnet brake torque. The motor overcomes the maximum magnet brake torque with the help of high hysteresis torque and eddy current torque, and ramps up towards the synchronous speed. The developed instantaneous electromagnetic torque is illustrated in Fig. 8. The hysteresis IPM motor has a starting torque. The developed torque is comprised of an average asynchronous torque and a pulsating torque. The average asynchronous torque combines the hysteresis torque, the eddy current torque and the magnet brake torque. Due to the presence of permanent magnet acting as a current source in the rotor, there is a pulsating torque superimposed on the average

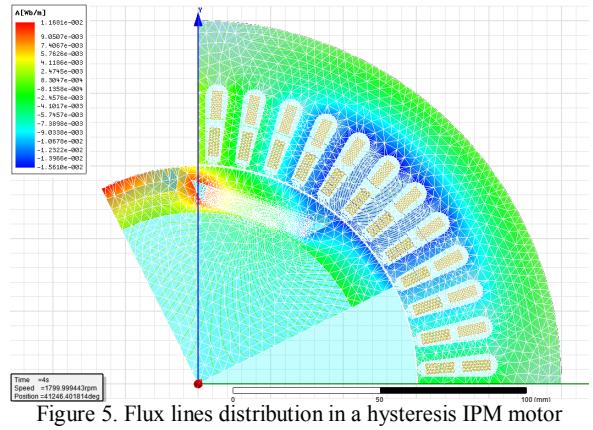


Figure 5. Flux lines distribution in a hysteresis IPM motor

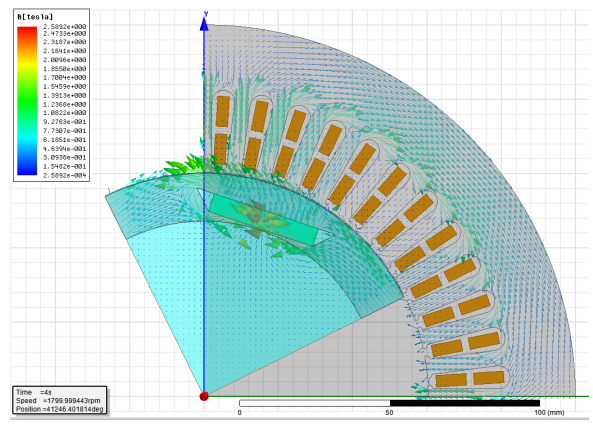


Figure 6. Flux density inside a hysteresis IPM motor

asynchronous torque. The pulsating torque assists the motor to a fast acceleration towards the synchronous speed. During the period of synchronization, the torque pulsation dies out and turns into a part of the synchronous torque. The eddy current torque also vanishes when the slip approaches zero. The magnet brake torque reverses its role and becomes a constituent member of the synchronous torque. The stator phase current for rated load is illustrated in Fig. 9. The starting current is only 2.5 times higher than the steady state rated current which is a significant improvement of conventional line start IPM motors [14]. In Fig. 10, the torque vs. speed trajectories for different loads are depicted. Because of the interaction between the permanent magnet flux and the currents inside the hysteresis ring, the line start hysteresis IPM motor suffers from a few number of torque reversals at low speed. Before reaching the synchronization, there is a series of slow pole-slips which appears in Fig. 10 as almost elliptical. The final synchronization trajectory follows a spiral that originates at the synchronous speed. The load torque was changed suddenly from no load to full rated load torque at t=2s. The rotor goes through some speed overshoots and undershoots for a short time and becomes synchronized again. Both the electromagnetic torque and the phase currents have gone through some oscillations for a very short time before reaching to higher steady state values. The motor starts with a constant hysteresis lag angle but goes though cyclic oscillations during the synchronization period before reaching to a steady state value. As the load torque becomes higher, the lag angle increases which results in a higher torque. Fig. 11 shows the constant flux linkages in phase A of the motor.

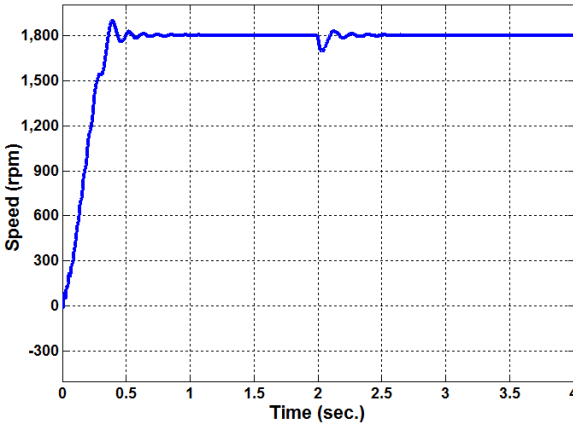


Figure 7. Run-up responses of a hysteresis IPM motor

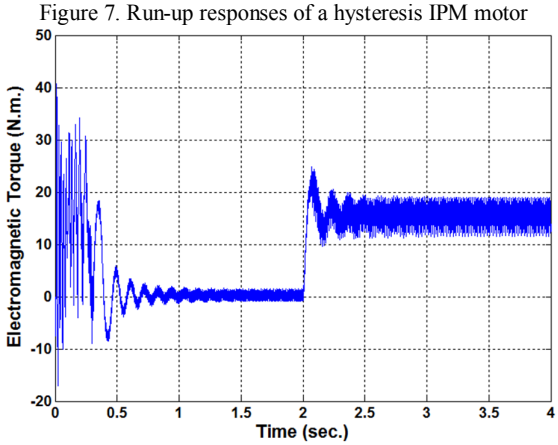


Figure 8. Developed torque of a hysteresis IPM motor

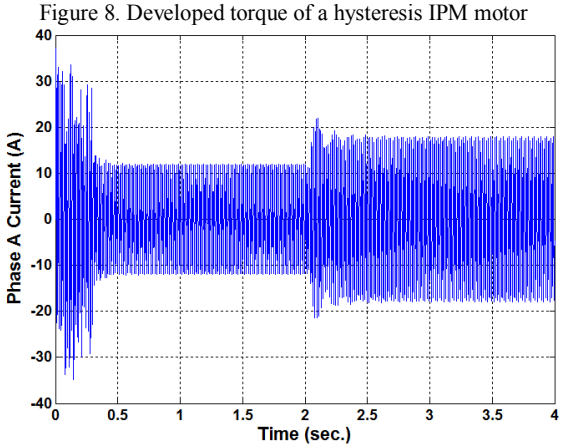


Figure 9. Phase current of a hysteresis IPM motor

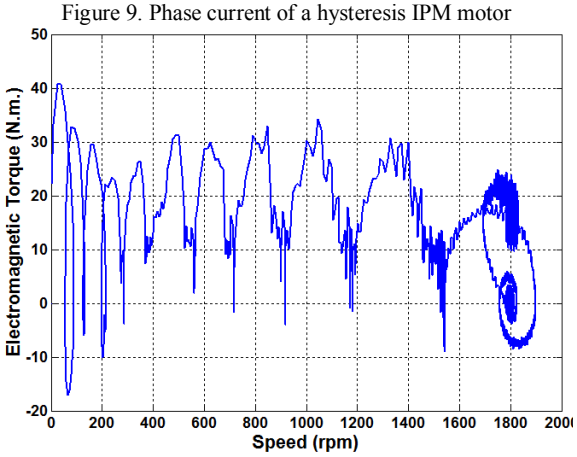


Figure 10. Torque- speed trajectories of a hysteresis IPM motor

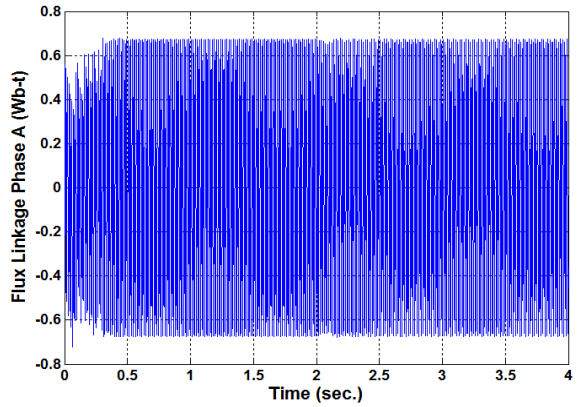


Figure 11. Magnetic flux linkage in phase A

VI. CONCLUSION

This paper presents the modeling and transient finite element analysis of a hysteresis interior permanent magnet (IPM) motor. *d-q* axis equivalent circuits have been developed in this paper considering both the eddy currents and the hysteresis phenomena in the rotor. Finite element transient analysis results of a hysteresis IPM motor are presented and explained in this paper. The simulation results exhibit that the proposed hysteresis IPM motor has good self-starting and synchronization capabilities.

REFERENCES

[1] M. A. Rahman and R. Qin., "Starting and Synchronization of Permanent Magnet Hysteresis Motors", *IEEE Transactions on Industry Applications*, Vol. 32, No. 5, pp. 1183-1189, September 1996.

[2] M. A. Rahman and R. Qin, "A permanent magnet hysteresis hybrid synchronous motor for electric vehicles", *IEEE Transactions on Industrial Electronics*, Vol. 44, No. 1, pp. 46–53, February 1997.

[3] J. Qian and M. A. Rahman, "Analysis of field oriented control for permanent magnet hysteresis synchronous motors", *IEEE Transactions on Industry Applications*, Vol. 29, No. 6, pp. 1156-1163, Nov. 1993.

[4] R. Qin and M. A. Rahman, "Magnetic equivalent circuit of PM hysteresis synchronous motor", *IEEE Transactions on Magnetics*, Vol. 39, No. 5, pp. 2998 – 3000, September 2003.

[5] K. Kurihara and M. A. Rahman, "Transient performance analysis for permanent-magnet hysteresis synchronous motor," *IEEE Transactions on Industry Applications*, Vol. 40, No. 1, pp. 135–142, January 2004.

[6] J. J. Nitao, E. T. Scharlemann and B. A. Kirkendall, "Equivalent Circuit Modeling of Hysteresis Motors", Lawerence Livermore National Laboratory , pp. 1-52, July 2009.

[7] M. A. Rahman, "History of interior permanent magnet motors," IEEE Industry Applications Magazine, vol.19, no. 1, pp. 10-15, January 2013.

[8] S. Miyairi and T. Kataoka, "A basic equivalent circuit of the hysteresis motor", *IEE Transactions of Japan*, Vol. 85, pp. 41-50, October 1965.

[9] S. Miyairi and T. Kataoka, "Analysis of hysteresis motors considering eddy current effects", *IEE Transactions of Japan*, Vol. 85, pp. 67-77, June 1966.

[10] T. Ishikawa and T. Kataoka, "Basic analysis of disc-type hysteresis motors," *Journal of Electrical Engineering Japan*, Vol. 101, No. 6, pp. 659-666, November 1981.

[11] M. A. Rahman, "Analytical models of polyphase hysteresis motors", *IEEE Transactions on Power Apparatus and Systems*, Vol. PAS-92, No. 1, pp. 237-242, January 1973.

[12] D. Lin, P. Zhou and A. Bergqvist, "Improved Vector Play Model and Parameter Identification for Magnetic Hysteresis Materials", *IEEE Transactions on Magnetics*, Vol. 50, No. 2, pp. , FEBRUARY 2014

[13] Maxwell Online Help: Release v15.0 (December 2013), ANSYS, Inc., [Online], Available: http:www. ansys. com

[14] S. F. Rabbi and M. A. Rahman, "Critical criteria for successful synchronization of line start IPM motors", *IEEE Journal of Emerging and Selected Topics in Power Electronics*, Vol. 2, No.2, pp. 348-358, June 2014.

Development of a Microcontroller-Based AC Voltage Controller with Soft Start Capability

Arifur Rahman*, Nayeem Ansari, Nazneen Ahmed, Kazi Mujibur Rahman and Md. Zahurul Islam

Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology, Dhaka-1205, Bangladesh
*arifurrahman.mail@gmail.com

***Abstract*—This paper describes the development of a prototype of a microcontroller-based phase angle controlled single-phase AC voltage controller that can efficiently control AC voltage and also accommodates soft start capability for single-phase induction motors. The output voltage of the controller is regulated to maintain a desired fixed RMS value and provide stabilized output by implementing a feedback control system. One microcontroller generates PWM signals in synchronism with the supply voltage to control the firing angle of thyristors while a second microcontroller remains dedicated for measuring the RMS value of the output voltage and sending that to the main microcontroller for the purpose of feedback control. The provision for soft starting of a load is also incorporated into this prototype.**

***Index Terms*—AC voltage controller, Automatic voltage regulation, Soft start**

I. Introduction

PWM-based AC voltage controllers are widely used in UPS and high power flexible AC transmission systems. This varying voltage output is used for dimming street lights, varying heating temperature in homes and industry, speed control of fans and winding machines and many other applications. These systems need switching elements which can bear high voltage.

Frequently, high power MOSFETS are used as the switching element. Their advantage is that they generate less lower-order harmonics. They have the physical limitation of maximum blocking voltage; excessive heating is also an issue for MOSFET based controllers. Amongst other options, an anti-parallel pair of SCRs has the advantage over a TRIAC in controlling highly inductive loads. Although TRIACs have the advantage of a comparatively simpler gate circuit, TRIACs have lower *dv/dt* ratings than SCRs and are available in only small ratings. Moreover, the reliability of a SCR is more than that of a TRIAC [1].

SCR is a reliable solution when it comes to AC voltage control at high power as they are available in higher ratings. However, SCR based system can only be controlled by phase angle control method. This method generates more harmonic distortions [2]. Moreover synchronization with AC supply is necessary when the control mechanism is provided by microcontrollers. The gating circuit required is relatively more complex.

The provision of feedback allows output voltage to be stable and maintain the same RMS value of voltage while there is change in the supply voltage [3]. The gain of the error signal of the feedback control system can be varied to obtain a slower or faster response of the system thus allowing soft start.

II. Overview Of The Design

The development of the prototype is mainly composed of several stages of design consideration: 1) anti-parallel SCR-based voltage controller, 2) Gate-drive circuitry for SCRs, 3) PWM generation through microcontroller, 4) Phase and frequency detection, 5) RMS value calculation for feedback and 6) Input-output Interfacing.

The functional block diagram of the system is shown in Fig. 1.

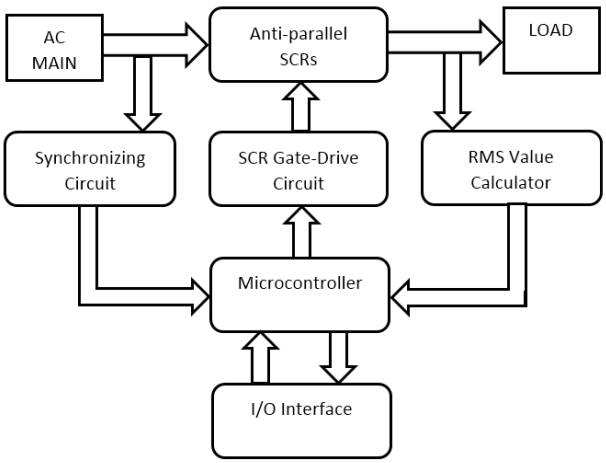


Fig. 1 Functional block diagram of the voltage controller

III. Hardware Implementation

A. Voltage Control Mechanism

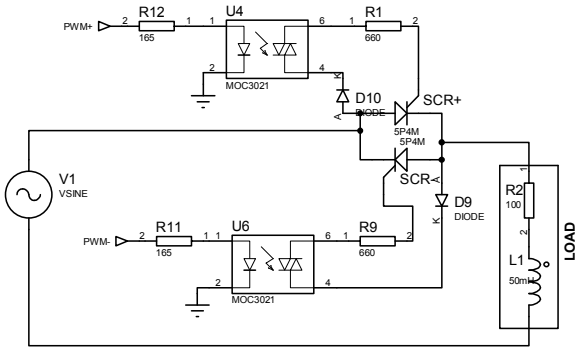


Fig. 2 The main circuit diagram for the prototype

A pair of anti-parallel SCRs is used as the switching device for controlling the AC voltage. The circuit diagram of the SCR based voltage control circuit is shown in Fig. 2. Due to the nature of the connection the two cathodes of the two SCRs corresponds to two different nodes. Thus each gate pulse must

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

be isolated from each other. This is achieved through optical coupling provided by MOC3021 optocouplers [4]. The mechanism is shown in Fig. 3.

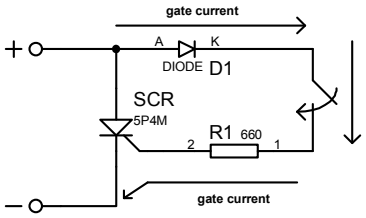


Fig. 3 Path of gate current when optocoupler is turned on

B. Synchronization

The PWM signals must match accurately in phase and frequency with the AC supply line. Even a very small deviation in frequency can gradually lead to the PWM signals becoming out of phase with the supply. A modified zero crossing detector is used for the purpose of synchronization. The circuit diagram is shown in Fig. 4.

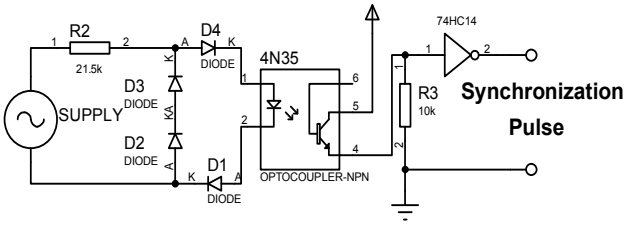


Fig. 4 Synchronization circuit, a modified zero crossing detector

The final output of this circuit is a PWM signal which should ideally have a logic level 0 at the positive half cycle and logic level 1 at the negative half cycle. However due to the presence of resistance in the input circuit the signal mark (logic level 1) starts a little earlier than the start of negative half cycle and ends a little later (by the same amount of time) than the end of the negative half cycle. The timing is shown in Fig. 5.

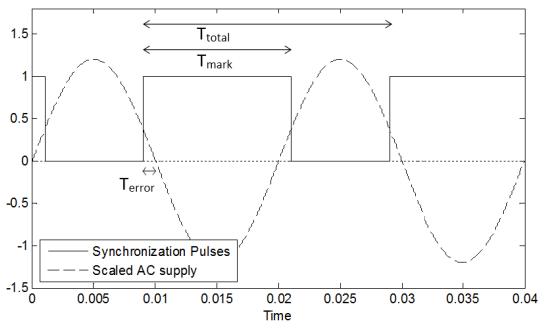


Fig. 5 Timing diagram of synchronization pulses

The error in time can be calculated by the equation (1) which is based on the relative positions of the AC supply and the synchronization pulses.

$$T_{error} = \frac{T_{mark} - T_{total}/2}{2} \quad (1)$$

T_{mark} and T_{total} is calculated using one of the timer resources available in the microcontroller and T_{error} is obtained from the above equation. T_{total} is also the period of the AC supply. Thus both frequency and phase information is obtained.

C. PWM Generation

To control the SCRs the PWM signals must fulfil certain criteria: 1) The PWM signals must be inverted (space comes before mark), 2) The frequency of the PWM must match exactly with the supply, 3) The phase information of the supply must be available and 4) The PWM frequencies must be twice the frequency of the AC supply.

An AVR microcontroller, ATmega32, was used for the purpose of PWM generation [5]. Even though it is possible to drive the two SCRs with the same PWM signal, a slight mismatch in timing can result in unwanted triggering in the next half cycle. Two PWMs working in an interleaved manner eliminates the risk of such triggering. Thus during the positive half cycle the PWM controlling the reverse SCR is set to be zero and during the negative half cycle the PWM controlling the forward SCR is set to zero. Fig. 6 shows the two PWM signal generated and the resultant output voltage alongside the supply voltage.

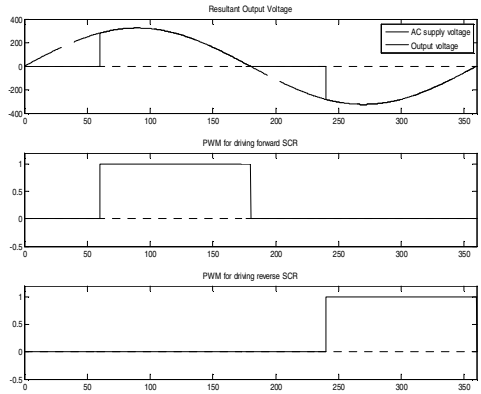


Fig. 6 PWM control signals

D. RMS Value Calculation:

The RMS value of the output cannot be determined by simply dividing the peak value by √2 due to the output voltage being a distorted sinusoidal wave [6]. In order to calculate the RMS value of the output voltage we have used another microcontroller, AVR ATmega8 [7]. This is due to the consideration that the burden of the excess calculation required for this step might hamper the time sensitive operation of the main microcontroller. If the output voltage is sampled over a full cycle of the supply voltage and N samples are obtained, the RMS value is calculated by the following equation.

$$V_{rms} = \sqrt{\frac{\sum_{i=1}^{N} V_i^2}{N}} \quad (2)$$

However since the analogue to digital converter of the AVR microcontroller only allows the conversion of positive voltages, the symmetric nature of the output is exploited and the output is sampled in one half cycle only. The result is then fed to the main microcontroller for the purpose of feedback.

E. Input/output Interface

For the provision of manual input to set the reference voltage value we have used a variable resistor. A 16x2 LCD module is interfaced for displaying various data related to the operation. The reference voltage value and the resultant output voltage are displayed on the LCD display. The frequency of

the supply voltage respective to the clock of the microcontroller is also displayed.

IV. Algorithm And Execution

When the device is started up it automatically measures the value of T_{total} and T_{error}. From Fig. 5 it is apparent that the rising edge of the synchronization pulse arrives at T_{error} time earlier than the end of the positive half cycle. Thus every rising edge means that there T_{error} time left until the negative clock cycle begins. If T_{PWM} is the time starting from the beginning of the PWM cycle then on every positive edge T_{PWM} should be equal to $(T_{total}/2-T_{error})$.

The two PWMs necessary are not generated separately. Rather, the same PWM is fed through two different pins of the microcontroller to deliver them to the respective SCR drive circuit at the appropriate time. After the end of each PWM cycle the pins are swapped. The flow chart of the synchronization process is shown in the Fig. 7.

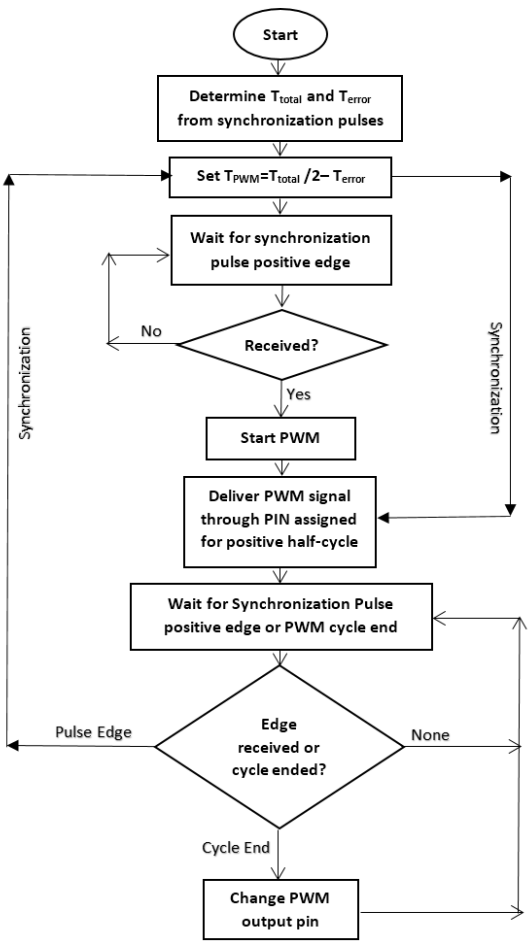


Fig. 7 Flow chart of re-synchronization

The target output voltage value is updated continuously. The target output voltage is determined by the manual input reference and the value of the output voltage obtained from feedback. A voltage error V_e(or V_{ep}) is calculated as,

$$V_e = V_{ref} - V_{out} \quad (3)$$

The error is not utilized directly rather a PID controller is implemented. The proportional, differential and the integral part of the error is calculated and each assigned a gain K_P, K_D and K_I respectively. The total error is also subjected to an overall gain of K. The new target output voltage is thus determined as,

$$V_{target} = V_{ref} + K * V_{etotal} \quad (4)$$

The value of the target firing angle required to achieve the particular value of RMS output voltage V_{target} is determined. For this purpose a look-up table of 64 data point is stored in the microcontroller. For any point that does not match with the desired value linear interpolation is used. The target duty cycle D is thus obtained from the target firing angle. Fig. 8 shows the flowchart for determining target duty cycle.

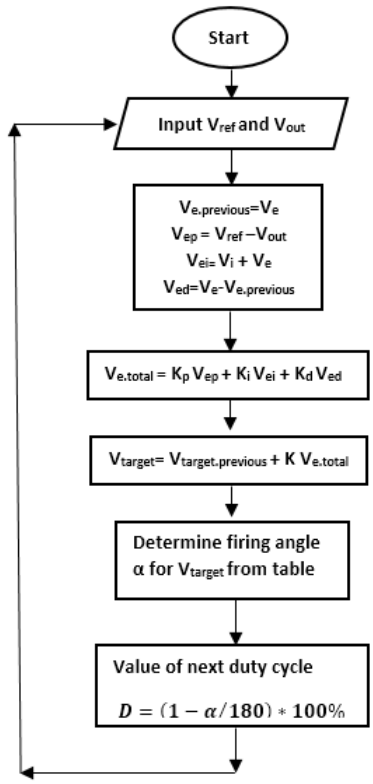


Fig. 8 Flow chart of duty cycle updating

V. Results and discussions

Theprototype was tested with pure resistive loads. Fig. 9 shows an input-output relationship where the input is given as the firing angle. The expected output is also plotted in the figure.

In Fig. 10 the output power is plotted with varying duty cycle. The relationship is non-linear in nature as the relationship between the voltage and the duty cycle. Due to almost a fixed amount of power loss of the operating circuit the efficiency of the device is very low when the power output is comparable to this small loss of power. The efficiency approaches unity as the load is increased and is in the range of any practical load. Fig. 11 shows the change of efficiency with change of output power.

Fig. 12, 13 and 14 shows the output power spectrum for 90˚, 120˚and 150˚ firing angles respectively. The contribution of harmonics towards the total power of the output is visibly high for the last two firing angles. So the preferred zone of operation should be firing angles less than or equal 90˚.

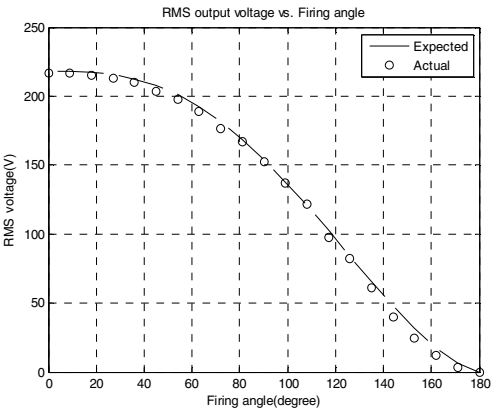


Fig. 9 RMS output voltage vs. firing angle

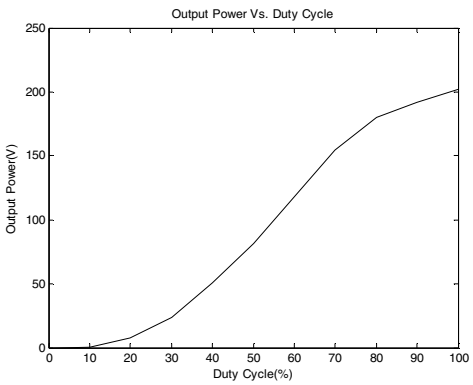


Fig. 10 Output power vs. duty cycle

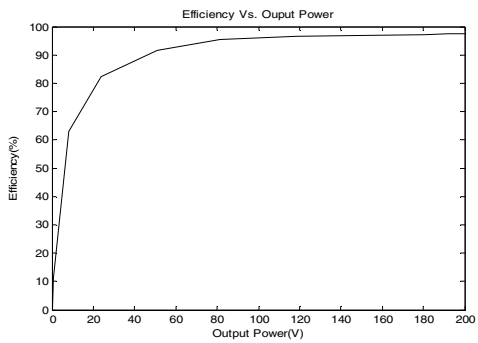


Fig. 11 Efficiency vs. output power

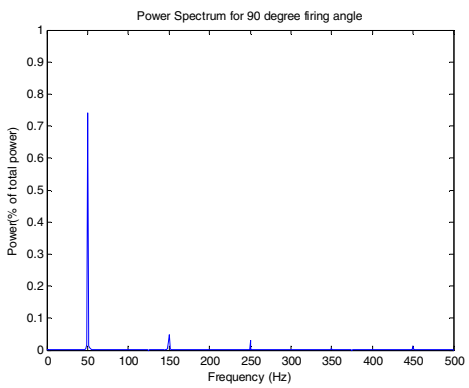


Fig. 12 Power spectrum for 90˚ firing angle

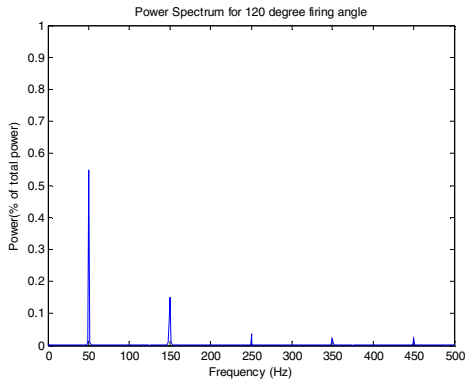


Fig. 13 Power spectrum for 120˚ firing angle

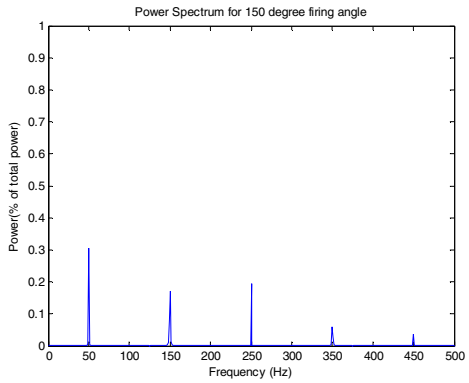


Fig. 14 Power spectrum for 150˚ firing angle

CONCLUSIONS

The feasibility and effectiveness of the proposed project is evaluated with simulation studies and practical real-time implementation. A microcontroller-based design can play a key role to increase its efficiency. Microcontroller based intelligent control allows the device to be more adaptive to different situations and is capable of responding to automation. Moreover, the device can be extended to work in a high power environment by simply replacing the SCR pair with a couple of high power SCRs. Filtering techniques could be applied to improve the design to reduce the harmonic content. However the presence of lower order harmonics makes the filtering difficult. Alternatively an adaptive range selector could be implemented that forces the firing angle to be less than 90^0 by reducing the input voltage by selecting from multiple transformers when necessary.

REFERENCES

[1] B. Paul, ”Industrial Electronics And Control”, PHI Learning Pvt. Ltd., Second Edition, 2009, pp.277-279.

[2] M. H. Rashid, "Power Electronics – Circuits, Devices, and Applications," Prenctice Hall Englewood Cliffs, Second Edition, 1993, pp.317-387.

[3] Nang, K.H. and L.O. Lwin, “Microcontroller based single phase automatic voltage regulator”, 3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT), Myanmar, 2010.

[4] MOC3021 Optoisolators, Texus Instruements, datasheet available at: http://www.ti.com/product/moc3021

[5] “Atmega32 Microcontroller data sheet”. Available: http://www.atmel.com/images/doc2503.pdf

[6] Guillermo Rico, “Tech Tip: Effective or RMS Voltage of a Sinusoid,” the Technology Interface, Spring 2006.

[7] “Atmega8 Microcontroller data sheet”. Available: http://www.atmel.com/devices/ATMEGA8.aspx

COMPARATIVE ANALYSIS OF CONTROL SCHEMES WITH HARMONIC CHARACTERISTICS FOR FUEL CELL SYSTEM

Md. Ashifur Rahman,[1,*] Mohammad Hasanuzzaman Shawon,[2] and Ikramul hasan Sohel[1]

[1] Department of Electrical and Electronic Engineering, Islamic University of Technology
Gazipur-1704, Bangladesh

[2] AGH University of Science and Technology
Krakow, Poland

[*]ashifur08@gmail.com

Abstract—**The concept of green and clean energy sources is the major concern for the researchers in the field of renewable and sustainable energy. Fuel cells are considered to be a promising source of electrical power but the stability of the output voltage is a question yet to be answered. In this paper a fuel cell generation model is shown as well as two stability control schemes (DVR and D-STATCOM) have been implemented in the time domain simulation of MATLAB/ SIMULINK. The comparison of these two schemes is also given for the analysis of the feasibility of the better scheme. The control schemes are analyzed mathematically and simulation is done to determine the stability of the system. A comprehensive analysis of harmonic characteristics is also shown for both the control schemes.**

Index Terms— green energy, fuel cells, DVR, D-STATCOM, MATLAB

I. INTRODUCTION

Power quality problems are of different types. One of the power quality problems is voltage sag that causes the voltage level instable in the transmission line [1]. Voltage sag can occur at any instant of time, with amplitudes ranging from 10%-90% and a duration for lasting half a cycle to one minute.

Another major problem is the fault case. These problems are minimized by DVR (dynamic voltage restorer), D-STATCOM (distribution static compensator) which are the most popular schemes to improve power system stability. These devices are based on voltage source converter principle.

Fuel cells are new sources of electrical energy. The energy produced from fuel cells is not stable. As it is a dc source it has to be converted to ac using inverter control strategy. Previously, fuel cell voltage converted to ac using inverter control strategy has been analyzed [2]. Power quality problem for this system has to be considered too. Symmetrical and Unsymmetrical faults may appear at any time and cause the instable condition which may make the system further worse. Previously, a comprehensive method for voltage stabilization scheme with DVR has been adopted [3]. By means of D-STATCOM a cost effective method is also approached. Separate Energy storage devices (SESD) based DVR is introduced and voltage sag is evaluated for a fuel cell system recently [4]. During different fault conditions how the fuel cell inverter system behaves also have been shown [5]. But the comparison of these two schemes for a fuel cell model has not been reported yet. In this paper, voltage sag is minimized using both DVR and D-STATCOM.

II. FUEL CELL MODEL

A. Principle

Fuel cells convert chemical energy of fuel into electrical energy. The fuel cell used for modeling is termed as Proton exchange membrane fuel cell (PEMFC). In fig. 1 the schematic diagram of PEM fuel cell is given. It is fed with hydrogen and air. There are few assumptions in the modeling. The electrode channels are assumed to be small enough that the pressure drop across them is negligible. The ratio of pressures between the inside and outside of the electrode channels is large enough to assume choked flow. The temperature of the fuel cell is stable. The gases are assumed to be ideal [6].

The typical anode and cathode reactions for a hydrogen fuel cell can be described as follows:

In the anode section, hydrogen gas oxidizes releasing electrons and creating H^+ ions.

Anode equation:

$$2H_2 = 4H^+ + 4e^- \tag{1}$$

At the cathode, oxygen reacts with electrons from the electrodes and H^+ ions from the electrolyte and form water. Here water is the waste product and taken out from the cell.

Cathode equation:

$$O_2 + 4e^- + 4H^+ = 2H_2O \tag{2}$$

$$O_2 + 2H_2 = 2H_2O + heat + electricity \tag{3}$$

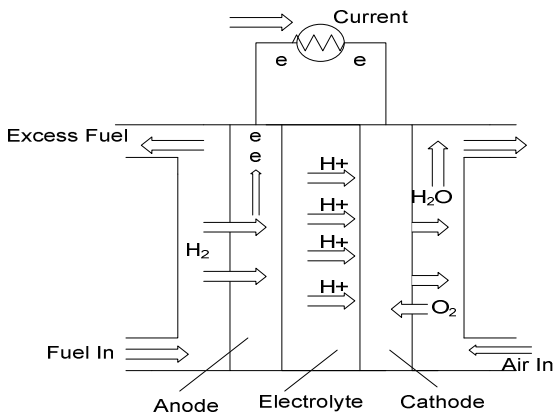


Fig. 1. Fuel Cell schematic diagram

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

B. Fuel Cell Model

The simplified model represents a particular fuel cell stack operating at nominal conditions of temperature and pressure. The parameters of the equivalent circuit can be modified based on the polarization curve obtained from the manufacturer datasheet. The tuning parameters of the simplified model are the value of the voltage at 0 and 1 A, the nominal and the maximum operating points. A diode is used to prevent the flow of negative current into the stack.

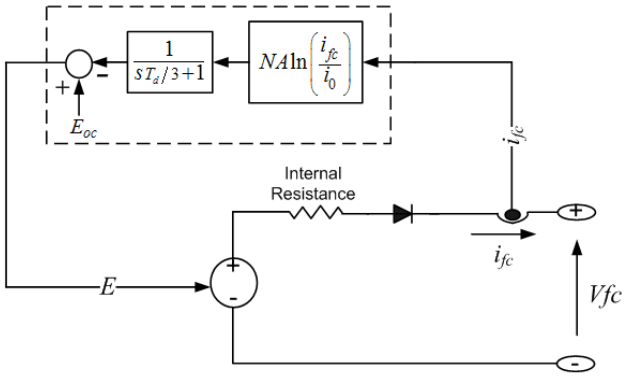


Fig. 2. Simplified model of a PEM Fuel Cell

The detailed model represents a particular fuel cell stack when the parameters such as pressures, temperature, compositions and flow rates of fuel and air vary. voltage (E_{oc}), the exchange current (i_0) and the Tafel slope (A). E_{oc}, i_0 and A are modified as follows:

$$E_{oc}=K_c E_n \tag{4}$$

$$i_0=\frac{zFk\,(p_{H_2}+p_{O_2})}{Rh}e^{\frac{-\Delta G}{RT}} \tag{5}$$

$$A=\frac{RT}{z\alpha F} \tag{6}$$

Where

R = 8.3145 J/(mol K)

F = 96485 A s/mol

z = Number of moving electrons

E_n = Nernst voltage, which is the thermodynamics voltage of the cells and depends on the temperatures and partial pressures of reactants and products inside the stack (V)

α = Charge transfer coefficient, which depends on the type of electrodes and catalysts used

P_{H2} = Partial pressure of hydrogen inside the stack (atm)

P_{O2} = Partial pressure of oxygen inside the stack (atm)

k = Boltzmann's constant = 1.38×10^{-23} J/K

h = Planck's constant = 6.626×10^{-34} J s

ΔG = Size of the activation barrier which depends on the type of electrode and catalyst used

T = Temperature of operation (K)

K_c = Voltage constant at nominal condition of operation

III. DVR Control Strategy

The basic structure of a DVR (Dynamic voltage restorer) is shown in Fig.3.

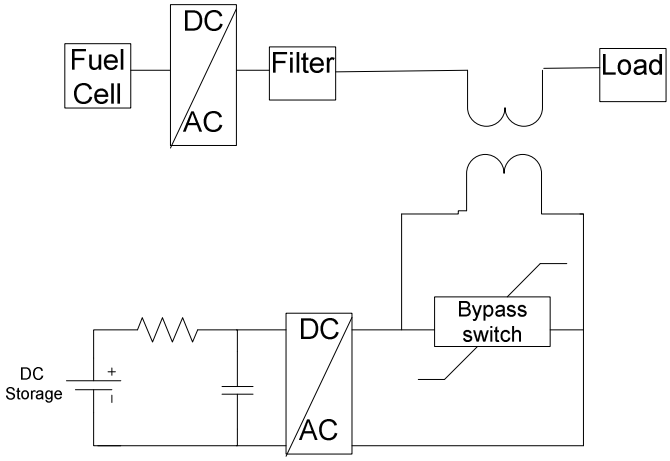


Fig.3. Basic Structure of Dynamic Voltage Restorer

Voltage injection transformers are used in DVR for voltage injection purpose. In a three-phase system, either three single-phase transformer units or one three phase transformer unit can be used. DVR has a large DC capacitor to ensure stiff DC voltage input to inverter. An Inverter system is used to convert dc storage into ac form [7]. Voltage source inverter (VSI) of low voltage and high current with step up injection transformer is used for this purpose in the DVR compensation technique. Filters are used to convert the inverted PWM waveform into a sinusoidal waveform. This is achieved by eliminating the unwanted harmonic components generated VSI action. Higher orders harmonic components distort the compensated output voltage. By-pass switch is used to protect the inverter from high currents in the presence of faulty conditions. Energy storage unit is responsible for energy storage in DC form. Flywheels, batteries, superconducting magnetic energy storage (SMES) and super capacitors can be used as energy storage devices.

A. Simulation result

In test case scenario, voltage is maintained at 1 per unit during stable operation. When the fault is applied voltage drops in case of system without DVR. But the system with DVR shows stable operation during this fault. So fuel cell voltage stability is enhanced using DVR strategy.

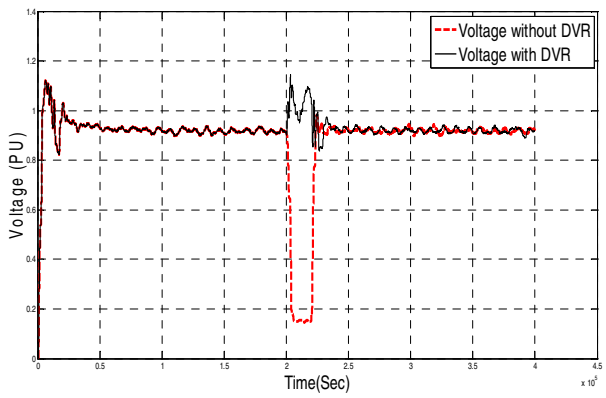


Fig. 4. Per Unit Fuel cell output voltage (ac)

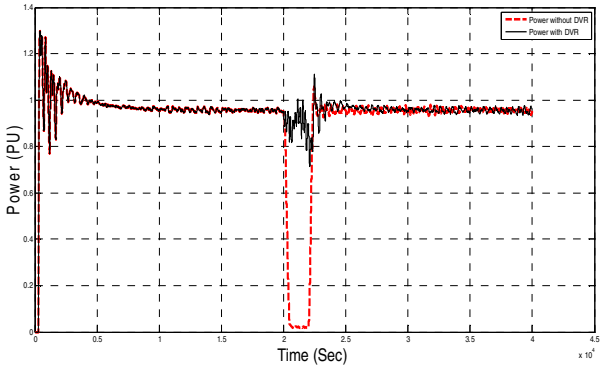


Fig. 5. Per Unit Fuel cell output power (ac)

IV. D-STATCOM CONTROL STRATEGY

A D-STATCOM (Distribution Static Compensator) consists of a dc energy storage unit, a two stage voltage source converter, a coupling transformer connected in shunt with the distribution network. The voltage source converter is essentially a dc to ac converter which converts the dc voltage of the energy storage device to three phase ac voltage. These ac voltages are in phase and coupled with the ac system through a coupling transformer [8].

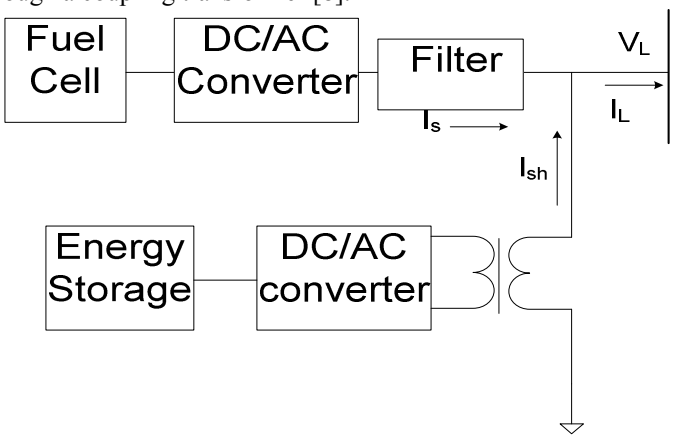


Fig. 6. D-STATCOM acting on the ac system

The shunt current I_{sh} corrects the voltage sags by adjusting the voltage drop across the system impedance Z_{th}. The shunt current in fig. 6 can be written as-

$$I_{sh} = I_L - I_s = \frac{V_{th} - V_L}{Z_{th}} \tag{7}$$

$$I_{sh}\angle\eta = I_L\angle -\theta - \frac{V_{th}}{Z_{th}}\angle(\partial - \beta) + \frac{V_L}{Z_{th}}\angle -\beta \tag{8}$$

The complex power injected from the D-STATCOM can be written as-

$$S_{sh} = V_L I_{sh}^* \tag{9}$$

The control scheme for the D-STATCOM follows the same principle as for DVR. When the shunt current I_{sh} is kept in quadrature with V_L, the desired voltage correction can be achieved without injecting any active power into the system. On the other hand, when the value of I_{sh} is minimized, the same voltage correction can be achieved with minimum apparent power injection into the system [8].

B. Simulation result

In test case scenario, voltage is maintained at 1 pu during stable operation. When the fault is applied the voltage drops in case of system without D-STATCOM. But the system with D-STATCOM shows stable operation during this fault. Although there exist some fluctuations in voltage level due to circuit breaker operation, it can be reduced using other protection schemes. So fuel cell voltage stability is enhanced using D-STATCOM strategy. In each case voltage and power recovery during fault condition using D-STATCOM are demonstrated by fig 7-10 respectively.

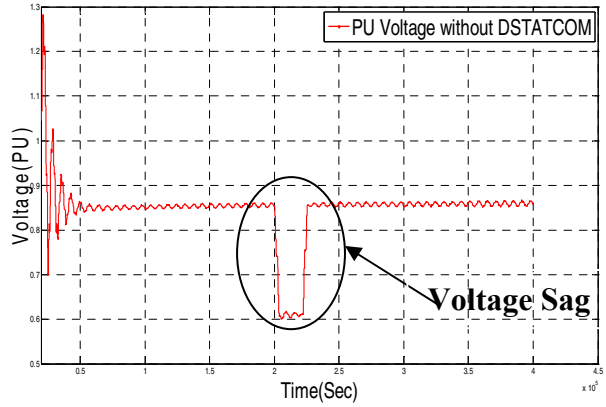


Fig. 7. Per Unit Fuel cell output voltage (ac) without D-STATCOM

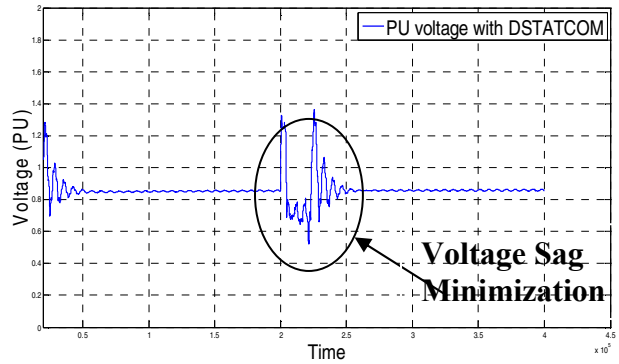


Fig. 8. Per Unit Fuel cell output voltage (ac) with D-STATCOM

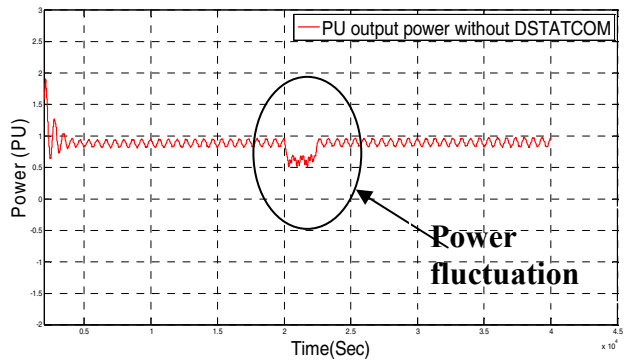


Fig. 9. Per Unit Fuel cell output power without D-STATCOM

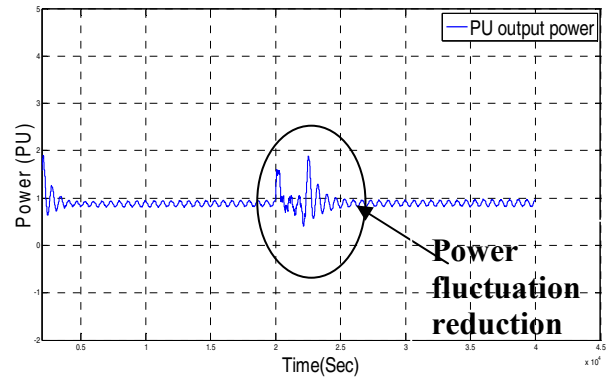


Fig. 10. Per Unit Fuel cell output power with D-STATCOM

V. COMPARATIVE HARMONIC ANALYSIS

THD is defined as the RMS value of the waveform remaining when the fundamental is removed. A perfect sine wave is 100%. Harmonic distortion is caused by the introduction of waveforms at frequencies in multiplies of the fundamental. If the fundamental frequency is 50 Hz then the 5th harmonic is 5x the fundamental frequency =250 Hz. Total harmonic distortion is the measure of the closeness to the sine wave. THD (total harmonic distortion) can be expressed by the following equation [9]

$$THD = \frac{\left[\sum (\text{Amplitudes of All Harmonics})^2\right]^{\frac{1}{2}}}{\text{Amplitude of Fundamental Harmonic}} \quad (10)$$

So, harmonic distortion is simply the ratio of the amplitude of the harmonic to the fundamental one.

At different DVR voltage the THD is calculated in harmonics order. At 1KV DVR dc storage voltage the THD is found to be 19.96%. For an increased dc storage voltage THD was also increased accordingly. The following figure shows the THD for DVR at 1KV.

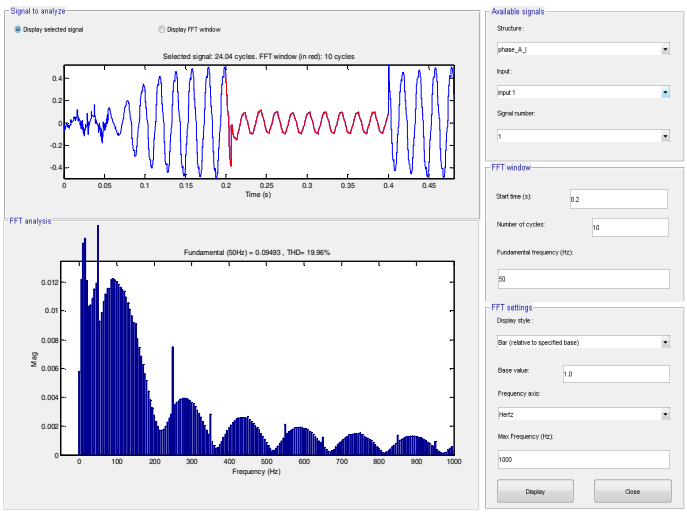


Fig.11. THD in Harmonics Order for DVR

For D-STATCOM the same procedure is followed. In case of D-STATCOM THD is found to be 15.53%. From the harmonic analysis it is shown that better THD has been found when D-STATCOM is used.

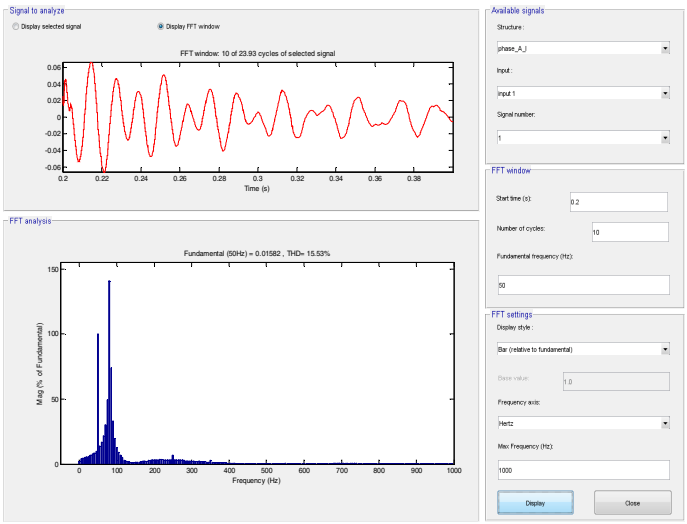


Fig.12. THD in Harmonics Order for D-STATCOM

THD is to be decreased to a certain level in order to decrease the pick current hence decreasing the power loss of the system. This is an important factor to be considered. LCL filter generally decrease the THD which is a scope of the future work.

VI. CONCLUSIONS

Both DVR and D-STATCOM are very effective stabilization schemes to reduce the voltage sags, voltage dips etc. The reduction of voltage sags and power injection to the system depends on the dc rating of the energy storage devices for both DVR and D-STATCOM. But from the simulation it is seen that D-STATCOM requires more active power injection to the system than DVR. D-STATCOM is also more costly stabilization scheme than DVR. DVR has more capabilities to regulate the voltage. There are other methods in FACTS devices such as FFS fundamental frequency switching methods. But PWM based method is better than these methods. Among the PWM methods, performance of DVR is found to be more superior to D-STATCOM considering all the facts.

REFERENCES

[1] Rioji Anahara, Sum10 Yokokawa and Masahiro Sakurai "Present Status and Future Prospects for Fuel Cell Power Systems" Proceedings of The IEEE, March 1993, Vol. 81, No. 3.

[2] Md. Ashifur Rahman, Mohammad Hasanuzzaman Shawon, Subarto Kumar Ghosh, Md. Mostafizur Rahman, "Analysis of Fuel Cell System including Inverter Control Strategy ," International Journal of Scientific & Engineering Research, Volume 4, Issue 6, June-2013 ISSN 2229-5518.

[3] H.P. Tiwari, Sunil Kumar Gupta "DVR Based On Fuel Cell: An Innovative Back-Up System," International Journal of Environmental Science and Development, Vol. 1, No. 1, April 2010 ISSN: 2010-0264.

[4] R.Vijayakumar, R.Subramanian, " Compensation of Voltage Variations In Distribution System by Using DVR Based Separate Energy Storage Devices," International Electrical Engineering Journal (IEEJ) Vol. 4 (2013) No. 1, pp. 1017-1026, ISSN 2078-2365

[5] Md. Ashifur Rahman, Mohammad Hasanuzzaman Shawon, Md. Mostafizur Rahman, Md. Shafayet Hossain, "Transient Stability Analysis of Grid Connected Fuel Cell System", European Scientific Journal (ESJ), June edition vol.9, No.16.

[6] Kanhu Charan Bhuyan, Sumit Kumar Sao, Prof. Kamalakanta Mahapatra, "Fuel Cell connected to Grid through Inverter."

[7] Sunil Kumar Gupta, H.P. Tiwari, Ramesh Pachar, " Estimation of DC Voltage Storage Requirements for Dynamic Voltage Compensation on Distribution Network using DVR," IACSIT International Journal of Engineering and Technology, Vol. 2, No. 1, February 2010, ISSN: 1793-8236

[8] S.V Ravi Kumar, S. Siva Nagaraju "Simulation Of D-STATCOM And DVR In Power Systems," ARPN Journal of Engineering and Applied Sciences, VOL. 2, NO. 3, JUNE 2007, ISSN 1819-6608.

[9] Sunil Kumar gupta, H.P. Tiwari, Rajesh Kumar, "Analysis of total harmonics distortion in DVR compensated system," Journal of Electrical Engineering.

[10] Md. Ashifur Rahman, Mohammad Hasanuzzaman Shawon, Subarto Kumar Ghosh, Md. Ariful Hoque Sojib, "Stability Enhancement of Fuel Cell Generation System Including DVR", International Conference on Informatics, Electronics & Vision (ICIEV), 2014, Print ISBN: 978-1-4799-5179-6, page 1 – 5.

Study of Vienna Rectifier and a Highly Efficient Single Phase Two Stage Inverter with low THD

Srabanty Ahmed Shaon, K. M. A. Salam

Department of Electrical and Computer Engineering
North South University
Dhaka, Bangladesh
srabanty_shaon@yahoo.com; asalam@northsouth.edu

***Abstract*—This study proposes a new circuit consisting of a Vienna rectifier and a single phase two stage inverter. The inverter consists with a combination of sinusoidal pulse width modulation (SPWM) and a square wave along with grid synchronization condition. The simulation work of this circuit has been done using PSIM. The new technique revealed that the low ripple, low THD below 0.1%, low cost and simple control providing a improved power factor of above 0.94.**

Keywords—Vienna Rectifier, Single phase two stage Inverter, Power Factor, Power electronics.

I. INTRODUCTION

The process of increasing the power factor without changing the loads or altering the voltage or the current to the original load is known as power factor improvement [1]. Among the rectifier topologies, VENNA rectifier[2-4] is an attractive choice for power factor improvement because its switch voltage stress is one half of the total output voltage so that fast switches such as MOSFETs can be used.The Vienna rectifier is useful wherever six switch converters are used for achieving sinusoidal mains current and controlled output voltage, when no energy feedback from load into the mains is required. The Vienna rectifier comprises a semiconductor switch in each phase leg of 1- phase Diode Bridge. It is a highly efficient method of high current, single-phase AC/DC conversion and is very useful for achieving unity power factor correction.[5].There are many processes to improve power factor. Vienna rectifier and Buck-Boost converter can improve the power factor of a three phase circuit [6]. Harmonics have a negative effect in theoperation of the electrical system and therefore, anincreasing attention is paid to their generation and control [7-8]. So, the study of Total Harmonic Distortion (THD)is important. Single phase two stage highly efficient inverter provides low harmonics, low cost, high efficiency, low THD and simple control [9]. To improve efficiency and noise performance, we have proposed the single phase two stage inverter instead of using Buck-Boost Converter.

II. VIENNA RECTIFIER

VIENNA rectifier is a three-switch rectifier which gives the DC output given in Fig.1. For controlling the rectifier it only need three switches rather five floating switches of other rectifiers. This switching system gives it more convenience to implement.

But in comparing this with single switch controlling is still complex. But the input current distortion of this rectifier is about 8.2% which is significantly less than the single switching and also than H-bridge and two switch implementation [10]. Basically the VIENNA rectifier works as a two-switch boost rectifier. One of the switches works at line frequency and two switches switched at high frequency. In 60° control block one switch is permanently on.

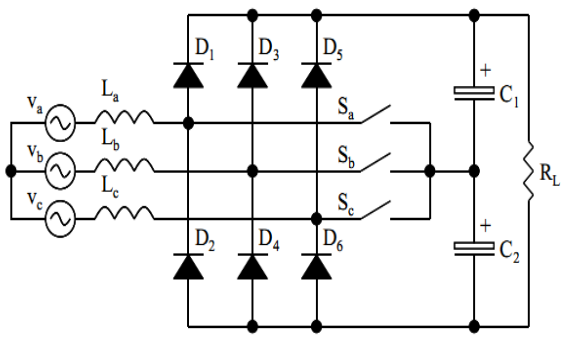


Fig.1. Vienna Rectifier

VIENNA rectifier can be explained as two independent boost rectifiers, one for boosting C1 and the other for boosting C2. Thus it can be seen that the minimum boost voltage over C1 and C2 will be the maximum line-to-line voltage of the input.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

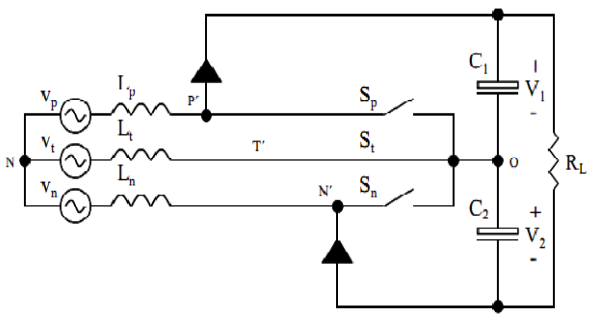


Fig.2. Equivalent circuit of VIENNA rectifier (60° control block) [11]

The equivalent representation of VIENNA rectifier for a 60° control block in one switch "on" condition is shown in the Fig.2. VIENNA rectifier has lower switch and diode currents than all of the other dual-boost rectifiers [12].For switch losses and diode losses it has the same harmonic distortion like others. But an extra convenience of the VIENNA rectifier is that modules are available having all the power stage bridge leg p, n and t-subscript denotes parts associated with the positive rail, parts associated with the negative rail and parts associated with a transitional period respectively. The small-signal model is derived for a negative duty cycle which is larger than the positive duty cycle, $d_N>d_P$. Then the process is repeated for the positive duty cycle which is larger than the negative duty cycle, $d_P>d_N$. The common parts of two switch cycle off periods $(1-d_N)$ and
$(1-d_P)$ can be associated.

III. DESIGN AND SIMULATION OF THE INVERTER:

This inverter power circuit design consists of a pair of two parallel MOSFET gates [9]. In conventional inverters only one type of switching technique is used. This design instead uses a combination of SPWM and square wave to reduce the switching loss by reducing the switching frequency.The sine wave is sampled from power grid by using transformer that steps down voltage from 220V to 5V.A high frequency triangle wave of 10 KHz frequency is used. Then the two signals are passed through a comparator to produce the SPWM signal.A square wave signal is used as the line frequency and is in phase with the SPWM.The square wave is transmitted through a NOT gate to produce a signal that is 180 degree out of phase of the original signal. The inverter requires four switching signals since it has four MOSFET switches [9].To create four signals an AND operation is performed between two sets of square wave signals and the SPWM. The four sets of switching signals can be categorizedin two groups. The first group contains MOSFETs Q_1 & Q_2while the second group contains MOSFETs Q_3 & Q_4. The gatepulses for switching of MOSFETs.

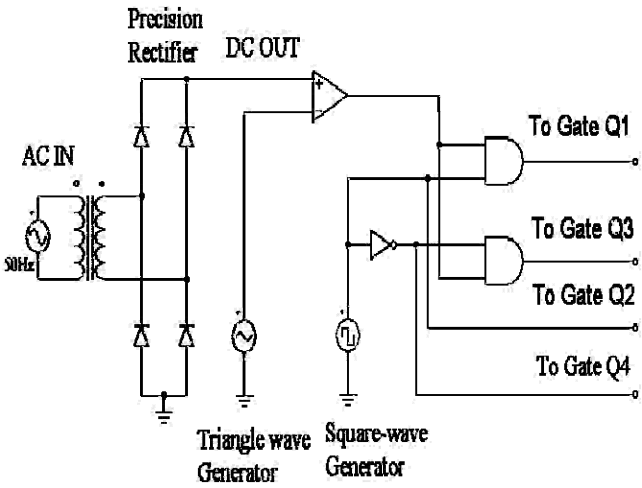


Fig.3.Control circuit of a single phase two stage inverter design

IV. SYSTEM MODEL

Fig.4 represents the model of a R-L load using single phase Vienna rectifier and single phase two stage highly efficient inverter.The Vienna Rectifier in which the input 30V p-p ac voltage is rectified dc voltage. The MOSFET are triggered using a Square pulse of 30V. The single phase two stage inverter, where the MOSFETs are triggered using a triangular pulse of 12V.We also use a L-C filter to minimize the distortion. The last part of this circuit is the R-L load.

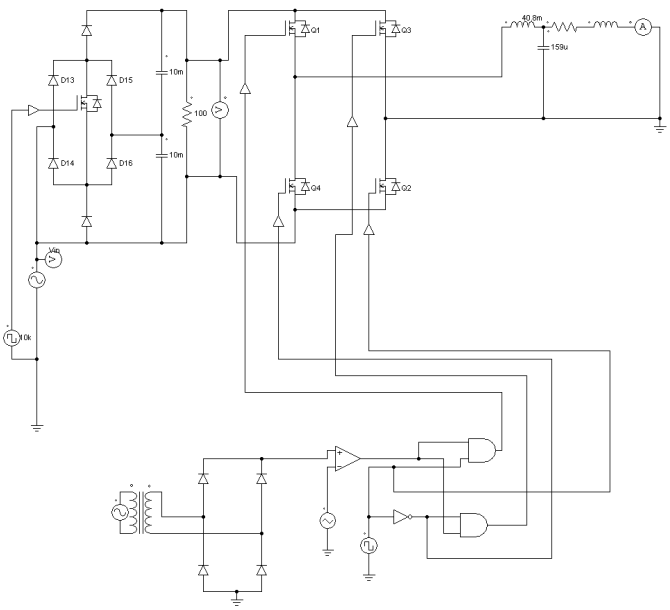


Fig.4. Schematic diagram of a Vienna rectifier connected to a single phase two stage highly efficient inverter for simulation in PSIM.

V. SIMULATION RESULT

Fig.5 shows the input voltage which is 30v p-p. Fig.6 exposes the output of Vienna.It revealed that the output is distorted and is difficult to measure the voltage. Fig.6and Fig.7showthe triggering pulses of Inverter and Vienna rectifier respectively. Fig.9 shows output current of the inverter. It contains ripple. To minimize this distortion we use a L-C filter. Then we get a almost ripple free current curve of the inverter in Fig.10. Same scenery has been achieved in the case of voltage. Fig.11 reveals a distorted voltage of the inverter before using filter while a almost pure sinusoidal voltage curve is achieved in Fig.12 after using the L-C filter. L-C filter is available in market at low cost. So, it is cost effective. L-C filter can be handled easily. So, it also provides simple control of the circuit.

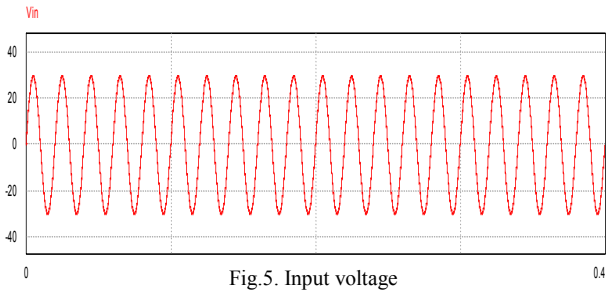


Fig.5. Input voltage

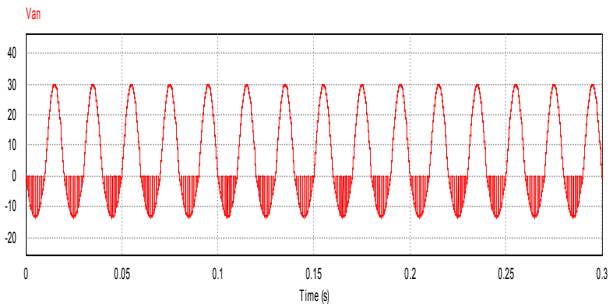


Fig.6. Output of Vienna rectifier

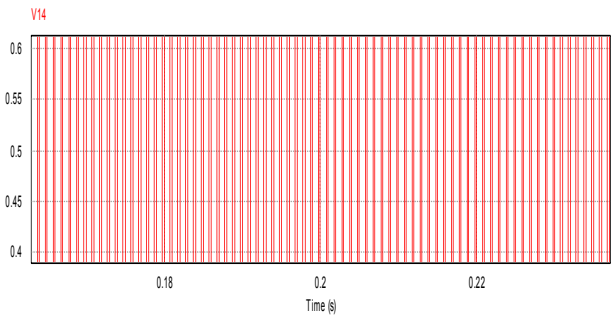


Fig.7. MOSFET's triggering pulse of the Inverter

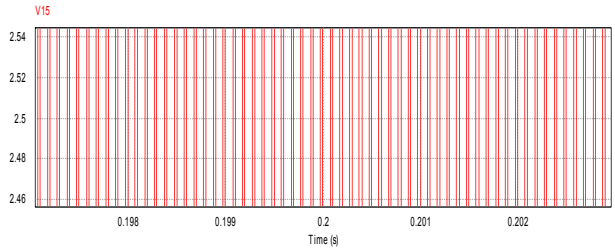


Fig.8. Triggering pulse of Vienna rectifier

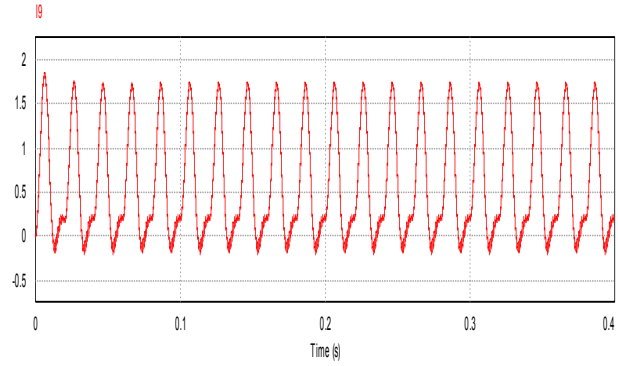


Fig.9. Output current of single phase two stage Inverter before filtering.

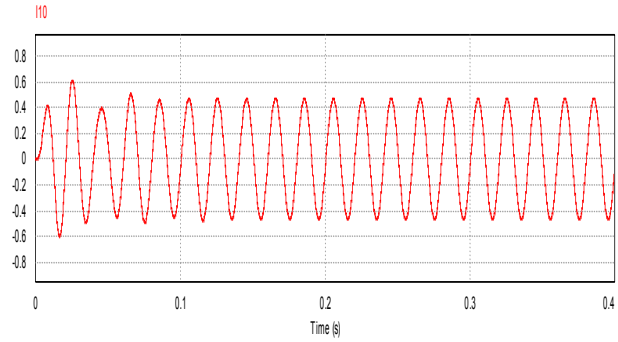


Fig.10. Output current of the inverter after filtering.

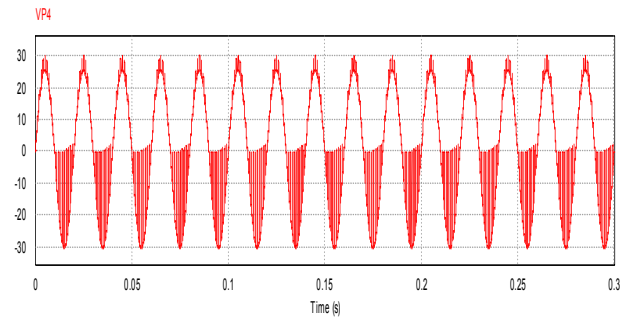


Fig.11. Output voltage of the single phase two stageInverter before filtering

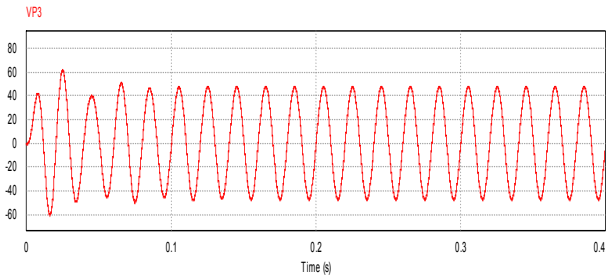


Fig.12. Output voltage of the inverter after using filter

VI. PERFORMANCE ANALYSIS

A vienna rectifier is useful to improve the power factor. The MOSFET switch controls the current by controlling the magnetization of the inductor. Switched on charges in the inductor, drives the current through the bidirectional switch. Deactivating the switch increases the current to bypass the switch and flow through the freewheeling diodes. This results in a negative voltage across the inductor and drains it. This describes the ability of the topology to control the current in phase with the mains voltage (Power Factor Improvement capability). Here a vienna rectifier is also used. So, an improved power factor of 0.9411 has been achieved. The output current and voltage of the single phase two stage inverter have been pure sinusoidal after using L-C filter. Before that there was ripple. It is also observed that the output power is sent at an angle of 45 degrees with total harmonic distortion of less than 0.1%.

VII. CONCLUSION

In this paper we have studied the model of a R-L load achieving low THD, low cost, improved power factor and a simple control. To fulfill the purpose we used a Vienna rectifier and a single phase two stage inverter. By analyzing the simulation result we got a power factor of above 0.94.The simulation result also shows that the Inverter produces a high quality ripple free sinusoidal AC curve with a voltage THD below 0.1% which is very much lower than IEEE 519 standard. So, the new circuit providing low cost, simple control, low THD, low ripple and improved power factormight be the application of modern power electronic system.

REFERENCES

[1] Malinowski, M., Jasinski, M. and Kazmierkowski, M. P.(2004) "Simple direct power control of three-phase PWM rectifier using space-vector modulation (DPC-SVM)", *IEEE Trans. Ind.Electron.*,vol. 51, no. 2, pp. 447–454.

[2] Kolar, J.W, Drofenik.; Zach, F.C. "Current handling capability of the neutral point of a three-phase/switch/level boost-type PWM(VIENNA)rectifier". PESC 96 Record. p.1329- 36 vol.2.

[3] Kolar, J.W; Zach, F.C. "A novel three-phase utility interfaceminimizing line current harmonics of high power telecommuncationsrectifiers modules", IEEE Trans on Industrial Electronics, vol. 44,IEEE, Aug. 1997. P.456-67.

[4] Kolar, J.W, Ertl, H; Zach, F.C. "Design and experimental investigationof a three-phase high power density high efficiency unity power-factorPWM (VIENNA) rectifier employing a novel integrated power semiconductor module".APEC'96, San Jose, CA. IEEE, 1996, p.514-23, vol.2.

[5] 'Single phase neutral linked Vienna rectifier: A solution for the mitigation of voltage sag in ASD fed induction motor' SelvarajA.,Paranjothi. S. R. and Jagadeesh B. Department of EEE, Rajalakshmi Engineering College, Chennai, India,VOL. 7,NO. 7, JULY 2012 ISSN 1819-6608ARPN Journal of Engineering and Applied Sciences.

[6] Navila Rahman Nadi, Srabanty Ahmed Shaon-'Power Factor Improvement usnig Vienna Rectifier and Buck Boost Converter', International Journal of Scientific & Engineering Research,Volume 4,Issue 7, July-2013 ISSN 2229-5518.

[7] J. Arrillaga, N. Watson "Power System Harmonics," John Wiley & Sons. Inc, Second Edition, 2003, ISBN: 0470851295.

[8] D. Paice "Power Electronic Converter Harmonics -Multipulse Methods for Clean Power," IEEE Press, Second Edition, 1996, ISBN: 078031137X.

[9] Ahmed Sony Kamal Chowdhury, M. Shamir Shehab, M. Abdul Awal, M. Abdur Razzak. "Single phase grid-connected photovoltaic inverter for residential application with maximum power point tracking",Informatics, Electronics & Vision (ICIEV), 2013 International Conference on date17-18 May 2013.

[10] Johann W. Kolm, Hans Ertl, ―Status of the Techniques of Three-phase Rectifier Systems with Low Effects on the Mains‖, 21st INTELEC, June 6-9, 1999, Copenhagen, Denmark .

[11] Chongming Qiao, Keyue M. Smedley, ―Three-Phase Unity-Power-Factor Star-Connected Switch (VIENNA) Rectifier With Unified Constant-Frequency Integration Control‖,IEEE transactions on power electronics, vol.18, no.4, July 2003.

[12] Chongming Qiao, Keyue M. Smedley, ―A General Three- Phase PFC Controller for Rectifiers With a Parallel-Connected Dual Boost Topology‖, IEEE transactions on power electronics, vol.17, no.6, November 2002.

Split Inductor Based Neutral Point Clamped Inverter using Improved Modulation Technique to Reduce Common Mode Voltage

Md. Jannatul Ferdous,[1,*] Shariar kabir,[2] Md. Rifat Alam Siddique[3] and Umme Salma [4]

[1,3]Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology
Dhaka-1205, Bangladesh

[2]Department of Electrical and Electronic Engineering, United International University

[4]Department of Electrical and Electronic Engineering, Chittagong University of Engineering and Technology

[*]j.ferdous63@yahoo.com

Abstract—**Due to high switching frequency and modulation strategy, neutral point clamped (NPC) inverter generates varying common mode voltage as well as leakage current in PV and wind turbine system. This leakage current causes electrical hazard, THD at the output voltage, EMI and ultimate pollution at the output power. To reduce this varying common mode voltage and leakage current, this paper proposes a Split Inductor based Neutral Point Clamped (SINPC) inverter with improved modulation technique. The proposed inverter can reduce common mode voltage and suppress leakage current comprehensively. In order to verify the effectiveness of the SINPC inverter with improved modulation, simulation has been carried out in PSIM software. Finally, the output wave shape of reduced common mode voltage and leakage current ensures the effectivity of the proposed circuit.**

***Index Terms*— Split Inductor NPC, common mode voltage, leakage current.**

I. Introduction

Voltage source inverters are widely used in renewable power processing stage like in photovoltaic (PV) system, wind turbine, fuel cell etc. Pulse width modulation (PWM), space vector modulation (SVM) etc. is the standard approaches to operate the inverter switches in order to generate the required high quality output voltage. But the switching frequency of the inverter in these techniques is usually very high. So due to this high switching frequency, some problems have been occurred in inverters connected with the PV systems or wind turbine. Among them, leakage current generation is one of the frequent problems usually observed in transformer less PV systems [1]-[4]. Due to the absence of the transformer, there is a galvanic connection between power grid and solar cell array. Not only that, there are stray capacitors to ground formed by the surface of the PV array. So when high frequency switching occurs, a varying common mode voltage will be generated in this system which creates a dangerous leakage current through the stray capacitance between the PV array and the ground. This leakage current cause's electrical hazard increases the total harmonic distortion (THD) and ultimate pollutes the output power of the inverter circuit. This problem not only occurs in transformer less PV system but also in DFIG based wind turbine where back-to-back PWM converter is used [3]-[7]. So our main target is to reduce this varying common mode voltage and leakage current which is generated during high frequency switching of the neutral point clamped (NPC) inverter.

In NPC inverter, Space Vector Modulation (SVM) is usually used to generate the output voltage. In this technique, a varying common mode voltage levels are being generated at different switching stages. If some of the harmful switching stages can be eliminated during the modulation period, then a minimize common mode voltage as well as leakage current can be achieved. In [4], some effort has been given to do that but in this work, the technique is implemented in a conventional NPC Inverter. So the varying common mode voltage still bears a very high magnitude (more than 500V peak to peak for 1200V DC input). In our paper we proposed a split Inductor based Neutral point clamped (SINPC) Inverter instead of conventional NPC inverter which reduces the varying common mode voltage comprehensively at the improved modulation technique. In addition to that, split inductor concept of the proposed inverter generates low leakage current in the system. In order to prove the effectiveness of proposed circuit, simulation has been carried out which has been demonstrated at the simulation part.

II. Common mode voltage And circuit Analysis

Fig.1 (a) shows the conventional NPC inverter and Fig 1(b) shows the proposed SINPC inverter for reduced common mode voltage at improved modulation Technique. L_{1a}, L_{1b},L_{2a},L_{2b},L_{3a} and L_{3b} are the split inductors in Fig.1(b).

From the figure, we can express the common mode voltage in mathematical form. The instantaneous common mode voltage -

$$V_{CM} = \frac{V_{ao} + V_{bo} + V_{co}}{3} \tag{1}$$

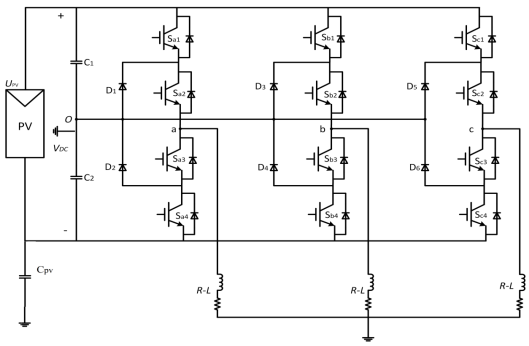
Where V_{ao},V_{bo},V_{co} are the instantaneous leg voltages of the inverter.

This varying common mode voltage creates a leakage current which flows through the capacitive element in a transformer less PV systems. For equal inductance, the magnitude of the leakage current is calculated as follows-

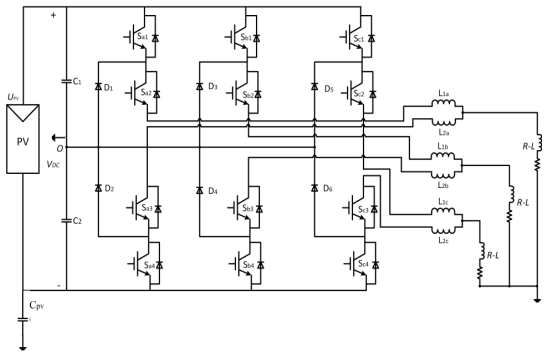
$$I_{leakagecurrent} = \frac{V_{CM}}{Z} \tag{2}$$

Where Z is the combined impedance of C_{PV} and equivalent inductance of the circuit. The equivalent common mode model is described in[3].

978-1-4799-4166-7/14/$31.00 ©2014 IEEE



(a)



(b)

Fig.1 (a) Conventional NPC inverter (b) Proposed SINPC inverter for improve modulation

III. Improved Modulation Technique

A very common SVM technique for three-level NPC inverter generates 27 switching states which is illustrated in Fig 3 and sum up at TABLE I [4]. In the table, a three digit number represents the switching stages of leg a, b and c. "0" means two upper switches of the leg is tuned off , "-1" means two lower switches are turned off, and "1" means first upper switch and last lower switch is off. It is evident from the table that, applying these vector generates $\pm V_{dc}/6$, $\pm V_{dc}/3$, $\pm V_{dc}/2$,0 common mode voltage which is varying in nature. The switching vectors which have been divided in to six two-level hexagons is shown in Fig 3.It is interesting in hexagon's figure that in each sector two or three vectors generating the same reference vector but leads to different common mode voltage. Like, $\pm V_{dc}/2(V_{14},V_{16},V_{22},V_4,V_{10},V_{12})$ or $\pm V_{dc}/3(V_1,V_3,V_9,V_{17},V_{23},V_{25})$ are the two vectors set creating the same reference vector but generating different common mode voltage. In order to minimize the variation in common mode voltage, we used the first switching vectors set in switching pattern. In addition to that V_0, V_{26} are the Zero switching vectors which produce $\pm V_{dc}/2$ common mode voltage but V_{13} is the another one which produce zero common mode voltage. We used V_{13} as zero switching vector. After eliminating these vectors, the six hexagon switching patterns now produce the common mode voltage only zero and $\pm V_{dc}/6$. So these help to reduce the variation of common mode voltage as well as leakage current. Now the switching patterns according to the improved modulation technique (which is discussed) are summarised in Fig.2 from Hex1 to Hex6.

Hex 1

Sector	Switching pattern
Sector 1	-111,-110,-100,-110,-111
Sector 2	-111,-110,110,-110,-111
Sector 3	-111,111,110,111,-111
Sector 4	-111,111,101,111,-111
Sector 5	-111,-101,101,-101,-111
Sector 6	-111,-101,-100,-101,-111

Hex 2

Sector	Switching pattern
Sector 1	110,-1-10,-110,-1-10,110
Sector 2	110,-1-10,1-10,-1-10,110
Sector 3	110,1-11,1-10,1-11,110
Sector 4	110,1-11,111,1-11,110
Sector 5	110,-111,111,-111,110
Sector 6	110,-111,-110,-111,110

Hex 3

Sector	Switching pattern
Sector 1	1-11,1-10,110,1-10,1-11
Sector2	1-11,1-10,0-10,1-10,1-11
Sector 3	1-11,0-11,0-10,0-11,1-11
Sector 4	1-11,0-11,011,0-11,1-11
Sector 5	1-11,111,011,111,1-11
Sector 6	1-11,111,110,111,1-11

Hex 4

Sector	Switching pattern
Sector 1	011,1-11,111,1-11,011
Sector 2	011,1-11,0-11,1-11,011
Sector 3	011,0-1-1,0-11,0-1-1,011
Sector 4	011,0-1-1,01-1,0-1-1,011
Sector 5	011,11-1,01-1,11-1,011
Sector 6	011,11-1,111,11-1,011

Hex 5

Sector	Switching pattern
Sector 1	11-1,111,101,111,11-1
Sector 2	11-1,111,011,111,11-1
Sector 3	11-1,01-1,011,01-1,11-1
Sector 4	11-1,01-1,00-1,01-1,11-1
Sector 5	11-1,10-1,00-1,10-1,11-1
Sector 6	11-1,10-1,101,10-1,11-1

Hex 6

Sector	Switching pattern
Sector 1	101,-111,-101,-111,101
Sector 2	101,-111,111,-111,101
Sector 3	101,11-1,111,11-1,101
Sector 4	101,11-1,10-1,11-1,101
Sector 5	101,-10-1,10-1,-10-1,101
Sector 6	101,-10-1,-101,-10-1,101

Fig.2 Switching pattern of improved modulation

TABLE I
27 Switching states for NPC Inverter

Vectors	Switching states	$V_{ao}(t)$	$V_{bo}(t)$	$V_{co}(t)$	$V_{com(t)}$
V_0	000	$-V_{dc}/2$	$-V_{dc}/2$	$-V_{dc}/2$	**$-V_{dc}/2$**
V_1	100	0	$-V_{dc}/2$	$-V_{dc}/2$	**$-V_{dc}/3$**
V_2	-100	$V_{dc}/2$	$-V_{dc}/2$	$-V_{dc}/2$	$-V_{dc}/6$
V_3	010	$-V_{dc}/2$	0	$-V_{dc}/2$	**$-V_{dc}/3$**
V_4	110	0	0	$-V_{dc}/2$	$-V_{dc}/6$
V_5	-110	$V_{dc}/2$	0	$V_{dc}/2$	0
V_6	0-10	$-V_{dc}/2$	$V_{dc}/2$	$-V_{dc}/2$	$-V_{dc}/6$
V_7	1-10	0	$V_{dc}/2$	$-V_{dc}/2$	0
V_8	-1-10	$V_{dc}/2$	$V_{dc}/2$	$-V_{dc}/2$	$V_{dc}/6$
V_9	001	$-V_{dc}/2$	$-V_{dc}/2$	0	**$-V_{dc}/3$**
V_{10}	101	0	$-V_{dc}/2$	0	$-V_{dc}/6$
V_{11}	-101	$V_{dc}/2$	$-V_{dc}/2$	0	0
V_{12}	011	$-V_{dc}/2$	0	0	$-V_{dc}/6$
V_{13}	111	0	0	0	0
V_{14}	-111	$V_{dc}/2$	0	0	$V_{dc}/6$
V_{15}	0-11	$-V_{dc}/2$	$V_{dc}/2$	0	0
V_{16}	1-11	0	$V_{dc}/2$	0	$V_{dc}/6$
V_{17}	-1-11	$V_{dc}/2$	$V_{dc}/2$	0	**$V_{dc}/3$**
V_{18}	00-1	$-V_{dc}/2$	$-V_{dc}/2$	$V_{dc}/2$	$-V_{dc}/6$
V_{19}	10-1	0	$-V_{dc}/2$	$V_{dc}/2$	0
V_{20}	-10-1	$V_{dc}/2$	$-V_{dc}/2$	$V_{dc}/2$	$V_{dc}/6$
V_{21}	01-1	$-V_{dc}/2$	0	$V_{dc}/2$	0
V_{22}	11-1	0	0	$V_{dc}/2$	$V_{dc}/6$
V_{23}	-11-1	$V_{dc}/2$	0	$V_{dc}/2$	**$V_{dc}/3$**
V_{24}	0-1-1	$-V_{dc}/2$	$V_{dc}/2$	$V_{dc}/2$	$V_{dc}/6$
V_{25}	1-1-1	0	$V_{dc}/2$	$V_{dc}/2$	**$V_{dc}/3$**
V_{26}	-1-1-1	$V_{dc}/2$	$V_{dc}/2$	$V_{dc}/2$	**$V_{dc}/2$**

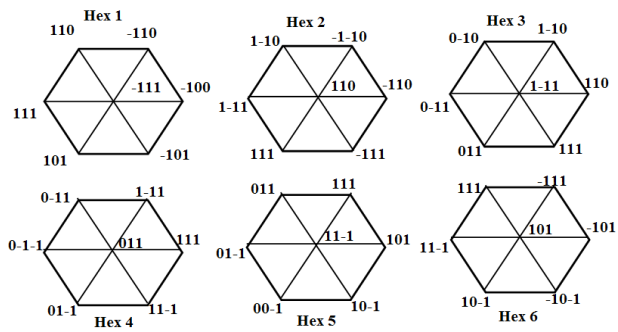


Fig. 3 Simplification of three level space vector diagram in two level space vector diagram

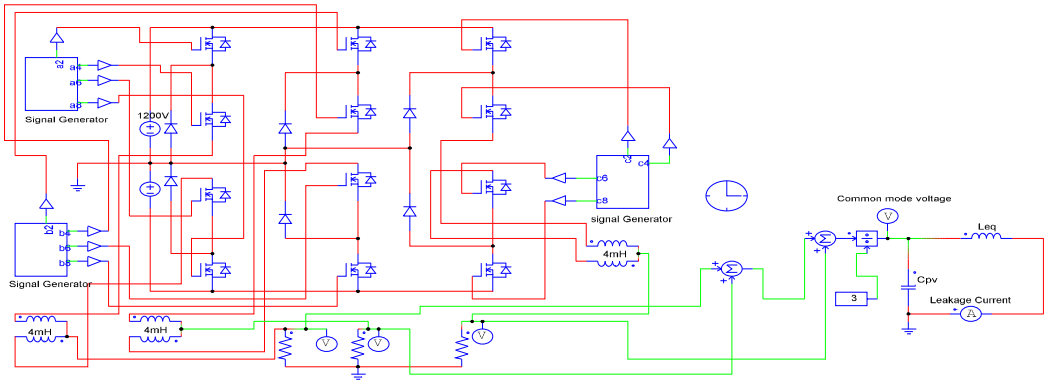
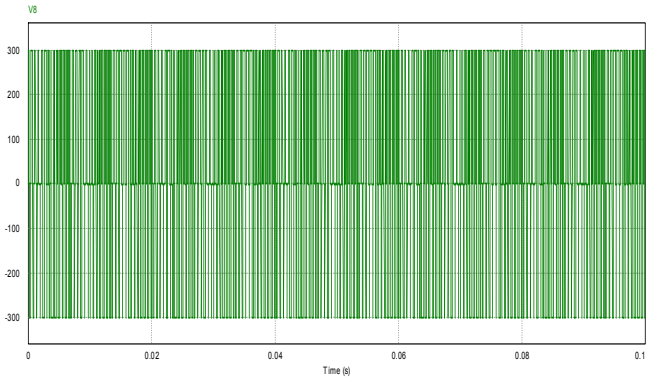


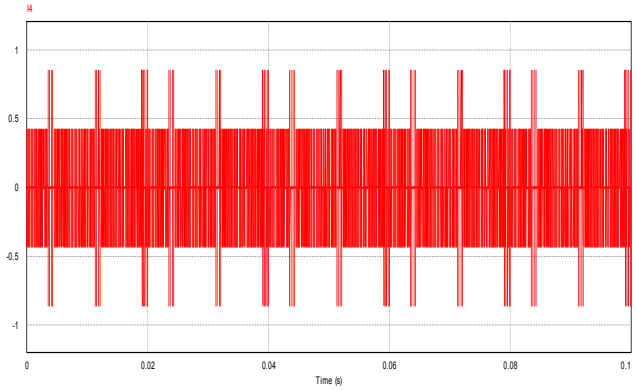
Fig.4 Simulated circuit diagram of SINPC in PSIM soft ware

IV. Simulations and Performance Appraisal

In this section, conventional NPC inverter and proposed SINPC inverter are simulated to display the validation of the proposal of reducing Vcm and leakage current. The switching pulse for 12 individual MOSFETs of both circuits is generated according to the aforementioned improved modulation technique. The simulation parameters are, Input DC voltage=1200V, Split inductor= 4mH, Cpv=0.1uF, Output voltage frequency=50Hz. The PV array was simulated with a DC voltage source and the values of split inductors for three phases are considered equal. The simulated Vcm and leakage current curves for conventional NPC inverter are depicted in Fig.5. It exhibits that Vcm has peak to peak 600V



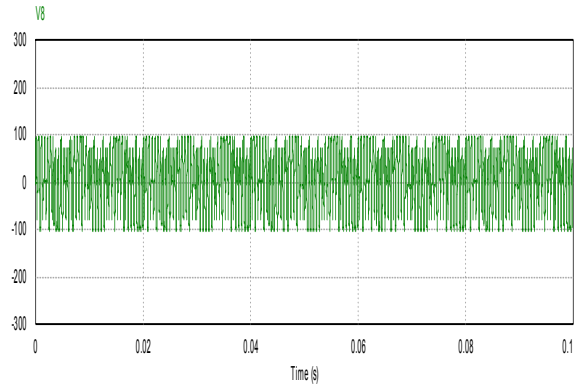
(a)



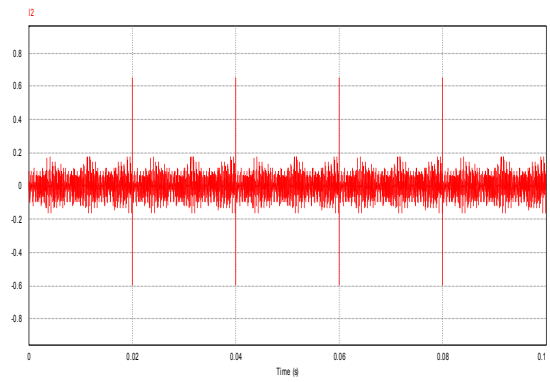
(b)

Fig.5 (a) Vcm of conventional NPC inverter (b) leakage Current

amplitude with large variation which is good enough to generate high leakage current. The measured leakage current through Cpv at this stage is 800mA. Similar analysis has been performed for SINPC inverter. Simulated results are presented in Fig.6 where it is noticed that the value of Vcm and leakage current reduced remarkably. Vcm falls down to 200V peak to peak where the leakage current becomes300mA. The comparison of the two curves validates the effectiveness of the proposed circuit. We have taken the output voltage wave shape across the load displayed in Fig.7(a). The FFT analysis of the proposed SINPC inverter and conventional NPC inverter has shown in Fig.7(b) and Fig.7(c) respectively. It is evident from the figure that SINPC inverter generates less harmonics (THD 0.176%) compare to conventional NPC inverter (1.23%)



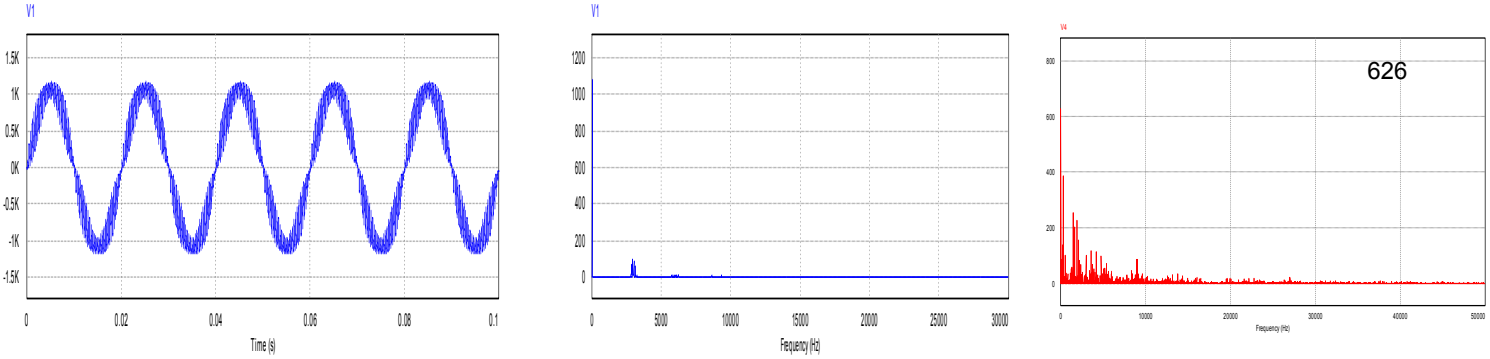
(a)



(b)

Fig.6 (a) Vcm of SINPC inverter (b) Leakage current

In the proposed SINPC Inverter,the value of split inductor is a pivotal parameter to reduce the Vcm and leakage current.So,we have measured the Vcm and leakage current by



(a) (b) (c)

Fig. 7 (a) Output voltage of SINPC inverter (b) FFT analysis of output voltage of SINPC inverter (c) FFT analysis of conventional NPC inverter

varying the split inductor value which is demonstrated in Fig.8 , Fig.9 and also sum up in TABLE II. It exhibits that the higher value of split inductor generates lower Vcm as well as lower leakage current. But larger values of split inductors will increase the overall inverter size, weight and loss. So we need to obtain an optimum value of split inductors which will be discussed elsewhere.

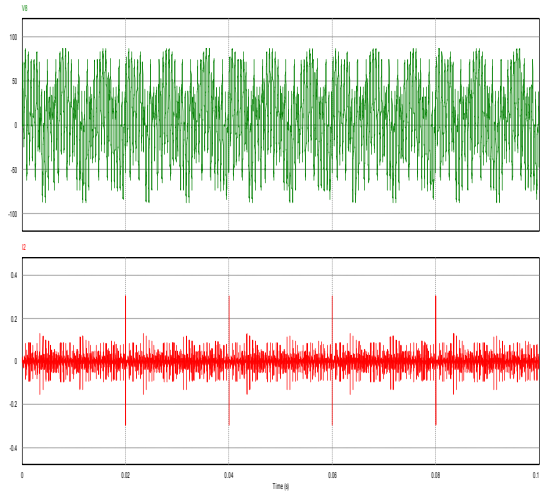


Fig. 8 Vcm and Leakage Current for 10mH Split Inductor

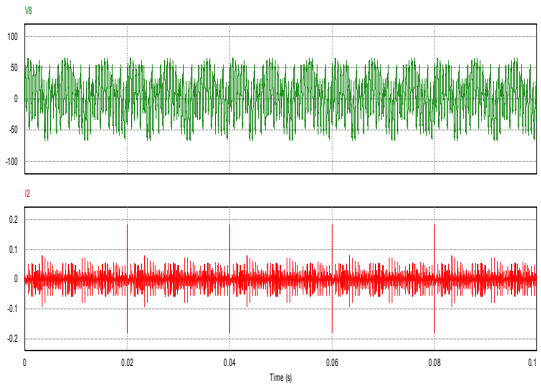


Fig. 9 Vcm and Leakage Current for 20mH Split Inductor

TABLE II. VCM AND LEAKAGE CURRENT VARIATION WITH SPLIT INDUCTANCE

Split Inductor Value	Common Mode Voltage (peak to peak)	Leakage Current (peak to peak)
4mH	200V(almost)	300mA(almost)
10mH	160V(almost)	200mA(almost)
20mH	120V(almost)	120mA(almost)

V. CONCLUSIONS

Variable common mode voltage Vcm as well as leakage current reduction strategy has been proposed in this paper. Based on the previous analysis we have observed that, Vcm usually depends upon the modulation technique and circuitry configuration. So we have applied a new modulation technique in a new split inductor based NPC inverter. Our analysis shows that due to the use of split inductors, the Vcm and leakage current reduce comprehensively compare to the conventional NPC inverter. The simulation shows that the new SINPC inverter also generates less higher order harmonics. So where the system is burdened with varying Vcm or leakage current problem, SINPC inverter with this improved modulation technique can resolve it adroitly.

REFERENCES

[1] T. Kerekes, R. Teodorescu, and M. Liserre, “Evaluation of three-phase transformerless photovoltaic inverter topologies,” IEEE Trans. Power Electron., vol. 24, no. 9, pp. 2202–2211, Sep. 2009.

[2] R. Gonzalez, E. Gubia, J. Lopez, and L. Marroyo, “Transformerless singlephase multilevel-based photovoltaic inverter,”IEEE Trans. Ind. Electron. vol. 55, no. 7, pp. 2694–2702, Jul. 2008.

[3] Huafeng Xiao, Shaojun Xie, “Transformerless Split- Inductor Neutral Point Clamped Three-Level PV Grid-Connected Inverter”, IEEE Transaction on Power Electronics, Vol. 27, No. 4, April2012

[4] M. Ebrahim Adabi, Abolfazl Vahedi, “Common mode Voltage Reduction Strategy for a DFIG with a Three Level Back to Back Converter”,IEEE conference on Power Electronics, Drive Systems and Technologies,2011.

[5] E. Gubra, P. Sanchis, A. Ursua, J. Lopez, and L. Marroyo, “Ground currents in single-phase transformerless photovoltaic systems,” Prog. Photovolt.: Res. Appl., pp. 629–650, May 2007.

[6] Jin.Woo.Jung, “Space Vector PWM Inverter,” Ph.D dissertation, Dep.Elect. Eng., Ohio University, Feb 2005.

[7] “Wind Farm- Detailed DFIG Model” Matlab SimPowerSystems, Demonstration files, www.mathworks.com

Simulation of Lightning Surge Characteristics on Reduced Scale Tower Model

Md. Salah Uddin Yusuf* and Mohiuddin Ahmad

[12]Department of Electrical and Electronic Engineering
Khulna University of Engineering & Technology
Khulna-9203, Bangladesh
*ymdsalahu2@gmail.com

Abstract—**This paper demonstrates the lightning surge response of a reduced scale four parallel vertical conductor model of tower for power systems including the effects of ground plane and without ground plane. Tower surge responses are analysed using Numerical Electromagnetic Code (NEC-2). Different types of model tower configuration are considered to determine, which will provide suitable structure of tower design for lightning. Besides, voltage, current and surge impedance are also evaluated to observe the lighting phenomena for the design of protective equipments so save our daily used sophisticated electronic equipments.**

Index Terms— **Lightning Surge, Method of Moment (MoM), Surge Impedance, Vertical Conductor, NEC-2.**

I. Introduction

Lightning is one of the most significant sources of overvoltages in overhead transmission lines. Transmission line faults caused by lightning strikes give serious damages, such as a massive blackout and instant voltage drop, on electric power systems. Therefore, the rational lightning protection measures should be adopted for ensuring a stable electric power system. Analysis of lightning surges is important for power systems, since lightning strokes on transmission towers, shield wires and phase conductors may cause insulation flashover which may lead interruption of continuous power transmission.

Several studies on surge analysis of the transmission lines have been carried out to evaluate the effect of the lightning protection measures and to estimate the lightning outrage rates [1]-[3] There are many works that have been devoted to the analysis of lightning surges on transmission towers. Experimental studies on this subject have usually been carried out in Japan and measured results have become an important base for computer simulations [4]. In experimental work done by Kawai surge response of transmission tower has been measured [1]. Later Ishii performed experiments on a full scale tower with phase conductors and ground wires [2] and as a result of these experiments a multistory tower model has been proposed for multiconductor analysis in EMTP. Series of experiments later have been continued on a different tower configuration [3] [4] to generalize the results obtained in previous works. On the other hand, in the area of numerical developments, Almeida and Correia De Barros [5] used finite element method for the EMTP simulation. Later Ishii and Baba [6] analyzed a large scale UHV transmission tower using moment method by solving electric field equations directly, and in their study the effect of slant elements, horizontal elements and crossarm has been investigated. Using the same method, they also investigated lightning surge characteristics of a transmission line comprising a tower, a shield wire and phase conductors with the help of computer program NEC-2 [7]. Next, in one of the works tower body and crossarms are modeled in details using short lines sections [8]. Recently, Mozumi et al. analyzed archorn voltages of a simulated 500 kV twin-circuit line [9]. In this paper, lightning surges on reduced scale four parallel conductor model a transmission tower are analysed using the numerical solution of the electric field integral equations used in NEC-2 are obtained.

For the present analysis, the (NEC-2) is employed. It is a widely used three-dimensional (3-D) electromagnetic modelling code based on the MoM in the frequency domain, and is particularly effective in analyzing the electromagnetic response of antennas or of other metallic structures composed of thin wires. A vertical conductor system needs to be decomposed into thin wire elements, and the position, orientation and the radius of each element constitute the input data, along with the description of the source and frequency to be analysed. In the analysis, all the elements in the systems are treated as perfect conductors. To solve the time varying electromagnetic response, Fourier transform and inverse Fourier transform are used. The study in this paper precedes the work done to obtain the surge characteristics of reduced scale model of tower struck by lightning with the help of NEC-2 in case of direct and indirect stroke.

II. Theoretical Formula of Surge Impedance

There is a theoretical formula of surge impedance [10] of a vertical conductor, in case with ground plane and without ground plane. Suppose that lightning surge strike on the vertical conductor whose height is h and radius is r. Then the surge current wave is reflected at the ground of the perfect conductor and returns to the top of the vertical conductor. Introducing the current reflectivity $\beta = 1$ and the magnetic field reflectivity $\gamma(\gamma_i, \gamma_r) = 0$, the theoretical formula of surge impedance which is very close to the well-known empirical formula of Dr. Hara, is obtained as follows:

$$Z = 60\left\{ln\left(\frac{h}{2r}\right) - \frac{1}{4}\right\} = 60\left\{ln\left(\frac{2\sqrt{2}\,h}{r}\right) - 1.983\right\} \quad (1)$$

Equation (1) gives the surge impedance of the vertical conductor just after the occurrence of the reflection of the traveling wave propagation down from the top of the structure. However if it is considered that $\beta = \gamma_i = \gamma_r = 1$, the potential $V(t)$ generated in the vertical conductor at h became,

$$V(t) = \frac{c\mu_0 I_0}{2\pi}\left\{ln\frac{(ct + 2r)}{2r} - \frac{ct}{2(ct + r)}\right\}$$

The above equation can be modified by substituting $ct = 2h$, where c is the velocity of light and assuming $h \gg r$ as follows:

$$Z = 60\left\{ln\left(\frac{h}{r}\right) - \frac{1}{2}\right\} = 60\left\{ln\left(\frac{2\sqrt{2}\,h}{r}\right) - 1.540\right\} \quad (2)$$

On the other hand, if there is no ground, the following formula is induced

$$V(t) = \int_0^{ct} (-E_i.dl) = \frac{c\mu_0 I_0}{2\pi}\left\{ln\frac{(ct+2r)}{2r} - \frac{ct}{2(ct+r)}\right\}$$

Substituting $ct = 2h$ and assuming $h \gg r$ in the above equation, we get

$$Z = 60\left\{ln\left(\frac{h}{r}\right) - \frac{1}{2}\right\} = 60\left\{ln\left(\frac{2\sqrt{2}\,h}{r}\right) - 1.540\right\} \quad (3)$$

This formula given by equation (2) is the same as equation (3). This is suitable for calculating the value of surge impedance single vertical conductor in case of with and without ground plane.

III. Surge Simulation of Reduced Scale Model

Lightning is responsible for severe electromagnetic stress in electric and electronic systems that frequently result in damages or operational outages. The evaluation of such a stress is very important for lightning protection. So in this research, lightning surge response for four parallel conductor model as reduced scale model of transmission tower has been taken into account. The surge impedance obtained from four parallel conductors is analyzed for both cases with ground plane and without ground plane.

Fig.1 shows the structures subjected to analysis. The single vertical conductor is shown in Fig. 1(a). The four parallel vertical conductor's model of a tower is shown in Fig.1 (b), which has a radius of 16.5 mm and 404 mm apart. A steep-front current having a rise time of 5 ns was injected into a current lead wire as shown in Fig.2. In Fig.2 the injected ramp also compared with the computed current at the input terminal. For the numerical analysis, the conductor system needs to be modeled with segments according to the modeling guidelines stated in the preceding chapter. It is not difficult to satisfy recommended conditions except the following one: $0.001\lambda < \Delta L < 0.1\lambda$ This requirement is too severe to be satisfied exactly, since the wider frequency range is needed here to investigate the transient characteristics of such a system. However the author found the following condition yielding reasonable computation results by trial and error process. Frequency range: 1.5625 MHz to 0.8000 GHz with 1.5625 MHz increment step and Segment length: 0.202 m to 0.300 m. This frequency range corresponds to the time range of 0 to 640ns with 1.25ns Increments.

A. *With Ground Plane*

The following Fig.3 shows the arrangement of the four parallel conductor's models to be analysed with ground case. To evaluate the voltage of the top of a structure, 10-kΩ resistance was inserted between the top of a structure and the end of the voltage measuring wire and the waveform of the current flowing through the resistance was calculated. The system of structures under the numerical analysis was postulated to be on the perfectly conducting ground. This postulation is used throughout in this thesis. Fig.4 shows the computed waveforms of the voltage and the current at the top of the four parallel conductors

If a travelling wave propagates along the tower at the velocity of light the reflected wave from the ground should return to the tower top after the round-trip time of a travelling

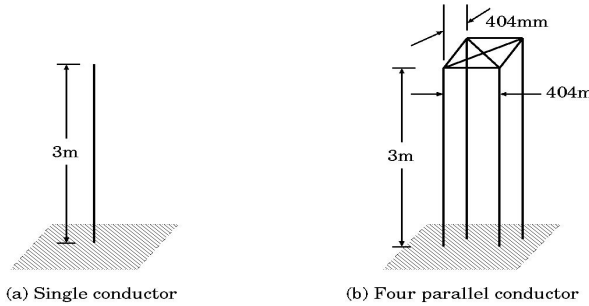


Fig.1 Structure subjected to analysis

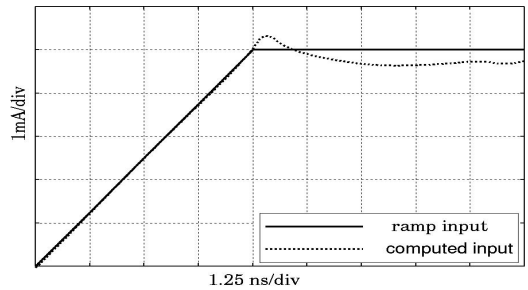


Fig.2 Comparison of injected and computed input waveforms for analysis

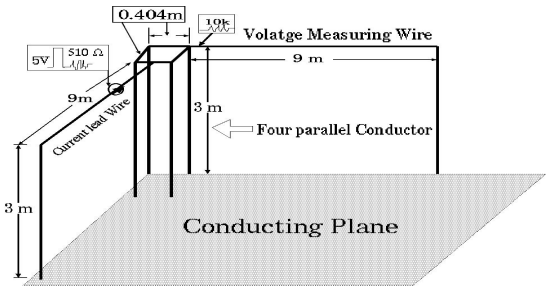
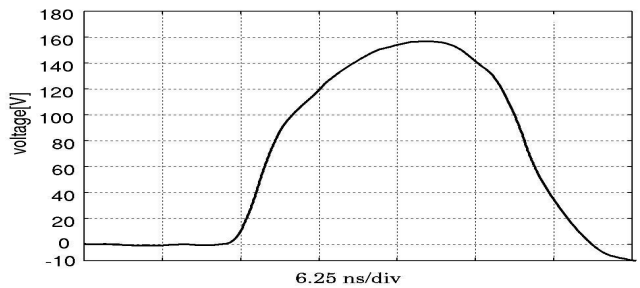
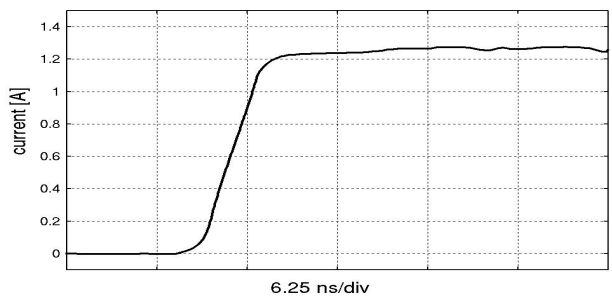


Fig.3 Arrangement of Four parallel conductor system for simulation with NEC-2 with ground plane.



(a) Voltage waveform



(b) current waveform

Fig.4 NEC-2 results of voltage at the top and current waveforms of the four parallel conductors with ground plane case.

return to the tower top after the round-trip time of a travelling wave in the tower which is 20 ns for this structure as shown in Fig.3. But form the computed result it is seen that voltage waveforms reach their peaks about 17 ns after the beginning which indicates that a negative voltage wave arrives at the top before the arrival of the reflected wave from the ground.

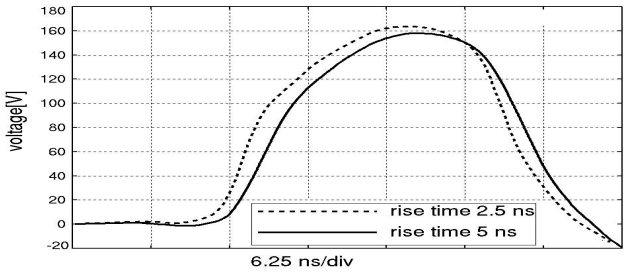


Fig.5 NEC-2 results of voltage waveforms at the top of the four parallel conductors with ground plane case for different rise time.

This phenomenon is due to the electromagnetic field of TEM mode illuminating the structure, i.e., the electromagnetic wave associated with a travelling wave in the current lead wire arrives at the structure foot simultaneously with the arrival at the top of a structure and a negative voltage wave is induced at the structure foot before the occurrence of the reflection of the travelling wave propagating down from the top of the structure. This induced wave is observed in the current waveform as shown in Fig.4 (b).

The analysis was carried out for different rise time as shown in Fig.5, which were 5 ns and 2.5 ns respectively. From the output it is seen that for small rise time the voltage of the structure at the top reaches the peak value earlier than other.

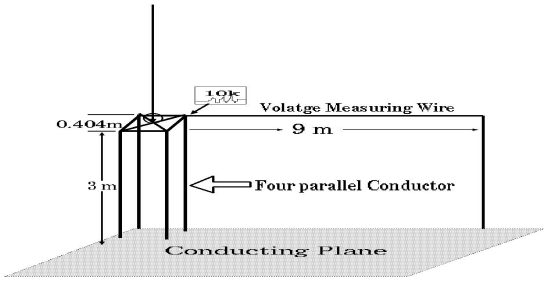
B. *Without Ground Plane*

The following Fig.6 (a), (b) and (c) show the arrangement of the four parallel conductor model to be analyzed without ground cases. On the other hand from numerical analysis by NEC-2 the computed waveforms of the voltage and the current at the top of the four parallel conductors are obtained as shown in Figs.7, 8 and 9. The rise time was set to be 5 ns as shown in the following Figs. In Fig.7 and 8, the voltage waveform rises exponentially and reaches their peak after 21 ns, which shows that in this case the electromagnetic wave propagates along the conductor with the velocity of light. Then it decreases downward as the reflection comes from the ground. But in Fig.9 the voltage waveform rises abnormally and reaches its peak after 25 ns, which is not similar to the other cases. As the output for structure of Figs.6 (a) and (b) show the characteristics like single vertical conductor so this structure is suitable for analysis and design. On the other hand, structure of Fig.6(c) is not suitable for design purposes.

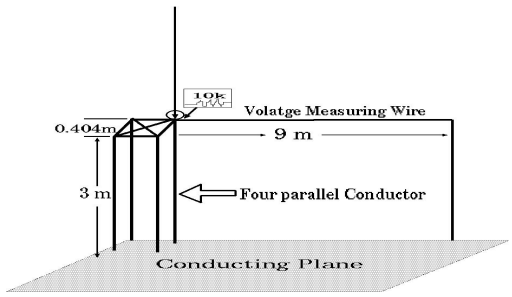
IV. Surge Impedance of the Simulated Model

The tower surge impedances are defined as the ratio of the instantaneous values of the voltage to the current at the moment of voltage peak. The computed tower surge impedance's using NEC-2 is summarized in TABLE I.

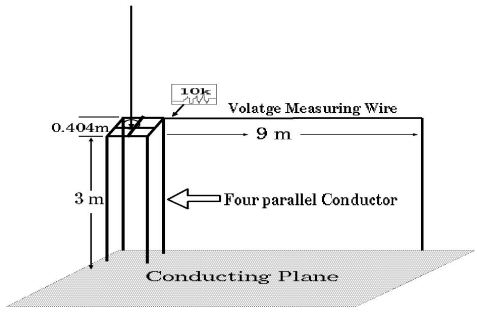
The difference of the surge impedance's depends on the arrangement that can be explained by the electric field associated with the steep-front currents flowing in the current lead wire, the voltage measuring wire or earth wire and the vertical conductor or tower. With vertically applied input or without ground plane case, the positive surge current from the pulse generator flowing into the vertical conductor and the negative current propagating up the vertical current lead wire produce the strong upward electric field. This makes the injected current split more into the voltage measuring wire and less into the vertical conductor. On the other hand, a strong horizontal electric field is produced by the negative current into the horizontal current lead wire. This makes the



(a) Injection is at the center top for cross configuration

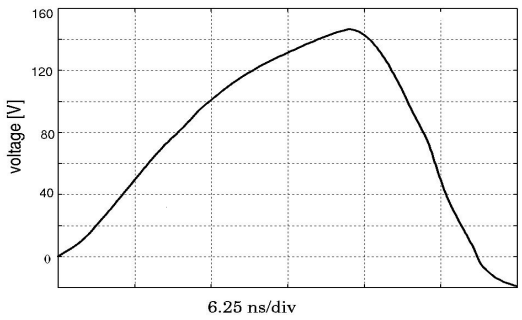


(b) Injection is at one end of top for cross configuration

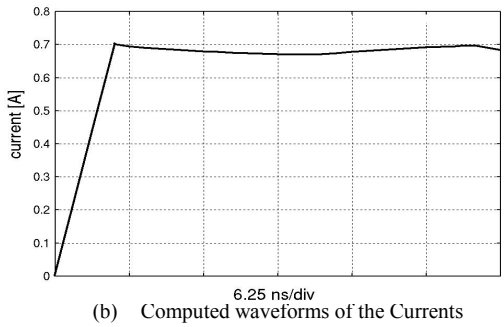


(c) Injection is at the junction of plus configuration

Fig.6 Arrangement of four parallel conductor system for simulation with NEC-2 in case without ground plane



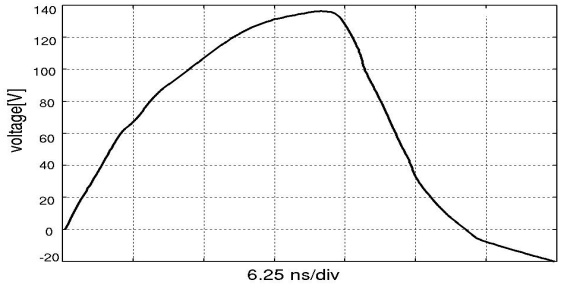
(a) Computed waveforms of the Voltage



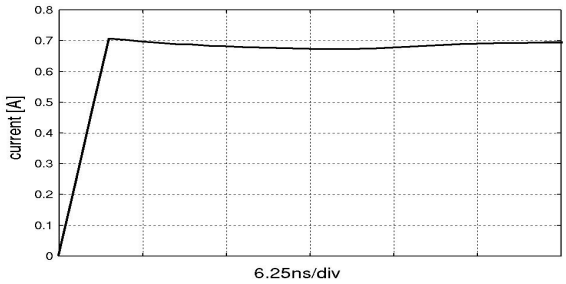
(b) Computed waveforms of the Currents

Fig.7. Computed waveforms of the voltage and the current at the top for Fig. 6(a)

injected current split more into the vertical conductor and less into the voltage measuring wire.

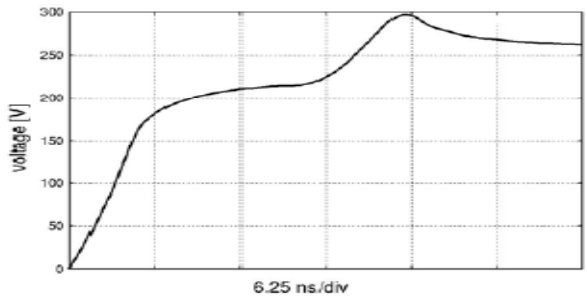


(a) Computed waveforms of the Voltage

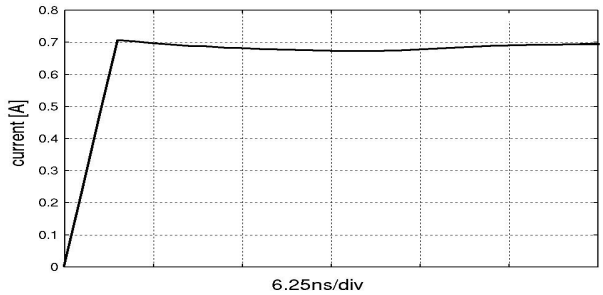


(b) Computed waveforms of the Currents

Fig.8 Computed waveforms of the voltage and the current at the top for Fig.6 (b)



(a) Computed waveforms of the Voltage



(b) Computed waveforms of the Currents

Fig.9 Computed waveforms of the voltage and the current at the top for Fig.6(c)

V. CONCLUSIONS

The surge characteristics obtained in this research are most likely to be the same as electromagnetic phenomena of lightning. These results can be used to estimate the tower surge impedance of the standard arrangement with a vertical current lead wire and horizontal voltage measuring wire as well as with ground plane in case of four parallel vertical conductor models. From the analysis Figs.6 (a) and 6 (b) as the models are recommended. This analysis will help us to configure the structure of control building and high tech building when highly sophisticated instruments and electronic equipments are installed.

TABLE I
SIMULATION RESULTS OF SURGE IMPEDANCE

Configuration for NEC-2	With ground plane (Ω)	Without ground plane (Ω)
For Fig.3 and Fig.6(a)	126	195
For Fig.3 and Fig.6(b)	126	205
For Fig.3 and Fig.6(c)	126	255

REFERENCES

[1] M. Kawai: "Studies of the surge response on a transmission tower," IEEE 1Tans. Power Apparatus and Systems, vol.83, no.1, pp.30-34, Jan. 1964.

[2] M. Ishii, T. Kawamura, T. Kouno, E. Ohsaki, K. Shiokawa, K. Murotani and T. Higuchi" Multistory transmission tower model for lightning surge analysis," IEEE 1Tans. Power Delivery, vol.6, no.3, pp.1327-1335, Jul. 1991.

[3] T. Yamada, A. Mochizuki, J. Sawada, E. Zaima, T. Kawamura, A. Ametani, M. Ishii and I; S. Kato: "Experimental evaluation of a UHV tower model for lightning surge analysis,"IEEE Trans. Power Delivery, vol.10, no.1, pp.393-402, Jan. 1995.

[4] G. D. Breuer, A. J. Schultz, R. H. Schlomann and W. S. Price: "Field studies of the surge response of a 345-kv transmission tower and ground wire," AIEE Tans. Power Apparatus and Systems, vol.76, pp.1392-1396, Feb. 1958.

[5] M. E. Almeida and M. T. Correia de Barros : "Tower modelling for lightning surge analysis using electro-magnetic transients program," IEE Proc. Gener. Transm. Distrb., vol.141,no.6, pp.637-639, Nov. 1994.

[6] Shen Zhi-heng, Zhou Hao, "Simulated calculation and analysis of lightning performance of UHV A.C. transmission line" 4th International Conference on Electric Utility Deregulation and Restructuring and Power Technologies (DRPT), pp. 250 – 253, July, 2011.

[7] M. Ishii and J. Hojo: "Evaluation of lightning fault rate of EHV transmission line based on lightning parameters derived from electromagnetic field observation,"Trans. IEE Japan, vol.111-B, no.5, pp.535-542, May 1991.

[8] H. Motoyama : "Experimental study of lightning surge characteristics on transmission tower and transmission lines using a scale model," CRIEPI Report, no.T96074, Jun. 1997.

[9] W. A. Chisholm, Y. L,Chow and K. D. Srivastava : " Lightning surge response of transmission towers," IEEE 1rans. Power Apparatus and Systems, vol.102, no.9, pp.3232-3242,Sep.1983.

[10] T. Hara, 0.Yamamoto, M. Hayashi and C. Uenosono : "Empirical formulas of surge impedance for single and multiple vertical cylinder," Tran.s. IEE Japan, vol.110-B, no.2,pp.129-137, Feb. 1990.

[11] http://www.nec2.org/other/nec2prt3.pdf.

Aberrant Amplification of Torsional Oscillating Torques in Turbine-Generator Shaft Due to Multiple EAF Operation

Md. Minarul Islam*, A. Hasib Chowdhury
Department of EEE, Bangladesh University of Engineering and Technology
Dhaka-1205, Bangladesh
*minarbuet@gmail.com

Abstract-**Electric arc furnace (EAF) acts as a source of sub-synchronous frequencies (SSFs) in the electrical network. If these SSFs lead to resonance, respective matched torsional modes of turbine-generator (T-G) shaft are excited causing torsional oscillating torques (TOTs) amplification. This may lead to gear damage, shaft fatigue lifetime reduction, severe damages to T-G shaft. This paper presents an assessment of TOTs amplification due to multiple EAF operations at the same PCC. The EAFs have been modeled using mathematical model of the electric arc, where resistances of EAFs varied deterministically in frequency range of 1.2-45.5 Hz to comport with realistic EAF. TOTs have been measured with/without resonant excitation of torsional modes. Eigenvalue analysis and FFT technique have been used to analyse torsional torques.**

Index Terms- **Electric arc furnace, sub-synchronous resonance, torsional oscillating torques, shaft fatigue**

I. Introduction

The dominant frequencies of torsional oscillating torques (TOTs) among masses of turbine-generator (T-G) shaft are in subsynchronous range [1]. Among the electrical loads, electric arc furnace (EAF) comports as a blatant source of subsynchronous frequencies (SSFs) [2]-[5] and may induce pulsating components in the air-gap torque of synchronous machine through which TOTs are amplified during torsional resonance (TR). The consequential torsional stress on T-G unit depends on EAF actions and how it is connected to grid, characteristics of T-G and utility power system.

Higher frequency components in air-gap torque have dominant effects on smaller components such as turbine blades and lower frequency components stimulate the TOTs in main shaft sections [6]. The random property of arc melting process and control system contribute to thermal and electrical dynamics which may affect supply system adversely [7]. In [8], Robert et el. discussed unusual power oscillations which were observed at two power plants located near large EAF.

Torsional amplification can lead to gear damage, shaft fatigue lifetime reduction, severe damages to T-G shaft [2]. Although the shaft could withstand most severe impact due to three-phase fault, it still can not guarantee long-term safety operations even under normal operating conditions if there occurs TR [3]. Mechanical failures have been consistently reported due to machine air-gap torques supplied by PWM drives [2]. The first and second T-G shafts failure due to subsynchronous resonance (SSR) were experienced at the Mohave plant in sourthern California [9].

Evidently, the impact on the torsional torque due to operation of multiple numbers of EAFs at same point of common coupling (PCC) or at close electrical proximity has not been reported in details yet. This paper presents an assessment of TOTs amplification and find out its severity due to operation of multiple EAFs at the same PCC.

II. Component Modeling

The major components such as EAF and air-gap torque of synchronous generator is presented in this section.

A. Electric Arc Furnace

A precise model for three-phase EAF is required to analyse its characteristics including SSFs. Due to nonstationary stochastic characteristics of steel making process using electric arc, it is very difficult to model EAF load appropriately. Furthermore, there are many external factors such as melting stage, scrap materials, electrode position, electrode arm control, the supply system voltage and impedance, which affect operation of EAF [10]. It is therefore desired to include all these factors in designing a model which can resemble realistic EAF. Hence, a dynamic model based on the principle of energy conservation of electric arc [11] has been considered for this simulation. The relationships between supplied current to furnace lead, i, arc radius, r, and arc voltage, v, can be expressed by the following equations:

$$C_1 r^n + C_2 r \frac{dr}{dt} = \frac{C_3}{r^{m+2}} i^2 \quad (1)$$

$$v = \frac{C_3}{r^{m+2}} i \quad (2)$$

Where C_1, C_2, C_3 are constants.

Fig. 1 shows V-I characteristic of electric arc which has been obtained by MATLAB simulation using equations (1) and (2), and this resembles to characteristic of realistic EAF.

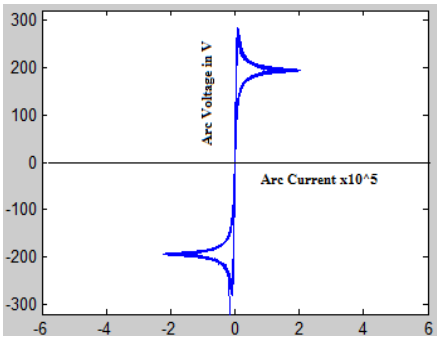


Fig. 1. Simulated V-I characteristic

Fig. 2 shows the simulation output of current waveform in phase-a at the primary side of furnace transformer. The waveform is modulated which can be expressed by the following equation.

$$i_a(t) = \sqrt{2} I_{rms} [1 + \sum I_n \sin(2\pi f_n t + \phi_n)] \sin(2\pi f_s t + \beta) \quad (3)$$

where, f_s is system fundamental frequency, I_n and f_n are amplitude and frequency of modulating current, and β is angle between feeder voltage and current.

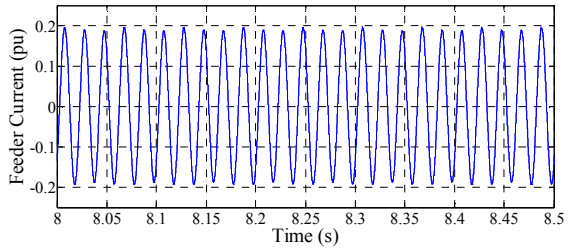


Fig. 2. Feeder current i_a at the primary side of furnace transformer

The frequency spectrum of the modulating current gives f_s+f_n and f_s-f_n, which are defined as super- and sub-synchronous frequencies, respectively. This current may disturb generator stator current which cause perturbation in air-gap torque. If SSR occurs, a large subsynchronous current may result from a small voltage induced by rotor oscillations. A rotor torque corresponding to this current enhances the rotor oscillation. If this torque become larger than mechanical damping, the coupled electromechanical system will experience increasing oscillations.

A. Air-Gap Torque of Synchronous Generator

The air-gap torque of synchronous generator, in *dq0* reference frame, can be expressed by the following equation.

$$T_e = \psi_d i_q - \psi_q i_d \tag{4}$$

where, $i_d = f(i_a, i_b, i_c)$, $i_q = f(i_a, i_b, i_c)$, $\psi_d = f(i_d, i_{fd}, i_{kd}, L)$, $\psi_q = f(i_q, i_{kq}, L)$, L includes leakage, self and mutual inductances. So it is seen that the air-gap torque is function of *dq*-axis flux linkages and currents. At synchronous operation, *dq*-axis parameters appeares as constant i.e. no change in T_e. But if there are any changes or disturbances in those parameters, T_e will be disturbed resulting in perturbation in TOTs.

III. Assessment of Torsional Mode Amplification

A test system consisting of steam turbine driven generator connected to an infinite bus and EAFs has been used for simulations (Fig. 3). In this paper, both single and multiple EAF operation has been considered. The T-G shaft consists of five masses: high presure (HP), intermediate presure (IP), and two low presure (LP_A and LP_B) turbine sections and the generator rotor section. The TOT signals are obtained by first using MATLAB Simulink. Torsional mode frequencies of the T-G system are then obatained by both, eigenvalue analysis of the system and FFT analysis of TOT signals.

B. Description of The Test System

The physical parameters of T-G shaft are given in [9]. The synchronous generator is rated as 3000 rpm, 2-pole, 24 kV, 555 MVA. A static exciter has been used with the generator. The electrical parameters in pu, based on rated value of the synchronous generator, are given in [9]. The generator is supplying power to the network via a step-up transformer (24kV/132kV, 555 MVA) through a transmission line of length 50 km. The 132 kV bus also supplies power to an EAF, located at 8 km distance from PCC, through a 132kV/24kV transformer.

Other transformers, one rated as 24kV/11kV, 200 MVA, three-phase and another rated as 11kV/.5kV, 50 MVA, 3 single phase units are also installed in furnace feeder line to step down voltage to furnace operating value. The impedances of the three-phase and single-phase transformers have been chosen as 0.008+j0.13 pu and 0.0006+j0.13 pu, respectively. Components of impedance- positive siquence: resistance (0.01165 ohm/km), inductance (0.8679 mH/km), capacitance (13.41 nF/km)-and zero sequence: resistance (0.2676 ohm/km), inductance (3.008 mH/km), capacitance (8.57 nF/km) have been used for transmission line.

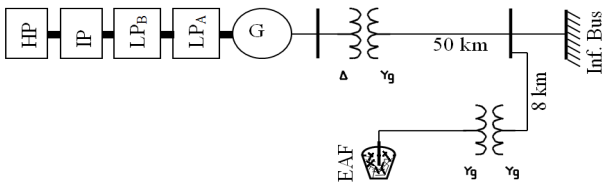


Fig. 3. Single line diagram of the test system

C. Determination of Torsional Mode Frequencies

The T-G system, shown in Fig. 3, can be represented by a simple lumped-mass model [9]. The equations of rotating motion of any rotor masses can be expressed by the following general formulae:

$$2H_j \frac{d\Delta\omega_j}{dt} = T_j + K_{j,j+1}(\delta_{j+1} - \delta_j) - K_{j-1,j}(\delta_j - \delta_{j-1}) - D_j(\Delta\omega_j) \tag{5}$$

$$\frac{d\delta_j}{dt} = \omega_0 \Delta\omega_j \tag{6}$$

where, j=1 to 5. $T_1 = -T_e$, $T_2 = T_{LPA}$, $T_3 = T_{LPB}$, $T_4 = T_{IP}$, $T_5 = T_{HP}$, H = mass inertia, K = rigidity coefficient, D = damping factor, ω = shaft speed, δ = rotor angle.

Equations (5) and (6) are linearized and expressed in state-space form for the eigenvalue analysis. For small disturbance, the change in electromagnetic torque can be expressed as $\Delta T_e = K_s \Delta\delta_1$, where K_s is the synchronizing torqe coefficient. The mechanical torques developed by the turbine section are assumed to be constant. Arranging the linearized equations in matrix from gives:

$$\begin{bmatrix} \Delta\dot{\omega}_i \\ \Delta\dot{\delta}_i \end{bmatrix} = \boldsymbol{A} \begin{bmatrix} \Delta\omega_i \\ \Delta\delta_i \end{bmatrix} \tag{7}$$

where $\boldsymbol{A}$ is the state matrix and function of H, D, K, K_s. The state variables are speed deviations $\Delta\omega_i$ and rotor angle deviation $\Delta\delta_i$, with i = 1 to 5. The eigenvalues of $\boldsymbol{A}$ are calculated. The imaginary part of eiginevalues gives the torsional mode frequency. Table I presents torsional modes and their frequencies.

TABLE I
Mode Frequencies and Torsional Modes

Eigenvalues λ	Mode Frequency (Hz)	Torsional Mode
$\lambda_{1,2}$ = -0.6931652 ± 285.44842383i	45.4	4
$\lambda_{3,4}$ = -0.5566807 ± 196.73304251i	31.3	3
$\lambda_{5,6}$ = -0.6441085 ± 156.21256768i	24.9	2
$\lambda_{7,8}$ = -0.3551595 ± 105.42518046i	16.8	1
$\lambda_{9,10}$ = -0.4069221 ± 3.4898816350i	0.55	0 (system mode)

As the T-G shaft has five masses, there are five modes of torsional oscillation. The mode frequencies are 0.55, 16.8, 24.9, 31.3 and 45.4 Hz. The mode frequencies can also be obtained from the FFT analysis of TOT signals as shown in Fig. 4. It is seen from Fig. 4 that the torsional mode frequencies are 1.2, 16.8, 24.9, 31.3, 45.4 Hz which are almost the same as obtained from eigenvalue analysis technique.

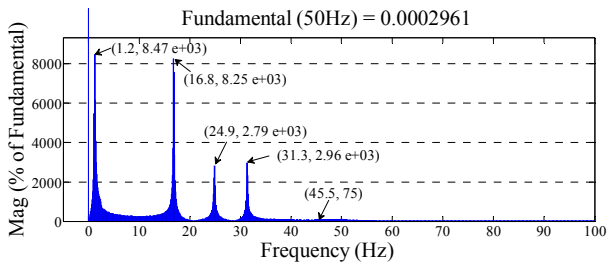


Fig. 4. FFT of torsional oscillating torque between generator and LP_A section

D. Simulation and Results

During the sinusoidal variation of arc resistance a 16.8 Hz signal along 60% modulation index is used simulation.

1)Arc Resistance and Voltage: Due to movement of electrodes and melting of scraps, the arc length changes which cause change in arc resistance. From Fig. 5 it is seen that arc voltage increases precipitately from its zero crossing point to arc ignition voltage, which signposts that the equivalent circuit of EAF comports as an open circuit.

A small leakage current flows through the foamy slag which presumed a constant high resistance implicated by the impulses in Fig. 5(a). When the arc is established, a transient process emerges in the voltage waveform at the beginning of arc melting process. The arc voltage drops exponentially from ignition voltage to a constant value. During arc established period, the electrodes and scraps are short circuited which results in almost zero resistance. After melting scraps the arc is extinguished and the arc voltage drops to arc extinction voltage.

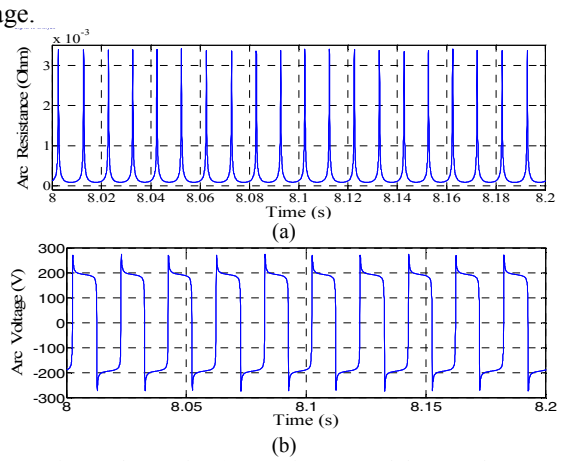


Fig. 5. Changes in (a) arc resistance, and (b) arc voltage

2)Feeder Current: The FFT of the feeder current, shown in Fig. 2, is presented in Fig. 6. It shows that the two frequencies 33.2 Hz and 66.8 Hz are present in feeder current waveforms. These frequencies are present because a 16.8 Hz signal was impregnated during resistance modulation.

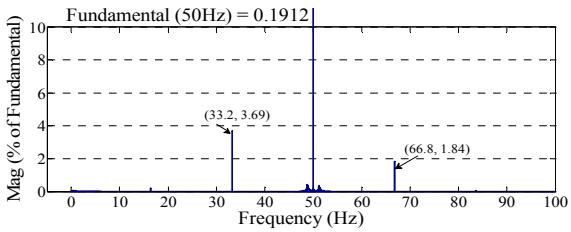


Fig. 6. FFT of i_a near furnace transformer

3) Analysis of Air-gap Torque: Fig. 7 shows air gap torques without/with resonance when there is EAF in the system. Fig. 7(a) shows a small variation of air-gap torque when there is no resonance. Fig. 7(b) shows amplification of this small variation of torque during resonance. This amplification is the result of perturbation of *d*-and *q*-axis current component due to operation of EAF in the system. The EAF produces a subsychronous current of frequency 16.8 Hz which perturb the air-gap torque.

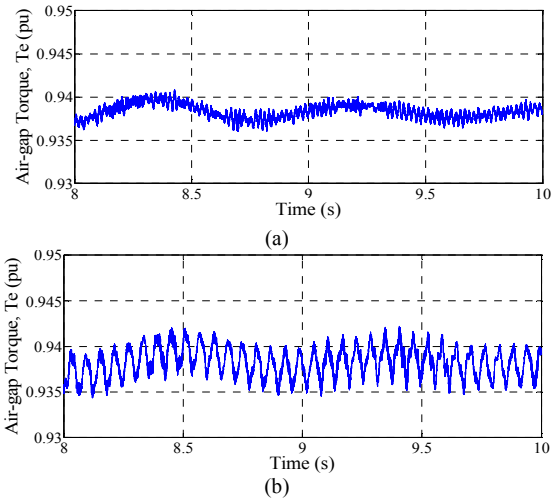


Fig. 7. Air-gap torque (T_e) (a) no resonance, (b) at resonance

4) Analysis of Torsional Torque Amplification: Fig. 8 shows the torsional torque transmission between generator and LP_A section of the turbine without/with resonance when there is an EAF in the power system. From Fig. 8(a), it is seen that the torsional torque reach to a steady value after 8 seconds of starting generator. But if the EAF produces any frequency equal or close to torsional mode frequency, the TOT between generator and LP_A mass is amplified as shown in Fig. 8(b).

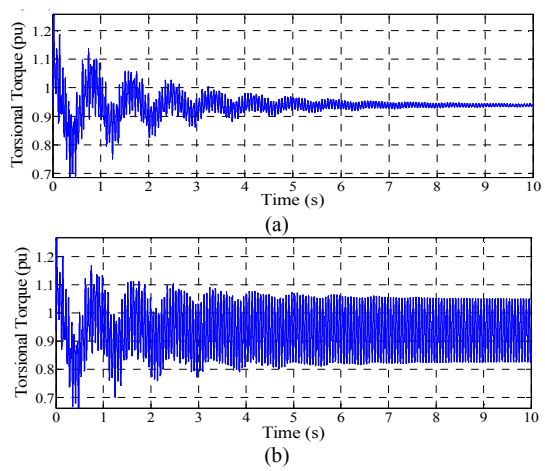


Fig. 8. Torsional torque (a) no resonance, (b) at resonance

E. Quantification of Torsional Torque

A paradigm of calculating amplified TOTs for a single EAF operation using FFT analysis has been expounded in Table II. Here, torque at 16.8 Hz resonance due to EAF has been compared with torque at non-resonance. It is seen that due to resonance, the TOT between generator and LP_A section of turbine is amplified by 4.57 times.

TABLE II
EXCITATION OF TORSIONAL MODE 1

Torque without EAF (pu)	Torque with EAF but resonance (pu)	Torque with EAF at resonance (pu)	Torque amplification
0.02443	a = 0.02533	b = 0.11569	b/a = 4.57

For the assessment of TOT amplification due to multiple EAF operation, for all EAFs, arc resistances have been adapted in such a way that the EAFs portray as current sources of frequencies in the range of 1.2-45.5 Hz, worst case conditions have been considered. TOT amplifications have been compared and shown in Fig. 9. For example, TOT amplification due to resonant excitation of two EAFs at same PCC is almost double to that of one EAF at PCC. As the number of EAFs increase at PCC there cause increased amplification of TOTs at T-G shaft.

Table III presents TOTs amplification with EAFs at PCC. It is seen that the average amplified TOT due to resonant excitation of mode 1 is 8.63 times of its non-resonant TOT. The average torque amplifications of mode 2, 3 and 4 are 12.3, 14.5 and 29.8 respectively. It seen that if one mode is excited it does not affect other modes.

The TOT amplifications due to one EAF at PCC have been compared with that of multiple EAFs at the same PCC in Table IV. From Table IV, it is seen that the TOT amplification, due to resonant excitation of two EAFs at same PCC, is almost double to that of one EAF at PCC. As the number of EAFs increase there cause increased amplification of TOTs.

TABLE III
TOTs AMPLIFICATION WITH TWO EAFs AT PCC

Resonance Condition	f(Hz) / Torque ratio	0	1.2	16.8	24.9	31.3
Resonant at 1.2 Hz	**HP-IP**	0.9965	1.4643	0.5733	0.2695	0.8496
	IP-LP$_B$	1.0061	1.4809	0.5751	0.2695	0.8481
	LP$_B$-LP$_A$	0.9984	1.4882	0.5757	0.2771	0.8527
	LP$_A$-Gen	0.9986	1.4872	0.6171	0.2611	0.7524
Resonant at 1.2 and 16.8 Hz	**HP-IP**	0.9983	1.4394	8.6193	1.1236	1.0122
	IP-LP$_B$	1.0039	1.4396	8.6454	1.1281	1.0217
	LP$_B$-LP$_A$	0.9979	1.4565	8.6348	0.7225	1.0011
	LP$_A$-Gen	0.9988	1.4759	8.6217	1.0939	1.0114
Resonant at 1.2, 16.8 and 24.9 Hz	**HP-IP**	1.0015	1.4521	8.6619	12.4713	0.9624
	IP-LP$_B$	0.9987	1.4458	8.6614	12.4462	0.9398
	LP$_B$-LP$_A$	0.9996	1.4472	8.6647	12.1719	0.9876
	LP$_A$-Gen	0.9983	1.4774	8.6583	12.3875	1.0171
Resonant at 1.2, 16.8, 24.9 and 31.3 Hz	**HP-IP**	1.0018	1.4267	8.6506	12.4908	16.0548
	IP-LP$_B$	0.9999	1.4356	8.6637	12.4668	14.2849
	LP$_B$-LP$_A$	1.0013	1.4639	8.6058	11.7248	14.3422
	LP$_A$-Gen	0.9982	1.4667	8.6335	12.4202	14.3131

TABLE IV
COMPARISON OF TOTs AMPLIFICATION

Mode	Number of EAF				
	One	Two	Three	Four	Five
0	1.24	1.46	1.64	1.83	2.01
1	4.55	8.63	12.42	15.86	19.48
2	6.8	12.3	17.04	22.1	26.71
3	7.4	14.5	19.65	24.61	29.82
4	16.3	29.8	41.46	50.08	58.98

Fig. 9 shows average amplified TOTs (torque ratios resonant to norresonant) due to operation of multiple EAFs. From the trends of the graph in Fig. 9, the torque ratio of a torsional mode due to operation of any number of EAF at the same PCC can be approximated. For example, if N_{eaf}= 8, then the torque ratio corresponding to mode 3 can be calculated using equation as follows:

$$T_{ratio(m3)} = 0.1833N^3_{eaf} - 1.9336N^2_{eaf} + 11.523N_{eaf} - 2.354$$
$$= 0.1833\times8^3 - 1.9336\times8^2 + 11.523\times8 - 2.354 = 59.97$$

This value indicates that mode 3 will be amplified by 59.97 times of its non-resonant TOT if 8 EAFs are connected to the same PCC and mode 3 is excited. From the analysis, it is deducible that the more the EAFs added to the same PCC and if resonant excitation occurs, the larger the amplitude of oscillation of TOT occurs.

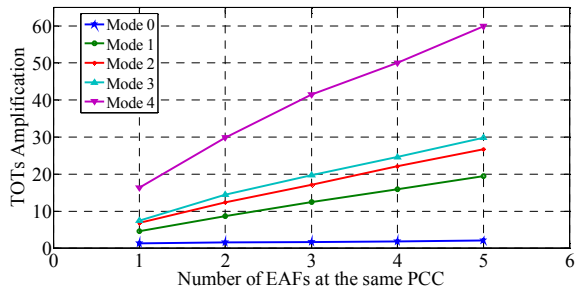


Fig. 9. TOTs amplification due to multiple EAFs at same PCC

IV. CONCLUSION

Subsynchronous frequencies engendered by EAF load stimulate torsional resonance in T-G shaft. Therefore, TOTs are amplified aberrantly which may induce abnormal torsional stress to the shaft system. It may reduce fatigue-life and even cause severe damages to T-G shaft. This paper presents a detailed assessment of impacts of multiple EAF operation at same PCC on nearby T-G system. It has been found that large amplification of TOTs occur when higher modes are excited. It has also been observed that the TOTs increase non-linearly to higher magnitude due to multiple EAF operation at the same PCC. During torsional excitation, TOTs absorb energy. Calculation of the amount of this energy and investigation of its source should be studied in future.

REFERENCES

[1] D. Lambrecht and T. Kulig, ‘‘Torsional performance of turbine generator shafts especially under resonant excitation’’, *IEEE Transactions on Power Apparatus and Systems*, Vol. PAS-101, No. 10, October 1982

[2] J. Song-Manguelle, C. Sihler, S. Schramm, ‘‘A General Approach of Damping Torsional Resonance Modes in Multi-megawatt Applications’’, *IEEE Transaction on Industry Application*, Vol. 47, No. 3, pp. 1390-1399, May-June 2011.

[3] J.I. Tsai, C.H. Lin, T.P. Tsao, ‘‘Assessment of long-term life expenditure for steam turbine shafts due to non-characteristic subharmonic currents in asynchronous links’’, *IEEE Transactions on Power Systems*, Vol. 19, No. 1, Pp. 507 - 516, 2004

[4] W.M. Lin, C.C. Tsai, C.H. Lin, T.P. Tsao, ‘‘Alleviating shaft torsional vibrations caused by electric arc furnaces for a low capacity turbine generator by using a flywheel coupler’’, *IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, pp. 1771-1775, 2010

[5] P. Solanics, K. Kozminski, M. Bajpai, J. Esztergalyos, E. Fennell, J. Gardell, C. Mozina, S. Patel, A. Pierce, V. Skendzic, P. Waudby, and J. Williams, ‘‘The Impact of Large Steel Mill Loads on Power Generating Units’’, *IEEE Transactions on Power Delivery*, Vol. 15, No. 1, pp. 24-30, January 2000

[6] D.G. Ramey, A.C. Sismour, G.C. Kung, ‘‘Important Parameters in Considering Transient Torques on Turbine-Generator Shaft Systems’’, *IEEE Transactions on Power Apparatus and Systems*, Vol. PAS-99, Issue: 1, pp. 311-317, 1980

[7] T. Zheng, E.B. Makram, ‘‘An Adaptive Arc Furnace Model’’, *IEEE Transactions on Power Delivery*, Vol. 15, No. 3, pp. 931-939, July 2000

[8] A. Robert, N. Janssens, M. Couvreur, P. Ledoux, G. Boon, H. Demaret, L.Z. Nedden, J. Deuse, J. Dubois, ‘‘Electromechanical Oscillations and Torsional Stresses of Generating Units near Arc Furnaces’’, cigre, http://www. cigre.org, 1998

[9] P. Kundur, *Power System Stability and Control*. New York: McGraw-Hill, 1994

[10] I. Vervenne, K.V. Reusel, R Belmans, ‘‘Electric Arc Furnace Modelling from a ‘‘Power Quality’’ Point of View’’, *3rd IEEE Benelux Young Researchers Symposium in Electrical Power Engineering*, Ghent, Belgium, April 2006

[11] E. Acha, A Semlyen, N. Rajakovic, ‘‘A Harmonic Domain Computational Package for Nonlinear Problems and its application to electric Arcs’’, *IEEE Transactions on Power Delivery*, Vol. 5, No. 3, pp. 1390-1397, July 1990

Vulnerable Operation of Turbine-Generator Due to EAF Loads in Bangladesh Power System

A. Hasib Chowdhury, Md. Minarul Islam*

Department of EEE, Bangladesh University of Engineering and Technology
Dhaka-1205, Bangladesh
*minarbuet@gmail.com

Abstract—**Steel mills in Bangladesh are concentrated in the Chittagong region. Some of these steel mills are installing electric arc furnace (EAF). EAFs generate sub-synchronous frequencies (SSFs) which may pose threat to nearby turbine-generator shaft by causing torsional resonance (TR). This paper presents an investigation into TR due to EAF operation in the Chittagong region of Bangladesh Power System (BPS). Multiple AC EAFs installed in two steel mills in close proximity has been considered. In the mathematical model of the EAF, electric arc resistances are varied deterministically in the frequency range of 1.2-45.5 Hz to meet the characteristics of realistic EAF. FFT technique has been used to find the torsional mode frequencies of the T-G shaft. Simulations show considerable TOTs amplification due to EAF operation. Simulations show possible T-G shaft damage in the Raozan and Shikalbaha power plants in the long run.**

Index Terms—**Electric Arc Furnace, Air-gap torque, Torsional resonance, Shaft fatigue, Bangladesh power system**

I. Introduction

The electric arc furnaces (EAF) may cause power quality problems in subsynchronous frequency (SSF) range, which is most prone to induce pulsating components in the air-gap torque of synchronous machine and through which the torsional oscillating torques (TOTs) are amplified during torsional resonance (TR) [1]-[4]. The TR imposes considerable stress on turbine shaft and blades [2]. The magnitude of torsional resonance depends on the proximity of stimulating frequency and natural frequency, on the magnitude and duration of the excitation and on the damping of the excited natural mode. Higher frequency components in air-gap torque have dominant effects on small components such as turbine blades and lower frequency components stimulate the TOTs in the main shaft sections [5]. Furthermore, unbalanced currents produced from EAF can contribute to generator rotor heating [3], [4].

TOTs amplification can lead to gear damage, shaft fatigue lifetime reduction, severe damages to turbine-generator (T-G) shaft [1]. The first and second T-G shafts failure due to TR were experienced at the Mohave plant in sourthern California [7]. Hence, interaction among EAF, other rotating electrical loads and T-G unit needs to be investigated with great importance.

Steel mills in Bangladesh are installing EAFs for better quality and larger quantity of steel production. The impact of steel mills, operation with EAF loads, on generating unit in Bangladesh Power System (BPS) has not been studied in details and reported yet. This paper presents the assessment of the impact of EAF operation on multi-stage T-G shaft in the Chittagong region of BPS. The TOT amplifications are quantified and its severity due to operation of additional EAFs are calculated.

II. Component Modelling

The major components such as electric arc furnace and air-gap torque of synchronous generator are modeled in this section.

A. Electric Arc Furnace

A precise model for three-phase arc furnace is required to analyse its characteristics. Because of its unpredictatable characteristics, it is very difficult to model EAF load appropriately. Different EAF models are available [8]-[9] and among them a dynamic model based on the principle of energy conservation of electric arc has been considered for the simulation of EAF. This model, described in [9], is expressed by equations (1) and (2), where i is EAF load current, v is arc voltage, r is arc radius, and C_1, C_2, C_3 are constants.

$$C_1 r^n + C_2 r \frac{dr}{dt} = \frac{C_3}{r^{m+2}} i^2 \tag{1}$$

$$v = \frac{C_3}{r^{m+2}} i \tag{2}$$

Fig. 1 shows the simulation output of current waveform in phase-a at the primary side of the furnace transformer. The waveform is modulated and can be expressed by,

$$i_a(t) = \sqrt{2} I_{rms}[1 + \sum I_n \sin(2\pi f_n t + \phi_n)]\sin(2\pi f_s t + \beta) \tag{3}$$

where, f_s is system fundamental frequency, I_n is amplitude of modulating current, f_n is frequency of the modulating current, β is angle between EAF feeder voltage and current. The frequency f_n could randomly spread all over the subsynchronous range. The frequency spectrum of the current gives f_s+f_n and f_s-f_n, which are defined as super and SSFs respectively. These frequencies spread all over the network. Hence an EAF can be considered as current source with SSFs. This current may disturb the stator current that may cause perturbation in generator air-gap torque.

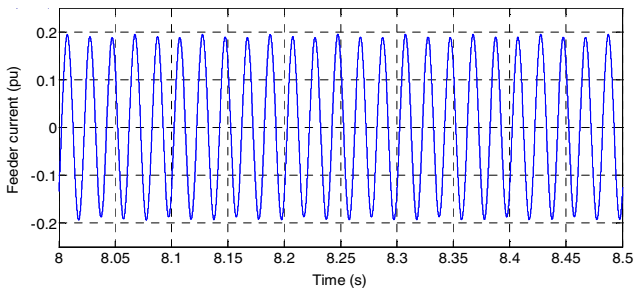


Fig. 1. Feeder current i_a at the primary side of furnace transformer

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

B. Air-Gap Torque of Synchronous Generator

The air-gap torque of synchronous generator, in *dq0* reference frame, can be expressed by equation (4).

$$T_e = \psi_d i_q - \psi_q i_d \tag{4}$$

where,

$$i_d = f(i_a, i_b, i_c)$$
$$i_q = f(i_a, i_b, i_c)$$
$$\psi_d = f(i_d, i_{fd}, i_{kd}, L)$$
$$\psi_q = f(i_q, i_{kq}, L)$$

L includes leakage, self and mutual inductances. Hence, air-gap torque is function of *dq*-axis flux linkages and currents. At synchronous operation, *dq*-axis parameters appeares as constant and results in no change in T_e. But if there are any changes or disturbances in those parameters, T_e will be disturbed which will result in perturbation in TOTs.

III. Application Area In BPS

A. Steel Mills in Chittagong

There are many steel mills in Chittagong region listed in Table I with their electrical load capacity. It is seen that three steel mills are using EAFs. Among them, BSRM and MKSBSL are using AC EAFs. To see the impacts on T-G shaft, two AC EAFs have been modeled for simulation. BSRM closer to Raozan power plant, is connected to grid at Hathazari bus and MKSBSL closer to Sikalbaha power plant, is connected to grid at Modern bus.

TABLE I
List Of Steel Mills In Chittagong Region

Steel mill name	Voltage level at PCC (kV)	Furnace type	Furnace load (MW)
H-steel	132	Induction + ladle	33
Sitalpur	132	Induction	10
BSRM	230	EAF + Ladle	87+19
MKSBSL	132	EAF + Ladle	13.6
Abul Khair	230	DC EAF + Ladle	140 + 24
Modern	132	Induction	30

B. Description of BPS Application Area

Fig. 2 shows the BPS grid network in the Chittagong region. There are three large power stations, namely Kaptai Hydro, Raozan Steam Turbine and Sikalbaha Steam and Gas Turbine Power Plants. Following assumptions/simplification have been made for the simulation.

- Two generating units at Raozan considered as a single unit of equal capacity.
- Two units at Sikalbaha considered as a single unit of equal capacity.
- Five units at Kaptai considered as one equivalent unit of equal capacity. Since impact of SSF in the slow speed water turbine is nominal, no investigation is carried out for these units.
- The physical parameters, *H*, *K* and *D* of T-G units, shown in Table II, are taken from different sources and IEEE(R) benchmark for SSR study [11].
- Five stages, HP, IP, LP_B, LP_A and generator, are considered for generating units at Raozan and Sikalbaha.
- Fixed excitation is considered for steam T-G units and controlled excitation is considered for hydro-T-G units.

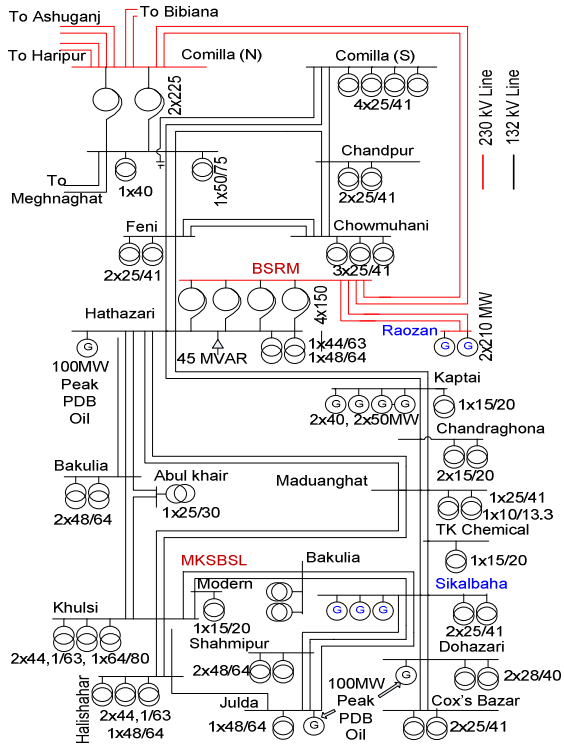


Fig. 2. 230 kV and 132 kV power grid of BPS in the Chittagong region [10]

TABLE II
Data For Raozan And Sikalbaha Generating Units

Mass	*H* (MW.s/MVA)	*K* (pu torque/elect.rad)	*D* (pu torque/pu speed deviation)
Generaotor	0.8	-	-
LP_A	1.116	51.92	2.4832
LP_B	1.081	63	0.40
IP	0.217	40.33	0.40
HP	0.117	18.17	0.32

The generator electrical parameters and transmission line data are given in Tables III to V. These data are available from the grid operator\ Power Grid Company of Bangladesh (PGCB) and the Bangladesh Power Development Board (BPDB).

TABLE III
Generator Data For Raozan Plant

Parameters	Value (PU)	Parameters	Value
Xd	1.83	*Xl*	0.1 pu
Xd'	0.219	R_a	0.003 pu
Xd''	0.133	*Tdo'*	5.5 s
Xq	1.65	*Tdo''*	0.163 s
Xq'	0.197	*Tqo'*	0.849 s
Xq''	0.12	*Tqo''*	0.016 s
Rotor type	Round rotor	No. of Pole	2

TABLE IV
Generator Data For Sikalabaha Plant

Parameters	Value (PU)	Parameters	Value
Xd	1.94	*Xl*	0.1 pu
Xd'	0.219	R_a	0.003 pu
Xd''	0.128	*Tdo'*	4.5 s
Xq	1.94	*Tdo''*	0.05 s
Xq'	0.27	*Tqo'*	1.041 s
Xq''	0.128	*Tqo''*	0.03 s
Rotor type	round rotor	No. of Pole	2

TABLE V
LINE IMPEDANCE DATA

Voltage level (kV)	Positive sequence components			Zero sequence components		
	R (Ω/km)	*L* (H/km)	*C* (F/km)	*R* (Ω/km)	*L* (H/km)	*C* (F/km)
230	Comilla-North to Hathazati substation					
	1.92E-2	4.22E-4	3.67E-8	4.36E-2	8.21E-4	2.18E-7
	Raozan power plant to Hathazari substation					
	1.74E-2	3.16E-4	3.68E-08	1.10E-1	1.16E-3	1.95E-8
132	1.01E-1	1.31E-3	1.33E-8	3.78E-1	3.93E-3	8.96E-9

IV. SIMULATION RESULTS

The impact of EAF loads on generating units at Raozan and Sikalbaha power plants have been investigated. For the purpose of analysis, EAFs have been considered to produce SSFs equal to torsional mode frequencies.

A. Torsional Mode Frequencies of T-G Shaft

FFT has been used to compute modes and its frequencies of TOT signals. As an example, FFT of torque signals between IP-HP section has been shown in Fig. 3 from which it is seen that the torsional mode frequencies are 14, 20.7, 26.1 and 37.9 Hz which are termed as mode 1, 2, 3 and 4 respectively.

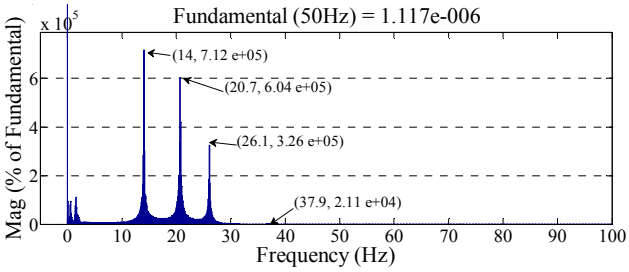


Fig. 3. FFT of torsional torque between IP-HP section in Raozan plant

B. Air-gap Torque

Fig. 4 shows FFTs of air-gap torque with EAF in the system. If the SSFs present in air-gap torque match frequencies of torsional mode, TOTs will be amplified. For the impact analysis, signals of SSFs such as 14, 20.7, 26.1 Hz have been generated by EAFs. Fig. 4 shows the presence of those frequencies in the air-gap torque.

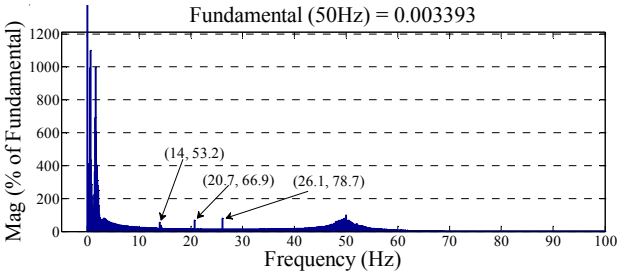


Fig. 4. FFT of air-gap torque with EAF in the system

C. Torsional Torques

Resonant excitation of torsional modes and the amplification of TOTs in the IP-HP section at Raozan and Shikalbaha plants are presented. Figs. 5(a) and 6(a) show that torsional torque transmitted between IP-HP sections of T-G shaft, when EAFs cause no TR, reach an almost steady value after some time of starting generator. On the other hand, Figs. 5(b) and 6(b) show the transmitted torque between IP-HP sections when EAFs cause TR. It is seen that amplitudes of TOTs are amplified abberently and that amplification at Raozan plant is greater than that at Shikalbaha plant. TOTs amplifiation in other sections of the T-G shaft show the same characteristics.

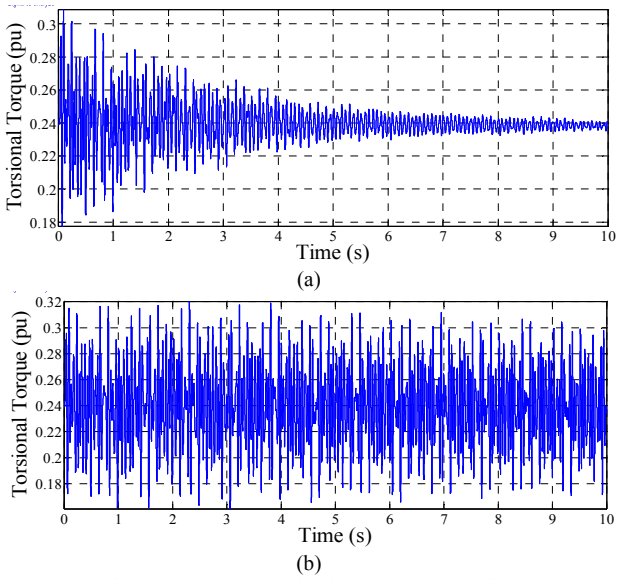


Fig. 5. Torsional torque in IP-HP section in Raozan power plant (a) no resonance (b) at resonance

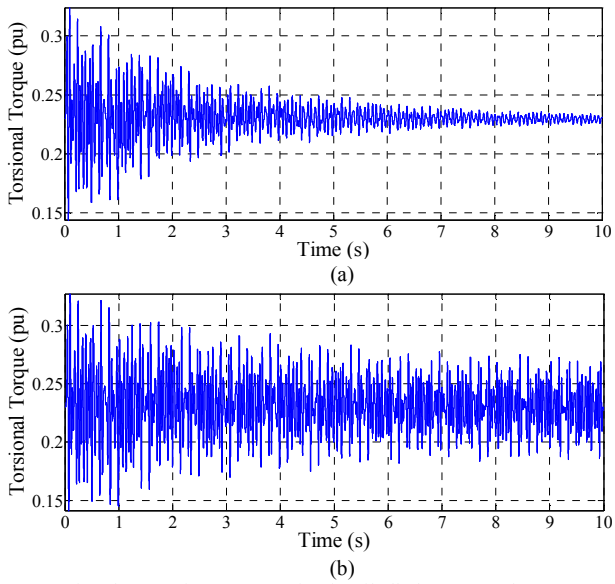


Fig. 6. Torsional torque in IP-HP section in Sikalbaha power plant (a) no resonance (b) at resonance

V. QUANTIFICATION OF TORSIONAL TORQUE

A. Torque Amplification with Existing EAFs

Magnitude of torsional torques have been quantified using FFT of TOT signals. Then, torsional torques at resonance have been compared with that of torques at nonresonance. Table VI and VII show the amount of torsional torque amplification due to TR at Raozan and Sikalbaha plants, respectively.

TABLE VI
TOTS AMPLIFICATION IN RAOZAN PLANT WITH EXISTING EAFS

Mode / Torque ratio between sections	0	1	2	3	4
HP-IP	.999	2.75	4.26	5.20	.923
IP-LP_B	1.00	2.75	4.26	5.21	.936
LP_B-LP_A	.998	2.76	4.16	5.21	1.34
LP_A-Gen	.999	2.75	4.26	5.19	.779

TABLE VII
TOTs AMPLIFICATION IN SIKALBAHA WITH EXISTING EAFs

Mode / Torque ratio between sections	0	1	2	3	4
HP-IP	.998	2.35	2.52	2.54	1.02
IP-LP_B	1.00	2.35	2.52	2.53	.778
LP_B-LP_A	.996	2.35	2.47	2.55	1.37
LP_A-Gen	.996	2.35	2.52	2.54	.262

From Table VI, it is seen, for example, that the amount of average amplified TOT due to resonant excitation of mode 2 is 4.26 times of its non-resonant TOT. From Table VII, it is seen that mode 2 is amplified, on an average, 2.52 times of its non-resonant value.

B. Torque Amplification with Additional EAFs

Considering future installations, four AC EAFs of equal capacity (100 MVA each) with similar characteristics have been assumed to be connected to the grid of BPS-Chittagong region: one at Abul Khair, one at Modern and the other two at Khulsi grid. TOTs at resonance are compared with that at nonresonance. Table VIII and IX show the amount of TOT amplification due to TR at Raozan and Sikalbaha plants, respectively.

TABLE VIII
TOTS AMPLIFICATION IN RAOZAN POWER PLANT WITH ADDITIONAL EAFS

Mode / Torque ratio between sections	0	1	2	3	4
HP-IP	1.0	5.74	8.66	10	1.09
IP-LP_B	1.0	5.83	8.70	10	1.13
LP_B-LP_A	1.0	5.81	8.43	10	2.44
LP_A-Gen	1.0	5.79	8.64	9.98	2.77

TABLE IX
TOTs AMPLIFICATION IN SIKALBAHA WITH ADDITIONAL EAFs

Mode / Torque ratio between sections	0	1	2	3	4
HP-IP	1.01	5.98	8.41	10	1.14
IP-LP_B	1.0	5.97	8.40	10	1.11
LP_B-LP_A	1.0	5.98	8.35	10	2.11
LP_A-Gen	1.0	5.98	8.37	10	2.25

Table VIII shows that when mode 3 is excited the TOT is amplified 10 times of its non-resonant value. Table IX shows similar amount of TOTs amplification in the T-G shaft at Sikalbaha plant. It can be seen that higher the mode is excited, higher the amplification occurs. TOT amplification with additional EAFs has been compared with that of existing EAFs in Table X which shows TOTs with additional EAFs are greater.

TABLE X
COMPARISON OF TOTS AMPLIFICATION BETWEEN SYSTEMS WITH ADDITIONAL EAFS AND EXISTING EAFS

Power plant	Mode	Torque amplification at present	Torque amplification with additional EAFs	Torque increment (%)
Raozan	1	2.7525	5.7925	110.5
	2	4.2350	8.6075	103.3
	3	5.2025	9.9950	92.1
Sikalbaha	1	2.3500	5.9775	154.4
	2	2.5075	8.3825	234.3
	3	2.5400	10.0000	293.7

If there are no protection schemes, the shaft of the T-G units at Raozan and Sikalbaha may not be broken/cracked immediately due to torsional resonance but will lose fatigue life. In the long run, the shaft system may get damaged.

VI. CONCLUSION

Electric arc furnaces in BPS are operating without prior study of their impact on T-G unit. EAF cause TR in T-G shaft. This paper focused on identification of occurance of TR. The three torsional modes of all generators at Raozan and Sikalbaha plants were excited by the simultaneous operation of EAFs. It is observed that the T-Gs in the Raozan plant faces a greater impact than those at Sikalbaha plant. Impacts of simultaneous operation of aditional EAFs in the region is also observed. Since the physical parameters of the turbines are not available and the data used in simulation are collected from different sources, the actual torsional mode frequencies in BPS may be different. To find the actual mode frequencies, acurate parameters of turbine section are required. The SSFs produced by EAFs may also be obtained from the field measurement data. Since there are no TR protection schemes in the BPS at present, it can be presumed that the T-G shafts at Raozan and Sikalbaha power plants may not get broken/cracked immediately due to TR, but will lose their fatigue life. In the long run, the shaft systems may get damaged.

REFERENCES

[1] J. Song-Manguelle, C. Sihler, S. Schramm, ‘‘A General Approach of Damping Torsional Resonance Modes in Multi-megawatt Applications’’, *IEEE Transaction on Industry Application*, Vol. 47, No. 3, pp. 1390-1399, May-June 2011.

[2] T.P. Tsao, J.I. Tsai, ‘‘Torsional interactions between an electrical arc furnace load and a turbine-generator set’’, *Proceedings of the 2004 IEEE International Conference on Electric Utility Deregulation, Restructuring and Power Technologies (DRPT)*, Vol. 2, pp. 627-632, 2004

[3] W.M. Lin, C.C. Tsai, C.H. Lin, T.P. Tsao, ‘‘Alleviating shaft torsional vibrations caused by electric arc furnaces for a low capacity turbine generator by using a flywheel coupler’’, *IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, pp. 1771-1775, 2010

[4] P. Solanics, K. Kozminski, M. Bajpai, J. Esztergalyos, E. Fennell, J. Gardell, C. Mozina, S. Patel, A. Pierce, V. Skendzic, P. Waudby, and J. Williams, ‘‘The Impact of Large Steel Mill Loads on Power Generating Units’’, *IEEE Transactions on Power Delivery*, Vol. 15, No. 1, pp. 24-30, January 2000

[5] D.G. Ramey, A.C. Sismour, G.C. Kung, ‘‘Important Parameters in Considering Transient Torques on Turbine-Generator Shaft Systems’’, *IEEE Transactions on Power Apparatus and Systems*, Vol. PAS-99, Issue: 1, pp. 311-317, 1980

[6] T. Zheng, E.B. Makram, ‘‘An Adaptive Arc Furnace Model’’, *IEEE Transactions on Power Delivery*, Vol. 15, No. 3, pp. 931-939, July 2000

[7] P. Kundur, *Power System Stability and Control*. New York: McGraw-Hill, 1994

[8] R. Collantes-Bellido, T. Gomez, ‘‘Identification and Modelling of a Three Phase Arc Furnace for Voltage Disturbance Simulation’’, *IEEE Transactions on Power Delivery*, Vol. 12, No. 4, pp. 1812-1817, October 1997

[9] E. Acha, A Semlyen, N. Rajakovic, ‘‘A Harmonic Domain Computational Package for Nonlinear Problems and its application to electric Arcs’’, *IEEE Transactions on Power Delivery*, Vol. 5, No. 3, pp. 1390-1397, July 1990

[10] Power Grid Company of Bangladesh (PGCB) Ltd, ‘‘Existing 230 kV and 132 kV grid network of Bangladesh’’, Available: http//www.pgcb.org.bd

[11] IEEE Working Group on Prime Mover and Energy Supply Models for System Dynamic Performance Studies, "Hydraulic Turbine and Turbine Control Models for Dynamic Studies," *IEEE Transactions on Power Systems*, Vol.7, No.1, pp. 167-179, February 1992

Effect of Intermediate Substation Configurations on Reliability of Power Transmission Corridor

A. Hasib Chowdhury, Partha Sarathi Sarker*, K.M. Rafiqul Islam, Syeda Tahsina Tanjil

Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology
Dhaka-1205, Bangladesh
*psseee@gmail.com

Abstract— **Reliability of power transmission corridor is of great importance for the reliability of the bulk power system. Any intermediate substation in the transmission corridor affects its reliability. Studies have been done on reliability of substation in isolation but its effect on transmission corridor has not been studied. Studies on transmission line reliability have also been carried out, but they focused on system connectivity neglecting reaction of system protection. This paper studies the reliability of a transmission corridor with intermediate substation having different configurations. Event tree analysis has been adopted for the purpose. The effect of substation configuration on reliability of the transmission corridor, both qualitatively and quantitatively, is analysed and compared.**

Index Terms—**Transmission corridor reliability, Substation configuration, Protection failure.**

I. Introduction

The main objective of a power system is to satisfy the load requirement continuously with reasonable assurance of quality. The ability to provide power continuously is ascertained by the term reliability. Generally power is generated at one place and then transmitted through transmission corridor to another place. Needs may arise to break in the transmission corridor and feed an intermediate substation. In that case the reliability of the transmission corridor will depend on the intermediate substation reliability. Whenever a fault occurs in a substation, the protection system isolates the faulty part. But there is a possibility that protection system may fail to react properly. As a result this protection failure may lead to cascading failures of the other components in the substation and eventually fault will propagate to the transmission corridor which will result in transmission failure through the corridor. In most of the traditional reliability studies, the assumption of protection systems being perfect, makes the analysis and calculations much easier, but may lead to unrealistic results. Reality has shown that failures of the protection can lead to serious outages of the substation [1] and protection failures have critical impact on the reliability of power system [2]. Therefore, the reliability of the corridor with protection failures in the substations is a major concern. Two main types of protection failures can happen [3]. One type of failure is the protection system's failure. This type of failure mainly refers to the situation where the protection system fails to see the fault within its operating zone or it fails to give a tripping signal to the circuit breakers. The other type of protection failure is that the circuit breaker cannot respond to the command in time.

This paper studies the effects of intermediate substation configuration subjected to protection failure on the reliability of power transmission corridor. It is assumed that the initial fault occurs on the transmission line/cable connected in the substation or the initial fault occurs on the substation components. It is further assumed that the protection system of the substation fails to operate properly. Different methods have been presented for the reliability assessment of power systems with protection failures. In this paper, the main method that is used to study the reliability of transmission corridor with intermediate substation is inspired by paper [3] and [4], which is the event tree.

II. Reliability of Power Transmission Corridor

A. Transmission Corridor Model

A power transmission corridor generally transmits electricity from a generation rich area to a load rich area. It may consist of overhead transmission line or underground cables or both. Transmission corridor with double circuit line has greater reliability. Bulk consumers generally prefer to get electricity from nearest part of the grid through line in line out (LILO) connection to avoid requiring long radial transmission lines. Any substation added to a transmission corridor also affects its reliability. A general arrangement of a single circuit LILO connection is shown in Fig. 1.

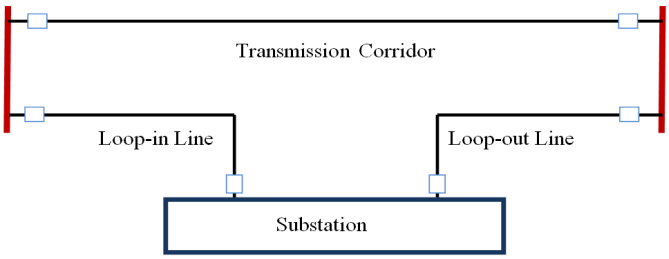


Fig. 1 Single circuit LILO connection with the consumer substation.

B. Technique to Assess Transmission Corridor Reliability

The reliability of a power system has two different aspects: system adequacy and system security [5]. System adequacy is assessed by static reliability, while system security is assessed by dynamic reliability. Most of the studies about power system reliability, are in the static reliability field, i.e. they only consider the connectivity of the system and does not consider disturbances such as load fluctuation and component outages exist in real world [6].

The post-fault phenomena in a substation are dynamic and are usually involved with the connectivity between the energy source and the load. These post-fault phenomena can be very complex depending on the system structure. To involve the system dynamic behaviour, such as the protection system response after a component failure, dynamic reliability evaluation must be applied [6].

There are two steps in dynamic reliability assessment technique. The first step is to select the state and the second

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

step involves calculating the state probability. These are explained below.

1) State Selection: State commonly refers to either the present condition of a system or entity, or to a governed entity. Effect of a unit on system reliability depends on its capacity and the probability of residing in each state. The states of a unit can be classified as shown in Fig. 2.

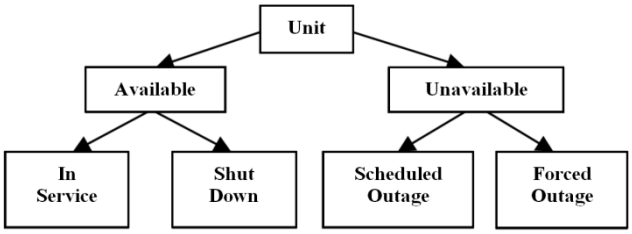


Fig. 2 Power system unit states.

There are two approaches of state selection: state enumeration and Monte Carlo simulation. State enumeration is more effective for a system with a relatively small number of components and/or low failure probabilities of components [7]. It involves- (a) enumeration of contingencies, including both circuit outages and unit outages, and (b) enumeration of electric load levels according to a specified electric load model. The objective of contingency enumeration is to identify the contingencies which may lead to unavailability [7].

The unavailability of a component is given by [8],

$$U = \frac{f \times MTTR}{8760} \tag{1}$$

where f is failure frequency and $MTTR$ is mean time to repair. A zone comprises of more than one component, failure in one or more components will cause the same effect. So, the components can be considered as series connected and unavailability of a zone comprised of such components can be calculated by [8],

$$U(unit) = \sum U(component) - \prod U(component) \tag{2}$$

2) State Probability Calculation: There are two main methods of probability calculation. They are fault tree analysis (FTA) and event tree analysis (ETA). FTA can build a diagram of all the elements that may contribute to a system failure, and then trace back the groups of elements that will necessarily lead to the system's undesired situation [9]. In the FTA, end result is the analysis starting point. Here, event is first defined for which causes are to be resolved. Causes are the starting points. Event is resolved into its immediate and causal event. Starting event is related to the causal events using appropriate Boolean logic. Stepwise resolution proceeds in this manner until primary causes are identified. Being an inductive analytical diagram, the event tree presents the system possible outcomes after the injection of an initiating event. In ETA initiating event such as loss or component failure is first defined which can have undesired consequences. Subsequent events are identified which define possible progressions of the initiating event. It displays sequence progression, end states and sequence dependencies across time. When applying the FTA in the reliability evaluation of power systems, it is mostly focused on the connectivity of the system, which is static reliability. For the reliability evaluation that includes protection failures, ETA offers a better solution.

III. Reliability of Transmission Corridor for Different Substation Configuration

A. The Considered System

In this paper three different substation configurations are considered for reliability evaluation of power transmission corridor with a single circuit LILO connection. The substation configurations are: 4/3 circuit breakers substation, one and a half circuit breakers substation and typical double busbar substation. The substation has six feeders, among which three at transmission voltage level and three fed by step down transformer. The system with three different substation configurations is shown Figs. 3 to 5.

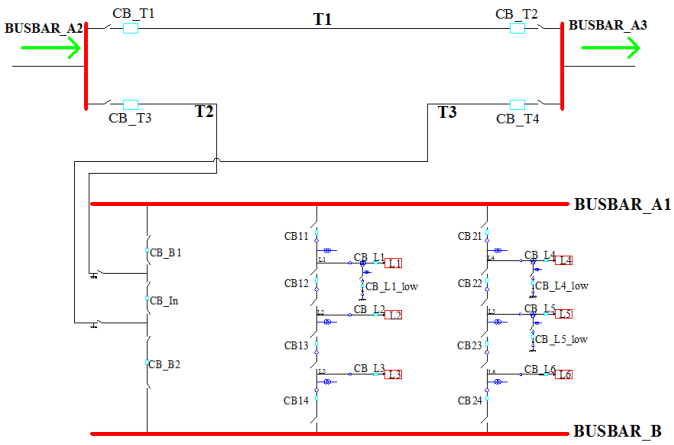


Fig. 3 Transmission corridor with intermediate 4/3 circuit breaker substation.

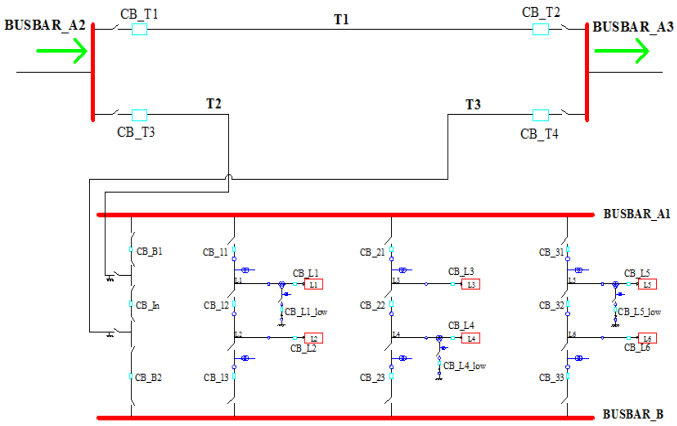


Fig. 4 Transmission corridor with intermediate one and a half circuit breaker substation.

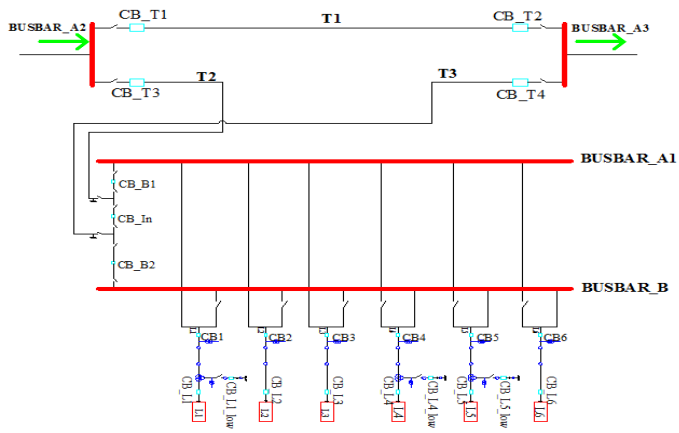
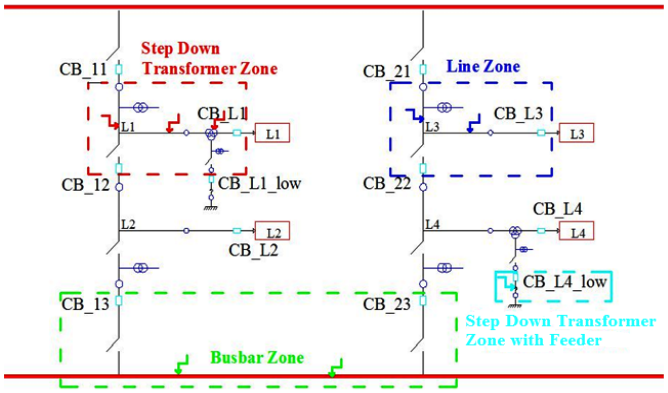


Fig. 5 Transmission corridor with intermediate typical double busbar circuit breaker substation.

B. Zone Division

Differential protection, distance protection and the busbar protection in a substation coordinate with each other to ensure overall system protection. Which of these protection systems reacts to a fault mainly depends on the fault location. The whole substation is divided into eleven zones. These eleven zones can be further classified into four types of zone: step down transformer zone, line zone, step down transformer zone with feeder and busbar zone. These four zones are shown in Fig.6 for one and a half circuit breaker substation.



ig. 6 Zone division for one and a half circuit breaker configuration.

C. Initiating Event Calculation

To calculate the resulting probabilities i.e. unavailability from the event tree, the unavailability of the initiating events must be calculated first. This can be done by using state enumeration. The failure statistics of different components of a substation used in this paper are listed in the Table I.

TABLE I
FAILURE STATISTICS OF DIFFERENT COMPONENT OF SUBSTATION

Component	Failure Frequency	Mean Time to Repair (Hour)
Overhead line	0.0031/km	8
Cable (including joints)	0.0063/km	600
Circuit breaker	0.003/comp.	24
Disconnecting/Earthing switch	0.003/comp.	8
Busbar	0.003/comp.	2
Transformer	0.05/comp.	24
Instrument transformer	0.0002/comp.	24
Surge arrestor	0.001/comp.	2
Compensating coil	0.004/comp.	3

The unavailability of a component is calculated using Eq. (1). Each substation zone consists of several components fail and failure of the components that are located in the same protection zone will cause the same effect. The unavailability calculation for L1 zone of one and a half circuit breaker substation system using Eq. (2) is shown in Table II.

TABLE II
UNAVAILABILITY CALCULATION FOR L1 ZONE

Zone Construction	Zone Data	Components Unavailability	Zone's Unavailability
CT	4	0.000000548	3.79057E-04
PT	2	0.000000548	
DS	2	0.00000274	
L1 Cable	0.5 km	0.000432	
Transformer	1	0.000137	
L1 150KV Line	0.3 km	0.00000283	
Circuit Breaker	2	0.00000822	

The unavailability of zones for the three substation configurations considered in the paper is presented in Table III.

TABLE III
UNAVAILABILITY OF ZONES FOR DIFFERENT SUBSTATION CONFIGURATION

Fault Zone	4/3 Configuration	One and a Half Configuration	Typical Configuration
L1	0.000379057	0.000379057	0.000367549
L1_Low	0.000010138	0.000010138	0.000010138
L2	0.008663564	0.008652604	0.008649864
L3	0.008661372	0.008661372	0.008649864
L4	0.000379057	0.000370289	0.000367549
L4_Low	0.000010138	0.000145768	0.000010138
L5	0.000381249	0.000379057	0.000367549
L5_Low	0.000010138	0.000010138	0.000010138
L6	0.008652604	0.008652604	0.008649864
Busbar A1	0.000033565	0.000044525	0.000033565
Busbar B	0.000033565	0.000044525	0.000044525

D. Event Tree Calculation

Event tree can be calculated using the unavailability value of the initiating event as input to the first stage which is the primary protection. The primary protection system and each circuit breaker associated with all protection stages are considered to have 0.999 and 0.9985 availability, respectively, under the condition that a fault has occurred in the system first. For each stage, success and failure is found by multiplying initiating event of that stage with the availability and unavailability of that stage, respectively. The event tree calculation for unavailability of transmission corridor with an intermediate substation having one and a half circuit breaker configuration for initiating fault at L1 zone is presented in Table IV followed by a description of the calculation process for Table IV.

TABLE IV
EVENT TREE FOR L1 ZONE OF ONE AND A HALF CONFIGURATION

Initiating Event	Protection		Success	Failure
	Availability	Unavailability		
Primary Protection				
3.79057E-04	0.999	0.001	3.78678 E-04	3.79057 E-07
Primary protection CBs (CB_11,CB_12,CB_L1,CB_L1_low)				
3.78678 E-04	0.994013487	0.005986513	3.76411 E-04	2.26696 E-06
Backup protection_1 CBs (CB_B1,CB_21,CB_31,CB_13,CB_L2)				
3.79057E-07	0.992522466	0.007477534	3.76222 E-07	2.83441 E-09
Backup protection_2 CBs (CB_22, CB_L3, CB_32, CB_L5, CB_L5_low, CB_T3, CB_B2, CB_23,CB_33,CB_In)				
2.83441E-9	0.985100846	0.014899154	2.79218 E-09	4.22303 E-11
Backup protection_3 CBs (CB_L4, CB_L4_low, CB_L6, CB_T4, CB_T1)				
4.22303E-11	0.992522466	.007477533725	4.19145 E-11	3.15778 E-13

When a fault occurs in the L1 zone, the unavailability of that zone initiates event tree and the primary protection to react. If it succeeds to react, its success initiates tripping of its associated circuit breakers CB_11, CB_12, CB_L1, CB_L1_low. If primary protection fails to operate then its failure initiates tripping of backup protection_1 circuit breakers CB_B1, CB_21, CB_31, CB_13, CB_L2. If any of

these circuit breakers fails to operate then their failure initiates backup protection_2 circuit breakers CB_22, CB_L3, CB_32, CB_L5, CB_L5_low, CB_T3, CB_B2, CB_23, CB_33, CB_In to trip. Again if any of these circuit breakers also fails then their failure initiates tripping of backup protection_3 circuit breakers CB_L4, CB_L4_low, CB_L6, CB_T4, CB_T1. Failure of a circuit breaker at this stage will lead to the fault reaching the transmission corridor and a stop in the power flow through the corridor. The failure value of the last stage is the unavailability of the transmission corridor for fault in L1 zone. Neighbouring substations are assumed to be perfect in this calculation.

Proceeding in this manner, the unavailability of the power transmission corridor for fault in each zone for all three substation configuration is calculated and presented in Table V. The last row of Table V presents transmission corridor total unavailability for the different substation configurations. Fig. 7 presents a comparison of transmission corridor unavailability for zone wise failure for all three substation configurations.

TABLE V
TRANSMISSION CORRIDOR UNAVAILABILITY WITH ZONE WISE FAILURE FOR DIFFERENT SUBSTATION CONFIGURATION

Zone	Unavailability for		
	4/3 Configuration	One and a Half Configuration	Typical Configuration
L1	2.479416E-13	3.157787E-13	4.442332E-14
L1_low	3.969823E-17	3.794823E-17	1.280807E-17
L2	2.771437E-14	5.769424E-12	1.305839E-12
L3	1.710548E-11	7.613844E-12	1.305839E-12
L4	2.656516E-13	3.698920E-13	5.548754E-14
L4_low	3.192428E-17	5.892843E-16	4.588047E-18
L5	9.153855E-16	3.157787E-13	5.548754E-14
L5_low	1.093726E-19	3.794823E-17	4.588047E-18
L6	5.057071E-12	8.643327E-12	1.565832E-12
Busbar A1	4.782894E-12	9.883873E-12	4.782894E-12
Busbar B	3.592550E-12	9.785035E-12	2.091223E-12
Total	3.108029E-11	4.269762E-11	1.120705E-11

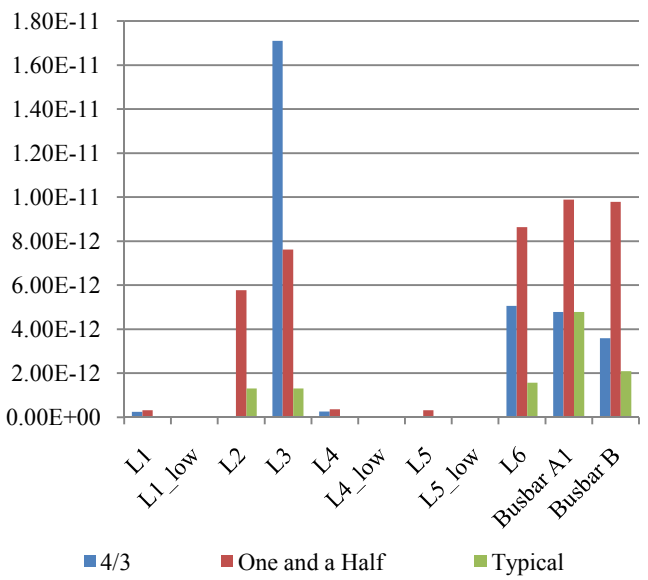


Fig. 7 Transmission corridor unavailability for zone wise failure for different substation configuration

From Table V and Fig. 7 it can be seen that for 4/3 system fault in L3 zone has the highest value of unavailability. But for most of the zones 3/2 substation configuration has the highest unavailability and typical substation configuration the lowest. It can be seen from the last row of Table V that typical substation configuration has the lowest total unavailability. That means if a fault occurs in the substation zone, the transmission corridor will be less venerable to the fault for intermediate substation having typical configuration. This is because in typical substation configuration there are less stages of protection and as a result number of associated circuit breaker is also less.

IV. CONCLUSIONS

In this paper, effect of different intermediate substation configurations on reliability of power transmission corridor subject to protection failure in the substation has been studied. Event tree analysis is used for unavailability calculation. In general, being an inductive graphical method, event tree analysis combines the calculation with system principles perfectly. The effects of substation configuration on reliability of the transmission corridor obtained both qualitatively and quantitatively.

It is observed that for greater reliability of a transmission corridor with intermediate substation typical substation configuration is the best choice. Moreover, typical substation is simple in design, cost effective and convenient to operate. There are scopes of further work including calculating the effects on power transmission corridor's reliability for different combination of circuit breaker failure for each stage of protection for both single and double LILO connection with an intermediate substation.

REFERENCES

[1] T. Johannesson, F. Roos, and S. Lindahl, "Reliability of Protection Systems – Operational Experience 1976-2002," 8th International Conference on Developments in Power System Protection, IEE, Vol. 1, pp. 303-306, April 5-8, 2004, Amsterdam, The Netherlands.

[2] P. M. Anderson, Power System Protection—Part IV: Reliability of Protection Systems, New York, USA: McGraw-Hill/IEEE Press, 1999.

[3] G. H. Kjølle, O. Gjerde, B. T. Hjartsjø, H. Engen, L. Haarla, L. Koivisto and P. Lindblad, "Protection System Faults - a Comparative Review of Fault Statistics," in 9th International Conference on Probabilistic Methods Applied to Power Systems, pp. 1-7, June 11-15, 2006, KTH, Stockholm, Sweden.

[4] L. R. Castro Ferreira, J. Pugh, P. A. Crossley, R. N. Allan and J. Goody, "Design and Reliability of Integrated Protection and Control Schemes," in Power System Control and Management, Conference Publication No. 421, pp. 87 - 91, 16-18 April 1996, London, UK.

[5] Roy Billinton, W. Li, Reliability Assessment of Electric Power Systems Using Monte Carlo Methods, New York, USA: Plenum Press, 1994.

[6] A. M. Rei, A. M. Leite da Silva, J. L. Jardim and J. C. de Oliveira Mello, "Static and Dynamic Aspects in Bulk Power System Reliability Evaluations," *IEEE Transactions on Power Systems*, vol. 15, no. 1, pp. 189-195, Feb. 2000.

[7] S. Meliopoulos, D. Taylor, C. Singh, Comprehensive Power System Reliability Assessment, PSERC Publication 05-13, April 2005.

[8] W. Li, *Risk Assessment Of Power Systems: Models, Methods, and Applications*, New York, USA: Wiley-IEEE Press, 2005.

[9] W. E. Vesely, N. H. Roberts, *Fault Tree Handbook*, Washington D.C., USA: US Nuclear Regulatory Commission, 1981.

A Voltage Flicker Severity Analysis Module for Multiple Electric Arc Furnace Operation

Abul Hasan Fahad,* Pertha Protim Dutta, and A. Hasib Chowdhury

Dept. of EEE, Bangladesh University of Engineering & Technology
Dhaka-1000, Bangladesh
*fahad.eee.buet@gmail.com

Abstract— **Electric arc furnaces (EAF) are very large electrical loads and they cause voltage flicker problem in the power system due to their non-linear and random electrical characteristics. Voltage flicker (cyclic and stochastic) causes annoying sensation experienced by human eye and is a very important power quality parameter. In this paper, a Matlab-Simulink based flicker severity analysis module has been proposed to analyse voltage flicker in the network under multiple EAF operation fed from a substation. In this module a new system-modelling strategy is proposed. The power system is modelled and simulated in Simulink and instantaneous voltage data is analysed by a separate Matlab program. With this module, a power system containing two parallel EAFs is modelled and simulated and the short-term flicker severity index of voltage at point of common coupling is calculated. This new module can be used very effectively to estimate the flicker of any system having multiple arc furnaces that are fed from a substation.**

Index Terms—**Electric arc furnace, voltage flicker, power quality, short term flicker severity**

I. Introduction

Voltage flicker is one of the important parameters of power quality and the only one based on human perception. Annoying sensation is experienced by human eye due to light flickering in lamps caused by rapid voltage fluctuations. Voltage flicker can be classified into two types- cyclic flicker and non-cyclic flicker. Cyclic flicker results from periodic voltage fluctuations whereas non-cyclic flicker involves random variations of power system voltage [1]. Potential flicker sources are- electric arc furnaces (EAFs), arc welders, start-up of large motors, compressors, power electronics devices and all rapidly-changing loads in general. Their impact is dependent on their power and location of connection to the network.

By far the largest flicker sources are EAFs which are usually connected to high voltage network. Operation of EAF imposes a serious strain to power supply network with its rapidly changing and highly irregular power consumption. Extensive studies on EAF has revealed that voltage variations as low as 0.5% in the frequency range of 6-10 Hz can cause objectionable light 'flicker' in ordinary incandescent lamps [2,3,4]. The flicker produced by EAFs depends on many different factors which may vary randomly, and exact prediction of flicker is, therefore, very difficult. Before connecting an EAF into the grid, its effect on the power system is thus difficult to estimate and consequently voltage flicker problems become unavoidable, if mitigation measures are not adopted. In most standard power system simulation software, EAF model for flicker study is not available. In this case, mathematical equations that model the behaviour of EAF and other components of the power system as accurately as possible are needed to forecast flicker in the system. There are several models of EAF described in literature for flicker study [2,5,6,8]. These models can be used to simulate the approximate real-time operating condition.

In this paper, a new flicker analysis module has been presented and a system-modelling technique has been developed to analyse flicker in a power system under multiple EAFs fed from a substation. The work is done in Simulink and Matlab. For flicker analysis, arc resistance model of the EAF has been adopted from [2]. Both cyclic and non-cyclic flicker are modelled by sinusoidal and stochastic modulation of arc resistance in the EAF. The simulation of the power system is conducted for ten minutes and voltage at the point of common coupling (PCC) is recorded at a sampling frequency of 5 kHz. The voltage data is then feed into a Matlab program which calculates short-term flicker severity (P_{st}) according to IEC flickermeter standard. Lastly, P_{st} for both cyclic and non-cyclic flicker phenomena are compared. The steps are shown in Fig.1.

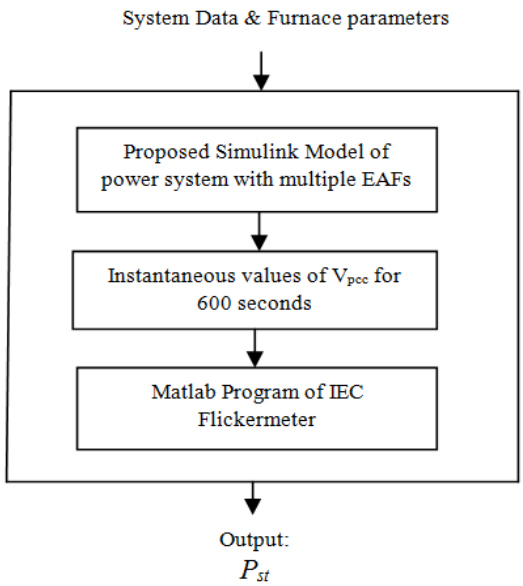


Fig. 1 Block diagram of proposed module

II. Voltage Flicker

Power system voltage varies over time due to perturbations that occur in the processes of electricity generation, transmission and distribution. Interaction of electrical loads with the network causes further deterioration of the electrical power quality. High power loads that draw fluctuating current, such as large motor drives and arc furnaces, cause low

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

frequency voltage variations i.e. voltage flicker. A voltage flicker waveform can be expressed as,

$$v(t) = \sqrt{2}\, V_{RMS}\left[1 + \frac{1}{2}\sum_n \Delta V_n \sin(2\pi f_n t + \phi_n)\right]\sin(2\pi f_s t) \quad (1)$$

Voltage flicker causes fluctuation in brightness of light sources which can cause significant physiological discomfort, physical and psychological tiredness, and even pathological effects for human beings. Flicker also causes problems with the stability of electrical devices and electronic circuits.

III. IEC FLICKERMETER

Flicker severity is expressed in terms of two parameters: short term flicker severity P_{st} and long term flicker severity P_{lt} [3]. Short term flicker severity (P_{st}) is determined by analysing the voltage at the point of common coupling. According to IEC standard 61000-4-15, which is accepted as the international standard on measuring flicker, the flickermeter is divided into 2 parts [3]:

- Task 1: Simulation of lamp-eye-brain response which is realized with the signal processing techniques such as filtering, squaring, and re-sampling
- Task 2: Statistical data evaluation of flicker level selected by CPF (cumulative probability function) as the standard defines.

Fig. 2 shows the block diagram of the IEC flickermeter. The process to calculate P_{st} is not simple and the statistical procedure is inevitable when calculating P_{st}. From the resulting momentary value of flicker the short term flicker value P_{st} is generated in the last block of Fig. 1 according to a statistical process over a predefined observation interval. Long term flicker P_{lt} is calculated as the cubic average of several P_{st} values.

The observation intervals and limiting values for P_{st} and P_{lt} as specified in IEC 61000-3-3 is presented in Table I [3].

TABLE I
ALLOWED LIMIT ON FLICKER SEVERITY VALUES

Value	Observation Interval	Limiting Value
P_{st}	10 min	1.0
P_{lt}	2 h	0.65

IV. ELECTRIC ARC FURNACE MODEL FOR FLICKER STUDY

The actual V-I characteristics and its piecewise linear model of an arc furnace is shown in Fig. 3. The arc ignition voltage V_{ig} and the arc ignition voltage V_{ex} are determined by the arc length during arc furnace operation [2]. Figure 2 shows actual EAF operating characteristics in solid line and approximate operating conditions in dotted line.

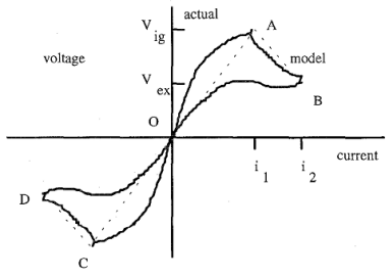


Fig. 3 Actual and piece-wise linear approximation of v-i characteristic of an arc furnace load [1]

Let R_1 and R_2 be the slopes of the lines OA and AB. Then the relationship between arc voltage and current for one cycle, according to [2,6] can be expressed as -

$$R_1 = \frac{V_{ig}}{i_1} \quad (2)$$

$$R_1 = \frac{V_{ig}^2}{P + \frac{V_{ig}^2}{R_2} - \frac{V_{ex}^2}{R_2}} \quad (3)$$

$$R_2 = \frac{V_{ex} - V_{ig}}{i_2 - i_1} \quad (4)$$

$$v(t) = \begin{cases} R_1 i(t); & 0 \le |i(t)| \le i_1 \\ R_2 i(t) + V_{ig}(1 - \frac{R_2}{R_1}); & i_1 \langle |i(t)| \langle i_2 \end{cases} \quad (5)$$

Figure 4 shows the simulated V-I characteristic of the model. As mentioned in [2], quantities V_{ig}, V_{ex}, R_2 are assumed not to change with load operating conditions and V-I characteristics are made dependent on the operating condition, P, of the arc furnace load.

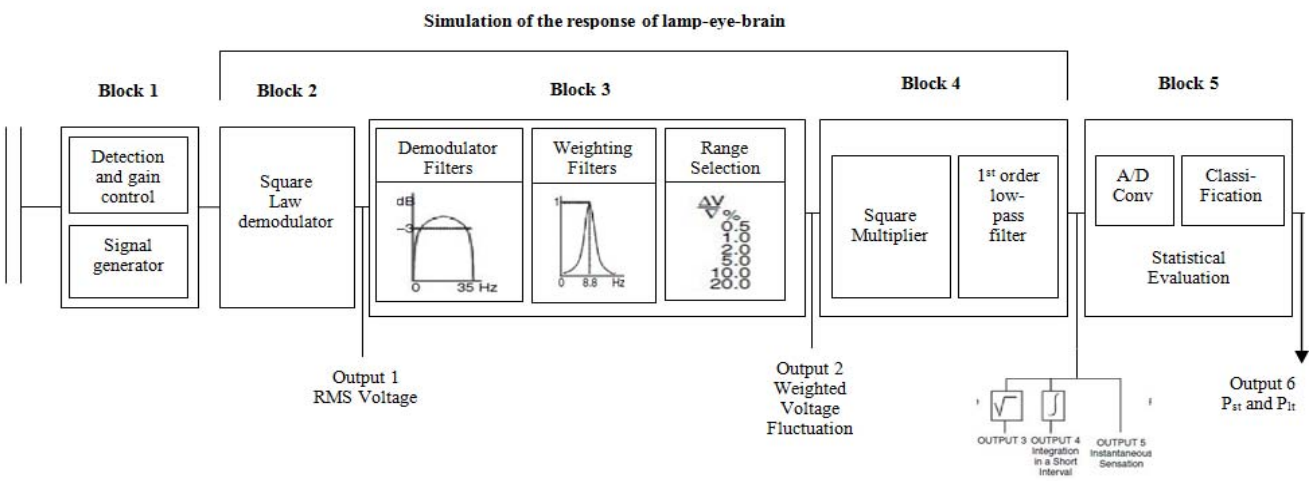


Fig. 2 Block diagram of the IEC flickermeter

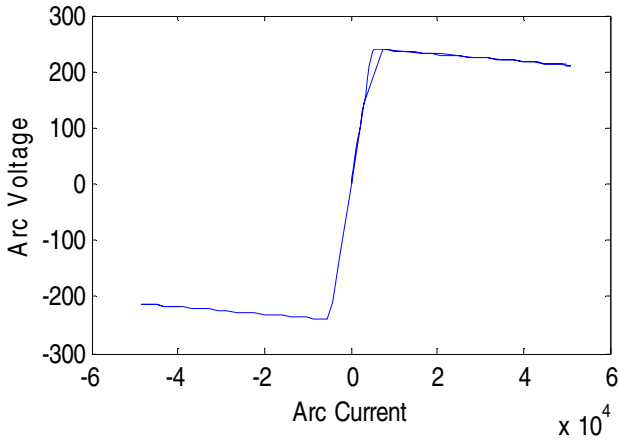


Fig. 4 Simulated V-I characteristic of the model

To analyze flicker generated by an arc furnace, the dynamic arc furnace model is needed. To generate the dynamic behaviour of the arc furnace using different models, the slope of the V-I curve in figure 4 should be changed. The change can be sinusoidal or stochastic. Stochastic modulation is done by superimposing the arc resistance of the EAF with band-limited noise of 5-12 Hz and sinusoidal modulation is done by modulating a sine wave of frequency 8.5 Hz multiplied by a "modulation index" on the constant value of arc resistance.

The arc resistance in the case of sinusoidal variation and in the case of band limited white noise variation are defined by Eqs. (6) and (7), respectively-

$$R_2(t) = R_2[1 + m\sin(\omega_f t)] \tag{6}$$

$$R_2(t) = R_2 + BLW \tag{7}$$

Where R_2 is a constant obtained from Eq. (2), ω_f is the flicker frequency, m is the modulation index and BLW is the band-limited (4-14 Hz) white noise with zero mean. The modulated value, $R_2(t)$ is such that the total input power requirement of the arc furnace is satisfied.

V. Proposed System Modelling Strategy

The power system under consideration is depicted in Fig. 5. It contains two arc furnaces, fed from a common substation transformer. The system equations for the system shown in Fig. 4 are given in equations (8) to (11).

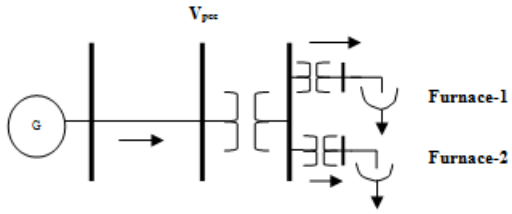


Fig. 5 Test System

$$V_{pcc} = V_1 + I_{arc,1} Z_{t1} \tag{8}$$

$$V_{pcc} = V_2 + I_{arc,1} Z_{t2} \tag{9}$$

$$I = I_{arc,1} + I_{arc,2} \tag{10}$$

$$V_{source} = V_{pcc} + IZ_s \tag{11}$$

This system of mathematical equations can be modeled in a block building fashion as shown in Fig. 6,where $Z_{t,1}$ & $Z_{t,2}$ are summation of line and transformer impedance of furnace-1 and furnace-2 respectively. Z_s is the Thevenin impedance at the source bus of the power system.

The unique aspect of the proposed model is that if the furnaces are located in nearby but different areas from the feeding point, then the change in impedance can be easily incorporated in the impedances indicated by $Z_{t,1}$ and $Z_{t,2}$.

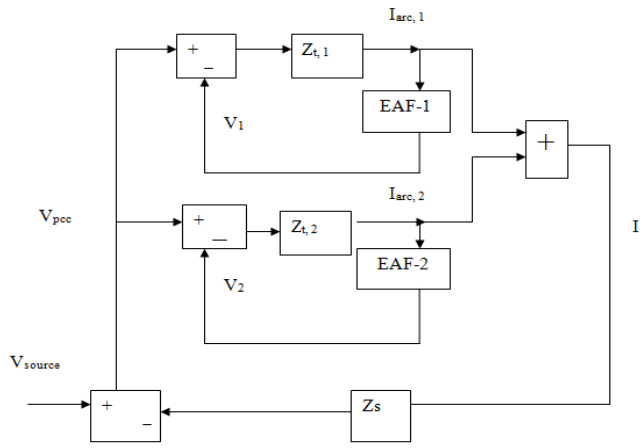


Fig. 6 Mathematical modelling of the Test System

Now, if a power system contains N EAFs fed from a substation, it can be accurately modeled and simulated by the proposed technique as shown in Fig. 7.

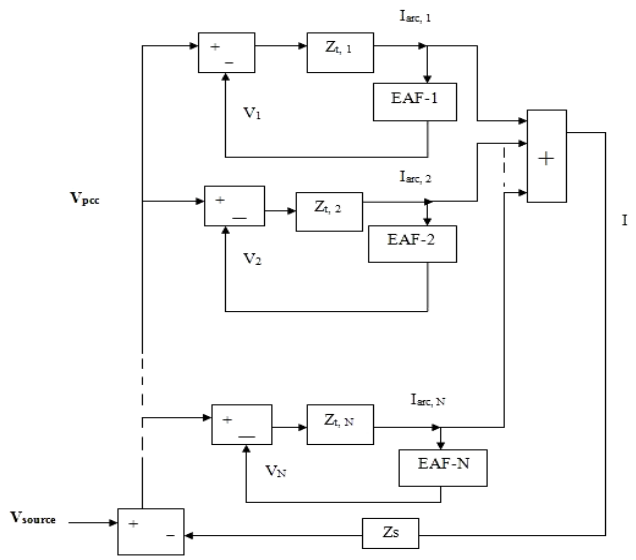
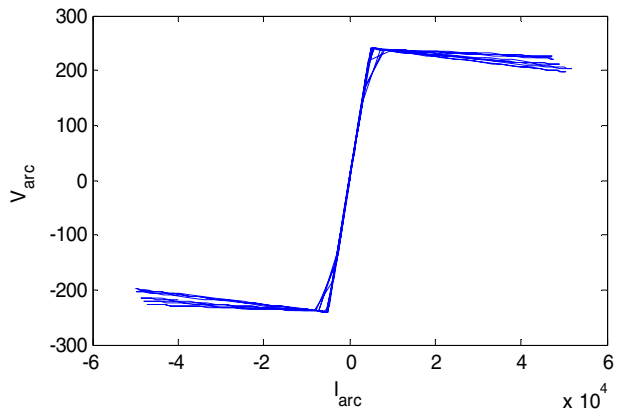


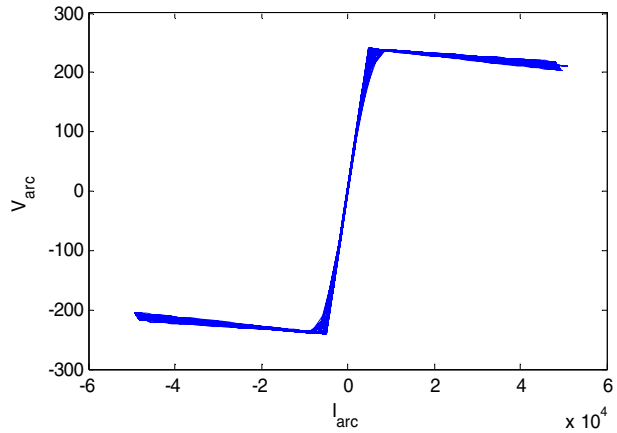
Fig. 7 Mathematical modelling of test system with N number of EAFs

VI. Simulation Results

The test system as shown Fig. 4 is simulated. The furnace parameters (V_{ig}, V_{ex}, i_1, i_2) are adopted and calculated according to Eqs. (2) to (4). Both the furnaces are assumed to have identical parameters and same power rating (87 MW). However, the system model can accommodate furnaces with different power ratings. The source impedance Z_s is 1.056e-4+j7.8e-4 ohms and impedance of each furnace transformer is taken 7.701e-4+j6.1398e-3 ohms (referred to the secondary of the furnace transformers). The simulation voltage data at PCC (V_{pcc}) is recorded and the acquired V_{pcc} data is then analysed in a Matlab program for P_{st} estimation [7]. Figures 8(a) and 8(b) show the dynamic V-I curve of the EAF under sinusoidal and stochastic modulation of the arc resistance, respectively. Figures 9(a) and 9(b) show the envelopes of the voltage at PCC with two EAFs operating simultaneously with sinusoidal and stochastic modulation of the arc resistance, respectively.

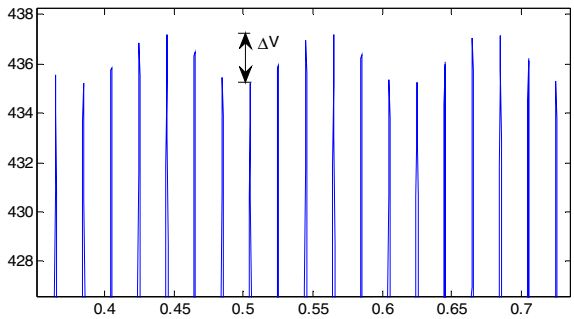


(a)

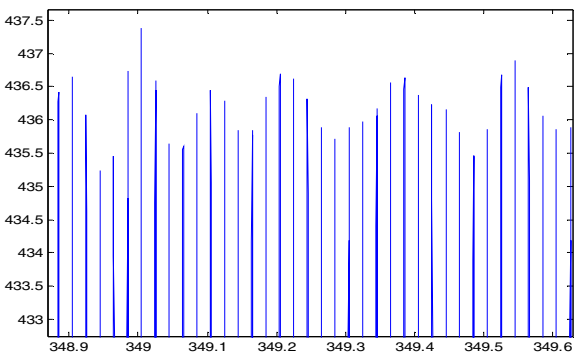


(b)

Fig. 8 Dynamic V-I curve of the EAF with arc resistance modulated (a) sinusoidally, (b) stochastically



(a)



(b)

Fig. 9 Envelope of V_{pcc} during simultaneous operation of two EAFs under (a) sinusoidal and (b) stochastic modulation of R_2

The instantaneous voltage at PCC is sampled at 5 kHz. The short term flicker severity (P_{st}) is then calculated using the flickermeter and listed in Table II.

TABLE II
SHORT TERM FLICKER SEVERITY

Furnace model	P_{st}
Sinusoidal modulation	3.3641
Stochastic modulation	1.5289

The simulation result shows that sinusoidal flicker results in higher P_{st} value than the stochastic modulation, which is in agreement with [9]. In addition, P_{st} value exceeds the limiting value of 1.0 in both cases, which indicates voltage flicker at the PCC will be perceptible to human eye.

VII. CONCLUSION

A new flicker analysis module along with a new system-modeling technique for flicker study in a power system containing multiple EAFs has been proposed. Together with the module, a representative power system has been modeled to analyze both sinusoidal and stochastic flicker in the system. IEC standard flickermeter has been used to estimate short term flicker severity index (P_{st}) of voltage at PCC. Dynamic V-I characteristics of EAF has been plotted along with voltage waveform at PCC and P_{st} obtained for both sinusoidal and stochastic flicker has been compared. It is evident that the module developed in this paper can be easily applied to predict voltage flicker in any power system under multiple electric arc furnace operation when the furnaces are fed from the same substation. A greater detailed model of EAF would give a more accurate result.

REFERENCES

[1] S.R.Mendis, M.T.Bishop, A.V.Do, D.M.Boyd, "Investigation of transmission system voltage flicker due to multiple ac and dc arc furnaces," *IEEE Trans. On Power Del.*,vol.10,no.1,Jaunary 1995

[2] S. Varadan, E.B. Makram, Adly A. Gigris, "A new time domain voltage source model for an arc furnace using EMTP," *IEEE Trans. On Power Del.*, vol 11, no. 3,pp.165-1690, July 1996

[3] S.H. Cho, J.A. Jung, G. Jang, S.H. Kwon, M.H. Kang, "Development of Matlab/Simulink Module for Voltage Flicker Simulation in Distribution Power Systems," *Journal of Electrical Engineering & Technology*, vol. 3, No. 3, pp. 314~319, 2008

[4] E.acha, A.Semlyen,N.Rajakovic, "A harmonic domain computational package for non linear problems and its application to electric arcs," *IEEE Trans. On Power Del.*,vol 5,no.3,pp 1390-1395,July 1990

[5] G.Manchur,C.C.Erven, "Development of a model for predicting flicker form electric arc furnaces," *IEEE Trans. On Power Del.*,vol.7,no.1,pp. 416-426,January 1992

[6] M. Islam, A.H. Chowdhury, "Comparison of dynamic resistance arc furnace models for flicker study," 7th *International Conference on Electrical and Computer Engineering* , December 2012, Dhaka, Bangladesh

[7] IEC Standard 61000-4-15 (Edition 1.1, 2003-02): *Electromagnetic Compatibility(EMC)–Part4 Testing and Measurement Techniques–Section15 Flickermeter: Functional and Design Specifications*

[8] O.Ozgun,A.Abur, "Flicker Study Using a Novel Arc Furnace Model," *IEEE Trans. On Power Del.*,vol.17,no.4,October 2002

[9] G. C. Montanari, M. Loggini, A. Cavallini, L. Pitti, D. Zanielli, "Arc furnace model for the study of flicker compensation in electrical networks," *IEEE Trans. On Power Del.*, vol. 8, pp. 2026–2036, October 1994.

[10] E. O'Neill-Carrillo, G. Heydt, E. J. Kostelich, S. S. Venkata, and A. Sundaram, "Nonlinear deterministic modeling of highly varying loads," *IEEE Trans. Power Del.*, vol. 14, pp. 537–542, Apr. 1999

[11] Deokar, S.A.; Waghmare, L.M.; Jadhav, G.N. "Voltage flicker assessment of induction motors used in the integrated water pumping station", *Power Electronics, Drives and Energy Systems (PEDES) & 2010 Power India, 2010 Joint International Conference on,* pp. 1 - 7

A Comparative Study of Short Term Flicker Severity Due to Electric Arc Furnace Operation

Pertha Protim Dutta,* Abul Hasan Fahad, and A.Hasib Chowdhury

Dept. of EEE, Bangladesh University of Engineering & Technology
Dhaka-1205, Bangladesh
*ppd024@gmail.com

Abstract— **Human eye is susceptible to rapid voltage fluctuations of power system, which is termed as 'voltage flicker'. Voltage flicker causes psycho-physical discomfort to humans. It is quantified by short term flicker severity index (P_{st}) according to IEC standards. Electric arc furnaces (EAF) are one of the main flicker sources in the power system. In this paper, voltage flicker caused by the operation of both single and multiple EAFs on the power system has been considered. A comparative analysis of flicker phenomena has been made by estimating P_{st} under variations of several modelling parameters and operating conditions.**

Index Terms— **Electric arc furnace, voltage flicker, power quality, short term flicker severity, band limited white noise**

I. Introduction

Power system voltage is subjected to rapid change or fluctuation due to the occurrence of disturbances in operating conditions. Voltage change is defined as a deviation in the R.M.S. voltage value with respect to a steady-state value averaged over some period of time [1]. The voltage deviations may or may not be periodical. When the fluctuations in voltage are subjected to luminosity and human eye comfort, the term 'flicker' is used and voltage flicker is classified into two types in the literature- cyclic flicker and non-cyclic flicker respectively, depending on the deviation type of voltage [2]. Flicker severity is a parameter used to quantify the intensity of flickering. Several severity indices exist in literature, and among these, short term flicker severity (P_{st}) is widely used.

In a power system, the most threatening flicker sources are electric arc furnaces (EAF). With its rapidly changing and highly irregular power-consumption characteristics, the flicker produced by electric arc furnaces deteriorates power quality of the network. Moreover, amount of flicker in the power system can vary with MW ratings, electrode geometry and thermodynamics of the EAFs. Hence, in this paper, a comparative approach is undertaken to observe flicker in the power system under several operating conditions such as- variation of number of furnaces in the system, variation of MW ratings of EAFs, relative change of the length of electrodes during arcing periods etc. For the purpose of the study, EAFs have been modelled as a source of both cyclic and non-cyclic flicker in the system. Flicker severity P_{st} has been calculated for all cases. The resultant values are then compared with allowed limits and conclusions drawn.

II. Flicker Injection by Electric Arc Furnaces

Electric arc furnace converts electrical energy to thermal energy creating arc and are widely used for smelting and refining metal scraps. As current drawn during the arcing period is extremely high (in the order of 10^4 A), the temperature rises to above 2000^0C which is enough for melting steels.

During EAF operation arcing is accomplished continuously. Electrodes are scrolled up and down in each period of arcing to vary arc length. Owing to irregular shaping of metal scraps, presence of non-homogeneity and impurities within phases and non uniformity of temperature inside furnace chamber, variation in arc length takes place from cycle to cycle of arc period. This complex metallurgical process is responsible for non-linearity in arc current which in turn, causes fluctuation in voltage at PCC and eventually leads to perceptible flicker sensation.

III. Flicker Severity and its Estimation

Flicker is basically characterized by two parameters obtained by means of the flicker measurement apparatus namely, short term flicker severity (P_{st}) and long term flicker severity (P_{lt}).

A. Short Term Flicker Severity

A unit of short term flicker severity, P_{st}, corresponds to the severity limit that should not be exceeded in order not to create malaise for an observer. P_{st} is obtained from $s(t)$ through a series of statistical evaluations as shown in Fig. 1 [5].

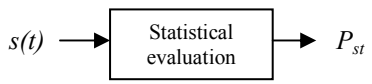


Fig. 1 Statistical analysis for P_{st} evaluation

B. Long Term Flicker Severity

The long term flicker severity, P_{lt}, is deduced from 12 consecutive values of P_{st} by Eq. (1).

$$P_{lt} = \sqrt[3]{\frac{\sum_{i=1}^{12} P_{st,i}^3}{12}} \quad (1)$$

The observation intervals and the limiting values for P_{st} and P_{lt} specified in the standard IEC 61000-3-3 [3] are presented in Table I.

TABLE I
Allowed Limit on Flicker Severity Values

Value	Observation Interval	Limiting Value
P_{st}	10 min	1.0
P_{lt}	2 h	0.65

C. Prediction of Flicker in Cyclic Flicker Phenomenon

Cyclic flicker results from periodic fluctuation of R.M.S voltage at the frequency sensitive to human perception at the point of interest. This variation may be sinusoidal, rectangular,

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

triangular etc. depending on the behavior of loads connected to system. For being periodic variation some direct equations to approximate P_{st} in cyclic flicker phenomenon have been proposed in literature.

From instantaneous flicker: In occurrence of sinusoidal fluctuation of system voltage short term flicker severity is approximated directly from instantaneous flicker sensation [4] by Eq. (2).

$$P_{st} = \sqrt{0.5096 * S_{\max}} \tag{2}$$

Where S_{max} is the maximum value of $s(t)$.

From relative voltage change: This method can be applied for rectangular voltage changes with the same magnitude and regularly distributed in time [1]. In this method P_{st} is given by

$$P_{st} = d/d_0 \tag{3}$$

where, d is relative percentage voltage change and d_0 indicates that for unity of flicker severity.

Analytical method: This method is presented in IEC 61000-3-3 [3]. It is based on calculating the flicker time, t_f, which is a magnitude with a time dimension describing the flicker impression caused by a single voltage variation. The expression of P_{st} appears as shown in Eq. (4).

$$P_{st} = \left(\frac{\Sigma t_f}{T_p}\right)^{\frac{1}{3.2}} \tag{4}$$

D. Estimation of P_{st} In Multiple EAF Operation

When two or more electric arc furnaces are operated simultaneously in a power network the cumulative P_{st} at PCC can be estimated using Eq. (5) [5].

$$P_{st,total} = \sqrt[n]{\sum_{i=1}^{m} P_{st,i}^{n}} \tag{5}$$

Where $P_{st,i}$ indicates the flicker severity due to single operation of each furnaces and n represents flicker summation factor, and it varies between 1 to 4 [6]. As the factor, n becomes smaller cumulative flicker severity at PCC becomes larger [5].

In case of parallel operation of identical EAFs (having same rating and modulation parameter), the summation factor can be determined directly using Eq. (6).

$$n = \frac{\ln(m)}{\ln(P_{st\,cum}) - \ln(P_{st\,ind})} \tag{6}$$

Where, m stands for number of furnaces, $P_{st,ind}$ denotes flicker severity due to standalone operation of each furnace.

IV. Electrical Arc Furnace Model for Flicker Study

The EAF is modelled as resistive load with different slopes during pre-arcing and arcing period for flicker analysis [7]. Hence, the EAF model appears as a piece-wise linear model. Fig. 2 shows actual EAF operating characteristics in solid line and approximate operating conditions in dotted line.

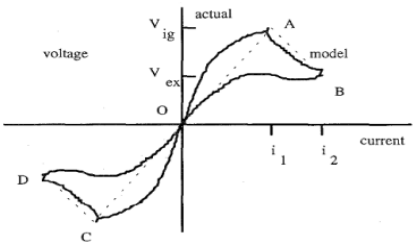


Fig. 2 Actual and piece-wise linear approximation of V-I characteristic of an arc furnace load [7]

Assuming R_1 and R_2 to be the slopes of the lines OA and AB in Fig. 2, the relationship between arc voltage and current for one cycle, is determined according to [7,8]. As voltage flicker occurs due to variation of arc length and since resistance is proportional to length of a conductor, the flicker phenomenon is simulated by varying R_2, where R_2 represents arc resistance during arcing period.

In case of cyclic flicker, fluctuation of sinusoidal system voltage is periodic, and this phenomenon is modelled by sinusoidal modulation of R_2 at flicker frequency [8]. For simulating non-cyclic flicker R_2 is varied by stochastic modulation. It is obtained by adding a band limited (5-12 Hz) Gaussian noise of zero mean with R_2 [7].

V. Test System and Modelling Parameters

In a practical system, EAFs are connected to a PCC through furnace transformers. During simulation EAFs are connected with a test system shown in Fig. 3.

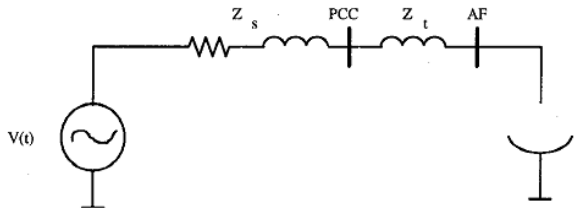


Fig. 3. A representative system for simulation [8]

In Fig. 3, Z_s includes source and transmission line impedance and Z_t accounts for the impedance of transformer. The selection of Z_t depends on the power ratings of EAFs.

For simulating the phenomena of multiple EAF operation and further comparison, two EAFs of different rating are selected. Single phase power in MW is chosen for each rating. EAF-1 is rated 29 MW and that of EAF-2 is 47 MW.

Modulation parameter plays a vital role both in cyclic and non-cyclic flicker. Choice of this parameter should be such that P_{arc} never exceeds the EAF rated power during furnace operation [7]. To investigate the effect of cyclic flicker, sinusoidal fluctuation of V_{pcc} has been considered as it occurs in most cases. In this phenomenon modulation index, m is considered to be that parameter. The value of allowable modulation index depends on system parameters.

Modulation parameter for stochastic flicker simulation is titled on power of random noise as it incorporates noise variance. Maximum allowable noise power increases with decreasing ratings of EAF. The reason lies in steeper V-I characteristics in smaller furnaces during arcing period.

VI. Results And Analysis

Simulations of the test system give the maximum allowable value of m to be 0.5 for all cases of cyclic flicker. Maximum allowable noise estimated for EAF-1 and EAF-2 is about $5.7\text{x}10^{-10}$ and $2.2\text{x}10^{-10}$, respectively. Simulation results and observations are presented in this section.

A. Variation of P_{st} With Rating of Furnaces

Simulating the representative system with single EAF short term flicker severity index is calculated for different power ratings. The result is tabulated in Table II.

It can be seen from Table II that for a fixed modulating index and modulation frequency P_{st} is larger for the furnace of higher rating. Flicker severity for non-cyclic flicker also shows similar trend.

TABLE II
VARIATION OF P_{ST} WITH EAF RATING

Furnace rating (MW)	P_{st}
29	1.9919
47	3.6751

B. Comparison of Furnace Rating with Modulation Index

For cyclic flicker a distinct feature is observed from Table III, and that is, allowable variation rate of arc length for keeping P_{st} under acceptable limit is to be reduced if an EAF of higher rating is to be used. It has been observed from simulations that similar comments can be made for parallel operation of EAFs.

TABLE III
VARIATION OF MAXIMUM VALUE OF m WITH EAF RATING

Furnace rating (MW)	Maximum value of m
29	0.25
47	0.144

C. Effect of P_{st} On Modulating Frequency

Human eye sensitivity is maximum when frequency of light flickering from incandescent lamp is between 6 Hz to 10 Hz. Variation in P_{st} with different frequencies in the flicker range is shown in Fig. 4. Furnace rated at 47 MW is chosen for the comparison and modulation index is taken as 0.5. It has been observed, as shown in Fig. 4, that P_{st} is maximum for a modulation frequency of 8.8 Hz.

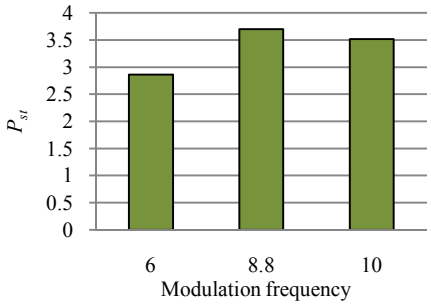


Fig. 4 Variation in P_{st} with different modulation frequencies

D. Effect On P_{st} Due to Multiple EAF Operation

Simultaneous operation of multiple arc furnaces in close proximity in a power system network causes greater severity in flicker phenomenon at PCC.

1) Sinusoidal Flicker Severity

Figure 5 gives a perceptible comparison to realise cyclic flicker severity at grid voltage owing to simultaneous operation of two EAFs over their individual operation. Flicker severity is also illustrated if ratings of EAFs connected in parallel were different.

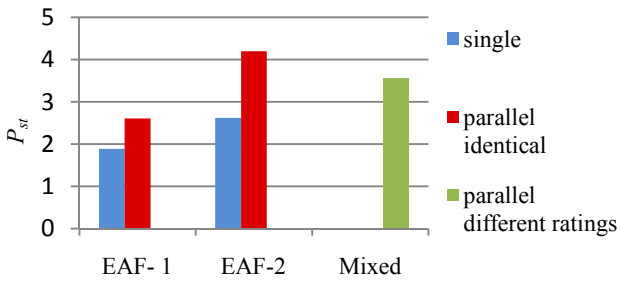


Fig. 5 Cyclic flicker severity at PCC for single and two EAFs operation

For all cases modulation index and the frequency of fluctuation is kept constant as 0.4 and 8.5 Hz respectively. Using equations (5) and (6) summation factor n for parallel EAF operation are found 1.94, 1.468 and 1.54, respectively. It implies that cumulative flicker level at PCC is more severe when two identical EAFs of smaller rating are operated simultaneously.

2) Stochastic Flicker Severity

Table IV presents the change in P_{st} due to parallel operation of two identical furnaces over their operation alone. Noise power is kept constant at maximum value for emphasizing this effect in particular. It can be seen that for single EAF operation P_{st} is within acceptable range, but it is not so for parallel operation. It is also observed that for multiple EAF operation Eq. (5) is satisfied but flicker summation factor is not fixed.

TABLE IV
P_{st} VALUES FOR SINGLE AND PARALLEL EAF OPERATION

EAF operation	P_{st}
Single	0.9371
Parallel	1.5289

E. Variation of Modulating Index in Parallel EAFs

In parallel operation of two identical EAFs, each having ratings of 29 MW, variation of modulation index m results in significant conclusion. This effect is shown in Table V.

TABLE V
EFFECT OF MODULATION INDEX ON P_{ST}

Modulation index		P_{st}
m_1	m_2	
0.145	0.145	0.9850
0.10	0.20	1.0189
0.05	0.24	0.9843

It can also be seen from Table V, from the first and last data rows, that the summation of m_1 and m_2 is constant for a given value of P_{st}. This property can be taken into account while installing a second EAF in a region in close proximity to the first EAF having identical rating. It is also notable that this rule doesn't follow if the ratings of the two EAFs are different.

F. Validation of Thumb Rule in Cyclic Flicker

Table VI reveals that the thumb rule depicted in Eq. (2) is applicable for all cases of cyclic flickers and it is not violated even for operation of multiple EAFs.

TABLE VI
P_{ST} AND $S(T)_{MAX}$ UNDER DIFFERENT OPERATING CONDITIONS

Furnace operation	m_1	m_2	P_{st}	S_{max}
Single EAF-1	0.4		1.5887	4.983
Parallel : two EAF-1	0.5	0.3	2.6973	14.3688
Single EAF-2	0.1		0.8951	1.5824
Parallel: one EAF-1 and one EAF-2	.5	.1	2.0487	8.2828

G. Effect Of Background Noise Power on Flicker Severity

P_{st} values against different background noise powers for parallel EAF operation are recorded in Fig. 6. The increasing nature of P_{st} with incremental noise power is obtained as

expected. Similar trends are observed for operation of single EAF of different ratings.

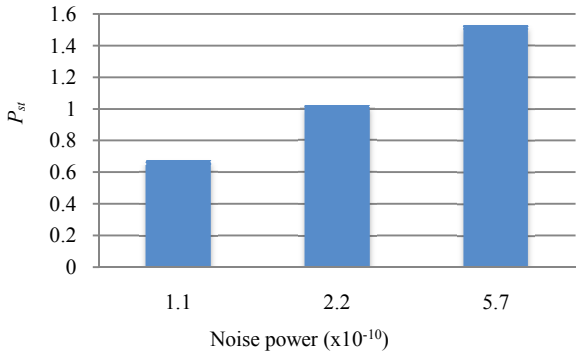


Fig. 6 Effect of background noise power on P_{st}

H. Flicker Severity Ratio of Single and Multiple EAF Operation

P_{st} ratio for parallel operation of two EAFs and their isolated operation is calculated and plotted against the corresponding modulating parameter *m,* for both sinusoidal and stochastic flicker phenomenon in Fig. 7. For multiple EAF operation the modulation parameter is considered same in each furnace.

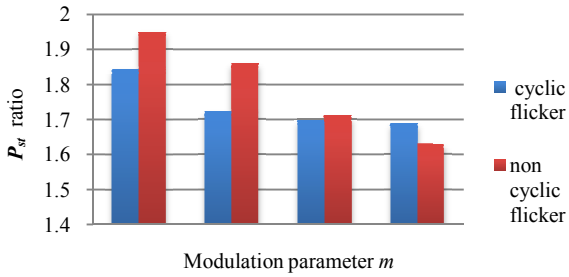


Fig. 7 P_{st} ratio for parallel and single EAF operation and for different modulation parameter *m*

Figure 7 shows that as the modulation parameter is increased, flicker ratio diminishes. It reveals that at lower rate of relative variation of arc length voltage flicker due to parallel operation of EAFs becomes prominent. The figure also indicates that in case of cyclic flicker the P_{st} ratio becomes less sensitive to variation of *m* at higher modulation indices. But for stochastic or non cyclic flicker such conclusions can't be drawn readily.

VII. CONCLUSION

In the paper flicker severity due to EAF operation has been discussed for both cyclic and non-cyclic flicker phenomena to make effective comparison between them. The effect of P_{st} on furnace ratings has been observed both for single and parallel operation of arc furnaces. Moreover, the effect of modulation index, modulation frequency and noise power on acceptable flicker limit has been depicted. It has been observed that relative variation of arc length is to be decreased for higher ratings of EAF for reliable operation. In case of parallel operation of identical EAFs, the cumulative level at PCC is more severe with smaller ratings than that of higher ratings. It is also remarkable that individual modulation indices have little effects on allowable P_{st} in sinusoidal flicker if the summation fulfils the condition. Comparison between cyclic and non-cyclic flicker reveals that the former is more threatening to power system than the other. For non cyclic flicker it is observed that no thumb rule is exactly satisfied.

REFERENCES

[1] Angelo Baggini, *Handbook of Power Quality*, New York, USA: Wiley, April 2008.

[2] S. R. Mendis, M. T. Bishop, A. V. Do, D. M. Boyd, "Investigation of transmission system voltage flicker due to multiple ac and dc arc furnaces," *IEEE Trans. On Power Del.*, vol.10, no.1, Jan. 1995.

[3] IEC Standard 61000-4-15 (Edition 1.1, 2003-02): *Electromagnetic Compatibility(EMC)–Part4 Testing and Measurement Techniques–Section15 Flickermeter: Functional and Design Specifications.*

[4] G. C. Montanari, M. Loggin, A. Cavallin, L. Pitti, D. Zaninelli, "ARC-furnace model for the study of flicker Compensation in electrical networks," *IEEE Transactions on Power Delivery,* vol. 9, no. 4, Oct. 1994.

[5] M. Maksić, I. Papič, "Network Flicker Levels due to Simultaneous Operation of Two Arc Furnaces," Proc. of 14th International Conference on Harmonics and Quality of Power (ICHQP, 2010), pp. 1-5, Bergamo, Italy, 26-29 Sept. 2010

[6] IEEE PES 01TP151, "Tutorial on voltage fluctuations and lamp flicker in electric power systems", 2001.

[7] S. Varadan, E.B. Makram, Adly A. Gigris, "A new time domain voltage source model for an arc furnace using EMTP," *IEEE Trans. On Power Del.*, vol 11, no. 3,pp.165-1690, July 1996

[8] M. Islam, A.H. Chowdhury, "Comparison of dynamic resistance arc furnace models for flicker study," 7th *International Conference on Electrical and Computer Engineering* , December 2012, Dhaka, Bangladesh

Reliability Evaluation of Grid Substations in Bangladesh Power System

*Md Iqbal Hossain, Rumana Binte Faruque, and A Hasib Chowdhury
Department of EEE, Bangladesh University of Engineering and Technology
Dhaka-1205, Bangladesh
*iqbalbuet18@gmail.com

***Abstract*— Determination of the reliability of a substation is important as it helps locate weak points that may be contributing to overall power system unreliability. This paper evaluates reliability of 230 kV grid substations in Bangladesh subject to protection failures. The problem belongs to the dynamic reliability evaluation field. Event tree analysis is used as the main method. Fault tree analysis is used to get the input of the event tree. Reliability of five grid substations are calculated. Ishurdi grid substation is found to have a mean time to failure of only six months, which will affect seriously the design and safety of the proposed nuclear power station at Rooppur sited closely nearby.**

***Index Terms*—Substation reliability, protection failure, event trees, Bangladesh power system**

I. Introduction

While one of the strongest points in a power system is the grid substation, it still contains weak points or points of failure that would lead to loss of load. Therefore, determining the reliability of a substation is important as it can help locate weak points that may be contributing to overall system unreliability.

The reliability of substation is dependent mainly on three factors: incorrect design, incorrect installation and service deterioration that leads to protection failure [1]. Several methodologies and computational models have been developed to evaluate the substations reliability. Most of the studies focus on the substation connectivity. The reaction of protection system is fully neglected. These methods use the minimum cut set technique to select the failure modes of the substation arrangements and classify them in outage categories. Mostly used methods are fault tree analysis (FTA) [2,3], event tree analysis (ETA) [2,3], Monte Carlo simulation and state enumeration [4,5]. Which method is to be used depends on the field of evaluation. To study the static reliability FTA, Monte Carlo simulation and state enumeration are used frequently. To evaluate the dynamic reliability ETA and sequential Monte Carlo are used.

This paper evaluates reliability of 230kV grid substations in Bangladesh power system subjected to protection failures. The reliability problem belongs to the dynamic reliability evaluation field. Design and installation mistakes are ignored and only the service deterioration is considered. Two types of protection failures are taken into account: protection system failures and associated circuit breaker (CB) failures. In this paper ETA is combined with FTA. ETA is the main method used and FTA is applied to get the input of the event tree [6,7].

II. Methodology

To reduce the number of event trees elements, components in the same protection zone are combined into one unit. When one or more than one of the components in a zone fails, the protection system will react. Consequently, FTA is needed to calculate the unavailability of this 'unit'. The result will be the initiating event probability for the ETA.

In ETA there are four states that occur based on the functionality of primary protection, CBs and back up protection [8].

State 1: When both the primary protection and CBs function well, there is no protection failure, and the fault will be cleared successfully

State 2: If the primary protection sees the fault, but the associated CBs fail, the CB failure function will be activated and trips the CBs next to the firstly associated ones.

State 3: When the primary protection fails, but the back-up protection works successfully, the fault will be cleared after a time delay.

State 4: When the primary protection and the back-up protection fail; the chance is so small that it is not considered. In this paper back-up CBs are assumed to be perfect.

When the events are independent of each other, the event trees can be calculated simply with the products of the probabilities along the tree branch. Thus the probability of any of the four states can be calculated by multiplying the probability of each event along the tree branch,

$$P(state1) = U(ini) \times A(pri) \times A(CB) \quad (1)$$

where $U(ini)$ is the unavailability of the initial event, $A(pri)$ is the availability of primary protection and $A(CB)$ is the availability of circuit breaker.

If the input data of the initiating event changed from its unavailability to its failure frequency while keeping the conditional probabilities of a protection failure and a circuit CB failure, the output of the tree will also change to a failure frequency,

$$P(state1) = f(ini) \times A(pri) \times A(CB) \quad (2)$$

where $f(ini)$ is the failure frequency of the initial event.

The concept pairs of reliability/failure probability and availability/unavailability are more or less the same. The parameters used in this paper are availability/unavailability.

Several assumptions are made first before analyzing the reliability of the substations.

- Components considered having a chance to fail: bus bar, CB, disconnecting switch, step-up and step-down transformers, current transformer, voltage transformer, overhead lines and cables connected to substation.
- Earthing switches and surge arrestors inside substation do not have a significant effect on the substation reliability and are neglected.
- Back-up protection systems and the back-up circuit breakers are assumed to be perfect.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

- Neighboring substations are assumed to be perfect.

A. *Effect of Substation Reliability On Respective Lines*

To realize the effect of reliability of a substation on respective lines consider a substation 'S/S 1' as shown in Fig. 1. The neighboring substation is 'S/S 2' which is assumed to be perfect. That is, 'S/S 2' is assumed to consist of CBs that will never fail.

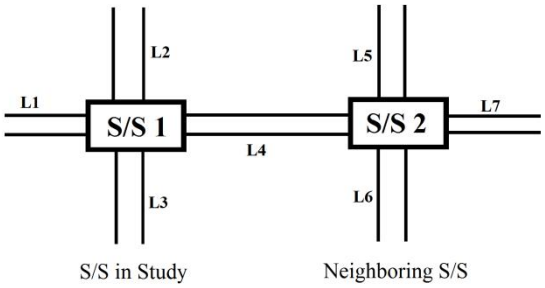


Fig. 1 Effect of substation reliability on respective Lines.

If 'S/S 1' fails then lines 1, 2, 3, and 4 are affected. Hence, reliability assessment of S/S 1 actually determines the unavailability (*U*), failure frequency (*f*), mean time to failure (*MTTF*) and mean time to repair (*MTTR*) of lines 1, 2, 3, and 4. These are reliability indices that are related to each other by Eqs. (3) and (4).

$$U = \frac{f \times MTTR}{8760} \tag{3}$$

$$f = \frac{1}{MTTR + MTTF} \tag{4}$$

In Eq. (3) MTTR is measured in hours, but in Eq. (4), both, MTTR and MTTF, are calculated in years.

B. *Reliability Input Data*

Failure statistics of different grid substation components including data for CBs, current transformer, potential transformer, and other switchgear components are not available for Bangladesh power system (BPS) network. Hence, the data from different sources are used [8,9]. Transmission line lengths and substation configuration data are taken from the Power Grid Company of Bangladesh (PGCB), the grid operator in Bangladesh.

III. Reliability Evaluation of BPS Grid Substations

There are a total of 104 grid substations in BPS. Among them 1 substation is of 400 kV HVDC, 17 substations are of 230/132 kV and 86 substations are of 132/33 kV [10]. In this paper reliability evaluation of five 230/132 kV substation is presented. These are Hasnabad, Hathazari, Ishurdi, Haripur and Rampura 230 kV grid substations. Detailed analysis is presented for the Hasnabad 230 kV grid substation.

Differential protection, distance protection and busbar protection coordinate with each other in a substation. Which of these protection systems reacts to a fault mainly depends on fault location. For the components that are within one protection zone, the effects on the protection will be the same. Therefore, the whole substation is divided into several zones. Hasnabad 230 kV grid substation has three types of zones.

1) *Bus bar Zone:* When the initiating fault is located in the zone of the busbar protection, as shown in Fig. 2, the busbar protection should react first, and trip the CBs: 808B, 818B, 828B and 838B. If any of these CBs fail, the CB failure function will be activated, and trip 805B, AT-1, 815B, AB-1, 825B, AB-2, 835B and AT-3. There is no back-up protection within the substation. If the busbar protection fails, the fault will not be seen by the distance protection within this substation. It will be the distance protection from the nearby substation that clears the fault. Therefore, the whole substation is assumed to be down.

The event tree for a fault within busbar 2 zone is shown in Fig. 3. When both the primary protection (busbar protection) and the associated CBs react successfully, there will be no lines isolated. When primary protection works successfully, but one of the associated CBs fails to trip, the next level of CBs will be tripped by the CB failure function. Line AT-1, Aminbazar-1, Aminbazar-2, AT-3 will be isolated in this case. If the busbar protection fails to see the busbar fault, the whole substation will be down. The situation where both primary and back-up protection fails is neglected, since back-up protection is assumed to be perfect in this study.

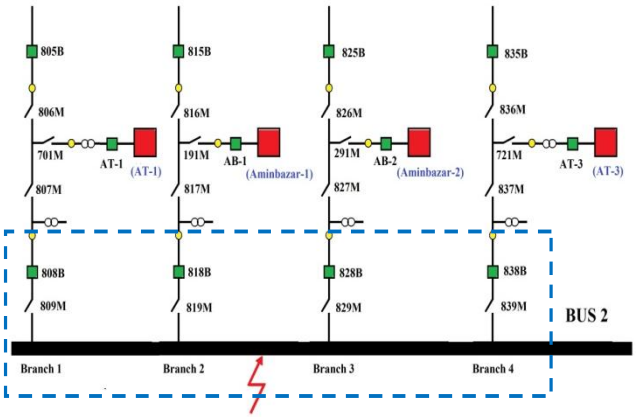


Fig. 2 Bus bar zone of Hasnabad 230 kV grid substation

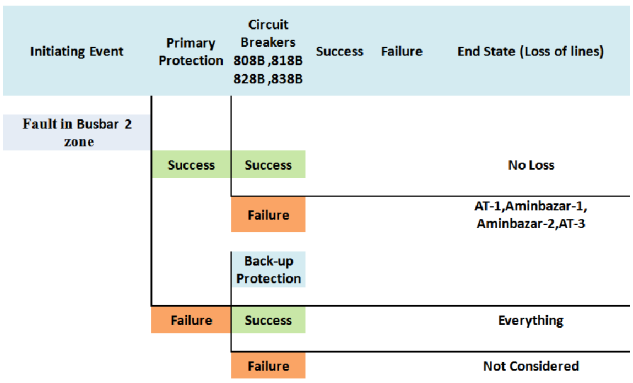


Fig. 3 Event tree for a fault in bus bar 2 zone of Hasnabad S/S

2) *Line Zone:* In Fig. 4, a fault occurring within a line zone is shown. When a fault occurs on the line, the field differential protection should react first. If the initiating fault is a line fault, line differential protection is the primary protection. Since the protection system for those two faults will trip the same CBs, they do not have a different effect on the reliability study. Therefore, these two zones are combined into a single line zone.

The event tree for a fault within Aminbazar-1 zone is shown in Fig. 5. In this case, the primary differential protection should react and trip CBs 815B, AB-1 and 818B and only Aminbazar-1 will be isolated. If one of the CBs fails, the CB failure function will trip 808B, 828B, 838B, 812B, MG-2 and Aminbazar-1, Meghnaghat-2 will be isolated. When differential protection fails to see the fault, the distance protection on the line should react, and trip 815B, AB-1 and 818B. Only Aminbazar-1 will be isolated in this case.

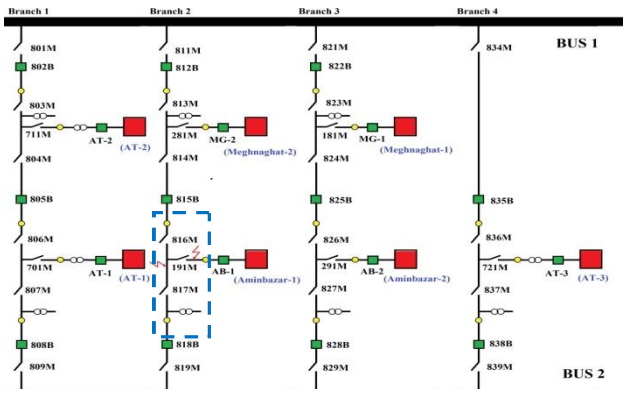


Fig. 4 Line zone of Hasnabad 230 kV grid substation

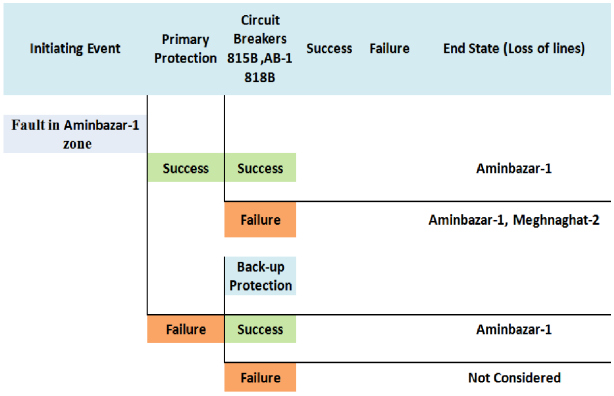


Fig. 5 Event tree for a fault in Aminbazar-1 zone of Hasnabad S/S

3) *Step-down Transformer Zone:* The step-down transformer zone is shown in Fig 6. If the fault is located on the line, field differential will react first. If the fault is a line fault, line differential protection is the primary protection. A fault lying in the transformer zone is protected by the transformer field differential protection. These different types of primary protection systems will trip the same CBs. Therefore, the three different zones are combined into one.

The event tree for the fault within the AT-1 zone is shown in Fig. 7. If an initiating fault occurs within this zone, the primary differential protection will trip 805B, AT-1, 808B and only AT-1 will be isolated. If one of those CBs fails, 818B, 828B, 838B, 802B, AT-2 will be tripped by the CB failure function. Both AT-1 and AT-2 will be isolated. When the primary protection fails, the distance protection should react and isolate AT-1.

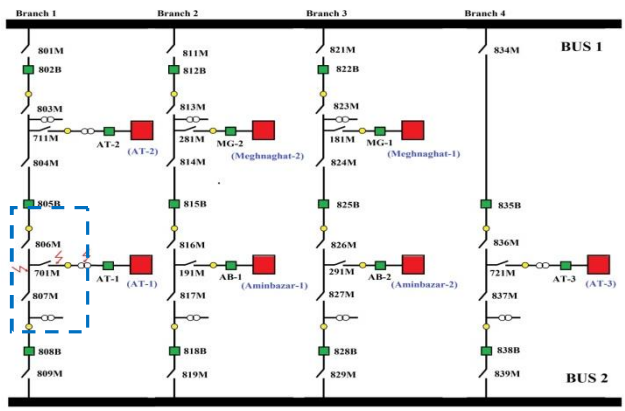


Fig. 6 Step-down transformer zone of Hasnabad 230 kV grid substation

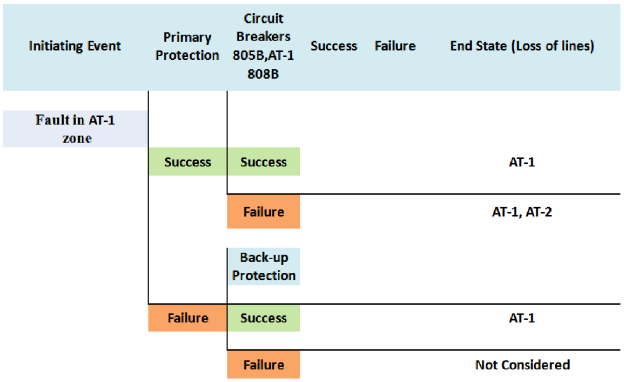


Fig. 7 Event tree for a fault in AT-1 zone

IV. Results And Analysis

To calculate the resulting probabilities and failure frequencies from the event tree, the unavailability and failure frequency of the initiating events is calculated first. This is done by using a small fault tree. When one or more than one of the components within a zone fails, this zone is regarded as failed. As such, every zone can be regarded as a small series connected system and its reliability can be calculated. The zone reliability for the Hasnabad grid substation is presented in Table I. These are the output of FTA and used as the input to ETA.

TABLE I
Zone Reliability

Zone name	Zone's failure frequency	Zone's unavailability
Aminbazar-1	0.0626	0.000058652
Aminbazar-2	0.0626	0.000058652
Meghnaghat-1	0.0668	0.000062124
Meghnaghat-2	0.0668	0.000062124
AT-1	0.06388	0.000364261
AT-2	0.06368	0.000363713
AT-3	0.06388	0.000364261
Busbar 1	0.015	0.000011645
Busbar 2	0.015	0.000011645

The event trees can be calculated using the method introduced in section III. By adding the failure statistics of the same end states, the frequency and probability of losing lines can be calculated. The results are presented in Table II.

TABLE II
Specific Line Failure Statistics

Lost line	Unavailability	Failure frequency	Mean time to failure (Year)	Mean time to repair (Hr)
Aminbazar1	6.06E-05	0.063	15.79	8.39
Aminbazar2	6.06E-05	0.063	15.79	8.39
Meghnaghat1	6.41E-05	0.067	14.81	8.32
Meghnaghat2	6.41E-05	0.067	14.81	8.32
AT1	3.68E-04	0.064	15.48	49.87
AT2	3.67E-04	0.064	15.52	49.95
AT3	3.64E-04	0.064	15.59	49.81

The unavailability and failure frequency of a line represents the probability/frequency that this line is isolated from the substation due to any component's fault. These are calculated using the event trees. The dominant component in reliability evaluation is the transmission line or cable. The longer the line, the greater the chance that a line fault occurs.

Meghnaghat1 and Meghnaghat2 have the longest transmission lines: 26 km. So their failure frequency is the highest. The MTTF represents the average time it takes before the specific line/generator is isolated due to a component's fault. The MTTR represents the average time it takes to locate and repair the fault, and then put the specific line back into operation.

A summary of the five 230 kV grid substations in terms of the most reliable zone and the zone of highest possibility to fail for each substation is presented in Table III. The zone which has highest MTTF or lowest failure frequency is the most reliable zone and the zone which have lowest MTTF or highest failure frequency has the highest failure possibility.

TABLE III
RELIABILITY OF ZONES OF RESPECTIVE SUBSTATIONS

Substation	Most Reliable zone	Most unreliable zone
Hasnabad	Aminbazar1, Aminbazar2	Meghnaghat1, Meghnaghat2
Hathazari	Raojan1, Raojan2	Comilla1
Ishwardi	T3	Ghorashal1, Ghorashal2
Haripur	Siddhirganj	Rampura1
Rampura	MT3	Ghorashal1

Figures 8(a) to 8(c) present the overall unavailability, failure frequency and MTTF, respectively, of the five grid substations.

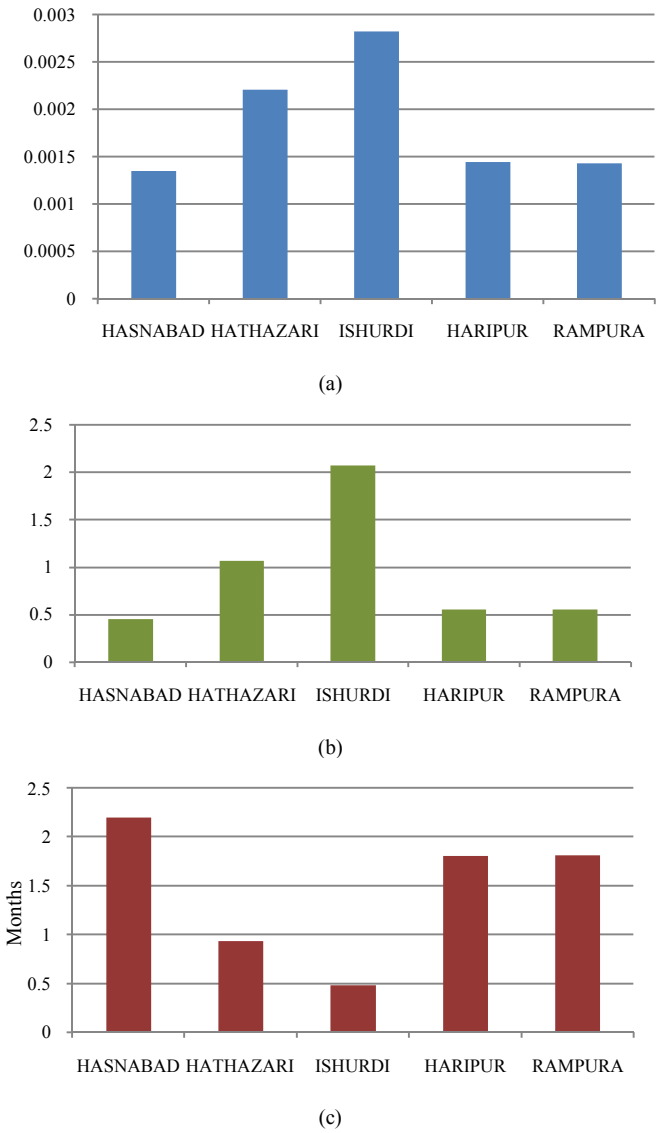


Fig. 8 Reliability indices for different grid substations: (a) unavailability, (b) failure frequency, (c) mean time to failure

The overall unavailability and failure frequency is calculated by adding up the unavailability and failure frequency of different zones. Then MTTF is calculated by Eqs. (3) and (4). It can be seen that the Ishurdi 230 kV grid substation has the highest unavailability of almost 0.003, followed by Hathhazari with an unavailability a little over 0.002. Hasnabad, Haripur and Rampura, all within or near the capital city of Dhaka, have similar level of unavailability of less than 0.0015. Hasnabad grid substation has the lowest failure frequency and the highest MTTF.

Ishurdi grid substation is the most vulnerable with the highest failure frequency of more than two failures a year and the lowest mean time to failure of a little less than six months. This will seriously affect the design and operational safety of the proposed nuclear power station at Rooppur sited closely nearby.

V. CONCLUSIONS

The reliability of five different 230kV substations of Bangladesh is analyzed and their unavailability, failure frequency and mean time to failure compared. The study will help to assume the ability of BPS grid substations to be connected to challenging energy sources like nuclear power plants. Throughout the analysis, it has been observed that the dominant component in reliability evaluation of a system is the transformers and line/cable, because their failure frequency is large and the lines with similar parameters and locations have similar failure statistics. In this paper, all the components within the substation are assumed to be operating properly, which means that maintenance is not considered. Maintenance may be included in future studies.

REFERENCES

[1] Roy Billinton, R.A., *Reliability Evaluation of Power Systems*. Second Edition, 1996, New York.

[2] L. R. C. Ferreira, J. Pugh, P. A. Crossley, R. N. Allan, and J. Goody, "Design and Reliability of Integrated Protection and Control Schemes," Proc. of Fourth International Conference on Power System Control and Management, pp. 87-91, London, UK, 16-18 April 1996.

[3] G. H. Kjolle, O. Gjerde, B. T. Hjartsjo, H. Engen, L. Haarla, L. Koivisto, and P. Lindblad, "Protection System Faults - a Comparative Review of Fault Statistics," Proc. of 9th International Conference on Probabilistic Methods Applied to Power Systems, pp. 1-7, Stockholm, Sweden, 11-15 June 2006.

[4] W. Li, *Risk Assessment of Power Systems: Models, Methods, and Applications*, New York, USA: IEEE Press, 2004.

[5] J. C. H. Schüller, J. L. Brinkman, P. J. Van Gestel, R.W. van Otterloo, *Methods for Determining and Processing Probabilities*, 2nd ed. The Hague: Gevaarlijke stiffen, 1997,

[6] L. Haarla, U. Pulkkinen, M. Koskinen, and J. Jyrinsalo, "A method for analyzing the reliability of a transmission grid," *Reliability Engineering and System Safety*, vol. 93, no. 2, pp. 277–287, Feb. 2008.

[7] L. Pottonen, "A Method for the Probabilistic Security Analysis of Transmission Grids," Ph.D. Thesis, Helsinki University of Technology, Helsinki, Finland, 2005.

[8] F. Wang, "Reliability Evaluation of Substations Subject to Protection Failures," M. Eng. thesis, Delft University of Technology, Delft, the Netherlands, July 2012.

[9] Canadian Electricity Association report, "Forced Outage Performance of Transmission Equipment, For the Period January 1, 1998 to December 31, 2002", 2004.

[10] (2014) PGCB website. [Online]. Available: http://www.pgcb.org.bd/

CMOS Compatible Generic Batch Process Towards Flexible Memory On Bulk Monocrystalline Silicon (100)

Mohamed Ghoneim, Jhonathan Rojas, Arwa Kutbee, Amir Hanna, and Muhammad Hussain*

Integrated Nanotechnology Lab, Electrical Engineering, Computer Electrical Mathematical Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

* MuhammadMustafa.Hussain@kaust.edu.sa

Abstract— **Today's mainstream flexible electronics research is geared towards replacing silicon either totally, by having organic devices on organic substrates, or partially, by transferring inorganic devices onto organic substrates. In this work, we present a pragmatic approach combining the desired flexibility of organic substrates and the ultra-high integration density, inherent in silicon semiconductor industry, to transform bulk/inflexible silicon into an ultra-thin mono-crystalline fabric. We also show the effectiveness of this approach in achieving fully flexible electronic systems. Furthermore, we provide a progress report on fabricating various memory devices on flexible silicon fabric and insights for completely flexible memory modules on silicon fabric.**

***Index Terms*— flexible electronics; mono-crystalline silicon; semiconductor industry; memory devices; electronic systems.**

I. Introduction

As new horizons and applications arise for flexible electronics, the opportunity for flexible systems enhances [1-9]. Applications ranging from biomedical to consumer electronics impose some specifications to be met in this new electronics generation, such as high flexibility, reliability, performance, and integration density. Great progress and milestones have already been achieved through different approaches utilizing polymeric substrates and polymeric devices [10-17]. However, there are many challenges along the way involving low carrier mobility, thermal instability, low integration density, etc. Other approaches exist, such as utilizing expensive silicon on insulator (SOI) or unorthodox silicon (111) wafers and peeling off the thin silicon nano-ribbons with or with-out devices [18-19]. This solves the thermal budget limitations partially (if the devices are already fabricated) and low performance associated with organic electronics but at the expense of the higher costs and limited integration density.

To this end, we provide a pragmatic approach to complement current research which transforms bulk silicon into a flexible fabric hosting devices [20-27]. Our approach follows an etch-protect-release-recycle process to peel off an ultra-thin flexible sheet with pre-fabricated devices from low cost bulk mono-crystalline silicon (100), commonly used in the semiconductor industry. This process capitalizes on the complementary metal oxide semiconductor (CMOS) technology as opposed to replace it while keeping the cost checked.

Furthermore, we utilize our demonstrated process to build up various device components required for building flexible memory modules, essential elements in all electronic systems. This is a necessary step towards taking flexible electronics research into demonstration of completely flexible stand-alone electronics systems.

This work is organized as follows; Section II discusses our generic transformative process for creating flexible Si fabrics; Section III provides review of the various demonstrated devices on flexible fabric required for building a flexible memory module; Section IV is for conclusions and insights related to the future of flexible electronics research.

II. Transforming Bulk Silicon Into a Flexible Ultra-Thin Semi-Transparent Silicon Fabric

To transform bulk devices into flexible ones, we have two approaches. The first is the "device first" approach where we build the devices on bulk then peel off a thin Si fabric containing the devices. The second is the "device last" approach where the thin silicon fabric is peeled off with only the edges acting as anchors to tie it to the wafer while device building processing steps are performed. The detailed process for each approach is as follows:

A. Device first approach

Fig. 1 shows the schematic for the process; starting with 1, a) showing a Si/hard mask (SiO_2 in our case)/ device stack. Then, etching deep trenches ("release holes") throughout the whole stack and deep into the Si up to the desired fabric thickness (Fig. 1, b) using photolithographic techniques and deep reactive ion etching. Next, using spacer formation technique a conformal atomic layer deposition (ALD) layer is deposited everywhere then vertically etched leaving only sidewalls protection (Fig. 1, c, d). Finally, the sample is put in XeF_2 for isotropic etching of Si from the bottom of release holes (only exposed Si area) until the etched regions merge, enabling the release of the thin device containing fabric (Fig. 1, d). Fig. 1, e shows a top view after the final step.

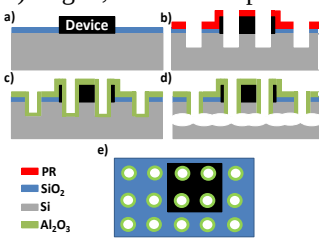


Fig. 1 Process flow fort the "device first" transformational approach.

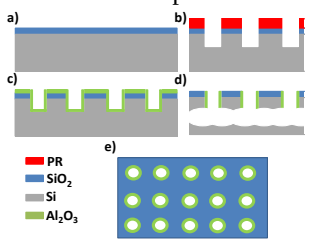


Fig. 2 Process flow fort the "device last" transformational approach.

B. Device last approach

Fig. 2 shows the fabric release process after which the devices can be built in the traditional way on the already flexible fabric.

Both approaches have associated perks compared to aforementioned techniques using polymeric devices or expensive inorganic substrate in terms of performance and cost. Performance wise, the devices are silicon based built on silicon using traditional optimized semiconductor processes that outperform organic electronics. Cost effectiveness comes from the fact that the wafers are mono-crystalline silicon (100) compared to the less common silicon (111) and the 10x more expensive SOI wafers. Moreover, the recyclability of the remaining bulk wafer after the removal of the top thin layer renders this solution more economical. On the other hand, the release holes network in this work consumes 19.6% (10 µm holes, 20 µm pitch, and 100 x 100 μm^2 devices) of chip area. This is still acceptable with respect to the international technology roadmap for semiconductors (ITRS) projections showing that 20% of active chip area will still be consumed with shallow trench isolation (STI) even with extreme scaling [28]. The pros and cons of the device first vs. the device last approaches are as follows. The device first approach follows standard CMOS processes until the devices are completely built, then release holes are etched-protected then the thin fabric containing devices is released. This makes it easier to handle during device fabrication and enables testing the bulk devices for functionality before and after flexing for comparison. However, this comes at the expense of fabricated devices passing extra processing steps (RIE, DRIE, ALD-300°C and XeF_2 gas for extended durations). On the other hand, the device last approach spares the devices the extra steps after their fabrication at the expense of more delicate handling constraints, relatively non planar released fabric, and the inability of probing bulk devices before flexing. Packaging of these devices can be done using SU-8 (an epoxy based polymer used in microelectronic applications) which possess chemical and thermal stability and excellent flexibility up to 75 µm thickness [29].

III. Review of Various Memory Components Built on Flexible Silicon Fabric and Performance Assessment

Using the two flexing approaches, various memory components (DRAM, ReRAM, FeRAM) were built on the flexible Si fabric. DRAM is composed of an array of memory cells hosting a select/access transistor and a stored charge in a capacitor representing bits (digital ‘0’ or ‘1’). DRAM has a volatile nature, i.e. loses information when no power is supplied, and has a small foot print due to its simple cell structure. ReRAM is an emerging non-volatile memory technology where resistance of material is dependent on the history of applied voltage. It has simple structure like DRAM but does not require continuous power supply to store the data and has fast switching capability. FeRAM is similar to ReRAM but it is a more mature technology (commercially available) where material polarization depends on the history of applied voltage. FeRAM has desirable low power operation, extended data retention periods, and survives millions of read/write cycles.

Memory arrays account for the majority of transistors (select/access transistors) in a CMOS system on chip. Therefore, MOSCAPs (basic gate stack of MOS transistors) is an essential memory component that needs to be assessed when addressing the roadmap towards flexible memory.

A. Dynamic random access memory (DRAM)

The DRAM is the volatile fast component of memory modules. Using the “device last” approach, metal/insulator/metal capacitors (MIMCAPs), basic building block of DRAMs and the structure suggested by technology roadmap for 65 nm DRAM nodes and below [30], were built. Fig. 3 shows the observed increase in normalized capacitance per unit planar area for released devices versus unreleased (details can be found in [21]). This observation indicates that the released devices can be used to build a flexible and denser DRAM relative to ones utilizing bulk devices.

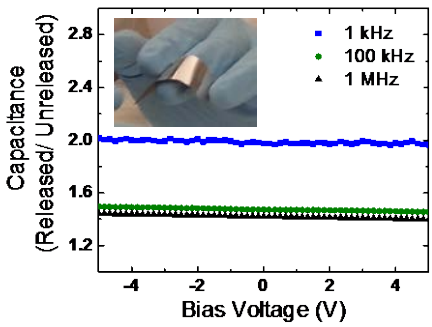


Fig. 3 Improvement in capacitance per unit planar area of released devices.

B. Resistive random access memory (ReRAM)

Conceptualized in 1971 by L. Chua and realized in 2008 in HP labs, memristors research has been gaining momentum especially as an alternative non-volatile memory (NVM) to CMOS based flash which is approaching scaling limits [31, 32]. This is mainly due to their simple structure and fast switching.The same MIMCAPs stack showed resistive switching at higher voltages which can be explained by a phenomenon similar to a soft breakdown after which the material is able to recover when applied voltage polarity is reversed. Figure 4 shows the memristive effect in MIM stack for bulk (unreleased) and flexible (released devices). The plot shows that there is no discernible degradation due to the flexing process as the curves overlap throughout the sweeping regime.

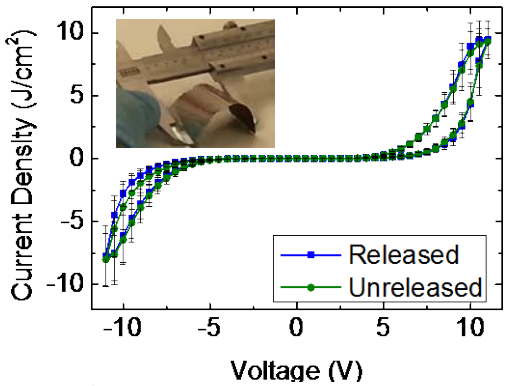


Fig. 4 Flexible and inflexible memristive devices observed hysteresis loops.

C. Ferroelectric random access memory (FeRAM)

Another form of NVM is FeRAM -- matured technology and widely available. Using the most widely used ferroelectric material, lead zirconium titanate (PZT), prepared using sol-gel technique, FeRAM cells were fabricated. Fig. 5 shows the fabrication process followed, which is simpler and different from our previous approach but a necessary step before performing our more elaborate flexing process. First, the

SiO_2/Si is sputtered with Pt electrode (Fig. 5, b). Then, PZT first layer was spun at 1500 rpm for 35 seconds, followed by a pyrolysis bake at 300 °C for 5 minutes; this is repeated for 4 layers (~280 nm) followed by a final anneal at 650°C for 30 minutes. Then, a Ti adhesion layer and Pt top electrode are sputtered and patterned. Finally, the top side of the sample is protected with photoresist and sample is flipped and etched using DRIE to make a thin Si sheet. The goal of this study is to make sure FeRAM devices will still function on a bent flexible Si fabric.

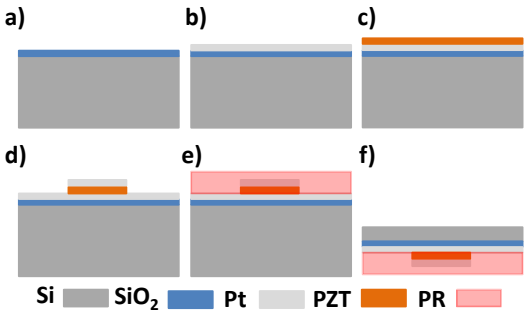


Fig. 5 Fabricating PZT FeRAM devices.

Figure 6 shows the P-E hysteresis loop for functioning bulk, flexible and bent at 10 mm bending radius ferroelectric devices.

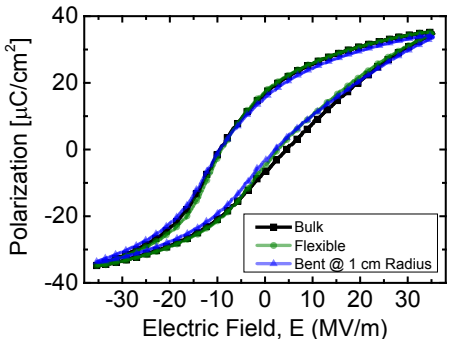


Fig. 6 P-E Hysteresis curves for bulk, flexible, and bent ferroelectric devices.

D. MOSCAPs

Transistors are essential elements in memory modules due to their significant role in crossbars arrays architecture [33, 34]. Select transistors are used to select the row and column of the cell to be read from or written into. Metal oxide semiconductor capacitors (MOSCAPs) are used to study the gate stack and predict the effects of various anomalies in these transistors.

State-of the art high-κ MOSCAPs were built on bulk wafer and then released using our "device first" approach. Figure 7, a) shows the C-V plot for bulk and flexible devices at different frequencies, indicating frequency imposed limitations. From the plot we observe that there is more frequency dispersion in the released devices which can be attributed to the release holes network passing through the active area of the devices. This can be avoided by scaling down active areas and varying the spaces between holes. This is still an area of improvement for our Si based flexible memory modules to function at high frequencies.

Transistors reliability is another area of concern and has been assessed for our MOSCAPs. Figure 7, b) shows the time dependent dielectric breakdown plot for released MOSCAPs. These plots are used to extrapolate the safe operational voltage of the devices that would pass the 10 years lifetime benchmark (industry standard). The lifetime projections showed 20% decrease in safe operation voltage for flexible devices versus unreleased. However, the minimum safe voltage was 5 V which is acceptable as today's ICs aim at 1 V supplies.

Furthermore, we have studied the effect of bending on the leakage current and the breakdown voltage of the devices (depicted in Figure 8). Figure 8, a) shows that the devices can survive up to 7 days while bent at 10 mm bending radius (0.125% nominal strain) if the operational voltage is below 85% of breakdown voltage. On the other hand, Figure 8, b) shows that the number of bending cycles has a fast degradation effect on the value of breakdown voltage of the devices. Hence, these devices would be more suitable for applications where they remain bent or conformal to a non-planar surface than an environment where it is repetitively bent and unbent (detailed discussion on mechanical reliability of flexible MOSCAPs can be found in Ref. [35]).

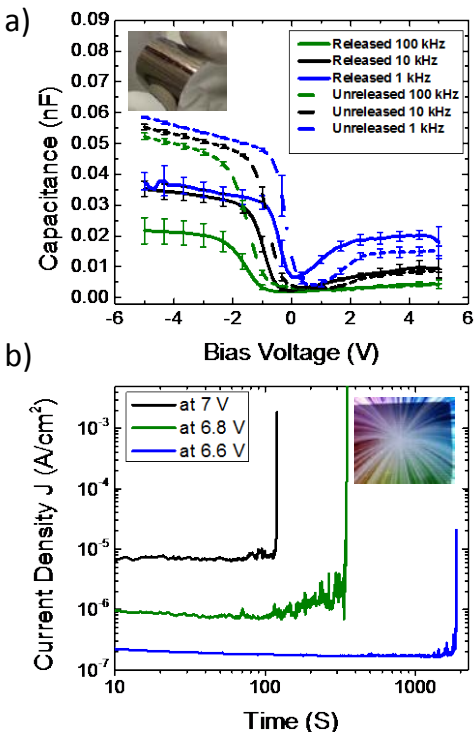


Fig. 7 MOSCAP plots, a) C-V plot for flexible and inflexible devices, b) time dependent dielectric breakdown study of flexible devices on Si fabric, inset showing semi-transparency of flexible silicon fabric.

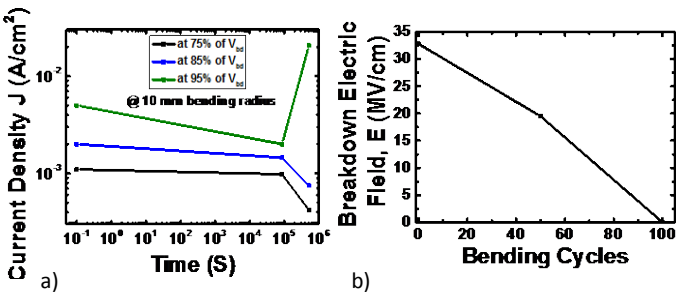


Fig. 8 Leakage plot as a function of bending time (a), and breakdown plot as a function of bending cycles (b).

IV. Conclusions

We have demonstrated a CMOS compatible low cost process to transform any traditional electronics into flexible and semi-transparent one. Using that generic batch process,

we have made all necessary elements required for silicon based fully flexible memory modules. Devices include DRAM cells, NVM cells (memristive and ferroelectric), and MOSCAPs to study the gate stack of select transistors to be used in crossbars arrays architectures, all built on thin flexible silicon fabric. This is an important step forward towards achieving completely flexible stand-alone electronic modules while capitalizing on the perks associated with the well-established silicon based semiconductor industry.

ACKNOWLEDGEMENTS

This work was supported by KAUST Office of Competitive Research Fund (OCRF) Competitive Research Grant CRG-1-2012-HUS-008. The authors would like to thank Mohammed Y. Al-Nassar for sharing his valuable knowledge of PZT film deposition conditions, and Mohammed A. Zidan for the invaluable discussions on resistive memory cells.

REFERENCES

[1] H. Zhou, J. H. Seo, D. M. Paskiewicz, Y. Zhu, G. K. Celler, P. M. Voyles, W. Zhou, M. G. Lagally, and Z. Ma, "Fast flexible electronics with strained silicon nanomembranes," *Sci. Rep.*, vol. 3, pp. 1-7, Feb. 2013.

[2] A. Lecavelier des Etangs-Levallois, M. Lesecq, F. Danneville, Y. Tagro, S. Lepilliet, V. Hoel, D. Troadec, D. Gloria, C. Raynaud, and E. Dubois, "Radio-frequency and low noise characteristics of SOI technology on plastic for flexible electronics, Solid-State Electronics," *Solid-State Electron.*, vol. 90, pp. 73-78, Dec. 2013.

[3] B. Nasr, D. Wang, R. Kruk, H. Rösner, H. Hahn, and S. Dasgupta, "Flexible Electronics: High-Speed, Low-Voltage, and Environmentally Stable Operation of Electrochemically Gated Zinc Oxide Nanowire Field‐Effect Transistors," *Adv. Funct. Mater.*, vol. 23, pp. 1729-1729, Apr. 2013.

[4] S. T. Han, Y. Zhou, C. Wang, L. He, W. Zhang, and V. Roy, "Flexible Electronics: Layer-by-Layer-Assembled Reduced Graphene Oxide/Gold Nanoparticle Hybrid Double-Floating-Gate Structure for Low-Voltage Flexible Flash Memory," *Adv. Mater.*, vol. 25, pp. 793-793, Feb. 2013.

[5] P. Sonar, T. j. Ha, T. R. B. Foong and A. Dodabalapur, "Push-Pull Based Novel π-Functional Polymeric Semiconductors for Printed Flexible Electronics," *223rd ECS Meeting*, Toronto, ON, Canada, 2013.

[6] S. Farsinezhad, A. Mohammadpour, A. N. Dalrymple, J. Geisinger, P. Kar, M. J. Brett, And K. Shankar, "Transparent anodic TiO_2 nanotube arrays on plastic substrates for disposable biosensors and flexible electronics," *J. Nanosci. Nanotechnol.*, vol. 13, pp. 2885-2891, Apr. 2013.

[7] L. A. Majewski, M. Grell, S. D. Ogier, and J. Veres, "A novel gate insulator for flexible electronics," *Org. Electron.*, vol. 4, pp. 27-32, Jun. 2003.

[8] M. C. McAlpine, R. S. Friedman, and C. M. Lieber, "High-performance nanowire electronics and photonics and nanoscale patterning on flexible plastic substrates," *Proc. IEEE*, vol. 93, pp. 1357-1363, Jul. 2005.

[9] Y. Sun and J. A. Rogers, "Inorganic Semiconductors for Flexible Electronics," *Adv. Mater.*, vol. 19, pp. 1897-1916, Jul. 2007.

[10] J. N. Tiwari, J. S. Meena, C.-S. Wu, R. N. Tiwari, M.-C. Chu, F.-C. Chang, and F.-H. Ko, "Thin-Film Composite Materials as a Dielectric Layer for Flexible Metal–Insulator–Metal Capacitors," *ChemPlusChem*, vol. 3, no. 9, pp. 1051-1056, Sep. 2010.

[11] M. K. Hota, M. K. Bera and C. K. Maiti, "Flexible metal–insulator–metal capacitors on polyethylene terephthalate plastic substrates," *Semicond. Sci. Technol.*, vol. 27, no. 10, pp. 105001-1-105001-6, Aug. 2012.

[12] A. Bag, M. K. Hota, S. Mallik and C. K. Maiti, "Graphene oxide-based flexible metal–insulator–metal capacitors," *Semicond. Sci. Technol.*, vol. 28, no. 5, pp. 055002-1-055002-7, Mar. 2013.

[13] A. Carlson, A. M. Bowen, Y. Huang, R. G. Nuzzo and J. A. Rogers, "Transfer Printing Techniques for Materials Assembly and Micro/Nanodevice Fabrication," *Adv. Mater.*, vol. 24, no. 39, pp. 5284-5318, Oct. 2012.

[14] S. A. Stauth and B. A. Parviz, "Self-assembled single-crystal silicon circuits on plastic," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 103, no. 38, pp. 13922-13927, Aug. 2006.

[15] T. Kim, Y. H. Jung, H.-J. Chung, K. J. Yu, N. Ahmed, C. J. Corcoran, J. S. Park, S. H. Jin, and J. A. Rogers, "Deterministic assembly of releasable single crystal silicon-metal oxide field-effect devices formed from bulk wafers," *Appl. Phys. Lett.*, vol 102, no. 9, pp. 182104-1-182104-4, May 2013.

[16] C. H. Lee, D. R. Kim, and X. L. Zheng, "Fabricating nanowire devices on diverse substrates by simple transfer-printing methods," *Proc. Natl. Acad. Sci. U.S.A.*, vol 107, no. 22, pp. 9950-9955, Jun. 2010.

[17] S. Kim, H. Y. Jeong, S. K. Kim, S.-Y. Choi and K. J. Lee, "Flexible Memristive Memory Array on Plastic Substrates," *Nano Lett.*, vol. 11, no. 12, pp. 5438-5442, Oct. 2011.

[18] D. Shahrjerdi, S. W. Bedell, "Extremely flexible nanoscale ultrathin body silicon integrated circuits on plastic", *Nano Lett.*, vol. 13, no. 1, pp. 315–320, Dec. 2012

[19] D. Shahrjerdi, S. W. Bedell, A. Khakifirooz, K. Fogel, P. Lauro, K. Cheng, J. A. Ott, M. Gaynes, D. K. Sadana, "Advanced Flexible CMOS Integrated Circuits on Plastic Enabled by Controlled Spalling Technology", in *Proc. IEEE Int. Electron Devices Meeting (IEDM)*, San Francisco, 2012, pp. 5.1.1-5.1.4.

[20] J. P. Rojas, G. T. Sevilla, and M. M. Hussain, "Structural and electrical characteristics of high-κ/metal gate MOSCAPs fabricated on flexible, semi-transparent silicon (100) fabric," *Appl. Phys. Lett.*, vol. 102, pp. 064102-1-064102-4, 2013.

[21] M. T. Ghoneim, J. P. Rojas, A. M. Hussain, and M. M. Hussain, "Additive advantage in characteristics of MIMCAPs on flexible silicon (100) fabric with release-first process," *Phys. Status Solidi RRL*, vol. 8, pp. 163-166, Feb. 2014.

[22] J. P. Rojas, M. T. Ghoneim, C. D. Young, M. M. Hussain, "Flexible High- κ/Metal Gate Metal/Insulator/Metal Capacitors on Silicon (100) Fabric," *IEEE Trans. Electron Dev.*, vol. 60, no.10, pp.3305- 3309, Oct. 2013.

[23] J. P. Rojas, G. A. Torres Sevilla, S. B. Inayat, A. M. Hussain, M. T. Ghoneim, S. Ahmed, M. M. Hussain, "Transformational Silicon Electronics" *ACS Nano*, vol. 8, no. 2, pp. 1468–1474, Jan. 2014.

[24] G. A. Torres Sevilla, C. E. Smith, J. P. Rojas, H. M. Fahad, M. M. Hussain, "Flexible and Transparent Silicon Based Sub-100 nm Non-planar 3D FinFET CMOS for Brain-inspired Computation" *Adv. Mater.* vol. 26, pp. 2794–2799, Mar. 2014.

[25] J. P. Rojas, G. A. Torres Sevilla, and M. M. Hussain, "Can We Build a Truly High Performance Computer Which is Flexible and Transparent?" *Sci. Rep.*, vol. 3, pp. 1-8, Sep. 2013.

[26] G. T. Sevilla, S. B. Inayat, J. P. Rojas, A. M. Hussain, and M. M. Hussain, "Energy Harvesters: Flexible and Semi-Transparent Thermoelectric Energy Harvesters from Low Cost Bulk Silicon (100)," *Small*, vol. 9, pp. 3916–3921, Dec. 2013.

[27] G. A. Torres Sevilla, M. T. Ghoneim, H. Fahad, J. P. Rojas, A. M. Hussain, M. M. Hussain, "Flexible Nano-Scale High-Performance FinFETs", *ACS Nano*, DOI: 10.1021/nn5041608, Sep. 2014.

[28] ITRS Reports 2013.

[29] E. Bosman, J. Missinne, B. Van, G. Van Steenberge, S. Kalathimekkad, J. Van Erps, I. Milenkov, K. Panajotov, T. Van Gijseghem, P. Dubruel, H. Thienpont, P. Daele, "Ultrathin optoelectronic device packaging in flexible carriers," *IEEE J. Sel. Topics Quantum Electron.*17, 617 – 628, Jan. 2011.

[30] K. Kim, "Technology for sub-50nm DRAM and NAND flash manufacturing," in *IEEE Int. Electron Devices Meet., Tech. Dig.*, Washington, DC, USA, 2005, pp. 323-326.

[31] L. O. Chua, "Memristor-The missing circuit element," *IEEE Trans. Circuit Theory*, vol. 18, no. 5, pp. 507- 519, Sept. 1971.

[32] D. B. Strukov, G. S. Snider, D. R. Stewart & R. S. Williams, "The missing memristor found," *Nature*, vol. 453, no. 7191, pp. 80-83, Mar. 2008.

[33] S. Jo, T. Chang, I. Ebong, B. Bhadviya, P. Mazumder, and W. Lu, "Nanoscale memristor device as synapse in neuromorphic systems," *Nano Lett.*, vol. 10, no. 4, pp. 1297–1301, Mar. 2010.

[34] K.-H. Kim, S. Gaba, D. Wheeler, J. M. Cruz-Albrecht, T. Hussain, N. Srinivasa, and W. Lu, "A Functional Hybrid Memristor Crossbar-Array/CMOS System for Data Storage and Neuromorphic Applications," *Nano Lett.*, vol. 12, pp. 389-395, Dec. 2011.

[35] M. T. Ghoneim, A. Kutbee, F. Ghodsi, G. Bersuker and M. M. Hussain, "Mechanical anomaly impact on metal-oxide-semiconductor capacitors on flexible silicon fabric," *Appl. Phys. Lett.*, vol. 104, no. 23, pp. 234104, Jun. 2014.

Ballistic performance comparison of InGaSb and InAsSb XOI nFET

Md. Nur Kutubul Alam, Muhammad Shaffatul Islam, Md. Raifqul Islam
Department of Electrical and Electronic Engineering
Khulna University of Engineering & Technology
Khulna-9203, Bangladesh
alamjhilam@gmail.com

***Abstract*—Ballistic performance of nanoscale InGaSb and InAsSb XOI n-FET is compared using non equilibrium greens function (NEGF) method. The simulation results demonstrate that the ballistic performance is opposite to what is expected if the devices were long channel MOSFET. Although InAsSb possesses higher bulk electron mobility, but in ballistic regime, InGaSb gives higher drain current along with better subthreshold characteristic and reduced threshold voltage, if the off state current for both material is made equal.**

Keywords—XOI;NEGF;Ballistic transport; InGaSb;InAsSb (key words)

I. INTRODUCTION

On demand of performance improvement and increased integration density, Gate-length of the Silicon MOSFETs has reduced from 10 µm to 22 nm in last 40 years, and it is now approaching its physical limit [1-3]. These ultra scaled devices suffer from various adverse effects including short channel and hot electron effects, parasitic capacitance, leakage current, transistor isolation etc. Getting rid of bulk silicon (Si) structure in which channel had been formed by inversion layer, a layer of single crystalline Si is laid over insulator. This Si layer serves as channel and the silicon-on-insulator (SOI) structure has drastically improved the device performance by suppressing these adverse effects [4,5]. Performance of this device could further be improved by replacing Si from the channel with a material having better electron transport properties. This would allow reduction in operating voltage without compromising the switching speed. Therefore, further downscaling may be feasible [6].

Group III-V compound semiconductors exhibit outstanding electronic, including high electron mobility, high peak saturation velocity [7,8]. Additionally, III-V semiconductors have a mature and large industry for solid-state lighting and transceivers for optical communications [8]. Therefore, integration of III-V semiconductors on Si platform would now be the 'royal road' for achieving high performance low power device fabrication. With that aim, in analogy to SOI, n-type InAs FETs were experimentally demonstrated on Si/SiO_2 substrate [9] and the structure is termed as "XOI" to represent compound semiconductor-on-insulator platform. Due to similar device structure, XOI also offers all the advantages of SOI. Specially, the suppression of short channel effect (SCE) makes the XOI a promising technology for sub-100-nm node and beyond [10].

But performance of a device in this ultra scaled dimension can be very different since the dominating factors are not the same. InAsSb long channel MOS shows better performance than that of InGaSb due to higher electron mobility [8, 11]. However, when the channel length goes to sub 100nm regime the electronic transport is ballistic. As a result, the drain current is determined by the density of states of the channel material as well as the electron velocity through the channel [12, 13]. In this paper, the performances of InAsSb and InGaSb XOI FETs are investigated using non equilibrium greens function (NEGF) method. It is found that InGaSb XOI FET shows better performance than that of InAsSb in all aspect.

II. QUANTUM BALLISTIC SIMULATION

A. Device structure

The device structure studied here is shown in fig. 1. Table-1 lists the numerical parameters used for the simulation. Two well established theories are there to study the ballistic performance namely "Top of the barrier (TOB)" model and "Non equilibrium Greens function (NEGF)". We used NEGF mode space approach in effective mass approximation in the environment of silvaco's Atlas device simulation tool.

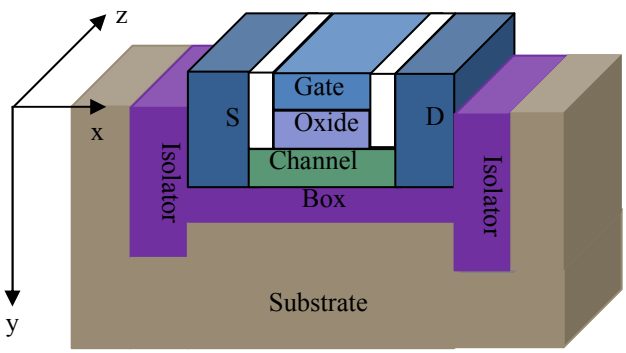


Fig. 1. XOI device structure under study. Detail of physical and material parameters are given in Table I.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

TABLE I. PARAMETERS USED IN THE SIMULATION OF THE DEVICE STRUCTURE SHOWN IN FIG. 1.

Parameter name	*Value*	*Description*
ε_{ins}	22	HfO_2 as gate oxide
t_{body}	3nm	It is the thickness of the channel. $In_{0.3}Ga_{0.7}Sb$ and $InAs_{0.7}Sb_{0.3}$ as channel material
t_{ins}	2.82 nm	Thickness of HfO_2 gate oxide, EOT = 0.5 nm
N_{body}	P type, $10^{15}/cm^{-3}$	Doping at the channel region laid right bellow the gate oxide.
N_{SD}	N type, $10^{20}/cm^{-3}$	Rest of the channel region, i,e region where N_{body} is not used.
L_{SD}	5nm	This is the extension of channel from that covered by gate oxide.
L_G	15nm	Channel region covered by gate oxide.
Φ_{metal}	4.1eV and 4.92073 eV	InGaSb and InAsSb XOI FETs respectively

B. NEGF theory

The NEGF method provides a rigorous formalism for quantum transport in nanoscale devices [13]. In this quantum modeling, schrodinger equation is solved using open boundary condition in the source and drain region (through which electron can come into or leave from the region under study). And it is solved self-consistently with the Poisson's equations to obtain the energy level, charge and current distribution [14].

NEGF are solved using two approaches namely real space and mode space method. Simulation in real space method is computationally very much costly. In case of the device where quantum confinement is strong (double gate MOS or XOI FET) the mode space approach can be used. The Hamiltonian in these device structures is separated in longitudinal and transverse directions; as a result, it greatly reduces the computational burden. Moreover, in the devices where the cross-section does not change, the sub-bands are not quantum-mechanically coupled to each other; the transport equations become essentially 1D for each sub-band that leads to use the uncoupled mode space approach [15]. The transformation of a real space effective mass Hamiltonian H_0 to a mode space is done by taking a matrix element between m^{th} and n^{th} wave functions of kth and l^{th} slices

$$H^{MS}_{mnkl} = < \psi^{k}_{m(y)} | H_0 | \psi^{l}_{n(y)} > \quad (1)$$

The transport equation in mode space is

$$\left(E - H^{MS}_{mnkl} - \Sigma^{R}_{S} - \Sigma^{R}_{D}\right) G^{R} = I \quad (2)$$

Here, E is the energy of carrier, Σ^{R}_{S} and Σ^{R}_{D} are retarded self energies carrying information about source and drain respectively, H_0 contains the information of the channel, and G^R is the retarded greens function, I is the identity matrix. After solving the above equations for each energy level, current density is calculated by [15]

$$J_x(x_i, y_j) = -\frac{2e}{hL_z\Delta_x} \sum_{k_z\sigma} \sum_{mn} \int Re(t_{i+1ijj} G^{<}_{mnii+1}\, \psi^{i}_{m}(y_j)\, \psi^{*i+1}_{n}(y_j) \frac{dE}{2\pi} \quad (3)$$

III. RESULT AND DISCUSSION

In our simulation, source and drain doping concentration is taken to be $10^{20}/cm^3$, and this doping is done throughout the region marked by L_{SD} in figure-1. Von Neumann boundary conditions for potential are used because the ballistic carrier do not create voltage drop near source and drain contact. Moreover the potential in source and drain are allowed to float, because the carrier may reflect backward and hence the total carrier concentrations in the contact can change with bias.

Figure-2 shows a comparison of I_D-V_G characteristics of InGaSb and InAsSb XOI FETs plotted at V_D=0.05V. The body thickness of the device is 3nm. Wave function penetration in the oxide region is also taken into account. To compare the performance, first the off current of the devices are made equal by tuning the gate metal work function, Φ_{metal}. It is clear from the figure that, the drain current, for InGaSb is higher than that for InAsSb at each gate bias. It is because, when device size shrinks down to the ballistic limit, carrier goes through the channel facing negligibly small or no scattering. Since current is defined as the amount of charge transferred per unit time, the ballistic current is simply determined by two factors, i) how much charge is there in the channel moving, ii) the transit time of the carrier through the channel [12]. The second factor also means the current is proportional to the carrier velocity. But the travelling carrier can have different allowed velocities through the channel, and to find the total current, summation of the currents given by each allowed velocities are to be taken. In another sense, integral over the entire energy range of travelling carriers must be taken when current components are expressed in terms of energy.

Amount of carrier in the channel depends on the density of states (DOS). For InGaSb and InAsSb materials the effective density of states are evaluated using linear interpolation from the values experimentally measured for the binaries [16], and found to be 1.93×10^{17} and 1.05×10^{17}, respectively.

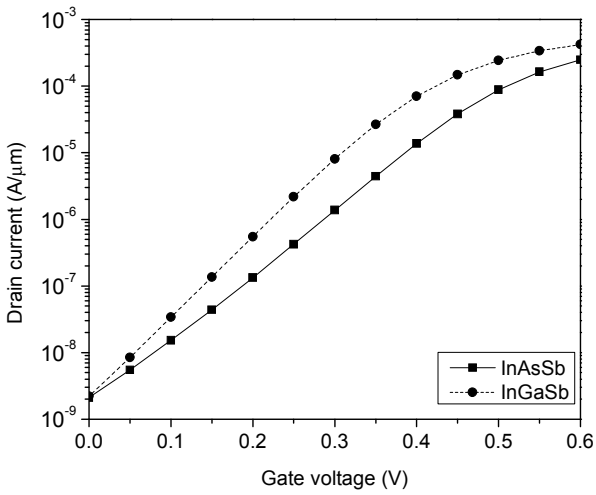


Fig. 2. I_D-V_G characteristics of 15nm InGaSb and InAsSb XOI nFETs. The structure of the devices is shown in fig.1.

Again, any heavier particle tends to move slowly, i,e the velocity of electron is inversely proportional to the effective mass (m^*). More accurately speaking, it is proportional to the inverse of square root of m^* for a given energy of travelling electron. The m^* for both materials are also calculated by linear interpolation and found to be m^*=0.039 and 0.026, for InGaSb and InAsSb respectively.
Therefore the ratio of effective DOS to m^* would indicate the amount of current through the device. This ratio for InGaSb is found to be 1.225 times higher than that of InAsSb. That is why InGaSb XOI FET gives higher current than InAsSb XOI FET.

Not only higher drain current, the I_D-V_G characteristic in fig. 2 also shows the drain current for InGaSb increases more sharply than InAsSb. This behaviour could be easily understood from semiclassical "Top of the barrier" or TOB model of ballistic transport. The highest point of electronic potential (band diagram) in the channle is called the "Top of the barrier", which is evaluated from self consistent electrostatics. According to this model, the drain current is given by [17]

$$I_D = \frac{W2q}{h}\int_{E_c}^{\infty}\sqrt{\frac{2m^*E}{\pi\hbar}}[f(E) - f(E - qV_d)]dE \quad (4)$$

Here, W is the width of the device, f(E) is the fermi function with respect to the chemical potential at the source, f(E-qV_d) is the fermi function with respect to drain chemical potential. E_C is the position of "TOB", and the above integral gives the on state current. Integral of the same equation, using the limit from negative infinity to "TOB" determines the subthreshold current, which is mainly leackage or tunnelling of carrier from source to drain [18]. That is why position of TOB actually determinces the subthreshold as well as off state current. This position of TOB is tuned by tuning gate metal work function Φ_{metal} to get same I_{OFF} for both material.

Equation (4) show that the drain current is proportional to the

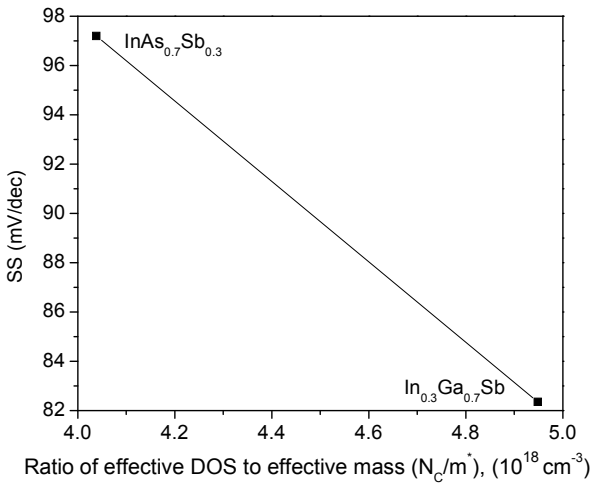


Fig. 3. Variation of Subthreshold slope (SS) with the ratio of effective DOS (density of states) to m^* (effective mass).

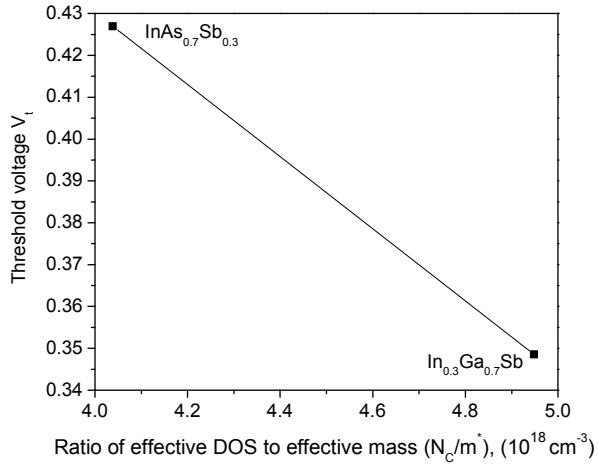


Fig. 4. Variation of threshold voltage with the ratio of effective density of states to effective mass. The workfunction the gate metal is tuned to give the same off current for InGaSb and InAsSb materials.

square root of the effective mass. Further, the rate of change of current with respect to gate bias also has similar proportionality relation. Since, InGaSb possess slightly higher effective mass, current increases much sharply leading to lower subthreshold slope (SS), and finally saturates at a higher value in case of InGaSb XOI FET. Fig. 3 shows the SS is reduced from 97.186 to 82.351 for changing the material from InAsSb to InGaSb.

Figure 4 shows the variation of threshold voltage of InGaSb and InAsSb XOI nFETs as a function of the ratio of effective density of states to effective mass. The threshold voltages are determined by substucting half of the applied drain bias from the point on voltage axis where the highest slope of the I-V curve cuts. It is found that InGaSb gives lower threshold voltage than InAsSb. The threshold voltage as well as the off current of the device depends on the barrier height at zero gate biases. This barrier height can be controlled by the work function of the gate metal. It is important that, we tuned the work function of the gate contact to make the off current equal for both cases so that other parameters can be compared.

IV. CONCLUSION

In this paper, we studied the ballistic performance of InAsSb and InGaSb XOI nFETs. The performance of the devices is compared for 15nm channel with 2.82 nm oxide thickness (0.5 nm EOT) using NEGF method. The source and drain regions are taken to be heavily doped (10^{20}/cm^3) and the channel is very lightly p type doped which can be considered as intrinsic. Gate metal work function is tuned to make the off state current equal for both of the material and the wave function penetration to the oxide is also taken into account. Source drain contacts are allowed to float to treat the carrier reflection accurately. We found that, in ballistic regime, the performance is opposite to what is expected for long channel MOSFET. InAsSb possess relatively higher bulk electron mobility, but in ballistic regime, InGaSb XOI FET gives higher drain current for all positive gate bias. With increased on state current, InGaSb also gives lower threshold voltage (V_T) and

subthreshold slope (SS). The study reveal that InGaSb would be the suitable material for low power high speed logic nanoscale FETs.

REFERENCES

[1] *The International Technology Roadmap for Semiconductors* (2012). [Online]. Available: http://www.itrs.net/Links/2012ITRS/Home2012.htm

[2] Isabelle Ferain, Cynthia A. Colinge & Jean-Pierre Colinge "Multigate transistors as the future of classical metal-oxide-semiconductor field-effect transistors*", Nature*, Vol.479, pp. 310-316, November 17, 2011.

[3] Farzin Assad, Zhibin Ren, Dragcia Vasileska, Supriyo Datta, Mark Lundstrom "On the Performance Limits for Si MOSFETs: A Theoretical Study", *IEEE Trans. Electron Devices*, vol. 47, No 1, pp.232-240, January 2000.

[4] G. K. Celler and Sorin Cristoloveanu, "Frontiers of silicon-on-insulator", *J. Appl. Phys.*, Vol.93, pp. 4955-4978, Number. 9, pp.4955-4978, May, 2003

[5] Hung-Sheng Chen and Sheng S. Li, "A Model for analyzing the interface properties of a semiconductor-insulator-semiconductor structure. I. Capacitance and Conductance techniques*", IEEE Trans. Electron Devices*, Vol.39, No.7, pp. 1740-1746, July 1992.

[6] Kuniharu Takei, Morten Madsen, Hui Fang, Rehan Kapadia, Steven Chuang, Ha Sul Kim, Chin-Hung Liu, E. Plis, Junghyo Nah, Sanjay Krishna, Yu-Lun Chueh, Jing Guo, and Ali Javey "Nanoscale InGaSb Heterostructure Membranes on Si Substrates for High Hole Mobility Transistors", *Nano Lett.,* Vol. 12. pp 2060–2066, March, 2012.

[7] Annesh Nainani, Ze Yuan, Tejas Krishnmohon, Brian R Bennett, J. Brad Boss, Mathew Reason, Mario G Ancona, Yoshio Nishi and Krishna C. Saraswat "$In_xGa_{1-x}Sb$ channel p-metal–oxide semiconductor filed effect transistors: Effect of strain and heterostructure design", *J. Appl. Phys.*, Vlo.110, Issue.1, pp. 0145031-0145039, July 2011.

[8] Jesus A. del Alamo*, "Nanometer-scale electronics with III-V compound semiconductors",* Nature, vlo.479, pp.317-323, November 2011.

[9] Hyunhyub Ko, Kuniharu Takei, Rehan Kapadia, Steven Chuang, HuiFang, Paul W. Leu,Kartik Ganapathi, Elena Plis, Ha Sul Kim, Szu-Ying Chen, Morten Madsen, Alexandra C. Ford, Yu-Lun Chueh, Sanjay Krishna, Sayeef Salahuddin & Ali Javey "Ultrathin compound semiconductor on insulator layers for high-performance nanoscale transistors" *Nature*, vol. 468, no. 7321, pp. 286-289, Nov. 2010

[10] Junji Koga, Shin-ichi Takagi and Akira Toriumi, "Influences of Buried-Oxide Interface on Inversion-Layer Mobility in Ultra-Thin SOI MOSFETs", *IEEE Trans. Electron Devices*, Vol.49, No.6, pp. 1042-1048, June 2002

[11] Hui Fang, Steven Chuang, Kuniharu Takei, Ha Sul Kim, Elena Pils, Ching-Hung Liu, Sanjay Krishna, Yu-Lun, Chueh and Ali Javey "Ultrathin-Body High-Mobility InAsSb-on-Insulator Field-Effect Transistors" *IEEE electron device letters*, vol.33, no.4, april 2012.

[12] Yan Guo, Kai-Tak Lam, Yee-Chia Yeo, and Gengchiau Liang, "Ultimate Performane Projection of Ballistic III-V Ultra-Thin-Body MOSFET", IEEE 5th International Nanoelectronic Conference (INEC), pp. 226-227, January 2013.

[13] NEGF theory 1 S. Datta. Electronic Transport in Mesoscopic Systems. Cambridge Univ.Press

[14] NEGF theory 2 R. Venugopal,Z. Ren, S. Datta, and M. S. Lundstrom "Simulating quantum transport in nanoscale transistors: Real versus mode-space approaches", *Journal of Applied physics*, Vol.92, No.7, October 2002

[15] NEGF theory 3 Mathieu Luisier,Andreas Schenk, and Wolfgang Fichtner "Quantum transport in two- and three-dimensional nanoscale transistors: Coupled mode effects in the nonequilibrium Green's function formalism", *Journal of Applied physics*, June 2006

[16] Silvaco's Atlas user manual, version 5.18.3.R

[17] Anisur Rahman, Jing Guo, Supriyo Datta, and Mark S. Lundstrom, "Theory of Ballistic Nanotransistors", *IEEE Trans. Electron Devices*, Vol. 50, No. 9, September 2003.

[18] Khairul Alam, Shinichi Takagi, and Mitsuru Takenaka, "Analysis and Comparison of L Valley Transport in GaAs, GaSb, and Ge Ultra-Thin-Body Ballistic nMOSFETs", *IEEE Trans. Electron Devices,* Vol. 60, Issue. 12, December 2013.

Influence of band parameter of gate dielectrics on the ballistic performance at same EOT

Md. Nur Kutubul Alam, Muhammad Shaffatul Islam, Md. Raifqul Islam
Department of Electrical and Electronic Engineering
Khulna University of Engineering & Technology
Khulna-9203, Bangladesh
alamjhilam@gmail.com

***Abstract*— Different high-k dielectrics have been supposed to provide same device performance if their physical thickness are not equal but equivalent, called "Equivalent oxide thickness" (EOT). For ultra thin body (UTB) devices like XOI, despite of EOT, conduction band offset (ΔE_C) at gate oxide-channel interface dominates the ballistic performance. In case of $In_{0.3}Ga_{0.7}Sb$ XOI nFET using Al_2O_3 and HfO_2 with 0.5 nm EOT, we show the threshold voltage decreases and the subthreshold slope (SS) increases with increase in ΔE_C.**

Keywords—XOI; NEGF; Ballistic transport; InGaSb; High K; gate dielectric.

I. INTRODUCTION

Transistors are now approaching to their physical limit since their size has been reduced from 10 µm to 22 nm in last four decades for the search of better performance [1-3]. At this extremely scaled size, various adverse effects like short channel and hot electron effects, parasitic capacitance, leakage current, problem of transistor isolation etc. arises. To solve these problems, silicon-on-insulator (SOI) structure has been adopted which drastically improved the device performance by suppressing these adverse effects [4,5]. Replacement of Si from the channel with III-V materials would further allow the lowering of operating voltage without sacrificing device performance due to their better transport properties [6,7 (Kuniharu, Nainani)]. That is why, taking the advantage of available process technology [8], n-type InAs FETs were experimentally demonstrated on Si/SiO_2 substrate by Javey et. al [9] and the structure is termed as "XOI" to represent compound semiconductor-on-insulator platform.

Not only the device architecture, but also the gate dielectric used is also a major issue. With decreasing gate length, area under gate dielectric decreases causing the oxide capacitance (C_{ox}) to decrease. That is why to maintain the control over channel, oxide thickness is needed to be decreased and thereby keeping required level of C_{ox}, I_{ON} and to control the threshold voltage roll off. In case of bulk Si MOSFET, thickness of gate oxide (SiO_2) is reduced over the years from 300nm for 10µm gate length (L_G) to 1.2nm for 65nm L_G. But use of that thin SiO_2 brings reliability issues due to high leakage current, dielectric breakdown, reduced channel mobility and so on [10,11]. One solution to this is replacement of SiO_2 with a material having higher dielectric constant. High-k material can be grown physically thicker and still providing the same oxide capacitance (C_{ox}) as thin SiO_2 does. In that case the thickness of high-K material is equivalent to the thickness of SiO_2, which is termed as "Equivalent oxide thickness", or "EOT". The thick high-k dielectric then offers required control over the channel with significant reduction in leakage current [11]. It has also been expected for different high-K material having different dielectric constant, if their EOT is the same, they would give same C_{ox} and therefore provide same device performance.

But in this study we found that for ultra thin body (UTB) devices like XOI, not the EOT solely control the device performance, but the band parameters like conduction band offset (ΔE_C) between gate dielectric and channel material has a great influence on the ballistic I-V characteristics. Using Non equilibrium Greens function (NEGF) method, we studied the ballistic performance of $In_{0.3}Ga_{0.7}Sb$ XOI nFET using Al_2O_3 and HfO_2 as gate oxide. In both cases, we used 0.5nm EOT. Wave function penetration to the oxide has also been taken into account. It is found that the threshold voltage decreases and the subthreshold swing (SS) increases with the higher value of ΔE_C. All other parameters affecting I-V characteristics (e,g gate metal work function) has been kept same in both cases.

II. QUANTUM BALLISTIC SIMULATION

A. Device structure

The device structure studied here is shown in fig. 1. Table-1 lists the numerical parameters used for the simulation. Two well established theories are there to study the ballistic performance namely "Top of the barrier (TOB)" model and "Non equilibrium Greens function (NEGF)". We used NEGF method in mode space approach with effective mass approximation in the environment of silvaco's Atlas device simulation tool.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

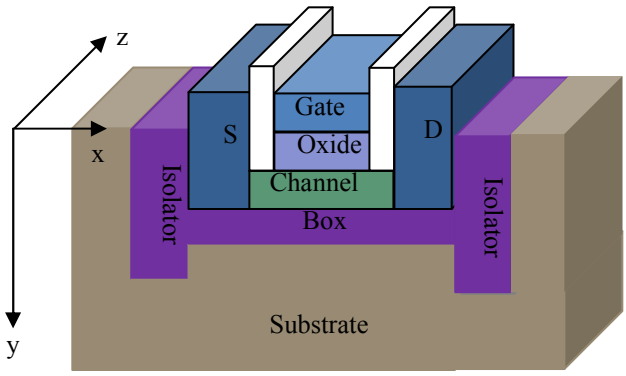


Fig. 1. XOI device structure under study. Detail of physical and material parameters are given in Table I.

TABLE I. PARAMETERS USED IN THE SIMULATION OF THE DEVICE STRUCTURE SHOWN IN FIG. 1.

Parameter name	***Value***	***Description***
ε_{ins}	22	HfO_2 as gate oxide
t_{body}	3nm	It is the thickness of the channel. $In_{0.3}Ga_{0.7}Sb$ and $InAs_{0.7}Sb_{0.3}$ as channel material
t_{ins}	2.82 nm	Thickness of HfO_2 gate oxide, EOT = 0.5nm
N_{body}	P type, $10^{15}/cm^{-3}$	Channel region laid right bellow the gate oxide.
N_{SD}	N type, $10^{20}/cm^{-3}$	Rest of the channel region, i,e region where N_{body} is not used.
L_{SD}	5nm	This is the extension of channel from that covered by gate oxide.
L_G	15nm	Channel region covered by gate oxide.

B. *NEGF theory*

The NEGF method provides a rigorous formalism for quantum transport in nanoscale devices [12]. In this quantum modeling, schrodinger equation is solved using open boundary condition in the source and drain region (through which electron can come into or leave from the region under study). And it is solved self-consistently with the Poisson's equations to obtain the energy level, charge and current distribution [13].

NEGF are solved using two approaches namely real space and mode space method. Simulation in real space method is computationally very much costly. In case of the device where quantum confinement is strong (double gate MOS or XOI FET) the mode space approach can be used. The Hamiltonian in these device structures is separated in longitudinal and transverse directions; as a result, it greatly reduces the computational burden. Moreover, in the devices where the cross-section does not change, the sub-bands are not quantum-mechanically coupled to each other; the transport equations become essentially 1D for each sub-band that leads to use the uncoupled mode space approach [14]. The transformation of a real space effective mass Hamiltonian H_0 to a mode space is done by taking a matrix element between m^{th} and n^{th} wave functions of kth and l^{th} slices

$$H_{mnkl}^{MS} = < \psi_{m(y)}^{k} | H_0 | \psi_{n(y)}^{l} > \quad (1)$$

The transport equation in mode space is

$$\left(E - H_{mnkl}^{MS} - \Sigma_S^R - \Sigma_D^R\right) G^R = I \quad (2)$$

Here, E is the energy of carrier, Σ_S^R and Σ_D^R are retarded self energies carrying information about source and drain respectively, H_0 contains the information of the channel, and G^R is the retarded greens function, I is the identity matrix. After solving the above equations for each energy level, current density is calculated by [14]

$$J_x(x_i, y_j) = -\frac{2e}{hL_z\Delta_x} \sum_{k_z\sigma} \sum_{mn} \int Re(t_{i+1ijj} G_{mnii+1}^{<} \psi_m^i(y_j) \psi_n^{*i+1}(y_j) \frac{dE}{2\pi} \quad (3)$$

III. RESULT AND DISCUSSION

In our simulation, source and drain doping concentration is taken to be $10^{20}/cm^3$, with zero gate underlap. Von Neumann boundary conditions for potential are used because the ballistic carrier do not create voltage drop near source and drain contact. Moreover the potential in source and drain are allowed to float, because the carrier may reflect backward and hence the total carrier concentrations in the contact can change with bias.

The I_D-V_G characteristic of the InGaSb XOI FET for two gate dielectrics is shown in Fig. 2. It is obtained at V_D=0.05V. The body thickness of the device is 3nm. Wave function penetration in the oxide region is also taken into account. It is found that for HfO_2, the drain current is lower than that for Al_2O_3, although they have same EOT. Generally if two gate oxides have the same EOT, the same oxide capacitance, and thereby same electrostatic control over the channel and hence identical device performance is expected. But in case of XOI structure, the deviation is found because of conduction band discontinuity (ΔE_C) between the oxide and the channel.

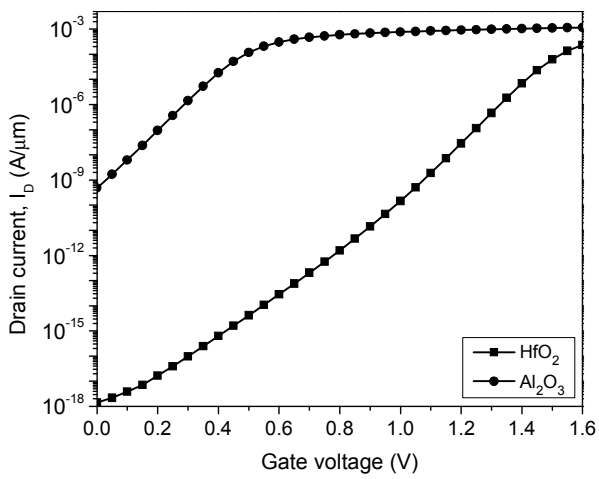


Fig. 2. I_D-V_G characteristics of 15nm InGaSb XOI nFETs using Al_2O_3 and HfO_2 as gate oxide. The structure of the devices is shown in fig. 1.

The ballistic current through any device depends on the fermi level at the source contact as well as on the highest value of conduction band along the channel, which is called "Top of the barrier" or TOB. If TOB lies above the fermi level at the source, it effectively blocs any current conduction. But if it lies bellow the source fermi level, then it modulates the current through the device.

In our simulation, fermi level at the source is kept fixed at zero eV. When the gate metal work function (Φ_{metal}) is kept fixed which is greater than 4 eV in our study, lower ΔE_C implies the higher difference between Fermi level at the source contact and the TOB at the channel. That further implies lowest number of energy states are available for conduction at every gate bias. Since lower number of states available for conduction leads to lower drain current, lower conduction band offset between oxide and channel leads to lower drain current, provided that Φ_{metal} is kept fixed. For $In_{0.3}Ga_{0.7}Sb$ channel, $\Delta E_C(Al_2O_3)$ = 2.88eV and $\Delta E_C(HfO_2)$ = 1.76eV [15]. That is why HfO_2 is giving lower drain current despite of same EOT.

Since the TOB is situated at different level for zero gate bias, we carried out the simulate for a wide range of gate bias to ensure the turn on of both devices. But in real life application both devices could be operated at a much less bias range by tuning the gate work function. The I-V curve shown in fig. 2 for HfO_2 can be shifted to the left by reducing Φ_{metal} used for HfO_2 based device. And, if it were choosen to be 1.12eV less than that used for Al_2O_3 based device (since $\Delta E_C\{Al_2O_3\}$-$\Delta E_C\{HfO_2\}$=1.12 eV), the two curves will be at closest position. To understand the effect of ΔE_C only, other influential factors like Φ_{metal} are so choosen that causes the energy band inside the oxide to start from the same maximum height.

Fig. 3 shows the change in subthreshold slope (SS) as a function of ΔE_C between oxide and the channel. It is found that the subthreshold slope decreases with decreasing ΔE_C.

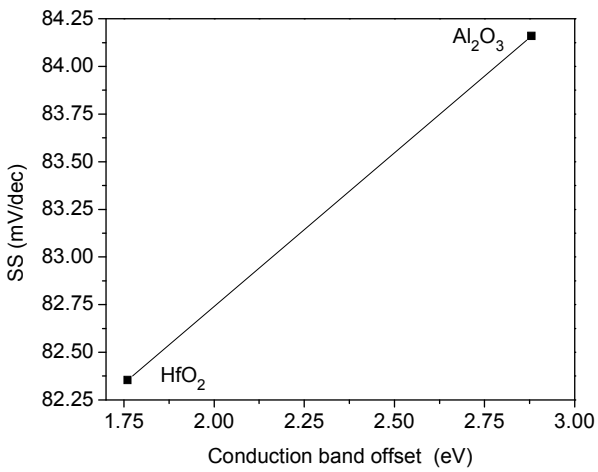


Fig. 3. Variation of Subthreshold slope (SS) for the use of HfO_2 and Al_2O_3 as gate oxide on InGaSb XOI nFET.

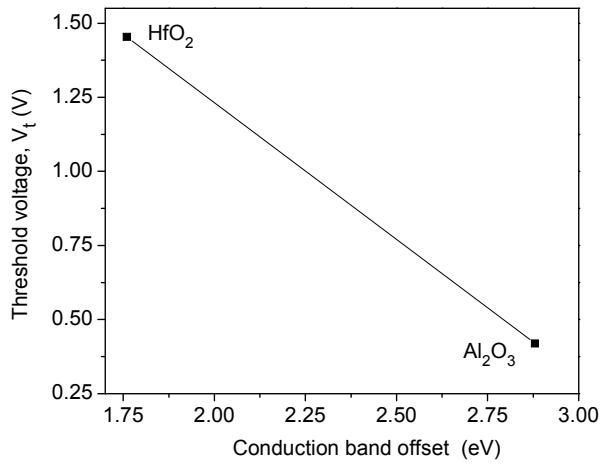


Fig. 4. Variation of threshold for the use of HfO_2 and Al_2O_3 as gate oxide on InGaSb XOI nFET.

The lower SS implies the sharp change of drain current with gate bias and the better electrostatic control of gate over the channel. Better control of gate also indicates the lower influence of drain voltage on I-V characteristic and hence supressed short channel effect (SCE), like drain induced barrier lowering.

Fig. 4 shows the variation of threshold voltage as a function of ΔE_C between oxide and the channel. It shows the threshold voltage decreases with increasing ΔE_C. For HfO_2 gate oxide, the threshold voltage found in our case is 1.454V, which is not acceptable any way. But it can be brought very close to zero by gate work function engineering, so it is not a problem at all. Our study reveals a great impact of ΔE_C on the SS and the threshold voltage.

IV. CONCLUSION

We analyzed the influence of conduction band offset, ΔE_C between the gate oxide and the channel on the performance of $In_{0.3}Ga_{0.7}Sb$-on-insulator or the XOI FET. We took Al_2O_3 and HfO_2 as gate oxides having 0.5 nm of equivalent oxide thickness for each case. The ballistic performance of the device is computed by NEGF method. The source and drain regions are taken to be heavily doped($10^{20}/cm^3$) and the channel is left undoped. Gate metal work functions are so chosen that the energy band at the gate metal-oxide interface lies at 2.92eV at zero gate bias. Wave function penetration to the oxide is also taken into account. Source drain contacts are allowed to float to treat the carrier reflection accurately. We found that the drain current increases with increase in ΔE_C. Also the threshold voltage (V_T) decreases and SS increases with increasing ΔE_C. SS of the device under study goes bellow 90mV/dec for both cases. So both of the design are acceptable for logic application. Although we get lower drain current for HfO_2 because of its lower ΔE_C with the InGaSb channel, it can be controlled by tuning the gate metal work function. That is why the overall performance of HfO_2 is better since it gives lower SS.

REFERENCES

[1] *The International Technology Roadmap for Semiconductors* (2012). [Online]. Available: http://www.itrs.net/Links/2012ITRS/Home2012.htm

[2] Isabelle Ferain, Cynthia A. Colinge & Jean-Pierre Colinge "Multigate transistors as the future of classical metal-oxide-semiconductor field-effect transistors*", Nature*, Vol.479, pp. 310-316, November 17, 2011.

[3] Farzin Assad, Zhibin Ren, Dragcia Vasileska, Supriyo Datta, Mark Lundstrom "On the Performance Limits for Si MOSFETs: A Theoretical Study", *IEEE Trans. Electron Devices*, vol. 47, No 1, pp.232-240, January 2000.

[4] G. K. Celler and Sorin Cristoloveanu, "Frontiers of silicon-on-insulator", *J. Appl. Phys.*, Vol.93, pp. 4955-4978, Number. 9, pp.4955-4978, May, 2003

[5] Hung-Sheng Chen and Sheng S. Li, "A Model for analyzing the interface properties of a semiconductor-insulator-semiconductor structure. I. Capacitance and Conductance techniques*", IEEE Trans. Electron Devices*, Vol.39, No.7, pp. 1740-1746, July 1992.

[6] Kuniharu Takei, Morten Madsen, Hui Fang, Rehan Kapadia, Steven Chuang, Ha Sul Kim, Chin-Hung Liu, E. Plis, Junghyo Nah, Sanjay Krishna, Yu-Lun Chueh, Jing Guo, and Ali Javey "Nanoscale InGaSb Heterostructure Membranes on Si Substrates for High Hole Mobility Transistors", *Nano Lett.,* Vol. 12. pp 2060–2066, March, 2012.

[7] Annesh Nainani, Ze Yuan, Tejas Krishnmohon, Brian R Bennett, J. Brad Boss, Mathew Reason, Mario G Ancona, Yoshio Nishi and Krishna C. Saraswat "$In_xGa_{1-x}Sb$ channel p-metal–oxide semiconductor filed effect transistors: Effect of strain and heterostructure design", *J. Appl. Phys.*, Vlo.110, Issue.1, pp. 0145031-0145039, July 2011.

[8] Jesus A. del Alamo*, "Nanometer-scale electronics with III-V compound semiconductors",* Nature, vlo.479, pp.317-323, November 2011.

[9] Hyunhyub Ko, Kuniharu Takei, Rehan Kapadia, Steven Chuang, HuiFang, Paul W. Leu,Kartik Ganapathi, Elena Plis, Ha Sul Kim, Szu-Ying Chen, Morten Madsen, Alexandra C. Ford, Yu-Lun Chueh, Sanjay Krishna, Sayeef Salahuddin & Ali Javey "Ultrathin compound semiconductor on insulator layers for high-performance nanoscale transistors" *Nature*, vol. 468, no. 7321, pp. 286-289, Nov. 2010

[10] G. Ribes, J. Mitard, M. Denais, S. Bruyere, F. Monsieur, C. Parthasarathy, E. Vincent, and G. Ghibaudo, "*Review on High- k Dielectrics Reliability Issues*", IEEETransactions on Device and Materials Reliability, vol. 5, no. 1, march 2005

[11] E. P. Gusev, V. Narayan, M. M. Frank, "*Advanced high-k dielectric stacks with polySi and metal gates: Recent progress and current challenges*", IBM Journal of Research and Development, Vol.50, Issue.4.5, Page(s).387-410, July 2006.

[12] NEGF theory 1 S. Datta. Electronic Transport in Mesoscopic Systems. Cambridge Univ.Press

[13] NEGF theory 2 R. Venugopal,Z. Ren, S. Datta, and M. S. Lundstrom "Simulating quantum transport in nanoscale transistors: Real versus mode-space approaches", *Journal of Applied physics*, Vol.92, No.7, October 2002

[14] NEGF theory 3 Mathieu Luisier,Andreas Schenk, and Wolfgang Fichtner "Quantum transport in two- and three-dimensional nanoscale transistors: Coupled mode effects in the nonequilibrium Green's function formalism", *Journal of Applied physics*, June 2006

[15] J. Robertson and B. Falabretti, "*Band offsets of high K gate oxides on III-V semiconductors*", J. Appl. Phys, 100, 014111 (2006)

Ballistic Transport Characteristic of InGaAs Quantum Well Surface Channel MOSFET Including Effects of Physical Device Parameter

*Abir Shadman, Ehsanur Rahman, Sudipta Romen Biswas, Kanak Datta and Quazi D.M. Khosru

Department of Electrical & Electronic Engineering,
Bangladesh University of Engineering & Technology
Dhaka-1000, Bangladesh
* abirshadman@gmail.com

Abstract—**In this paper, impact of device & process parameter variation on quantum ballistic Current-Voltage (I-V) characteristics of a surface channel, High K stack gate Quantum Well MOSFET is simulated. Physical device parameters like channel thickness, gate dielectric thickness and process parameters like doping density have direct effects on quantum ballistic current. We use mode space approach with NEGF formalism to simulate Current-Voltage (I-V) characteristics. Short Channel effects (SCE) are studied from the simulation for these variations. Observed effect is scaling dielectric & channel thickness results in better subthreshold slope & Drain induced barrier lowering at the cost of On-current. By increasing doping concentration, ballistic current can be improved. However with increasing doping density, SCE effects are compromised.**

***Index Terms*— Quantum Well MOSFET, 2D Electrostatics, III-V Semiconductors, Delta Doping, Ballistic Transport**

I. Introduction

Si CMOS has reached almost bottleneck because of practical & fundamental considerations [1], [2]. So, new and alternate device structures and device materials are in great demand for research purposes. III-V materials and their associated device structures [3] have emerged as potential candidates. Because they can be used for low power, high frequency operations. III-V materials are mostly direct bandgap materials, have higher mobility due to smaller effective masses. Performance of III-V MOSFET is superior to Si counterpart. A 55nm Quantum Well device structure that is surface channel with high K stack as gate dielectric has been reported in recent literature [4]. Effect of various device and process parameters like doping density, channel thickness, gate dielectric thickness on ballistic performance of the device structure is observed in this paper. The simulation approach is evolved from Nanomos 2.5[5]. III-V HEMT device structures have simulated with same approach [6]. We follow the same approach for the quantum ballistic simulation of the structure in [4].

II. Device Structure And Simulation Approach

We developed a device simulator to characterize the device in Fig. 1. The device has 5nm delta-doped n type $In_{0.52}Al_{0.48}As$ layer separated from channel by 1nm $In_{0.52}Al_{0.48}As$ spacer. High-k stack consists of 0.5nm Al_2O_3 & 6.5nm HfO_2. In real devices, the current flow is two- dimensional from the raised source to the drain through the doped heterostructure stack.

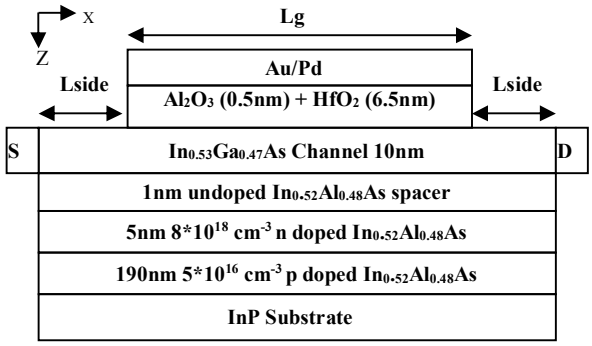


Fig. 1. Rectangular intrinsic simulation region where quantum ballistic transport dominates, used in this study.

Rather than attempting to simulate the contacts and the associated metal semiconductor contact resistance, we placed ideal contacts at the two ends of the channel to simplify the structure. Although this approach does neglect some practical aspects of device operation like source access issue, source starvation etc. it provides us with an upper level projection of device performance.

Fig. 1 shows the rectangular 'intrinsic' simulation region used in this study. Here, Lg refers to the gate length; Lside refers to the distance from gate to n+ layer edge of the device. To maintain numerical stability we have considered length of Lside region to be equal of Lg. The total simulation region is divided into *Nx*Nz* number of nodes, where *Nx* and *Nz* refers to the total number of nodes in X and Z direction. The 2D simulation starts by solving the Poisson equation with the charge from an initial flat band as the guess potential. The updated potential leads to a new set of spatial charge distribution. This process is repeated until desired convergence is achieved. The ballistic current is then readily calculated with the NEGF formalism in mode space for each subband. For current calculation, we consider 2 subbands in the QW channel. The Poisson equation is solved using proper boundary conditions applying a non-linear algorithm [7]. In this approach, the carrier concentration $n_{m,n}$ at node [m, n] is expressed in terms of the to-be-solved potential $V_{m,n}$ and the quasi-Fermi potential energy, Fn[8]:

$$n_{m,n} = N_C \Im_{1/2}\left[\frac{(F_n)_{m,n}+qV_{m,n}}{k_B T}\right] \quad (1)$$

Where $\Im_{1/2}$ stands for the Fermi-Dirac integral of order ½ and N_C is the effective density of states. The quasi-Fermi potential is pre-calculated from the input potential and charge density:

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

$$(F_n)_{m,n} = -q(V_{old})_{m,n} + k_B T \Im_{1/2}^{-1}\left(\frac{n_{m,n}}{N_C}\right) \quad (2)$$

We can obtain the charge distribution by solving the 1D effective mass Schrödinger equation along the Z directed slice in the 2D discretization mesh

$$-\frac{h^2}{2m_z^*}\frac{d^2}{dz^2}\Psi_i(x,z) - qV(x,z)\Psi_i(x,z) = E_i(x)\Psi_i(x,z) \quad (3)$$

Where, m_z^* is the electron effective mass in the z direction, $\Psi_i(x,z)$ and $E_i(x)$ are the wavefunction and Eigen energy for subband i at slice x. Using an uncoupled mode space approach, we divide this 2D problem into two 1D problems. We find out the total number of subbands in the quantum well channel and calculate charge density in each of these subbands using NEGF formalism. If a is the grid spacing in X direction, using the Green's function formalism [9], the 2D charge density matrix of mode i is obtained as

$$n_{2D} = \int_0^\infty \frac{1}{a\hbar}\sqrt{\frac{m_y^* k_B T}{2\pi^3}}\left[\Im_{-1/2}(\mu_S - E_l)A_S + \Im_{-1/2}(\mu_D - E_l)A_D\right]dE_l \quad (4)$$

Where

$$A_S = G\Gamma_S G^+, \qquad A_D = G\Gamma_D G^+$$

are the spectral density functions due to source/drain contacts [10]. Γ_S and Γ_D are self energy terms for source & drain. Physically this term determines the electron exchange rates between the source/drain reservoirs and the active device region. But in general it can be viewed as the measure of interaction strength due to any perturbation source. Although the device itself may be in non-equilibrium state, electrons are injected from the equilibrium source/drain reservoirs. G is the retarded green function [7]. The 3D charge density at slice x can be obtained by multiplying $n_{2D}(x)$ by the wavefunction probability and summing for all the subbands:

$$n(x,z) = \sum_i n_{2D}(x,i)|\Psi_i(x,z)|^2 \quad (5)$$

Once self-consistency is achieved, the intrinsic ballistic current is calculated as

$$I_D = \int_0^\infty \frac{q}{\hbar^2}\sqrt{\frac{m_y^* k_B T}{2\pi^3}}\left[\Im_{-1/2}(\mu_S - E_l) - \Im_{-1/2}(\mu_D - E_l)\right]T_{SD}(E_l)dE_l \quad (6)$$

Where, $T_{SD}(E_l)$ is the transmission coefficient as a function of the longitudinal energy calculated from the Green's function.

III. Results & Discussions

A. Effect of Channel thickness

Fig. 2 shows the effect of channel thickness variation on intrinsic Id-Vd characteristics for the 55nm gate length device structure. The figure shows Id-Vd characteristics for 3 different channel layer thicknesses: 9nm, 10nm, 11nm. Here gate voltage varies from 0.6V to 1.4V in step of 0.4V and drain voltage varies from 0.0V to 1V.From Fig. 2 it is observed that, as we increase channel thickness, the drain

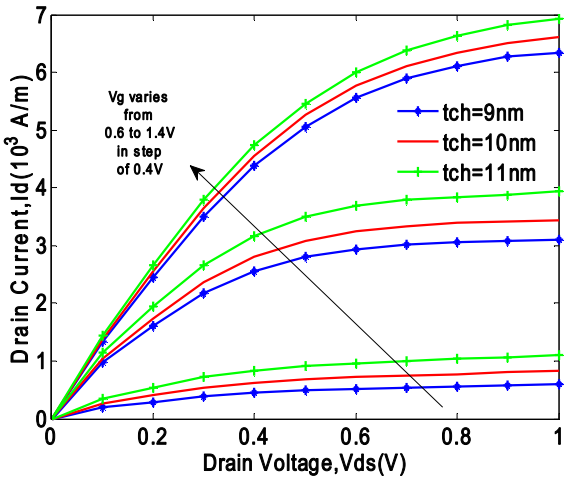


Fig. 2. Id-Vds characteristics for different channel thickness conditions

current increases. Increasing channel thickness allows accumulation of more carriers in the channel. Fig. 3 shows Id-Vg characteristics (log scale) for various channel widths at a drain bias of 0.5 volt. The logId-Vg characteristics give us an idea of the subthreshold performance of the device. Higher channel thickness leads to higher current in subthreshold region and therefore subthreshold swing (SS) and drain induced barrier lowering (DIBL) parameters increase.

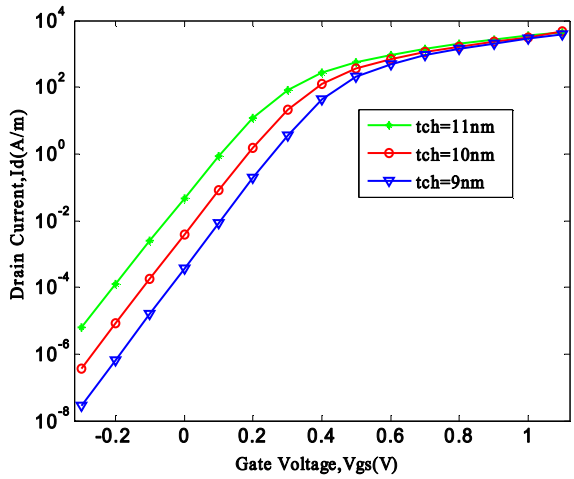


Fig.3. Id-Vgs characteristics for different channel thickness conditions (log scale). Vd=0.5V

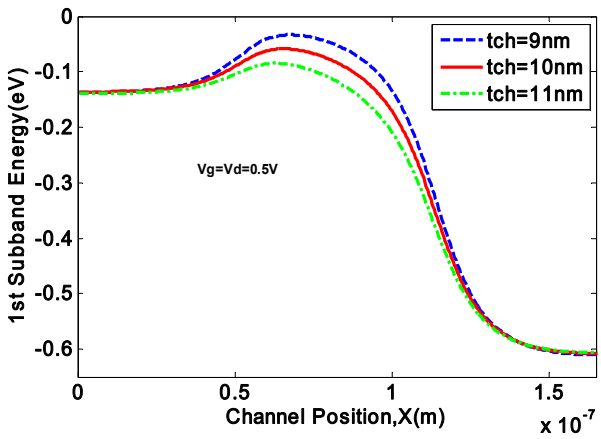


Fig. 4. 1st subband energy at different Channel thickness conditions.

Fig. 4 shows the potential barrier lowering in the channel with higher channel thickness which also supports higher drain current at increased channel thickness.

B. Effect of Gate Oxide thickness

Fig. 5 shows intrinsic I_d-V_d characteristics for the 55nm device under variable oxide thickness condition. As the oxide thickness increases, the drain current increases. This phenomenon may be attributed to 2D electrostatics inside the device and the weakening control of gate voltage over the energy barrier in the channel for the device. As oxide thickness is increased, for the same applied gate bias, the voltage drop across oxide increases and therefore control of gate voltage over the potential barrier in the channel weakens. So DIBL effect becomes more prominent in the channel region. This leads to an increase in the drain current in the ballistic regime. The barrier lowering effect can be observed from the Fig. 6. In this figure we show the 1st subband energy in the channel of the device.

Fig. 7 presents I_d-V_{gs} characteristics for the device (log scale). For higher oxide thickness, drain current increases more rapidly with increasing gate voltage which is due to greater lowering of the potential barrier in the channel at higher oxide thickness. At sub threshold region, device with higher oxide thickness shows shallower slope and therefore higher SS. Due to weaker gate control over the potential barrier, the device with higher oxide thickness would show degraded DIBL effects compared to the device with lower oxide thickness.

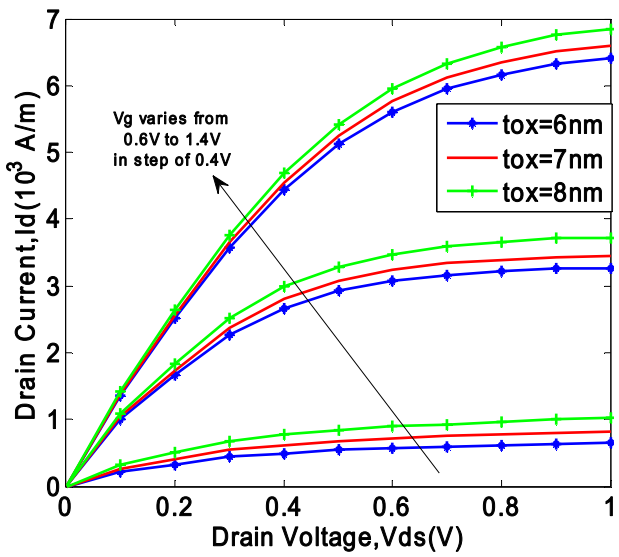


Fig. 5. Id-Vd characteristics for the device structure at different oxide thickness

C. Effect of Doping Concentration

Fig. 8 shows the I_d-V_d characteristics of the quantum well structure at different doping concentration of the 5 nm n type high doped region. As we increase the doping concentration, current in the channel increases. This phenomena can be explained from the fact that carrier in the channel is supplied from the nearby high n doped region. As we modulate the channel potential barrier by gate voltage, more and more carriers are injected into the channel from the high doped region. A higher doping concentration in the high doped region means it can supply greater amount of charge in the

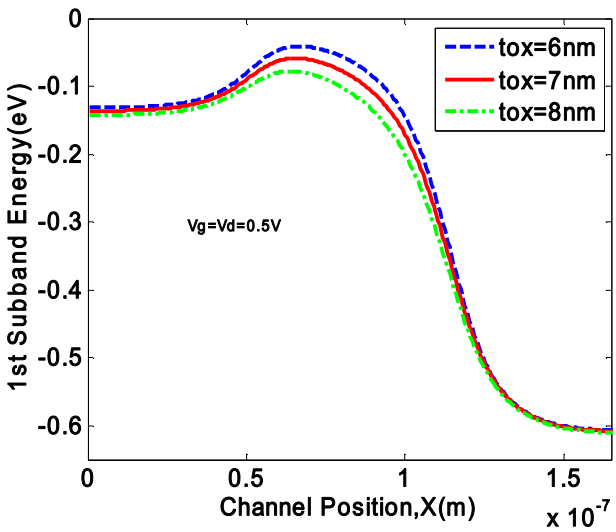


Fig. 6 1st subband energy at different Oxide thickness conditions. At Vg=Vd=0.5V

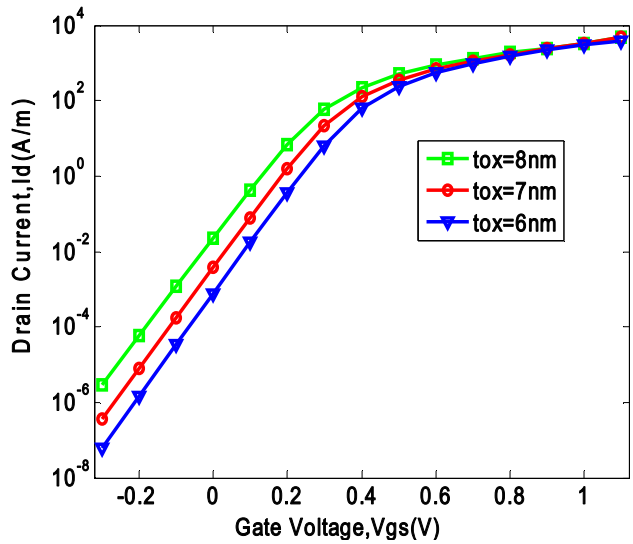


Fig. 7. Id-Vgs characteristics for different oxide thickness conditions (log scale). Vd=0.5V

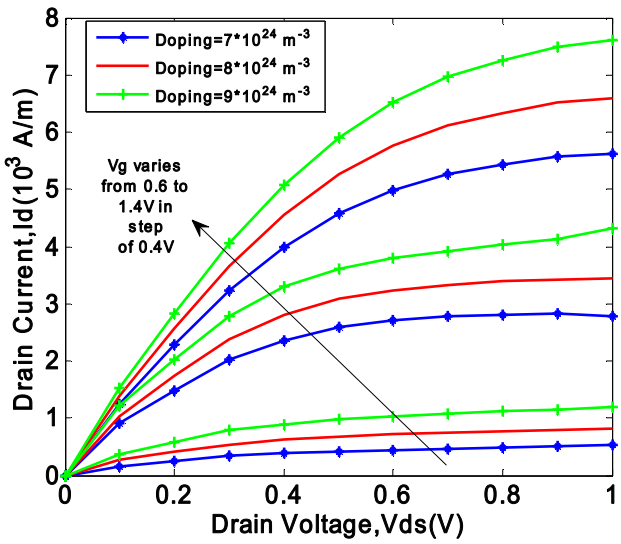


Fig. 8. Id-Vd characteristics for the device structure at different Doping concentrations.

channel for a given gate and drain bias. A higher carrier density in the channel eventually results in greater lowering of the potential barrier in the channel as evident from Fig. 9. It results in higher drain current as observed from Fig. 8.

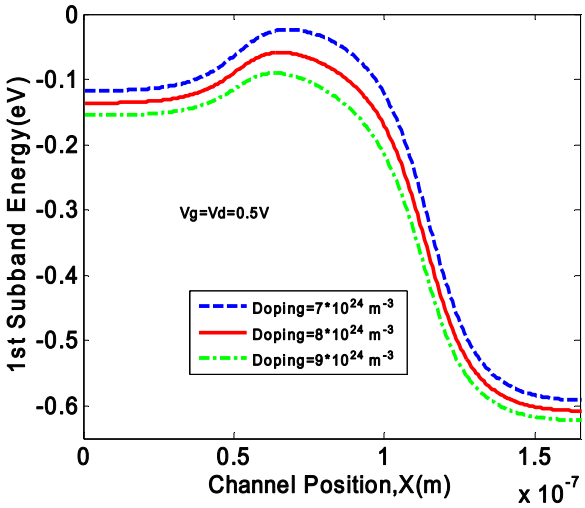


Fig. 9. 1st subband energy at different Doping Concentration. At Vg=Vd=0.5V

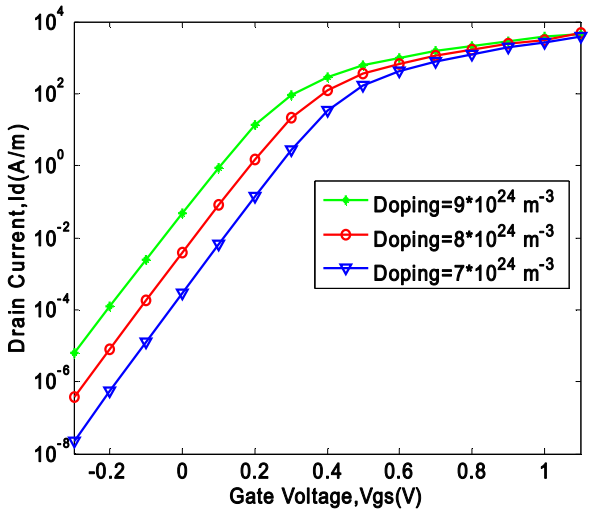


Fig. 10. Id-Vgs characteristics for different doping concentration conditions (log scale). Vd=0.5V

The Log I_d-V_{gs} characteristics have been shown in of Fig 10. It is observed from the figure that a higher doping concentration results in higher change in drain current with gate bias which is due to the same reason explained before. At sub threshold region, device with higher doping concentration shows shallower slope and therefore higher SS. Also increasing doping concentration lowers threshold voltage as observed from the Fig. 10.

IV. Conclusions

In this work, we report simulation result for physical & process parameter variation on quantum ballistic current for a 55 nm InGaAs surface channel quantum well MOSFET. Among these three parameters, process parameter (Doping concentration) produced more observable result. Reducing InGaAs channel improves short channel performance of the device, but we get a compromised ballistic On-current. Since these quantum ballistic current is usually 2-3 times the experimental one, we didn't measure accurate values of SS swing & DIBL. However we captured the trend on SCE effects with this scaling. Increasing gate dielectric material thickness improves ballistic device performance but leads to degraded short channel performance. Doping density, as expected, results in higher on-current.

Acknowledgment

The authors are thankful to Ms. Urmita Sikder & Md. Hasibul Alam for their help in simulation.

References

[1] International Technology Roadmap for Semiconductors [Online] Available: http://public.itrs.net

[2] Oktyabrsky, Ye, et. al. Fundamentals of III-V Semiconductor MOSFETs. Springer US, 2010.

[3] J.A.del Alamo, "Nanometre-scale electronics with III-V compound semiconductors," Nature, vol. 479, pp. 317-323, November 2011.

[4] M. Egard et. al., "High Transconductance Self-Aligned Gate-Last Surface Channel $In_{0.53}Ga_{0.47}As$ MOSFET", IEEE IEDM, p.303 (2011).

[5] Z.Ren, R.Venugopal, S.Goasguen, S.Datta, M.Lundstorm, "nanoMOS 2.5: A two-dimensional simulator for quantum transport in double-gate MOSFETs," Electron Devices, IEEE Transactions on, vol 50, pp. 1914-1925, Sept. 2003.

[6] Y. Liu, and M. Lundstrom, "Simulation of III-V HEMTs for high-speed low-power logic applications," ECS Transactions 19.5 (2009): 331-342

[7] Z. Ren, Nanoscale MOSFETs: Physics, Simulation, and Design, Ph.D.Thesis, Purdue University, West Lafayette, IN (2001).

[8] Liu, Yang, and Mark Lundstrom. "Simulation of III-V HEMTs for highspeed low-power logic applications." ECS Transactions 19.5 (2009): 331-342

[9] S. Datta,"Nanoscale device modeling: the Green's function method." Superlattices and Microstructures 28.4 (2000): 253-278.

[10] S. Datta, Quantum Transport: Atom to Transistor, Cambridge University Press(2005)

[11] M. S. Lundstrom, "Elementary scattering theory of the MOSFET," IEEE Electron Device Lett., vol. 18, pp. 361–363, Nov. 1997.

Quantum Ballistic Simulation Study of $In_{0.7}Ga_{0.3}As/InAs/In_{0.7}Ga_{0.3}As$ Quantum Well MOSFET

Sudipta Romen Biswas*, Kanak Datta, Abir Shadman, Ehsanur Rahman, and Quazi D. M. Khosru

Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology, Dhaka–1000, Bangladesh

*sudipta_r_biswas@yahoo.com

***Abstract*— In this work, quantum ballistic simulation study of a novel III-V $In_{0.7}Ga_{0.3}As/InAs/In_{0.7}Ga_{0.3}As$ Quantum Well MOSFET is presented. To simulate the device in quantum ballistic regime, nonequilibrium Green's function formalism has been used. 2D Poisson and Schrodinger equations are solved in self-consistent manner taking into account 2D electrostatics and other quantum mechanical effects. Strong carrier confinement in the $In_{0.7}Ga_{0.3}As/InAs/In_{0.7}Ga_{0.3}As$ quantum well allows the application of efficient mode space approach in quantum ballistic simulation. Simulation results for the QW device with 30 nm gate length are reported. At the same time, effect of gate length variation on the quantum ballistic characteristics is explored.**

***Index Terms*—Quantum Well MOSFET, Delta Doping, III-V Semiconductors, Ballistic Transport, 2D Electrostatics**.

I. Introduction

With tremendous improvement in lithography and scaling technology, smart optimization, Moore's law [1] had gone unabated for the last four decades [2]. However, scaling of planar silicon CMOS transistors is reaching its fundamental limits due to various practical limitations [3]. Scaling of planar MOSFET structures brings a number of challenges like increased short channel effects, increased gate leakage and off-state leakage, increased power dissipation etc. With excellent carrier transport properties and improved 'on' state performance over silicon, III-V compound semiconductors like: GaAs, InGaAs and InAs and their associated devices structures have emerged as potential candidates for high speed, low power logic operations [4]. III-V quantum well device structures with extremely scaled high-k gate dielectric material exhibit excellent 'on' state performance and are considered suitable for future low power logic implementation. A novel III-V quantum well device that incorporates a thin pure InAs sub-channel and extremely scaled HfO_2 dielectric material has been reported in recent literature [5]. This paper focuses on the quantum ballistic simulation study of the device structure reported in [5]. The effect of gate length scaling on ballistic performance is also explored and presented.

II. Device Structure and Simulation Approach

In this work, we simulate a simplified form of the device structure reported in [5]. Here the channel consists of high mobility $In_{0.7}Ga_{0.3}As/InAs/In_{0.7}Ga_{0.3}As$ materials. A 1 nm Si δ-doped layer is placed just 5 nm below the bottom $In_{0.7}Ga_{0.3}As$ channel layer. High-k material HfO_2 is used as gate dielectric. To improve interface properties and reduce

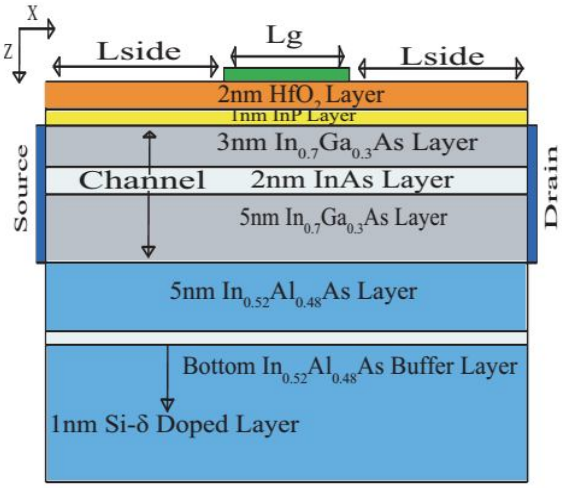


Fig. 1 Rectangular intrinsic simulation region where quantum ballistic transport is expected to dominate, used in this study. Simulation region is similar to the region applied in [7].

interface scattering, 1 nm InP layer is placed just beneath the gate oxide layer. The bottom $In_{0.52}Al_{0.48}As$ buffer layer is over 300 nm thick. Fig. 1 shows the simplified intrinsic simulation region used in this study. In real HEMT or QW MOSFET devices , the current flow usually takes place from raised drain through heterostructure stack, laterally through the channel and then through the raised source contact [7]. However, ideal contacts are considered at two ends of the channel. We hereby only simulate the region where ballistic transport is expected to dominate. In Fig. 1, Lg refers to the gate length, Lside refers to the distance from gate to n^+ edge of the device. The total simulation region is divided into N_x * N_z number of nodes, where N_x and N_z refers to the total number of nodes in X and Z direction. However, the bottom $In_{0.52}Al_{0.48}As$ layer is about 300 nm thick. To reduce total number of nodes, we apply uniform grid spacing up to a certain distance along Z direction where more accurate results are required. After that, we apply nonuniform grid spacing. At the same time, to maintain numerical stability we have considered length of Lside region to be 44 nm. The doping in the Si δ-doped region is kept at $1\sim2\times10^{12}$ cm^{-2}.

The simulation approach used in this study evolves from nanoMOS2.5 [6]. Similar simulation approach had already been applied for simulation study of III-V HEMTs [7], [8]. In this approach, 2D Poisson equation is solved using Finite Difference Method (FDM) method with proper boundary conditions by applying a nonlinear algorithm [9]. Charge distribution is obtained by solving the 1D effective mass

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

Schrodinger equation along the Z directed slice in the 2D discretization mesh:

$$-\frac{\hbar^2}{2m^*}\frac{d^2}{dz^2}\Psi_i(x,z)-qV(x,z)\Psi_i(x,z)=E_i(x)\Psi_i(x,z)$$

where m^* is the electron effective mass in the Z direction, $\Psi_i(x,z)$ and $E_i(x)$ are the wavefunction and eigenenergy for subband i at slice x. Strong confinement of carriers in the $In_{0.7}Ga_{0.3}As/InAs/In_{0.7}Ga_{0.3}As$ quantum well allows the implementation of uncoupled mode space approach in simulation that divides this 2D problem into two 1D problems. The total number of subbands in the $In_{0.7}Ga_{0.3}As/InAs/In_{0.7}Ga_{0.3}As$ quantum well is determined and carrier density in the channel is determined using NEGF formalism [10]. The 2D carrier density for mode i is obtained as:

$$n_{2D}(x,i)=\int_0^\infty \frac{1}{a\hbar}\sqrt{\frac{m_y^* k_B T}{2\pi^3}}$$
$$[\Im_{-1/2}(\mu_S-E_l)A_S+\Im_{-1/2}(\mu_D-E_l)A_D]dE_l$$

where,

$$A_S=G\Gamma_S G^+ \qquad and \qquad A_D=G\Gamma_D G^+$$

are the spectral density functions for source and drain contacts respectively [10], G is the retarded Green's function and a is the grid spacing in X direction. Using the wavefunction profiles obtained from the solution of Schrodinger equation along Z direction, a spatial charge distribution profile is obtained and used for the solution of Poisson equation:

$$n(x,z)=\sum_i n_{2D}(x,i)\left|\Psi_i(x,z)\right|^2$$

Once self-consistency is achieved, the intrinsic ballistic current is calculated as:

$$I_D=\int_0^\infty \frac{q}{\hbar^2}\sqrt{\frac{m_y^* k_B T}{2\pi^3}}$$
$$[\Im_{-1/2}(\mu_S-E_l)-\Im_{-1/2}(\mu_D-E_l)]T_{SD}(E_l)dE_l$$

And

$$T_{SD}(E_l)=trace(\Gamma_S G\Gamma_D G^+)$$

where $T_{SD}(E_l)$ is the transmission coefficient from the source to drain contact as a function of the longitudinal energy calculated from the Green's function [10].

III. Results and Discussion

Fig. 2 shows the conduction band profile and the normalized wavefunction for the 1st subband along gate confinement direction for a bias condition Vg=0.3V and Vd=0.5V. From this figure, we can observe that the carrier wavefunction is well confined in the $In_{0.7}Ga_{0.3}As/InAs/In_{0.7}Ga_{0.3}As$ quantum well which validates the application of mode space approach. Fig. 3 shows the 1st subband profile in the channel as a function of gate bias voltage at Vd=0.5V. With increased gate bias, the top of the barrier energy is lowered which leads to the injection of more carriers from the source contact. In this simulation, two subbands in the $In_{0.7}Ga_{0.3}As/InAs/In_{0.7}Ga_{0.3}As$ quantum well are considered. Fig. 4 shows the 1st subband profile in the channel at Vg=0.2V and at different intrinsic drain bias voltages. The figure reveals a slight lowering of the top of the barrier energy with increasing drain bias. At the same time, subband

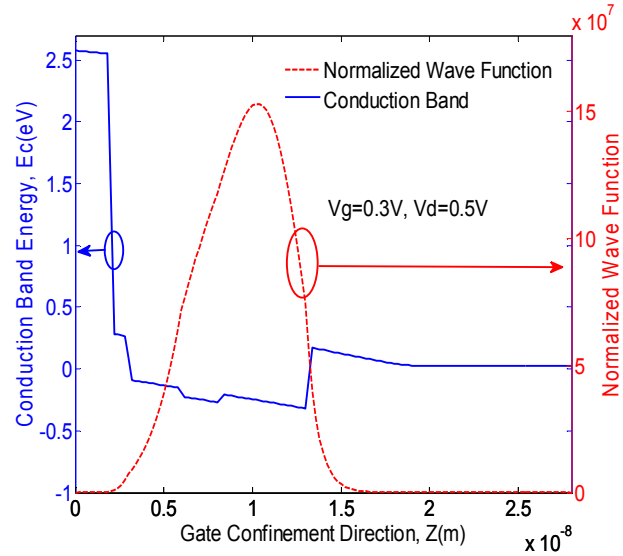


Fig. 2 Normalized wavefunction and the conduction band profile along the gate confinement direction for the 1st subband at Vd=0.5V and Vg=0.3V.

energy at the drain contact end also decreases with increasing drain bias voltage. Fig. 5 shows the 1st subband profile along with transmission coefficient obtained from simulation at Vg=0.2V and Vd=0.5V. As seen from the figure, transmission coefficient does not show an entirely step like nature. The figure also reveals that, the transmission coefficient rises to a value of 1 at just about the top of the barrier energy. At the same time, due to the consideration of two subbands in the quantum well, the combined transmission coefficient reaches a value of 2 as seen from the inset figure at the bottom left corner. Fig. 6 shows the energy resolved current density for the 1st subband obtained from simulation. This figure reveals contribution of carrier at different energies to the overall current density of the device. From this figure, it can be seen that carriers having energies above the top of the barrier energy, contribute to the overall current density. Fig. 6 also reveals a little fraction of under the barrier transport which may be attributed to the transmission coefficient not having an abrupt step like nature.

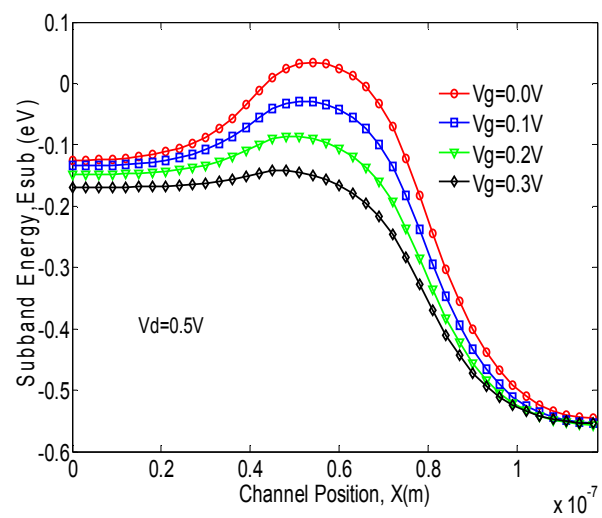


Fig. 3 1st subband profile in the channel at different intrinsic gate bias conditions for an intrinsic drain bias voltage of Vd=0.5V. With increasing gate voltage the energy barrier height between the source and the channel is lowered which leads to the injection of more carriers into the channel.

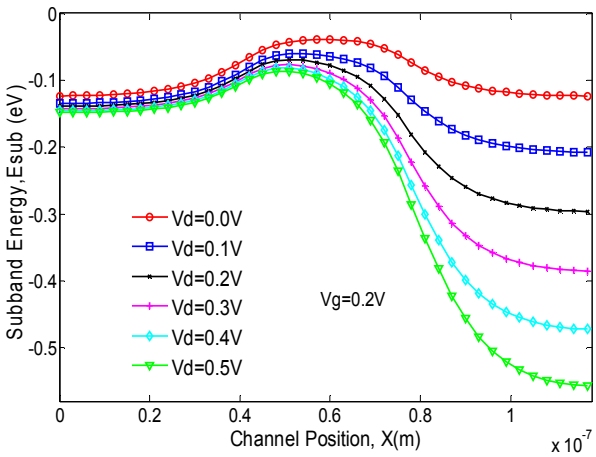


Fig. 4 1st subband profile in the channel at different intrinsic drain bias conditions for an intrinsic gate bias voltage of Vg=0.2V. With increasing drain voltage a slight lowering of the energy barrier height between the source and the channel can be seen.

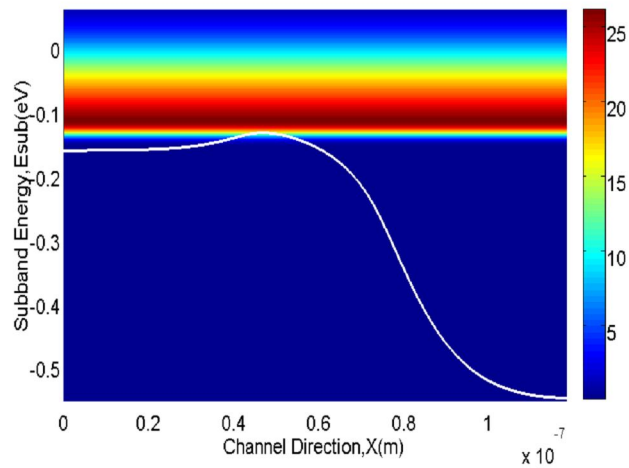


Fig. 6 Energy resolved current density for 1st subband at Vg=0.3V and Vd=0.5V. Carriers with energy higher than top of the energy barrier contribute to total ballistic current. Some under the barrier conduction is also observed.

Fig. 7 shows the intrinsic Id-Vd characteristics for the simulated device structure with Vg varying in 0.05V step. With increasing gate bias conditions, intrinsic drain current increases. MOSFETs operate by modulating the energy barrier between the source and the channel [11]. With increasing gate bias voltage, this energy barrier height is lowered which leads to injection of more carriers from the source and therefore increased energy separation between the source fermi energy level and the top of the barrier energy. This phenomenon leads to an increase in the drain current with increased gate bias voltage. Short channel effects and 2D electrostatic effects are also observed as the drain current gradually increases with increasing drain bias voltage. At Vg=0.4V and Vd=0.5V the simulated intrinsic current is about 2.5 times of the reported experimental results [5]. This discrepancy may be attributed to disregarding scattering effects which becomes significant at high gate bias conditions. Fig. 8 shows the intrinsic Id-Vg characteristics for the simulated device with 30 nm gate length. Here, we present Id-Vg characteristics for two drain bias voltages. The figure also shows the Id-Vg characteristics in log scale.

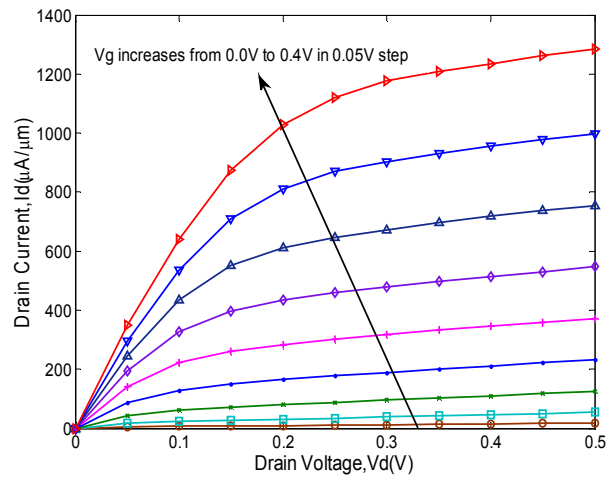


Fig. 7 Intrinsic Id-Vd characteristics for the simulated device structure at different gate bias conditions.

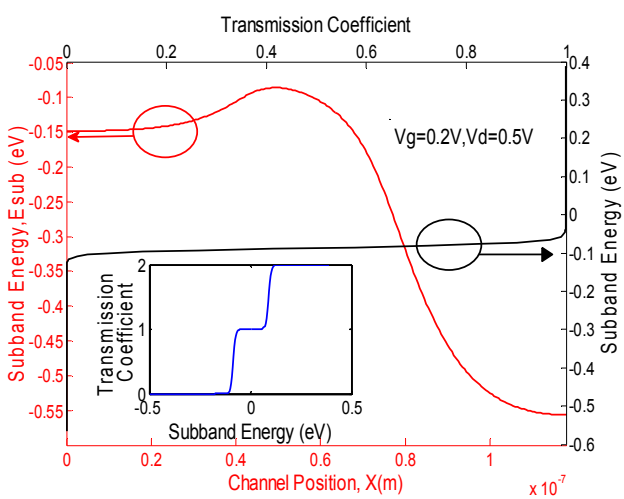


Fig. 5 1st subband energy profile along with transmission coefficient for intrinsic gate bias of 0.2V and drain bias 0.5V. The inset figure at bottom left shows transmission coefficient for 1st and 2nd subband combined at same bias.

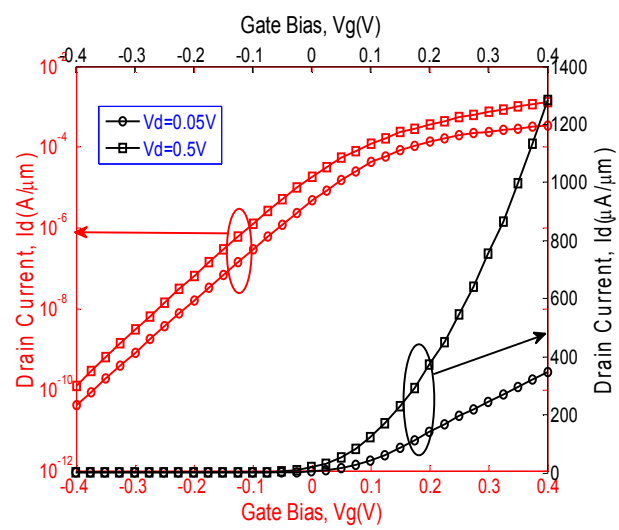


Fig. 8 Intrinsic Id-Vg and logId-Vg characteristics for the simulated device structure at two intrinsic drain bias conditions.

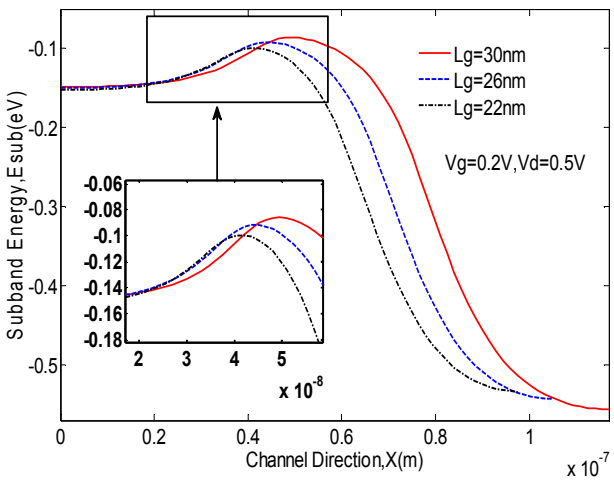


Fig. 9 1st subband energy profile at different gate length conditions at Vg=0.2V and Vd=0.5V. The lowering of the energy barrier between the source and the channel can be observed from these profiles.

Fig. 9 shows the effect of gate length on the 1st subband energy profiles in the channel at Vg=0.2V and Vd=0.5V. The figure reveals a slight lowering of the energy barrier height between source and the channel with decreasing gate length. This lowering of energy barrier increases the difference between source fermi energy and top of the barrier energy and therefore drain current increases.

Fig. 10 shows the intrinsic Id-Vd characteristics at three different gate length conditions. The figure reveals an increment in drain current with gate length scaling. The separation between the intrinsic Id-Vd curves is more observable at high drain bias regime which suggests increasing short channel effects with gate length scaling. Fig. 11 shows the effect of gate length scaling on the intrinsic Id-Vg characteristics in both normal and log scale at Vd=0.5V. From this figure, we see that device with lower gate length has a higher subthreshold current. Also, it can be observed that Subthreshold Swing (SS) increases and threshold voltage decreases with gate length scaling.

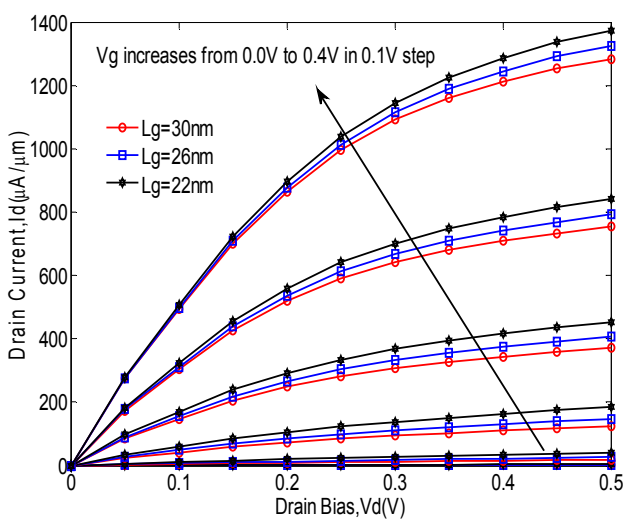


Fig. 10 Intrinsic Id-Vd characteristics at different gate length conditions for the simulated device structure. Vg varies from 0.0V to 0.4V in 0.1V step. Scaling results in increased drain current. At the same time degraded short channel performance can also be observed from the figure.

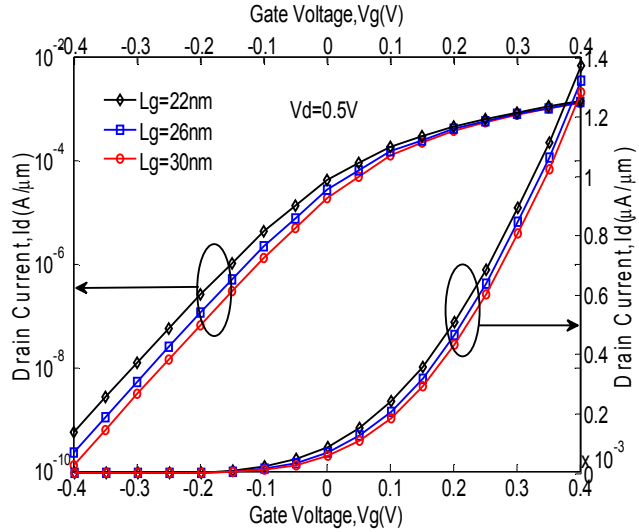


Fig. 11 Intrinsic Id-Vg and logId-Vg characteristics at different gate length conditions at intrinsic drain bias Vd=0.5V. Drain current in subthreshold region increases with decreasing gate length. Increment in Subthreshold Swing (SS) with gate length scaling can be observed as well.

IV. Conclusion

In this work, we report quantum ballistic simulation results of a novel 30 nm quantum well device using mode space approach. The quantum ballistic current is calculated using NEGF formalism. The simulated intrinsic ballistic current overshoots experimentally reported values [5] as scattering is not considered in quantum ballistic simulation. The effect of gate length scaling on the ballistic device operation and subthreshold operation is studied and presented. Scaling of gate length improves quantum ballistic current. However, simulation also reveals lowering of threshold voltage, degraded short channel performance with gate length scaling.

References

[1] G. E. Moore, "Cramming more components onto integrated circuits," in *Proc. IEEE*, vol. 86.1, pp. 82-85,Jan. 1998.

[2] S. E. Thompson, and S. Parthasarathy, "Moore's law: the future of Si microelectronics," *Materials Today*, vol. 9.6, pp. 20-25, June 2006.

[3] S. Oktyabrsky and P. Ye, Ed., *Fundamentals of III-V Semiconductor MOSFETs*, Springer US, 2010.

[4] J. A. Del Alamo, "Nanometre-scale electronics with III-V compound semiconductors," *Nature*, vol. 479, pp. 317-323, Nov. 2011.

[5] J. Lin, D. A. Antoniadis, and J. A. del Alamo, "Sub-30 nm InAs QuantumWell MOSFETs with self-aligned metal contacts and Sub-1 nm EOT HfO_2 insulator," *IEEE International Electron Devices Meeting*, pp. 32.1.1 -32.1.4, Dec. 2012.

[6] Z. Ren, R. Venugopal, S. Goasguen, S. Datta, and M. S. Lundstrom, "nanoMOS 2.5: A two-dimensional simulator for quantum transport in double-gate MOSFETs," *IEEE Transactions on Electron Devices*, vol. 50.9,pp. 1914-1925, Sept. 2003.

[7] Liu, Yang, and Mark Lundstrom. "Simulation of III-V HEMTs for highspeed low-power logic applications," *ECS Transactions*, vol. 19.5, pp.331-342, May 2009.

[8] N. Neophytou, T. Rakshit, and M. S. Lundstrom. "Performance analysis of 60-nm gate-length III–V InGaAs HEMTs: Simulations versus experiments," *IEEE Transactions on Electron Devices*, vol. 56.7, pp. 1377-1387, July 2009.

[9] Z. Ren, "Nanoscale MOSFETs: Physics, Simulation, and Design," PhD. Diss., Purdue University, West Lafayette, IN, Oct. 2001.

[10] S. Datta, "Nanoscale device modeling: the Green's function method," *Superlattices and Microstructures*, vol. 28.4, pp. 253-278, Oct. 2000.

[11] M. S. Lundstrom, and Z. Ren. "Essential physics of carrier transport in nanoscale MOSFETs," *IEEE Transactions on Electron Devices*, vol. 49.1, pp. 133-141, Jan. 2002.

Nano-bridge Enabled Three-Dimensional Gate-all-around Field Effect Transistors

Jin Yong Oh,[1] Hyun-June Jang,[2] Won-Ju Cho,[2] Jong-Tae Park,[3] and M. Saif Islam[1,*]
[1]Department of Electrical and Computer Engineering and Center for Nano and Micro Manufacturing
University of California, Davis, CA, 95616, USA
[2]Department of Electronic Materials Engineering, Kwangwoon University, Seoul , 139–701, Republic of Korea
[3]Department of Electronics Engineering, University of Incheon, Incheon, 406–772, Republic of Korea
[*]sislam@ucdavis.edu

Abstract— **We demonstrate the integration of nanowires to design gate-all-around field-effect-transistors (GAA-FETs) in the shape of horizontally aligned 3-dimensional semiconductor nano-bridges. The transistors exhibited a high on/off-current ratio, low sub-threshold swing and very low off-currents. In addition to designing low-voltage non-volatile memory cells, the photosensitivity of the FETs is employed to demonstrate several analog and digital applications such as electro-optical OR gate circuit elements and frequency doublers. Based on this approach, nanowires with high mobility can be integrated on Si platform to enable the next generation high performance nanowire FETs.**

***Index Terms*—transistors, nanowires, 3D FETs**

I. Introduction

Field effect transistors (FETs) enabled by one-dimensional nanowires grown by synthetic *bottom-up* methods have attracted attention because they are expected to present a pathway to overcome the roadblocks to Moore's law and can enable gate-all-around (GAA) transistors. However, lack of reproducibility in their growth, alignment into arrays, and contact formation to enable scalable fabrication of transistors is challenging. As a practical way of overcoming the obstacles for the integration of nanowire FETs, in-place growth techniques have been proposed [1-3]. Amongst them, fabrication of nanowire bridges is particularly attractive because it allows for direct growth of crystallographically aligned nanowires on pre-patterned semiconductor electrodes as well as their electrically and metallurgically robust linkages to the electrodes (as shown in Fig. 1) [3-6]. For FETs, the bridged nanowires are then designated as channels, and source/drain electrodes can be readily patterned along with the pre-patterned electrodes. Consequently, compared to any other methods, this route provides much simpler and more practical ways for preparing aligned nanowires and further fabricating FETs.

Here, we demonstrate GAA-FETs based on horizontally aligned nano-bridges with a high on/off-current ratio, low sub-threshold swing and very low off-currents. In addition to designing low-voltage non-volatile memory cells, the photosensitivity of the 3-dimensional (3D) FETs is employed to demonstrate several analog and digital applications.

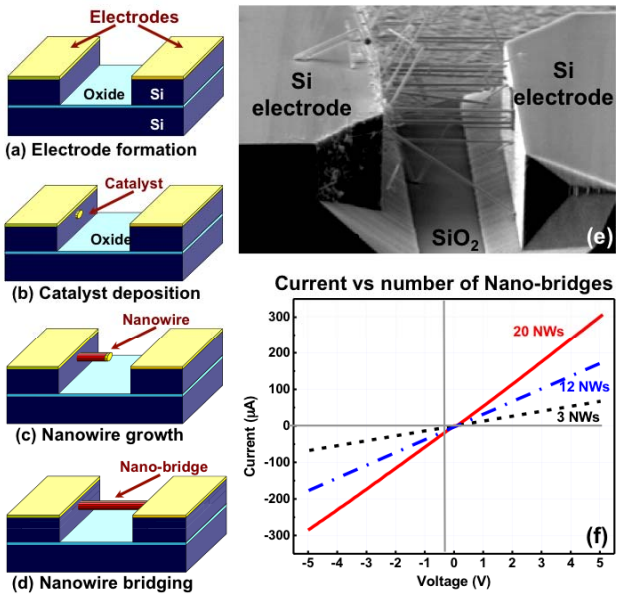


Fig. 1. (a-d) Schematics of nano-bridges synthesis method. (e) SEM micrograph of multiple Si nano-bridges formed between two vertical Si surfaces. (e) Highly linear I-V characteristics of the bridged structures different number of nanowires between electrodes.

II. GAA-FET Fabrication

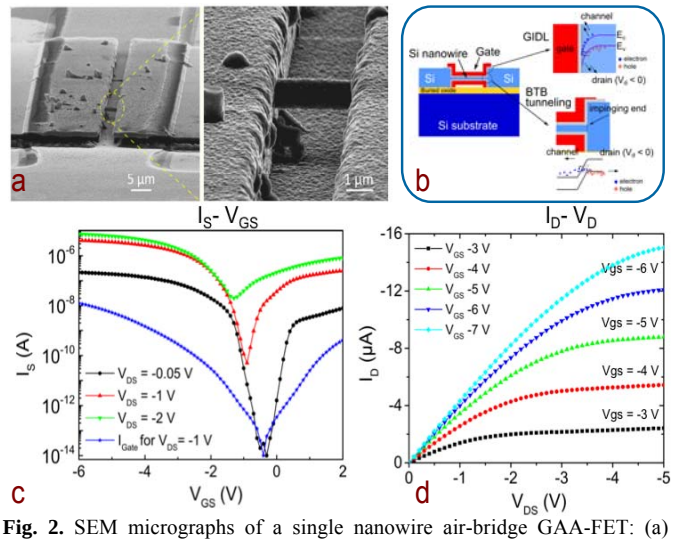


Fig. 2. SEM micrographs of a single nanowire air-bridge GAA-FET: (a) Bird's-eye view and zoomed-in view of the nanowire channel. The nanowire channel is suspended between two electrodes and surrounded by polysilicon gate electrodes. (b) Schematics showing device cross-section and band-diagram. (c) Measured current (I_S) -voltage (V_{GS}) for three drain voltages (V_{DS}). (d) Output characteristics (I_D-V_{DS}), collected from a nanowire channel GAA-FET with a 3 μm long and 120 nm thick channel.

To fabricate an array of integrated FETs, silicon source/drain electrodes, were patterned on p^+-(110) silicon (5×10^{18} cm^{-3}) in SOI, using KOH solutions or reactive ion etching (RIE) after photolithography. On the sidewall surfaces, photolithography was carried out to create small openings,

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

where nanowires are expected to grow. In those openings, the poly-L-lysine was treated and then gold colloids (200 nm, from Sigma Aldrich), as a catalyst to aid nanowire growth, were deposited. Consecutively, the substrates were cleaned with acetone, a piranha solution (H_2SO_4:H_2O_2=2:1), HF acid, and DI water. Silicon nanowires were grown using a cold-wall CVD tool (First Nano®) with SiH_4, B_2H_6, HCl, and H_2 gases. Growth temperature and pressure were respectively 575°C and 13 torr. Diameters and lengths of nanowires were determined by sizes of gold nanoparticles and growth time, respectively. Gold tips and impurities along silicon nanowire surfaces were removed using a gold etchant TFA without damaging the nanowires prior to additional processes. Detailed procedures for creating nanowire bridges are presented in Ref 14c. For creating GAA-FETs, the nanowire bridges, were oxidized at 900°C for 10 min to grow 10 nm thick gate oxide. On the oxidized nanowires, p+ polysilicon was conformally deposited and delineated to the gate electrodes using photolithography and wet etching. For the fabricating flash memory cells, on the other hand, tunneling barrier layers of SiO_2 (2 nm)/Si_3N_4 (3 nm)/SiO_2 (2 nm), a charge trap layer of HfO_2 (5 nm), and a blocking layer of Al_2O_3 (15 nm) were sequentially deposited on nanowire bridges. The detailed deposition scheme and structural analysis of the tunneling barrier are presented elsewhere [6, 7]. Using the lift-off technique, platinum thin film pads were fabricated to create Ohmic contacts on the silicon electrodes.

In our nano-bridge enabled FETs, we synthesized lateral bridging nanowires with both Ti and Au catalysts in between two vertical walls of a trench [6, 8-12]. Fig. 2a illustrates Au-nucleated Si nanowires growing laterally from one (111)-oriented trench face toward the opposing face. To form functioning devices such as FETs or memory, the electrodes must be isolated from each other, from electrodes of other devices, and from the substrate, which might contain other electronic components. For Si nanowire devices, the electrical and thermal isolation can be provided by forming the electrodes from the heavily doped, (110)-oriented, top silicon layer of a silicon-on-insulator (SOI) structure, in which a thin layer of SiO_2 separates the Si device layer from the mechanically supporting underlying Si wafer, as schematically illustrated in Fig. 2a. Removing the entire thickness of the top silicon layer between the electrodes provides the necessary electrical isolation (Fig. 2b). Metal catalyzed nanowires are grown and bridged between two isolated electrodes in the same method as described above.

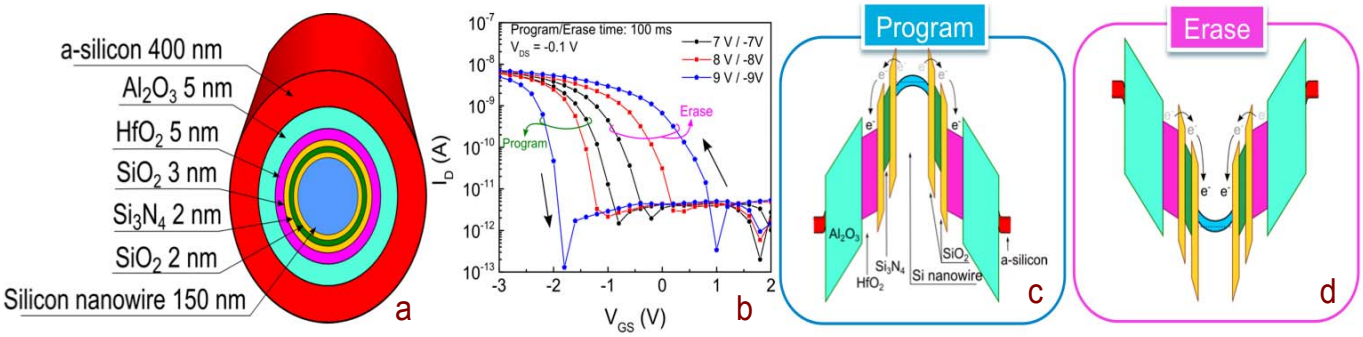


Fig. 3. (a) Schematic diagram of the nanowire channel with tunneling barriers, trap layers and the gate in a VARIOT type nanowire air-bridge memory cell. (b) Semi-logarithmic plot of current ($|I_D|$) versus V_{GS} after program and erase operations. For ±7 V, ±8 V, and ±9 V sweeps at V_{DS}= -0.1 V, memory window range increases from 0.5 V to 1.3 V and to 2.7 V, respectively; arrows represent sweep/hysteresis direction. Energy band diagrams of the memory cell for (c) programming and (d) erasing.

III. Device Characterization

GAA-FET Based Memory and Logic: To demonstrate a proof-of-concept device, we fabricated an array of GAA-FETs with Si nanowire air-bridged channels, grown by VLS technique. The horizontally aligned nanowire air-bridges led to the integration of a GAA-FET array without rigorous processes for aligning and contacting the nanowires. The devices exhibited excellent performance including high on/off-current ratios, low sub-threshold swings and off-current, as shown in Fig. 3. Statistical evaluation on electrical properties of the GAA-FETs confirmed that our fabrication method possesses a promising capability of integrating any number of nanowire FET devices. In addition, the demonstration of the low voltage operational VARIOT type non-volatile memory cells reveals that our device fabrication scheme has high versatility to prepare other functional nanodevices beyond FETs.

Demonstration of GAA Phototransistors for Electro-optical OR Gate Circuit and Frequency Doubler Applications: We also demonstrated photosensitive 3-D silicon nanobridge GAA-FETs integrated in an array using a high-throughput fabrication technique, as shown in Fig. 4. Off-current of the nanobridge GAA-FET increases by as high as three orders of magnitude with wide-spectral illumination. The device also showed phototransistors characteristics in which output current was linearly modulated with the incident illumination. Photosensitive off-current of the GAA-FET allowed for the successful demonstration of digital and analog applications, such as an electro-optical OR gate circuit and an illumination-enabled frequency doubler. When they are associated with more photosensitive compound semiconductor nanowires such InP and GaN, the nanobridge GAA-FETs can be exploited for realizing high performance multifunctional devices including optoelectronic transceivers as well as sensors, as is proposed in this proposal.

In order to address the challenges in continued scaling of CMOS and increase their performances, one proposed solution is to introduce a new channel material in which charge carriers travel at a much higher velocity than those in silicon. This would allow a reduction in voltage without a loss of speed or other performance and this is why attention is turning to III–V compound semiconductors. Such materials are attractive because of their high electron mobility,

exceptional thermal and chemical stability, high electrical endurance, very high reliability and high-quality interface with gate dielectrics. Currently silicon transistors are going through new gate architectures, channel materials and switching mechanisms. The trend in transistor scaling has already led to a change in gate structure from two dimensions (2D) to three (3D), to avoid the issues in miniaturization such as high off-state leakage current and the short-channel effect.

Compared to the conventional planar gate FETs, multiple gate FETs—such as tri-gate, pi-gate, Ω-gate, and surrounding gate FETs—have improved controllability over the channel in such a way that they effectively suppress the off-current. Owing to enhanced gate controllability, multiple gate FETs aid in low power consumption and downscaling of CMOS devices. The idea of gate-all-around (GAA) transistors, in which the gates are wrapped around nanowire channels, provide the best possible electrostatic gate control [6]. At present, planar, fin architectures using III–V materials are being explored as alternative fast channels on silicon (Si). Although Si is widely used as the channel materials in such devices when devices are designed on Si wafers, using other materials such as InP or GaN channels on silicon, however, has not been widely investigated because of difficulties in integrating free-standing III-V nanostructures on silicon.

If we were to deposit III-V layers directly on Si, the layers will not be continuous and will grow as widely separated nano-dots. However, researchers have shown that even with large lattice mismatch with Si, small cross-section allows the growth of III-V nanowires on a Si surface, accommodating this lattice mismatch [11,13]. Transistors enabled by one-dimensional nanowires have attracted attention because they are expected to present a pathway to overcome the roadblocks to Moore's law and can enable gate-all-around transistors. However, lack of reproducibility in their growth, alignment into arrays, and contact formation to enable scalable fabrication of transistors is challenging.

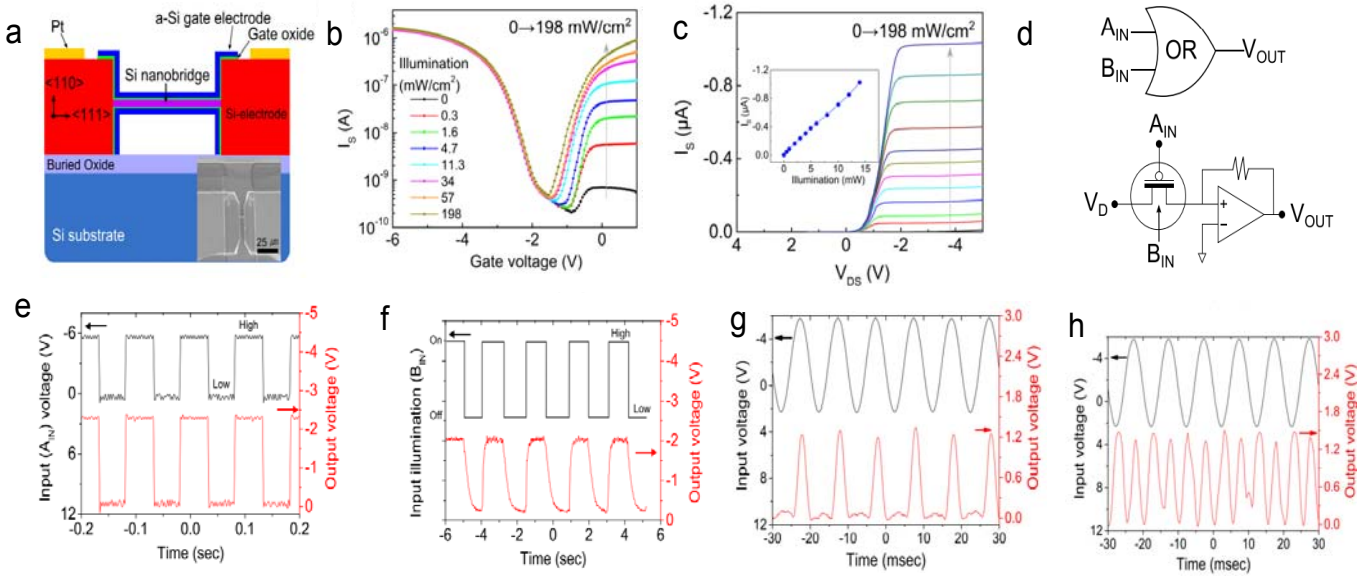


Fig. 4. (a) A 2D schematic diagram of a Si nano-bridge GAA-phototransistor and its SEM micrograph shown in the inset. (b) Measured I_S-V_{GS} curves with illumination of different intensities from 0-198 mW/cm^2. (c) Output characteristics (I_S-V_{DS}) with photo-gating. The inset in (c) shows a linear increase of photocurrent collected for V_{DS} = -2.5 V with increasing illumination intensities. Note that while the off-current in (b) drastically increases with illumination, the on-current increases by 10 %, and subthreshold slopes and threshold voltages remain unaffected by illumination. (d) Symbolic representation of an electro-optical OR gate and its circuit diagram. (e-f) Ouput voltages with an electrical input (A_{IN}) (e) in dark and optical input (B_{IN}) (f) with low A_{IN}. It should be noted that a discernable delay in the fall-time of the output pulses are caused by the latent light from the tungsten-halogen bulb. Illumination-regulating frequency doubling characteristics of the circuit illustrated in Figure e-f. Input and output voltages measured (g) without illumination and (h) with illumination intensity of 190 mW/cm^2. Fast-Fourier-transform (FFT) of output signals of (g) and (h) show that the input signal of 100 Hz is doubled to the output of 200 Hz.

IV. Conclusions

In conclusion, an array of GAA-FETs with Si nano-bridge channels grown via a high-throughput *bottom-up* (VLS) technique has been demonstrated. The horizontally aligned nano-bridges led to the integration of nanowire GAA-FETs without rigorous processes for aligning and contacting the nanowires. The devices exhibited high on/off-current ratios, low sub-threshold swings and off-current. Off-current of the nanobridge GAA-FET increases by three orders of magnitude with wide-spectral illumination. The device also showed phototransistors characteristics in which output current was linearly modulated with the incident illumination. Photosensitive off-current of the GAA-FET allowed for the demonstration of digital and analog applications, such as an electro-optical OR gate circuit and an illumination-enabled frequency doubler. When they are associated with more photosensitive compound semiconductor nanowires such as ZnO and InP, the nanobridge GAA-FETs can be exploited for realizing multifunctional devices including optoelectronic transceivers as well as sensors. Statistical evaluation on electrical properties of the GAA-FETs confirmed that the presented fabrication method possesses a promising capability of integrating nanowire FETs.

Acknowledgment

This work is partially supported by NSF grant # CMMI-0547113 and CMMI-1235592, Korea Basic Science Research Program # 2011-0021826, and a research grant from Kwangwoon University in 2013.

REFERENCES

[1] H. Yan, H. S. Choe, S. W. Nam, Y. J. Hu, S. Das, J. F. Klemic, J. C. Ellenbogen, and C. M. Lieber, "Programmable nanowire circuits for nanoprocessors," *Nature,* vol. 470, pp. 240-244, Feb 10 2011.

[2] A. Heinzig, S. Slesazeck, F. Kreupl, T. Mikolajick, and W. M. Weber, "Reconfigurable Silicon Nanowire Transistors," *Nano Letters,* vol. 12, pp. 119-124, Jan 2012.

[3] M. S. Islam, S. Sharma, T. I. Kamins, and R. S. Williams, "A novel interconnection technique for manufacturing nanowire devices," *Applied Physics a-Materials Science & Processing,* vol. 80, pp. 1133-1140, Mar 2005.

[4] M. S. Islam, S. Sharma, T. I. Kamins, and R. S. Williams, "Ultrahigh-density silicon nanobridges formed between two vertical silicon surfaces," *Nanotechnology,* vol. 15, pp. L5-L8, May 2004.

[5] J. Y. Oh, H. J. Jang, W. J. Cho, and M. S. Islam, "Highly sensitive electrolyte-insulator-semiconductor pH sensors enabled by silicon nanowires with Al2O3/SiO2 sensing membrane," *Sensors and Actuators B-Chemical,* vol. 171, pp. 238-243, Aug-Sep 2012.

[6] J. Y. Oh, J. T. Park, H. J. Jang, W. J. Cho, and M. S. Islam, "3D-Transistor Array Based on Horizontally Suspended Silicon Nano-bridges Grown via a Bottom-Up Technique," *Advanced Materials,* vol. 26, pp. 1929-1934, Mar 2014.

[7] J. Y. Oh and M. S. Islam, "Nanobridge gate-all-around phototransistors for electro-optical OR gate circuit and frequency doubler applications," *Applied Physics Letters,* vol. 104, Jan 13 2014.

[8] M. S. Islam, M. R. Haque, C. M. Oh, Y. Wang, and B. H. Park, "A common-path optical coherence tomography based electrode for structural imaging of nerves and recording of action potentials," *Photonic Therapeutics and Diagnostics Ix,* vol. 8565, 2013.

[9] M. S. Islam, S. Sharma, T. I. Kamins, and R. S. Williams, "Nano-bridging: A massively parallel self-assembly technique for interconnecting nanowire sensors," *Nanosensing: Materials and Devices,* vol. 5593, pp. 101-111, 2004.

[10] J. Y. Oh, J. T. Park, and M. S. Islam, "Fabrication of Ambipolar Gate-All-Around Field-Effect Transistors using Silicon Nanobridge Arrays," *Nanoepitaxy: Materials and Devices V,* vol. 8820, 2013.

[11] S. S. Yi, G. Girolami, J. Amano, M. S. Islam, S. Sharma, T. I. Kamins, and I. Kimukin, "InP nanobridges epitaxially formed between two vertical Si surfaces by metal-catalyzed chemical vapor deposition," *Applied Physics Letters,* vol. 89, Sep 25 2006.

[12] S. Sharma, T. I. Kamins, M. S. Islam, R. S. Williams, and A. F. Marshall, "Structural characteristics and connection mechanism of gold-catalyzed bridging silicon nanowires," *Journal of Crystal Growth,* vol. 280, pp. 562-568, Jul 1 2005.

[13] V. J. Logeeswaran, A. Sarkar, M. S. Islam, N. P. Kobayashi, J. Straznicky, X. Li, W. Wu, S. Mathai, M. R. T. Tan, S. Y. Wang, and R. S. Williams, "A 14-ps full width at half maximum high-speed photoconductor fabricated with intersecting InP nanowires on an amorphous surface," *Applied Physics a-Materials Science & Processing,* vol. 91, pp. 1-5, Apr 2008.

Study on Gate Dielectric for Ballistic III-V MOSFET: Influence of band parameter keeping same EOT

Muhammad Shaffatul Islam ,Md. Nur Kutubul Alam,Md. Raifqul Islam
Department of Electrical and Electronic Engineering
Khulna University of Engineering & Technology
Khulna-9203, Bangladesh
Email: mashru.islam@yahoo.com, alamjhilam@gmail.com, islambit@yahoo.com

***Abstract*—Ballistic performance of nanoscale $In_{0.3}Ga_{0.7}Sb$ double gate n-MOSFET is studied for different gate oxides considering same Equivalent oxide thickness (EOT). Non equilibrium greens function method is utilized under the environment of well-known SILVACO's ATLAS device simulation package to carry out the simulation. Wave function penetration to the oxide is taken into account in the simulation. The results obtained from the simulation indicate that EOT is not solely control the device performance. Rather, a strong correlation is found between the I-V characteristics and the conduction band offset (ΔE_C) of the channel and dielectric materials. It is also found that the threshold voltage decreases and the subthreshold swing increases with the higher value of ΔE_C. It reveals that for nanoscale MOSFET design, another tunable variable is found to ensure best performance.**

***Keywords*—InGaSb MOS; Double gate; NEGF; Gate oxide; high-k**

I. INTRODUCTION

Owing to the matured process technology Si has been using as an attractive material for commercial device fabrication [1]. For the last 30 years [2], CMOS transistor has been drastically scaled down to meet up the demand of high speed computation. In recent years, further scaling down has become a challenging issue due to the physical limits [3] of Si-based MOS device technology. When the dimension of Si transistors are scaled downed to a certain limit, it causes short channel effects (SCEs), enhanced parasitic resistance, scattering at oxide-silicon interface and thereby degrade the device performance [4]. To overcome these limitations, III-V materials are being studied to replace Si from the channel and are expected to give better performance due to their superior transport properties [5].

According to ITRS roadmap III-V material will replace Si from the channel by 2018[6,7]. Among the III-V materials, InGaSb is a promising candidate for the channel of future MOSFET as it shows highest electron and hole mobilities [8]. Apart from the channel material, gate oxide of MOSFET is also a major issue. When the channel length of a device is scaled down, oxide thickness is also to be scaled to maintain the required level of C_{ox}, I_{ON} and to control the threshold voltage roll off, that is, to keep the control over the channel. Thickness of SiO_2, which is the preferred gate insulator for Si MOSFETs, is reduced over the years from 300nm for 10μm technology to 1.2nm for 65nm technology. When SiO_2 thickness is scaled down, key dielectric properties degrades and device suffers from high leakage current, dielectric breakdown, reduced channel mobility and so on [9,10]. These problems can be solved by replacing SiO_2 with a material having higher dielectric constant. High-k material can be grown physically thicker than SiO_2which has the same electrostatic effects that a thin SiO_2 of thickness equal to "Equivalent oxide thickness, EOT" has. The thick high-k dielectric offers required control over the channel with significant reduction in leakage current [10].

It is well established that to achieve the required control over the channel at the small length targeted by ITRS, multi gate device structure is necessary. Keeping all these issues in mind, in this paper, we studied the performance of $In_{0.3}Ga_{0.7}Sb$ double gate MOSFET for n-type operation considering two different dielectric materials Al_2O_3 and HfO_2 with same EOT of 1nm. The effects of conduction band offsets and gate oxides over current-voltage characteristics as well as related terms like subthreshold swing are investigated. Since drain current is normally calculated by quantum mechanical approaches in ballistic range, we used non equilibrium greens function (NEGF) to carry out the simulation.

II. QUANTUM BALLISTIC SIMULATION

A. Device structure

The proposed device structure is shown in figure-1 and the numerical parameters used in our study are summarized in Table-1. SILVACO's Atlas device simulation package is used to determine the ballistic performance of the device. We used NEGF mode space in effective mass approach.

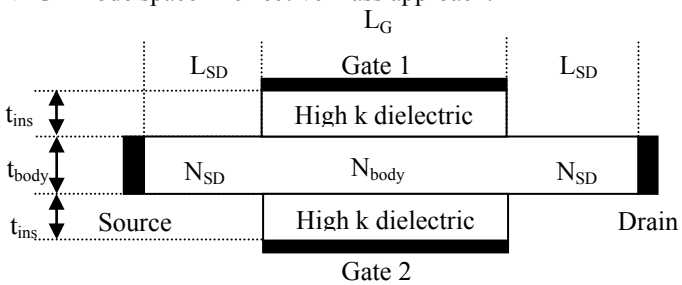


Fig. 1. Device structure of InGaSb double gate nMOSFET. Some the simulation parametersares are shown in TABLE I.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

TABLE I. PARAMETERS USED IN THE SIMULATION OF THE DEVICE STRUCTURE SHOWN IN FIG. 1.

Parameter	*Value*	*Description*
V_{dd}	0.05V	
ε_{ins}	9.3 (for Al_2O_3) 22 (for HfO_2)	Al_2O_3 and HfO_2 as gate oxide
t_{body}	3nm	$In_{0.3}Ga_{0.7}Sb$ channel
t_{ins}	2.385 nm (for Al_2O_3) 5.64nm (for HfO_2)	EOT=1nm for both oxide
N_{SD}	N type, $10^{20}/cm^{-3}$	Within the region L_{SD}
N_{body}	0	Within the region L_G
L_{SD}	5nm	
L_G	15nm	
Φ_{metal}	3.52eV	Gate metal workfunction

B. NEGF theory

The NEGF method provides a rigorous formalism for quantum transport in nanoscale devices [11]. This quantum modeling relies on self-consistent solution of Greens function and Poisson's equations to obtain the energy level, charge and current distribution [12]. The Greens function in fact is a modified schrodinger equation in which the quantum mechanical effect is introduced considering that electrons travel like wave or wave packet from source to drain through the channel.

NEGF are solved using two approaches namely real space and mode space. Simulation in real space method is computationally very much costly. In case of the device where quantum confinement is strong (double gate MOS or XOI FET) the mode space approach can be used. The Hamiltonian in this situation is separated in longitudinal and transverse directions; as a result, it greatly reduces the computational burden. Moreover, in a device structure where the cross-section does not change, the sub-bands are not quantum-mechanically coupled to each other; the transport equations become essentially 1D for each sub-band that leads to use the uncoupled mode space approach [13]. The transformation of a real space effective mass Hamiltonian H_0 to a mode space is done by taking a matrix element between m^{th} and n^{th} wave functions of kth and l^{th} slices

$$H^{MS}_{mnkl} = < \psi^{k}_{m(y)} | H_0 | \psi^{l}_{n(y)} >$$

The transport equation in mode space is

$$\left(E - H^{MS}_{mnkl} - \Sigma^R_S - \Sigma^R_D\right) G^R = I$$

Here, E is the energy of carrier, Σ^R_S and Σ^R_D are retarded self energies carrying information about source and drain respectively, $G^R(E)$is the retarded greens function. After solving the above equations for each energy level, current is calculated by [13]

$$J_x(x_i, y_j) = -\frac{2e}{hL_z\Delta_x}\sum_{k_z\sigma}.\sum_{mn}\int Re(t_{i+1ijj}G^{<}_{mnii+1}\,\psi^{i}_{m}(y_j)\,\psi^{*i+1}_{n}(y_j)\frac{dE}{2\pi}$$

III. RESULT AND DISCUSSION

In our simulation, source and drain doping concentration is taken to be $10^{20}/cm^3$, and this doping is done throughout the region marked by L_{SD} in figure-1. Von Neumann boundary conditions for potential are used because the ballistic carrier do not create voltage drop near source and drain contact. Moreover, the potential in source and drain are allowed to float, because the number of carrier reflected backward and hence the total carrier concentration in the contact can change with bias. The body thickness of the device is 3nm. Wave function penetration in the oxide region is also taken into account.

Figure-2 shows a comparison between the I_D-V_G characteristics plotted at V_D=0.05V for the double gate MOS device having two gate dielectrics. It is found that the drain current is higher for Al_2O_3 than that of HfO_2, although they have same EOT. Generally it is expected that if two oxides have the same EOT, they will provide same oxide capacitance, thereby same electrostatic control over the channel and hence identical device performance. But in case of double gate MOSFET, the deviation is found because of conduction band discontinuity (ΔE_C) between the oxide and the channel.

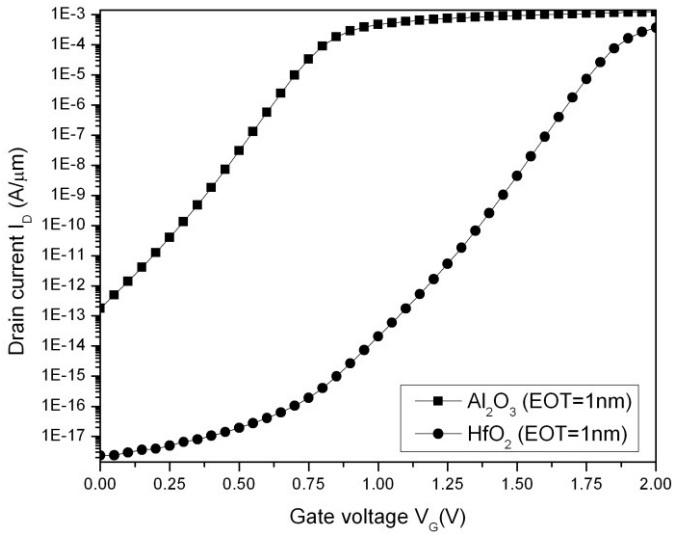


Fig. 2. I_D-V_G characteristics of double gate MOSFET for different gate dielectric material. The structure of the device is shown in fig-1

To understand the effect of ΔE_C, the conduction band diagram across and along the channel at V_D=0.05V and V_G=2V are estimated and plotted in figures 3 (a) and (b), respectively. It is clear that at very low gate bias, with keeping the gate metal work function (Φ_{metal}) fixed, lower ΔE_C implies the higher difference between the channel conduction band minima and Fermi level at the source contact, which is set at "zero". The opposite is found for high gate bias as shown in fig. 3(a). These band positions relative to source contact fermi level implies lowest number of energy states are available for conduction if ΔE_C gets lower. Since lower number conducting states leads to lower drain current, lower conduction band offset between oxide and channel leads to lower drain current,

provided that Φ_{metal} is kept fixed. For $In_{0.3}Ga_{0.7}Sb$ channel, $\Delta E_C(Al_2O_3)$ = 2.88eV and $\Delta E_C(HfO_2)$ = 1.76eV [14]. That is why HfO_2 is giving lower drain current despite of same EOT.

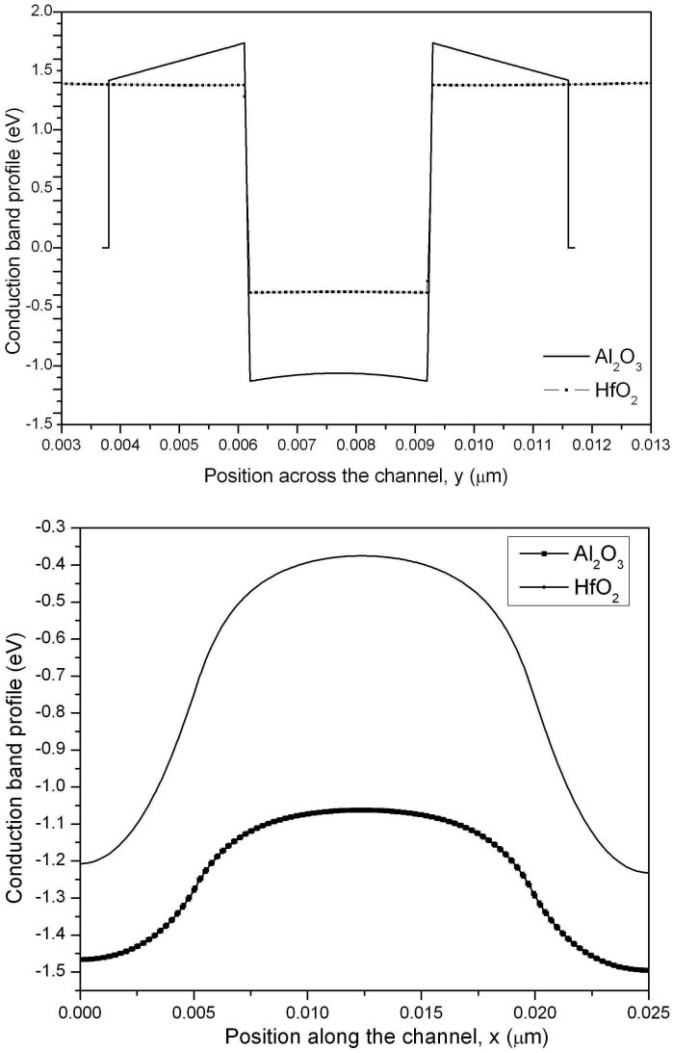


Fig. 3. Conduction band profile of the double gate MOSFET (a) across the channel (b) along the channel, at V_D=0.05V and V_G=2V.

It is also found in figure 3(b) that the conduction band profile does not start from the same position, not from zero level either for two different gate oxides. It is because in ballistic transport source and drain contacts are allowed to float to take into account any reflection of carrier during transport.

Since the threshold voltage can be tuned by tuning the gate metal work function (Φ_{metal}), the I-V curve for HfO_2 can be braught to the left by reducing Φ_{metal}. If Φ_{metal} for HfO_2 based device is choosen to be 1.12eV less than that used for Al_2O_3 device, the I-V curves will be at closest position, because $\Delta E_C(Al_2O_3)$-$\Delta E_C(HfO_2)$=1.12. But to compare the impact of ΔE_C only, all other device parameters need to be same.

Figure-4 shows the change in subthreshold swing (SS) as a function of ΔE_C between oxide and the channel. It is found that the subthreshold swing decreases with decreasing ΔE_C. Lower SS implies the sharp change of drain current with gate bias and the better electrostatic control of gate over the channel. Better control of the gate also indicates the lower influence of the drain voltage on I-V characteristic and hence supressed SCE, like drain induced barrier lowering. Theoriticaly the lowest possible SS at room temperature is 60mV/dec which means the best electrostaic control of MOSFET. In practice, SS bellow 90mV/dec is acceptable for logic application.

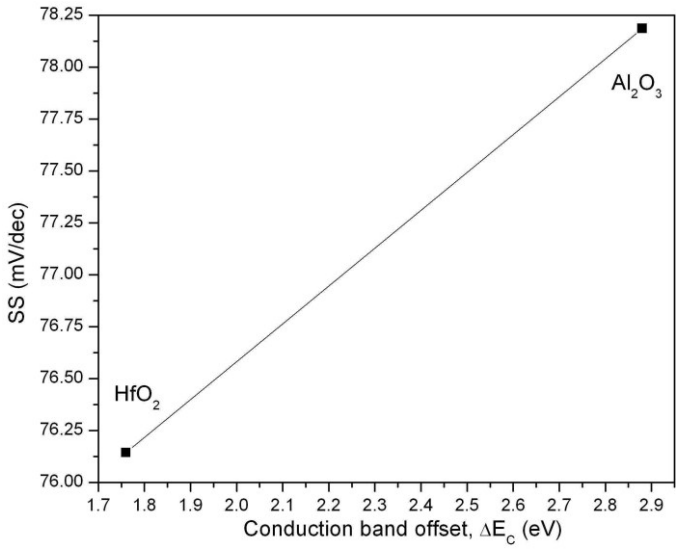


Fig. 4. Variation of subthreshold swing(SS) as function of gate oxidedielectric constant

IV. CONCLUSION

In this paper, we analyzed the influence of ΔE_C between the gate oxide and the channel on the performance of $In_{0.3}Ga_{0.7}Sb$ nanoscale double gate MOSFET. We took 15nm channel with Al_2O_3 and HfO_2 gate oxides having 1nm of equivalent oxide thickness and simulated the performance of the device by NEGF method. The source and drain regions are taken to be heavily doped($10^{20}/cm^3$) and the channel is left undoped. Effective gate metal work function is taken to be 3.52eV and the wave function penetration to the oxide is also taken into account. Source drain contacts are allowed to float to treat the carrier reflection accurately. We found that the drain current increases with increase in ΔE_C if the work function of the gate remains unchanged. Also the threshold voltage (V_T) decreases and SS increases with increasing ΔE_C. SS of the device under study goes bellow 90mV/dec for both case and it is a rule of thumb that SS bellow 90mv/dec is acceptable for logic application. Although we get lower drain current for HfO_2 because of its lower ΔE_C with InGaSb, it can be controlled by tuning the gate metal work function. That is why the overall performance of HfO_2 is better since it gives lower SS.

REFERENCES

[1] Sebastien Nuttinck, "*Ultrathin-Body SOI Devices as a CMOS Technology Downscaling Option: RF Perspective*" IEEE Transactions On Electron Devices, Vol.53, No.5, May 2006

[2] Yasuyuki Nakajima, Seiji Horiguchi, Masanari Shoji and Yasuhisa Omura "*Physical origin and characteristics of gate capacitance in siliconmetal-oxide-semiconductor field-effect transistors*" Journal of Applie Physics, Vol.83, Number.9, 1 May, 1998.

[3] Yasuyuki Nakajima, Seiji Horiguchi, Masanari Shoji and Yasuhisa Omura "*Physical origin and characteristics of gate capacitance in siliconmetal-oxide-semiconductor field-effect transistors*" Journal of Applie Physics, Vol.83, Number.9, 1 May, 1998

[4] Muhammad Shaffatul Islam, Md. Nur Kutubul Alam, Md. Rafiqul Islam "Self-Consistent Quasi Static CV Characterization of $In_xGa_{1-x}Sb$ Buried Channel n-MOSFET", 5th International Nanoelectronics Conference (INEC), 2013, PP.94-96.

[5] Farzin Assad, Zhibin Ren, Dragcia Vasileska, Supriyo Datta, Mark Lundstrom "*On the Performance Limits for Si MOSFETs: A Theoretical Study*" IEEE Transactions On Electron Devices ,vol. 47,No 1, pp.232-240 January 2000.

[6] Annesh Nainani, Ze Yuan, Tejas Krishnmohon, Brian R Bennett, J. Brad Boss, Mathew Reason, Mario G Ancona, Yoshio Nishi and Krishna C. Saraswat "*$In_xGa_{1-x}Sb$ channel p-metal–oxide semiconductor filed effect transistors: Effect of strain and heterostructure design*" Journal of Applied Physics 110,014503 (2011)

[7] The international technology roadmap for semiconductiors, 2012

[availble] online:http://www.itrs.net/Links/2012ITRS/Home2012.htm

[8] Aneesh Nainani Toshifumi Irisawa, Ze Yuan, Yun Sun, Tejas Krishnamohan, Matthew Reason, B rian R. Bennett, J. Brad Boos, Mario G. Ancona, Yoshio Nishi, Krishna C. Saraswat "*Development of high-k dielectric for antimonides and a sub 350°C III–V pMOSFET outperforming Germanium",*International Electron Devices Meeting (IEDM), 6-8 Dec.2010, Page(s): 6.4.1-6.4.4

[9] G. Ribes, J. Mitard, M. Denais, S. Bruyere, F. Monsieur, C. Parthasarathy, E. Vincent, and G. Ghibaudo, "*Review on High- k Dielectrics Reliability Issues*", IEEETransactions on Device and Materials Reliability, vol. 5, no. 1, march 2005

[10] E. P. Gusev, V. Narayan, M. M. Frank, "*Advanced high-k dielectric stacks with polySi and metal gates: Recent progress and current challenges*", IBM Journal of Research and Development, Vol.50, Issue.4.5, Page(s).387-410, July 2006.

[11] S. Datta. Electronic Transp ort in Mesoscopic Systems. Cambridge Univ.Press

[12] R. Venugopal,Z. Ren, S. Datta, and M. S. Lundstrom "*Simulating quantum transport in nanoscale transistors: Real versus mode-space approaches*", Journal of Applied physics, Vol.92, No.7, October 2002

[13] Mathieu Luisier,Andreas Schenk, and Wolfgang Fichtner "*Quantum transport in two- and three-dimensional nanoscale transistors: Coupled mode effects in the nonequilibrium Green's function formalism*", Journal of Applied physics, June 2006

[14] J. Robertson and B. Falabretti, "*Band offsets of high K gate oxides on III-V semiconductors*", J. Appl. Phys, 100, 014111 (2006)

A High Accuracy Continuous Wavelet Function Approximation for Implantable Device Applications

Qingyun Ma and Mohammad Rafiqul Haider*
Department of Electrical and Computer Engineering
The University of Alabama at Birmingham
Birmingham, Alabama 35294-4551
*mrhaider@uab.edu

***Abstract*—Wavelets as one of the most important signal processing tools are widely used in biosignal research. The continuous wavelet transform based on an analog circuit implementation offers power-efficient real-time signal processing. This paper presents a complexity reduction method of an analog wavelet filter design. The Chebyshev polynomial with higher accuracy approximation is utilized in the wavelet design. In order to simplify the circuit implementation, the transfer function is transformed into the multiplication of biquads. For the proof-of-concept, a Mexican hat wavelet filter circuit is implemented as the multiplication of biquads. The measurement results show that the proposed design process offers higher accuracy approximation and less complex method for the implementation of analog wavelet transform circuit.**

***Index Terms*—Chebyshev approximation, continuous wavelet function, implantable system, Taylor's approximation.**

I. Introduction

Various types of biosignals originating from the human body are being extensively used for emerging clinical applications including but not limited to mind-driven rehabilitation, metabolic monitoring, seizure detection, etc[1]. Biosignals such as Electrocardiograph (ECG), Electroencephalograph (EEG) are typical examples of biosignals characterized by transients and non-stationarity such as spiking and small oscillation [2].

In order to process the vital physiologic phenomena in the biosignals, the need of a biopotential recording system is indispensable. The conventional biopotential recording system includes low-noise amplifier, filter, analog-to-digital converter and signal processing circuitry [3]. The signal processing unit as one of the most important parts in the biopotential recording system offers various types of signal analyses such as Fourier Transform, Hilbert Transform, and Continuous Wavelet Transform (CWT).

The CWT is considered as one of the most useful technique in signal processing that offers the localization in both time and frequency for transients, aperiodic and non-stationary signals. The CWT utilizes a wavelet function that acts as a low pass filter (Gaussian wavelet) or a bandpass filter (Mexican hat wavelet and Morlet wavelet). When a signal is applied to such a filter, the wavelet function facilitates the analysis of the frequency contents of the signal at any particular instance of time. For wavelets with different mother wavelet functions and dilation parameters, CWT provides time and frequency information of the input signal, which allows for noise reduction [4], spike detection [5], and seizure detection [6].

The conventional CWT is based on software which requires high data rate analog-to-digital (ADC) converter to minimize the loss of information in the biosignal. The high performance ADC usually has high power consumption. However, for applications requiring real-time processing, such as implanted or mobile biomedical devices, low-power consumption is a priority. Hence, the signal processing block should be implemented in hardware instead of software. Even in hardware-based implementation, analog circuits can offer reduced power consumption compared to equivalent digital circuits [7]. CWT based on an analog filter implementation offers a power-efficient real-time signal processing such as feature extraction and skip detection [8].

Various types of low-power analog CWTs are reported in the recent years - LC ladder, Gm-C filter [8], and log-domain filter structures based CWT [7]. The padé approximation and log-domain based CWT design offers less order of the approximation and low-power consumption [7]. The 7th order Taylor approximation and LC ladder with active inductor greatly decreases the power consumption and the circuit complexity.

This paper presents a method on the implementation of the CWT, with the motivation of reducing the complexity of the circuit implementation. A 4th order Chebyshev polynomial is used to increase the accuracy of the Fourier transform of the Mexican hat wavelet function. The transfer function of the wavelet is reconstructed into the multiplication of two biquad filters to further simplify the circuit implementation. For the proof of concept, a circuit implementation of Mexican hat wavelet based on the proposed method is tested. The results show that the proposed design process offers higher accuracy approximation and less complex method in analog wavelet transform circuit implementation.

In section II, a brief background of wavelet transformation and the method of Chebyshev approximation with different interval is explained. The analog circuit implementation of the wavelet function is shown in section III. Section IV shows the simulation and measurement results. Finally the conclusion is drawn in section V.

II. Continuous Wavelet Transform and Chebyshev Polynomial

A wavelet transform is considered as one of the most widely used mathematical function in signal processing. Wavelet analysis overcomes the inability of Fourier analysis to capture the transients of the time location specific features within a signal. Comparing with the method of Short Time Fourier Transform, wavelets can simultaneously provide the temporal and spectral information of a signal [9]. Wavelets have dilation

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

and location parameters, which result in functions that are localized in both time and frequency. The dilation parameter facilitates the wavelet function with a variable window width. In the low-frequency zone, the long time windows are used to get a finer low-frequency resolution whereas in the high frequency zone, the short time windows are used to get finer time information.

Wavelet transformation as a time-frequency analysis method can be performed in continuous time as well as in discrete time domain. The CWT can be defined as,

$$T(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\psi^* \left(\frac{t-b}{a}\right) dt \tag{1}$$

where $\psi^*(t)$ is the complex conjugate of the mother wavelet function $\psi(t)$, a is the dilation parameter, and b is the location parameter of the wavelet.

The Gaussian function based mother wavelet (Gaussian wavelet and Mexican hat) are the popular choices for biosignal analysis. Therefore, in this paper we demonstrate a Mexican hat wavelet design. The Mexican hat wavelet is the second derivative of a Gaussian function, and the ideal impulse response is as shown in Fig. 1. The equation of Mexican hat in time domain can be expressed as,

$$\psi_t = \frac{2}{\sqrt{3\sigma}\pi^{\frac{1}{4}}} \left(1 - \frac{t^2}{\sigma^2}\right) e^{\frac{-t^2}{2\sigma^2}}. \tag{2}$$

where σ is the dilation factor. The corresponding Fourier transform of the Mexican hat wavelet is ,

$$H(j\omega) = \sqrt{\frac{8}{3}}\pi^{\frac{1}{4}}\sigma^{\frac{5}{2}}\omega^2 e^{\frac{-\sigma^2\omega^2}{2}}. \tag{3}$$

After substituting s into $j\omega$ in (3) and adding a delay $e^{-\tau s}$ into (3), the expression becomes,

$$H(s) = \frac{-\sqrt{\frac{8}{3}}\pi^{\frac{1}{4}}\sigma^{\frac{5}{2}}s^2}{e^{\tau s - \frac{-\sigma^2 s^2}{2}}}. \tag{4}$$

Since (4) has exponential term in the denominator, the transfer function cannot be realized into a standard form. Hence the exponential term needs to be expanded. The most common exponential expansion is Taylor using Maclaurin's series and Padé approximation. Maclaurin's series is easy to implement. However it is very accurate when σ is near zero, but becomes increasingly inaccurate when σ increases. The padé approximation offers less order of polynomial compared to Taylor, however it has higher complexity and unsuitability. To achieve the high accuracy approximation with less order, the Chebyshev polynomial expansion is used.

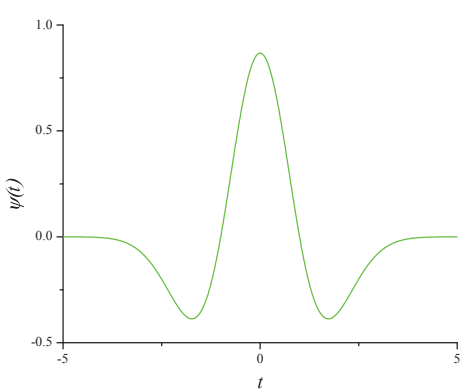


Fig. 1. The plot of ideal Mexican hat mother wavelet.

Chebyshev polynomials, named after Pafnuty Chebyshev, are a sequence of orthogonal polynomials that can be written as monic polynomials of nth degree whose deviation from zero is as small as possible. Using series based on orthogonal Chebyshev polynomials on approximating large systems of equations gives better results in terms of accuracy than the typical methods such as the mean square approximation or the Taylor approximation [10].

The recurrence relation of Chebyshev polynomials with arbitrary interval a can be found as,

$$\left.\begin{aligned} T_0(x) &= 1 \\ T_1(x) &= \frac{x}{a} \\ T_2(x) &= \frac{2x}{a}T_1(x) - T_0(x) \\ &\vdots \\ T_n(x) &= \frac{2x}{a}T_{n-1}(x) - T_{n-2}(x) \end{aligned}\right\} \quad for\ x \in [-a, a]. \tag{5}$$

The weight function $C_k(x)$ in the interval of $[-a, a]$ is,

$$C_k(x) = \int_{-a}^{a} W(x) \cdot T_n^2(x)\, dx$$
$$where \tag{6}$$
$$W(x) = \frac{1}{\sqrt{a^2 - x^2}}.$$

The coefficients of the Chebyshev polynomials corresponding to the function $f(x)$ can be found as,

$$c_{o_n} = \frac{1}{C_n(x)} \int_{-a}^{a} W(x) \cdot T_n(x) \cdot f(x)\, dx. \tag{7}$$

III. The Analog Continuous Wavelet Transform Implementation

The design procedure of the Continuous Wavelet Transform can be found in Fig. 2. The design steps involve selecting the mother wavelet, taking the Fourier transfer of the desired wavelet, adding delay, applying Chebyshev approximation, calculating the zeros and poles of the transfer function, reconstructing the transfer function, and finally implementing the resultant transfer function by using biquads.

To process the EEG signal, the low frequency Mexican hat wavelet is needed. With a dilation factor $\sigma = 0.2$ and delay $\tau = 0.8$, the center frequency of the wavelet is found to be 1 Hz. The filter order used represents a trade-off among the accurate approximation of the desired function, the circuit complexity and the power consumption. To optimize the performance and the circuit complexity, the 4th order approximation is used.

To compare the performance of the Chebyshev approximation, same order of the Taylor expansion is used to approximate the wavelet function. Fig. 3 shows the percentage error of the 4th order Chebyshev and the 4th order Taylor approximations of the denominator in (4) with $\sigma = 0.2$ and $\tau = 0.8$ in the interval of -3 to 3. The maximal error of Taylor approximation reaches to 40%, however the maximal error of Chebyshev approximation is only 6%.

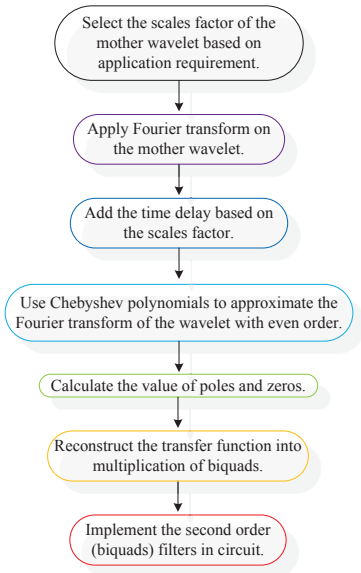


Fig. 2. The flow chart of the design procedure of the Continuous Wavelet Transform

The 4th order Chebyshev approximation of (4) with $\sigma = 0.2$ and $\tau = 0.8$ is as shown below,

$$H(s) = \frac{-0.0389s^2}{\begin{array}{l}0.011999853794280714727977842784763s^4 \\ +0.082948231213368508144515537331379s^3 \\ +0.296199087631179461112108776256 31s^2 \\ +0.769154217916028601932191636573 56s \\ +1.001894393842919726497764365547\end{array}}. \tag{8}$$

The approximated transfer function can be implemented by LC ladder, Gm-C, or log-domain structures. However, constructing a high order filter based on (8) is not obvious. Due to the complexity of higher order transfer function (such as the function in (8)) implementation in analog domain, a new transfer function has been reconstructed by using the values of poles and zeros of the transfer function. The values of four poles and two zeros can be calculated from (8). They are $s_1 = -2.91 - j0.71$, $s_2 = -2.91 + j0.71$, $s_3 = -0.551 - j3.011$, $s_4 = -0.551 + j3.011$, and $z_1 = z_2 = 0$. The reconstrudted transfer function can be rewritten as,

$$H(s) = \frac{s^2}{s^2 + 36.568138s + 354.208258} \cdot \frac{-0.0389}{s^2 + 6.92407s + 369.90179795}. \tag{9}$$

The new transfer function (9) indicates that the Mexican hat wavelet can be easily implemented by cascading a second-order high pass filter and a second-order low pass filter. The corresponding block diagram to the reconstructed transfer function is shown in Fig. 4. The first stage is a second-order high pass filter with two zeros at the origin. The second stage is a second-order low pass filter that generates the third and fourth poles for the system. The last stage is a normalized gain control stage. The first and the second stage can be implemented either by passive structures, such as LC ladder and RC filter, or by active structures, such as signle-amplifier biquad filter, butterworth biquad, and Sallen-Key circuits *et al*; and other MOSFET based structures, such as Gm-C filter and log-domain filter.

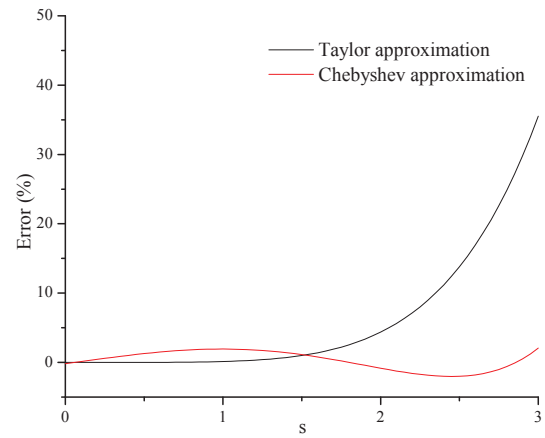


Fig. 3. The maximal error of Taylor approximation reaches to 40%, however the maximal error of Chebyshev approximation is only 6%.

IV. Simulation and Experimental Results

An OTA based 4th Mexican hat wavelet filter has been designed and simulated by Cadence. Due to very limited microchip fabrication facility, a proof-of-concept of the proposed method is implementation by using AD8044. The proposed method can be easily implemented by using three OPAMPs based units - a high pass filter, a low pass filter, and an inverting amplifier.

Fig. 5 is the impulse response of the Mexican hat wavelet measured by oscillator scope. The upper plot is the input impulse generated by function generator with an amplitude of 1 V, pulse width of 100 ms and delay of 500 ms . The lower plot is the output of the Mexican hat wavelet corresponding to the impulse signal with amplitude of 1.2 V.

Fig. 6 shows the comparison of measured result of the impulse response with the simulation result. Although the passive components have value variation and the mismatch, the measured results show the great similarity to the simulation result. The experimental result shows that the realization of the wavelet transform by using biquads structure has a great tolerance on components value variation.

V. Conclusion

This paper presented a method on the implementation of the continuous time wavelet transformation by using the biquads,

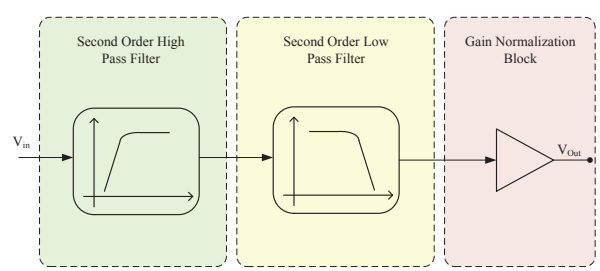


Fig. 4. The block diagram related to (9).

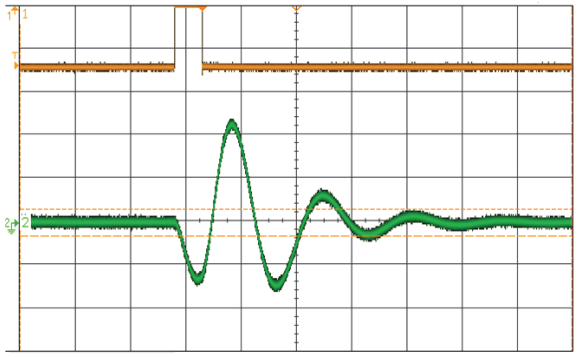


Fig. 5. The upper plot is the input impulse with amplitude of 1 V, 100 ms pulse width and 500 ms delay. The lower plot is the impulse response output of the Mexican hat wavelet.

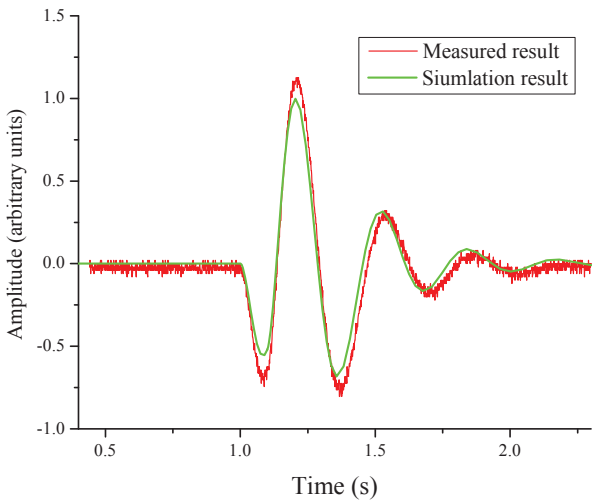


Fig. 6. The comparison of measured result of the impulse response versus the simulation results.

with the motivation of simplifying the circuit implementation. The Chebyshev approximation of Mexican hat wavelet shows better accuracy compared to Taylor expansion in 4th order. The maximal error of Taylor approximation reaches to 40%whereas the maximal error of Chebyshev approximation of the denominator is only 6%. A new transfer function has been reconstructed into the form of multiplication of biquads to simplify the circuit implementation of higher order transfer function. As a proof of concept, a circuit implementation of Mexican hat wavelet has been implemented. The experimental result and the simulation result show a close match. The proposed the method offers approximation with higher accuracy and easier way to implement the analog continuous wavelet transform circuit.

References

[1] L. R. Hochberg, M. D. Serruya, G. M. Friehs, J. A. Mukand, M. Saleh, A. H. Caplan, A. Branner, D. Chen, R. D. Penn, and J. P. Donoghue, "Neuronal ensemble control of prosthetic devices by a human with tetraplegia," *Nature*, vol. 442, pp. 164–171, 2006.

[2] M. Unser and A. Aldroubi, "A review of wavelets in biomedical applications," *Proceedings of the IEEE*, vol. 84, no. 4, pp. 626–638, 1996.

[3] Y. Perelman and R. Ginosar, "An integrated system for multichannel neuronal recording with spike/lfp separation, integrated a/d conversion and threshold detection," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 1, pp. 130 –137, January 2007.

[4] J. Lewis and C. Burrus, "Approximate continuous wavelet transform with an application to noise reduction," in *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, vol. 3, 1998, pp. 1533–1536 vol.3.

[5] M. Faezipour, T. Tiwari, A. Saeed, M. Nourani, and L. Tamil, "Wavelet-based denoising and beat detection of ecg signal," in *Life Science Systems and Applications Workshop, 2009. LiSSA 2009. IEEE/NIH*, 2009, pp. 100–103.

[6] H. Vavadi, A. Ayatollahi, and A. Mirzaei, "A wavelet-approximate entropy method for epileptic activity detection from eeg and its sub-bands," *Journal of Biomedical Science and Engineering*, vol. 3, no. 1, pp. 1182–1189, October 2010.

[7] S. A. P. Haddad and W. A. Serdijn, *Ultra Low-Power Biomedical Signal Processing*. Springer, 2009.

[8] A. Casson and E. Rodriguez-Villegas, "A 60 pw g_mC continuous wavelet transform circuit for portable eeg systems," *Solid-State Circuits, IEEE Journal of*, vol. 46, no. 6, pp. 1406–1415, 2011.

[9] M. Sifuzzaman, M. Islam1, and M. Ali, "Application of wavelet transform and its advantages compared to fourier transform," *Journal of Physical Sciences*, vol. 13, pp. 121–134, October 2009.

[10] C. W. Clenshaw, "A comparison of "best" polynomial approximations with truncated chebyshev series expansions," *Journal of the Society for Industrial and Applied Mathematics: Series B, Numerical Analysis*, vol. 1, pp. pp. 26–37, 1964. [Online]. Available: http://www.jstor.org/stable/2949762

Dermatological Disease Detection using Image Processing and Artificial Neural Network

Rahat Yasir,* Md. Ashiqur Rahman, and Nova Ahmed

Electronic and Computer Engineering Department,
North South University,
Dhaka, Bangladesh
*rahat.anindo@live.com

***Abstract*— Skin diseases are among the most common health problems worldwide. In this article we proposed a method that uses computer vision based techniques to detect various kinds of dermatological skin diseases. We have used different types of image processing algorithms for feature extraction and feed forward artificial neural network for training and testing purpose. The system works on two phases- first pre-process the colour skin images to extract significant features and later identifies the diseases. The system successfully detects 9 different types of dermatological skin diseases with an accuracy rate of 90%.**

***Index Terms*— Skin disease, artificial neural network, computer vision, clustering, dermatology.**

I. Introduction

Dermatology is the branch of medicine dealing with the hair, nails, skin and its diseases, It is a specialty with both medical and surgical aspects A dermatologist takes care of diseases, in the widest sense, and some cosmetic problems of the skin, scalp, hair, and nails. Human skin is one of the most unpredictable and difficult terrains to automatically synthesize and analyze due to its complexity of jaggedness, tone, presence of hair and other mitigating features. In a developing country like Bangladesh it is expensive for a large number of people to go to dermatologist for their skin disease problem. Every year a large number of populations in the developing countries like Bangladesh suffer due to different types of skin diseases. So, it is very necessary for both the patients and dermatologists to have an automated skin disease detection system especially in developing countries. Even though there have been several researches conducted to detect dermatological skin diseases using Computer Vision based techniques but almost every one worked for only 2-3 diseases. In our wok we have worked to detect 9 different types of skin diseases. They are Eczema, Acne, Leprosy, Psoriasis, Scabies, Foot Ulcer, Vitiligo, Tinea Corporis and Pityriasis Rosea. We have used 8 different types of algorithms for image pre-processing (YCbCr, grey image, sharpening filter, median filter, smooth filter, binary mask, histogram and sobel operator). Our system will take 10 different features from image pre-processing results and user inputs (liquid type, liquid colour, elevation, duration, feeling, gender, age). These features are used for training and testing purpose of our feed forward artificial neural networks (ANN). Using artificial neural networks (ANN) as knowledge base appears to be a promising method for diagnosis and possible treatment routines. In this article next we will discuss in brief about the related works regarding this topic, architecture, methodology, pre-processing algorithms and learning algorithm used in our proposed method. Then we will discuss about result and efficiency of our system.

II. Related Work

Detecting different types of skin diseases from colour image is a very challenging task in computer vision. Finding out different features from the colour skin images of the infected area of different skin diseases and detecting them with a high accuracy rate is the primary purpose of this research.

Researchers are working on several algorithms that can be used to detect different types of skin diseases. Kabari et al [6] crated a artificial neural networks system that predict diagnosis and routine treatment for skin diseases patients and their accuracy rate is *90%.* M. SHAMSUL et al [11] proposed an automated dermatological diagnostic system. They have used different pre-processing algorithms like ours and used feed forward back propagation artificial neural networks for training and testing purpose. Shuzlina et al [12] elaborates a prototype with back propagation neural network (BPNN) to assist the dermatologist accuracy rate was *91.2%.* Nidhal et al [15] proposed a psoriais diseases detection system using feed forward artificial neural networks. Nasir et al [2], Shang et al [7] and Nibaran et al [13] compared between different features of different types of algorithms. Hashim et al [1] used different matlab tools for image processing and Wahab et al [3] presented a texture classification system. Gerald et al [4] have introduced an approach that produces accurate overlays of thermal and visual medical images. Chang et al [8] proposed an automatic facial skin defects detection and recognition system with an accuracy rate of *98.0%.*

In this research similar approach like [6],[11] is used in order to detect different types of skin diseases from colour image. The main difference between their system and our system is, they have considered region of interest (ROI) as image extracted feature but we have considered average colour code of infected area, shape, and area size and also some inputs from user as features like elevation of the infected area, liquid colour, liquid type etc.

III. Architecture and Methodology

The system works on two phases- first pre-process the colour skin images to extract significant features and later identifies the diseases. At first we are using the colour skin images and then apply 8 different image processing algorithm on it to find some visual pattern and significant features like average colour code of infected area, infected area size in case of pixels and shape or edge detection of an infected area. Then we use user inputs like gender, age, duration, liquid type,

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

liquid colour, elevation and feeling. We train user input values along with colour skin image extracted features to train and test into a feed forward back propagation artificial neural network to identify the dermatological disease. Here we have shown the complete architecture of the system through a diagram.

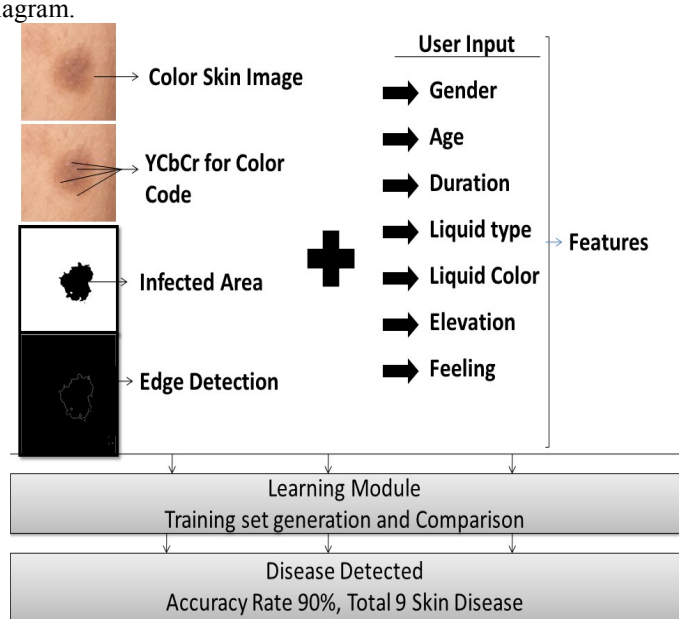


Figure 1: Complete Architecture of the system.

IV. Data Collection

Our first task was to collect necessary data's (image, information) of patients. For fulfilling this purpose we went to Sir Salimullah Medical College and Mitford Hospital, Dhaka, Bangladesh, Department of Dermatology. We took images of total 128 patients of 9 different diseases from the dermatology department.

Table 1: Data Collection Table

Disease Name	Number of Patients	Images taken
Eczema	7	28
Acne	25	152
Leprosy	5	24
Psoriasis	12	99
Scabies	40	277
Foot Ulcer	6	35
Vitiligo	15	62
Tinea Corporis	10	66
Pityriasis Rosea	8	32

A specialized doctor was present to validate and record the external data of the patients such as disease history, feeling in diseased part of body, elevation of the diseased region etc.

Table 2: Camera Specification

Camera	Sony Cyber shot DSC-W550
Focusing	The camera was focused manually to adjust the variable nature of natural light
Image Resolution	14.1 Megapixels
Shutter Speed	1/20 to 1/125 seconds (based on natural lighting)

V. Pre-processing Algorithms

We pre-processed the images and trained the resulted images in our feed forward back-propagation neural network. Eight different types of algorithms were used, they are grey image, sharpening filter, median filter, smooth filter, binary mask, histogram, YCbCr and sobel operator. The algorithms were implemented sequentially.

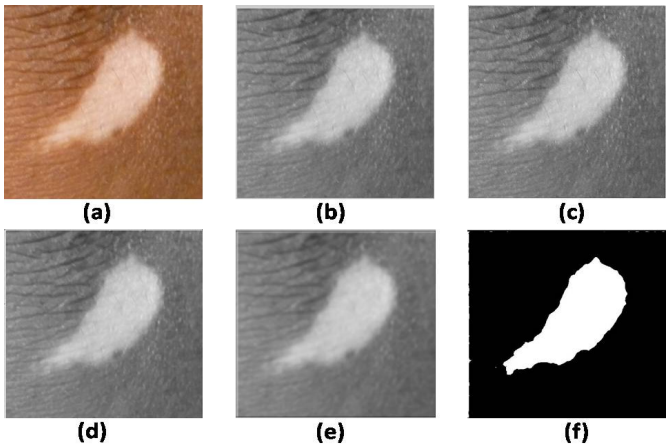


Figure 2: Result of Pre-processing algorithms. (a) Original Colour skin image, (b) Grey Scale image conversion, (c) sharpening filter, (d) Median filter, (e) Smooth Filter, (f) Binary Mask

The normal image of the infected area is converted into a greyscale image at first. Sharpening filter is applied to the greyscale image to get shaper details of the infected area. Median filter is used after sharpening filter to remove noise. We used 5*5 matrix here for median filter. Smoothing filter is applied after media filter. The idea of mean (smoothing) filtering is simply to replace each pixel value in an image with the mean ('average') value of its neighbours, including itself. Binary image was generated from the mean filtered image and distribution of colour of binary image was showed by histogram. YCbCr was used to extract average colour code of the infected area from the binary image. Sobel operator was applied to binary image to detect edge of the infected area.

Figure-2 shows the original colour skin image and results of implementing different algorithms.

VI. Feature Extraction

Two types of features were extracted are given below:

1. Features extracted from images (colour, area, edge)
2. Features extracted from user input (liquid type, liquid colour, elevation, duration, feeling, gender, age)

1. Features Extracted from Images

The first feature extracted from colour image is the colour code of the infected area of the skin. We used YCbCr algorithm to find out the colour code of the infected area from the binary image. The area or size of the infected skin is important to classify the disease. It is done by first getting the histogram graph of the binary image. After we get the histogram of the binary image we get to know the pixel of the whole area. Then we multiplied the area in pixel with the ratio of pixel to finally get the area or size of the infected skin in mm^2.

Detecting the edge of the infected area gives us the shape, as it is very important to detect the diseases. To detect the edge we have used sobel operator here. Binary mask splits the main infected area from the whole image and applying sobel

operator in the binary image gives us the shape of the area by detecting the edge.

2. Features Extracted from User Input:

The features we collected from patient's history are liquid type, liquid colour, elevation, duration, feeling, gender, age. Three different types of liquid type (blood, pus, watery) are considered. Liquid colour parameters for blood, pus and watery are red, yellowish/white and transparent respectively.

During data collection we measured elevation of the infected area for different diseases. The average approximate sizes of elevation for all diseases are 0.1cm-7cm depending on the situation expect leprosy patients because the elevation of infected area is very high for leprosy. We saw different types of feeling of patients for different diseases like itching, pain, burning sensation, numbness, headache, fever etc. Gender, age, duration these features are collected from normal patient's survey. A specialized dermatologist helped us to validate the survey we conducted on the patients.

VII. Learning Algorithm

We have used feed-forward back-propagation neural network training to perform this step. We validated and tested our system using the tenfold cross validation process. The virtue of using a cross validation technique is that there are no overlapping of the test data and training data, making the system testing results viable and dependable. We have trained our feed forward back-propagation neural network with 10 different features, 3 of them are extracted from the image pre-processing algorithms and 7 of them are from user inputs. In this system we have used 100 neurons in our hidden layer to get the best result from the system. Some sample feature values are given below that we have used in our input layer.

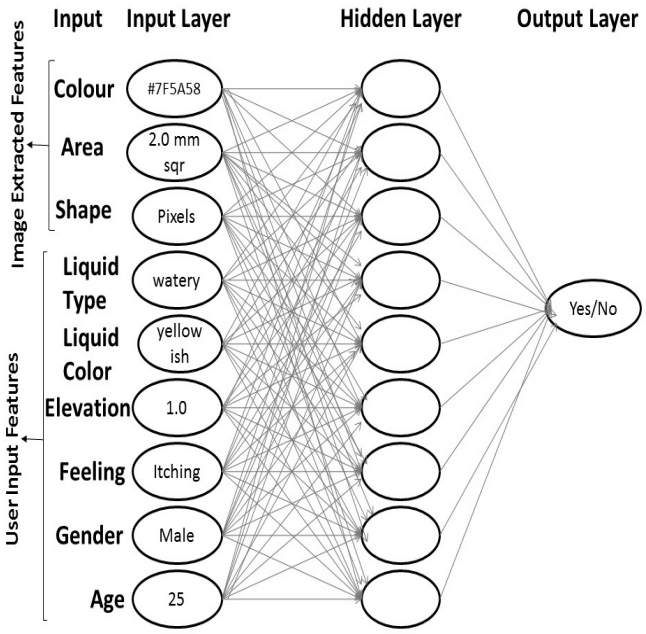


Figure 5: Feed forward back propagation artificial neural network diagram

In the figure-5 we can see that we have trained our feed-forward back-propagation neural network by 10 features. Features like colour, area, edge/shape are extracted from colour skin images and rest of the features like liquid type, liquid colour, elevation, feeling, gender, age are extracted from user input.

VIII. Result

In this research 775 colour skin images were used of 128 dermatological disease patients. The proposed system can successfully detect 9 different dermatological diseases with an accuracy of 90%. We have used 15 % of our colour skin images for testing purpose, 10% of images for validation purpose and 75% of images for training purpose. The supervised system works well than semi-supervised and unsupervised system. Detection rate of our supervised system is 90% where the detection rate of semi-supervised system is 88% and 85% for unsupervised system.

In the figure-6 we can see different detection rate for 9 different diseases. Detection rate of diseases that has low elevation in the infected area like foot ulcer, vitiligo and psoriasis are very high like 97%, 97% and 91%. Where the detection rate of diseases that has low elevation in the infected area are comparatively low which is between 85-88%. The whole processes of this proposed system starts with taking a colour skin image and pre-process the image by applying 8 different algorithms. Applying pre-processing algorithms we get three image extracted features which we use in our feed forward back-propagation neural network along with seven other user input features. Out of 9 dermatological diseases our system shows best accuracy rate for foot ulcer and vitiligo and our system did not perform according to our expectation for acne and tinea corporis.

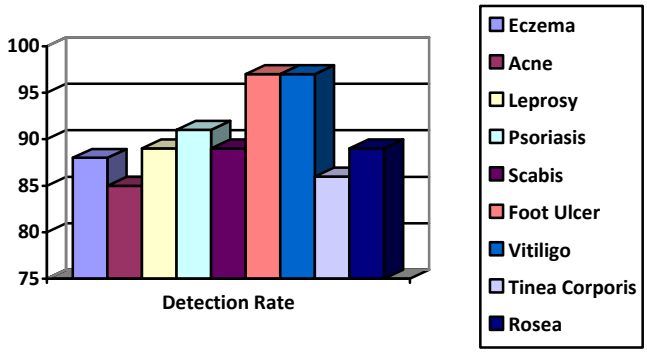


Figure 6: Detection rate of different skin diseases

In the Table- 3, we provided sample colour skin images of nine dermatological diseases, approximate elevation rate of each dermatological disease, total image used to train and test our system, successfully disease detected images and accuracy rate for each dermatological disease.

We can see that our system performed well for diseases that has low elevation in the infected area. For this type of disease images, system can generate binary image accurately and sobel operator can detect the shape of the infected area more precisely.

Table 3: Disease detection rate

Disease Name	Sample image	Elevation (cm)	Total image	Disease detected	Detection rate %
Eczema		0.1-0.3	28	23	88%

Acne		0.2-0.4	152	140	85%
Leprosy		1-7	24	20	89%
Psoriasis		0.1-0.2	99	90	91%
Scabies		0.15-0.25	277	246	89%
Foot Ulcer		NO	35	33	97%
Vitiligo		NO	62	60	97%
Tinea Corporis		0.1-0.2	66	52	86%
Rosea		NO	32	23	89%

IX. Conclusion

In this article we presented an automated system for detecting various kinds of skin diseases. At first it was not easy for us as we were trying to analyse human skin characteristics since human skin is one of the most difficult surface to analyse. It is unique and novel since our dataset was not acquired from secondary sources; rather it was a work of months of toil in an actual hospital of Bangladesh making the dataset a standard dataset which is completely new in perspective of Bangladesh. The system examines an image of infected human skin and detects the disease with an accuracy rate of 90%. We have tested a total number of 775 skin images for 9 diseases.

Acknowledgment

We are thankful to the Director Brig Dr. Zakir and Dr. Md. Masudul Hoque, Register, Dermatology & Veneriology Department of Sir Salimullah Medical College, Dhaka for helping us to collect Colour skin image data from the Medicine Department of the Hospital. We are also thankful to Adnan Firoze and M.Shamsul Arifin for giving their research data for testing our developed system.

References

[1] Hadzli Hashim, Rozita Jailani, and Mohd Nasir Taib, "A Visual Record of Medical Skin Disease Imaging Using MATLAB Tools", 2002 Student Conference on Research and Development Proceedings.
[2] Rozita Jailanil, Mohd Nasir Taib and Saadiih Sulaiman, "Colour Space for Psoriasis Skin Diseases Analysis", AsiaSENSE 2003 SENSOR, pg. 263 - 268
[3] Nesreen Abdel Wahab, Manal Abdel Wahed, Abdallah S. A. Mohamed, "Texture Features Neural Classifier of Some Skin Diseases", ©2004 IEEE
[4] Gerald Schaefer, Roger Tait and Shao Ying Zhu, "Overlay of thermal and visual medical images using skin detection and image registration", Proceedings of the 28th IEEE EMBS Annual International Conference New York City, USA, Aug 30-Sept 3, 2006
[5] Günter Schreier, *Member, IEEE,* Dieter Hayn, Peter Kastner, Silvia Koller, Wolfgang Salmhofer, and Rainer Hofmann-Wellenhof, "A Mobile-Phone based Teledermatology System to support Self-Management of Patients suffering from Psoriasis", 30th Annual International IEEE EMBS Conference Vancouver, British Columbia, Canada, August 20-24, 2008
[6] L. G. Kabari and F. S. Bakpo, *Member, IEEE,* "Diagnosing Skin Diseases Using an Artificial Neural Network", c_2009 IEEE
[7] Shang Keke, Zhou Peng, Li Guohui, "Study on Skin Colour Image Segmentation Used
by Fuzzy-c-means arithmetic", 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2010)
[8] Chuan-Yu Chang, Heng-Yi Liao, "Automatic Facial Skin Defects Detection and Recognition System", 2011 Fifth International Conference on Genetic and Evolutionary Computing
[9] Hatice Çataloluk, Metin Kesler, "A Diagnostic Software Tool for Skin Diseases with Basic and Weighted K-NN", ©2012 IEEE
[10] T. Happillon, D. Sebiskveradze, V. Vrabie, O. Piot, P. Jeannesson, M. Manfait, C. Gobinet, "FCM Parameter Estimation Methods: Application to Infrared Spectral Histology of Human Skin Cancers", 20th European Signal Processing Conference (EUSIPCO 2012) Bucharest, Romania, August 27 - 31, 2012
[11] M. Shamsul Arifin, M. Golam Kibria, Adnan Firoze, M. Ashraful Amin, Hong Yan, "Dermatological Disease Diagnosis Using Colour-skin Images", Proceedings of the 2012 International Conference on Machine Learning and Cybernetics, Xian, 15-17 July, 2012
[12] Shuzlina Abdul-Rahman, Ahmad Khairil Norhan, Marina Yusoff, Azlinah Mohamed, Sofianita Mutalib, "Dermatology Diagnosis with Feature Selection Methods and Artificial Neural Network", 2012 IEEE EMBS International Conference on Biomedical Engineering and Sciences | Langkawi | 17th - 19th December 2012
[13] Nibaran Das, Anabik Pal, Sanjoy Mazumder, Somenath Sarkar, Dwijendranath Gangopadhyay, Mita Nasipuri, "An SVM based skin disease identification using Local Binary Patterns", 2013 Third International Conference on Advances in Computing and Communications
[14] Jianbo Gao, Jun Zhang, Matthew G. Fleming, Ilya Pollak, Armand B. Cognetta, "Segmentation of Dermatoscopic Images by Stabilized Inverse Diffusion Equations", 1998 IEEE
[15] Nidhal K. Al Abbadi, Nizar Saadi Dahir, Muhsin A. AL-Dhalimi and Hind Restom, "Psoriasis Detection Using Skin Colour and Texture Features", Journal of Computer Science 6 (6): 626-630, 2010, ISSN 1549-3636, © 2010 Science Publications
[16] Mr. Harvinder Singh, Prof (Dr). J.S. Sodhi, " Image Enhancement using Sharpen Filters", International Journal of Latest Trends in Engineering and Technology (IJLTET)

Direct Strain Estimation for Ultrasound Elastography Using Peak Frequency Ratio of Radiofrequency Echo Signals

Abu Shahir Md. Khalid Hasan,* Raziur Rahman and Mohammad Ariful Haque
Department of Electrical and Electronic Engineering, Bangladesh University of Engineering Technology
Dhaka,Bangladesh
*khalidhasan27@yahoo.com

***Abstract*—Ultrasound elastography is a medical imaging technique to visualize the tissue stiffness using a conventional ultrasound machine. To compute the elastogram, the radiofrequency echo signals are acquired before and after a small applied deformation. Then the local tissue displacement field is estimated by searching the peak of the cross-correlation function between small segments of pre- and post- deformation signals. Finally, the corresponding strain distribution is imaged by differentiating the displacement field. However, the pre- and post- signals are not exactly the shifted versions of one another and hence the displacement estimation is inaccurate and the resulting strain image is highly noisy. In this paper, we propose a novel technique to estimate the strain image from the ratios of the peak frequencies of the pre- and post- deformation signals. Unlike other time/frequency domain reported schemes, our proposed technique does not use cross-correlation or any other similarity measures and directly estimate the strain in the power spectral density domain. Simulation experiments are conducted at various strain conditions and the results show that the proposed method can be an effective and computationally efficient algorithm for ultrasound elastography.**

***Index Terms*—Elastography, Strain Estimation, Power Spectrum Density, Peak Frequency Ratio**

I. Introduction

Tissue stiffness is utilized by clinicians for centuries to aid clinical diagnosis. Ultrasound Elastography is a sophisticated non-invasive alternative technique to estimate tissue stiffness. Measurement of biological tissue stiffness/elasticity is done for the purpose of pathological diagnosis as tissue stiffness or relative strain contains vital information for detection and classification of tumors in different organs such as breast [1], prostate [2], thyroid [3] and liver [4]. Normal tissue and benign tumors are typically elastic or soft and can be compressed easily by applied stress whereas malignant tumors are just about inelastic. A number of algorithms have been developed to estimate the strain distribution in elastography which have different image quality and numerical efficiency. The objective of our work is to get an improved strain image with a computationally efficient algorithm.

Strain imaging methods developed so far can be broadly classified into two groups: gradient based strain estimation and direct strain estimation. In the gradient based methods, strain image is generated from the derivative of the displacement field. The displacement can be estimated in various ways, namely time-domain cross-correlation [5], the phase root seeking (PRS) techniques. In the PRS method [6-7], the displacement/time-shift is calculated by seeking the phase-root of the zero-lag cross-correlation between the corresponding analytic pre-and post-compression windowed RF signals. The gradient based method by nature has the deficit of high noise corruption [8] and therefore may result in poor signal-to-noise ratio (SNR) strain image. Moreover, due to internal tissue compression, post-compressed RF echo signal can't be exact delayed or shifted version of pre-compressed signal [9]. To reduce the effect of this compression, temporal stretching [10] by an appropriate factor is performed on the post-compressed RF signal. Though the stretching operation improves the SNR of the strain image, the performance of the technique is highly dependent on proper choice of the stretching factor. On the other hand, in direct strain estimation method [9], the post-compressed signal is stretched by an interpolation technique and cross-correlated with the pre-compressed RF signal. The stretching factor is varied iteratively until the cross-correlation coefficient is maximized. The strain image obtained from direct method is superior to that of the gradient based method, but due to multiple stretching and matching, the former one requires significantly greater amount of computation.

In this paper, a novel ratio-based elastography method is proposed where strain is estimated by the ratio of peak frequencies of the pre- and post-compressed signals in the power spectrum density (PSD) domain. The ratio-based method has the advantage of computational efficiency as no temporal stretching and cross-correlation operation are required. The auto-correlation operation of pre- and post-compressed signals for obtaining the PSD reduces noise and increases robustness of the technique. Moreover, an adaptive search window is used to find the corresponding frequency peaks of the pre- and post-compressed signals that further improves the strain estimation accuracy. Simulation results show that the proposed technique produces better strain image as compared to the conventional PRS method.

II. Signal Model

The echo of ultrasound RF signals, as shown in Fig. 1, corresponding to pre- and post-compressed tissue can be modeled as

$$y_1(t) = x_1(t) + v_1(t) = x(t) * p(t) + v_1(t) \quad (1)$$

$$y_2(t) = x_2(t) + v_2(t) = x(t/a) * p(t) + v_2(t) \quad (2)$$

where $x(t)$ is the 1-D scattering function of the elastic target and $p(t)$ is the impulse response of the ultrasonic system [11]. Here, $v_1(t)$ and $v_2(t)$ are uncorrelated noise signal, $y_1(t)$ and $y_2(t)$ are the pre- and post-compression RF echo signals, respectively, and '*' indicates the convolution operation. Here

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

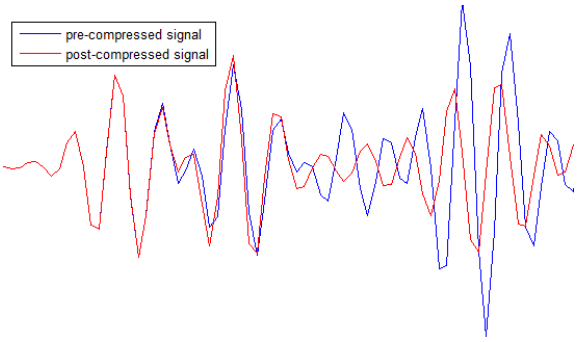


Fig. 1. Pre- and post-compressed signal in the time domain

the post-compressed signal is modeled as scaled version of the pre-compressed signal with compression factor $1/a$. The strain can be estimated as

$$s = 1 - a \tag{3}$$

where $a \leq 1$ and $s \ll 1$. Our objective is to estimate s utilizing the pre- and post-compressed signals $y_1(t)$ and $y_2(t)$.

III. Proposed Peak Frequency Ratio-based Strain Imaging Method

A. Principle of operation

In our proposed method, we estimate strain from the pre- and post-compressed signals in the PSD domain. In order to estimate the PSD, we need autocorrelation operation followed by Discrete Fourier Transform (DFT) of the pre- and post-compressed signals. The auto correlation of the pre-compression RF echo signals can be expressed as

$$r_{y_1y_1}(\tau) = r_{x_1x_1}(\tau) + r_{x_1v_1}(\tau) + r_{v_1x_1}(\tau) + r_{v_1v_1}(\tau) \tag{4}$$

Now auto correlation of pre-compressed elastic target is

$$\begin{aligned} r_{x_1x_1}(\tau) &= x_1(\tau) * x_1(-\tau) \\ &= [x(\tau) * p(\tau)] * [x(-\tau) * p(-\tau)] \\ &= [x(\tau) * x(-\tau)] * [p(\tau) * p(-\tau)] \\ &= r_{xx}(\tau) * r_{pp}(\tau) \end{aligned} \tag{5}$$

Since the pre-compressed signal and the noise are uncorrelated, the cross correlation between these two signals can be approximated to zero. Again, the auto correlation of the noise signal is a delta function. Therefore, equation (4) can be rewritten as

$$r_{y_1y_1}(\tau) = r_{xx}(\tau) * r_{pp}(\tau) + \delta(\tau) \tag{6}$$

Using the Wiener-Khintchine Theorem, the PSD of the pre-compressed RF echo can be expressed as

$$S_{y_1y_1}(\omega) = S_{xx}(\omega)S_{pp}(\omega) + U(\omega) \tag{7}$$

Similarly, the auto correlation and power spectral density of post-compressed RF echo is expressed as

$$r_{y_2y_2}(\tau) = r_{xx}(\tau/a) * r_{pp}(\tau) + \delta(\tau) \tag{8}$$

$$S_{y_2y_2}(\omega) = aS_{xx}(a\omega)S_{pp}(\omega) + U(\omega) \tag{9}$$

Comparing equation (7) and (9), it is seen that the frequencies of the post compressed echo are expanded with the ratio of the inverse of the compression factor. Now, the ratio of peak frequencies of the pre- and post-compressed PDS can estimate the strain distribution of the pathological tissue. The

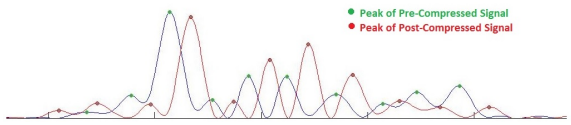


Fig. 2. Peaks of Pre and Post-compressed Signal.

importance of autocorrelation can also be observed from above equations where the noises have become an unit function. As a result, it will have a little effect on peak determination of the pre- and post- compression spectrum.

B. Ratio Estimation

Fig. 2 presents a typical PSD of small segments of pre- and post-compressed signals. The peak frequencies are indicated in the figure. The compression factor can be estimated from the ratio of the adjacent peak frequencies of the pre- and post-compressed signals. In our work, we have estimated the inverse of the compression factor using all the peak frequencies as follows:

$$\hat{a} = \frac{1}{N}\sum_{i=1}^{N}\frac{\omega_{post(i)}}{\omega_{pre(i)}} \tag{10}$$

where N is number of peaks in the PSD of the selected window block. ω_{pre} and ω_{post} represent the frequencies of the peaks of the pre- and post-compressed PSD, respectively. We have calculated the strain from the estimated $\hat{a}$. For better accuracy, only the peaks over a certain threshold magnitude has been chosen to estimate the strain distribution. There are two irregular cases for finding peaks in the pre- and post-compressed signals. In the first case, for one peak in pre-compressed signal, there are more than one peaks in post-compressed signal. Then, the first post-peak is used for corresponding pre-peak. In other case, where more than one pre-peaks are found for one post-peak, last pre peak is used to calculate the ratio.

C. Adaptive Search Window for Finding the Corresponding Peak Frequencies

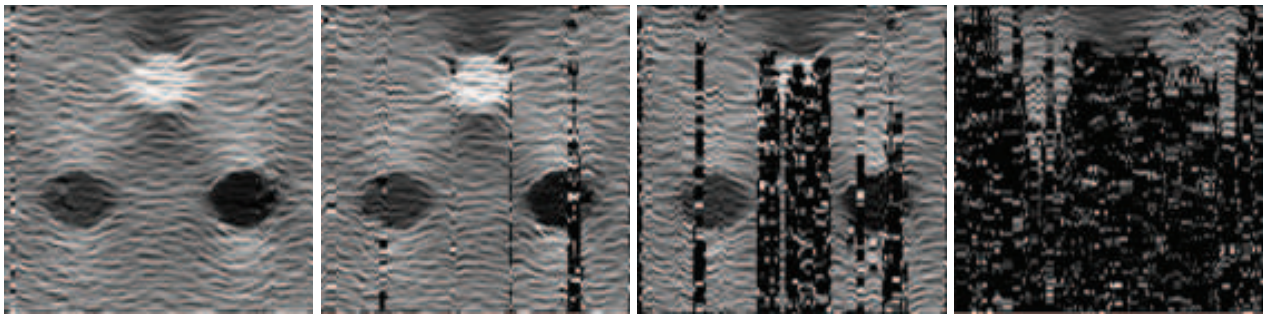
In order to estimate $\hat{a}$ in equation (10), we need to know the corresponding peak frequencies of pre- and post-compressed signals. To find the corresponding expanded peaks of the pre-compressed signal in the post-compressed spectrum, we have to set a frequency range. For example, if pre-peak is in frequency ω_k, then the frequency range to find its post-peak is set as

$$\omega_k \leq \omega \leq \omega_k * a_{rr}. \tag{11}$$

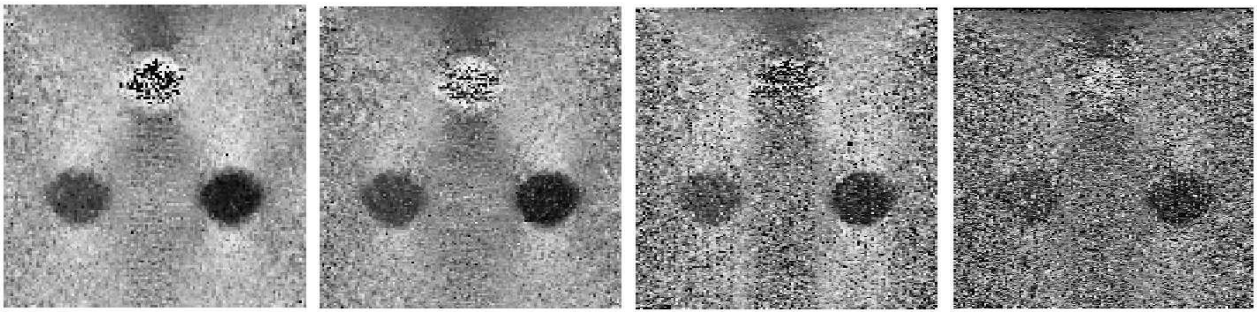
Here, a_{rr} refers to search window range to find the peak. At low strain, a fixed a_{rr} can be used. However, at high strain, the fixed range does not work. We have developed an adaptive algorithm for automatic detection of the search window range as follows:

$$a_{rr(k)} = \frac{1}{3}\sum_{i=0}^{2} a_{rr(k-1-i)} + \left(\frac{a_{rs}}{m} - 1\right). \tag{12}$$

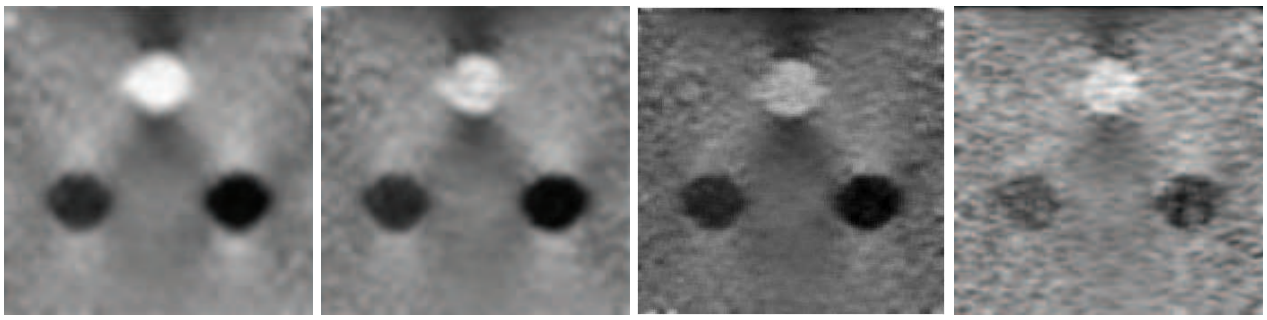
Here, $a_{rr}(k)$ refers to the adaptive range for the kth window block, m is the total number of blocks processed so far and a_{rs} is the sum of estimated ratios of the m blocks. Here, the first part of equation (12) accounts for the variation of range depending on recent estimates of a_{rr} and the second part



(a) Phase Root Seeking



(b) Estimated Strain from Ratio of Peak Frequencies in Fixed Frequency Range



(c) Proposed Peak Frequency Ratio-based Strain Imaging Method

Fig. 3. Strain images of the FEM simulation phantom generated by different methods for 2,3,4 and 6 percent compression.

provides an extra boundary based on the cumulative average value of estimated a_{rr}.

D. High Compression offset for search window range

When the applied deformation is very high, the corresponding post-peaks may go beyond the search window range obtained from equation (12). Therefore, an offset is added to the adaptive search window range, when high compression ratio is detected. By incorporating this offset value, the strain image can be generated for as high as 6% deformation. The modified equation can be simply stated as

$$a_{rr(k)} = \frac{1}{3}\sum_{i=0}^{2}\hat{a}_{k-1-i} + \left(\frac{\hat{a}_{rs}}{m} - 1\right) + HCO \qquad (13)$$

where HCO is the high compression offset. Finally the obtained strain image is smoothed using a Gaussian filter kernel.

IV. Result

In this section, we present simulation results to evaluate the performance of our proposed peak frequency ratio based strain imaging method. We also compare the results with that of the PRS method [7]. A rectangular phantom of dimension $40mm \times 40mm$ was used for the study. The phantom was simulated using finite element model (FEM) simulation software ANSYS (ANSYS, Inc., Canonsburg, PA) and Field II [12]. An ultrasound RF data were generated by Field II simulation program (10 MHz transducer center frequency and 1.5 mm beam width). The total number of samples and scan-lines were 26749 and 128, respectively.

Fig. 3 presents the strain images obtained using our proposed and PRS methods at different applied compressions. The elastograms in Fig. 3(a) are generated by the PRS method. The method works well for low strain, but for higher strain, strain estimation deteriorates significantly. In Fig. 3(b) and (c), strain images are shown for the proposed ratio-based method considering fixed and variable search window, respectively. For fixed search window method, strain estimation has been improved compared to the PRS method, thought the image is quite noisy at higher strain. Fig. 4 explains the reason of this noisy strain image. It is seen that with increasing strain, the frequency distance between pre and post peaks increases. Now if a predetermined fixed search window is used, some post peaks may go beyond the search window, especially at higher strain. As a result, the strain estimation is erroneous and the image gets noisy. The problem is overcome using an

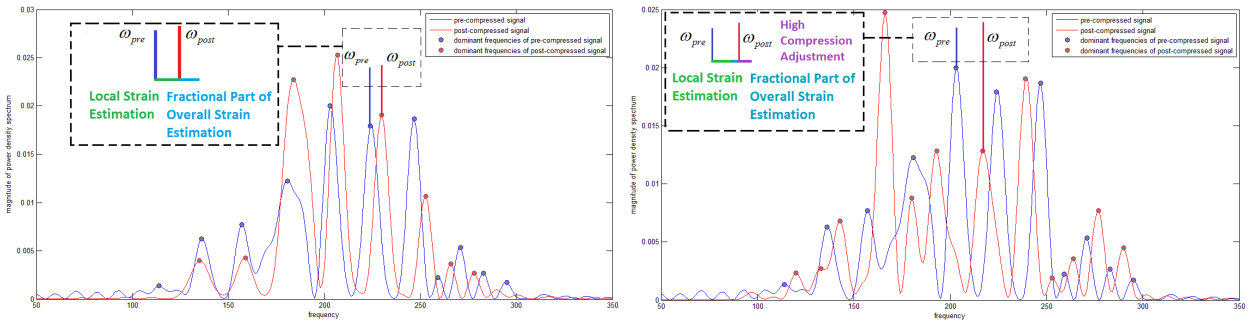


Fig. 4. Power Density Spectrum of a block of data for 2 and 6 %.

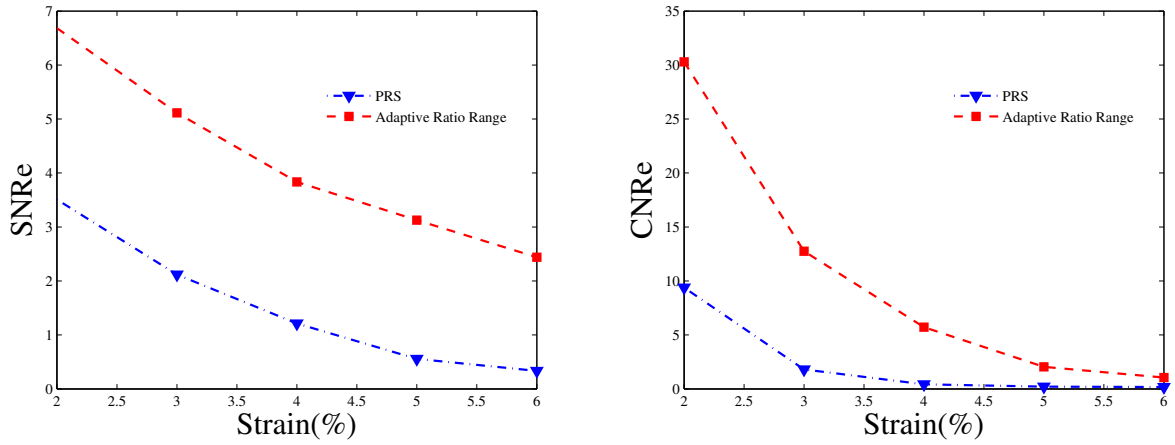


Fig. 5. Performance Comparison of Different Methods using Numerical performance Metrics for FEM simulation Data(a)SNRe vs. Applied Strain (b)CNRe vs. Applied Strain.

adaptive search window along with a high compression offset as shown in Figure 3(c). Now, the strain estimation works quite well for up to 6 % applied strain.

The quantitative performances of the proposed and PRS methods were evaluated by estimating the elastographic signal-to-noise ratio (SNRe)[13] and elastographic contrast-to-noise ratio (CNRe)[14] from the obtained strain images. The results are presented in Fig. 5 for 2% to 6% applied strain. It is observed that the proposed method performs significantly better than the PRS methods in terms of SNRe and CNRe.

V. CONCLUSION

In this paper, a novel algorithm for elastography has been proposed to estimate strain distribution using ratio of dominant frequencies of pre- and post-compressed signals. The technique does not require any cross-correlation and/or temporal stretching operations like any other reported techniques. As a result, it is computationally efficient. Simulation results show that the propose technique provides better strain image compared to the conventional gradient based technique.

REFERENCES

[1] B. S. Garra, E. I. Cespedes, J. Ophir, S. R. Spratt, R. A. Zuurbier, C. M.Magnant, and M. F. Pennanen, "Elastography of breast lesions: initial clinical results," *Radiology.*, vol. 202, no. 1, pp. 79-86, 1997.

[2] K. K ONIG, U. Scheipers, A. Pesavento, A. Lorenz, H. Ermert, andT. Senge, "Initial experiences with real-time elastography guided biopsies of the prostate," *The Journal of urology.*, vol. 174, no. 1, pp. 115-117, 2005.

[3] C. Vorl ander, J. Wolff, S. Saalabian, R. H. Lienenl uke, and R. A. Wahl, "Real-time ultrasound elastographya noninvasive diagnostic procedurefor evaluating dominant thyroid nodules," *Langenbeck's Archives of Surgery*, vol. 395, no. 7, pp. 865-871, 2010.

[4] M. Friedrich-Rust, M.-F. Ong, E. Herrmann, V. Dries, P. Samaras,S. Zeuzem, and C. Sarrazin, "Real-time elastography for noninvasiveassessment of liver fibrosis in chronic viral hepatitis," *American Journal of Roentgenology*, vol. 188, no. 3, pp. 758-764, 2007.

[5] J. Ophir, I. CQspedes, H. Ponnekanti, Y. Yazdi, and X. LI, "Elastography: a quantitative method for imaging the elasticity in biological tissues," *Ultrason. Imaging*, vol. 13, pp. 111-134, 1991.

[6] A. Pesavento, C. Perrey, M. Krueger, and H. Ermert, "A time-efficient and accurate strain estimation concept for ultrasonic elastography using iterative phase zero estimation," *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on*, vol. 46, no. 5, pp. 1057-1067,1999.

[7] J. E. Lindop, G. M. Treece, A. H. Gee, and R. W. Prager, "Phase-based ultrasonic deformation estimation," *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on*, vol. 55, no. 1, pp. 94-111, 2008.

[8] Varghese T, Konofagou EE, Ophir J, Alam SK and Bilgen M., "Direct strain estimation in elastography using spectral cross-correlation," *Ultrasound in Med and Biol* , vol. 26, no. 9, pp. 1525-1537, 2000.

[9] Alam SK, Ophir J, Konofagou EE, "An adaptive strain estimator for elastography," *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on* , vol. 45, no. 2, pp. 461-472, 1998.

[10] Cespedes I, Ophir J, Ponnekanti H, Maklad N., "Elastography: Elasticity imaging using ultrasound with application to muscle and breast in vivo.," *Ultrason Imaging*, vol. 15, no. 2, pp. 73-88, 1993.

[11] M. K. Hasan, E. M. A. Anas, S. K. Alam, and S. Y. Lee, "Direct Mean Strain Estimation for Elastography Using Nearest-Neighbor Weighted Least-Squares Approach in the Frequency Domain," *Ultrasound in Medicine and Biology*, vol. 38, no. 10, pp. 1759-1777, 2012.

[12] J. A. Jensen, "Field: A program for simulating ultrasound systems," *10th Nordic-Baltic Conf. Biomedical Image*, vol. 13, no. 4, pp. 351-353, 1996

[13] I. C espedes and J. Ophir, "Reduction of image noise in elastography," *Ultrasonic Imaging*, vol. 15, no. 2, pp. 89-102, 1993.

[14] T. Varghese and J. Ophir, "An analysis of elastographic contrast-to-noise ratio," *Ultrasound in medicine & biology*,vol. 24, no. 6, pp. 915-924, 1998.

Bessel K-Form Parameters in the Dual Tree Complex Wavelet Transform Domain for the Detection of Epilepsy and Seizure

Anindya Bijoy Das* and Mohammed Imamul Hassan Bhuiyan
Department of Electrical and Electronic Engineering(Bangladesh University of Engineering and Technology, Bangladesh)
Email: *anindya149@yahoo.com

***Abstract*—In this paper, a statistical analysis of EEG signals is carried out in the dual tree complex wavelet transform (DT-CWT) domain. It is shown that Bessel k-form(BKF) pdf can suitably model the DT-CWT sub-bands and the BKF parameters in various DT-CWT sub-bands can discriminate various types of EEG data effectively. Next these parameters are utilized by the SVM-based classifiers to classify the EEG data. The classification performance is studied for three clinically relevant cases including healthy vs seizure, non-seizure vs seizure and inter-ictal vs ictal recordings. The proposed method provides 100% accuracy with 100% sensitivity and 100% specificity in all the cases. In addition, in comparison to several state-of-the-art algorithms, the proposed method has also been shown to be computationally fast.**

***Index Terms*—Electroencephalogram(EEG), Seizure, Dual Tree Complex Wavelet Transform(DT-CWT), Bessel K-Form(BKF), Support Vector Machine(SVM).**

I. Introduction

About 1% of the world's population are affected with epilepsy, which is characterized with recurrent seizures [1]. It has serious health implications as well as involve economic loss since it may cause injury, sometimes death and constricts the work capacity of epileptic patients. Traditionally, epilepsy diagnosis for seizure detection is carried out by electroencephalographers, highly trained professionals, who observe long term records, hours to days, even weeks of electroencephalogram(EEG) signals. Obviously, this is an expensive and time-consuming process and may be prone to error due to its subjective nature. Thus, automated detection of seizures from EEG signals can assist the electroencephalographers in this respect by relieving off this laborious job of monitoring EEG signals. In additon, it may provide solutions for therapeutic devices such as the cranially implanted neuro-stimulators, such as Responsive neuro-stimulators(RNS) [2].

Among various algorithms in the literature, the time-frequency-based ones ([3]- [9]) have been shown to be highly promising for automatic detection of seizures. One of the main reasons behind the success of these methods may be the brain dynamics being spread over a number of time-frequency EEG sub-bands [5]. The Dual Tree Complex Wavelet Transform (DT-CWT) has been recently introduced as a highly efficient time-frequency transform having improved directionality, limited redundancy, shift invariance and less computational-complexity [10]- [11]. It is shown in [8] and [9] that the higher order statistical moments in the dual tree complex wavelet transform(DT-CWT) sub-bands can discriminate and classify the EEG data into various classes such as seizure and non-seizure. The higher-order moments can be obtained using the underlying probability density function (pdf) that can model the statistics of the data accurately. The first objective of this paper is to investigate whether the Bessel K-Form (BKF) is a suitable pdf to model the DT-CWT sub-bands of EEG data and to see whether its parameters can discriminate the EEG data. The motivation for using the BKF pdf arises from its ability to model the heavy-tailed statistics of non-linear data effectively [12]- [13]. The second goal is to develop Support Vector Machine (SVM)-based classifiers to effectively classify the EEG signals wherein the BKF parameters extracted from the various DT-CWT sub-bands are utilized as features. The analysis is carried out using a publicly available benchmark EEG database [14], [15]. The performance of the SVM-based classifiers is studied for a number of clinically relevant classification cases and compared to that of several other algorithms available in the EEG literature.

II. Dual Tree Complex Wavelet Transform

In this Section, a brief introduction of the DT-CWT is provided. Traditional Discrete wavelet transform(DWT) offers good time-frequency localization. However, it is not redundant, lacks shift-invariance and gives directional information only in three directions. The undecimated wavelet transform(UWT), albeit shift-invariant, has similar directional information and is computationally expensive. The DT-CWT provides a reasonable solution in this respect. It is shift invariant while being only 2^d times redundant for a d-dimensional signal as of the DWT and offers directional information in six directions, thus being more efficient in time frequency localization of signal features. The DT-CWT has two DWT trees. The first tree represents the real part of the complex coefficients. whereas the second represents the imaginary parts.

Several wavelet functions are available in the literature [10]. In this paper, the Farras wavelets [16] are employed for performing DT-CWT decomposition.Fig. 1 shows the plots of two sample EEG segments for eight seconds from the datasets D (top left) and E (top right) in the first row and the plots of the corresponding first level DT-CWT coefficients in the second and the third row.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

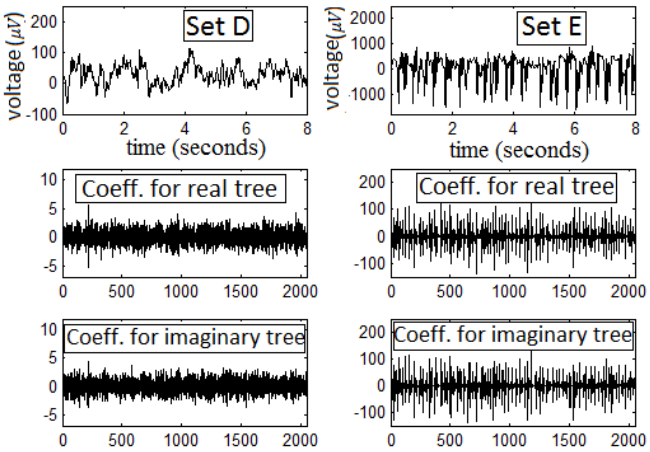


Fig. 1. Sample EEG Signals and the corresponding DT-CWT coefficients

III. Analysis of EEG in DT-CWT domain

The EEG database used in this paper has a considerable volume with the collection of five different types of recordings (Sets A, B, C, D and E) providing the statistical significance of the analysis. The database has been widely used in the EEG literature ([3]- [9]). The sampling frequency of this EEG database is 173.61 Hz, and thus, the highest frequency component is 86.8 Hz. But, the frequency range of an EEG signal spans over 0 to 60 Hz. Thus, the frequencies greater than 60 Hz may be considered as noise [5] and those are removed by using a 6th order Butter-worth filter.

Next, the band-limited signals are subjected to a 4 level DT-CWT decomposition. After four levels of decomposition of a signal **x**(0-60 Hz), the components z4 (0-4 Hz), y4 (4-8 Hz), y3 (8-15 Hz), y2 (15-30 Hz), and y1 (30-60 Hz) are obtained [8]. Reconstructions of these five components using the inverse DT-CWT approximately correspond to the five physiological EEG sub-bands delta, theta, alpha, beta, and gamma respectively [5]. Since, each DT-CWT coefficient has two parts, real and imaginary, the 4-level decomposition yields ten sub-bands in total (five for real and five for imaginary). The DT-CWT coefficients for any sub-band are characterized with higher order statistical moments such as variance, skewness and kurtosis. The dispersion, asymmetry and peakedness of a data-set are well described by these moments. These higher-order moments of the data can be obtained using the under-lying probability density function (pdf) that can model the statistics of the data accurately. Thus, it might be interesting to see whether the parameters of that suitable pdf can be used to differentiate EEG data. In this paper, the Bessel k-form (BKF) pdf is proposed as a suitable prior for modeling the EEG data in the DT-CWT domain. The BKF pdf is unimodal and symmetric around the mode (the mode necessarily not being zero). The BKF pdf is expressed as:

$$f(x) = \frac{1}{\sqrt{\pi}\Gamma(p)} \left(\frac{c}{2}\right)^{-\frac{p}{2}-\frac{1}{4}} \left|\frac{x}{2}\right|^{p-\frac{1}{2}} K_{p-\frac{1}{2}} \left(\sqrt{\frac{2}{c}}|x|\right) \quad (1)$$

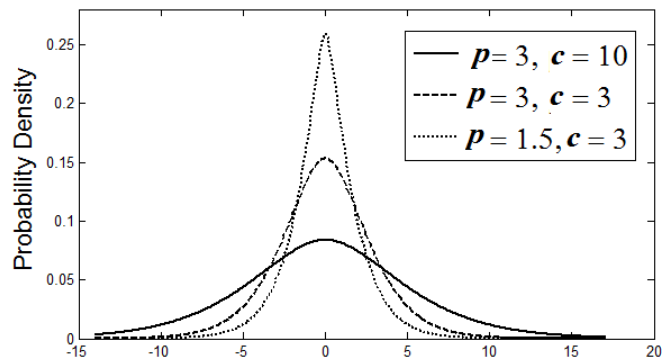


Fig. 2. Changes in shape of the BKF pdf for various values of its parameters

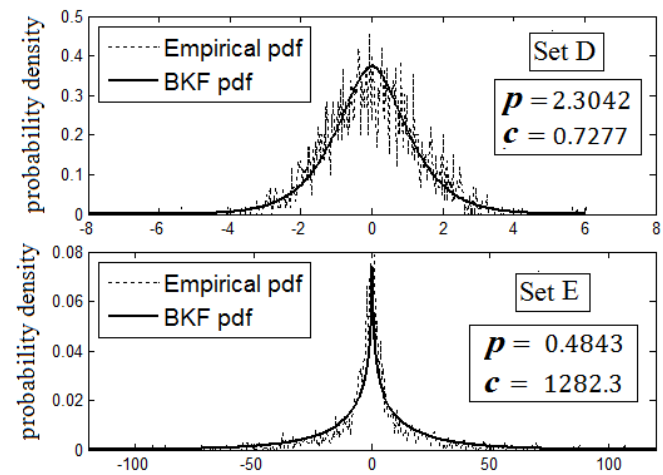


Fig. 3. Plots of the empirical pdf with the BKF pdf for the sub-band (y1,1) of two sample EEG signals along with the values of respective BKF parameters

where K_v is the modified Bessel function of the second kind of order v, and p and c are the BKF parameters where $p, c > 0$. For a BKF pdf, p controls the peakedness of the pdf and c controls the dispersion. Fig. 2 shows the plots of the pdf for various values of the parameters p and c. The values of the parameters p and c are estimated [12] from the different EEG sub-bands using the following equations-

$$p = \frac{3}{kurtosis(x) - 3} \quad (2)$$

$$c = \frac{variance(x)}{p} \quad (3)$$

Fig. 3 shows the plots of the empirical pdfs and the corresponding BKF pdfs of the y1 sub-band (from the real tree) for two sample EEG signals from Set D and Set E. It is seen that the BKF pdfs fit the empirical ones quite well.

For analysis, the parameters p and c are estimated from all the DT-CWT sub-bands of 100 data segments for each of the five sets. Table I provides the corresponding p-values obtained from the ANOVA analysis. Note that, the real and imaginary trees are indicated by tree 1 and 2, respectively. For example, (z4,1) means the z4 sub-band (0-4 Hz) of the real tree or (y3,2) means the y3 sub-band (8-15 Hz) of the imaginary tree.

TABLE I
P-VALUES FOR ANOVA ANALYSIS FOR THE BKF PARAMETERS

Signals	p-values for p	p-values for c
Band-limited	0.2386	0.0014
(y1,1)	0.4609	1.3456e-15
(y1,2)	0.2674	8.6940e-18
(y2,1)	0.5713	1.0294e-30
(y2,2)	0.3521	4.2977e-30
(y3,1)	0.4972	2.0112e-19
(y3,2)	0.4298	2.2076e-19
(y4,1)	0.5132	2.2392e-04
(y4,2)	0.2835	1.6270e-07
(z4,1)	0.8919	2.1929e-05
(z4,2)	0.4678	4.1029e-04

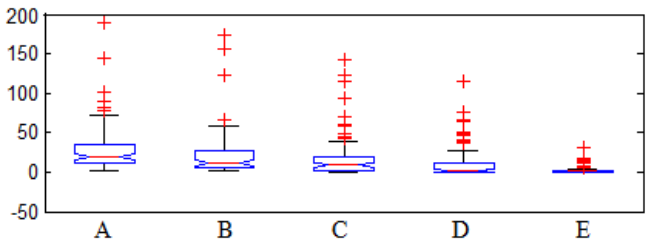


Fig. 4. Box plots for p for the real tree of y1 sub-bands

It is seen from Table I that the p-values for c are smaller than those for p in all the cases which indicates a better discrimination capability of c than p to classify the EEG data into different classes. However, the p-values for the band-limited segments are not significantly small enough to distinguish among the EEG data. Among the box plots, in Fig. 5, there is a more obvious and precise indication than in Fig. 4 for the discrimination of the seizure recordings (Set E) from the non-seizure ones (Sets A,B,C and D) which also confirms the superior discrimination capability of the parameter c over p. Thus, the BKF parameter, c, extracted from the DT-CWT sub-bands may be used as the features in the classifier for discriminating the EEG signals.

IV. THE PROPOSED SVM-BASED CLASSIFICATION

In this Section, the proposed classification procedures are described. In this paper, Support Vector Machine (SVM), proposed by Vapnik [17], is used for classifying various types of EEG signals. It is based on statistical learning theory

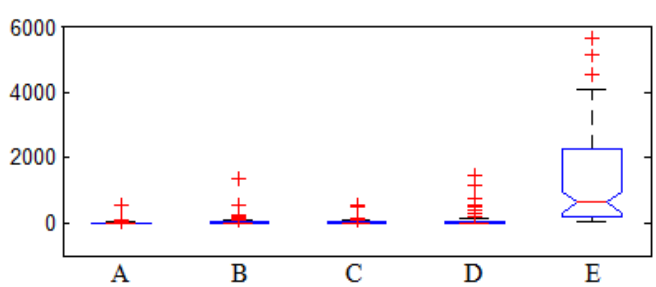


Fig. 5. Box plots for c for the real tree of y1 sub-bands

where the nonlinear separable samples are projected onto a higher dimensional space using kernel functions. Subsequently, the classification is performed by locating the optimal separating hyper-plane in the projected space. The various kernel functions are studied empirically and the radial-basis function(RBF) kernel provides the best result.

In order to perform feature extraction from an EEG record, each data-set is first divided into 16 non-overlapping segments due to the non-stationary nature of EEG data [3]. As there are 100 EEG data-sets from each of the Sets A,B,C,D and E, 500×16=8000 segments are generated in total. After that, the BKF parameter, c is calculated from each of the DT-CWT sub-bands for all the segments. For example, Set E has 100×16=1600 segments, giving 1600×10=16000 sub-bands, thus 16000 values of c are obtained for Set E. Finally, for a particular set, half of the segments chosen randomly are used for training, the rest half for testing in a SVM classifier.

For the five sets of EEG records described earlier, three different cases of classification problems are considered. The cases are chosen based on their clinical relevance and their widely use in various papers in the literature. In Case I, Set A and Set E are classified into the healthy and seizure classes respectively. For Case II, Set D and Set E are classified into the inter-ictal and ictal classes, respectively. It separates the ictal recordings from the inter-ictal ones which is related to the detection of the onset of seizures. In Case III, segments from Sets A, B, C and D are combined together to form the non-seizure class and Set E is classified seizure class.

The performance criteria of the proposed classification method in classifying different types of EEG data are-

$$Sensitivity = \frac{TP}{TP+FN} \times 100\% \tag{4}$$

$$Specificity = \frac{TN}{TN+FP} \times 100\% \tag{5}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \tag{6}$$

where TP, TN, FP and FN stand for true positive, true negative, false positive and false negative events, respectively [3].

The performance of the proposed method is first studied using the features from band-limited (BL) EEG signals only, and a particular DT-CWT sub-band. Table II shows the performance(per) for all the classification cases in terms of sensitivity (sen), specificity (spec) and accuracy (acc) using the BKF parameter c estimated from different sub-bands and band-limited segments. It is seen that in almost all the cases the high frequency sub-bands like y1 or y2 provide better performances than the low frequency ones like y4 or z4. Next, Table III shows the performance for the combinations of features from different sub-bands. It is seen that 100% accuracy is achieved along with 100% sensitivity and 100% specificity for all three cases using features from y1 to z4 sub-bands. As DT-CWT consists of two trees which indicates that for sub-band level, two values of c are obtained. Thus, when $5 \times 2 = 10$ features are utilized, it provides 100% accuracy for

all the cases. However, for Case I, six features from sub-bands y1,y2 and y3 also provide 100% accuracy.

TABLE II
PERFORMANCE FOR VARIOUS CASES USING FEATURES FROM BAND-LIMITED SIGNAL AND A SINGLE SUB-BAND

Cases	Per	Names of the Used Segments					
		BL	y1	y2	y3	y4	z4
Case I A,E	Sen	89.56	91.13	96.81	93.63	86.88	88.13
	Spec	99.44	99.75	98.81	99.75	99.69	98.31
	Acc	94.5	95.44	97.81	96.69	93.78	92.22
Case II D,E	Sen	83.19	92.81	97.13	95.75	89.56	85.94
	Spec	89.63	96.56	98	97.81	95.81	95.13
	Acc	86.41	94.69	97.56	96.78	92.69	90.53
Case III ABCD,E	Sen	87.31	91.44	95	92.75	89.13	88.25
	Spec	94.66	93.94	96.69	95.56	95.50	92.52
	Acc	93.19	93.44	96.35	95	94.23	91.66

TABLE III
PERFORMANCE FOR VARIOUS CASES USING FEATURES FROM MULTIPLE SUB-BANDS

Cases	Per	Names of the Used Segments			
		y1 to y3	y2 to y4	y3 to z4	y1 to z4
Case I A,E	Sen	**100**	98.4625	96.9375	**100**
	Spec	**100**	99.75	99.5625	**100**
	Acc	**100**	99.1063	98.25	**100**
Case II D,E	Sen	99.5	97.75	94.75	**100**
	Spec	99.125	98.9375	98.875	**100**
	Acc	99.3125	98.3438	96.8125	**100**
Case III ABCD,E	Sen	99	97.8125	93.6875	**100**
	Spec	99.375	98.3281	96.6094	**100**
	Acc	99.3	98.225	96.0250	**100**

TABLE IV
COMPARISON OF CLASSIFICATION PERFORMANCE OBTAINED FOR VARIOUS ALGORITHMS

Cases	Methods	Classifier	No. of Features	Acc(%)
Case I	Tzallas [4], (2007)	ANN	16	100
	Fathima [6], (2011)	Lin. Classifier	3	96.9
	Alam [3], (2013)	ANN	3	100
	Proposed	SVM	6	**100**
Case II	Liang [7], (2010)	SVM	16	98.74
	Alam [3], (2013)	ANN	3	100
	Proposed	SVM	10	**100**
Case III	Tzallas [4], (2007)	ANN	16	97.73
	Liang [7], (2010)	SVM	16	98.51
	Alam [3], (2013)	ANN	3	100
	Proposed	SVM	10	**100**

The performance of the proposed classifier is compared with that of several existing methods in Table IV. Along with providing 100% accuracy, the proposed method is also computationally fast. It is implemented in MATLAB in a desktop computer with a core-to-duo processor of 2.66 GHz and 2 GB RAM. The proposed method requires 0.003∼0.005 seconds to extract features. On the contrary, the method of [3] needs 0.2∼0.6 seconds. The method in [4] requires 3.8∼4.09 seconds to extract features from a 23.6 seconds segment.

V. CONCLUSION

In this paper, EEG signals have been analyzed in the DT-CWT domain and the suitability of using the BKF pdf to model the DT-CWT sub-bands has been explored. It has been shown that the BKF parameters can discriminate the EEG data. SVM-based classifiers have been developed to classify EEG signals using BKF parameters computed from the DT-CWT sub-bands. For three classification cases, the proposed SVM-classifiers have been shown to provide 100% accuracy, sensitivity and specificity. Furthermore, the proposed method is computationally faster in comparison to several state-of-the-art algorithms. However, to confirm the use of the proposed method for clinical applications, experiments with long-term EEG recordings are necessary. It is expected that the proposed classifier would demonstrate similar performance since it employs the underlying statistics of EEG signals in time-frequency domain and thus, the high variability of background and seizure characteristics is avoided.

REFERENCES

[1] S. Liang, Y. Chen, Y. Wang, P. Chen, C. Yang and H. Chiueh,"A hierarchical approach for online temporal lobe seizure detection in long-term intracranial EEG recordings," *Journal of Neural Engineering*, 2013.

[2] M. T. Salam, M. Sawan and D. K. Nguyen,"Low-Power Implantable Device for Onset Detection and Subsequent Treatment of Epileptic Seizures," *Journal of Healthcare Engineering*, vol. 1, no. 5, 2010.

[3] S M S. Alam and M. I. H. Bhuiyan, "Detection of Seizure and Epilepsy using Higherorder Statistics in the EMD Domain," *IEEE Transactions on Information Technology in BioMedicine*, vol. 17, pp. 312-318, 2013.

[4] A. T. Tzallas, M. G. Tsipouras, and D. I. Fotiadis, "Automatic Seizure Detection Based on Time-Frequency Analysis and Artificial Neural Networks," *Computational Intelligence and Neuroscience, Hindawi Publishing Corporation*, vol. 2007, no. 18, August,2007.

[5] H. Adeli, S. G.Dastidar and N. Dadmehr , "A Wavelet-Chaos Methodology for Analysis of EEGs and EEG Sub-Bands to Detect Seizure and Epilepsy," *IEEE Transactions on Biomedical Engineering*, vol:54, 2007.

[6] T. Fathima, Y. U. Khan, M. Bedeeuzzaman, and O. Farooq,"Discriminant Analysis for Epileptic Seizure Detection", in Proc. *International Conference on Devices and Communications*, 2011.

[7] S. F. Liang, H. C. Wang and W. L. Chang, "Combination of EEG Complexity and Spectral Analysis for Epilepsy Diagnosis and Seizure Detection", *EURASIP Journal on Advances in Signal Processing*, vol. 2010, article id. 853434, Hindawi Publishing Corporation, 2010.

[8] A.B. Das, M.I.H. Bhuiyan and S.M.S. Alam,"Statistical Parameters in the Dual Tree Complex Wavelet Transform Domain for the Detection of Epilepsy and Seizure", Proc. of *International Conference on Electrical Information and Communication Technology*, February, 2014.

[9] A. B. Das, M. I. H. Bhuiyan and S. M. S. Alam,"A Statistical Method For Automatic Detection of Seizure and Epilepsy in the Dual Tree Complex Wavelet Transform Domain", Proc. of IEEE *International Conference on Informatics, Electronics and Vision(ICIEV-2014)*, Bangladesh.

[10] W. Selesnick, R. G. Baraniuk, and N. Kingsbury, "The dual tree complex wavelet transform," *IEEE Signal Processing magazine*, vol. 22, no. 6, pp. 123-151, November 2005.

[11] S.M.M. Rahman, M.O. Ahmad and M.N.S. Swamy, "Video Denoising Based on Inter-frame Statistical Modeling of Wavelet Coefficients," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.17, no.2, pp.187-198, Feb. 2007

[12] A. Srivastava, A.B. Lee, E.P. Simoncelli and S.C. Zhu, "On Advances in Statistical Modeling of Natural Images," *Journal of Mathematical Imaging and Vision*, vol. 18, pp.1733, 2003.

[13] J. M. Fadili and L. Boubchir, "Ananytical form for a Bayesian wavelet estimator images using the Bessel K-Form densities," *IEEE Trans. On Image Processing*, vol. 14, no. 2, 2005.

[14] EEG time series download page [Online] Available: http://epileptologie-bonn.de/cms/front_content.php?idcat=193&lang=3&changelang=3

[15] Ralph G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," vol. 64, no. 6, November, 2001.

[16] Wavelet Software: http://eeweb.poly.edu/iselesni/WaveletSoftware/

[17] V. Vapnik, "The Nature of Statistical Learning Theory", 1995.

Connected Component Based ROI Selection to Improve Identification of Microcalcification from Mammogram Images

Md. Enamul Haque,[1*] K. M. Imtiaz-Ud-Din,[2] Md. Muntasir Rahman,[3] and Aminur Rahman[4]
[1]Department of Computer Engineering, King Fahd University of Petroleum and Minerals
Dhahran-31261, Kingdom of Saudi Arabia
[2]Department of Computer Science and Engineering, East West University
Dhaka-1219, Bangladesh
[3,4]Department of Computer Science and Engineering, Islamic University of Technology
Board Bazar, Gazipur-1704, Bangladesh
*enamul_cse@yahoo.com

***Abstract*—In this paper, we propose automated image processing method for mammogram image analysis to enhance the classification of malignant and benign clustered micro calcifications. Mini-Mias, one of the most renowned mammogram dataset, is used for our experiment. Region of interest (ROI) is selected using connected component labeling method. Shape and textural features are computed and used in five different classifiers to validate the correctness of labeling method usage for this specific research. Experimental results show that Random Forest and Bagging classifier can produce best classification accuracy among them. We used Receiver operating characteristic (ROC) analysis to asses and differentiate the classification performance from different methods. Random Forest and Bagging, both classifiers provide more than 99% accuracy.**

***Index Terms*—Connected Component Labeling, Breast Cancer Detection, Mammogram Analysis, Image Processing**

I. Introduction

Breast cancer is one of the prevalent diseases observed among the women and the incident rate is in upward trend in recent years. There are several ways to identify breast cancer using the images from the radiologists. According to the physicians and past researchers, mammography is the most effective technique to detect this cancer at early stage. There are some research on this issue focusing on CAD (Computer Aided Diagnosis), mammogram, ultrasound and MRI (Magnetic Resonance Imaging) and clinical prototypes are built as well to improve the diagnostic performance. In this paper, we extract region of interest from Mini-Mias database and compute several features to use as feature vector for classifiers. In our approach, we apply connected component labeling to find the ROI from each mammogram and construct a feature extractor to collect features from them.

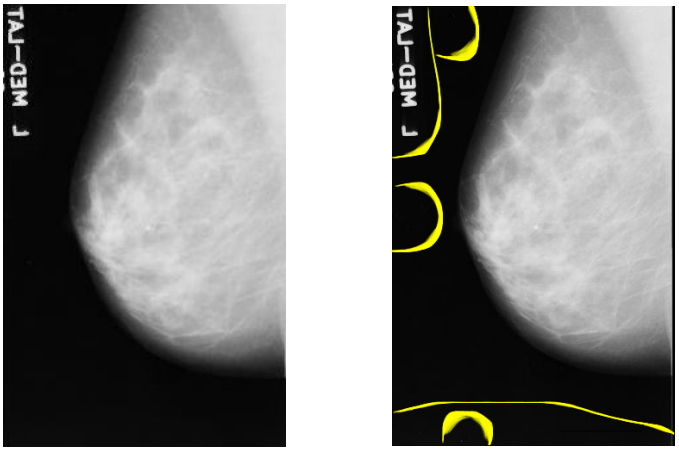
The following is a brief analysis of other contemporary research on classification of micro calcification from mammograms. Solka et al. [1] emphasized on the high dimensional features through two techniques, namely, dimensionality reduction and discriminant analysis. They used different version of contrast, smoothness, edge strength, gradient, moment, compactness, shape, and average grey level moment as features for the breast images. Later they performed Fisher Linear Discriminant (FLD) and Principal Component Analysis (PCA) to reduce dimensionality and followed Bayesian approach for classification. They show that FLD performs better than PCA while classification. Hung et al. [2] mentioned neural network to diagnose breast cancer in more depth using resampling techniques like bootstrap.They focus on building an integrative approach to build neural network models to resolve the issues of model selection, feature selection and function approximation. Their goal was to produce more accurate classification for posterior probabilities. Giger et al. [3] explained the necessity of the CAD technologies for determining the cancer indication. She mentioned three techniques, MRI, Mammography and Ultrasound and explained the comparison among them in brief. Their observation concludes that Mammogram data analysis provides better indication about the cancer detection method. Haddadnia et al. [4] used both statistical and thermal features to diagnose the breast cancer from medical infrared images. They calculated mean, variance, skewness, kurtosis and entropy to form a combinatorial model with the help of artificial neural network and genetic algorithm. It was mentioned that thermal pattern and kurtosis are the main parameters for diagnosing breast cancer form medical imaging.

Hassanien et al. [5] showed various machine intelligence techniques to diagnose breast cancer from medical images. They used type-II fuzzy algorithm to improve the quality of MRI images and then performed segmentation using pulse coupled neural networks to extract region of interests. Finally, wavelet transformation was performed on those regions to get the original information to process via support vector machine and decide whether they represent cancer or not. Huang et al. [6] used vessel analysis in 3D power doppler ultrasound images of 113 solid breast masses. Vessel analysis estimates the difference between blood vessels of tumor and non tumor areas. Six features including volume of vessels, vascularity index in normal tissue, and vascularity index in surrounding region of the tumor within 2 cm were measured; later neural network was applied to classify the tumors. The receiver operating characteristic (ROC) curve analysis was used to evaluate the performance. Lee et al. [7] proposed an efficient classifier with feature selection based on fuzzy entropy for pattern classification which was formulated from Shanon entropy.They performed their experiment in Iris and Wisconsin Breast Cancer Diagnostic Database(WBCDB).

The rest of the paper is organized as follows: A brief

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

discussion on region of interest (ROI) selection method, mathematical formulation, and description on connected component technique and its application in our research is presented in section II. Performance evaluation and results are discussed in section IV. Finally, section V concludes and suggests future direction of our research.



(a) (b)

Fig. 1. (a) Raw X-ray mammogram from MiniMias Dataset (image number mdb029.pgm). (b) Generalized marking of the objects detected by the connected component labeling algorithm.

II. ROI Selection Using Connected Component

In this paper, the mammogram image [8] classification is separated in two groups with seven classes, i.e., an image can be either clustered with micro-calcification or not. These two classes are known as malignant and benign by the radiologists. It is not obvious that a breast tissue with malignant micro-calcification is sure to have some cancer in it.

The first step is to preprocess the mammogram images which is the raw image data from the X-ray machine. The image contains additional data(mammogram date, difference in background and foreground due to the imbalance of the light rays, shade, spots etc.) that can be put aside while detecting the lesion grand. But the question comes how computers can ignore these noise? To begin, let vector $x_i^j \subset R^N$ denote the set of points for different objects within an image. In addition, i and j refers number of objects and pixel values within each objects those are not null and have distance less or equal to some predefined threshold.

$$x_i^j = \{1, \ldots, j\} \tag{1}$$

Most of the previous work was performed with the help of image cropping using tools like Adobe Photoshop that can invalidate useful pixel values while conversion from one format to another. Thus, we avoid using those methods to resize the image. In general, it is evident from the mammogram images that the largest area covered must preserve the breast gland. So, our aim is to look for the available objects within the image first, which makes the problem into an object detection problem. If we look into Fig.1 (b), it is noticeable that the yellow marks are showing different smaller objects within the image area. White letters within the textual part are treated individually while leaving them out in the preprocessing step, but it is shown in a group for the simplicity. Additionally, the images are considered to be black and white for this experiment.

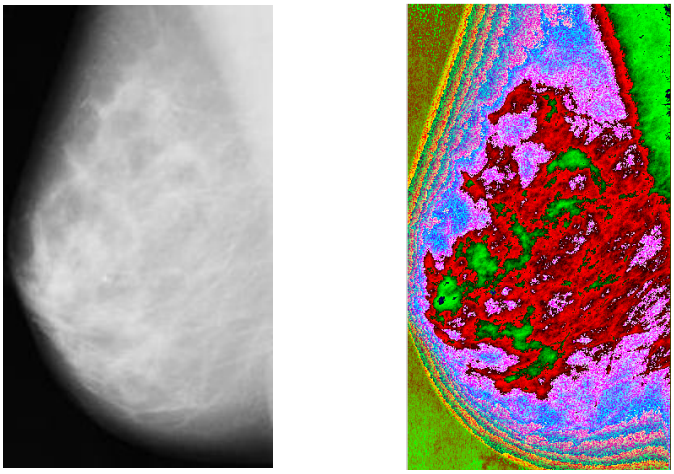
Mammograms are taken from different angles to ensure that no area from the tissue is being abandoned, later the X-ray film is examined by specialists for further verification. Now, our aim is to simulate the specialists verification knowledge through CAD (Computer Aided Diagnosis) systems.

Before going in detail, first we should know what connected sets determine? Mathematically, A region $R \subset S$ is defined as connected under $c(s)$ if for all $s, r \in R$, there exists a sequence of M pixels, $s_1, \ldots, s_M$ such that $s_1 \in c(s), s_2 \in c(s_1), \ldots, s_M \in c(s_{M-1}), r \in c(s_M)$. Which mean there must be a connected path from s to r. Now, as we are concerned about grayscale image which has values between 0 to 255 in each pixel, we can start from any pixel which has value 0 (black pixel treated as background) and start to grow each connected set. Additionally, we chose the neighborhood system to select the nearest pixels for region growing purpose. Usually, there are two neighborhood systems, i.e., 4-point neighborhood and 8-point neighborhood method denoted as N_4 and N_8 respectively.[9]

	N	
W	S_0	E
	S	

NW	N	NE
W	S_1	E
SW	S	SE

Fig. 2. Demonstration of neighborhoods of a pixel. S_0 is at the center when 4-point neighborhood is considered. S_1 is the central pixel surrounded by eight neighboring pixels. The four neighborhood of $N_4(x, y)$ of pixel (x, y) is surrounded with pixels $(x-1, y), (x+1, y), (x, y+1), (x, y-1)$. Similarly, $N_8(x, y)$ of pixel (x, y) includes with neighboring pixels of $(x-1, y), (x-1, y+1), (x-1, y-1), (x+1, y), (x+1, y+1), (x+1, y-1), (x, y+1), (x, y-1)$. N,E,W,S and other four derived values denotes the side of the center pixel.



(a) (b)

Fig. 3. (a) Cropped image without removing any pixel information from the region of interest. (b) Colormap applied for better visualization of abnormal tissue image in the ROI.

8-points neighborhood pixel method is suitable for use in RGB images where objects have different combination of colors and the pixels values vary too much in general. On the contrary, gray scale images can have only values between 0 to 255 and most dominant colors are black and white, thus the difference between white and black sectors are very much prevalent to distinguish. As a result 4-point neighborhood method works well to determine the connected objects with

ease, but there is some relation with the threshold value which has to be checked each time the region is grown [10] from the source S_0 (Fig.2). Now we can formulate that the connected set, $c(s) = \{r \in \delta_s : |S_r - P_s| < Th\}$, where, r is the pixel position to be tested, δ_s is the set of neighboring pixels, $S_r - P_s$ refers to the distance between pixel r and each neighbors, and Th is the threshold value that can not be crossed. However, two or more objects may overlap in some scenarios, which can be difficult to differentiate for gray scale images compared to RGB ones. Connected component algorithm can well identify the presence of the overlapping objects by computing the intensities of the color level, but it is not easy in our case as the black color is considered to be the background and the only important values for pixels are those which are non zero values.

III. Description of Mammogram Dataset and Feature Selection

A. Mammogram Data Set

We performed our experiment using a dataset prepared by the Department of Physics at the Royal Marsden Hospital of London, UK and it is popular by its name Mini-Mias mammogram repository [11]. The data contains 322 images of 161 patients, each one having left and right mammograms. All the images size were adjusted by padding zeroes (black background) to make 1024×1024 pixels. Each image is accompanied with some predefined information, i.e., center of the image, radius, coordinate start location etc. It was also mentioned whether calcification is present centrally or in a scattered manner. Normal and benign classes are mentioned along with other six abnormality types. The origin of the abnormality is provided with their axis values and the radius of the calcification spread. For example, the calcification origin is at (500,200) and the radius is given as 312, which means the abnormality is approximately spread around the point (500,200) with the radius of 312 pixel units.

B. Feature Extraction and Processing

For automated classification, we used total ten features of both shape and texture category and saved in a vector. For example, if an image, A has radius, r and center, c as features, then the feature vector of image A is $V(A) =< r, c >$. The features extracted from the processed mammograms are variance, standard deviation, median, mode, average gray scale value, average contrast, smoothness, skewness, uniformity, and entropy. The first four features are statistical measures and the latter six are texture features. Average gray-scale value is computed from the first moment which gives the idea about the distribution of the gray-level pixels. Average contrast is measured by the standard deviation σ, which is the square root of the second moment. Smoothness [12] provides the approximating function to capture significant patterns from the pixel data by removing the noise. We used the below equation for smoothing purpose.

$$R = 1 - \frac{1}{1 + \sigma^2} \tag{2}$$

Skewness can be positive, negative or both to quantify the bias of the random variable (here pixel) from the ideal viewpoint.[13] It is measured using the third moment. Equation 3 shows the correspondence among third central moment (μ_3), mean (μ), standard deviation (σ), and the expectation operator (E).

$$\gamma_1 = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{\mu_3}{\sigma^3} = \frac{E[(X-\mu)^3]}{(E[(X-\mu)^2])^{3/2}} \tag{3}$$

Uniformity of the mammogram measures the sum of the squared relative frequencies p_i of the gray-scale values and it is maximum when image remain constant. Entropy E measures the predictability using Equation 4.

$$E = p_i \log_2 p_i \tag{4}$$

We masked [14] the original mammograms after processing them using connected component algorithm in order to compute these features for further classification. As discussed in the next section, several classifiers were run to identify the classification accuracy parameters using the feature vector.

IV. Evaluation

Statistical and texture features were extracted from the Mini-Mias dataset using Matlab programs. Microcalcification detection was performed using decision tree, bagging, decision table, and locally weighted learning classifier available in the Weka Machine Learning Algorithms Toolkit [15]. This experiment was evaluated using 10-fold cross validation for the classifiers to split the input feature vector into both training and testing set.

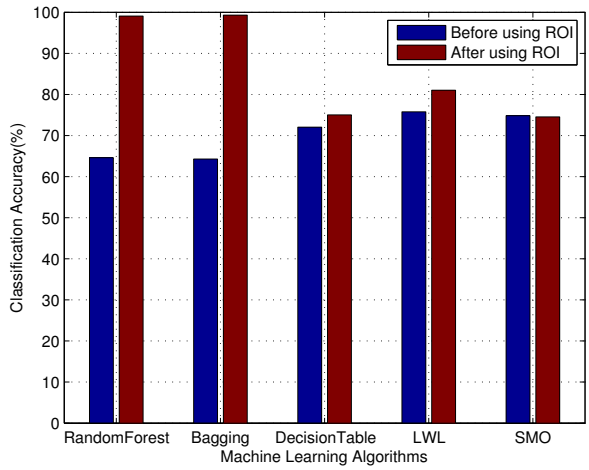


Fig. 4. Classification accuracy and type of the algorithms used with the data vectors extracted from the processed mammograms. Random Forest and Bagging provides similar accuracy both before and after the usage of labeling.

A. Results and Discussion

Classifiers were tested with several configuration using the dataset containing 322 samples from the MiniMias. Experimental result shows that Bagging [16] and Random Forest [17] classifier provides the maximum accuracy (Fig. 4) among the chosen five algorithms. Random Forest is of type decision tree classifier, which uses multiple decision trees at the training time and the mode class among them becomes the desired output class (99.06% accuracy). Random forest of 10 trees, each constructed while considering 5 random features and the out of bag error was 0.3571. Bagging was used with 100% bag size, J48 pruned tree, and seed value of 1.00. It provides 99.37%

TABLE I. TP (True Positive), FP (False Positive) and ROC (Receiver Operating Curve) values shows strong class label detection. (Sample shown for Bagging classifier)

TP Rate	FP Rate	ROC Area	Class
1.00	0.00	1.00	CIRC
1.00	0.017	1.00	NORM
0.929	0.00	1.00	MISC
1.00	0.00	1.00	ASYM
1.00	0.00	1.00	ARCH
1.00	0.00	1.00	SPIC
0.96	0.00	1.00	CALC
0.994	0.011	1.00	Weighted Avg.

accuracy for the classification. Particularly, Bagging classifier reduces the variance among the output class using sampling with replacement. Locally weighted learning (LWL) algorithm used decision stump as classifier, Linear nearest neighbor as neighbor search algorithm, and the value for KNN weighted function as 1 to select all the neighboring values (80.43% accuracy). Decision table majority classifier used leave-one-out cross validation method, forward best first search, and RMSE (Root mean squared error) as evaluation measure. SMO (Sequential Minimal Optimization) used Polykernel as kernel function along with 0.001 as tolerance parameter and provided 74.53% accuracy.

Table I shows the ROC (Receiver operating characteristics) [18], TP (True positive), and FP (False positive) values for each of the classes for Bagging classifier. The weighted ROC value is 1.00, which means the maximum accurate performance is achieved with this classification. TP rate and FP rate helps perceive the misclassification error rate of the classifier. For example, let there are two classes of A, B and TP rate of 0.8 and 0.9 respectively (FP rate is 0.2 and 0.1 as it is the complement of TP rate). So, it means 80% of class A data is picked as class A and the rest 20% is misclassified as class B. Thus, higher value of TP rate for a particular class defines higher classification accuracy. (TP and FP rate ranges between 0 and 1)

On the other hand, we can notice significant improvement of classification accuracy for both Random Forest and Bagging classifier when compared with non-connected component labelled mammograms. One of the major reason for this improvement is normalizing the frequency distribution of the pixels, which is ensured by the labeling operation. Additionally, all the four classifier provides better result than before, except SMO which shows a bit less. SMO uses Support vector machine (SVM), which is best known for classification of noisy data of two class. In our experiment, six classes were present and there were some overlapping of feaures of different classes that can make the classification task confusing for SVM.

Table I shows seven different class labels provided in the dataset, which can further be narrowed down to two class problem (Normal and Abnormal) [19]. These six other classes, except normal category, defines different type of malignant tissue growth within breast gland.

V. Conclusions

Using regression model and decision tree can enhance the classification accuracy in great detail, compared to other algorithms for this specific mammogram classification research. However, our major goal was to augment the mammography image processing to collect the desired region from them, which in the end helps exposing the abnormality or severity of breast cancer via computer aided applications. Thus, we had to go through the classification process as well to prove that the detection accuracy improves if our ROI selection method is followed. This experiment was performed using 10-fold cross validation with labeled dataset.We will use larger dataset and narrow down the seven class problem into two class classification problem using layered architecture in our future research. The first layer will separate the seven classes. The second class will leave *Normal* category intact and run the classification on the rest six classes. We will use our method on separate testing and training data as well, which can surely help diagnose early stage cancer identification.

References

[1] Solka, Jeffrey L., et al. *The detection of micro-calcifications in mammographic images using high dimensional features.* Computer-Based Medical Systems, 1994., Proceedings 1994 IEEE Seventh Symposium on. IEEE, 1994.

[2] Hung, Ming S., Murali Shanker, and Michael Y. Hu. *Estimating breast cancer risks using neural networks.* Journal of the Operational Research Society 53.2 (2002): 222-231.

[3] Giger, Maryellen L. *Computer-aided diagnosis of breast lesions in medical images.* Computing in Science & Engineering 2.5 (2000): 39-45.

[4] Haddadnia, Javad, Maryam Hashemian, and Kazem Hassanpour. *Diagnosis of Breast Cancer using a Combination of Genetic Algorithm and Artificial Neural Network in Medical Infrared Thermal Imaging.* Iranian Journal of Medical Physics 9.4 (2013): 265-274.

[5] Hassanien, Aboul Ella, et al. *Breast Cancer Detection and Classification Using Support Vector Machines and Pulse Coupled Neural Network.* Proceedings of the Third International Conference on Intelligent Human Computer Interaction (IHCI 2011), Prague, Czech Republic, August, 2011. Springer Berlin Heidelberg, 2013.

[6] Huang, Yan-Hao, et al. *Diagnosis of Solid Breast Tumors Using Vessel Analysis in Three-Dimensional Power Doppler Ultrasound Images.* Journal of Digital Imaging (2013): 1-9.

[7] Lee, Hahn-Ming, et al. *An efficient fuzzy classifier with feature selection based on fuzzy entropy.* Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 31.3 (2001): 426-432.

[8] Jeffreys, M., et al. *Initial experiences of using an automated volumetric measure of breast density: the standard mammogram form.* (2014).

[9] He, Li-Feng, Yu-Yan Chao, and Kenji Suzuki. *An algorithm for connected-component labeling, hole labeling and Euler number computing.* Journal of Computer Science and Technology 28.3 (2013): 468-478.

[10] Lu, Xiaoqi, et al. *The study and application of the improved region growing algorithm for liver segmentation.* Optik-International Journal for Light and Electron Optics 125.9 (2014): 2142-2147.

[11] Suckling, John, et al. *The mammographic image analysis society digital mammogram database.*, (1994).

[12] Tian, Jie, et al. *Medical image processing and analysis.* Molecular Imaging. Springer Berlin Heidelberg, 2013. 415-469.

[13] Schlüter, Steffen, et al. *Image processing of multiphase images obtained via Xray microtomography: A review.* Water Resources Research 50.4 (2014): 3615-3639.

[14] Kim, Young Jae, et al. *Body fat assessment method using CT images with separation mask algorithm.* Journal of digital imaging 26.2 (2013): 155-162.

[15] Witten, Ian H., and Eibe Frank. *Data Mining: Practical machine learning tools and techniques.* Morgan Kaufmann, 2005.

[16] Dong, Liuhuan, Yuan Yuan, and Yudong Cai. *Using Bagging classifier to predict protein domain structural class.* Journal of biomolecular structure & dynamics 24.3 (2006): 239.

[17] Gray, Katherine R., et al. *Random forest-based similarity measures for multi-modal classification of Alzheimer's disease.* NeuroImage 65 (2013): 167-175.

[18] Hanley, James A., and Barbara J. McNeil. *The meaning and use of the area under a receiver operating characteristic (ROC) curve.* Radiology 143.1 (1982): 29-36.

[19] Russo, Jose, et al. *Methodological Approach to Tissue Microarray for Studying the Normal and Cancerous Human Breast.* Techniques and Methodological Approaches in Breast Cancer Research. Springer New York, 2014. 103-118.

Motor Imagery EEG Signal Classification Scheme Based on Entropy of Intrinsic Mode Function

Md. Toky Foysal Talukdar, Shahnewaz Karim Sakib, Celia Shahnaz, and Shaikh Anowarul Fattah[1]
Department of Electrical and Electronic Engineering (EEE)
Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh
Email: [1]fattah@eee.buet.ac.bd

***Abstract*—In this paper, an efficient scheme is proposed for classification of electroencephalogram (EEG) data in to different motor imagery (MI) tasks based on empirical mode decomposition (EMD). The EEG data recorded from each channel is first decomposed into a set of intrinsic mode functions (IMFs) by using the EMD analysis. In view of extracting discriminative information from the EEG signals corresponding to different MI tasks, we propose to utilize the entropy of band-limited IMF. Instead of considering all IMFs or first few IMFs, in the proposed method only the first IMF is chosen because of its low variance. In order to reduce the dimension of the feature vector consisting of entropy values from all channels, principal component analysis is performed. For the purpose of classification, train and test datasets are formed as per leave one out cross validation scheme and then linear discriminant analysis (LDA) is carried out. Simulation is performed on publicly available MI dataset IVa of BCI Competition-III to classify the MI data in to two classes, namely right hand and right foot MI tasks. It is observed that the proposed classification scheme not only offers significant reduction in feature dimensionality but also provides satisfactory classification accuracy.**

***Index Terms*—Brain computer interface, classification, electroencephalogram, emperical mode decomposition, feature extraction, linear discriminant analysis, motor imagery.**

I. Introduction

Electroencephalography (EEG) has become the most popular and widely studied non-invasive brain-computer interface (BCI) due to its excellent temporal resolution, usability and low setup costs. EEG-based BCIs use electrical brain activity in assisting, augmenting, and repairing human cognitive functions. The greatest challenge in BCI systems is to correctly and efficiently identify different EEG signals of different motor imagery (MI) tasks. Recently, MI EEG signal classification has received much attention by several researchers [1]-[4]. With the ease of availability of high quality low cost EEG leads, one major concern now-a-days is the huge amount of data to be handled in case of multi-lead EEG signal analysis. Different spatial filters can be used for EEG channel reduction, where a major drawback is the manual selection of regularized parameter. For example, sparse spatial filter optimization method, introduced in [1], can offer EEG channel reduction at the expense of manual parameter selection. Similarly, the regularized common spatial pattern (CSP) algorithm, reported in [2], involves two regularization parameters which are not selected optimally. Four different approaches for MI task classification using EEG signals are proposed in [3] based on CSP which itself is very sensitive to noise and often over fits with small training sets. The regularization and aggregation techniques introduced in [4] solve the parameter determination problem by cross-validation method but only in a small sample setting. In [5], least squares support vector machine (LS-SVM) based clustering technique is proposed for feature extraction but the parameters are not selected optimally. Although a high level of cross validation accuracy is reported in this method, a prior knowledge on the class labels is assumed to be available during cross-correlation based feature extraction. Considering the non-stationary nature of EEG signal, empirical mode decomposition (EMD) is also applied for EEG analysis in different applications [6]-[8]. Recently in [8], the potential of EMD analysis in MI data handling is discussed with a note that relevant features must be carefully selected for getting better accuracy.

The objective of this paper is to develop a robust classification scheme based on EMD analysis of the EEG signal corresponding to different MI tasks. The EMD operation is first performed on EEG data of each channel to obtain a set of intrinsic mode functions (IMFs). Next, the entropy of band pass filtered first IMF of each channel data is computed and a feature vector is formed. Because of a large number of channels, the feature dimension becomes large and thus PCA based dimension reduction is performed. Finally, linear discriminant analysis is carried out for classifying the EEG data into two MI tasks. Performance of the proposed method is validated by using a publicly available MI EEG dataset IVa [9]-[10] of BCI Competition III.

II. Data Acquisition

In this paper, most widely used publicly available EEG dataset IVa of BCI Competition III is considered for the purpose of classification [9]-[10]. This dataset was recorded from five healthy subjects labeled aa, al, av, aw, and ay. In this dataset, EEG signals from 118 electrodes were collected based on the international 10/20 system. There were 280 trials for each subject, namely 140 trials for each task per subject. During the each trial, each subject was required to perform either of the two (right hand and right foot for IVa) MI tasks for 3.5 seconds. The down-sampled data at 100 Hz is generally used although the original sampling rate was 1000 Hz. Duration of the recording of the MI task is 3.5 seconds.

III. Proposed Method

A. Empirical Mode Decomposition (EMD)

The EMD is a time-frequency analysis tool for nonlinear and nonstationary signals [6]-[7]. It is a fully data driven technique

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

with which any complicated data set can be decomposed into a finite and often small number of intrinsic mode functions (IMF). A function is considered to be an IMF if it satisfies two conditions; first, in the whole data set, the number of local extrema and that of zero crossings must be equal to each other or different by at most one and second, at any point, the mean value of the envelope defined by the local maxima and that defined by the local minima should be zero. The systematic approach to decompose the data $x[n]$ into IMFs, known as the "sifting" process, can be briefly summarized as follows [7]:

- Identify all local extrema of the signal. Next, interpolate maxima to define an upper envelope E_U and interpolate minima to define a lower envelope E_L. Set $h_1[n] = x[n] - m_1$; $m_1 = \{E_L[n] + E_U[n]\}/2$.
- Iterate on h_1 until a stopping criterion is satisfied. Upon stopping one has an IMF C_1. The extraction of subsequent IMFs works by iterating on $r_1 = x - C_1$ and so on until an outer loop stopping criterion is satisfied. The result is $x[n] = C_1 + C_2 + \ldots\ldots + C_N + r_N$.
- Let after w cycles of operation, $h_w[n] = h_{w-1}[n] - m_w$ becomes an IMF. Repeating the above process for L times, L no. of IMFs is obtained along with the final residue $r_L[t]$. A popular stopping criteria for the sifting process is to have the value of standard difference (SD) within a threshold as:

$$SD = \sum_{n=1}^{N} \frac{\left|h_{w-1}[n] - h_w[n]\right|^2}{h_w[n]^2}, \qquad (1)$$

where w and $w-1$ are index terms indicating two consecutive sifting processes. Thus the decomposition process is stopped since $r_L[n]$ becomes a monotonic function from which no more IMF can be extracted. To this end, for L level of decomposition, $x[n]$ can be reconstructed as

$$x[n] = \sum_{k=1}^{L} c_k[n] + r_L[n] \qquad (2)$$

Here $c_1, c_2, \ldots, c_L$ represent the IMFs. In this work, EEG data of each channel is decomposed into 8 IMFs.

For the purpose of feature extraction, it would be very difficult to consider all IMFs, which will not only increase the computational burden but also reduce the feature consistency. In many cases, first few IMFs are chosen. However, in this paper, it is decided to consider only one of the eight IMFs. For the purpose of analysis, first, the variance of the IMF data pattern is measured. In Fig. 1(a), the variance of the first and second IMFs are plotted for different trials considering a fixed channel (chammel no. 1). In a similar way in Fig. 1(b), the variance of the first and second IMFs are plotted for different channels considering a fixed trial (trial no. 1). In Fig. 1(c), variation of energy over different trials for a fixed channel is shown for the first three IMFs. It can be clearly observed that the variance of the first IMF is is lower in most of the cases in comparison to that of the second IMF and its energy is also comparatively lower. Hence, in the proposed method only the first IMF is considered for feature extraction.

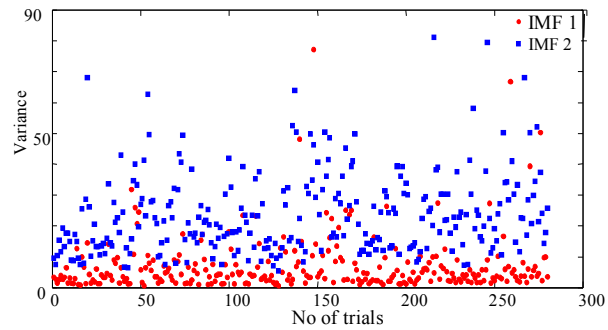


Fig. 1(a). Variance of the first and second IMFs over different trials for a fixed channel.

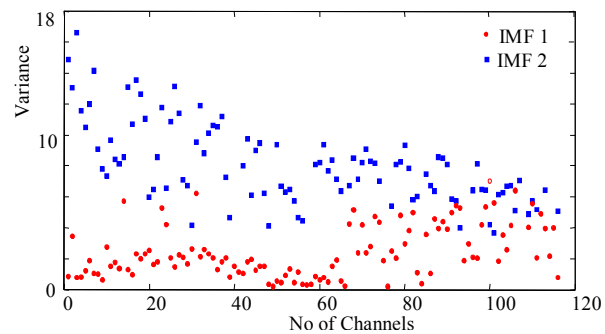


Fig. 1(b). Variance of the first and second IMFs over different channels for a fixed trial.

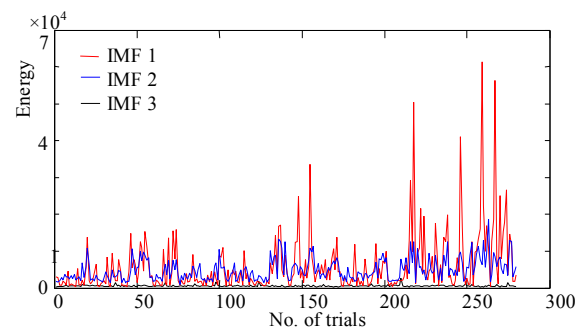


Fig 1(c). Variation of energy over different trials for fixed channel for first three IMFs.

Next the frequency content of the first IMF is investigated. In low frequency region, it is very difficult to discriminate the data for two different MI tasks. Alternatively, it is expected to find more distinguishable characteristics in the high frequency region for the first IMF. It is to be mentioned that such a scenario may not be true for other IMFs. Hence, the first IMF is band limited using a high pass filter. The frequency spectra of a sample first IMF is shown without and with band limitation in Figs. 2(a) and 2(b), respectively. It is observed that the spectra exhibit more distinguishable information in the high frequency range as expected.

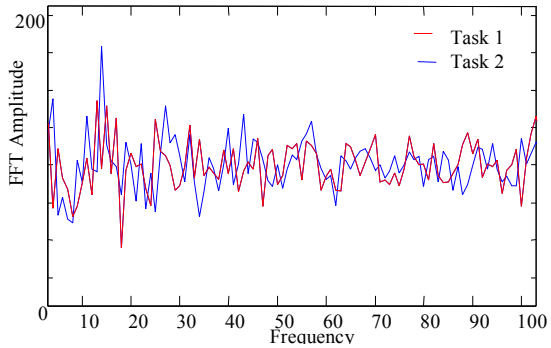


Fig. 2(a). Frequency spectra corresponding to the first IMF of two tasks without using band limiting operation.

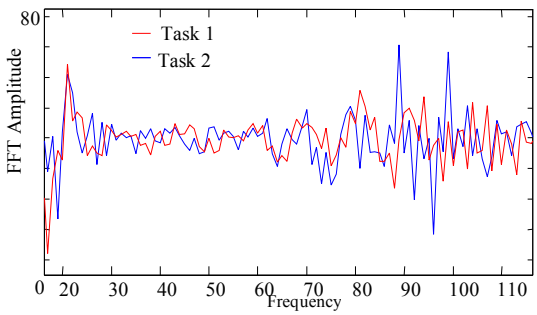


Fig. 2(b). Frequency spectra corresponding to the first IMF of two tasks after band limiting operation.

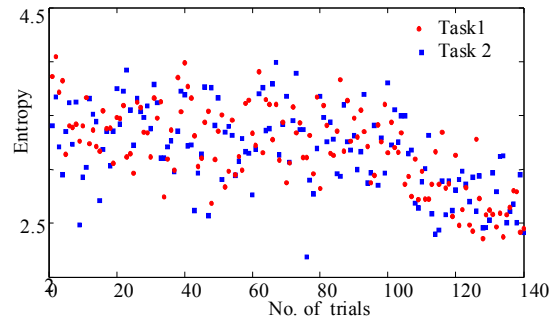


Fig 3(a). Variation of entropy values over different trials for fixed channel (channel 1).

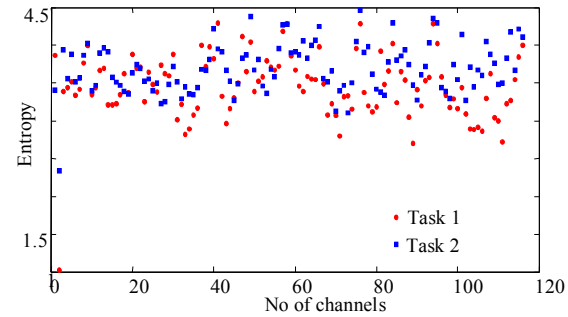


Fig 3(b). Variation of entropy values over different channels for fixed trial (trial 1).

B. Proposed Feature: Entropy of Band-limited IMF

Entropy of a discrete random variable X with (N+1) number of possible values $\{x_0, x_1, x_2, \ldots, x_N\}$ is defined as

$$H(X) = E(I(X)), \tag{3}$$

where $E(.)$ denotes the expectation operator and $I(X)$ represents the information content. For a particular value x_i of X, the information content can be expressed as

$$I(X) = -\log_2(p(x_i)), \tag{4}$$

where $p(x_i)$ denotes the probability of occurrence of x_i. Using (4), the entropy in (3) can be rewritten as

$$H(X) = -\sum_{i=0}^{N} p(x_i) \times \log_2(p(x_i)). \tag{5}$$

Here $p(x_i)=n_i/N$ and where n_i is the number of occurrence of coefficients in a bin in proximity of x_i value among the N number of values, i.e. $\sum_i n_i=N$. In the proposed method, entropy of the band limited first IMF is considered as feature. In Fig. 3(a), variation of entropy values over different trials for a fixed channel (channel no. 1) is shown. Next in Fig. 3(b), variation of entropy values over different channels for a fixed trial (trial no. 1) is presented. From these figures, it is difficult to comment on the quality of the extracted feature. In order to improve the feature quality, instead of considering the entire duration of an IMF in entropy calculation, we propose to compute entropy in a small duration. The band limited first IMF is windowed into 10 overlapping frames with 50% overlap and the entropy is calculated in each window. An average value of entropy over then frames is considered as a desired feature. It is observed that instead of computing entropy once on the whole duration, computing it in all small windows and then considering the average value significantly enhances the feature quality. Entropy values obtained from all the channels are used to form the desired feature vector.

C. Dimensionality Reduction Using PCA

As in the feature vector, entropy values from all channel are considered, the feature dimension becomes large. Next, principal component analysis (PCA) is carried on to reduce the feature dimension. The PCA is a very well-known and efficient orthogonal linear transformation [11]. It transforms the original p-dimensional feature vector into the L-dimensional linear subspace spanned by the leading eigenvectors of the covariance matrix of feature vector in each cluster ($L<p$). PCA is theoretically the optimum transform for given data in the least square sense. For a data matrix $\mathbf{X}^T$ with zero empirical mean, PCA transformation is given by

$$\mathbf{Y}^T=\mathbf{X}^T\mathbf{W}=\mathbf{V}\mathbf{\Sigma}^T \tag{6}$$

where the matrix $\mathbf{\Sigma}$ is a $m\times n$ diagonal matrix with nonnegative real numbers on the diagonal and $\mathbf{W}\mathbf{\Sigma}\mathbf{V}^T$ is the singular value decomposition of $\mathbf{X}$. Implementation of PCA on the derived feature space not only reduces the feature dimension it also helps in obtaining better feature characteristics. In order to demonstrate the effectiveness of using the PCA, feature values corresponding to two tasks obtained after performing the PCA are shown in Fig. 4. It can be observed that the proposed features after performing PCA exhibit better separability between the two tasks.

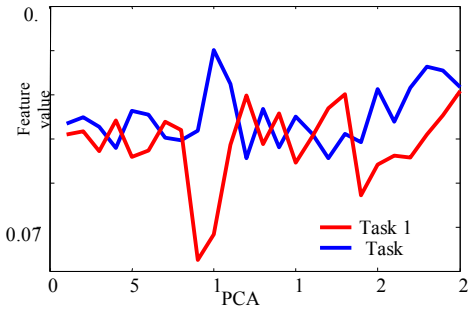


Fig 4. Comparison of feature values obtained after PCA.

D. Classification

In the proposed method, linear discriminant analysis (LDA) is used to classify the testing EEG signals among two different classes (tasks). In LDA, the objective function is based on determining a projection direction **w** to maximize the Fisher's discriminant defined as

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_b \mathbf{w}}{\mathbf{w}^T \mathbf{S}_w \mathbf{w}} \tag{7}$$

where $\mathbf{S}_w$ and $\mathbf{S}_b$ are within- and between-class scatter matrices, respectively [12]. Classification performance is measured using leave-one-out cross validation where the test is performed for all 280 datasets (both training and testing) for each subject.

IV. Simulation Results

For the purpose of testing with MI EEg data, sample values are first imported into EEGLAB GUI [13]. A reference electrode TP10 is chosen during data import. Without a reference electrode, it is possible that minimum level of 40 dB noise is accumulated in channel information [13], [14]. Typical references in the EEG recording are TP10, Cz or channel on the nose tip. EEGLAB is also used to separate frames of two classes from a jumbled dataset. In the band limiting operation on the first IMF, high frequency band is selected as 70-120 Hz. After PCA the feature dimension is reduced to 20. Finally, features obtained after PCA is used in the LDA to classify the two tasks via leave one out cross validation. Accuracy is measured based on the correctly identified number of trials among all 280 trials. In Table I, comparative results of the classification accuracy are presented for dataset IVa for the five subjects as well as the overall mean accuracy [1], [14]. It is clearly observed that the performance of the proposed method is significantly better in cases of subject av and aw, which is very promising in subject specific applications. However, the performance in case of subject al is relatively poor and needs further investigation. Overall the proposed methodology provides competitive average classification accuracy compared to all the reported algorithms.

V. Conclusion

A simple but efficient MI classification scheme is developed based on EMD analysis. It is found that use of only the first IMF can provide satisfactory classification performance when entropy of the band-limited IMF is considered as the feature. Because of considering only the entropy value of the selected IMF, the feature dimension is drastically reduced. Moreover, use of PCA provides further control on feature dimension reduction as well as significant improvement in the derived feature quality. Because of the quality of the extracted features, even simple LDA based classifier can provide satisfactory classification performance. It is observed that the proposed entropy based method offers a very satisfactory classification performance for all subjects in comparison to some of the existing methods.

TABLE I. PERFORMANCE COMPARISON OF DIFFERENT METHODS

Method	**Classification Accuracy (%)**					
	aa	**al**	**av**	**aw**	**ay**	***Average***
Proposed	72.64	75.17	73.36	88.51	83.59	78.65
CSP [14]	66.96	89.29	52.25	47.77	52.38	61.79
CSSP [14]	79.46	92.86	52.55	91.52	51.59	73.60
BSSFO [14]	79.46	94.64	57.65	91.96	53.57	75.46
FBCSP [14]	69.64	80.36	47.96	55.36	48.41	60.35
DCSP [14]	69.64	82.14	54.08	50.89	48.41	61.03
SSRCSP [1]	70.54	96.43	53.57	71.88	75.39	73.56
TRCSP [1]	71.43	96.43	63.27	71.88	86.9	77.98
WTRCSP [1]	69.64	98.21	54.59	71.88	85.32	75.93
SRCSP[1]	72.32	96.43	60.2	77.68	86.5	78.63

Acknowledgement

The authors would like to express their gratitude to the Department of EEE of BUET for providing constant support.

References

[1] X. Yong, R. K. Ward, and G. E. Birch, "Sparse spatial filter optimization for EEG channel reduction in brain-computer interface," in Proc. ICASSP 2008, pp. 417-420.

[2] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularized common spatial patterns with generic learning for EEG signal classification," in Proc. IEEE Int. Conf. EMBS, Sep. 2-6, 2009.

[3] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms," IEEE Trans. Biomed. Eng., vol. 58, no. 2, pp. 355-362, Feb. 2011.

[4] H. Lu, H. L. Eng, C. Guan, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularized common spatial patterns with aggregation for EEG classification in small-sample setting," IEEE Trans. Biomed. Eng. Vol. 57, no. 12, pp. 2936-3945, Dec. 2010.

[5] S. Siuly and Y. Li, "Improving the separability of Motor Imagery EEG signals using a cross correlation-based least square support vector machine for brain-computer interface," IEEETrans. Neural Syst. Rehabil. Eng.,vol.20, no. 4, pp. 526–538,Jul. 2012.

[6] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, E. H. Shih, Q. Zheng, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis," Proc. Royal Soc. London, A 454, pp. 903-995, 1998.

[7] P. Flandrin, G. Rilling, and P. Goncalves, "Empirical mode decomposition as a filter bank," IEEE Signal Process. Lett, vol. 11, pp. 112–114, 2004.

[8] A Gupta and R. K. Agarwal, "Relevant feature selection from EEG signal for mental task classification," Advances in Knowledge Discovery and Data Mining, pp 431-442, 2012

[9] BCI Competition III [Online]. Available: http://www.bbci.de/ competition/iii B.

[10] K. R. Blankertz *et. al.*, "The BCI competition III: Validating alternative approaches to actual BCI problems, "IEEE Trans. Neural Rehabil. Eng., vol. 14, no. 2, pp. 153-159, 2006.

[11] I.T. Jolloffe, "Principal component analysis," SpringerVerlag, Berlin, 1986.

[12] R.O. Duda and P.E. Hart, *Pattern Classification*, John Wiley, 2001.

[13] A. Delorme, T. Fernsler, H. Serby, and S. Makeig, "EEGLAB Tutorial," vol. 601b, pp.25, April 12, 2006. Copyright University of San Diego California.

[14] Heung-Il Suk, Seong-Whan Lee, "A novel bayesian framework for discriminative feature extraction in brain computer interfaces", IEEE Trans. On Pattern Analysis and Machine Intelligence, vol. 35, no. 2, Feb. 2013.

Efficient Event Related Oscillatory Pattern Classification for EEG Based BCI Utilizing Spatial Brain Dynamics

Simanto Saha[1] and Khawza I. Ahmed[2]
Department of Electrical and Electronic Engineering, United International University (UIU)
Road 8/A, Dhanmondi, Dhaka-1209, Bangladesh
[1]simantosaha@gmail.com [2]khawza@eee.uiu.ac.bd

***Abstract*—This paper features the spatial characteristics of the brain towards brain-computer interface (BCI) research. A study on motor imagery (MI) based BCI has been carried out and important implications are identified. Common Spatial Pattern (CSP) is applied to the EEG signals before proceeding to the classification. The primary focus of this research is to utilize the spatial dynamics of the brain to develop BCI with reduced number of electrodes which contribute to the motor imagery tasks with optimal impact. It is observed that computational cost can be reduced drastically by selecting channels from specific regions of interests (ROIs) of the brain without compromising the classification accuracy making BCI efficient. Here, we have reported the best classification accuracies 72.5% and 97.1% which are achieved for two subjects ('av' and 'ay', respectively, in the dataset IVa in the BCI competition III) using less number of electrodes.**

***Index Terms*—Brain Computer Interface (BCI), Motor Imagery (MI), Electroencephalogram (EEG), Common Spatial Pattern (CSP), Brain Dynamics.**

I. Introduction

The research on brain-computer interface (BCI) has got significant pace in its way lately. The objective of BCI is to develop a sophisticated system that allows people to communicate with computer without stimulating any muscles. The BCI system enables physically disabled people, e.g., having Tetraplegia, to communicate with computer or to control artificial limb. Also, there are other applications of BCI such as lie detection, brain fingerprinting, mode assessment and mind control, etc. There are few techniques of signal acquisition such as functional magnetic resonance imaging (fMRI), magnetoencephalogram (MEG), electroencephalogram (EEG), electrocorticogram (ECoG), etc., which may be used in BCI. The fMRI and the MEG are brain imaging techniques which give good spatial resolution but poor temporal resolution. The EEG is the recording of the electrical activity of the brain recorded from the scalp using noninvasive electrode montages. The EEG gives good spatial and temporal features. The ECoG is also a process of acquiring signal from the brain by placing the electrodes directly on the exposed brain surface. However, ECoG needs invasive surgery to place electrode montages in exposed brain surface.

The EEG based BCI is very popular due to its cost efficiency and non-invasive signal acquisition process. In this paper, we have used EEG signal as response from the brain for motor imagery (MI) tasks. In the development of BCI, motor imagery tasks (MI) are proved to be very effective. According to the Jeannerod (1994, 1995), brain's response to the conscious motor imagery (MI) task is functionally equivalent to the brain's response to the preparation of the unconscious motor execution (ME) [1]. In this paper, we have studied the classification of event related EEG pattern for motor imagery (MI) tasks. The BCI research trend based on motor imagery (MI) task is reported in [2]-[5]. The EEG signal needs to be properly preprocessed to implement BCI system. It is quite challenging as EEG signal is highly nonlinear in nature. We have applied the common spatial pattern (CSP) filter as a preprocessing technique. The CSP is very popular method in BCI field as it is used in many of the previous works [2]-[5]. We have validated our proposed algorithm using both CSP and Regularized Common Spatial Pattern (RCSP). The RCSP is more robust as it can reduce the effect of outliers [5]. The outliers are the patterns deviated from the general pattern of the data. In [2], the RCSP with aggregation is proposed for small sample settings. A few modified CSP algorithms are also described in [3]. It is observed that both CSP and RCSP algorithms have been proven to be very effective in increasing the classification accuracy.

The goal of the proposed experiment is for finding a way to employ less number of electrodes in BCI system without compromising accuracy so that the computational cost can be reduced. We have demonstrated a comparative study concerning number of electrodes as well as the computational cost in BCI system. Our experiment attributes to computational latencies which is distinctive from the orthodox BCI literatures. In our experiment, we have proposed to consider the spatial brain dynamics effectively. The experiments in previous literatures have emphasized on the classification in different settings, e.g., in the presence of outliers [5], in case of small training trials set [2] or comparison between the potentiality of multiple preprocessing methods [3], etc. Although our experiment aims to classification problem for BCI, we have emphasized a great deal on the spatial dynamics of the brain to minimize the number of electrodes as well as the computational cost. An approach to important channel (actual EEG montage) selection is proposed in [4] based on CSP features. In [4], the CSP is applied to the EEG data and optimal channels (after CSP is applied) for different classes are used to correlate with the actual EEG montages. Thus, the actual EEG electrodes are selected for BCI purpose which have more discriminative features for two classes. In contrast, we have proposed to simply use the spatial brain dynamics to reduce the number of electrodes used rather to consider statistical features provided by CSP as proposed in [4].

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

II. Channel Selection Featuring Brain Dynamics

The BCI research tends to be collaborative. Engagement of different fields such as Neuroscience, Mathematics, Statistics, and Engineering, etc., is necessary for BCI development. In this paper, we have utilized the spatial dynamics (i.e., knowledge of Neuroscience) of the brain to sort out electrodes which have optimal impact for the specific task. Though the CSP gives best channels for different classes irrespective of actual EEG montages, the spatial dynamics can give idea about the suitable electrode montages. The assumption behind this study is that the channels placed on the specific brain regions where the impact of the event related EEGs are mostly significant, will help the algorithm to give optimal performance with less computation.

There is significant number of works studying spatial areas activated during motor imagery (MI) tasks. In [1], the general concepts about motor imagery task (MI) are explained elaborately. An investigation based on fMRI has been carried out in [6] and specific regions are reported during motor imagery (MI) task. Also, source activation in different brain areas is shown in [7]. These studies show some regions frequently activated during MI. The specific areas are Primary Motor Cortex (M1), Supplementary Motor Area (SMA) and Premotor Cortex (PMC) [1]. These areas are also reported in [6] as well. According to the source activation shown in [7], lateral PMC is mostly activated during MI. Besides, there are significant activation in SMA and parietal lobe during MI.

III. The Dataset and The Experiment Overview

To evaluate the proposed algorithm, we have used the dataset IVa from the BCI Competition III [8]. The data was recorded from five healthy subjects while sitting on a comfortable chair with arms resting on armrests. The signal was recorded using 118 channels (Extended 10/20 System) at a sampling rate of 1 KHz. We have used the data which are downsampled to 100 Hz. The subjects were given visual cues to perform each of three MI tasks (left hand, right hand and right foot). The data is provided for only two classes (right hand and right foot). The data consists of 280 trials for each subject. Each trial consists of 3.5s of recording for any of the two classes.

We split our experiment into four cases listed in Table I considering the dataset described earlier. Each case deals with electrodes from specific regions. In Case-I, we have evaluated the data using all the available electrodes (118 channels). In Case-II, we choose 18 electrodes from the specific regions reported in [1]-[7] in the left hemisphere. In the Case-III, we choose 18 electrodes from the motor area in both hemispheres. Finally in Case-IV, we choose only 9 electrodes from the sensorimotor areas in left hemisphere which is mostly activated during MI as shown in [1]-[7]. In the last section, we consider the following assumptions featuring the spatial dynamics of the brain:

(a) The left hemisphere of the brain controls the movements in the right part of the body.
(b) The source activation during the MI is mostly in the sensorimotor (i.e. M1, SMA, PMC etc.) areas [1]-[7].

IV. Spatial Filtering Method

In this paper, we have applied CSP and RCSP as preprocessing techniques. The CSP gives a projection matrix which is used to transform the EEG data linearly into a space in which the maximal discriminative features between two classes are seen. In case of RCSP, regularization of different parameters are introduced and we can get better accuracies by tuning the parameters. The motivation behind studying RCSP is that its robustness to outliers [5] as well as its performance in small sample settings [2]. The following sections describe briefly CSP and RCSP[2].

TABLE I
Study Of Spatial Brain Areas

Case	No. Of Channels	Spatial Brain Areas
I	118	Whole Brain
II	18	Sensorimotor Areas + Parietal Lobe (Left Hemisphere)
III	18	Sensorimotor Areas (Both Hemisphere)
IV	9	Sensorimotor Areas (Left Hemisphere)

A. Sample Based Covariance Matrix Estimation For CSP

At first, the CSP method was applied to EEG regarding MI hand movements by Ramoser in [9]. CSP is very sensitive to outliers and artefact as reported in [2] and [5]. As EEG signals have very challenging SNR, more robust methods are needed to improve performance. We use RCSP which is more robust compared to CSP [2], [5].

EEG signal is represented by $\mathbf{E}$ and of size $N \times P$, where N is the number of channels and P is the number of samples per trial. In traditional CSP algorithm, the sample based covariance matrix estimation is required. The sample covariance matrix of a trial $\mathbf{E}$ is normalized with the total variance as [2], [9].

$$\mathbf{S} = \frac{\mathbf{E}\mathbf{E}^T}{\text{trace}\left[\mathbf{E}\mathbf{E}^T\right]} \tag{1}$$

where ^{T} denotes the transpose of a matrix.

Considering M trials are available for training in each class for a subject, indexed by m as $\mathbf{E}_{(\mathbf{c},\mathbf{m})}, m = 1, 2, ..., M$

$$\bar{\mathbf{S}}_c = \frac{1}{M}\sum_{m=1}^{M}\mathbf{S}_{(\mathbf{c},\mathbf{m})} \tag{2}$$

where $c \in \{1, 2\}$ represents either of the two classes of the trial depending on the MI task for the right hand or right foot movement.

The discriminative spatial patterns in CSP are calculated based on this sample based covariance matrix estimation based on (2). The next section will introduce regularization in CSP.

B. Covariance Matrix Estimation With Regularization

Regularization works by biasing the covariance estimation away from their sample based values towards more physically plausible values, which reduces the variance of the sample based estimates while tending to increase bias [2]. This is done using one or more regularization parameters (i.e., β and γ in this paper).

Now, the regularized average spatial covariance matrix for each class is defined as

$$\widehat{\sum}_c(\beta,\gamma) = (1-\gamma)\widehat{\mathbf{\Omega}}_c(\beta) + \frac{\gamma}{N}\text{trace}\left[\widehat{\mathbf{\Omega}}_c(\beta)\right]\cdot\mathbf{I} \tag{3}$$

Here, β and γ are the two regularization parameters ($0 \leq \beta, \gamma \leq 1$) and $\mathbf{I}$ is an identity matrix. $\widehat{\mathbf{\Omega}}_c(\beta)$ is comprised of

covariance matrix for the trials from the specific subjects as well as generic trials and given by

$$\widehat{\mathbf{\Omega}}_c(\beta) = \frac{(1-\beta).\mathbf{S}_c + \beta.\widehat{\mathbf{S}}_c}{(1-\beta).M + \beta.\widehat{M}} \tag{4}$$

In that case, $\mathbf{S}_c$ is the sum of the sample covariance matrices for all M training trials for class c :

$$\mathbf{S}_c = \sum_{m=1}^{M} \mathbf{S}_{(c,m)} \tag{5}$$

Also, $\widehat{\mathbf{S}}_c$ is the sum of the sample covariance matrices for $\widehat{M}$ generic training trials with covariance matrix $\mathbf{E}_{(c,\hat{m})}$ for class c:

$$\widehat{\mathbf{S}}_c = \sum_{\hat{m}=1}^{\widehat{M}} \mathbf{S}_{(\mathbf{c},\hat{\mathbf{m}})} \tag{6}$$

Here, $\mathbf{S}_c$ and $\widehat{\mathbf{S}}_c$ are normalized and analogous to the sample covariance matrix mentioned in (1). The objective of $\widehat{\mathbf{S}}_c$ is to reduce the variance in the covariance matrix estimation and may produce more reliable results.

The regularization parameter β controls the shrinkage of the training samples covariance matrix estimate to the pooled estimate. The other regularization parameter γ controls the degree of shrinkage towards a multiple of the identity matrix, with the average eigenvalue of $\hat{\mathbf{\Omega}}_\mathbf{c}(\beta)$ as the multiplier. This second shrinkage has the effect of decreasing the larger eigenvalues while increasing the smaller ones. Actually, this works well in case of small training samples set as in [2].

C. Feature Extraction In RCSP

The regularized composite spatial covariance is formed and factorized as [2]

$$\widehat{\sum}(\beta,\gamma) = \widehat{\sum}_1(\beta,\gamma) + \widehat{\sum}_2(\beta,\gamma) = \widehat{\mathbf{U}}\widehat{\wedge}\widehat{\mathbf{U}}^T \tag{7}$$

Here, $\widehat{\mathbf{U}}$ denotes the matrix eigenvectors and $\widehat{\wedge}$ denotes the diagonal matrix of corresponding eigenvalues. The eigenvalues are assumed to be sorted in descending order throughout this paper [2], [9].

Finally, the projection matrix is formed as [2]

$$\widehat{\mathbf{W}} = \widehat{\mathbf{B}}^T\widehat{\wedge}^{-1/2}\widehat{\mathbf{U}}^T \tag{8}$$

where $\widehat{\mathbf{B}}$ denotes the matrix of eigenvectors for the whitened spatial covariance matrix and defined as

$$\widehat{\mathbf{B}} = \widehat{\wedge}^{-1/2}\widehat{\mathbf{U}}^T\widehat{\sum}_c(\beta,\gamma)\widehat{\mathbf{U}}\widehat{\wedge}^{-1/2} \tag{9}$$

In RCSP, an input trial E is projected as [2]

$$\widehat{\mathbf{X}} = \widehat{\mathbf{W}}^T\mathbf{E} \tag{10}$$

To get the most discriminative features for both classes, the optimal channels are to be selected from the leftmost and rightmost channels. For example, the first channel represents the most distinguished features for class 1 and the last channel represents the most distinguished features for class 2. As channel selection converges to the central channel of the $\widehat{\mathbf{X}}$, the features becomes poor which may hardly distinguish different classes. It is to be noted that, RCSP becomes traditional CSP when $\beta = \gamma = 0$.

V. Experimental Design And Validation

In this section, we will discuss the experimental steps and elaborate the parameters' specific values which we have used in this study.

A. Experimental Design

The experimental flow of the proposed study is described in Fig.1. We have used N channels as described in Table I. At first, the set of signals are bandpass filtered to get rid of the low frequency artifacts such as eyeblinks. Then, we have applied spatial filtering to increase the classification accuracy. From channels containing filtered signals, we select 4 optimal channels, i.e., first two channels for class 1 and the last two channels for class 2. After selecting 4 optimal channels, we have extracted subband energy and subband entropy from all the four optimal channels. Afterwards, the features are used for classification. We have used Neural Network (NN) as classifier.

B. Validation Of Proposed Method

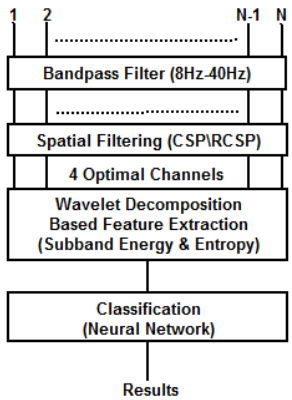


Fig. 1. Block Diagram Of Our Proposed Algorithm

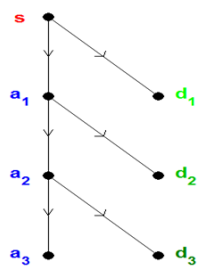


Fig. 2. Wavelet Decomposition Tree Upto Level 3

In this experimental study, we have proposed the spatial dynamics to be studied to minimize the number of channels as well as computational cost. We have used 120 trials for the training phase, 60 trials for the validation phase and 100 trials for the testing phase. For each trial, we have taken 3secs of data after 0.5secs of the visual cues. The experiment is divided into four *cases* as described in Table I. First, the data is band pass filtered from 8-40Hz using Butterworth filter of order 10. The Butterworth filter is used as it has magnitude response which is almost flat in the pass band. Second, the data is preprocessed using CSP or RCSP as discussed in the previous section. The first and last two channels are selected which show optimal variances for class 1 and class 2, respectively.

Then, wavelet based subband energy and subband entropy (Shannon Entropy) are extracted as features. The signals are decomposed up to level 3 using wavelet transform as shown in Fig. 2. Thus, each of the four channels is decomposed into four subbands. In each of the subbands, energy and Shannon entropy have been calculated. Finally, Neural Network is used as classifier. We have used a two layer feed forward network with 12 hidden neurons at the hidden node. This setup is applied to the all four *cases* as in Table I. In this experiment, we have used different combination of values of β and γ as follows

$\beta = (0, 0.001, 0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9)$

$\gamma = (0, 0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9)$

We have tested all possible combination of values from the above sets. These values are selected from previous experiments described in [2] and [3].

TABLE II
VALIDATION OF PROPOSED ALGORITHM : CLASSIFICATION ACCURACIES

Case	Method	aa	al	aw	av	ay	Avg.
I	CSP	83.6%	97.9%	93.2%	70.4%	91.8%	87.38%
	RCSP	86.8%	97.9%	95.7%	70.4%	92.9%	88.78%
II	CSP	80.4%	96.1%	92.5%	66.8%	89.3%	85.02%
	RCSP	85.4%	97.1%	93.2%	70.7%	93.6%	88%
III	CSP	81.1%	95.7%	92.5%	67.9%	93.2%	86.08%
	RCSP	87.1%	96.4%	93.6%	70%	93.9%	88.2%
IV	CSP	82.9%	94.6%	87.1%	71.8%	93.9%	86.06%
	RCSP	85.4%	94.3%	88.2%	72.5%	97.1%	87.5%

The results for all four *cases* under this study are shown in Table II. It is observed that the performance of the proposed method varies across different subjects. This is due to the fact that the quality of the signal regarding MI depends on the subjects' ability to simulate MI task in brain. In this experiment, the best result is 97.9% for subject 'al' in Case-I and the worst result is 66.8% for subject 'av' in Case-II. In Case-IV, we have the best results for subjects 'av' and 'ay' which are 72.5% and 97.1% respectively although we have used only 9 selective channels. For the subjects 'aa', 'al', and 'aw', the best results are 87.1% (Case-III), 97.9% (Case-I) and 95.7% (Case-I), respectively. But, these best results for subjects 'aa' and 'al' are comparable with the results reported in Case-IV. Only for subject 'aw', the difference between the best result and the result from Case-IV is a little high. However, if we look at the average results for different subjects, we can easily notice that the best average result is 88.78% (Case-I) and the average result for Case-IV is 87.5%. These are quite comparable results although we have reduced the number of electrodes significantly. Therefore, we can conclude that we can reduce the computational cost by engaging spatial brain dynamics in BCI development. The computational latencies of all four cases are reported in Table III. The difference between the studies are the number of channels (electrodes) which are used to verify the proposed algorithm. Therefore, we have accounted for the computational latencies between the input of the bandpass filter and the output of the spatial filter as shown in Fig. 1. The remaining computational complexities are quite equivalent as in Fig. 1. The subject specific computational latencies are shown in Table III. In each of the cases, the computational latencies are observed 10 times (i.e, 10 simulation runs for each). Then, we have calculated the average over all the 10 runs. Also, the averages for each of the studies are shown in Table III. The average computational latency in Case-I is significantly higher compared to the average computational latency of Case-IV. Therefore, it is feasible to implement an BCI system with less number of electrodes without compromising performance significantly. In this way, computational cost can be reduced drastically by selecting channels from specific Regions Of Interests (ROIs) of the brain.

TABLE III
COMPUTATIONAL LATENCIES (IN SECS): COMPARISON BETWEEN STUDIES

	Study 1	Study 2	Study 3	Study 4
aa	22.33	4.56	4.39	2.89
al	20.80	4.26	4.21	2.76
aw	20.98	4.41	4.18	2.75
av	22.03	4.21	4.21	2.73
ay	20.61	4.32	4.19	2.73
Avg.	21.35	4.35	4.24	2.77

This experiment has been performed in a personal workstation with the following configurations: CPU-2.5GHz, RAM-4GB, Windows 7 Home Premium, MATLAB 2013b.

VI. FUTURE DEVELOPMENT

In this experimental study, we have studied spatial brain dynamics and have tried to correlate these facts with BCI development. This paper presents the significance of the spatial dynamics which contributes to reducing the computational cost in BCI system. We have considered previous literature based on the spatial reasoning of the brain during MI tasks. But, the fact is that the brain dynamics varies a lot from subject to subject. Thus, subject specific brain dynamics may play important role in subject specific BCI development. To do this, employment of both the brain imaging techniques (i.e., fMRI) and the EEG is necessary. In that case, it is possible to extract useful spatial information by studying on the source activation areas during MI (i.e., by analyzing the the fMRI images for specific subject). Based on this, EEGs from few selected channels can be employed to extract features resulting in reduction in the computational complexities in BCI.

REFERENCES

[1] Martin Lotze, Ulrike Halsband, "Motor imagery", *Journal of Physiology-Paris*, Volume 99, Issues 4 June 2006, Pages 386-395, ISSN 0928-4257.

[2] Haiping Lu, How-Lung Eng, Cuntai Guan, K.N. Plataniotis, and A.N. Venetsanopoulos, "Regularized Common Spatial Pattern With Aggregation for EEG Classification in Small-Sample Setting", *IEEE Trans. on Biomedical Engineering,* Vol. 57, No. 12, pp. 2936-2946, Dec. 2010.

[3] Lotte, F.; Cuntai Guan, "Regularizing Common Spatial Patterns to Improve BCI Designs: Unified Theory and New Algorithms," *Biomedical Engineering, IEEE Transactions on*, vol.58, no.2, pp.355,362, Feb. 2011.

[4] Yijun Wang; Shangkai Gao; Xiaorong Gao, "Common Spatial Pattern Method for Channel Selelction in Motor Imagery Based Brain-computer Interface," *Engineering in Medicine and Biology Society*, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the , vol., no., pp.5392,5395, 17-18 Jan. 2006.

[5] X. Yong, R. K. Ward, and G. Birch, "Robust common spatial patterns for EEG signal preprocessing," in *EMBC 2008*, Vancouver, August 2008, pp. 20872090.

[6] Martin Lotze , Pedro Montoya , Michael Erb , Ernst Hlsmann , Herta Flor, Uwe Klose, Niels Birbaumer, Wolfgang Grodd, "Activation of Cortical and Cerebellar Motor Areas during Executed and Imagined Hand Movements: An fMRI Study" *Journal of Cognitive Neuroscience*, September 1999, Vol. 11, No. 5, Pages 491-501

[7] Solodkin Ana, Hlustik Petr, Chen, E. Elinor, Small, Steven L., "Fine Modulation in Network Activation during Motor Execution and Motor Imagery" *Cereb Cortex*. 2004 Nov;14(11):1246-55. Epub 2004 May 27.

[8] Blankertz, B.; Muller, K.; Curio, G.; Vaughan, T.M.; Schalk, G.; Wolpaw, J.R.; Schlogl, A.; Neuper, C.; Pfurtscheller, G.; Hinterberger, T.; Schroder, M.; Birbaumer, N., "The BCI competition 2003: progress and perspectives in detection and discrimination of EEG single trials," *Biomedical Engineering, IEEE Transactions on*, vol.51, no.6, pp.1044,1051, June 2004.

[9] Ramoser, H.; Muller-Gerking, J.; Pfurtscheller, G., "Optimal spatial filtering of single trial EEG during imagined hand movement," *Rehabilitation Engineering, IEEE Transactions on*, vol.8, no.4, pp.441,446, Dec 2000

Biomass Quality Analysis for Power Generation

A.H.M. Kamal,[1,*] Md. Mahbubur Rahman,[1] Md. Mohaiminul Haque,[1] Salman Sadiq,[1] and Aminul Hoque[1]

[1]Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology
Dhaka-1205, Bangladesh
*jimmybuet@gmail.com

Abstract— **Quality analysis of biomass is an important parameter to use it as an effective energy source. It is necessary to understand because the performance and efficiency of biomass based power plant machineries are interrelated with biomass quality. In this paper, the efficiency of biomass utilization has been realized from quality perspective. Moreover, biomass quality variation has been analysed with respect to different weather conditions.**

***Index Terms*—Biomass, biomass quality, power generation, NCV, EMC.**

I. Introduction

Every year the whole world produces about hundreds of tons of biomass in various forms lending itself a popular source of energy after coal, oil and natural gas. The most obvious benefit of biomass based energy is that it is a renewable source and also environment friendly. Worldwide biomass ranks fourth as an energy resource, providing approximately 14% of the world's energy. In developing countries like Bangladesh, it can be as high as 35% of the primary energy supply [1]. Quality control and analysis of the biomass material are essential to both the home heat generator and large industrial users. Biomass has different characteristics and understanding the composition of it can help to decide which conversion technique will be best for power generation. Biomass quality also changes with seasonal variation. So, it is another important factor to be considered in the case of energy extraction process from biomass.

In this paper, biomass characteristics with its chemical composition have been presented. Moreover, the comparison between biomass and fossil fuel has been discussed on the basis of their function as fuel. For the analysis of biomass quality, Equilibrium Moisture Content (EMC) of woody biomass has been calculated by using an empirical equation which is an approximation of the relationship among EMC, temperature and relative humidity. However, seasonal quality variation and corresponding efficiency of biomass energy source have been obtained by analysing the yearly weather variation in Dhaka, Bangladesh.

II. Biomass as Energy Source

The fundamental process by which energy is stored in biomass is photosynthesis. It is the process where light energy from the sun is captured and stored in the chemical bonds of the tissues of living plants [2]. The photosynthetic efficiency is the fraction of light energy converted into chemical energy during photosynthesis in plants and algae. Photosynthesis can be described by the simplified chemical reaction-

$$H_2O + CO_2 + \text{energy} \rightarrow C_6H_{12}O_6 + O_2$$

$C_6H_{12}O_6$ represents carbohydrates such as sugars, cellulose and lignin [3].

Biomass energy can be converted into different forms of energy like chemical, mechanical or electrical .The process of converting energy varies from biomass type and specific properties like moisture and ash content of biomass. Net Calorific Value (NCV) or Lower Heating Value (LHV) of biomass changes significantly with the variation of moisture content in it [4]. When biomass is converted thermally in the presence of excess amounts of air (combustion), the total energy released in the form of heat is called its heating value or calorific value. Generally, biomass has heating value ranging from 15 MJ/kg to 25 MJ/kg [5]. Fig. 1 shows the variation of NCV of some biomass samples with the change in moisture content where NCV decreases with the increase of moisture content.

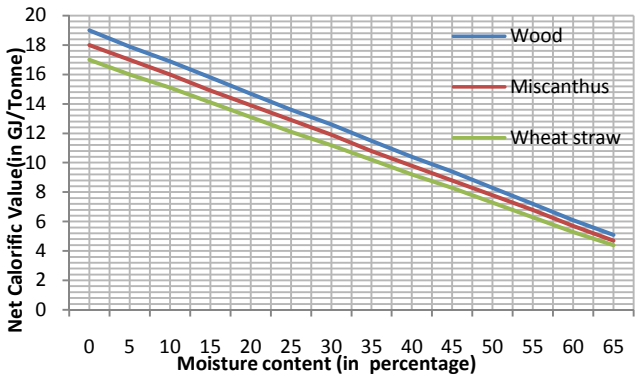


Fig. 1 Calorific value vs. moisture content of some typical biomass.

III. Chemical Composition of Biomass

Biomass is carbon based and composed of a mixture of organic molecules containing hydrogen, usually including atoms of oxygen, often nitrogen. It also contains small quantities of other atoms, including alkali, alkaline earth and heavy metals [6]. Table I contains general chemical composition of biomass in percentage weight (dry condition) [7].

TABLE I
General Chemical Composition of Biomass

Element	Weight (%)
Carbon	44-51
Hydrogen	5.5-6.7
Oxygen	41-50
Nitrogen	0.12-0.6
Sulfur and Others	0-0.2

Biomass differs from coal (solid fossil fuel) in many important ways, including the organic, inorganic, energy content and physical properties. Relative to coal, biomass

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

generally has less carbon, more oxygen, more silica and potassium, less aluminium and iron, lower heating value, higher moisture content and lower density and friability [8]. Fig. 2 shows the comparison of chemical composition between biomass and coal.

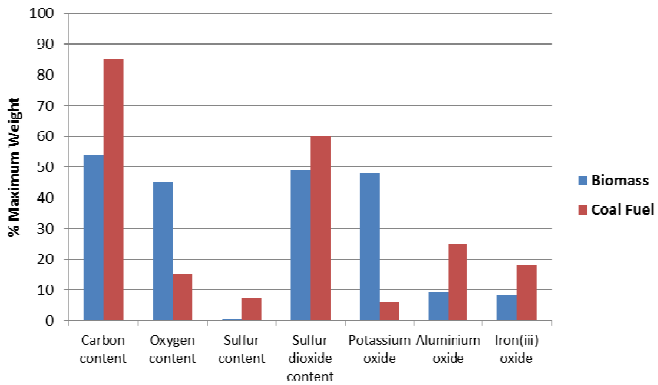


Fig. 2 Chemical composition comparison between biomass and coal.

IV. Quality Analysis

Biomass qualities are affected by mainly two factors. These are internal or structural factors and external or environmental factors.

A. Internal Factors

Internal factors of biomass are mainly the constitutional factors such as fixed carbon, volatile substances, ash content, water content etc.

B. External Factors

For biomass quality analysis external factors are the most important parameters to be considered. Among the external factors, moisture content is the important one. Moisture content can be determined by two methods [9].

1) Wet basis: In the case of wet basis calculation,

$$Moisture\ content = \frac{Mass\ of\ water\ in\ the\ biomass}{Total\ mass\ of\ biomass} \times 100\%.$$

2) Dry basis: In the case of dry basis calculation,

$$Moisture\ content = \frac{Mass\ of\ water\ in\ the\ biomass}{Total\ mass\ of\ biomass\ without\ water} \times 100\%.$$

Moisture content of biomass largely depends on temperature and humidity variation of the environment. The moisture content of wood is directly related to the humidity and temperature of the surrounding air. The Equilibrium Moisture Content (EMC) occurs when the wood has reached a water content equilibrium with its environment and is no longer gaining or losing moisture. The EMC value of biomass in different seasons of the year is necessary to realize the efficiency of biomass utilization. EMC can be calculated by using the Hailwood-Horrobin equation [10],

$$EMC\ (\%) = \frac{1800}{W}\left[\frac{kh}{1-kh} + \frac{klh + 2k^2lmh^2}{1 + klh + k^2lmh^2}\right] \quad (1)$$

Where, h = Relative humidity (fractional), T = Temperature (in degree Fahrenheit), $W = 330 + 0.452T + 0.00415T^2$, $k = 0.791 + 4.63 \times 10^{-4}\ T - 8.44 \times 10^{-7}T^2$, $l = 6.34 + 7.75 \times 10^{-4}\ T - 9.35 \times 10^{-5}T^2$, $m = 1.09 + 2.84 \times 10^{-2}\ T - 9.04 \times 10^{-5}T^2$.

Using (1), EMC (%) vs. temperature for different values of humidity has been plotted by MATLAB R2008a software.

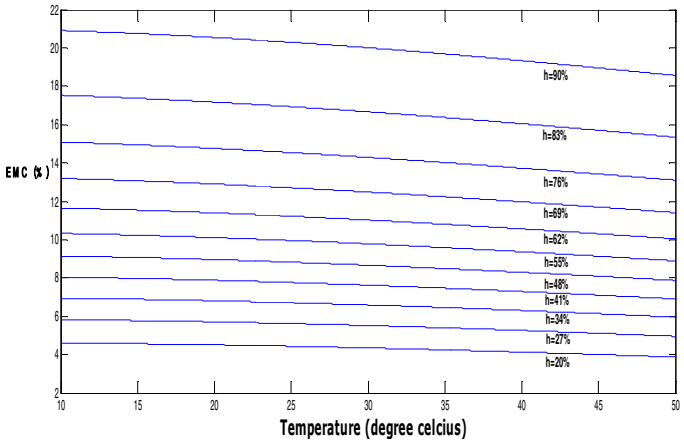


Fig. 3 EMC (%) vs. temperature graph for different humidity.

Using the graph of Fig. 3, EMC (%) value fluctuation due to temperature and humidity variation of different months (year 2013) of Dhaka, Bangladesh has been determined. Fig. 4 shows the EMC (%) fluctuation graphically.

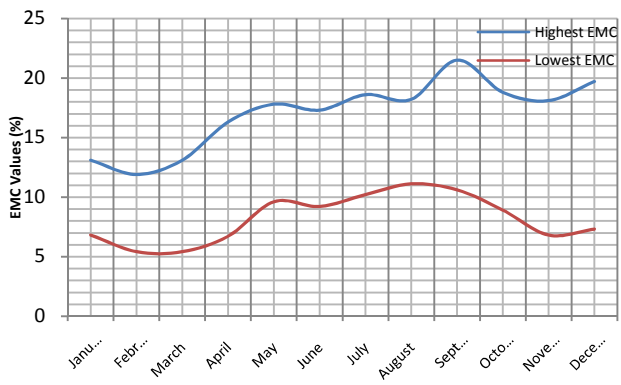


Fig. 4 EMC (%) values vs. different months.

From Fig. 4 it is observed that, EMC (%) value falls from the month of January and reaches its bottom at the month of March. After that, from March to May its shows increasing behaviour of EMC (%). During the period of May to August it becomes almost constant. In the mid of September to October EMC (%) value shows abrupt behaviour and reaches its peak value. From October to November its decreasing rate is similar to the month of January to February.

In biomass based power generation, total system efficiency depends on EMC (%) of biomass. If EMC (%) increases, NCV of biomass decreases (Fig. 1) as well as the efficiency of power generating components like boiler, turbine etc. From the analysis of Fig. 4 it has been found that, probability of high efficient operation is greater at the month of February-March than any other months throughout the year. It has been also found that lower efficiency occurs at the month of September-October.

As it is observed that, efficiency is higher at a particular time in a year whereas lower at other periods, so for ensuring continuous power flow from biomass based power plant, sufficient amount of backup power should be provided at the period of lower efficiency and less backup is needed at the time of higher efficiency.

V. Conclusions

Biomass quality analysis should be done at priori before using it as fuel in the power generation system because the efficiency of energy extraction from biomass depends on its quality. From the study, it has been found that efficiency of biomass energy source is high at specific period of year and low at the remaining time. For this reason, the weather condition should be kept in mind for the utilization of biomass as energy source.

Acknowledgment

Heartiest gratitude to Saleh Md. Main Uddin, Superintending Engineer, Waste Management Department; A.H.M. Abdulla Harun, Assistant Engineer, Waste Management Department of Dhaka South City Corporation, Md. Mesbahul Karim, Superintending Engineer, Waste Management Department of Dhaka North City Corporation and Mir Farida Shaheen, Professional Assistant, Climate Division, Dhaka for providing necessary documents and data on biomass and weather related information respectively.

References

[1] Ayhan Demirbas, *Biorefineries: For Biomass Upgrading Facilities*, 1st ed. London, England: Spinger-Verlag, 2010.

[2] (2005) Stanford University Global Climate & Energy Project, Annual Report, An Assessment of Biomass Feedstock and Conversion Research Opportunities. [Online]. Available: http://gcep.stanford. edu/pdfs/assessments/biomass_assessment.pdf

[3] (2013) Photosynthetic efficiency. [Online]. Available: http://en.wikipedia.org/wiki/Photosynthetic_efficiency

[4] (2014) Calorific value as a function of moisture content - Biomass Energy. [Online]. Available: www.biomassenergycentre.org

[5] Sergio C. Capareda, "Introduction to Biomass Energy Conversions," vol. 1, pp.210-211,2013.

[6] (2013) Biomass Energy. [Online]. Available: http://www. biomassenergycentre.org.uk/

[7] Stanislav V.Vassilev, David Baxter, Lars K.Andersen & Christina G.Vassileva. (2009) An overview of the chemical composition of biomass .[Online]. Available: http://www.sciencedirect.com/ science/article/pii/S0016236109004

[8] Ayhan Demirbas. (2003) Combustion characteristics of different biomass fuels. [Online]. Available: http://www.sciencedirect.com/ science/article/pii/ S0360128503000789

[9] (2013) Characteristics of Biomass as a Heating Fuel. [Online]. Available: http://pubs.cas.psu.edu/FreePubs/PDFs/ub043.pdf

[10] (2013) Equilibrium moisture content. [Online]. Available: http://en.wikipedia.org/wiki/Equilibrium_moisture_content

A New Technique to Improve the Linear Generator Designed for Oceanic Wave Energy Conversion

Omar Farrok[1,*] and Md. Mortuza Ali[2]

[1] Department of Electrical and Electronic Engineering, Ahsanullah University of Science and Technology (AUST)
141-142 Love road, Tejgaon I/A, Dhaka-1208, Bangladesh

[2] Department of Electrical and Electronic Engineering, Rajshahi University of Engineering & Technology (RUET)
Talaimary, Rajshahi-6204, Bangladesh

*omarruet@gmail.com

Abstract—**In this paper a new technique is proposed to improve the performance of the linear permanent magnet generator (LPMG) which is basically a point absorber. This paper also represents analysis, design and simulation of a LPMG for improvement of oceanic wave energy conversion (WEC). Different output voltage, current and power waveforms of the conventional and proposed LPMG are analyzed, simulated and compared by MATLAB software by constructing mathematical models of WEC, LPMG and output stage. Various parameters like magnitude, wavelength and phase of oceanic wave are observed to make proper simulation by changing different parameters. It is applicable almost all of the point absorbers and offers the great advantage to get relatively more even power distribution which is a major concern for WEC.**

Index Terms—**Linear generator, point absorber, oceanic wave energy, translator.**

I. Introduction

In recent years WECs are the centre of attention in the renewable energy field because it has the feature of very high energy density. The wave energy resource is substantial, and the total resource around the world is estimated 10 TW in the open sea, which is comparable to the world's total power consumption [1]. Different types of WEC devices have been proposed for development, such as Archimedes Wave Swing (AWS), Oscillating Water Column (OWC), Pelamis structure, Wave Dragon, Mighty Whale etc. ([2]–[4]). Ocean power technology companies own the commercial product based on a point-capture wave power generation system with AWS submerged under the water for power generation [5]. Traditional wave power generation methods have the disadvantages of complex and expensive system structures and are hard to maintain and possess low transformation efficiency ([6], [7]). Hence, the generated electricity from such systems is not cost-effective. The AWS consists of a hollow cylinder and a lid fixed on the sea-bottom filled with air, called floater as shown in Fig.1(a). The floater moves in the vertical direction. When the wave comes, the floater sinks because of increase of the weight of the water above, while the pressure of the air in the cylinder increases and vice-versa. In a power buoy, from the reciprocating motion of the floater moved linearly, the wave energy is extracted that convert the wave energy to electrical energy. A point absorber is a floating structure with components that move relative to each other due to wave action utilizing the rise and fall of the wave height at a single point for energy conversion as shown in Fig.1(b). The measurement of power buoy and accessories are given here as an example and need to be varied according to different situation. The relative up and down bobbing motion caused by passing waves is used to drive electromechanical or hydraulic energy converters to generate power. Internal construction of a conventional LPMG is shown in Fig.1(c).

At present, research studies on direct-drive methods have been carried out by direct capture of wave energy as it is proved to be more efficient compared to the traditional generators ([7]–[10]). The LPMGs are employed to directly transform the low speed mechanical energy into electricity, thus eliminating intermediate mechanical translators, a crank or linkage etc. or converters. Although the LPMGs have relatively large force-to-volume ratio and own high efficiency, the involvement of LPMGs results in complicated winding scheme or sophisticated arrangement and assembly of LPMGs [11].

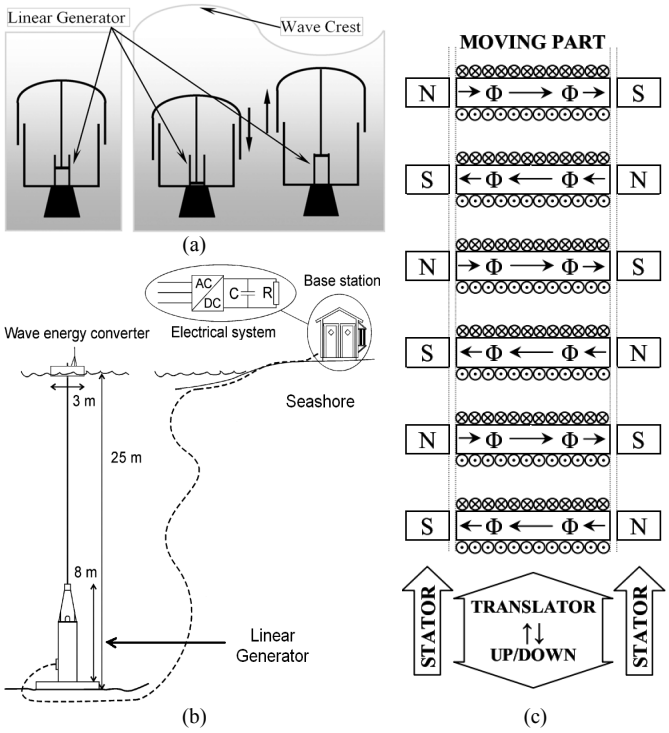


Fig. 1(a) LPMG shown in an AWS configuration, (b) LPMG mounted with power buoy, (c) LPMG internal construction.

In the latest report from the International Energy Agency (IEA), 81 different wave energy projects including linear generator (LG) and WEC were examined and out of them 13 were estimated to have reached the stage of having a full- or near-full-scale prototype at sea [12].

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

II. Conventional Linear Generators

Different internal construction of conventional LPMGs are shown in Fig.2(a–c). When a magnet moves in relation to an electromagnetic coil, this changes the magnetic flux passing through the coil, and thus induces the flow of an electric current. Like rotary generators it essentially contains stator and translator which works similar as rotor. The translator and the stator can either have the same vertical length as shown in Fig.1(c) that mean the active surface will vary during cycle of motion or one of them can exceed the other in vertical length as shown in Fig. 2(c). If the surplus in length is equal to the pitch length the active surface will be constant.

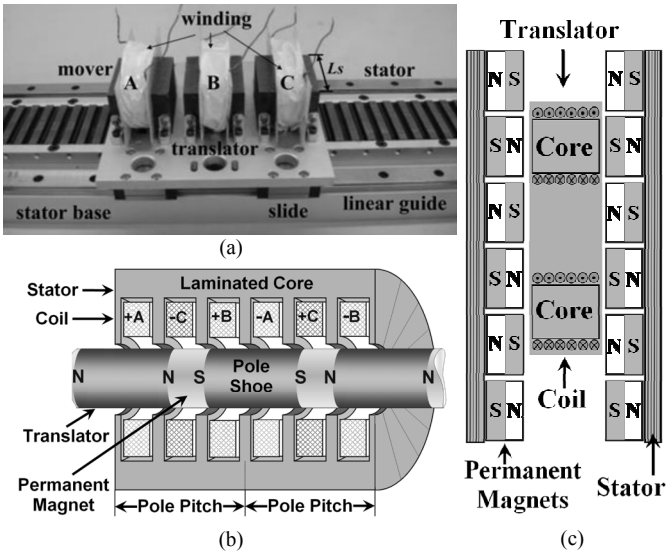


Fig. 2(a) Conventional flux switching LPMG, (b) Conventional tubular LPMG, and (c) Conventional LPMG with longer stator.

However, the longer translator has some advantages on longer stator though both of these give similar output waveforms. A practical output waveform of electrical power is shown in Fig.3 for an example [12].

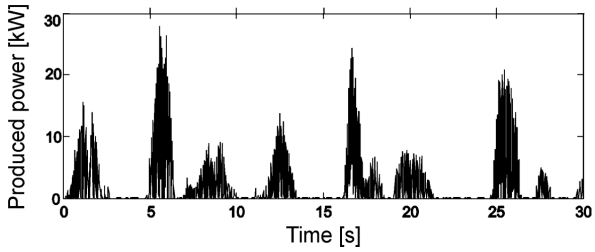


Fig. 3 Electrical power output of a conventional LPMG.

All conventional LPMGs produced discontinuous output power. From Fig.3 it is clear that it gives negligible output power from 2 to 3 seconds approximately in each cycle.

III. Model of WEC and LPMG

The up and down movement produced by an ideal wave can be approximated by a sine function, resulting a cosine-shaped speed variation, with maximum speeds reaching up to 2m/sec, depending on the sea state. The speeds are usually between 0.5m/sec to 1m/sec. The linear generator should be chosen considering these speed values and the working conditions. The induced voltage is found according to Faraday's law of electromagnetic induction as given in (1)

$$E = -N\frac{d\overline{\Phi}}{dt} \tag{1}$$

Here E is induced voltage in the coil, N is turn number of wire in coil and $\overline{\Phi}$ is average flux passing through coil. As vertical wave displacement is very similar to sinusoidal function, assuming sinusoidal wave the vertical displacement of wave and velocity can be expressed as given in (2) and (3) respectively.

$$y(t) = H_m \sin\left(\frac{2\pi}{T}t \pm \theta_i\right) \tag{2}$$

$$v(t) = H_m \frac{2\pi}{T}\cos\left(\frac{2\pi}{T}t \pm \theta_i\right) \tag{3}$$

Here $y(t)$ and $v(t)$ represents position or displacement of wave and velocity respectively. H_m, θ_i and T represents height, phase angle and period of oceanic wave respectively. In this model H_m, θ_i and T can be varied to get wide range of input variable.

The proposed LPMG is consisted of a possible special new translator as shown in Fig.4(a) and a stator with windings. Flux switching method is used for generation of electricity. According to Fig.4(b) when translator moves at a particular time it aligns at position-1 and flux direction is upward through first pole shoe. At position-2 the flux direction is reversed as it is downward through the same pole shoe. All the cases vertical velocity of any type of translator mounted with floater or something like that is very low at the two ends, hence generating negligible amount of electricity as shown in Fig.3 as an example as the output power is proportional to speed or velocity. A new design of pole shoes so far is also implemented which is able to increase rate of change of flux with proper mounting with translator, thus generate more electrical power even with low velocity because although the displacement per unit time is low two consecutive translator tips are closer near both ends compared to the regular distance around the middle position.

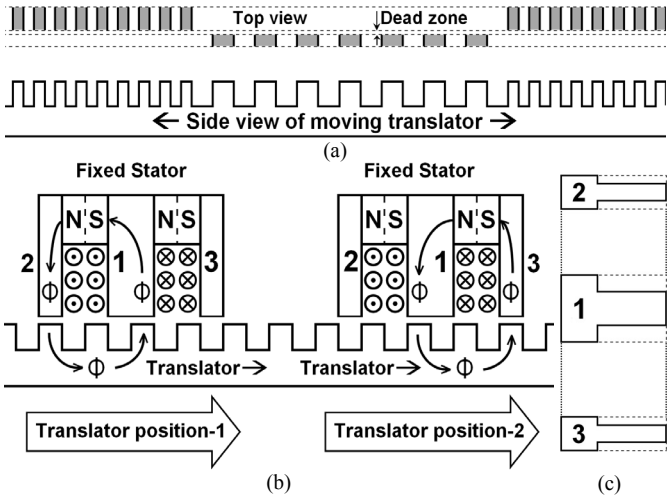


Fig. 4(a) Top view and side view of the proposed translator, (b) Direction change of flux in stator when position is changed, and (c) Bottom view of proposed pole tips of stator.

The Proposed LPMG block diagram of model is shown in Fig.5. Different parameters of wave energy like wave length, height, phase, reaction force, cogging force, friction is considered which can be given as preset value or it can be varied during simulation. Variables can be changed during simulation using function with preset value or can be given in the model by another function named "Random" connected with some other blocks that can produce random height and wavelength of wave energy to produce more realistic outputs.

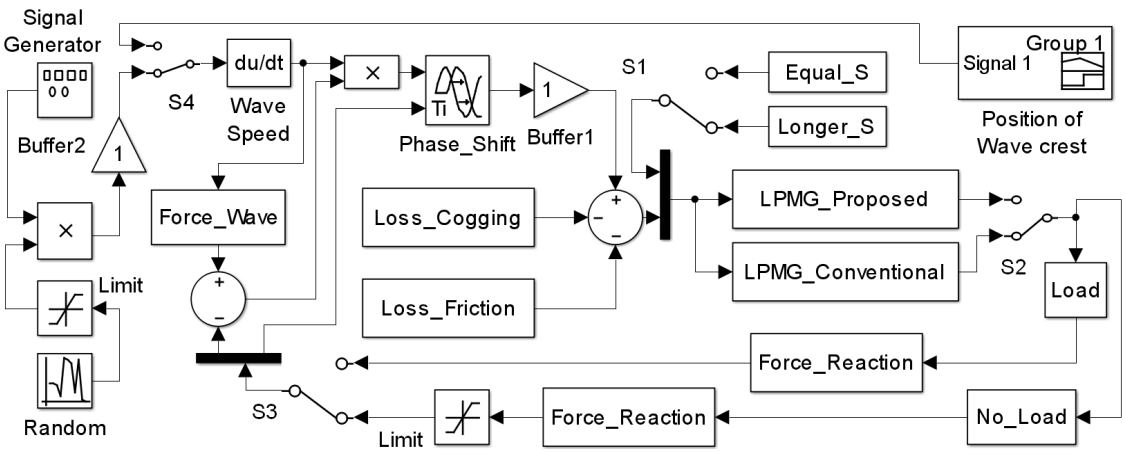


Fig. 5 Block diagram of simulation model of the conventional and proposed WEC and LPMG.

IV. SIMULATIONS

In the simulation model cogging force, core loss and frictional loss is considered as the function of velocity with some coefficients. Reaction force and loading effect depends on floater, output electrical power and current mainly. Selector switches S1 to S4 are used in this model to increase its flexibility and multimode selection purpose. For an example, equal stator and longer stator mode can be selected by S1 represented as Equal_S and Longer_S respectively. Fig.6 represents vertical oceanic wave displacement and velocity which are the inputs for the conventional and proposed model according to (2) and (3). Fig.7 shows the voltage and current waveforms of a conventional LPMG and Fig.8 shows the same for the proposed LPMG for a given resistive load. Fig.9 shows generated voltage of a conventional LPMG with respect to magnetic flux density represented as B in tesla. As flux density varies from 0 to 1.2 wb/m^2 according to the model parameters which is much less compared to voltage scale so, it is zoomed to 100 times so that is can be easily visualize in the same scale of voltage. Fig.10 shows generated voltage versus B for the proposed LPMG. Fig.11 and Fig.12 represents phase and magnitude of no load and full load voltage waveforms of a conventional and the proposed LPMG respectively. Full load voltage is less than no load voltage not only for the regulation of LPMG but also for the reaction force created by the wave, hence there is some phase difference. The output electrical power of a conventional and the proposed LPMG is shown in Fig.13 and Fig.14 respectively for equal translator length and the same are shown in Fig.15 and Fig.16 respectively for longer translator length than stator where the pitch length is equal or greater than wave amplitude. It is clear that average output electrical power is more in the proposed LPMG and also the ripple factor is significantly low than the conventional LPMG. Fig.7 to Fig.16 are cropped from 0 to 10 second from full simulation time of 20 and 40 second to easily visualize and compare the waveform, peak values and phases.

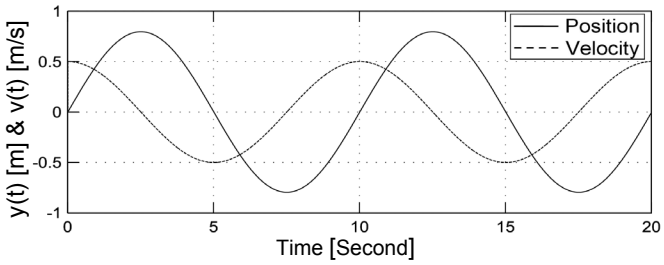


Fig. 6 Displacement & vertical velocity of oceanic wave.

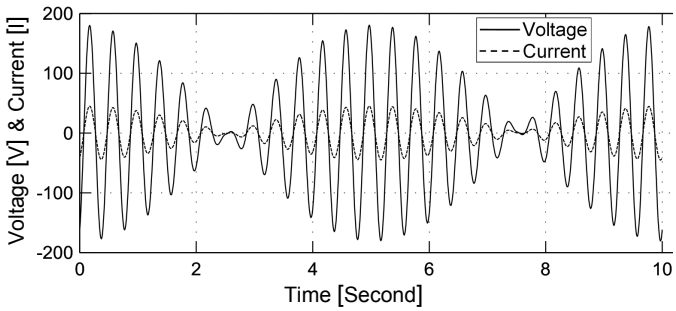


Fig. 7 Voltage & current waveforms of a conventional LPMG.

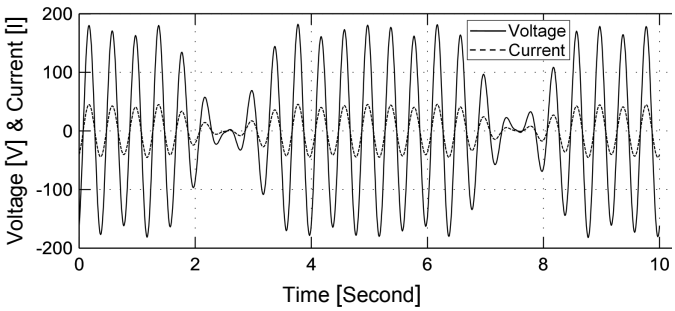


Fig. 8 Voltage & current waveforms of the proposed LPMG.

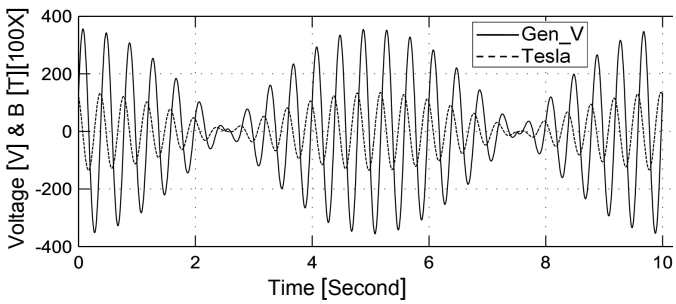


Fig. 9 Voltage & flux density waveforms of a conventional LPMG.

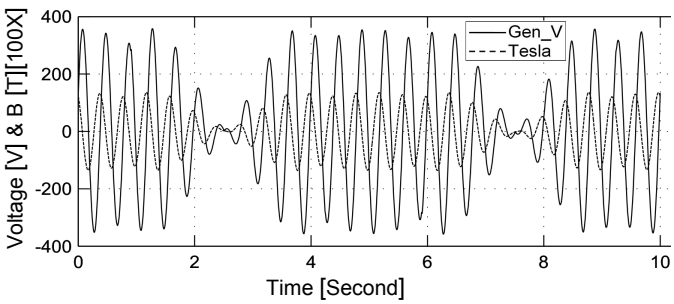


Fig. 10 Voltage & flux density waveforms of the proposed LPMG.

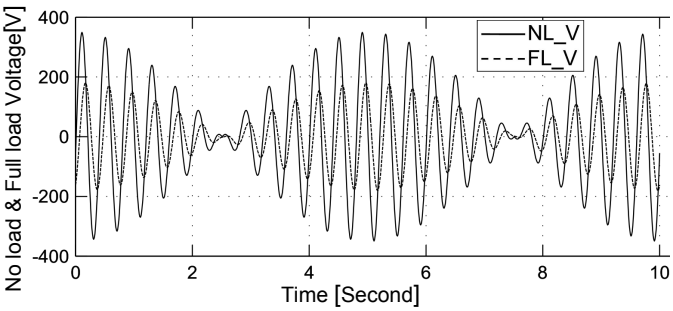


Fig. 11 No load & Full load voltage waveforms of a conventional LPMG.

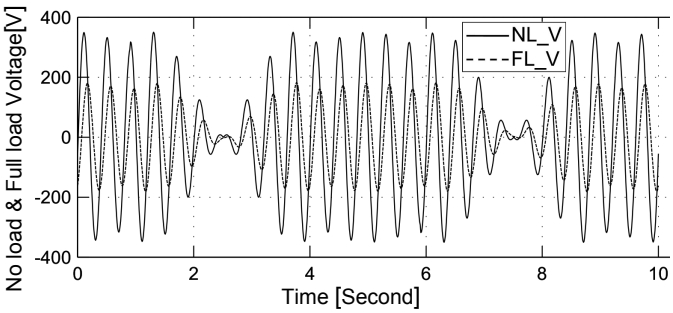


Fig. 12 No load & Full load voltage waveforms of the proposed LPMG.

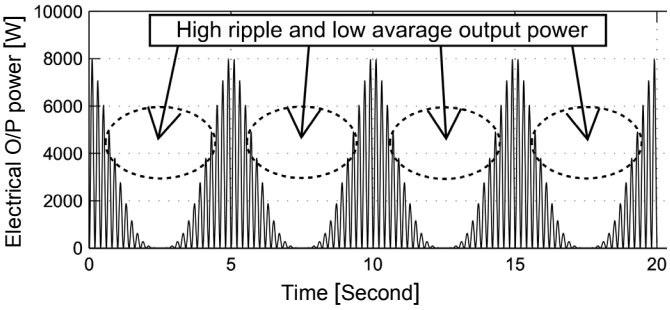


Fig. 13 Output instantaneous power waveform of a conventional LPMG of equal length of translator and stator.

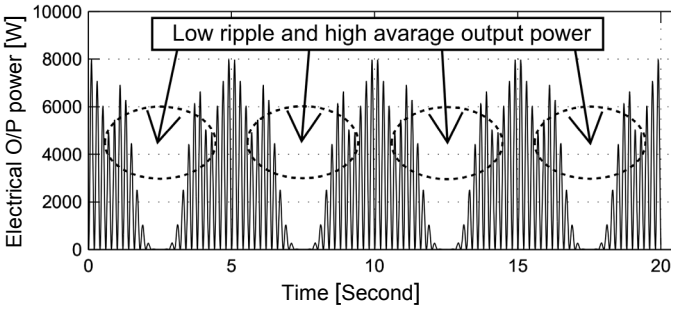


Fig. 14 Output instantaneous power waveform of the proposed LPMG of equal length of translator and stator.

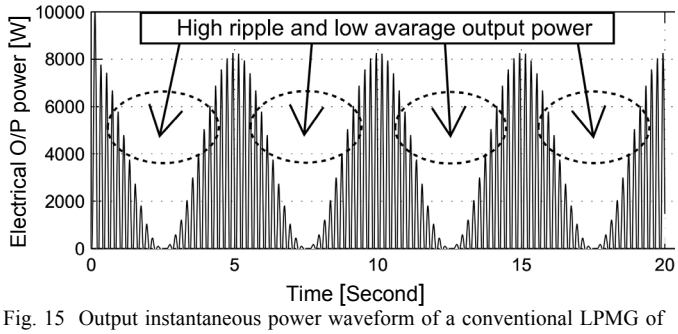


Fig. 15 Output instantaneous power waveform of a conventional LPMG of longer length of translator than stator.

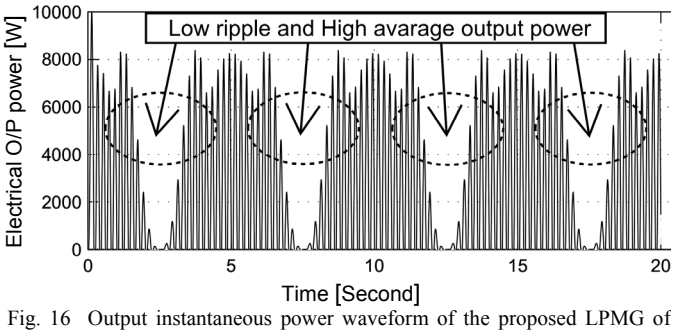


Fig. 16 Output instantaneous power waveform of the proposed LPMG of longer length of translator than stator.

V. Conclusions

Though in commercial LPMG multiple windings are used, in this model comparison of only one winding for both the conventional and proposed one is explained but it can also simulate multiple coil windings. Generally all of the LPMG connected to grid is at first connected to power converter to make a dc power then supply ac electrical power to grid by means of a power inverter with additional devices. As all of the conventional LPMG produces negligible electrical power for long interval of time in each half cycle of wave energy so, it is not possible to supply an acceptable level of constant real power to grid without electrical storage with large capacity because of the high fluctuating electrical power. This major drawback of all the LPMG is reduced in a great extent with the proposed design which can extract electrical power from wave energy while the vertical velocity of wave is low.

References

[1] Feng Wu, Xiao-Ping Zhang, Ping Ju, and Michael J. H. Sterling, "Modeling and Control of AWS-Based Wave Energy Conversion System Integrated Into Power Grid," *IEEE Transactions on power systems*, vol. 23, no. 3, pp. 1196–1204, Aug. 2008.

[2] R. Henderson, "Design, simulation, and testing of a novel hydraulic power take-off system for the Pelamis wave energy converter," *Renew. Energy*, vol. 31, no. 1, pp. 271–283, Jan. 2006.

[3] J. P. Kofoed, P. Frigaard, E. Friis-Madsen, and H. C. Sorensen, "Prototype testing of the wave energy converter wave dragon," *Renew. Energy*, vol. 31, no. 2, pp. 181–189, Feb. 2006.

[4] D. L. O'Sullivan and A. W. Lewis, "Generator selection and comparative performance in offshore oscillating water column ocean wave energy converters," *IEEE Trans. Energy Convers.*, vol. 26, no. 2, pp. 603–614, Jun. 2011.

[5] H. Polinder, M. E. C. Damen, and F. Gardner, "Linear PM generator system for wave energy conversion in the AWS," *IEEE Trans. Energy Convers.*, vol. 19, no. 3, pp. 583–589, Sep. 2004.

[6] A. Nasiri, S. A. Zabalawi, and D. C. Jeutter, "A linear permanent magnet generator for powering implanted electronic devices," *IEEE Trans. Power Electron.*, vol. 26, no. 1, pp. 192–199, Jan. 2011.

[7] S. W. Zhao, N. C. Cheung,W. Gan, and J. Yang, "High-precision position control of a linear-switched reluctance motor using a self-tuning regulator," *IEEE Trans. Power Electron.*, vol. 25, no. 11, pp. 2820–2827, Nov.2010.

[8] C. Cheun, "TSK-type self-organizing recurrent-neural-fuzzy control of linear microstepping motor drives," *IEEE Trans. Power Electron.*, vol. 25, no. 9, pp. 2253–2265, Sep. 2010.

[9] C. Cheun, "Supervisory interval type-2 TSK neural fuzzy network control for linear microstepping motor drives with uncertainty observer," IEEE Trans. Power Electron., vol. 26, no. 7, pp. 2049–2064, Jul. 2011.

[10] J. Prudell, M. Stoddard, E. Amon, T. K. A. Brekken, and A. Jouanne, "A permanent-magnet tubular linear generator for ocean wave energy conversion," *IEEE Trans. Ind. Appl.*, vol. 46, no. 6, pp. 2392–2400, Nov. 2010.

[11] C. Kral, A. Haumer, and S. Lee, "A practical thermal model for the estimation of permanent magnet and stator winding temperatures," *IEEE Trans. Power Electron.*, vol. 29, no. 1, pp. 455–464, Jan. 2014.

[12] Cecilia Boström, Rafael Waters, Erik Lejerskog, Olle Svensson, Magnus Stålberg, Erland Strömstedt, and Mats Leijon, "Study of a Wave Energy Converter Connected to a Nonlinear Load," *IEEE Journal of oceanic engineering*, Vol. 34, No. 2, pp. 123–127, Apr. 2009.

Power Generating Slabs: Lost energy conversion of human locomotive force into electrical energy

* Rajesh Kumar Datta, [1], Sazid Rahman[2]

[1] Department of Electrical and Electronic Engineering, Chittagong University of Engineering and Technology, Bangladesh
[2] Department of Electrical and Electronic Engineering, Chittagong University of Engineering and Technology, Bangladesh
*rajeshcuet10@gmail.com

***Abstract*— In the world of modern technology, newer sources of energy and new methods of power generation are two important area of interest for researchers and engineers. Recently a new method is established to generate power from the lost energy of a human during his/her locomotive period. A piezoelectric sensor based costly product is available in some developed countries which can generate power from human locomotive force, but it is not suitable for countries like Bangladesh where power demand is very high but economy is not highly developed. In this paper we will explain the design and construction method of a "Power Generating Slab" which can be used to generate power by establishing anywhere of the walking zone of human (roads, stairs, pavements, dais etc). This is done in a low cost process with locally available equipments (micro-generator, iron plate, rack pinion gear, free-wheeling system, spring etc). These slabs can easily sustain human weight when people walk along them. With a small deformation of spring, the weight (mechanical energy) of a human body is converted into electrical energy by rotating a micro-generator with the help of a 'rotating shaft' coupled with it. From each foot step almost 10V-12V is generated. This energy is stored in a rechargeable battery which can be used as a power source to drive loads. Assembly of some power generating slabs can give better result. This is a method which is reusable and it will run for a longer period.**

***Index Terms*— Power Generating Slab, Mechanical force, Locomotive force, Foot step, Electrical energy.**

I. Introduction

Lost energy conversion of human locomotive force into electrical energy is a relevant and important proposal of power production for highly populated countries of the world. For a healthy human being, walking is the most common activity in daily life. During the movement of a person, energy is lost in the form of impact or vibration on the walking surface. [4] We can convert this lost energy in the usable form of power. The method we have implemented to make "Power Generating Slab" needs some simple mechanical set ups. These slabs can be embedded in the places where there are continuous rush of people; like roads, footpaths, railway platform, stairs etc. When any person will step on the top plate of the slab, the body weight will compress the springs under the slab and the top plate will dip down a little bit. This small downward displacement of the top plate will rotate a micro generator's shaft fitted at the bottom plate and produce electrical energy. When the person will take off his foot from the slab, the springs will regain the position of the top plate. This produced electrical energy will be stored to use later. More human movement will generate more power from these slabs. In overcrowded countries like Bangladesh this methodology will be very much efficient and helpful to reduce power crisis.

II. Design Principle of Power Generating Slabs

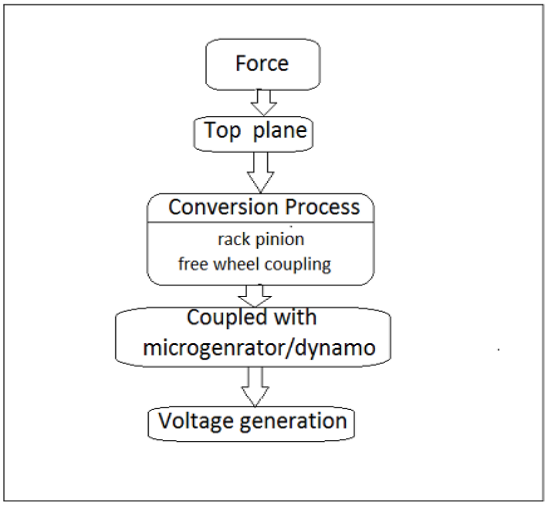


Fig 1: Block diagram of Power Generating Slab

III. Construction and Specification

Top Plane:

This is the upper portion of the slab. In our design we have made this portion with the iron plate. It can be built with other materials too. The main condition of this plate is to sustain the weight of any person walking along it. Footsteps will work as mechanical force.

Length: 50cm, Width: 30cm, Material: Iron

Fig 2 : Top Plane

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

Rack and pinion:

This part of the device is attached with the top plate. The mechanical force is converted into a rotational force with the help of rack and pinion. Starting point of rack is welded with the iron plate and the bottom is attached with a pinion along with a shaft.

Length (Rack): 18cm, Radius (Pinion): 03cm

Fig 3: Rack

Fig 4: Pinion

Springs:

Below the top plane layer there are 4 helical springs [5]. These springs are one of the most important parts of this design. When the top plate is loaded with human foot step these springs are distorted slightly and when the load is removed they recover their original shape. This displacement does not make any discomfort in the human walking zone.

Spring type: Helical, Number of springs: 04, Length: 3cm, Maximum weight tested: 90 kg, Maximum displacement: 1cm.

Fig 5: Helical Spring

Shaft with freewheeling bearing:

A shaft is coupled with the pinion and the gear mechanism. The shaft is clamped between the iron frame works. And the freewheeling type bearings are attached on the shaft. The free-wheeling bearings rotate only in one direction.

Length: 50cm

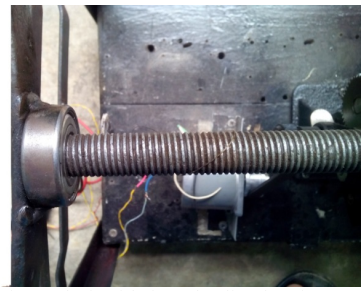


Fig 6: Shaft with freewheeling bearing

Gear mechanism:

This mechanism can be used in many different forms. The arrangement we have used for our device is attached with the shaft with the help of sprocket arrangement. There is a smaller output shaft in this mechanism which converts the total applied force into the rotational force.

Sprocket diameter: 12.5cm

Fig 7: Coupling system with Sprocket

Fig 8: Gear arrangements

Micro- Generator (dynamo):

In this slab we have used a 12V, 0.7A dynamo as micro generator. A dynamo, another form of electrical generator produces direct current. This dc current is produced with the help of a commutator. In dynamo there are rotating coils and a magnetic field. When the shaft of the dynamo rotates with the help of any external force a pulsating direct current is produced because of Faraday's law of induction[3].The constant magnetic field is provided by a stator and a set of armature turn in that magnetic field. When the armature cuts the magnetic field, an alternating electric current is induced in the wire. Commutator works like a rotary switch. Connection of the windings is reversed to the external circuit by the commutator when the direction of the potential reverses. As a result the direction of the output current remains always same. Thus pulsating direct current is produced instead of alternating current from the dynamo.

Fig 9: Coupled dynamo with a belt

Total mechanical setup:

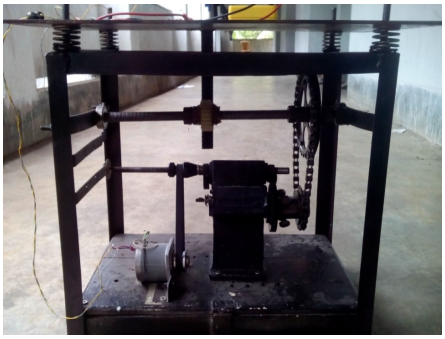


Fig 10: Power Generating Slab

Output:
Maximum voltage in one foot step: 12v

Voltage type: Pure dc

Fig 11: Output Voltage

IV. Charging Arrangements

The produced energy from the slab must be stored in a rechargeable battery. So a charge controller is needed at the output of the slab. For charging of a battery we have used a constant voltage method. The voltage level produces from slab varies time to time. So a charge controller is needed to keep a constant voltage to charge a battery. The controller we have used can charge a battery at a constant voltage of about 13.8V. The voltage that is generated from the human foot step is firstly stored in a capacitor. This capacitor can be of any value as much as desired. The charge controller circuit keep the voltage constant at the output. This helps to charge a battery properly. When the battery is fully charged, charge controller does not pass any voltage to the battery terminal.

V. DATA

Number footsteps	Average power/foot step		
	Capacitor voltage	current	Average Power/foot step
5	32V dc	0.7A dc	4.48 Watt
10	61V dc	0.7A	4.27 Watt
15	85Vdc	0.7A	3.96 Watt

This data table shows the relationship between the number of footsteps and average produced power. This observation is done with a person of 60kg weight.

.

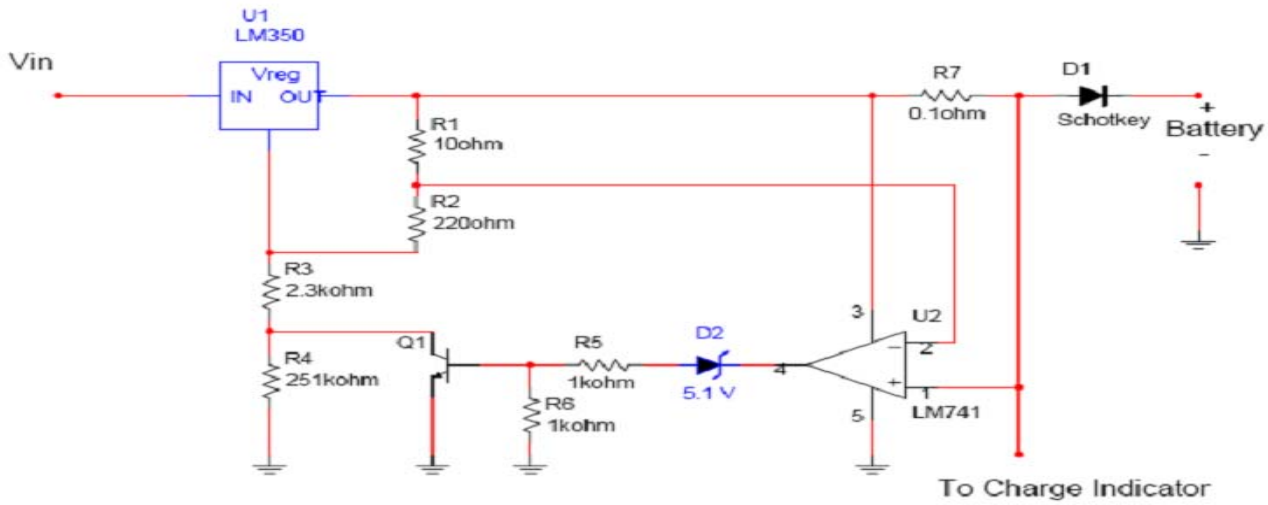


Fig 12: Charge controller circuit [1]

V. RELATION BETWEEN FORCE AND VOLTAGE

The relation between mechanical force and electrical voltage is proportional in these slabs. The force is applied in the form of human weight. More the weight is applied, more the voltage is generated. The force strength is depended on the spring's strength. Springs consume the mechanical energy obeying Hook's Law in their elasticity limit

$$F = -kx$$

Where

x is the displacement of springs from its normal length.

F is the resulting force vector, the magnitude and direction of the restoring force the spring exerts.

k is the spring constant , a constant that depends on the spring's material and construction[7].

This force **F** is applied to the freewheeling bearing which produces a rotational energy. The amount of energy stored in a flywheel is proportional to the square of its rotational speed.

$$E_k = \frac{1}{2} I \omega^2$$

Where,

ω is the angular velocity

I is the moment of inertia of the mass about the centre of rotation. [8]

Thus the rotational energy is converted into electrical energy with dynamo rotation.

So, the relation is actually

Force ∞ rotational energy ∞voltage

VI. ADVANTAGES

1. Lower cost.
2. Can be made with locally available materials.
3. Power can be generated in many forms.
4. Easy construction.

VII. APPLICATIONS

1. For emergency backup power.
2. To run small electrical equipments at the time of load shedding.
3. Street light
4. For charging purpose.
5. Different ways of applications like stairs, classroom daises, parks etc.

VIII. DIFFERENCE WITH EXISTING SYSTEM

The existing system in market is based on the piezoelectric sensor which is a costly material. So the price of the total device is very high ([2], [6]). And the materials to make the slab are not available in all places. In our method, we have used the materials which are easily available in all places and the mechanical system can be easily made in workshops. Our design can be reused and if any kind of hamper is occurred it can be easily solved. The flooring system is a prime concern of this project. We have successfully made the system in which the flooring will not be a discomfort to walk.

IX. CONCLUSION

In modern world the demand of renewable energy is increasing rapidly. Power generating slab doesn't need any fuel to produce power. It just uses the energy that is lost by a human during walking. These slabs can be helpful in reducing power crisis. If these slabs are produced industrially and further research is done on materials and the design, the efficiency will be a lot better. Then these slabs can be viewed as a next promising source of generating power.

REFERENCES

[1] Abhik datta , *Design of a Lead Acid Battery Charger System;* National Institute of Technology, Rourkela, 2009

[2] http://urbantimes.co/2012/10/footsteps-power-city-sustainably-pavegen-paving-tiles-smart/

[3] http://en.wikipedia.org/wiki/Dynamo

[4] S.S.Taliyan, B.B. Biswas, R.K. Patil and G. P. Srivastava, T.K. Basu , "*Electricity From Footsteps*" , research article of general interest ,ISSUE NO 313 • MAR. - APR. 2010

[5] http://engg-learning.blogspot.com/2011/03/helical-spring.html

[6] Pramathesh.T, Ankur.S "*Piezoelectric Crystals : Future Source of Electricity" International Journal of Scientific Engineering and Technology (ISSN : 2277-1581) Volume 2 Issue 4, pp : 260-262*

[7] http://en.wikipedia.org/wiki/Hooke's_law

[8] http://en.wikipedia.org/wiki/Rotational_energy

Minimization of Output Fluctuations of Hybrid Power System by Using Modified Pitch Controller

M.R.I. Sheikh,[1] D. Datta,[2,*] and M.S.R. Ashrafi[3]

[1,2,3]EEE Department, Rajshahi University of Engineering & Technology
Rajshahi-6204, Bangladesh
[*]dristidatta@gmail.com

Abstract—**The rapid depletion of fossil fuel resources and environmental concern has given awareness on generation of renewable energy resources. Among the various renewable resources, wind energy seems to be promising solution to provide reliable power supply with improved system efficiency. As wind turbine output is proportional to the cube of wind speed, the wind turbine generator output fluctuates due to wind speed variation. This study investigates the influences of a new pitch controller & governor control system model on frequency of power system, when variable load has been considered. It is observed that system frequency can be maintained in acceptable level by using newly designed pitch controller & governor control system when several interconnected synchronous generators (SGs) are operated in different governor control modes with high wind power penetration, when the total capacity of SGs is considered as 100 MVA. Real wind speed data have been considered for this study. PSCAD/EMTDC software is used for simulation analyses.**

Index Terms—**Wind turbine; new pitch controller; variable load; system frequency; wind power capacity; governor control system model.**

I. Introduction

Global energy consumption has increased by leaps and bounds to improve our living standards. But it is a matter of great regret that, the world has limited amount of fossil fuel and nuclear power resources ([1], [2]). So, the growth of wind energy around the world in recent years has been consistently impressive.

In the conventional operation of wind power generators, when the wind speed is between the rated speed and the cut out speed, the wind power generator output is controlled at the rated value by a pitch control system [3]. On the other hand, when the wind speed is between the cut in speed and the rated speed, the blade pitch angle is maintained constant (= 0 deg), in general for the wind turbine to capture the maximum power from the wind turbine [4]. If the power capacity of wind generators becomes large, wind generator output can have an influence on the power system frequency ([5]-[8]).

From the previous study ([1], [3], [4]), it is seen that when single SG or several SGs are connected with networks, only governor control system model & conventional pitch controller cannot maintain the power system frequency to the desired level and severe situation occurs when wind power penetration becomes 10% of the total capacity. But this study investigates the influence of a new modified pitch controller & governor control system and it is observed that system frequency can be maintained in acceptable range when several interconnected SGs are operated with different governor control modes where the load is considered variable and wind power capacity is 10% of the total capacity.

In [5] the authors applied a new pitch control system where the reference power is considered variable. Thus the frequency can be controlled up to a certain limit of wind generator capacity by losing some amount of energy. However, in our study energy can be the same, which is more economical. In fact voltage can not be controlled by their pitch control system.

As the wind power penetration increases day by day, this study plays a vital role for power system application. It is seen from the study that the modified pitch control and governor control system model can maintain the frequency as well as voltage in acceptable limit with high wind power penetration. As a result, energy storage devices can be avoided which reduce the total cost of the system.

II. Model System For Simulation Analysis

The model system used in the simulation analysis is shown in Fig. 1. The model system consists of five wind generators (IG), a hydropower generator, HG (a salient pole synchronous generator, SG1), two thermal power generators, TG (cylindrical type synchronous generators, SG2 and SG3), a nuclear power generator, NG (a cylindrical type synchronous generator, SG4) and three loads. SG1 and SG3 are operated under load frequency control (LFC), SG2 is under governor-free (GF) control and SG4 is under load limit (LL) operation [9] which are all clearly indicated in the figure below.

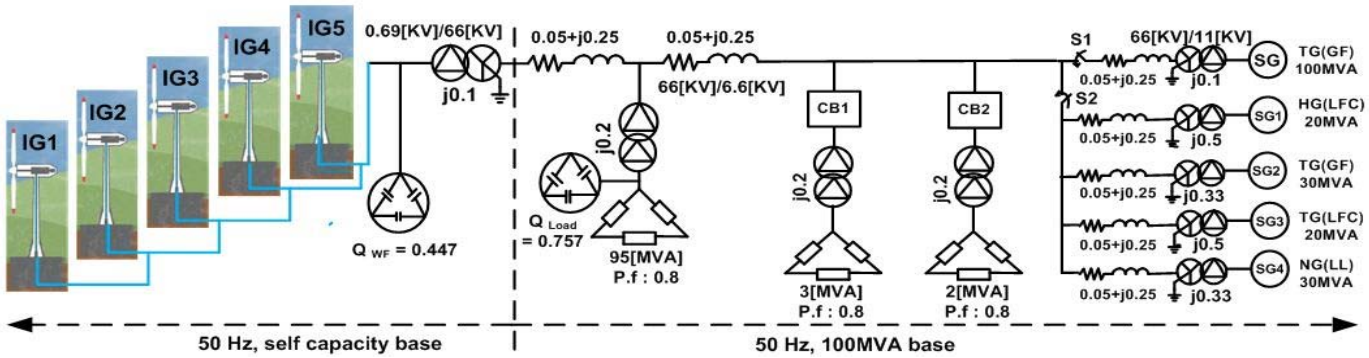


Fig. 1 Model circuit.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

Q_{WF} and Q_{Load} are capacitor banks where Q_{WF} is used at the terminal of IG to compensate the reactive power demand of wind generator at steady state. The value of the capacitor is chosen so that the p.f. becomes unity, when the wind generator operated in the rated condition. On the other hand, Q_{Load} is used at the terminal of load to compensate the voltage drop by the impedance of transmission lines. Here circuit breakers CB1 and CB2 are used to control the load as a variable load. Core saturations of induction generator and synchronous generators are not considered for simplicity. The initial power flow and initial conditions are taken from [3].

III. Synchronous Generator Model

A. Governor for Hydro and Thermal Generators[3]

The governor models used in the simulation analyses are shown in Fig. 2 and Fig. 3 respectively. The values of 65M and 77M of different SGs are shown in Table I. Where, Sg: the revolution speed deviation (pu); 65M: the initial output (pu); 77M: the load limit (65M + rated MW output × PLM (%)); PLM: the spare governor operation (%); Pm: the turbine output (pu). PLM for SG2 is set 5% and for SG4 PLM is set 20% because the nuclear generator output (SG4) is controlled constantly (LL, load limit operation).

For governor-free (GF) operation, When PLM > 0

65M = the initial output (pu)

77M = 65M + rated MW output × PLM (%)

For load limit (LL) operation, When PLM < 0

65M = 77M + rated MW output × PLM (%)

77M = the initial output (Pu)

Sg is set zero for SG1 and SG3 because these generators are operated under LFC to control frequency fluctuations with a relatively long period.

B. Load Frequency Control Mode [4]

In the Load Frequency Control (LFC) model, the output power signal is sent to each power plant when the frequency deviation is detected in the power system. After that, governor output value (65M) of each power plant has changed by LFC signals and then the output of power plant is also changed. The LFC model is taken from [4].

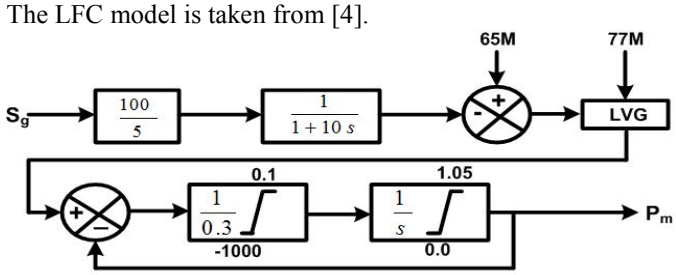


Fig. 2 Model of hydro governor.

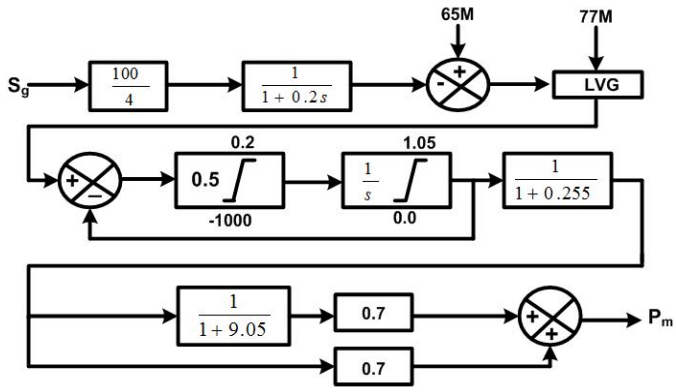


Fig. 3 Model of thermal governor.

TABLE I
Values of 65M and 77M for Multiple Generators

SG1 (Hydro)		
Frequency Control	65M	77M
LFC	LFC signal	1
SG3 (Thermal)		
Frequency Control	65M	77M
LFC	LFC signal	1
SG2 (Thermal)		
Frequency Control	65M	77M
GF	Starting Value=0.653 Running adjustment If f > 50.2 then 65M = 0.70 If f < 49.8 then 65M = 0.62	0.767
SG4 (Nuclear)		
Frequency Control	65M	77M
LL	0.96	0.80

IV. Modified Pitch Controller

The newly designed pitch controller is shown in Fig. 4 and its purpose is to maintain the output power of the wind generator at rated level by controlling the blade pitch angle of turbine blade when the wind speed is over the rated speed. Generally, the blade pitch operation system is complicated, but this paper simulates the pitch operation system by using a first order time delay system with time constant T_w = 5 sec. In addition, the pitch angle can not be changed instantly due to the rotational inertia of blade and mechanical limitations. Therefore, the rate of change of pitch angle is limited to 10 degrees per second in the simulations. Also this controller can detect the frequency levels whether it is beyond or bellow the acceptable level. Thus this controller can also generate new pitch angle for the wind turbine if the frequency is beyond the acceptable level. Beside this, the modified pitch controller can adjust the initial input to the thermal governor i.e. 65M (shown in Fig. 2 and Table I) by comparing the following (1) and (2) to maintain frequency to the desired acceptable range. The comparator and control circuit as shown in Fig. 5 is used to adjust the initial input i.e. 65M for the governor with the addition to pitch controller.

If f > 50.2 then 65M = 0.70 (1)

Or If f < 49.8 then 65M = 0.62 (2)

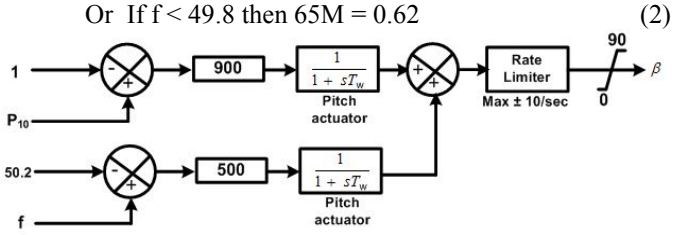


Fig. 4 Modified pitch control system.

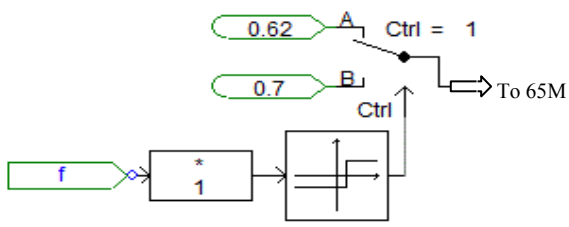


Fig. 5 PSCAD control box.

V. Simulation Analyses

In this paper, simulation analyses have been compared by using both conventional and newly designed pitch controller and the analyses have been carried out to investigate the performance of the new pitch controller & governor control system with high wind power penetration using real wind speed data, which are measured in Hokkaido Island, Japan. Simulation combination shows in Table II. In this study, it is considered that five wind generators experience different five wind data as shown in Fig. 6. The simulation analyses have been performed by using PSCAD/EMTDC [10].

TABLE II
Simulation Pattern

No. of Cases	SG	Rating (MVA)	Total Capacity (MVA)
Case 1	TG(GF)	100	100
Case 2	HG(LFC) TG(GF) TG(LFC) NG(LL)	20 30 20 30	100

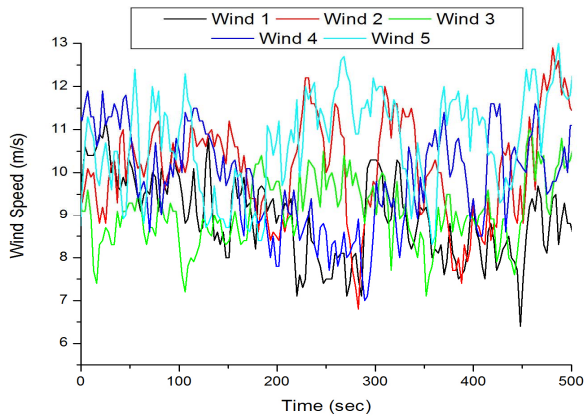


Fig. 6 Wind speed data.

Fig. 7 shows the grid voltage, where the fluctuation is comparatively less for modified pitch controller then conventional pitch controller. Fig. 8 shows the total load that varies between 74MW to 80 MW and Fig. 9 represents the synchronous generator power output which fluctuates more at the time of load change. From Fig. 10 it is seen that when modified pitch controller is used, frequency fluctuation is comparatively less but energy remains almost the same (from Fig. 11) and the variation of pitch angle for wind 5 corresponds to IG5 is shown in Fig. 12.

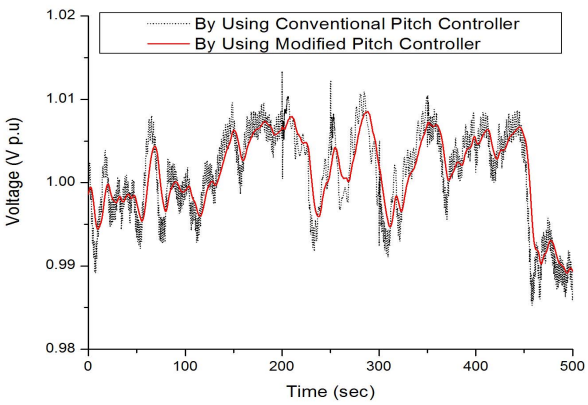


Fig. 7 Response of output voltage [case 1].

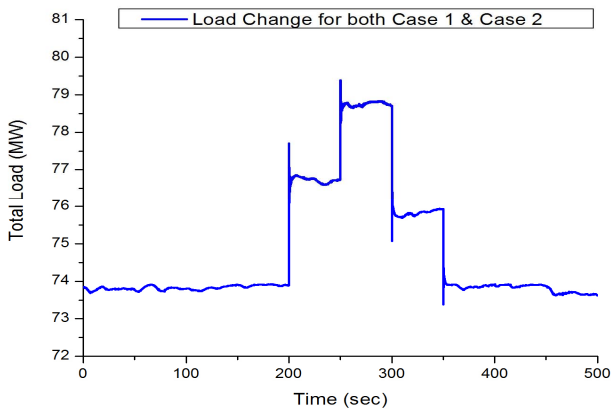


Fig. 8 Response of load change.

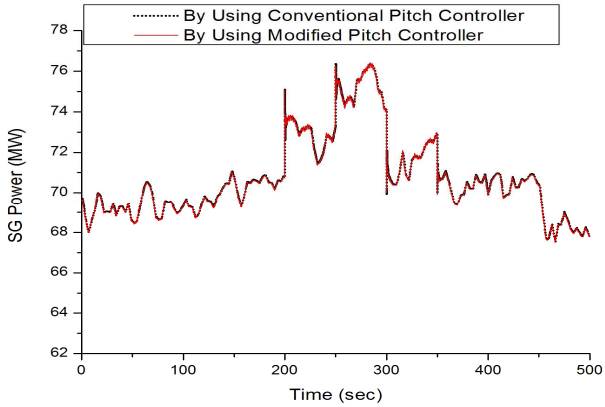


Fig. 9 Response of SG power [case 1].

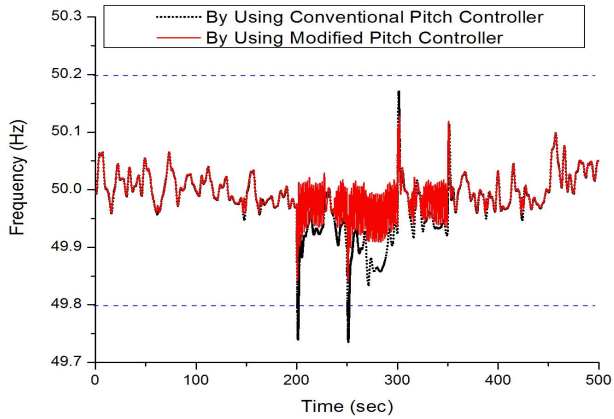


Fig. 10 Response of output frequency [case 1].

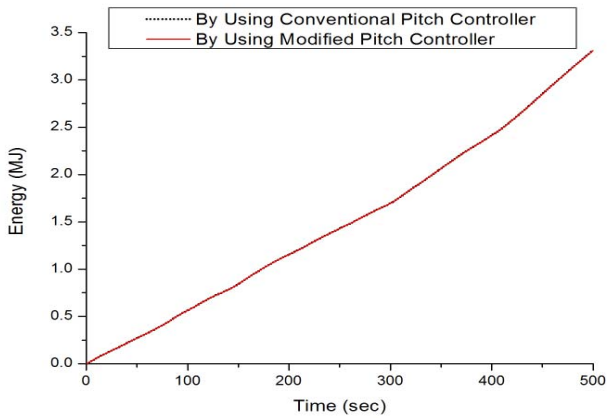


Fig. 11 Response of output energy [case 1].

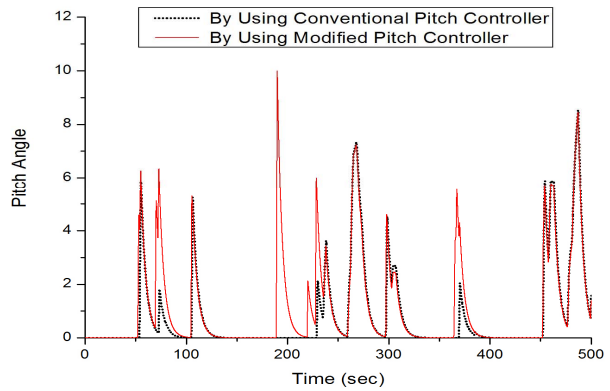


Fig. 12 Response of pitch angle for IGs [case 1].

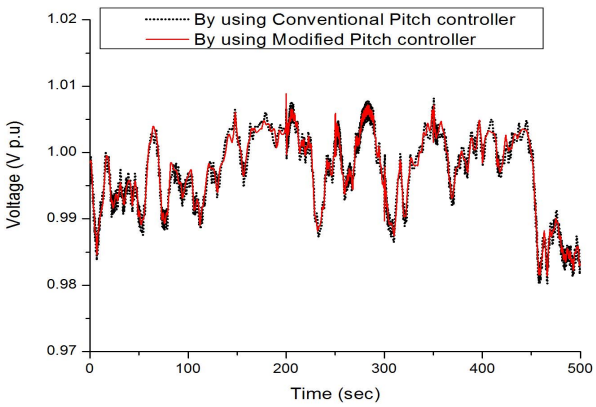


Fig. 13 Response of output voltage [case 2].

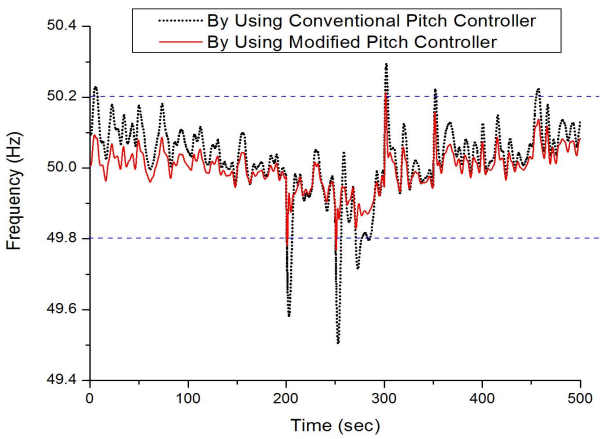


Fig. 14 Response of output frequency [case 2].

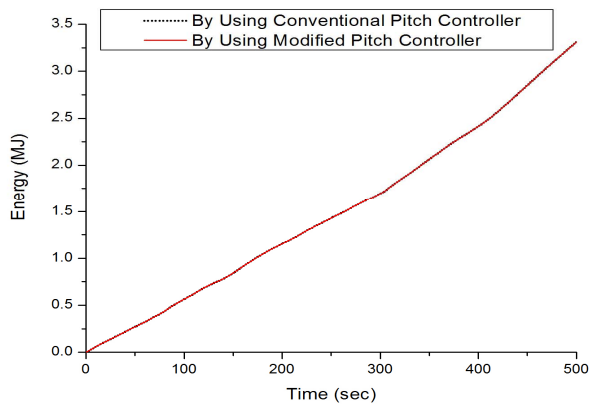


Fig. 15 Response of output Energy [case 2].

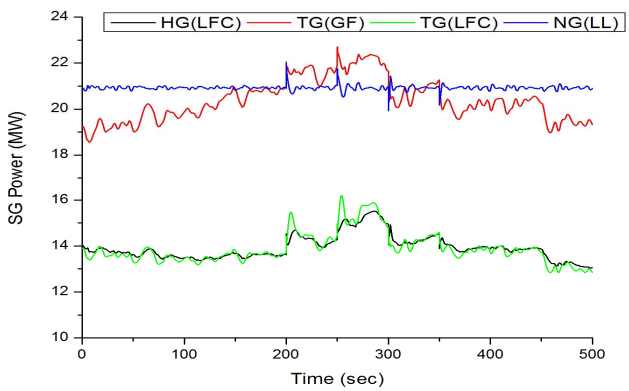


Fig. 16 Response of SGs real power [case 2].

Fig. 13 and Fig. 14 represent the output voltage and frequency which fluctuate less without any energy loss (from Fig. 15) when modified pitch controller is used. The outputs of four SGs are shown in Fig. 16.

VI. Conclusions

A permissible range of power system frequency deviation provided by the most of the developed countries is within ± 0.2 [Hz]. As frequency is one of the measures to determine the quality of electric power, this study focuses on the influences of a modified pitch controller and governor control system model when variable load and different types of governor control system are considered. It is confirmed from the above simulation results that the modified pitch control and governor control system are able to maintain the frequency fluctuations within acceptable level when high wind penetration is considered. It is also cleared that the newly designed control system can maintain frequency without any loss of energy. Thus energy storage device can be neglected.

References

[1] E. Haque and M.R.I Sheikh, "Grid frequency analysis with the issue of high wind power penetration", *International Conference on Electrical Information and Communication Technology,* KUET, Khulna, Bangladesh, Dec. 2013.

[2] M.R.I. Sheikh, M.S.R. Ashrafi, and D. Datta, "Co-ordinated Frequency Control of Power System with High Wind Power Penetration", *International conference on Electrical Engineering and Information & Communication Technology,* MIST, Dhaka, Bangladesh,12-14 Apr. 2014.

[3] M.R.I. Sheikh, M.M. Haque, and A. Hossain, ''Performance of governor system on minimizing frequency fluctuations with wind power generation'', *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 2, no. 1, pp. 46--56, Feb. 2012.

[4] M.R.I. Sheikh, R. Takahashi, and J. Tamura, ''Study on frequency fluctuations in power system with a large penetration of wind power generation'', *International Energy Journal* , vol. 1, no. 1, Mar. 2011.

[5] T. Yamazaki, R. T. Murata, T. Tamura, Fukushima, E. Ssano, K. Shinya, and T. Matstumoto, ''Smoothing control of wind generator output fluctuations by new pitch controller'', *IEEJ Transactions on Power and Energy*, vol. 129, no. 7, pp. 880-888,Jul. 2009.

[6] C Luo, and B. T Ooi, ''Frequency deviation of thermal power plants due to wind farms'', *IEEE Transaction on Energy Conversion*,'' vol. 21, no. 3, pp. 708-716, Sep. 2006.

[7] C. Carrillo, A.E. Feijoo, J. Cidras, and J. Gonzalez, ''Power fluctuation in an isolated wind plant'', *IEEJ Transactions on Power and Energy,* vol. 19, no. 1, pp. 217-221, Mar. 2004.

[8] H. Umeki, R.Takahashi, T. Murata, and J. Tamura, ''Smoothing control of wind generator output fluctuations by new pitch controller'', *The Int. Conf. of Electrical Machines and Systems(ICEMS),*LS3C-1,2006.

[9] Technical Reports, ''Standard models of electrical power system,'' *IEE of Japan*, vol. 754, pp. 40-43.

[10] Manitoba HVDC Research Center. 1994. PSCAD/EMTDC Manual.

Hybrid Power System: Biomass and Diesel

A.H.M. Kamal,[1,*] Md. Mahbubur Rahman,[1] Md. Mohaiminul Haque, [1] Salman Sadiq, [1] and Aminul Hoque[1]

[1]Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology
Dhaka-1205, Bangladesh
*jimmybuet@gmail.com

Abstract—**A hybrid energy system consists of two or more renewable or conventional energy sources used together to provide efficient power supply. In this paper, a new and evolving hybrid power system has been designed by integrating different types of biomass and diesel. This system can ensure an uninterrupted power generation irrespective of the weather condition and supply of biomass. This system can also ensure maximum utilization of biomass energy sources which will reduce the dependency on the limited fossil fuels like oil and coal those are environmentally harmful.**

***Index Terms*—Hybrid power system, biomass, diesel, renewable diesel.**

I. Introduction

Hybrid energy technology is the latest innovation in the field of energy sector which is often done either to reduce the cost of generating electricity from fossil fuels or to provide backup for a renewable energy system that ensures continuity of power supply when the renewable energy source fluctuates. Hybrid energy usually consists of one conventional energy source and another would be renewable energy source like solar photovoltaic, wind, biomass, hydro etc.

However, a lot of requirements have to be considered first for the implementation of hybrid system. The most important factors are time, location and power demand. Location associates information about climate, energy sources availability and environment conditions. This information is very important to decide what kind of renewable source can be chosen [1]. Biomass-diesel combination can be a very effective choice because biomass is an easily available, eco-friendly renewable source of energy and it does not vary considerably with the change of weather, temperature or season. Moreover, it can address limitations in terms of fuel flexibility, efficiency, reliability, emissions and economics. In this paper, the proposed biomass – diesel hybrid system design is more applicable where different types of biomass are available and can easily be processed. This can emerge as a pragmatic solution to the energy crisis for any agricultural and overpopulated countries where amount of solid waste generation is abundant.

II. Energy Sources

There are mainly two types of energy sources, non-renewable energy sources and renewable energy sources.

A. Non-renewable Energy Sources

Non-renewable energy sources are not replaced or replaced very slowly by natural processes such as oil, coal and coal derivatives, natural gas, diesel and radioactive atoms (uranium). Most of them dispose great amount of green-house gases while burning and responsible for global warming and climatic change.

B. Renewable Energy Sources

Renewable energy is generally defined as energy that comes from resources which are continually replenished on a human timescale such as sunlight, wind, rain, tides, waves, biomass and geothermal heat [2]. Biomass is a renewable source of energy because the waste residues will always exist in terms of scrap wood, mill residuals or forest resources. As plants use atmospheric CO_2 in the process of photosynthesis which completes the carbon cycle, so it is renewable. The carbohydrate produced by photosynthesis is in fact the source of biomass.

III. Biomass Quality Analysis

Quality analysis of biomass is an important parameter to use it as an effective energy source. Biomass qualities are affected by mainly two factors- internal factors and external factors.

A. Internal Factors

Internal factors are the constitutional factors that affect the variation of biomass heating value radically. These are fixed carbon, volatile substances, ash content and water content.

Composition of twenty different biomass samples has been observed which can be used as fuel for the power generation [3], [4].

TABLE I
Different Types of Biomass and Their Composition

Biomass Type	Fixed Carbon (%)	Water Content (%)	Ash Content (%)	Volatile Matter (%)
1.Wood(dry)	14.5	7.8	0.2	77.5
2.Sugarcane	11.1	10.4	1.9	76.6
3.Coconut shells	22.0	4.4	3.1	70.5
4.Musturd husk	22.0	5.6	3.9	68.5
5.Bamboo whole	15.2	13.0	0.8	71.0
6.Wheat straw	16.3	10.1	6.4	67.2
7.Rice straw	14.4	7.6	18.6	59.4
8.Rice husks	17.2	10.6	16.1	56.1
9.Miscanthus grass	14.0	11.4	2.7	71.9
10.Buffalo gourd	12.3	10.0	4.2	73.5
11.Eucalyptus bark	15.1	12.0	4.2	68.7
12.Wood residue	12.2	26.4	4.0	57.4
13.Sawdust	9.3	34.9	0.7	55.1
14.Fir mill residue	6.5	62.9	0.2	30.4
15.Forest residue	7.3	56.8	1.4	34.5
16.Furniture waste	11.8	12.1	3.2	72.9
17.Wood agri-residue	12.7	30.3	2.3	54.7
18.Biomass mixture	16.5	8.8	11.4	63.3
19.Mixed waste paper	6.8	8.8	7.6	76.8
20.Land-clear wood	7.0	49.2	8.4	35.4

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

B. External Factors

Extensive care should be taken for external factors. Among these external factors, moisture content is the important one. Most boilers are designed to operate with fuel of limited moisture range. If the fuel is used outside this range, the boiler will either operate inefficiently leading to increased emissions or trip out the control system and fail to function at all. It is very important to find out the specification of correct moisture content range of boiler from the manufacturer. In the case of purchasing wood fuel by weight, moisture content is critical parameter to know how much are paying for energy [5].

Based on biomass quality analysis and availability of sources, biomass can be classified into four categories-

1) Biomass Type-1: Biomass having moisture content less than 30% has been classified as Biomass Type-1.

2) Biomass Type-2: Biomass having internal moisture content more than 30% has been classified as Biomass Type-2.

3) Biomass Type-3: All types of sewage and human or animal waste that are used to produce biogas have been classified as Biomass Type-3. Biogas is produced by anaerobic digestion with anaerobic bacteria or fermentation of biodegradable materials such as manure, sewage, municipal waste etc. It is primarily methane (CH_4) and carbon dioxide (CO_2) and may have small amounts of hydrogen sulphide (H_2S), moisture and siloxanes [6]. The gases methane, hydrogen and carbon monoxide (CO) can be combusted or oxidized with oxygen. This energy release allows biogas to be used as a fuel.

4) Biomass Type-4: Special type of biomass which can produce renewable diesel has been classified as Biomass Type-4. It can be produced either by Engineered Bacteria (*Escherichia coli*) or by Hydropyrolysis [7]. This green diesel can be used alone or blended with petrodiesel in any proportions as fuel.

TABLE II
BIOMASS CLASSIFICATION FOR HYBRID SYSTEM DESIGN

Biomass Type	Description	Examples
1	moisture content less than 30%	wood, wheat straw, rice husk, coconut shell, sugarcane etc.
2	moisture content more than 30%	sawdust, fir mill residue, forest residue, agricultural residue, land clearing wood etc.
3	sewage, human or animal waste	cow dung, poultry litter etc.
4	producing green diesel	vegetable oil, animal fat (algae, jatropha etc.)

Fig. 1 shows the scatter-diagram of different biomass samples where some lies in Biomass Type-1 zone and other in Biomass Type-2 zone.

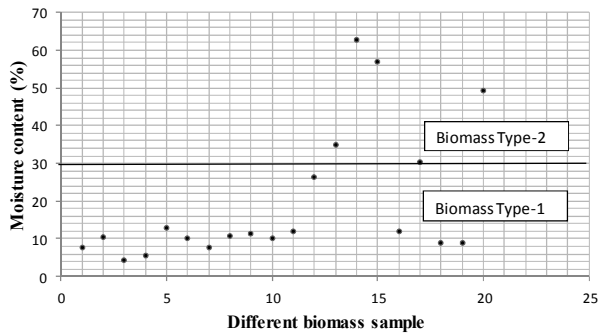


Fig. 1 Scatter-diagram of different biomass samples.

IV. DESCRIPTION OF THE DESIGN

Fig. 2 shows the design of a hybrid power system where different components have been integrated to produce electricity from biomass and diesel with reliability and efficiency.

Biomass Type-1 which is readily combustible is provided into the combustion unit with sufficient air supply. On the

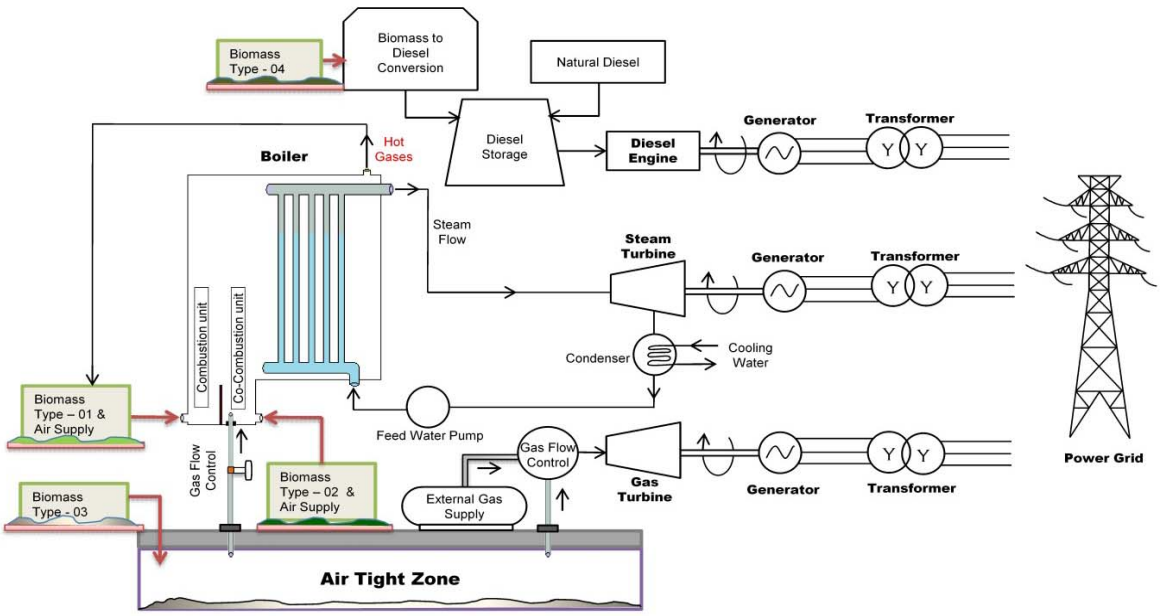


Fig.2 Biomass-diesel hybrid power system design.

other hand, Biomass Type-2 and biogas from airtight dome are provided into the co-combustion unit with sufficient air supply. In air tight zone, biogas is produced using Biomass Type -3 by anaerobic digestion process or fermentation. Biogas produced in this zone is supplied to both co-combustion unit and gas turbine via gas flow control valves.

Due to biomass quality variation, heat generation rate in co-combustion chamber differs significantly. To maintain constant heat generation in co-combustion unit, a valve senses temperature and pressure of the chamber and allows the effective flow rate of gas using the feedback. Another valve is interfaced between air tight zone and gas turbine. An external gas supply is also connected to the control valve to ensure necessary flow rate of gas in the gas turbine.

Steam from boiler is used to drive the steam turbine. A condenser is used for the purpose of condensing the exhaust steam from the turbine to reuse in the cycle and to maximize turbine efficiency by maintaining proper vacuum. A pump is used to feed the condensed water into the steam boiler. Moreover, the exhaust hot gases from the boiler unit are utilized for drying up Biomass Type-1 hence reducing the inherent external moisture content in it.

To ensure scheduled power supply to the power grid, a diesel backup system has been introduced in the design. Biomass Type-4 is used to produce diesel known as 'renewable diesel' by different bio-chemical processes. It is both non-toxic and renewable as it essentially comes from plants and animals which can be replenished through farming and recycling.

Other conventional components of power generation system are- generator, transformer, diesel engine, control unit etc.

V. Conclusions

The proposed hybrid design presented in this paper is less susceptible to weather variation than solar and wind based power system because solar energy is always dependent on intensity of day light whereas wind energy varies significantly with seasonal variation. The integration of the presented hybrid system with solar or wind power system can provide a cost effective solution throughout the year. Moreover, as the hybrid power system contains conventional components like boiler, turbine, generator etc., it is easy to implement practically.

References

[1] E. F. F. Ribeiro, A. J Marques Cardoso, and C. Boccaletti, "Uninterruptible Energy Production in Standalone Power Systems for Telecommunications," International Conference on Renewable Energies and Power Quality (ICREPQ'09), Valencia (Spain),15 th to 17 th April, 2009.

[2] (2013)The myth of renewable energy | Bulletin of the Atomic Scientists. [Online]. Available: Thebulletin.org. 2011-11-22.

[3] Stanislav V.Vassilev, David Baxter, Lars K.Andersen & Christina G.Vassileva (2009) "An overview of the chemical composition of biomass". [Online]. Available: http://www.sciencedirect.com/ science/article/pii/S0016236109004

[4] Ayhan Demirbas (2003) "Combustion characteristics of different biomass fuels". [Online]. Available: http://www.sciencedirect.com/ science/article/pii/ S0360128503000789

[5] (2013)Moisture content-Biomass Energy Center. [Online]. Available: www.biomassenergycentre.org

[6] (2014)Biogas.[Online].Available: http://en.wikipedia.org/wiki/Biogas

[7] (2014)Diesel Substitutes: Biodiesel and Biomass-based Diesel. [Online].Available:http://www.betalabservices.com/biofuels/biomass-based-diesel.html

Design and Optimal Cost Analysis of Hybrid Power System for Kutubdia Island of Bangladesh

Amit Kumar Sikder,[1,*] Nishat Anjum Khan,[1] and Aminul Hoque[1]
[1] Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka-1205, Bangladesh
[*] amitkumarsikder@gmail.com

Abstract— **Hybrid power system is the newest idea of utilizing renewable energy resources for generating electricity. Power generation largely depends on fossil fuel and as the total amount of reserved fossil fuel is decreasing we have to search for new resources for power generation. Hybrid power system usually demonstrates lower cost of production and higher reliability than the power system only using one renewable source. Geographically, Bangladesh is situated in an area having many renewable resources. Utilization of these resources could help to minimize the power crisis of Bangladesh. As Bangladesh has a large coastal area, we can have moderate wind speed and tidal wave throughout the year. A hybrid power system utilizing wind and tidal energy will be effective to generate power in the coastal areas. This paper shows the design and optimal cost analysis of a hybrid power system using wind and tidal energy in the Kutubdia island of Bangladesh.**

Index Terms—**Hybrid power, Optimal cost, Energy resources, Operating cost.**

I. Introduction

Bangladesh is a developing country having a population over 160 million [1]. The total demand of electricity is increasing with time and energy crisis have become a severe problem for Bangladesh. Total demand of electricity has been raised to over 7000 MW. The electricity demand is increasing day by day which is also expected to be doubled within 2020 [2]. Only 31.2% of the total population is connected to grid electricity (80% of urban and 23% of rural households), with the vast majority being deprived of a power supply [3, 4].The main source of power generation is using natural gas. But the total reserve of gas is decreasing. So, the search for new source of energy is essential in the context of Bangladesh.

.

Hybrid Power Generation, as the name implies, combine two or more mode of power generation together usually using renewable technologies. Hybrid power systems are designed for the generation of electrical power. They are generally independent of large centralized electric grids and are used in remote areas. Hybrid systems by definition contain a number of power generation devices such as wind turbines, photovoltaic, micro-hydro and/or fossil fuel generators. Hybrid power systems range from small systems designed for one or several homes to very large ones for remote island grids or large communities.

Geographically, Bangladesh is situated between 20.30 and 26.38° north latitude and 88.04 and 92.44° east longitude. In the southern portion of the country, the Bay of Bengal is situated which is a big source of tidal energy. We have many rivers flowing over the country and meet the Bay of Bengal. In these outfalls, tidal height and tidal flow both are high and suitable for harnessing energy. Wind flow is also reasonable in these coastal areas of Bangladesh. So, it is feasible to implement a hybrid power system in the coastal area of Bangladesh. Implementation of hybrid power system in the coastal area could solve energy crisis to some extent.

II. Energy Resources for Hybrid System

To design an optimal hybrid energy system it is important to select suitable energy resources. Bangladesh is blessed with many renewable energy resources and since Kutubdia is in the coastal region, so tidal and wind resources are selected for this work as there is availability of high tidal range and wind velocity. The optimal design is performed with the HOMER (Hybrid Optimization Model for Electric renewables) software and economic system is obtained to find the feasibility of our proposed system.

A. Wind Resources

The wind velocity over Bangladesh differ from 3m/s to 9 m/s at different height during month March to September and according to Bangladesh Metrological Department (BMD) the wind speed is high during March to September monsoon and low during October to February in Bangladesh [5]. Kutubdia is a coastal area along the line of Bay of Bengal. In the area near Bay there is enough wind speed to produce electricity commercially. The wind speed data at Kutubdia is collected from LGED [6] at height 20m. Figure 1 shows the monthly average wind speed at Kutubdia over a year.

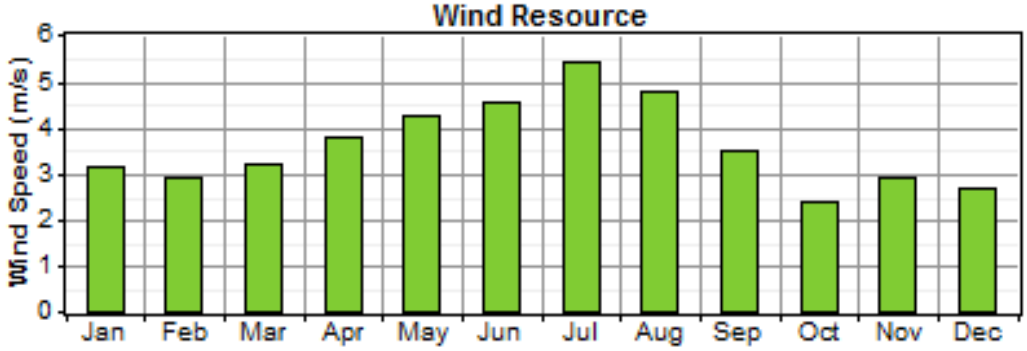


Figure 1: Wind speed variation per month.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

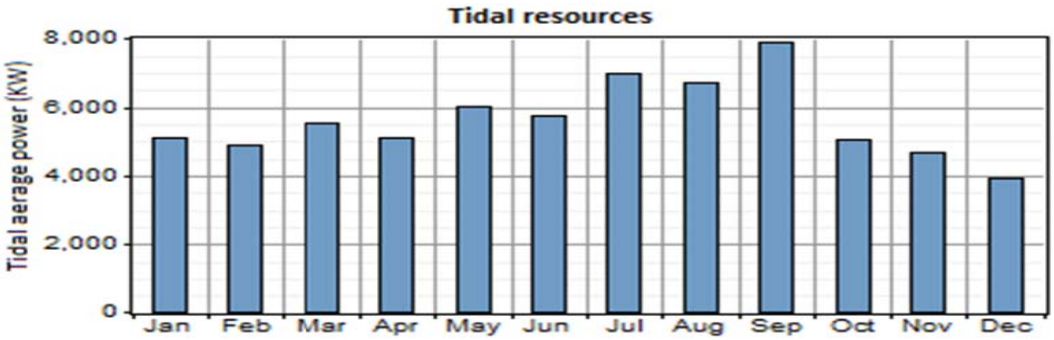


Figure 2: Average tidal power per month.

Wind energy is a special kind of kinetic energy in air as it flows. The power extracted by wind energy can be calculated by $P=1/2\rho AV^3$ formula [7]. Where, A is the area perpendicular to direction of flow (m^2), V is the wind velocity (m/s) and ρ is the air density (kg/m^3). Only a part of this power can be extracted. Under ideal condition maximum power that can be extracted is 59.3% of available power [5]. For wind speed calculation the weibull (k) constant, autocorrelation factor, diurnal pattern strength and hour of peak wind speed is needed in HOMER. For the selected site k is considered 2.17, autocorrelation measures the randomness of the wind which is considered 0.914. Diurnal pattern measures how strongly wind speed depends on the time of the day. This is considered 0.161 and hour of peak wind speed is considered 15 hours which considered the time of the day to be windiest on average throughout the year.

B. Tidal Resources

The energy extracted from the tides can be obtained from both the vertical water movements associated with the rise and fall of the tides. Potential energy or the kinetic energy which is the result of the roughly horizontal water motions termed as tidal currents. For these reasons tidal power harnessing technology can be categorized into two main types, tidal barrages and tidal current turbines [8]. Tidal barrages technology is selected for our designed system. For tidal barrage system we can extract energy from tide by two processes, one way tidal generation and two way tidal generation system [8, 9].

Two way generation system is selected for this proposed project, since the average output power getting from this system is more than one way generation system. The total power for two way power generation is $\eta \times 0.113\ AR^2$ [10]. Where, A is the basin area (m^2), R is the tidal range (m) and η is the efficiency of the turbine. Figure 2 shows the monthly averaged tidal power at Kutubdia over a year.

III. Design of Hybrid System

A hybrid energy system for Kutubdia is designed where tidal and wind power generation is combined with diesel generator. Diesel generator has been chosen because of its availability and its low cost comparing to other generator in Bangladesh. A hybrid system consists of primary energy sources which work with parallel to the secondary energy units. HOMER has been used to optimize the best efficient system for Kutubdia considering different load and tidal-wind and diesel combination. Figure 3 shows a design of hybrid system as implemented in HOMER.

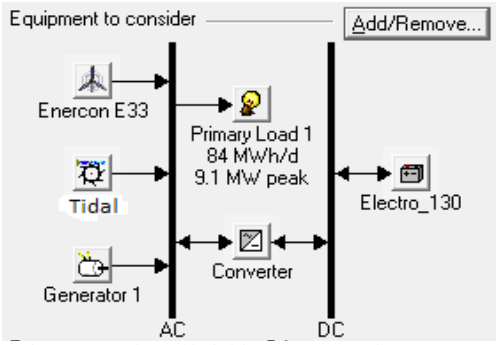
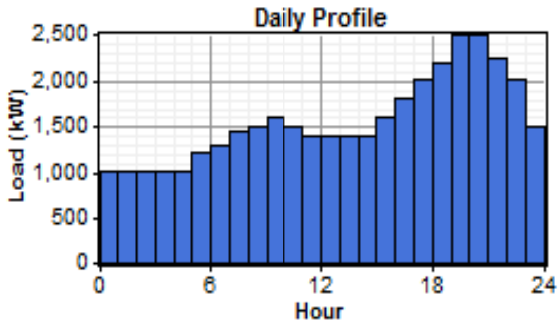


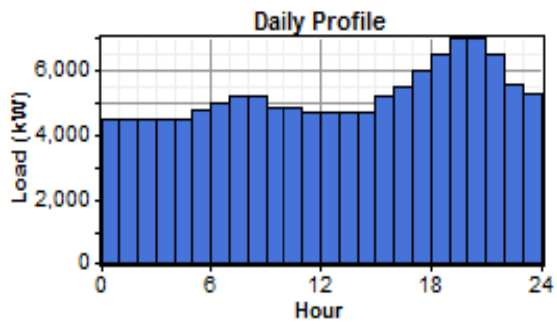
Figure 3: Design of hybrid power system in HOMER.

A. Electric Load Profile

An electrical load profile is the basis for design of hybrid energy system in HOMER. Load profile is actually the determination of electric load in an area according to a definite time period. Kutubdia is a coastal area of Bangladesh. The density of population is 580.6/km^2 and total population is around 150000 [11]. Each household is considered to consume 3 energy efficient bulbs (20W), 2 ceiling fans (60W). [12] The daily average uses of bulbs and fans are 8-9 hours and 12-13 hours respectively during the summer time and it varies from summer to winter. The total family is assumed 35000 and the load profile for January (winter) and May (summer) is shown in Figure 4



(a)



(b)

Figure 4: Load profile for winter (a) and summer (b)

B. Wind Turbine

The average wind speed in Kutubdia is measured 3.637m/s at 20 m height. From this we can measure wind speed at 50m which is 4.61m/s [13]. this velocity differs from summer to winter so cut in speed of wind turbine should be around 3m/s. Enercon E48 is suitable for low wind velocity and its cut in speed is 2.5m/s at height 50. Its rated power is 880KW AC and works in 50Hz which is suitable for Bangladesh. Enercon offers 15 years comprehensive operation and maintenance warranty reducing risk during the wind turbine operation [14]. The installation cost of Enercon E48 is $500000 and it offers most cost effective project even at site of difficult access. The lifetime of Enercon E48 is considered 25 years [14].

C. Tidal Turbine

In HOMER there is no block for tidal turbine, so the hydro block is modeled in such a way that it works as a tidal turbine block. For this the available head is assumed 101.9m, the designed maximum flow rate is assumed 4500L/s and efficiency is considered 100% so that the steam flow rate is equal to the tidal output power.

D. Diesel Generator

Diesel generator is used for the backup generation for this design of hybrid system. Diesel generator is worked in parallel with wind and tidal energy system to increase the efficiency. The installation cost of diesel generator is $18000 with operating cost $0.5/hr and minimum load 30% is provided by generator. The price of diesel is used $0.85.

E. Converter and Battery

A converter is used for AC-DC converter. $310 is considered for the cost of 2kW converter and efficiency is assumed 75%. The operating and maintenance cost is assumed $50/yr and lifetime for a converter is considered 15 years. Since in Bangladesh maximum rated capacity is 130-150Ah so a battery named electro_130 is created in HOMER with nominal capacity 130Ah and operating voltage 12 volt which has maximum energy storage 1.56kwh. The cost is considered $190 for per battery.

IV. Simulation Result and Analysis

The designed hybrid system will provide power to all the people in Kutubdia throughout year. Simulated result of HOMER gives an optimized result to meet the load demand in Kutubdia.

A. Optimization Result

The optimization result in homer is based on COE (cost of energy). The leveled COE for the best feasible result is $0.054/kWh, the total NPC (Net Present Cost) is $21,186,998 and operating cost is $84,062/yr. The detailed optimization cost for best possible combination of tidal-wind-diesel is shown in Figure 5 and Table I.

B. Simulation Result

The optimization result getting from HOMER simulation gives the best possible combination of energy resources to meet the demand. The optimized combination of tidal-wind-diesel is 22% of wind, 78% of tidal and very few of diesel generator which is near to 0%. Figure 6 shows the contribution of energy resources in the total power supply.

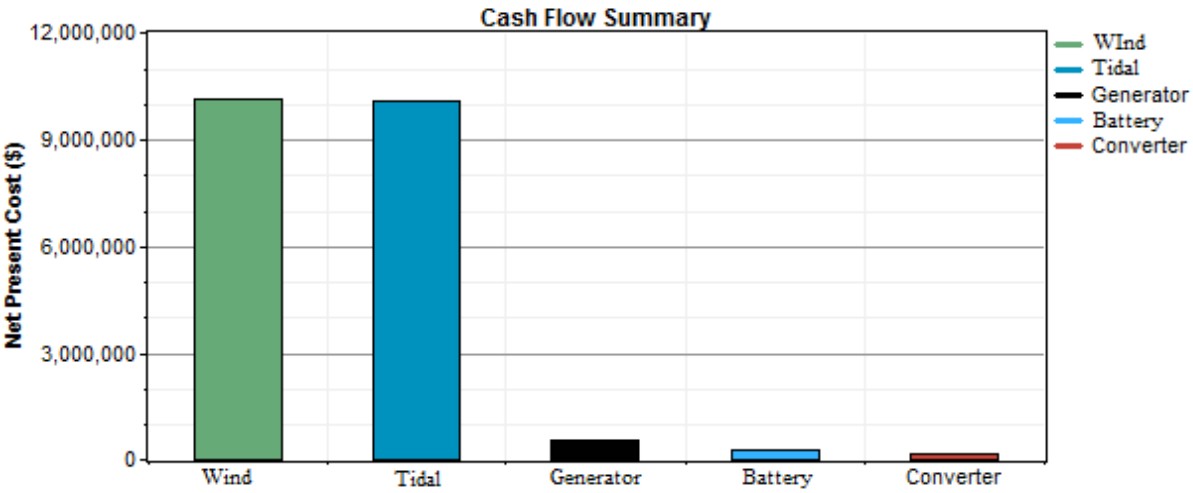


Figure 5: Cash flow bar chart

TABLE I. Optimizing Cost of Hybrid Power Designed System

Energy component	Initial Capital($)	Replacement ($)	Operating and Maintenance Cost($)	Fuel Cost ($)	Total cost ($)
Wind turbine	10,000,000	0	127,834	0	10,127,835
Tidal	10,000,000	0	63,917	0	10,063,917
Diesel Generator	5,400	1,831	1,685	556,198	564,498
Battery	76,000	122,083	76,700	0	274,784
Converter	31,000	61,042	63,917	0	155,959

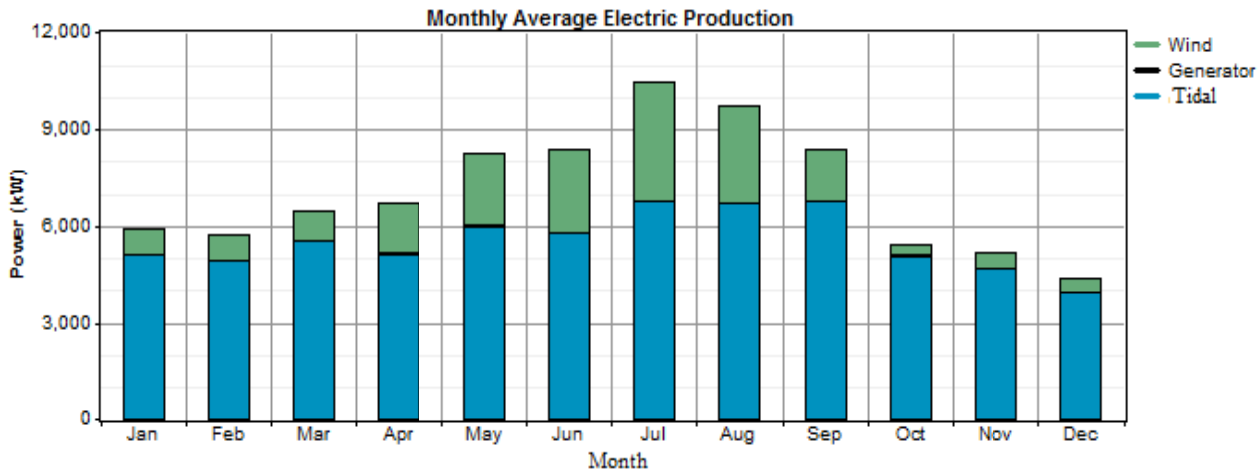


Figure 6: Monthly average power generation in the designed system

C. *Environmental Analysis*

Due to the reduction of fuel consumption the proposed design of hybrid system reduces the gas emission. The emission of the gases is calculated by HOMER software are given in Table II

TABLE II. ENVIRONMENTAL EFFECTS ON HYBRID POWER DESIGNED SYSTEM

Pollutant	**Emission (kg/yr)**
Carbon dioxide	134,794
Carbon monoxide	333
Unburned hydrocarbons	36.9
Particulate matter	25.1
Sulfur dioxide	271
Nitrogen oxides	2,969

V. CONCLUSION

Bangladesh is blessed with renewable resources and using it in power generation can be really effective in solving energy crisis of the country. The remote coastal areas which are not connected in the national grid can be supplied with electricity by implementing hybrid power system. Due to high diesel cost only diesel based power generation is not economically feasible. At the same time due to the intermittent nature of renewable sources and high initial cost of renewable energy technology, large scale power generation with only tidal or wind turbine is not economically feasible. So, a hybrid power system with more than one renewable energy source is economically suitable and sustainable system. According to the simulation results the projected hybrid system is economically feasible as decentralized or grid interactive system. We can install such power system for harnessing energy from renewable sources.

REFERENCES

[1] BPDB *Annual Report*, Bangladesh Power Development Board 2010-2011.

[2] *Household Income and Expenditure Survey*, Bangladesh Bureau of Statistics, Ministry of Planning, 2010.

[3] *Statistical yearbook of Bangladesh 2007*, Bangladesh Bureau of Statistics, Ministry of Planning, 2008.

[4] M. H. Alam and D. Manfred, "*Hybrid systems for decentralized power generation in Bangladesh*," Center for Development Research (ZEF), University of Bonn, Bonn, 2009.

[5] A. Zubair, A.A. Tanvir and M.M. Hasan, "Optimal Planning of Standalone Solar-Wind-Diesel Hybrid Energy System for a Coastal Area of Bangladesh," *International Journal of Electrical and Computer Engineering,* Vol. 2, No. 6, December 2012, pp. 731-738.

[6] "LGED Tide Tables", Department of Hydrography, Local Government Engineering Department, Under the Ministry of Local Government, Rural Development & Co- operatives.

[7] J. Park, *The Wind Power Book*, Chapter 1, Fundamental, 13-21

[8] Khaligh and O. C. Onar, *Energy Harvesting, Solar, Wind and Ocean Energy Conversion System*, Chapter 3, Tidal Energy Harvesting, 173-180, 2010

[9] G.N. Tiwari and M.K. Ghosal, *Renewable Energy Resources: Basic Principles and Applications*, Oxford, UK: Alpha Science International Ltd., April 2005.

[10] T. Ali, S.D. Gupta, I.Z. Arnab and S.M. Ferdous, "An Analysis of the Potentiality of Tidal Power in Swandwip by Using One Way & Two Way Power Generation Technology: An Ideal Model," *International Journal of Renewable Energy Research*, Vol. 2, Nov.3, 2012

[11] Kutubdia-City Population [online] Available: http://www.citypopulation.de/php/bangladesh-admin.php?adm2id=2245. (Accessed on 01 June, 2014)

[12] Power Consumption Table [online]. Available: http://www.absak.com/library/power-consumption-table. (Accessed on 01 June, 2014)

[13] Wind Gradient [Online]. Available: http://en.wikipedia.org/wiki/Wind_gradient (Accessed on 01 June, 2014)

[14] Turbine Option: Enercon E48 [online]. Available: http://www.locogen.com/products-services/wind-energy/turbine-options/enercon-e33.html (Accessed on 01 June, 2014)

Fault Detection of Brahmanbaria Gas Plant using Neural Network

M.T. Sowgath,[1,*] and S Ahmed[2]

[1] Assistant Professor, Chemical Engineering Department, BUET
Dhaka, Bangladesh

[2]Senior Gas Plant Operations Specialist, Chevron Bangladesh Ltd, Bangladesh

[*]mstanvir@che.buet.ac.bd

Abstract—**In recent years, several accidents in pioneer gas processing industries led industries to put emphasis on real-time fault detection. Neural Network (NN) based fault (abnormal situation) detection technique played an important role in monitoring industrial safety. In this work, an attempt has been made to study the fault detection of Brahmanbaria gas processing plant using multi layered feed forward NN based system. NN based fault detection system is trained, validated and tested using data generated using the dynamic model. Preliminary results show that NN based method is able to detect the faults of Brahmanbaria Gas processing plant for fewer no of faults.**[*]

Index Terms— **Neural Network, Fault detection, GasProcessing Plant, Industrial Safety Management.**

I. Introduction

Natural gas (NG) plays an important role in different sectors such as power generation sector, fertilizer, industrial and commercial sector in Bangladesh [1-2]. There are many gas producing fields in Bangladesh. Although process safety technology has been gradually implemented over the years, several accidents happened in gas field industry in Bangladesh like BGFCL, Nikko, Occidental, Tullow, and Chevron. Massive blowout took place in Occidental and the similar incident happened at Tengratila by Nikko. Major gas leakage was found in Titas Gas Field (BGFCL) [3-5]. Still, the process plants face difficulties due to lack of proper monitoring. Bangura gas plant was shut down for lacking proper monitoring [3-5] and the Bibiyana gas plant was shut down to repair a leaking gasket. Large gas processing plant consists of thousands of measurements, hundreds or thousands of controllers and many recycle streams. To reduce the frequency and consequences of accidents of chemical process operation, classical methods for fault detection are implemented. However, classical methods for fault detection in industries depend on human supervision and expertise. Again, operator of modern plant is saturated by alarm of supervisory control and data acquisition (SCADA) and Distributed control system (DCS) and difficult to manage the multiple fault simultaneously. Therefore, the above incidents clearly show the necessity of pro-active real time techniques for fault detection. A number of real-time monitoring techniques for chemical and petrochemical processes, such as extended kalman filter, signed diagraph, qualitative simulation, statistical method and neural network has been suggested in literature [6-12]. The use Neural Networks (NNs), in all aspects of process engineering activities, such as modeling, design, optimization and fault diagnosis, error detection, data reconciliation etc. has been considerably increased in recent years [8,11,12]. To the best of our knowledge, there has been no work reported in literature on using NN based fault detection system for gas processing plants in Bangladesh.

In this work, the steady state model of Brahmanbaria gas processing plant is developed using Aspen HYSYS and is validated using real plant data. Then, the dynamic model of the plant is developed from steady state model to understand the real plant behaviour and plant disturbances. Finally, an attempt has been made to study the fault detection of Brahmanbaria gas processing plant using NN based system. Early malfunction of liquid control valve in the plant is distinguished by using NN based Fault detection system from various parameters.

II. Brahmanbaria Gas Process Plant

Aspen HYSYS is user friendly software for simulation of chemical plants and oil refineries [13]. In this process (Fig.1) gas steam from air cooler flows through two phase vertical separator where, liquid separates from the gas. The overhead product gas passing through gas-gas exchanger, the gas goes to another two phase vertical separator where the bottom liquid product goes to liquid storage tank preheated by steam heater. The heavier hydrocarbon collected in the liquid storage tank is pumped to a distillation column with twenty trays.

III. Neural Network Fault Detection Scheme

Any non-linear relationship between input and output of a system can be captured effectively using NNs. NNs are consist of a large number of primitive computational elements called neuron. NN based fault detection are based on classification of historic process knowledge. The development of the NN based system involves selecting suitable architecture to differentiate between the fault of the process with the normal condition. The steps of NN based fault detection system in this work is shown in Fig. 2 and Fig. 3. NN is trained, validated and tested are similar to stated as [10] is shown in Fig. 4 and Fig. 5. Historic data of 10 key variables which affects the different mode of operations (types of faults and normal) is fed to the input layer and mode of operations fed to the output layer of the NN. NN is trained, validated and tested using dynamic data. Several multi layered feedforward neural networks with varying configurations and Levenberg Marquardt back propagation algorithm are also employed in the training and testing process of the to obtain optimum neural network architecture.

[*] Part of the work is presented at 62nd Canadian Chemical Engineering Conference

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

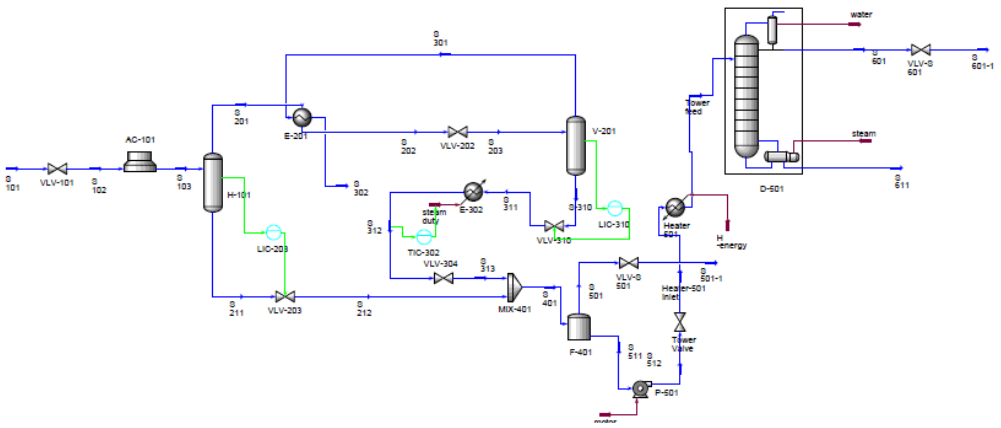


Fig. 1 Process Flow Diagram the Brahmanbaria Gas processing Plant.

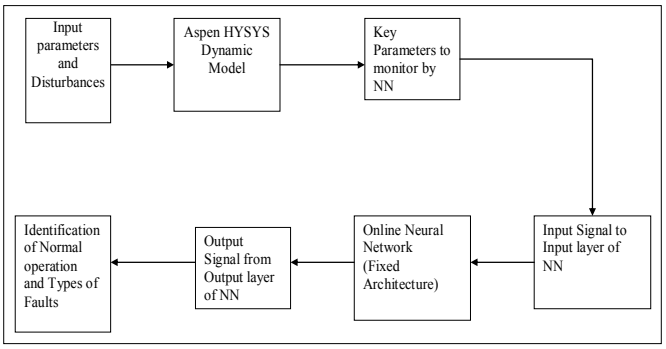


Fig. 2 NN based fault detection system

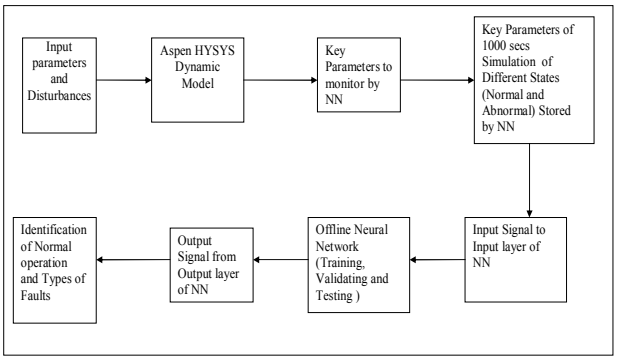


Fig. 3 Offline NN based fault detection

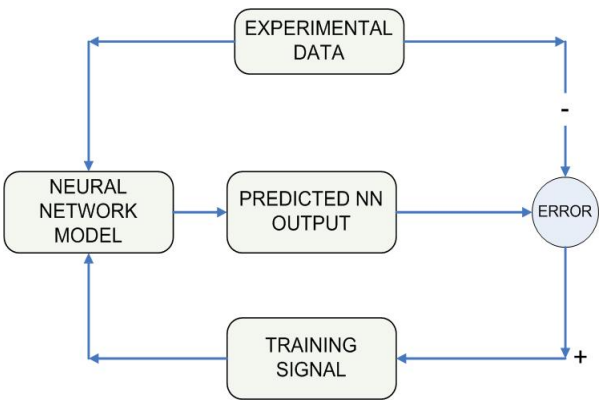


Fig. 4 Neural Network Backpropagation Training Scheme.

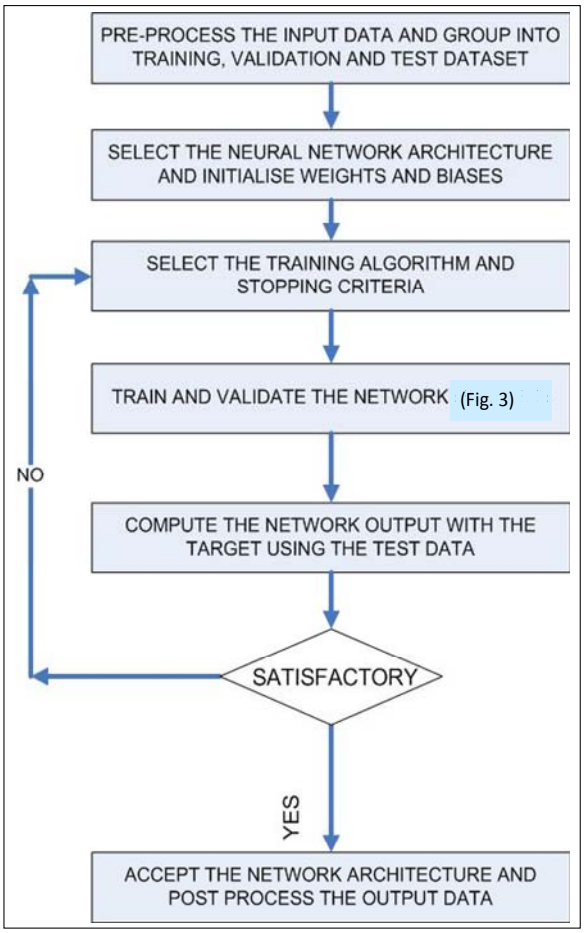


Fig. 5 The steps of NN model development.

IV. Results And Discussion

Table I shows the design specification of the Brahmanbaria Gas processing Plant Model (Fig. 1) within Aspen HYSYS. 1700 sample data for key parameters from the normal situation to each type of faults are simulated using Aspen HYSYS simulator. The key data (Table III) of different normal operation mode using Aspen HYSYS simulator are shown in chart forms from Fig. 6 and sample fault (Table II) in the process plant generated using dynamic model are shown in chart forms from Fig. .7. However, visually distinguished such data pattern is quite difficult.

Those data for each key parameters are monitored by NN based fault detection system and fed to neural network input layer (Table III). Types of Operation Mode/Disturbance Criteria (Table II) is integrated with Output layer Neuron of NN based system. Neuron of output layer for each states value equal to 1 and for absent of the states value equal to 0 (Table IV). Data sets are sorted to avoid uneven distribution. All data are scaled to the symmetrical range of -1 to +l.

TABLE I
INPUT-OUTPUT PARAMETER

Stream Name	**S 101**	**S 302**	**Heater-501 Inlet**
Temperature (°F)	150	98	62
Pressure (psia)	3000	1024	19
Mass flow (lb/hr)	57112	39789	11314
Stream Name	**Tower feed**	**S 601**	**S 611**
Temperature (°F)	62	14	350
Pressure (psia)	19	14	16
Mass flow (lb/hr)	11314	10113	1201

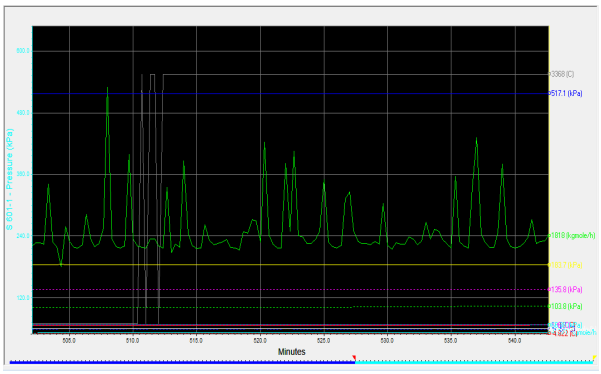


Fig. 6 Normal operation mode by HYSYS

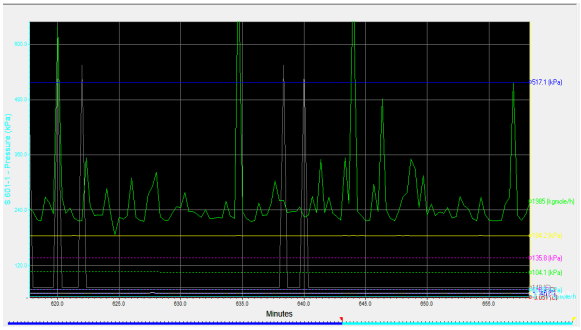


Fig. 7 Tower valve disturbance HYSYS model

TABLE II
DIFFERENT PARAMETERS MONITOR BY NN TO IDENTIFY OPERATION STATES

Different Parameters and Different Layer Neuron			
Monitor Variables and Input layer Neuron of NN	Tower feed Molar Flow	Tower feed Temperature	Tower feed - Pressure
	S 601-1 - Molar Flow	S 601-1 - Temperature	S 601-1 - Pressure
	S 611-1 - Molar Flow	S 611-1 - Temperature	S 611-1 - Pressure
	S 302 - Molar Flow		S 302 - Pressure
	S 401 - Molar Flow	S 401 - Temperature	S 401 - Pressure

TABLE III
TYPES OF OPERATION MODE/DISTURBANCE CRITERIA

Different states (normal and abnormal) and Output layer Neuron	**Operation Mode/Disturbance Criteria**		
	Normal operation	Tower valve open	Tower valve close
		VLV-203 & 304 open	VLV-203 & 304 close

TABLE IV
NN ARCHITECTURE FOR DIFFERENT CONDITIONS

Condition	**Normal**	**Abnormal**			
		Tower valve open	**Tower valve close**	**VLV-203 & 304 open**	**VLV-203 & 304 close**
NN output layer	$\begin{bmatrix}1\\0\\0\\0\\0\end{bmatrix}$	$\begin{bmatrix}0\\1\\0\\0\\0\end{bmatrix}$	$\begin{bmatrix}0\\0\\1\\0\\0\end{bmatrix}$	$\begin{bmatrix}0\\0\\0\\1\\0\end{bmatrix}$	$\begin{bmatrix}0\\0\\0\\0\\1\end{bmatrix}$

Training of NN Fault detection system is shown in Fig. 8. The statistical regression plots of (Fig. 9 to Fig. 10) between predicted and target data of different operation mode are plotted. Network architecture is updated until the regression value is almost close to 1. Optimum network is found for this work with 9 neurons in hidden layer.

However, If the setpoint of normal operation is changed whose response is not similar to the training pattern network, NN fault detection might detect as normal operation. Hyperbolic tangent functions are used in the input and hidden nodes. Linear functions are used in the hidden layer and output nodes.

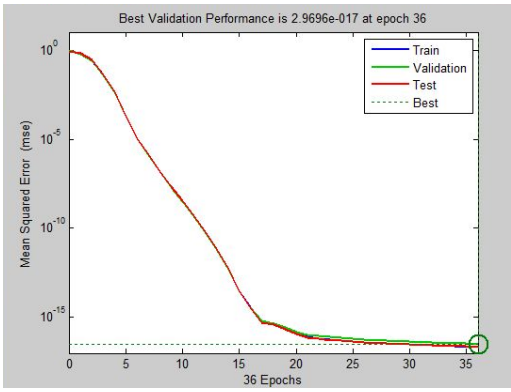


Fig. 8 Training of NN Fault detection system

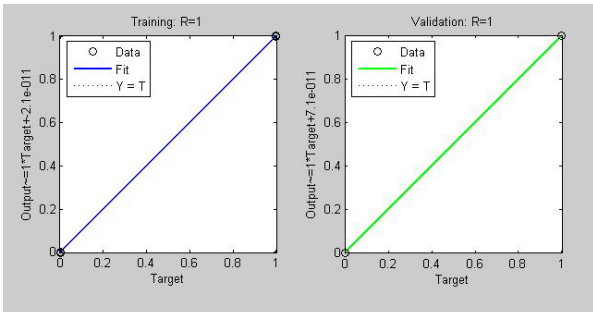


Fig. 9 Regression analysis of NN predicted data with Fault

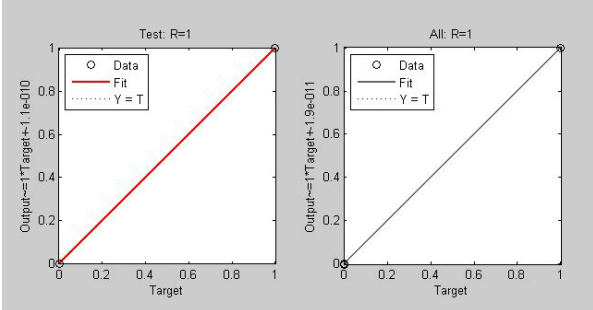


Fig. 10 Regression analysis of NN predicted data with Fault

Predictions by optimum NN within the training range follow the expected trends and it is within the engineering accuracy (Fig. 11 and Table V). This probe that optimum network able to predict types of fault (here liquid control valve failure for product quality loses) even when the network is with new inputs (test data).

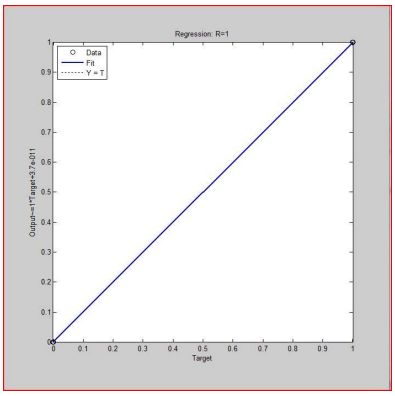


Fig. 11 Regression analysis of NN predicted test data

V. Conclusions

The priority of fault detection is increasing. Brahmanbaria Gas Plant behaviour and disturbances are studied using HYSYS dynamic model. Disturbance data are shown in chart form. A feedforward NN based fault detection was developed to identify the fault (disturbance) and no fault (normal) in plant operation. The NN based fault detection system was be trained, validated and tested using the dynamic model data. preliminary results show that NN based fault detection able to identify realistic fault output.

Authors believe that, NN based fault detection will help to avoid accident events and productivity losses in Gas processing industry in Bangladesh and help operators to identify and monitor multiple fault in real-time.

Acknowledgment

We express our gratitude to the Professor I.M.Mujtaba, University of Bradford, UK in neural networks whose his support has been a inspiration to us.

References

[1] M. Rahman, and M. Tamim, and L. Rahman "Analysis of natural gas consumption by industrial sector of Bangladesh," *Journal of Chemical Engineering*, *IEB*, vol. 27, pp. 1–7, June 2012.

[2] M.T. Sowgath, Z. M Hossain, and S. Kawsher, "Maximising the Heat Content of Sales Gas at Kailashtilla Gas Processing Plant " *Chemical Engineering Transactions*, vol. 24, pp. 313-316, 2011.

[3] (2013) The daily-sun website. [Online]. Available: http://www.dailysun. com/ access at July 10 2014./

[4] (2008) The daily-sun website. [Online]. Available: http://www.thedailystar.net/newDesign/news-det / access at July 10 2012./

[5] (2008) The Financial Express. [Online]. Available: http:// http://www.thefinancialexpress-bd.com/2008/0/ access at July 10 2012./

[6] D. M. Himmelblau, *Fault Detection and Diagnosis in Chemical and Petrochemical Processes*, Elsevier Scientific Publishing Company, Amsterdam, Oxford, New York, 1978.

[7] D. M. Himmelblau and M. A. Hussain, "Fault Detection and Diagnosis in Chemical and Petrochemical Processes", Vol-8 Berlin, Amsterdam: Elsevier Scientific Publishing Company, 1978.

[8] I. M. Mujtaba and M. A. Hussain, Ed., *Neural Networks and Other Learning Technologies in Process Engineering*, London, Imperial College Press, 2004, Vol. 3.

[9] M.T.,Hagan H.B. Demuth and M.N. Beale, *Neural Network Design*,: 1st edition PWS Publishing Company of International Thomson Publishing Inc, 1996.

[10] M.S. Tanvir and I.M., Mujtaba Neural network based correlations for estimating temperature elevation for seawater in MSF desalination process, Elsevier, 2005, 195, 251-72.

[11] Y. Maki, and K. A. Loparo, "A Neural Network Approach to Fault detection and Diagnosis in Industrial Process," *IEEE Transaction on Control Systems technology*, vol. 5, No. 6 pp. 529–541, Nov. 1997.

[12] V. Venkatasubramanian, R. Rengaswamy, S. N. Kavuri, and K. Yin, "A Review of process fault detection and diagnosis Part III: Process history based methods," *Computers and Chemical Engineering*, vol. 27, pp. 327-346, 2003.

[13] HYSYS, HYSYS 3.2 user guide, Hyprotech Ltd, http://www.hyprotech.com/

TABLE V

Sample Fault Scheme Result for Test Data

	Tower feed - Temperature	S 601-1 - Pressure	Tower feed - Pressure	S 601-1 - Temperature	S 611-1 - Temperature	S 611-1 - Pressure	S 302 - Molar Flow	S 302 - Pressure	S 401 - Temperature	S 401 - Pressure	Output Neuron Value					Output Neuron Value by NN				
	[F]	[psia]	[psia]	[F]	[F]	[psia]	[MMSCFD]	[psia]	[F]	[psia]	1	0	0	0	0	1	4.03374E-09	-9.98884E-10	-5.25759E-10	-2.23052E-09
Normal Operation	121	10	19	214	225	20	24	75	101	10	1	0	0	0	0	1	4.87022E-09	-1.01484E-09	-1.02606E-09	-2.48704E-09
	121	10	19	214	225	20	24	75	104	10	1	0	0	0	0	1	8.74083E-10	3.62786E-10	-3.46233E-10	-5.62695E-10
	119	10	17	188	225	20	25	75	104	10	1	0	0	0	0	1.000000003	6.49216E-09	-2.05814E-09	6.919E-10	-7.83192E-09
	119	10	17	188	225	20	24	75	99	10	1	0	0	0	0	0.999999994	5.27751E-09	3.90782E-09	-1.00769E-09	-2.55304E-09
Tower valve full open fault	79	14	191	10	225	20	24	75	10	68	0	1	0	0	0	3.80625E-10	1.000000001	2.16216E-09	4.48077E-10	-4.317E-09
	79	14	191	10	225	20	23	75	10	68	0	1	0	0	0	5.33552E-10	0.999999986	2.48718E-09	-4.40562E-10	1.15456E-08
	79	14	191	10	225	20	25	75	10	68	0	1	0	0	0	2.36126E-10	1.000000014	1.75591E-09	1.50443E-09	-1.78391E-08
	76	12	155	10	296	20	24	75	10	68	0	1	0	0	0	8.636E-11	1.000000003	-1.92406E-10	-2.2108E-09	-8.29511E-10
	76	12	155	10	296	20	24	75	10	68	0	1	0	0	0	1.00077E-10	1.000000003	-2.23888E-10	-1.98415E-09	-5.09009E-10
Tower valve full close fault	180	7	-76	10	225	20	24	75	10	68	0	0	1	0	0	3.22709E-09	-3.89884E-09	1.000000001	-1.69263E-09	1.03893E-09
	179	7	-76	10	225	20	24	75	10	68	0	0	1	0	0	3.15343E-09	-4.48098E-09	1.000000001	-1.78324E-09	2.20213E-09
	178	7	-76	10	225	20	24	75	10	68	0	0	1	0	0	3.79229E-09	-5.05028E-09	1.000000003	-1.91119E-09	2.02801E-10
VLV-203&304 full open fault	373	25	81	10	260	20	24	75	69	374	0	0	0	1	0	3.5087E-10	2.13279E-10	1.59257E-10	0.999999999	9.77164E-11
	373	25	81	10	260	20	24	75	69	374	0	0	0	1	0	6.38787E-10	-1.54718E-10	-1.23796E-10	1	-5.15153E-11
	373	25	81	10	260	20	24	75	69	374	0	0	0	1	0	2.169E-10	-7.39016E-10	-7.82877E-10	1	1.06334E-09
VLV-203&3	150	18	216	10	225	20	24	75	10	166	0	0	0	0	1	2.84952E-10	1.39294E-08	1.93263E-11	-1.36459E-08	0.999999999
	150	18	216	10	225	20	23	75	10	166	0	0	0	0	1	4.5815E-10	2.62054E-08	3.29405E-10	-1.66684E-08	0.99999999
	149	18	218	10	225	20	24	75	10	93	0	0	0	0	1	-4.74787E-12	-2.54849E-09	1.70005E-11	-2.0985E-10	1.000000003
	149	18	218	10	225	20	24	75	10	93	0	0	0	0	1	-1.26436E-11	-3.24281E-09	3.33034E-12	-9.03236E-11	1.000000003

Interconnected Systems with "Mixed" Passivity and Negative-Imaginary Properties: A Stability Analysis in Discrete-time

Md. Arifur Rahman*, Abdullah Al Mamun, Sajal K. Das
Department of Electrical and Computer Engineering, National University of Singapore, Singapore
Email: *a0092557@nus.edu.sg

***Abstract*—This paper briefly presents a stability analysis of interconnected systems with mixed passivity and negative-imaginary properties in discrete-time. These results are analogous to previously established continuous-time case. As discretization of the system or discrete-time design is an important issue for control system design, this paper provides a base for discretization of "mixed" system with passivity and negative imaginary properties. Numerical example and simulation results demonstrate the effectiveness of the proposed analysis.**

***Keywords*—Passivity, Negative-imaginary, Stability**

I. Introduction

The passivity theorem and Negative-imaginary theorem are important the results in the theory of stability of input-output systems. Passivity theorem states that the negative feedback interconnection between two stable linear time invariant systems is stable when both systems are passive and one of them is strictly passive with finite gain [1], [2]. According to Negative-imaginary theorem, the positive feedback interconnection between two systems where one system is negative-imaginary and other system is strictly negative-imaginary is stable if and only if the DC loop gain is less than one [3], [4].

However, it is often seen that many practical systems do not always belong to the specific class over the whole interval of frequency [5]-[8]. There are many systems which belong to a mixed class with different properties at different frequency intervals. For example, Fig. 1 shows the experimental frequency response of a PZT micro-actuator ([15], [16]) where it can be observed that the system is not purely passive or negative-imaginary for the whole frequency range, but it is a "mixed" system.

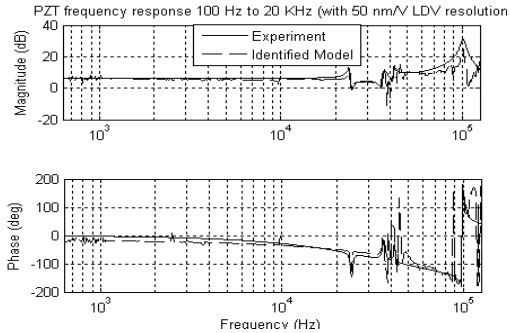


Fig. 1: experimental frequency response of a PZT micro-actuator

Many other systems may have "mixed" passivity and small-gain properties [5], [6] or "mixed" negative-imaginary and small-gain [7] properties. An unconditional stability analysis for systems with a "mixed" passivity and small-gain properties using negative-feedback is presented in [5] whereas [8] shows conditional stability between systems with "mixed" negative-imaginary and small-gain properties using positive feedback. Small-gain stability analysis can be applied and implemented to any kind of "mixed" system where the system does not hold only negative-imaginary or passivity property. The small-gain theorem states that the feedback interconnection of two stable systems is stable if the product of the gain of the systems at each frequency is less than one [1]. However, the use of the small-gain theorem makes the controller gain low which leads to low level of damping of the structures resonant modes in vibration control applications.

Nowadays, engineers or researcher rarely work with continuous-time systems in the industry or in the academic labs. In order to implement the control design for high speed system through some RTI tool like dSPACE, a discrete-time representation of the system must be required. Thus, it is to be established whether discrete-time systems inherit fundamental properties of the continuous-time systems from which they are derived. System discretization has currently become an issue of importance once again, and several papers [9]–[14] have recently appeared on this topic, particularly in the switched systems community. The purpose of this paper is to lay the foundation for future studies concerning the discretization of "mixed" systems with passivity and negative-imaginary properties and establishing a stability analysis in discrete-time for this types of mixed systems.

The rest of the paper is organized as follows. Section 2 presents the mathematical definitions of passivity; negative-imaginary and "mixed" properties and also highlights some preliminary results. Feedback stability analysis is presented in section 3. Section 4 demonstrates the numerical example and simulation results and section 5 presents the summary of the paper in conclusion.

II. Mathematical Definitions and Preliminary Results

Considering $y = xe^{j\theta}$ is a complex number where $x = |y|$ and $\theta = \angle y$ and assuming that $\theta = \omega T_s$ where T_s and ω are the sampling interval in seconds and signal frequency in rad/sec respectively.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

Passivity and Negative Passivity **: (i)** A bounded-input bounded-output (BIBO) stable, discrete-time system with square, proper, real-rational transfer function matrix $G(z)$ is said to be input-output strictly passive over $[-m,-n]\cup[m,n]$ if there exists $\alpha,\beta\geq 0$ such that

$$-\alpha G^*(e^{j\theta})G(e^{j\theta})+G^*(e^{j\theta})+G(e^{j\theta})-\beta I\geq 0 \quad (1)$$

for all $\theta\in[-m,-n]\cup[m,n]$ where $0\leq m\leq n\leq\pi$.
(ii) A BIBO stable, discrete-time system with square, proper, real-rational transfer function matrix $H(z)$ is said to be input-output strictly negative passive over $[-m,-n]\cup[m,n]$ if there exists $\alpha,\beta\geq 0$ such that

$$-\alpha H^*(e^{j\theta})H(e^{j\theta})-H^*(e^{j\theta})-H(e^{j\theta})-\beta I\geq 0 \quad (2)$$

for all $\theta\in[-m,-n]\cup[m,n]$ where $0\leq m\leq n\leq\pi$.
The systems $G(z)$ and $H(z)$ are called input strictly passive and input strictly negative passive respectively if (1) and (2) are satisfied respectively for $\alpha=0$; output strictly passive and output strictly negative passive respectively if (1) and (2) are satisfied respectively for $\beta=0$; and passive and negative passive respectively if (1) and (2) are satisfied respectively for $\alpha=\beta=0$.
Negative Imaginary (NI)**:** A BIBO stable, discrete-time system with square, proper, real-rational transfer function matrix $G(z)$ is said to be input-output strictly negative-imaginary over $[-m,-n]\cup[m,n]$ if there exists $\alpha,\beta\geq 0$ such that

$$-\alpha\theta^2G^*(e^{j\theta})G(e^{j\theta})+j\theta G(e^{j\theta})-j\theta G(e^{j\theta})-\beta\geq 0 \quad (3)$$

for all $\theta\in[-m,-n]\cup[m,n]$ where $0\leq m\leq n\leq\pi$.
The system $G(z)$ is called input strictly negative-imaginary and input strictly negative passive respectively if (3) is satisfied respectively for $\alpha=0$; output strictly negative-imaginary if (3) is satisfied for $\beta=0$; and negative-imaginary if (3) is satisfied for $\alpha=\beta=0$.
"Mixed" NI and Passivity or Negative Passivity**: (i)** A BIBO stable, discrete-time system with square, proper, real-rational transfer function matrix $G(z)$ is said to have the "mixed" passivity and negative-imaginary properties if for each frequency $\theta\in[-\pi,\pi]$, there exists $\alpha,\beta\geq 0$ such that either a) $-\alpha G^*(e^{j\theta})G(e^{j\theta})+G^*(e^{j\theta})+G(e^{j\theta})-\beta I\geq 0$ or b) $-\delta\theta^2G^*(e^{j\theta})G(e^{j\theta})+j\theta G(e^{j\theta})-j\theta G(e^{j\theta})-\rho I\geq 0$ or both where $\alpha,\beta,\delta,\rho\geq 0$
(ii) A BIBO stable, discrete-time system with square, proper, real-rational transfer function matrix $G(z)$ is said to have the "mixed" negative passivity and negative-imaginary properties if for each frequency $\theta\in[-\pi,\pi]$, there exists $\alpha,\beta\geq 0$ such that either a) $-\alpha G^*(e^{j\theta})G(e^{j\theta})-G^*(e^{j\theta})-G(e^{j\theta})-\beta I\geq 0$ or b) $-\delta\theta^2G^*(e^{j\theta})G(e^{j\theta})+j\theta G(e^{j\theta})-j\theta G(e^{j\theta})-\rho I\geq 0$ or both where $\alpha,\beta,\delta,\rho\geq 0$
Let us now consider the following example from [17]

$$M(s)=\frac{10s}{s^3+10.02s^2+256.2s=2560} \quad (4)$$

M(s) is a system with "mixed" negative-imaginary and passivity properties [17], [18]. Now using zero-order hold discretization method $M(s)$ is discretized to $M_1(z)$ and $M_2(z)$ for sampling time $T=0.02$ sec and $T=0.2$ sec respectively.

$$M_1(z)=\frac{0.0019z^2-0.00012z-0.0017}{z^3=2.717z^2+2.554z-0.8184} \quad (5)$$

And

$$M_2(z)=\frac{0.0307z^2+0.00011z-0.0319}{z^3+1.857z^2+0.7263z-0.1348} \quad (6)$$

The nyquist plots in Fig. 3 shows that $M_1(z)$ is a discrete-time system with "mixed" passivity and negative-imaginary properties where $M_2(z)$ is a discrete-time system with "mixed" passivity, negative passivity and negative-imaginary properties. $M_1(z)$ holds the same properties as the $M(s)$ where $M_2(z)$ fails to preserve the same properties after discretization. Thus discretization may change the properties of a system from its continuous-time case.
Now some preliminary results are presented to establish the main results

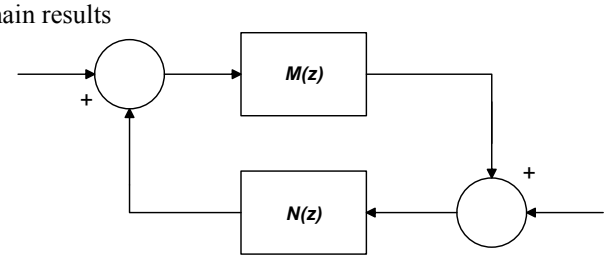


Fig. 2: Interconnection of system with positive feedback

Theorem 1**:** *According to [19], in discrete-time if $M(z)$ and $N(z)$ are strictly proper, real-rational transfer function matrices, then feedback interconnection of Figure 2 is stable if the nyquist plot of $\det[I-M(e^{j\theta})N(e^{j\theta})]$ for $-\pi\leq\theta\leq\pi$ does not make any encirclements of the origin.*

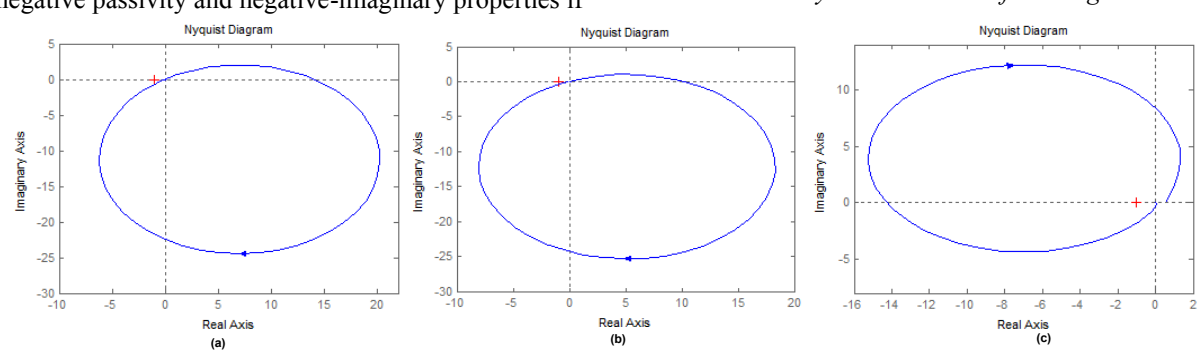


Fig. 3: For positive frequencies (a) Nyquist diagram of $M(s)$, (b) nyquist diagram of $M_1(z)$, and nyquist diagram of $M_2(z)$

The nyquist plot of $\det[I - M(e^{j\theta})N(e^{j\theta})]$ is included in the group of nyquist plots of $\det[I - \frac{1}{\Upsilon}M(e^{j\theta})N(e^{j\theta})]$, where $\Upsilon \in [1, \infty)$. Suppose the nyquist plot of $\det[I - \frac{1}{\Upsilon}M(e^{j\theta})N(e^{j\theta})]$ encircles the origin at least once which means that if Υ is continuously decreased from a large value towards 1, then there must be at least one Υ_0 and one θ_0 for which $\det[I - \frac{1}{\Upsilon}M(e^{j\theta})N(e^{j\theta})]$=0. Therefore, to avoid the encirclement of the origin, the sufficient condition is that for all $\Upsilon \in [1, \infty)$ and all $\theta \in [-\pi, \pi]$, $det[I - \frac{1}{\Upsilon}M[I - \frac{1}{\Upsilon}M(e^{j\theta})$ $N(e^{j\theta})] \neq 0$.

***Lemma 2*:** Suppose that $M(z)$ and $N(z)$ are BIBO stable, square, strictly proper, real-rational transfer function matrices. At some $\theta_0 \in [-\pi, \pi]$, if $M^*(e^{j\theta_0}) + M(e^{j\theta_0}) > 0$ and $-N^*(e^{j\theta_0}) - N(e^{j\theta_0}) \geq 0$, then $\det[I - M(e^{j\theta_0})N(e^{j\theta_0})] \neq 0$.

Proof: Since $M^*(e^{j\theta_0}) + M(e^{j\theta_0}) > 0$ and $-N^*(e^{j\theta_0}) - N(e^{j\theta_0}) \geq 0$, hence real parts of $\lambda_i[M(e^{j\theta})] > 0$ and real parts of $\lambda_i[-N(e^{j\theta})] \geq 0$, where $\lambda_i[.]$ denotes the ith eigen value [20]. Therefore $M(e^{j\theta})$ is nonsingular. Since $M^*(e^{j\theta_0}) + M(e^{j\theta_0})$ and $M^{-*}(e^{j\theta_0}) + M^{-1}(e^{j\theta_0})$ are Hermitian congruent [21], hence $M^{-*}(e^{j\theta_0}) + M^{-1}(e^{j\theta_0}) > 0$ which therefore implies that $M^{-*}(e^{j\theta_0}) - N^*(e^{j\theta_0}) + M^{-1}(e^{j\theta_0}) - N(e^{j\theta_0}) > 0$. Therefore, real parts of $\lambda_i[M^{-1}(e^{j\theta_0}) - N(e^{j\theta_0})] > 0$ for all λ_i and $det[M^{-1}(e^{j\theta_0}) - N(e^{j\theta_0})] \neq 0$ which entails that $det[I - M(e^{j\theta_0})N(e^{j\theta_0})] \neq 0$. Since $det[I - M(e^{j\theta_0})N(e^{j\theta_0})] = \det[M(e^{j\theta_0})\det[M^{-1}(e^{j\theta_0}) - N(e^{j\theta_0})]]$, $M(e^{j\theta_0})$ is nonsingular.

***Lemma 3*:** Suppose that $M(z)$ and $N(z)$ are BIBO stable, square, strictly proper, real-rational transfer function matrices. At some $\theta_0 \in [-\pi, \pi]$, if $j\left(M(e^{j\theta_0}) - M^*(e^{j\theta_0})\right) > 0$ and $j\left(N(e^{j\theta_0}) - N^*(e^{j\theta_0})\right) > 0$ then $\det[I - M(e^{j\theta_0})N(e^{j\theta_0})] \neq 0$

Proof: The two suppositions can be written as $jM(e^{j\theta_0}) + (jM(e^{j\theta_0})^* > 0$ and $\left(jN(e^{j\theta_0})\right)^{-1} + \left(jN(e^{j\theta_0})\right)^{-*} > 0$. Hence, $\det[I - M(e^{j\theta_0})N(e^{j\theta_0})] = \det[I + (jM(e^{j\theta_0}))$ $(jN(e^{j\theta_0}))] = \det[(jM(e^{j\theta_0}) + (jN(e^{j\theta_0})^{-1}]$ $\det[(jN(e^{j\theta_0})] \neq 0$.

III. Feedback Stability

***Theorem 4*:** If $M(z)$ and $N(z)$ are BIBO stable, square, strictly proper, real-rational transfer function matrices, then for the feedback interconnected system in Fig. 2, the interconnected system is finite-gain stable if there exists there sets of frequency range: (i) $\boldsymbol{\varphi_{NI}}$, a set consists of $\theta \in [-\pi, \pi]$ over which both $M(z)$ and $N(z)$ have negative-imaginary property (ii) $\boldsymbol{\varphi_{P-NP}}$, a set consists of $\theta \in [-\pi, \pi]$ over which $M(z)$ has passivity property and $N(z)$ has negative passivity property (iii) $\boldsymbol{\varphi_{NP-P}}$, a set of consists of $\theta \in [-\pi, \pi]$ over which $M(z)$ has negative passivity property and $N(z)$ has passivity property.

Proof: The proof can be divided into two parts: (a) For $M(z)$ and $N(z)$ belong to set $\boldsymbol{\varphi_{NI}}$, it is to be shown that $\det[I - \frac{1}{\Upsilon}M(e^{j\theta_0})N(e^{j\theta_0}) \neq 0$ all $\Upsilon \in [1, \infty)$ (b) For $M(z)$ and $N(z)$ belong to set $\boldsymbol{\varphi_{P-NP}}$ or $\boldsymbol{\varphi_{NP-P}}$, it is to be shown that $\det[I - \frac{1}{\Upsilon}M(e^{j\theta_0})N(e^{j\theta_0}) \neq 0$ all $\Upsilon \in [1, \infty)$.

(a) Since $M(z)$ and $N(z)$ belong to set $\boldsymbol{\varphi_{NI}}$, there exists $\alpha_i, \beta_i > 0$, such that $-\alpha_1\theta^2 M^*(e^{j\theta})M(e^{j\theta}) + j\theta M(e^{j\theta}) - j\theta M(e^{j\theta}) - \beta_1 \geq 0$ and $-\alpha_2\theta^2 N^*(e^{j\theta})N(e^{j\theta}) + j\theta N(e^{j\theta}) - j\theta N(e^{j\theta}) - \beta_2 \geq 0$ respectively, which entails that $j\theta\left(M(e^{j\theta_0}) - M^*(e^{j\theta_0})\right) > 0$ and $j\theta\left(N(e^{j\theta_0}) - N^*(e^{j\theta_0})\right) > 0$. Therefore, $\frac{1}{\sqrt{g}}j\theta\left(M(e^{j\theta_0}) - M^*(e^{j\theta_0})\right) > 0$, for $g > 0$. Hence, according to Lemma 3, $\det[I - \frac{1}{\Upsilon}M(e^{j\theta_0})N(e^{j\theta_0}) \neq 0$.

(b) Let us consider $M(z)$ and $N(z)$ belong to set $\boldsymbol{\varphi_{P-NP}}$, then there exists $\alpha_i, \beta_i > 0$ such that $-\alpha_1 M^*(e^{j\theta})M(e^{j\theta}) + M^*(e^{j\theta}) + M(e^{j\theta}) - \beta_1 I \geq 0$ and $-\alpha_2 N^*(e^{j\theta})N(e^{j\theta}) - N^*(e^{j\theta}) - N(e^{j\theta}) - \beta_2 I \geq 0$ respectively which implies that $M^*(e^{j\theta_0}) + M(e^{j\theta_0}) > 0$ and $-N^*(e^{j\theta_0}) - N(e^{j\theta_0}) \geq 0$. Therefore, $\frac{1}{\sqrt{g}}M^*(e^{j\theta_0}) + \frac{1}{\sqrt{g}}M(e^{j\theta_0}) > 0$ and $-\frac{1}{\sqrt{g}}N^*(e^{j\theta_0}) - \frac{1}{\sqrt{g}}N(e^{j\theta_0}) \geq 0$, for $g > 0$. Hence from Lemma 2, $\det[I - \frac{1}{\Upsilon}M(e^{j\theta_0})N(e^{j\theta_0}) \neq 0$. If, $M(z)$ and $N(z)$ belong to set $\boldsymbol{\varphi_{NP-P}}$, the above proof is vice versa.

IV. Numerical Example and Simulation

We consider the transfer function models $M_1(z)$ in (5) and let us consider another continuous time system as follows:

$$N(s) = \frac{0.1s^2 + 0.2s}{s^3 + 10.02s^2 + 356.2s + 3560}$$

Now using the sampling interval 0.02 sec we discretize the above model

$$N_1(z) = \frac{0.0018z^2 - 0.003542z + 0.001734}{z^3 - 2.678z^2 + 2.522z - 0.8184}$$

The nyquist plot of $N_1(z)$ is shown in Fig. 5(a). If we see the properties of $M_1(z)$ and $N_1(z)$ from Fig. 3(b) and Fig. 5(a), then we can summarize as shown in Fig. 4. Three frequency regions are here. In the first region, $N_1(z)$ has the negative passivity property where $M_1(z)$ has passivity property, in the second region both $M_1(z)$ and $N_1(z)$ have negative-imaginary property and $M_1(z)$ has negative passivity property where $N_1(z)$ has passivity property in the third region. Therefore positive feedback interconnection of $M_1(z)$ and $N_1(z)$ is stable. Step response (Fig. 5(b)) of the closed-loop interconnected system also confirms the stability.

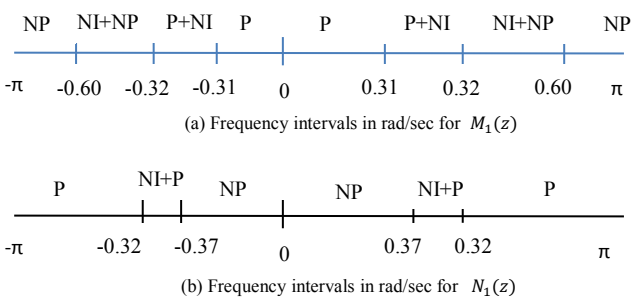


Fig. 4: Frequency intervals (NI: Negative-imaginary, P: Passivity, NP: negative passivity)

V. Conclusion

This paper proves the stability of discrete-time interconnected systems with mixed passivity and negative-imaginary properties. Discretization method may change the properties of the system than from its continuous-time case. Therefore Finite-gain stability theorem is proposed here in discrete time for system with "mixed" negative–imaginary and passivity properties. Numerical example and simulation validates the theorem.

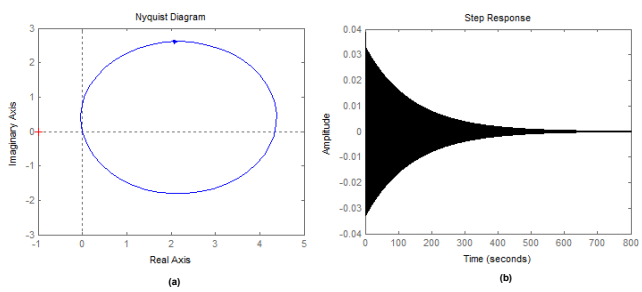


Fig. 5: (a) nyquist plot of $N_1(z)$ (b) step response of closed loop system

Acknowledgement

This work is supported by the Singapore National Research Foundation (NRF) under CRP Award No. NRF-CRP-4-2008-06.

References

[1] C. A. Desoer and M. Vidyasagar, *Feedback Systems: Input-Output Properties*. New York: Academic Press, 1975.

[2] M. Green and D. J. N. Limebeer, *Linear Robust Control*, E. Cliffs,Ed. NJ: Prentice-Hall, 1996.

[3] A. Lanzon and I. R. Petersen, "A modified positive-real type stability condition," *in Proc. European Control Conference*, Kos, Greece, Jul. 2007, pp. 3912–3918.

[4] Lanzon, A.; Petersen, I.R., "Stability Robustness of a Feedback Interconnection of Systems With Negative Imaginary Frequency Response," *IEEE Transactions on Automatic Control*, vol.53, no.4, pp.1042-1046, May 2008.

[5] W. M. Griggs, B. D. O. Anderson, and A. Lanzon, "A "mixed" small gain and passivity theorem in the frequency domain," *Systems and Control Letters*, vol.56, no.9-10, pp. 596–602, Sept.-Oct. 2007.

[6] W. M. Griggs, B. D. O. Anderson, A. Lanzon, and M. C. Rotkowitz, "Interconnections of nonlinear systems with "mixed" small gain and passivity properties and associated input-output stability results," *Systems and Control Letters*, vol. 58, no. 4, pp. 289–295, Apr. 2009.

[7] S. Patra and A. Lanzon, "Stability analysis of interconnected systems with"mixed" negative-imaginary and small-gain properties," *IEEE Transactions on Automatic Control*, vol. 56, no. 6, pp. 1395 – 1400, Jun. 2011.

[8] S. K. Das, H. R. Pota, and I. R. Petersen, "Resonant control of atomic force microscope scanner: A "mixed" negative-imaginary and small-gain approach," *in Proc. American Control Conference*, 2013, pp. 5476–5481.

[9] A. Pietrus and V. M. Veliov, "On the discretization of switched linear systems", *Systems & Control Letters*, vol. 58, no. 6, pp. 395-399, Jun. 2009.

[10] A. Zappavigna, P. Colaneri, S. Kirkland and R. Shorten, "Essentially negative news about positive systems", *Linear Algebra and its Applications*, vol. 436, no. 9, pp. 3425-3442, May 2012.

[11] F. Borrelli, A. Bemporad, M. Fodor and D. Hrovat, "An MPC/hybrid system approach to traction control", *IEEE Transactions on Control Systems Technology*, vol. 14, no. 3, pp. 541-552, May 2006.

[12] P. Falcone, H. E. Tseng, F. Borrelli, J. Asgari and D. Hrovat, MPC based yaw and lateral stabilisation via active front steering and braking, Vehicle System Dynamics, vol. 46, supp. 1, pp. 611-628, 2008.

[13] R. Shorten, M. Corless, S. Sajja and S. Solmaz, "On Pade approximations, quadratic stability and discretization of switched linear systems", *Systems & Control Letters*, vol. 60, no. 9, pp. 683689, Sept. 2011.

[14] R. G. Sanfelice and A. R. Teel, "Dynamical properties of hybrid systems simulators", *Automatica*, vol. 46, no. 2, pp. 239-248, Feb. 2010.

[15] Md. Arifur Rahman and Abdullah Al Mamun, "Nonlinearity Analysis, Modeling and Compensation in PZT Micro Actuator of Dual-stage Actuator System", accepted for the publication *in Proc. IEEE International Conference on Control and Automation, Taiwan, 2014*.

[16] Md. Arifur Rahman, Abdullah Al Mamun, Kui Yao, Yohanes Daud "Particle Swarm Optimization based Modeling and Compensation of Hysteresis of PZT Micro-actuator used in High Precision Dual-stage Servo System", accepted for the publication *in Proc. IEEE International Conference on Mechatronics and Automation*, China, 2014.

[17] Das, S.K.; Pota, H.R.; Petersen, I.R., "Stability analysis for interconnected systems with "mixed" negative-imaginary and passivity," *in Proc. 3rd Australian Control Conference*, Nov. 2013, pp.445-449.

[18] T. Glad and L. Ljung, *Control Theory: Multivariable and Nonlinear Methods*, London and New York: Taylor & Francis, 2000.

[19] K. Zhou, J. C. Doyle, and K. Glover, *Robust and Optimal Control*, N. Upper Saddle River, Ed. Prentica Hall, 1996.

[20] P. Lancaster and M. Tismenetsky, *The Theory of Matrices with Applications*, San Diego and London: Academic Press, 1985.

[21] B. Noble and J. W. Daniel, *Applied Linear Algebra*, Englewood Cliffs, NJ: Prentice Hall, 1988.

Evaluating the Performance of Optimal Control based Sensor Deployment Algorithms for Realistic Terrain Model

Taynan Sabri Helal, Pritom Mozumdar and Lutfa Akter
Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology, Dhaka-1000, Bangladesh
Email: lutfaakter@eee.buet.ac.bd

***Abstract*—The extensive use of sensor networks in detection and surveillance systems has made optimal sensor deployment a key issue. A lot of sensor deployment techniques have been devised. While a number of mobile sensor networks consider the presence of obstacles and other terrain features in the region of interest, most static sensor networks either consider an idealistic basic terrain model or model the obstacles as straight lines. The aim of this paper is to extend upon a previously proposed optimal sensor deployment technique and incorporate into it a more realistic terrain model. This paper considers only those terrain features which can be modeled by convex polygons. Optimal sensor deployment together with a realistic terrain model makes sensor deployment more practical and robust. Simulation results show that the the number of sensors, their locations and the node density change in presence of obstacles.**

***Index Terms*—Sensor deployment, Optimal control, Realistic terrain**

I. Introduction

Sensor networks are essential for the design and implementation of detection and surveillance systems. Applications of sensor networks include military, environmental, medical, industrial and even home applications. There are many aspects of a sensor network. An important and interesting one is that of sensor deployment. Sensor deployment is the intelligent positioning of sensor nodes in the region of interest (*RoI*) to detect or monitor a physical phenomenon referred to as the *target*. The *RoI* is the area where the target is likely to be detected and monitored.

The issue of sensor deployment has been extensively studied and a number of deployment techniques have been devised. The focus of these techniques is coverage as well as connectivity. A number of computational geometry based deployment techniques has appeared in literature. In [1] the deployment of both static and mobile sensors is considered using a deployment technique called *Bidding Protocol*. The initial deployment is random. Static nodes calculate coverage holes and place *bids* proportional to it. A mobile node has a *base price* proportional to the coverage hole it leaves when it moves. If a *bid* is higher than the *base price*, the mobile sensor moves to heal that coverage hole. In [2], a technique called *The Minimax Algorithm* is used to deploy mobile sensors.The initial deployment is random. The aim is to move a sensor to a point called *Minimax point*. The variance of distances from *Minimax point* to Voronoi vertices is minimum. This means a better coverage within the detection radius of the sensor.

Deployment techniques based on *Virtual Forces, (VF)* also appear in literature. In [3], it is suggested that a sensor is subject to three virtual forces: an attractive force exerted by preferential areas, a repulsive force exerted by obstacles and a force (attractive/repulsive) exerted by neighboring sensors. The initial deployment is random. If the coverage is inadequate, the *Virtual Forces Algorithm (VFA)* is executed by the cluster head to find the net force on each sensor and their new locations. When the threshold coverage is obtained, the cluster head communicates the final positions resulting in a one-time redeployment. In [4], *VFA* is modified on the basis that the original *VFA* leaves coverage holes even after convergence. It is stated that the forces exerted on a sensor by its neighbors cancel each other so that the sensor reaches equilibrium even though some area still remains uncovered. It is proposed that only forces exerted by the *adjacent sensors* should be considered instead of neighboring sensors. The sensors which are connected in Delaunay triangulation of the initial random deployment are defined as *adjacent sensors*.

In [5], a grid based *RoI* with exponential sensing model is considered. The overall miss probability due to a new sensor is calculated for all the candidate grid points. The sensor is deployed at a grid point with minimum overall miss probability according to the algorithm *MIN_MISS* which considers uniform detection requirements. In [6], it is shown that the logarithmic collective miss probability for the grid points is Linear Shift Invariant (LSI). The sensor deployment problem is formulated as an integer linear programming problem which is NP-hard. The sensors are deployed using a heuristic algorithm *DIFF_DEPLOY* which can be used for both uniform and differentiated detection requirements.

In [7] a control theoretic approach is used to solve the sensor deployment problem. By using the concept of control theory and optimization, the optimal locations of the sensors are found. Optimal sensor deployment is minimalistic in nature as it ensures the use of minimum number of sensors. Minimum number of sensors means minimized cost, reduced computation and communication and less sensor management issues (such as replacing faulty ones) while still providing threshold coverage. Minimum number of sensors also ensures that the sensor network is stealthy (if necessary) and that additional sensors can be deployed only for increased reliability.

Among the mobile sensor networks, *VFA* considers the presence of obstacles. Among the static sensor networks, [5]-[7] consider obstacles. However, these obstacles are modeled by straight lines parallel to either of the axis. The effect of obstacles of different shapes on routing protocols is considered in [8]. In [9], the effect of obstacles having the shape of

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

polygons on both coverage and connectivity is considered.

In this paper, the work in [7] is extended and the effect of *two dimensional (2D)* obstacles on sensor deployment and coverage is studied. The algorithms presented in [7] is modified and simulated. The original algorithms ensure optimal deployment. Modification to these algorithms does not compromise optimality. Rather the effects of a more realistic terrain model is incorporated into it.

II. System Model

A. Modeling the Sensor

A probabilistic sensor detection model is used. More specifically the exponential detection model is used:

$$p_{detect}(t,s) = \begin{cases} e^{-\alpha d(t,s)} & \text{if } d(t,s) \leq r_d \\ 0 & \text{if } d(t,s) > r_d \end{cases} \quad (1)$$

Here r_d is the sensing range and α is a parameter which depends on the sensor. These factors can be found out from experiments. The distance between the target(t) and the sensor(s) is denoted by $d(t,s)$. Both $d(t,s)$ and r_d are measured in units of grid points. It is assumed that the sensors follow the line-of-sight (LOS) principle.

B. Modeling the RoI and Terrain

The *RoI* is modeled as a $2D$ grid of points denoted by $\mathcal{G}$. There are N_x points along one axis and N_y along the other giving a total of $N_x \times N_y$ points. Terrain in its simplest sense means the features of an area. In our case, it means the features of the *RoI*. It includes natural features such as vegetation, forest, water bodies, hilly areas, rocks, uneven surfaces and so on. It also includes man made structures such as buildings and walls. For indoors walls, furniture and other equipments can be considered as terrain features.

A basic terrain model assumes a plain *RoI*. That is a flat, open and unobstructed *RoI*. A basic terrain model is ideal for sensor deployment. However, in reality there may be obstacles present in the *RoI* and so deploying sensors assuming a basic terrain model is not appropriate. If the terrain features are known beforehand then that knowledge should be used to modify the sensor deployment accordingly. In this paper, knowledge about the terrain will be incorporated with existing deployment algorithms. Since the *RoI* is *2D* so only those terrain features which can be modeled by *2D* obstacles are considered. The *2D* obstacles in turn are modeled by convex polygons. Obstacles of other shapes can be approximated using convex polygons.

C. Problem Formulation and Solution

The problem formulation is stated as: To find the location of sensors in an *RoI* with a realistic terrain model such that the effective squared error (eSE) is minimized after each sensor deployment. The eSE is the square of the error between the achieved and required overall logarithmic miss probabilities (logarithmic miss probability is defined as logarithm of (1-detection probability)) at those grid points for which achievement is less than the requirement. The problem formulation is similar to that stated in [7]. The modification in this paper is that a more realistic terrain model is used by taking into account *2D* obstacles.

In [7], the formulated problem is solved by a control theoretic approach. The sensor deployment problem is at first converted to a control theory problem, specifically as a linear-quadratic regulator (LQR). The maximum value of the control vector in the k_{th} step is the location of the k_{th} sensor. The solution is based on the concept of dynamic optimization. The optimality conditions are found out using method of Lagrange multipliers and then applying discrete Karush-Kuhn-Tucker (KKT) conditions. Finally, the optimal solution is found by the sweep method. In this paper, the same solution procedure is used. The mathematical equations used for problem formulation and solution are given in [7] and not repeated here due to page limitation. The relevant terms used in this paper are defined in Table I.

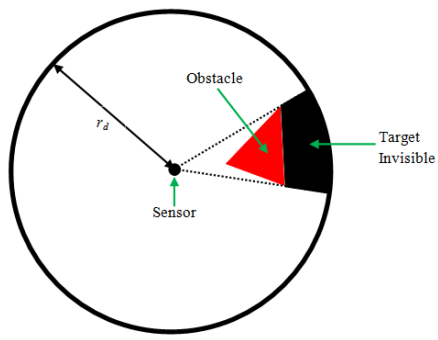


Fig. 1. The presence of an obstacle limits visibility of the sensor

TABLE I
NOTATIONS

$\mathbf{B}$	Node detection model ($N_xN_y \times N_xN_y$)
$\mathbf{p}_d^{req}$ / $\mathbf{p}_m^{req}$	Detection probability requirement vector / Miss probability requirement vector ($N_xN_y \times 1$)
$\mathbf{m}_{req}(= \ln(\mathbf{p}_m^{req}))$	Logarithmic miss probability requirement vector ($N_xN_y \times 1$)
$\mathbf{D}$	Deployment vector ($N_xN_y \times 1$)
$\mathbf{u}_k$	Control vector at the k_{th} step ($N_xN_y \times 1$)
$\mathbf{x}_k$	Difference between achieved and required overall logarithmic miss probabilities after k_{th} sensor deployment ($N_xN_y \times 1$)
$\mathbf{G}_k$	Feedback gain matrix for k_{th} step ($N_xN_y \times N_xN_y$)
$\mathbf{R}$, $\mathbf{Q}$, $\mathbf{Q_f}$	Diagonal weighting matrices ($N_xN_y \times N_xN_y$)

III. Deployment Algorithm

In [7], two deployment algorithms are proposed and simulated. These are the *Optimal Control Based Algorithm* (it will be referred to as *Optimal_Control* henceforth) and *Max_Deficiency*.

In *Optimal_Control*, the optimal location of the sensor is found out from the equation:

$$\mathbf{u}_k = -\mathbf{G}_k\mathbf{x}_k \quad (2)$$

The index corresponding to the maximum value of the control vector $\mathbf{u}_k$ gives the optimal location of the k_{th} sensor. Here $\mathbf{x}_k$ is the difference or error between achieved and required overall logarithmic miss probabilities and $\mathbf{G}_k$ is the feedback gain matrix given by the sweep method. In *Max_Deficiency*, a sensor is deployed at the grid point where the difference between the the required and achieved detection probability is maximum. The difference between required and achieved detection probability is referred to as *deficiency*. In this paper, the two algorithms are modified to deploy sensors in a more

realistic terrain. The modifications to the two algorithms are now discussed.

A. Optimal_Control

The original algorithm appears in [7]. The modifications are made to the inputs of the algorithm. These include matrix $\mathbf{B}$, column vectors $\mathbf{p}_d^{req}$/$\mathbf{p}_m^{req}$ and matrices $\mathbf{R}$, $\mathbf{Q}$ and $\mathbf{Q_f}$.

1) Matrix $\mathbf{B}$*:* The matrix $\mathbf{B}$ is the node detection model. The individual elements can be written as $b(r,c)$. If a sensor is at (i,j), target is at at (x,y) then $b(r,c) = \ln(1 - p_{detect})$ where p_{detect} is given by (1), $r = (x-1)N_y + y$ and $c = (i-1)N_y + j$.

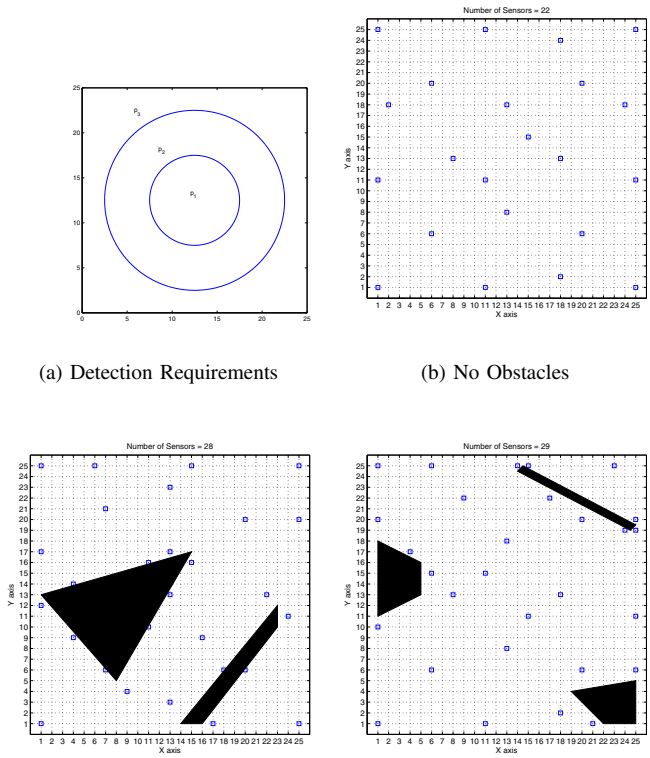
The effect of obstacles is shown in Fig. 1. The sensors are assumed to work on the LOS principle. A target is detected if it is within the the detection radius r_d and LOS exits with the sensor. In the absence of obstacles, targets at all grid points within the detection range are detected according to (1). In the presence of obstacles, if a target is at any grid points beyond the obstacle (such as the shaded region in Fig. 1), LOS does not exist and the target is *invisible* to the sensor. For multiple obstacles, more grid points will become *invisible* to the sensor. Hence for these points p_{detect} becomes zero making the corresponding entries of $\mathbf{B}$ zero as well. Also obstacles themselves occupy grid points in the *RoI*. These grid points are referred to as *inaccessible* points. It is assumed that neither the sensor nor the target can be placed at these grid points. Thus, the corresponding entries in $\mathbf{B}$ are set to zero.

2) Column vectors $\mathbf{p}_d^{req}$*,* $\mathbf{p}_m^{req}$ *and* $\mathbf{m}_{req}$*:* The value of $\mathbf{p}_d^{req}$/$\mathbf{p}_m^{req}$ is specified by the user. The following modifications are made: $\mathbf{p}_d^{req}(inaccessible) = 0$ and $\mathbf{p}_m^{req}(inaccessible) = 1$ leading to $\mathbf{m}_{req}(inaccessible) = 0$. It is assumed that the *inaccessible* points will not be monitored and so there is no need to have any detection requirements for these points.

Algorithm 1 Modification to the Algorithm

% The target (t) is at (x,y), the sensor (s) is at (i,j) and the set of grid points occupied by the obstacles is *inaccessible*
% Modification to matrix $\mathbf{B}$
if $(i,j) \in$ *inaccessible* $||$ $(x,y) \in$ *inaccessible* **then**
  $b(r,c) = 0$
else if $d(t,s) \leq r_d$ && *LOS exists between t and s* **then**
  $b(r,c) = \ln(1 - p_{detect})$ % where p_{detect} is given by (1)
else
  $b(r,c) = 0$
end if
% Modification to column vectors $\mathbf{p}_d^{req}$, $\mathbf{p}_m^{req}$ and $\mathbf{m}_{req}$
$\mathbf{p}_d^{req}(inaccessible) = 0$
$\mathbf{p}_m^{req}(inaccessible) = 1$, $\mathbf{m}_{req}(inaccessible) = 0$
% Modification to matrices $\mathbf{R}$, $\mathbf{Q}$ and $\mathbf{Q_f}$
$\mathbf{Q}(inaccessible) = \mathbf{Q_f}(inaccessible) = 0$
$\mathbf{R}(inaccessible)$ set to a large value

3) Matrices $\mathbf{R}$*,* $\mathbf{Q}$ *and* $\mathbf{Q_f}$*:* Matrices $\mathbf{R}$, $\mathbf{Q}$ and $\mathbf{Q_f}$ are diagonal weighting matrices. They are given by $\mathbf{R}(i) = (\mathbf{m}_{req}(i)/\mathbf{1}^T\mathbf{m}_{req})^{-1}$ and $\mathbf{Q}(i) = \mathbf{Q_f}(i) = (\mathbf{R}(i))^{-1}$. These weighting matrices ensure that the solution is biased towards grid points having higher detection requirements. Modification of $\mathbf{m}_{req}$ at the *inaccessible* points automatically modifies these weighting matrices such that $\mathbf{Q}(inaccessible) = \mathbf{Q_f}(inaccessible) = 0$ while $\mathbf{R}(inaccessible)$ is set to a large value. This reduces the bias towards the *inaccessible* points and hence sensor deployment on the available grid points is emphasized.



(a) Detection Requirements (b) No Obstacles

(c) Obstacle Set A (d) Obstacle Set B

Fig. 2. Simulation Profile and Sensor Deployment for First Experiment

For convenience, again the modification part to *Optimal_Control* algorithm is summarized in Algorithm 1.

The algorithm *Optimal_Control* comprises of two parts. First is the algorithm named *Sweep method* with inputs $\mathbf{B}$ and cost matrices $\mathbf{Q}$, $\mathbf{Q}_f$ and $\mathbf{R}$. In this paper, the *Sweep method* is executed and $\mathbf{G}_k$ are found out using modified $\mathbf{B}$, $\mathbf{Q}$, $\mathbf{Q}_f$ and $\mathbf{R}$. The second part is the actual deployment algorithm with inputs $\mathbf{B}$ and $\mathbf{p}_d^{req}$. In this paper, the deployment algorithm is carried out using modified $\mathbf{B}$ and $\mathbf{p}_d^{req}$ and sensor positions are determined from (2). As a result, the deployment vector $\mathbf{D}$ gives the optimal sensor positions taking into account the obstacles.

B. Max_Deficiency

The original algorithm appears in [7]. The inputs of the algorithm are matrix $\mathbf{B}$ and column vector $\mathbf{p}_d^{req}$. The modification to $\mathbf{B}$ and the column vector $\mathbf{p}_d^{req}$ are the same as that for *Optimal_Control*. Using modified inputs, the most *deficient* grid points in the presence of obstacles are found out and sensors are deployed according to deployment vector $\mathbf{D}$.

IV. Simulation Results

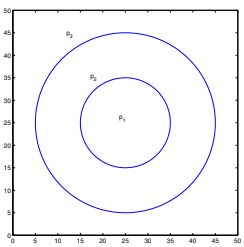
In this section, sensor deployment in presence of obstacles is compared to sensor deployment in a basic terrain model with no obstacles. The sensor parameters used are $\alpha = 0.2$ and $r_d = 7$.

In the first experiment, the effect of obstacles on sensor deployment by *Optimal_Control* is simulated. The *RoI* is modeled as a 25×25 grid. The detection requirements in

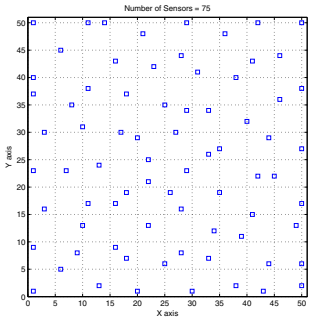
TABLE II
FIRST EXPERIMENT RESULTS

Detection Requirements	No Obstacles	Obstacle Set A	Obstacle Set B
$p_1 = 0.85, p_2 = 0.75, p_3 = 0.65$	22	28	29
$p_1 = 0.75, p_2 = 0.65, p_3 = 0.55$	21	24	25
$p_1 = 0.65, p_2 = 0.65, p_3 = 0.65$	17	22	25

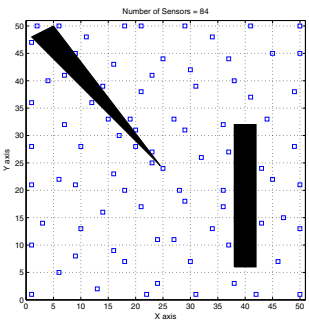
the *RoI* is given in Fig. 2(a). These detection requirements are specified by the user and can be varied. As an example, Fig.2 shows the sensor deployment for $p_1 = 0.85$, $p_2 = 0.75$ and $p_3 = 0.65$. In Fig. 2(b), a basic terrain model is assumed with no obstacles. The number of sensors deployed is 22. In Fig. 2(c) and Fig. 2(d), the presence of $2D$ obstacles are considered to model the features of the terrain. It is seen that both the number of sensors deployed and their locations in each case are different from the basic terrain model. For obstacle set A 28 sensors are needed and for obstacle set B 29 sensors are needed. One other observation is that the increased number of sensors are deployed over a smaller area. In the basic terrain model in Fig. 2(b), the sensors are deployed all over the *RoI* and provide coverage throughout. However, in Fig. 2(c) and Fig. 2(d) since the obstacles themselves occupy a certain amount of area the sensors are deployed over a reduced area. This means that the node density increases in the presence of obstacles. It can be observed in Fig. 2(c) and Fig. 2(d) that a greater number of nodes are deployed close to the obstacles. The detection requirements are varied and the number of sensors deployed recorded in Table II.



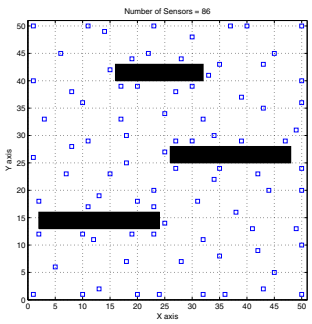
(a) Detection Requirements



(b) No Obstacles



(c) Obstacle Set A



(d) Obstacle Set B

Fig. 3. Simulation Profile and Sensor Deployment for Second Experiment

In the second experiment, the effect of obstacles on sensor deployment by *Max_Deficiency* is simulated. Due to the computational complexity of *Optimal_Control*, large grid sizes cannot be simulated. Since *Max_Deficiency* can simulate larger grid sizes, the *RoI* in the second experiment is a 50×50 grid. The detection requirements in the *RoI* is given in Fig. 3(a). As an example, Fig. 3 shows the sensor deployment for $p_1 = 0.85$, $p_2 = 0.75$ and $p_3 = 0.65$. In Fig. 3(b), a basic terrain model is assumed with no obstacles. The number of sensors deployed is 75. In the presence of obstacles, the number of sensors is 84 and 86 as shown in Fig. 3(c) and Fig. 3(d), respectively. Similar to the first experiment, the number of sensors, their locations and the node density are different compared to a basic terrain model. It is also observed that the node density increases in Fig. 3(c) and Fig. 3(d) with greater number of nodes deployed close to the edges of the obstacles. The detection requirements are varied and the number of sensors deployed recorded in Table III.

TABLE III
SECOND EXPERIMENT RESULTS

Detection Requirements	No Obstacles	Obstacle Set A	Obstacle Set B
$p_1 = 0.85, p_2 = 0.75, p_3 = 0.65$	75	84	86
$p_1 = 0.75, p_2 = 0.65, p_3 = 0.55$	66	73	69
$p_1 = 0.65, p_2 = 0.65, p_3 = 0.65$	69	72	72

V. CONCLUSION

In this paper, the problem of sensor deployment in a realistic terrain model is studied and simulated. The study of optimal sensor deployment proposed in a previous work is modified to take into account the terrain features of the *RoI*. The algorithms *Optimal_Control* and *Max_Deficiency* are modified and simulated. It is found that the number of sensors deployed are affected by presence of obstacles. The locations of the sensors in case of a realistic terrain model are found out. It is also seen that the node density increases over the *RoI* especially along the edges of the obstacles.

REFERENCES

[1] G. Wang, G. Cao, and T. LaPorta, "A bidding protocol for deploying mobile sensors," in *Network Protocols, 2003. Proceedings. 11th IEEE International Conference on*. IEEE, 2003, pp. 315–324.

[2] G. Wang, G. Cao, and T. La Porta, "Movement-assisted sensor deployment," *Mobile Computing, IEEE Transactions on*, vol. 5, no. 6, pp. 640–652, 2006.

[3] Y. Zou and K. Chakrabarty, "Sensor deployment and target localization based on virtual forces," in *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, vol. 2. IEEE, 2003, pp. 1293–1303.

[4] X. Yu, W. Huang, J. Lan, and X. Qian, "A novel virtual force approach for node deployment in wireless sensor network," in *Distributed Computing in Sensor Systems (DCOSS), 2012 IEEE 8th International Conference on*. IEEE, 2012, pp. 359–363.

[5] Y. Zou and K. Chakrabarty, "Uncertainty-aware and coverage-oriented deployment for sensor networks," *Journal of Parallel and Distributed Computing*, vol. 64, no. 7, pp. 788–798, 2004.

[6] J. Zhang, T. Yan, and S. H. Son, *Deployment strategies for differentiated detection in wireless sensor networks*. IEEE, 2006, vol. 1.

[7] A. Ababnah and B. Natarajan, "Optimal control-based strategy for sensor deployment," *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, vol. 41, no. 1, pp. 97–104, 2011.

[8] I. Chatzigiannakis, G. Mylonas, and S. Nikoletseas, "Modeling and evaluation of the effect of obstacles on the performance of wireless sensor networks," in *Proceedings of the 39th annual Symposium on Simulation*. IEEE Computer Society, 2006, pp. 50–60.

[9] Y.-C. Wang, C.-C. Hu, and Y.-C. Tseng, "Efficient deployment algorithms for ensuring coverage and connectivity of wireless sensor networks," in *Wireless Internet, 2005. Proceedings. First International Conference on*. IEEE, 2005, pp. 114–121.

Robust Control of Robot Manipulator based on Estimation of Upper Bounds on Parametric Uncertainty

Amol A. Khalate,[1] Rajeeb Dey,[2] and Goshaidas Ray[1,*]

[1] Department of Electrical Engineering, Indian Institute of Technology
I.I.T Kharagpur, West Bengal, India.

[2] Department of Electronics and Communication, Sikkim Manipal Institute
West Bengal, India.

[*]gray@ee.iitkgp.ernet.in

Abstract— **This paper proposes a robust control law for trajectory tracking problem of an n-link robot manipulator subjected to parametric uncertainty. The control law is derived based on Lyapunov theory, which guarantees asymptotic stability of the system. The upper bounds on the parametric uncertainties are determined with estimation law and an additional control input is updated according to the estimated value. The proposed methods improve tracking performance and provide less control effort with no chattering effect in control torques. The simulation results of 2-link robot manipulator are presented to validate the effectiveness of the proposed method.**

Index Terms: **Parametric uncertainty, Lyapunov's stability theorem, Bound estimation, Robust control.**

I. Introduction

In past two decades, the various optimal schemes have been developed to solve robotic manipulator problem [1]-[3]. Trajectory tracking problem is the most significant and fundamental task in control of robotic manipulator. The major source of uncertainty in robot manipulator is parametric uncertainty, which arises due to variations in payload and mass of links. The adaptive control approach proposed in [4] is very attractive to deal with the parametric uncertainty in robot manipulator. The novelty of the approach in [5] lies in the fact that uncertainty bounds information are necessary to derive the control law and the uniform ultimate boundedness of the tracking error depend only on the inertia parameters of robot manipulator. In [6], [7] a robust controller is designed based on estimation of upper bounds on parametric uncertainty, but in [7] priory knowledge about the upper bounds on parametric uncertainty is not required.

This paper is concerned with the design of a robust control law for trajectory tracking control of robot manipulator based on the approaches in [5] and [7]. The proposed method introduces an additional control input that uses the estimate of upper bounds of uncertainty; and the estimator is an exponential function of maximum singular value of a matrix that is related with the robot kinematics, inertia parameters and tracking error. The designed robust control law comprises with two terms, one act as a dynamic feed-forward to ensure robustness to the parametric uncertainty and the other acts as a simple PD feedback term to regulate the transient behavior. The proposed controllers significantly improve the tracking performance and also provide less control effort with no chattering effect in input torques.

II. Dynamic Model and Robust Control

An n-DOF robot manipulator is described by

$$M(q)\ddot{q}+C(q,\dot{q})\dot{q}+g(q)=\tau \text{ with } q\in\Re^n,\ \tau\in\Re^n \quad (1)$$

where $M(q)$ is $n\times n$ inertia matrix, the vector $C(q,\dot{q})\dot{q}$ represents centrifugal and coriolis forces, and $g(q)$ is the $(n\times 1)$ vector of gravitational forces. The above dynamic equation (1) is linearly parameterizable [5]

$$M(q)\ddot{q}+C(q,\dot{q})+g(q)=Y(q,\dot{q},\ddot{q})\theta=\tau \quad (2)$$

where θ is a constant $(px1)$ -dimensional vector of robot parameters and Y is an $(n\times p)$ matrix of known functions of the generalized coordinates and their higher derivatives.

It is assumed that $\exists$ an unknown bound on parametric uncertainty such that

$$\|\tilde{\theta}\|=\|\theta-\theta_0\|\le\rho \quad (3)$$

where θ_0 represents fixed robot parameters. In order to control dynamic behaviour of the system properly, we need to estimate ρ in an efficient manner. For any specific trajectory, $q_d,\dot{q}_d,\ddot{q}_d$ are desired position, velocity, acceleration and the corresponding position and velocity error are defined as $\tilde{q}=q-q_d$, $\dot{\tilde{q}}=\dot{q}-\dot{q}_d$. The following reference velocity and acceleration are defined as

$$\dot{q}_r=\dot{q}_d-\Lambda\tilde{q},\ \ \ddot{q}_r=\ddot{q}_d-\Lambda\dot{\tilde{q}} \quad (4)$$

and reference velocity error is given as

$$\sigma=\dot{q}-\dot{q}_r=\dot{\tilde{q}}+\Lambda\tilde{q} \quad (5)$$

The nominal control law τ_0 is expressed as

$$\tau_0=M_0(q)\ddot{q}_r+C_0(q,\dot{q})\dot{q}_r+g_0(q)-K\sigma=Y(q,\dot{q},\dot{q}_r,\ddot{q}_r)\theta_0-K\sigma \quad (6)$$

where the gain matrices K and Λ are positive definite diagonal matrices. Now, we define the control torque τ in terms of nominal control vector τ_0 as

$$\tau=\tau_0+Y(q,\dot{q},\dot{q}_r,\ddot{q}_r)u=Y(q,\dot{q},\dot{q}_r,\ddot{q}_r)(\theta_0+u)-K\sigma \quad (7)$$

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

Where, u is an additional control input that is designed to achieve robustness to parametric uncertainty represented by $\tilde{\theta} = \theta - \theta_0$. From (7) and (1), after simplification we get,

$$M(q)\sigma + C(q,\dot{q})\sigma + K\sigma = Y(q,\dot{q},\dot{q}_r,\ddot{q}_r)(\tilde{\theta} + u) \quad (8)$$

Theorem 1: Consider the upper uncertainty bound ρ is known priori and if $\exists$ $\hat{\rho}(t)$ an estimate of ρ, and an additional control law $u(t) = -\beta\,\hat{\rho}(t)$ (where β is $(px1)$ dimensional unit vector) such that control law (7) stabilizes the closed-loop system as well as improves the performance of the tracking error when the estimator is designed as

$$\hat{\rho}(t) = \begin{cases} \dfrac{2\sigma^T Y\beta}{s_{\max}}\left(e^{(-1/2)\int s_{\max}\,dt} - e^{-\int s_{\max}\,dt}\right) + \rho & if\ \ s_{\max} > \varepsilon \\ \dfrac{2\sigma^T Y\beta}{\varepsilon}\left(e^{(-1/2)\int s_{\max}\,dt} - e^{-\int s_{\max}\,dt}\right) + \rho & if\ \ s_{\max} \le \varepsilon \end{cases}$$

where $s_{\max}$ = maximum singular value of $(Y^T\sigma\sigma^T Y)$ and $\tilde{\rho}(t) = \rho - \hat{\rho}(t)$.

Proof: Let the Lyapunov function as

$$V(\sigma, q, \tilde{\rho}(t)) = \frac{1}{2}\sigma^T M(q)\sigma + \frac{1}{2}\tilde{q}^T B\tilde{q} + \frac{1}{2}\tilde{\rho}(t)\psi(t)\tilde{\rho}(t) \quad (9)$$

where B is positive definite diagonal matrix and $\psi(t) > 0\,(scalar)$. The $\dot{V}(t)$ along the trajectories is

$$\dot{V} = \sigma^T M(q)\dot{\sigma} + \frac{1}{2}\sigma^T \dot{M}(q)\sigma + \tilde{q}^T B\dot{\tilde{q}} + \frac{1}{2}\tilde{\rho}(t)\dot{\psi}(t)\tilde{\rho}(t) + \tilde{\rho}(t)\psi(t)\dot{\tilde{\rho}}(t) \quad (10)$$

Using the skew-symmetric properties $\left[\dot{M}(q) - 2C(q,\dot{q})\right] = 0$, and assigning $(B - 2\Lambda K) = 0$, one can rewrite (10) as

$$\dot{V} = -\dot{\tilde{q}}^T K\dot{\tilde{q}} - \tilde{q}^T \Lambda K \Lambda\tilde{q} + \sigma^T Y(\tilde{\theta} + u(t)) + \frac{1}{2}\tilde{\rho}(t)\dot{\psi}(t)\tilde{\rho}(t) + \tilde{\rho}(t)\psi(t)\dot{\tilde{\rho}}(t) \quad (11)$$

With $\tilde{\theta} \le \beta \cdot \rho$ and $u(t) = -\beta \cdot \hat{\rho}(t)$, the above equation reduces to

$\dot{V} \le -\dot{\tilde{q}}^T K\dot{\tilde{q}} - \tilde{q}^T \Lambda K \Lambda\tilde{q} < 0$ (sin$ce\ K > 0$) if the rest of the part of (11) is assigned to zero i.e.

$$\sigma^T Y\beta + \dot{\psi}(t)(\rho - \hat{\rho}(t)) - \psi(t)\dot{\hat{\rho}}(t) = 0 \quad (12)$$

We defined $\psi(t)$ and its time derivative as follows

$$\psi(t) = e^{\int S_{\max}\,dt},\ \dot{\psi}(t) = S_{\max} e^{\int S_{\max}\,dt} \quad (13)$$

Substituting $\psi(t)$ and $\dot{\psi}(t)$ in (12) and subsequently multiplying both sides by $e^{(-1/2)\int S_{\max}\,dt}$ and rearranging, we get

$$\frac{d}{dt}\left[e^{(1/2)\int s_{\max}\,dt}\hat{\rho}(t)\right] = \sigma^T Y\beta \cdot e^{(-1/2)\int s_{\max}\,dt} + \frac{1}{2}s_{\max} \cdot e^{(1/2)\int s_{\max}\,dt}\rho \quad (14)$$

Integrating both sides of (14), and then multiplying the final expression by $e^{(-1/2)\int S_{\max}\,dt}$, we have

$$\hat{\rho}(t) = -\frac{2\sigma^T Y\beta}{s_{\max}} e^{-\int S_{\max}\,dt} + \rho + c \cdot e^{(-1/2)\int S_{\max}\,dt} \quad (15)$$

If we choose initial condition $\hat{\rho}(0) = \rho$, then integration constant is obtained as $c = \dfrac{2\sigma^T Y\beta}{s_{\max}}$ and (15) is rewritten as,

$$\hat{\rho}(t) = \frac{2\sigma^T Y\beta}{s_{\max}}\left(e^{(-1/2)\int s_{\max}\,dt} - e^{-\int s_{\max}\,dt}\right) + \rho$$

If the reference velocity tracking error ($\sigma = \dot{q} - \dot{q}_r = \dot{\tilde{q}} + \Lambda\tilde{q}$) approaches to zero, the corresponding value of $s_{\max}$ also approaches to zero. This problem may be eliminated using the modified estimation law as

$$\hat{\rho}(t) = \begin{cases} \dfrac{2\sigma^T Y\beta}{s_{\max}}\left(e^{(-1/2)\int s_{\max}\,dt} - e^{-\int s_{\max}\,dt}\right) + \rho & if\ \ s_{\max} > \varepsilon \\ \dfrac{2\sigma^T Y\beta}{\varepsilon}\left(e^{(-1/2)\int s_{\max}\,dt} - e^{-\int s_{\max}\,dt}\right) + \rho & if\ \ s_{\max} \le \varepsilon \end{cases} \quad (16)$$

where ε is small positive number.

Consequently, the additional control input is

$$u(t) = -\beta\,\hat{\rho}(t) \quad (17)$$

Noting $\hat{\rho}(t)$ is obtained in such way that (12) is satisfied, and substituting (17) in (11), we get $\dot{V} = -\dot{\tilde{q}}^T K\dot{\tilde{q}} - \tilde{q}^T \Lambda K \Lambda\tilde{q} < 0$. Hence, the closed-loop system is asymptotically stable and improves the tracking errors $\tilde{q}$ and $\dot{\tilde{q}}$ depending upon the choice of Λ. □

Remark-1: In the above proof if we choose different positive scalar function $\psi(t)$, we can get different estimation law. Now by choosing $\psi(t) = e^{\int \sigma^T Y\beta\,dt}$, we have derived different estimation law and controller is defined in following theorem.

Theorem 2: Let the upper uncertainty bound ρ is known priori and the estimate of ρ is defined as

$\hat{\rho}(t) = 2\left(e^{(-1/2)\int \sigma^T Y\beta\,dt} - e^{-\int \sigma^T Y\beta\,dt}\right) + \rho$, and the additional control input $u(t) = -\beta\,\hat{\rho}(t)$ ensure the asymptotic stability of the system and the trajectory tracking errors $\tilde{q}$ and $\dot{\tilde{q}}$ converge to zero.

Proof: Using $\psi(t) = e^{\int \sigma^T Y\beta\,dt}$, $\dot{\psi}(t) = \sigma^T Y\beta e^{\int \sigma^T Y\beta\,dt}$ and adopting the procedures given in theorem-1, one can prove the above theorem and details derivations are omitted.

Remark-2: Both the proposed theorems are the modified version of the algorithm described in [7]. It may be noted that the expressions for the uncertain bound ρ is modified to achieve improved tracking error as well as to get less input torque compared to the results obtained in [5] and [7].

III. DESIGN STRATEGY

In [5], the additional input is defined as

$$u = \begin{cases} -\rho\dfrac{Y^T\sigma}{\|Y^T\sigma\|} & \text{if } \|Y^T\sigma\| > \varepsilon;\ or\ -\rho\dfrac{Y^T\sigma}{\|Y^T\sigma\|} & \text{if } \|Y^T\sigma\| \le \varepsilon \end{cases} \quad (18)$$

where $\varepsilon > 0$ and the upper uncertainty bound is defined as

$$\rho = \left(\sum_{i=1}^{p} \rho_i^2 \right)^{1/2} \tag{19}$$

In previous section, derived additional control input $u(t)$ uses a scalar parameter $\hat{\rho}(t)$ as measure of uncertainty, but this may lead to excessive gains [5]. Assigning different gains to the each component of $u(t)$, one can modify the proposed estimation laws in both theorems if $\exists$ an upper bound on each parametric uncertainty as $\left|\tilde{\theta}_i\right| \le \rho_i, \quad i = 1, 2, \cdots, p.$

$$\hat{\rho}_i(t) = \begin{cases} \dfrac{2\sigma^T Y \beta}{s_{\max}} \left(e^{(-1/2)\int s_{\max}\,\mathrm{d}t} - e^{-\int s_{\max}\,\mathrm{d}t} \right) + \rho_i & if \;\; s_{\max} > \varepsilon_i \\ & ; \varepsilon_i > 0 \\ \dfrac{2\sigma^T Y \beta}{\varepsilon} \left(e^{(-1/2)\int s_{\max}\,\mathrm{d}t} - e^{-\int s_{\max}\,\mathrm{d}t} \right) + \rho_i & if \;\; s_{\max} \le \varepsilon_i \end{cases} \tag{20}$$

$$u_i(t) = -\beta \hat{\rho}_i(t) \tag{21}$$

Similarly, each component of estimation law in theorem-2 can also be modified as

$$\hat{\rho}_i(t) = 2\left(e^{(-1/2)\int \sigma^T Y \beta \mathrm{d}t} - e^{-\int \sigma^T Y \beta \mathrm{d}t} \right) + \rho_i; \; u_i(t) = -\beta \hat{\rho}_i(t) \tag{22}$$

If we don't have any priori knowledge about upper bound on uncertainty ρ then one can compute the proposed control law in theorem-1 as

$$\dot{\hat{\rho}}(t) = \left\{ \sigma^T Y \beta \left(2e^{-\int s_{\max}\,\mathrm{d}t} - e^{(-1/2)\int s_{\max}\,\mathrm{d}t} \right) \right. ; \; if \;\; s_{\max} > \varepsilon \;\; or \;\; 0 \;\; if \;\; s_{\max} \le \varepsilon$$

$$u(t) = -\beta\, \hat{\rho}(t) \tag{23}$$

Similarly, control law in theorem-2 can also be computed as

$$\dot{\hat{\rho}}(t) = \sigma^T Y \beta \left(2e^{-\int \sigma^T Y \beta \mathrm{d}t} - e^{(-1/2)\int \sigma^T Y \beta \mathrm{d}t} \right); \; u(t) = -\beta\, \hat{\rho}(t) \tag{24}$$

IV. Simulation Results

To show the effectiveness of the proposed controllers, a 2-link planar manipulator is considered in Fig.1. The matrices and vectors ($M(q)$, $C(q,\dot{q})$, $g(q)$ and $\tau = Y(q,\dot{q},\ddot{q})\theta;\; \theta \in \Re^6$) can be computed using the following parameters [6].

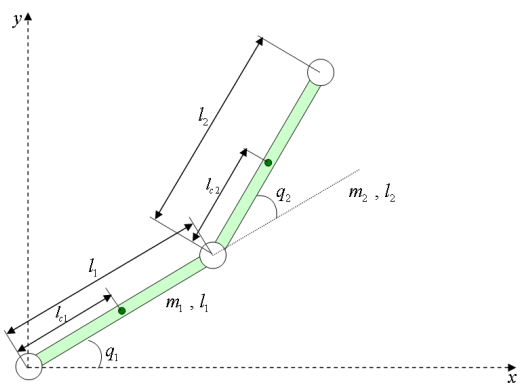


Fig.1. Two-link planar robot.

Using the values from Table I, we can get θ_i as shown in Table II.

TABLE I
Parameters of the Unloaded Arm

m_1	m_2	l_1	l_2	l_{c1}	l_{c2}	I_1	I_2
10	5	1	1	0.5	0.5	10/12	5/12

TABLE II
θ_i for the Unloaded Arm

θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
8.33	1.67	2.5	5	5	2.5

TABLE III
Nominal Parameter Vector θ_0

θ_{01}	θ_{02}	θ_{03}	θ_{04}	θ_{05}	θ_{06}
13.33	8.96	8.75	5	10	8.75

TABLE IV
Uncertainty Bounds ρ_i

ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6
5	7.29	6.25	0	5	6.25

We assume that due to unknown load carried by second link of the robot, the parameters m_2 , l_{c2} and I_2 changes to $m_2 + \Delta m_2$, $l_{c2} + \Delta l_{c2}$ and $I_2 + \Delta I_2$, respectively. A robust controller is designed to ensure asymptotic stability and good tracking performance in presence of parametric uncertainties.

$$0 \le \Delta m_2 \le 10; \qquad 0 \le \Delta l_{c2} \le 0.5; \qquad 0 \le \Delta I_2 \le \frac{15}{12}.$$

The nominal parameter vector θ_0 , shown is in Table III, is obtained after choosing the mean value for the range of possible θ_i . Table IV shows uncertainty bounds ρ_i for each parameter, which is difference of Table III and Table II.

With the uncertainty range and nominal parameters θ_0 , it is possible to compute upper uncertainty ρ as

$$\rho = \left(\sum_{i=1}^{6} \rho_i^2 \right)^{1/2} = \left(\sum_{i=1}^{6} (\theta_{0i} - \theta_i)^2 \right)^{1/2} = 13.46$$

A 2-link robot manipulator is discretized with a sampling time (T) =0.005 sec. and desired trajectory for both the joints are considered as $q_d = 1 - \cos(at)$, $\dot{q}_d = a \sin(at)$ with $a = \dfrac{2\pi}{5}$. To illustrate the novelty of the proposed methods, results of [5] are compared with the proposed controllers stated in theorems 1 and 2. We have chosen $K = diag(75 \;\; 50)$ and $\Lambda = diag(40 \;\; 15)$ for all the controllers. Simulation results based on [5] are presented in Fig.2 where tracking error is about 9.9×10^{-3} rad. for joint-1, and 14.7×10^{-3} rad. for joint-2. The results obtained (in figs.3 and 4) based on theorems-1 and 2 are compared with that of [5]. It has been observed that the both the proposed methods ensure good tracking performance (using theorem-1 tracking error is observed about 2.7×10^{-5} rad. for joint-1, and 2.8×10^{-5} rad. for joint-2 (see Fig.3) whereas tracking error is about 3.1×10^{-5} rad. for joint-1, and 5.6×10^{-5} rad. for joint-2 when theorem-2 is employed (see Fig.4)) as well as require very less control torque compared to the results presented in [5]. Results based on proposed methods are also revealing better performance than that of [7]. Further, it may be noted that the chattering effect on the input torques is reduced to zero (see figs. 3 and 4).

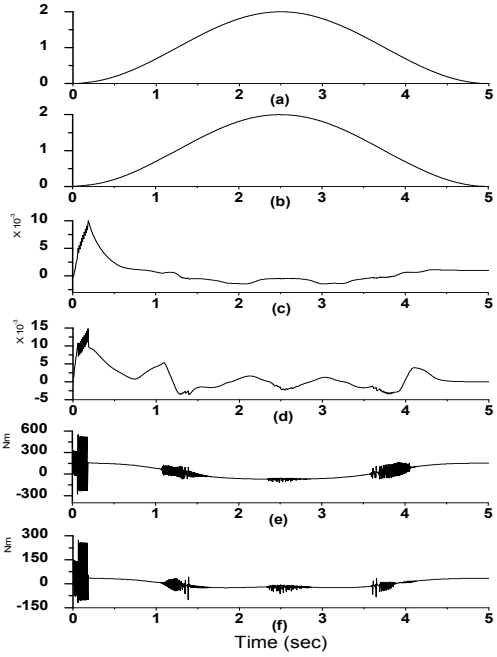


Fig.2. System response [5] when $\varepsilon = 0.1$: (a) Joint-1 position q_1 (b) Joint-2 position q_2 (c) Tracking error $\tilde{q}_1$, (d) Tracking error $\tilde{q}_2$, (e) Torque input τ_1 to joint 1, (f) Torque input τ_2 to joint 2.

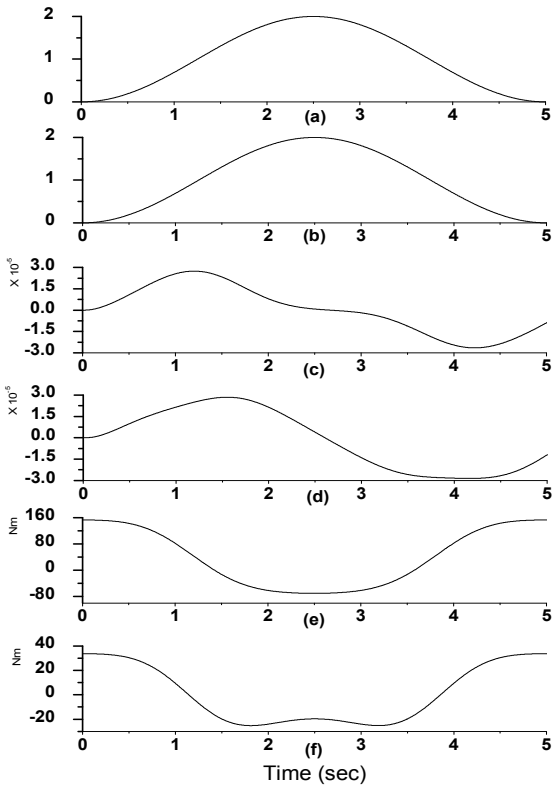


Fig.3. System response based on theorem-1: (a) Joint-1 position q_1 (b) Joint-2 position q_2 (c) Tracking error $\tilde{q}_1$, (d) Tracking error $\tilde{q}_2$, (e) Torque input τ_1 to joint 1, (f) Torque input τ_2 to joint 2.

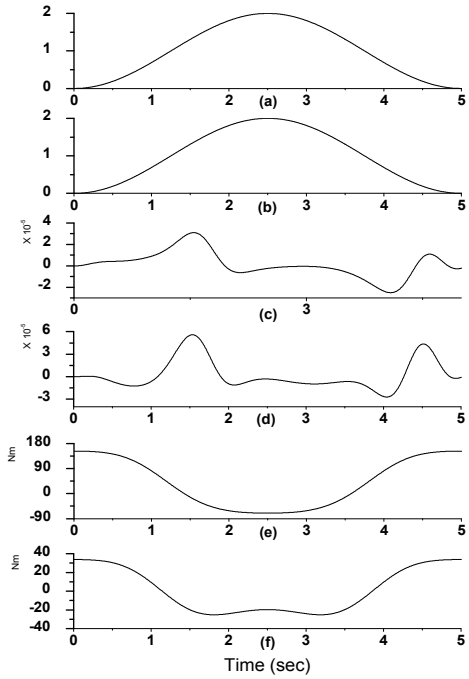


Fig.4. System response based on theorem-2: (a) Joint-1 position q_1 (b) Joint-2 position q_2 (c) Tracking error $\tilde{q}_1$, (d) Tracking error $\tilde{q}_2$, (e) Torque input τ_1 to joint 1, (f) Torque input τ_2 to joint 2.

V. CONCLUDING REMARKS

In this paper, two robust controllers for robot manipulator have been proposed by following the work of [5] and [7]. Stability analysis of such system is derived based on Laypunov function along with the estimated value of upper bounds on parametric uncertainty. Introducing new exponential functions yield improved parameter convergence rate of the proposed estimation laws and also ensure significant good tracking performance. It is also noticed that a reduction in torque magnitude with no chattering effect in input torques. The simulation results validate the effectiveness of the proposed designed techniques.

REFERENCES

[1] J. P. Ferreira, M. M. Crisostomo, and A. P.Coimbra, "Adaptive PD controller modeled via support vector regression for a biped robot," *IEEE Trans. on Control Systems Technology, vol.21, no.3, pp. 941-949, 2013.*

[2] M Galicki, "An adaptive regulator of robotic manipulators in the task space," *IEEE Trans. Automat. Contr., vol.53, no.4, pp. 1058-1062, 2008.*

[3] S. Hsu, and L. Fu, " A fully adaptive decentralized control of robot manipulators," *Automatica, vol.42, pp. 1761-1767, 2006.*

[4] J. J. E. Slotine, W. Li, "On the adaptive control of robotic manipulators," *Int. J Robotic Research, vol. 6 (3), pp. 49-59, 1987.*

[5] M. W. Spong, "On the robust control of robot manipulators," *IEEE Trans. Automat. Contr., vol. 37, no.11, pp. 1782-1786, 1992.*

[6] K. M. Koo, J. H. Kim, "Robust control of robot manipulator with uncertainty," *IEEE Trans. Automat. Contr., vol. 39 (6), pp. 1230-1233, 1994.*

[7] R. Burkan, and I. Uzmay, "Upper bounding estimation for robustness to the parameter uncertainty in trajectory control of robot arm," *Robotics and Autonomous Systems, vol. 45, 2, pp. 99-110, 2003.*

Design of Non-fragile Controller for Motion Control of Robot Manipulator

K. J. Binoo,[1] Sabyasachi Sengupta,[1],* and G Ray[1]
[1] Department of Electrical Engineering, Indian Institute of Technology
Kharagpur-721302, West Bengal, India.
*gray@ee.iitkgp.ernet.in

***Abstract*—This paper presents a design of PI type non-fragile controller for trajectory tracking and disturbance rejection of robot manipulators. The state equation of a robot manipulator is transformed to Canonical form and subsequently an error dynamic equation is obtained. Stability analysis and convergence of trajectory tracking error are investigated. Simulation results illustrating the performance of the proposed control algorithm for trajectary tracking of a robot manipulator with high precision tracking and improved convergence are presented.**

***Index Terms*—Canonical form, end effector, uncertainties,error dynamics,Riccatti equation, Optimal control.**

I. Introduction

In past two decades, the various optimal schemes have been developed to solve robotic manipulator problem [1]-[3]. Trajectory tracking problem is the most significant and fundamental task in control of robotic manipulator. The major source of uncertainty in robot manipulator is parametric uncertainty, which arises due to variations in payload and mass of links. With the increase in the order of such systems, the modeling of system has become a complicated issue and it is impossible to find an exact mathematical model of such systems. Imprecise Model of system leads to the devastating failure of controllers. A robust controller poses a challange for the control systems designers and it works even with actual system deviation from designed model. It is, therefore, quite important to design robust control strategies that ensure to track desire position in presence of uncertainties and payload variations and design techinques of such controllers can be found in [4]-[5].

This paper deals with the design of a non-fragile controller for trajectory tracking of an uncertain robot manipulator modifying control approach discussed in [6]. To control an such uncertain robot manipulator, one needs to deal with uncertainties involved in payload and friction. Our aim is to design a non-fragile controller by introducing the error equation in state space form, where the uncertainty is a function of state and its derivatives. It has been shown that the proposed trajectory tracking control algorithm improves the tracking performance compared with that of [6] in the presence of uncertainties in the system dynamics i.e. it demonstrates the movement of end effector tracks the desired trajectory in much less time in presence of disturbances in the control input channel.

II. Model Description and Problem Formulation

The dynamics of a robot manipulator is described by [6]

$$M(q)\ddot{q} + V(q,\dot{q}) + F(\dot{q}) + G(q) = \tau \tag{1}$$

where q consists of the joint positions, τ is the generalized forces, $M(q)$ the inertia matrix, $V(q,\dot{q})$ the coriolis/centripetal vector, $G(q)$ the gravity vector, and $F(\dot{q})$ the friction vector. Let

$$N(q,\dot{q}) = V(q,\dot{q}) + F(\dot{q}) + G(q)$$

There are uncertainties in $M(q)$ and $N(q,\dot{q})$ due to unknown load on the manipulator and unmodeled frictions and these are assumed to be bounded by the following equation.

$0 < M_{min}(q) \leq M(q) \leq M_0(q),\ \|N(q,\dot{q}) - N_0(q,\dot{q})\| \leq \eta_{max}(q,\dot{q})$ with $\eta_{max}(0,0) = 0$
Where, $M_{min}(q), M_0(q)$ and $\eta_{max}(q,\dot{q})$ are positive definite matrices and scalar quantity.

The objective of the proposed controller is to regulate the motion of robot manipulator with a view to track a desired trajectory $q_d(t)$, and also rejects the effect of disturbance on the system performance. The position and velocity error vector of n-link robot manipulator is defined as

$$\begin{aligned} e_{pi} &= q_i - q_{i,d} = x_i - x_{i,d} \\ e_{vi} &= \dot{q}_i - \dot{q}_{i,d} = \dot{x}_i - \dot{x}_{i,d}, \text{for } i = 1,2,3.....n. \end{aligned} \tag{2}$$

The dynamics of the error equation can be written as

$$\dot{e} = \begin{bmatrix} \dot{e}_{p(t)} \\ \dot{e}_{v(t)} \end{bmatrix} = \begin{bmatrix} 0 & I \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{e}_{p(t)} \\ \dot{e}_{v(t)} \end{bmatrix} + \begin{bmatrix} 0 \\ \ddot{q} \end{bmatrix} - \begin{bmatrix} 0 \\ \ddot{q}_d \end{bmatrix} \tag{3}$$

where $\dot{e}_p = [\dot{e}_{p1}, \dot{e}_{p2}.....\dot{e}_{pn}]^T$; $\dot{e}_v = [\dot{e}_{v1}, \dot{e}_{v2}.....\dot{e}_{vn}]^T$.
Using the expression $\ddot{q}$ from (1), one can rewrite the above equation as

$$\dot{e} = Ae + \begin{bmatrix} 0 \\ I \end{bmatrix} M(q)^{-1}[\tau - N(q,\dot{q})] - \begin{bmatrix} 0 \\ \ddot{q}_d \end{bmatrix} \tag{4}$$

where $B = [0^T \quad I]^T$. Now, the concept of inverse dynamics of an n-link robot manipulator is utilised to express the control law in the form

$$u(t) = M_0(q)^{-1}[\tau - N_0(q,\dot{q})] \tag{5}$$

where $u(t)$ represents a new control law. Using (5) in (4) and after some mathematical manipulations and simplification, one can obtain the following error vector expression.

$$\dot{e} = Ae + B(I + h(x))u(t) + Bf(x,\dot{x}) - B\ddot{q}_d \tag{6}$$

Where $h(x) = M(x)^{-1}M_0(x) - I \geq 0$ and $f(x,\dot{x}(t)) = M(x)^{-1}[N_0(x,\dot{x}) - N(x,\dot{x})]$
It may be noted that the error equation (6) is driven by the control input and the uncertainties $(h(x)\ and\ f(x,\dot{x}))$. Further, it may be noted that due to the special structures of A and B, a solution of non-fragile control problem can be formulated in simpler framework based on [6].

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

III. Proposed PI type non-fragile control strategy for a class of nonlinear systems

Let us introduce a new state vector defined by

$$z(t) = \int_0^t e_p(t)dt;\ \dot{z} = e_p(t) = [I\ \ 0]e(t) \tag{7}$$

where $e_p(t) = q(t) - q_d(t)$ and $q(t)$ is the actual position of n-link robot manipulator and $q_d(t)$ is the desired link position.

Augmenting (7) with (6), we get the final expression as

$$\begin{bmatrix} \dot{e} \\ \dot{z} \end{bmatrix} = \begin{bmatrix} A & 0 \\ C & 0 \end{bmatrix} \begin{bmatrix} e \\ z \end{bmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix} ((I + h(x))u + f(x, \dot{x})) - \ddot{q}_d) \tag{8}$$

One can rewrite (8) as

$$\dot{e_{au}} = \overline{A}e_{au} + \overline{B}(I + h(x))u(t) + \overline{B}f(x, \dot{x}) - \overline{B}\ddot{q}_d$$

where $e_{au} = [e^T\ z^T]^T$; $\bar{B} = [B^T\ \ 0^T]^T$ and $C = [I\ \ 0]_{n\times 2n}$. It is assumed that $f(x, \dot{x})$ is norm bounded by a known function, f_{max}. Our main goal is to design a linear robust controller $u(t) = -[K_1\ K_2]e_{au}$ based on minimizing quadratic performance index such that the system (8) is globally asymptotically stable for all admissive perturbation $f(x, \dot{x}(t))$ and uncertainties involved in input matrix $B(I+h(x))$. Control problem of such uncertain system described above can be solved with a non-fragile controller under some assumptions given below.

A. Design of Non-fragile Controller

Consider a class of nonlinear system (8) and each term associated with the expression is explained in the previous section.

It may be noted that the for all uncertainties involved in (8) can be bounded by a nonnegative function f_{max} and h_{max} such that $\|f(x, \dot{x})\| \leq f_{max}$ and $\|h(x)\| \leq h_{max}$. The formulation of proposed control strategy is quite different from the work of [6]-[7] and one can formulate the solution of non-fragile control problem into a standard linear quadratic control problem framework that will ensure asymptotic stablity of the closedloop system (8) for all uncertainties.

Recall the augmented system (8) and the corresponding linear feedback control law

$$u(t) = -Ke_{au}(t) = -(K_1e(t) + K_2z(t))$$

such that the performance index defined by the cost functional

$$J_{au} = \int_0^\infty (f^2{}_{max} + h^2{}_{max} + e_{au}(t)^TQe_{au}(t) + u(t)^Tu(t))dt \tag{9}$$

is minimised. It may be noted that the cost function consists of four terms. The first two terms stand for the cost of uncertainty compensation; third part for the cost of regulation; and $u(t)^Tu(t)$ for the cost of control effort. The stability analysis of the robust controlled system is derived by considering the theorem stated below.

Theorem 1: For the augmented system (8), there exists a control law $u(t)$ that minimizes a cost function (9) and simultaneously guarantees the stability analysis of the uncertain augmented system (8).

Proof: We need to design a control law $u(t) = u_0(t)$ such that the cost function for the augmented system (8) is minimised by bringing any initial error state $e(0) \neq 0$ to $e(\infty) = 0$. Based on the Hamilton-Jacobi- Bellman equation a condition is given below

$min_u[f^2{}_{max} + h^2_{max} + e^T_{au}Qe_{au} + u^Tu + (\frac{\partial V}{\partial e_{au}})^T(\overline{A}e_{au} + \overline{B}u - \overline{B}\ddot{q}_d)] = 0$ to derive the optimal control law. If $u = u_0$ is the solution to the control problem then,

$$f^2{}_{max} + h^2_{max} + e^T_{au}Qe_{au} + u^Tu + (\frac{\partial V}{\partial e_{au}})^T(\overline{A}e_{au} + \overline{B}u - \overline{B}\ddot{q}_d)] = 0 \tag{10}$$

Differentiating the above equation with respect to u_0, we get

$$2 * u_0^T + (\frac{\partial V}{\partial e_{au}})^T\overline{B} = 0 \tag{11}$$

We now show that the equilibrium point $e_{au} = 0$ of the augmented system (8) is globally asymptotically stable for all uncertainties. This can be proved by introducing a Lyapunov function defined as

$V(e_{au}(t)) > 0, e_{au}(t) \neq 0;\quad V(e_{au}(t)) = 0, e_{au}(t) = 0$ and subsequently we need to show that $\dot{V}(e_{au}(t)) \leq 0, e_{au}(t) \neq 0$ Now,

$$\begin{aligned} \dot{V}(e_{au}(t)) &= \left(\frac{\partial V(e_{au}(t))}{\partial e_{au}(t)}\right)^T \left(\frac{\partial e_{au}(t)}{\partial t}\right) \\ &= \left(\frac{\partial V(e_{au})}{\partial e_{au}}\right)^T [\overline{A}e_{au} + \overline{B}(u_0 + h(x)u_0) + \overline{B}f(x, \dot{x}) - \overline{B}\ddot{q}_d] \\ &= \left(\frac{\partial V(e_{au})}{\partial e_{au}}\right)^T [(\overline{A}e_{au} + \overline{B}u_0 - \overline{B}\ddot{q}_d) + \overline{B}h(x)u_0 + \overline{B}f(x, \dot{x})] \end{aligned} \tag{12}$$

Using (10) and (11) in (12) , we get

$$\begin{aligned} &= (-f^2_{max} - h^2_{max} - e^T_{au}Qe_{au} - u_0^Tu_0) \\ &-2u_0^Th(x)u_0 - 2u_0^Tf(x, \dot{x}) \\ &= -f^2{}_{max} - h^2_{max} + f^T(x)f(x) - (f(x) + u_0)^T(f(x) + u_0) \\ &-e^T_{au}Qe_{au} - 2u_0^Th(x)u_0 \\ &= -(f^2{}_{max} - f(x)^Tf(x)), \text{since } h(x) \geq 0 \end{aligned} \tag{13}$$

Thus, the conditions of local Lyapunov stability theorem are satisfied.

■

B. Design Steps for Implementing the Control Law

Systematic algorithmic steps to obtain the control law for the system (8) are given below:

Algorithmic steps:

Step-1: Transform the dynamic system (1) into the augmented error dynamic equation (8).

Step-2: The uncertainties $f(x, \dot{x})$ and $h(x)$ involved in (1) is bounded by the expression $f^T(x, \dot{x})f(x, \dot{x}) \leq f_{max} \leq x^T_{au}P_1x_{au} \leq e^T_{au}Pe_{au}$, and $h^T(x)h(x) \leq h_{max} \leq x^T_{au}P_2x_{au} \leq e^T_{au}\bar{P}e_{au}$, where $P_1,\ P_2 > 0,\ x_{au} = [x^T\ \dot{x}^T]^T; e_{au} = [e^T\ z^T]^T$

Note: $x^T_{au}P_1x_{au} \leq e^T_{au}Pe_{au}$, $x^TP_2x \leq e^T_{au}\bar{P}e_{au}$ and equality sign in both cases are valid when state tracks the desired trajectory perfectly.

Step-3: For the augmented system (8), find a feedback control law $u = u_0$ that minimises the cost function $\int_0^\infty (e^T_{au}Pe_{au} + e^T_{au}\bar{P}e_{au} + e^T_{au}Qe_{au} + u^Tu)dt$. Obtain W by solving the algebraic Riccati equation

$$\overline{A}^TW + W\overline{A}^T + (P + \bar{P} + Q) - W\overline{B}\overline{B}^TW = 0 \tag{14}$$

Note that the term $e_{au}^T(P+\bar{P})e_{au}$ is introduced in the performance index to compensate the effect of uncertainties involved in the system dynamics.

Step-4: The control law is computed using the relation $u_0 = -\overline{B}^T W e_{au} = -(K_1 e + K_2 Z)$. The designed control law can be applied in (1) to study trajectory performance and also to investigate the effectiveness of the proposed control algorithm.

IV. Simulations

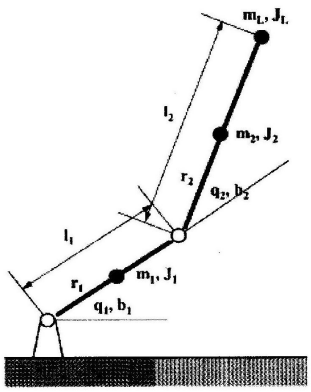


Fig. 1. A Two-link Robot Manipulator

Fig.1 shows a two-link rigid robotic manipulator [8] and its model parameters are given below: The robot manipulator has two joint variables, i.e two angles q_1 and q_2. The inertia matrix is

$$M(q) = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \tag{15}$$

Where

$$\begin{aligned} M_{11} &= J_1 + m_1 r_1^2 + J_2 + m_2(l_1^2 + r_2^2 + 2l_1 r_2 cosq_2) \\ &\quad + J_L + m_L(l_1^2 + l_2^2 + 2l_1 l_2 cosq_2) \\ M_{12} &= M_{21} = J_2 + m_2(r_2^2 + l_1 r_2 cosq_2) + J_L + m_L l_2^2 \\ M_{22} &= J_2 + m_2 r_2^2 + J_L + m_L l_2^2 \end{aligned} \tag{16}$$

The coriolis/centripetal vector is

$$V(q,\dot{q}) = \begin{bmatrix} V_1 & V_2 \end{bmatrix}$$

Where

$$\begin{aligned} V_1 &= (m_2 l_1 r_2 + m_l l_1 l_2)(\dot{q}_2^2 + 2\dot{q}_1\dot{q}_2)\, sinq_2 \\ V_2 &= (m_2 l_1 r_2 + m_l l_1 l_2)\dot{q}_2^2\, sinq_2 \end{aligned}$$

and assumed the gravity vector $G(q) = 0$

The friction vector is

$$F(q,\dot{q}) = \begin{bmatrix} b_1\dot{q}_1 \\ b_2\dot{q}_2 \end{bmatrix};$$

$$\text{Disturbance vector } d = \begin{bmatrix} 0.78 \\ 1.56 \end{bmatrix} \tag{17}$$

The constant matrices $\bar{A}$, $\bar{B}$ of the augmented system (see, e.g. in (8)) are already discussed in sections II and III. The following parameter values are considered for obtaining the simulation results.

$$m_1 = 0.41kg.;\ m_2 = 0.0985kg.;$$

$$J_1 = 72.6kgm/rad/sec.^2;\ J_2 = 19.4kgm/rad/s^2;$$

$$r_1 = 0.15m;\ r_2 = 0.08m; l_1 = 0.203m;$$

$$l_2 = 0.15m;\ b_1 = 0.236kgm/rad/sec.$$

$$b_2 = 0.59kgm/rad/sec.;\ m_L = 0.296\epsilon kg.;$$

$$J_L = 609.88\epsilon^2 kgm/rad/sec^2.$$

where $\epsilon \in [0,1]$' For trajectory tracking of two link robot manipulator the following initial data have been considered.

In this case, considering the augmented error dynamic equation (8), the associated quadratic bound for the the uncertainty terms $f(x,\dot{x})$ and $h(x)$ is computed combinely as

$$P^* = P + \bar{P} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2808 & 190 & 0 & 0 \\ 0 & 0 & 190 & 100 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The augmented state weightimg matrix is selected as $Q = 5I_{6\times 6}$. Solving the augmented algebraic Riccati equation (14) of the error dynamic equation (8), we obtained the following PI type control strategy

$$\begin{aligned} u_1 &= -34.7e_{p1} - 1.0e_{p2} - 54.95e_{v1} - 2.75e_{v2} - 10z_1 \\ u_2 &= -e_{p1} - 20.16e_{p2} - 2.741e_{v1} - 15.26e_{v2} - 10z_2 \end{aligned}$$

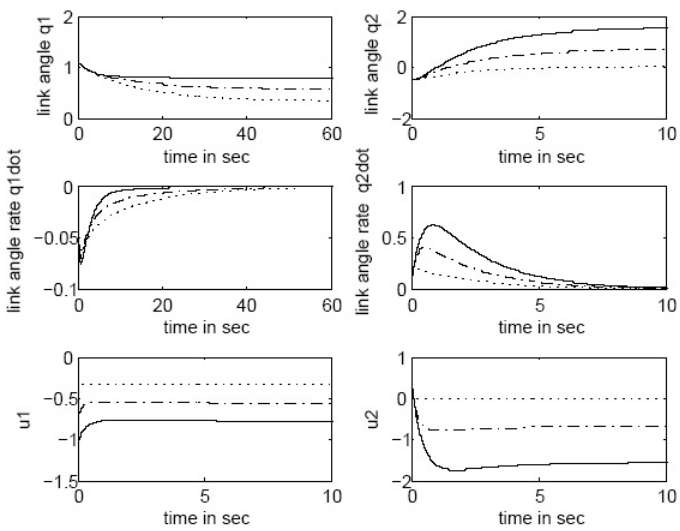


Fig. 2. Response based on [6] with disturbance $d = [0.78\ 1.56]^T$
$\cdots\cdots$ $m_L = 0$; $-\cdot-\cdot-$ $m_L = 0.15kg$; —— $m_L = 0.3kg$

Figs. 2 and 3 show the responses of two-link robot manipulator with changes in payload masses and constant external disturbances. It is obseved that the proposed method drives the link positions at the origin (i.e. $q_1 = q_2 = 0$). Figs. 4 and 5 depict the performance of the proposed controller for tracking a given trajectory $q_{d1}(t) = 1 + 0.1 * sin(2dt)$ and $q_{d2}(t) = 1 + 0.1 * sin(2dt)$ when the initial conditions $q_1 = 90^0$ and $q_2 = -60^0$ are considerd. Simulation results show that the proposed tracking controller appears to be

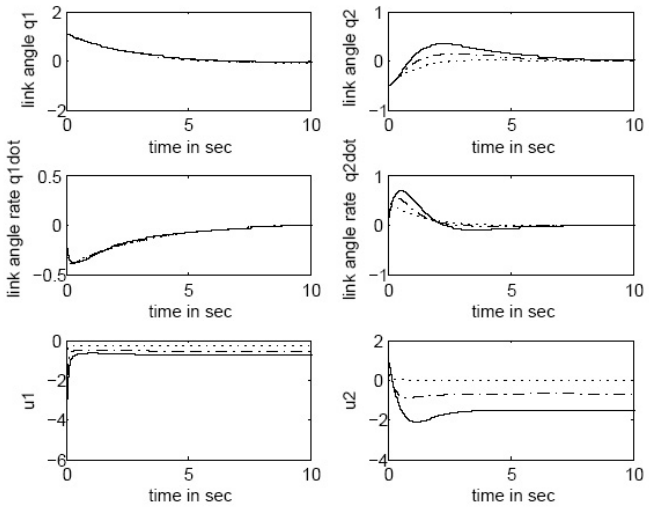


Fig. 3. Response of proposed method with disturbance $d = [0.78\ 1.56]^T$
$\cdots\cdots$ $m_L = 0$; $- \cdot - \cdot -$ $m_L = 0.15kg$; —— $m_L = 0.3kg$

useful to diminish the effects of payload uncertainties, external disturbances. The proposed controller outperforms than that of [6].

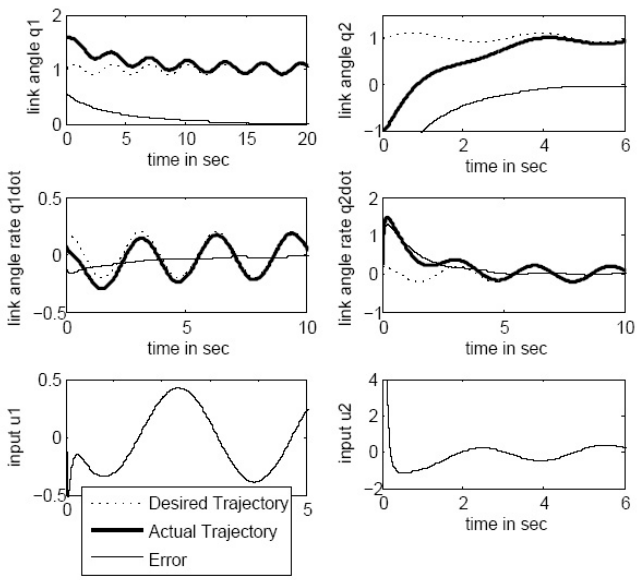


Fig. 4. Tracking response for m=0.3kg. based on the method of [6]

V. Concluding Remarks

A non-fragile controller is designed for an n-link robot manipulators based on an optimal controller approach. This control strategy can be used effectively (i) to move robot manipulator from any initial position to desired position $(q_d, \dot{q}_d) = (0, 0)$ with changes in payload masses and contant external disturbances, (ii) to track a desired trajectory. The proposed controller is designed in the framework of Lin et al. method [6] by introducing a new vector. In fact, this control law acts as a PI type non-fragile controller which is more effective in presence of constant disturbance and uncertainties. Figs 4 and 5 show that the output tracks the reference trajectory

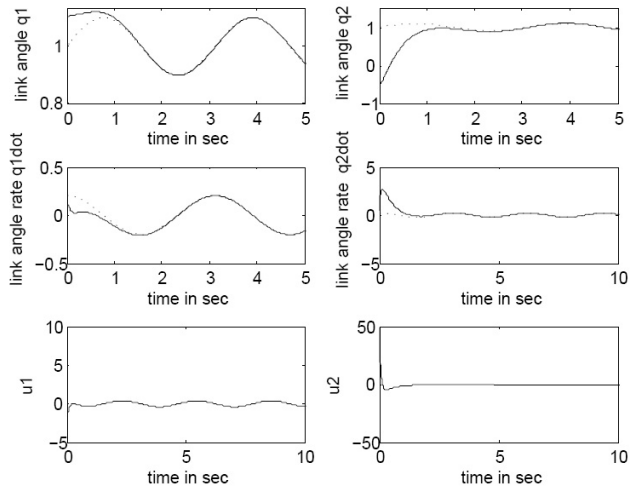


Fig. 5. Tracking response for m=0.3kg. with the proposed method
$\cdots\cdots$ Desired trajectory; ——— Actual trajectory

more perfectly and better results are obtained compared with that of Lin et al. [6] in presence of external disturbances and uncertainties in model.

References

[1] J. P. Ferreira, M. M. Crisostomo, and A. P.Coimbra, "Adaptive PD controller modeled via support vector regression for a biped robot,"*IEEE Trans. on Control Systems Technology*, vol.21, pp. 941-949, 2013.

[2] M. Galicki, "An adaptive regulator of robotic manipulators in the task space," *IEEE Trans. Automat. Contr.,* vol.53, pp. 1058-1062, 2008.

[3] S. Hsu, and L. Fu, " A fully adaptive decentralized control of robot manipulators," *Automatica,* vol.42, pp. 1761-1767, 2006.

[4] D. Zhao, S. Li, Q. Zhu, and F. Gao, "Robust finite-time control approach for robotic manipulators," *IET Control Theory and Applications,* vol.4, pp. 1-15, 2010.

[5] H. Pan, and M. Xin, " Nonlinear robust and optimal control of robot manipulator," *Nonlinear Dynamics,* vol.76, pp. 237-54, 2014.

[6] F.lin and R. Brandt, "An optimal control approach to robust control of robot manipulators," *IEEE Trans. on Robotics and Automation*, Vol.14, 1998.

[7] F.Lin,R.Brandt and J.Sun. "Robust control of nonlinear systems:Compensating for uncertainty,"*Int.J.Control* ,Vol. 56 pp.1453-1459, 1992.

[8] T.S.Chung, "An inherent stability problem in cartesian compliance and an alternative structure of compliance control,"*IEEE Transactions on Robotics and Automation*,Vol.7, pp.21-30, 1991

Self-Localization of a Mobile Robot using Monocular Vision of a Chessboard Pattern

Parvez Ahmmed*, Zabir Ahmed, Mohammad Ishfaque Jahan Rafee, M A Awal, Sajid Muhaimin Choudhury
Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology, Dhaka–1205, Bangladesh
Email: *p.ahmmed.bd@ieee.org

***Abstract*—A computationally fast visual approach of autonomous robot localization with high degree of accuracy is demonstrated. Relative position and orientation of an autonomous robot with respect to a fixed reference point is determined by minimal processing of a monocular image employing projective geometry resulting in very low computational requirements. A three-layered sequential algorithm i.e. Detection–Pixel Level Analysis–Sub Pixel Accuracy, is used to speed up the process and to reduce computational bulk. The position co-ordinate of a point in a plane is obtained by analyzing the perspective projection of a reference structure on a monocular image perceived by a camera placed at that point. A chessboard pattern is suitably chosen as the reference structure which is placed at the origin of the plane.**

***Index Terms*—chessboard, localization, projective geometry, monocular vision, mobile robot, perspective view.**

I. Introduction

Localization is of fundamental importance in autonomy of a mobile robot [1]. In any environment where human involvement is hazardous such as in space or in some disaster-prone area where an autonomous robot is launched from a reference position, it would require a self-localization technique in order to track its reference so that it may return back to the original position. This necessitates a faster localization method which will run continuously without hampering the other processes running on the computer.

Various researches have been conducted on localization techniques utilizing different sensory inputs like laser range finders and sonar sensors. Substantial works have been done using vision as primary sensor, e.g. stereopsis uses two or more image capturing devices to obtain a depth map of its surroundings [2]. Some of these approaches especially stereo-matching algorithms require fairly high computational power to obtain nearly real-time data with appreciable level of accuracy. Again, these spatial sensors and image pairs for the stereo vision provide relative distance between the robot and some obstacle or boundary. So these methods are particularly useful in obstacle avoidance and path planning. Moreover, maintaining the relative spatial position of the two cameras used in the binocular vision is very difficult.

However, in many cases robot's location relative to a certain reference point at any instant of time might be of greater importance for path and action planning. For example, if a mobile robot is to explore a bounded region in space in the vicinity of a reference point or zone, availability of instantaneous spatial co-ordinates and information of its pose, orientation can greatly reduce its decision making complexities.

This paper presents such a method of robot localization which utilizes the geometry of a reference structure in a captured monocular image to obtain robot's position co-ordinates. A 4-by-4 chessboard with nine internal corner points is used as the reference structure providing an operational field of nearly 180°. Despite the broader 360° operational field of other localization techniques based on color-space image processing of 3D reference structures [3], the proposed approach is superior due to its relative insensitivity to varying light conditions and lesser processing requirements. Use of only one camera is another advantage of the proposed method. The mathematical treatment involved in this paper employs some basic rules of perspective projection which is a subset of projective geometry [4].

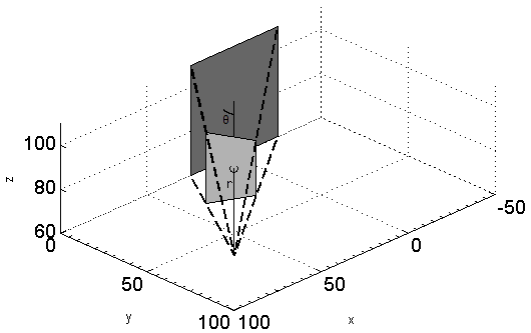


Fig. 1. Perspective view of a square-shaped reference placed at origin

II. Localization

The instantaneous co-ordinate of the robot is proposed to be estimated from the perspective view of a square placed at the origin of a horizontal plane. In order to analyze the perspective shape of the square perceived by a camera which is placed in that plane passing through the center of the reference square, identification of the four corners will suffice. From these four corners, the value of (r, θ) of the polar co-ordinate of the camera in Fig. 1 is to be calculated.

A. Reference Selection

While choosing a unique but easily detectable reference pattern, a number of procedures were adopted. The prime objective is to identify the four corners of the reference square structure placed at a dynamic environment in a reasonably short time.

At first, four colored-LEDs were chosen to be placed at the corners of the square, but the glowing LEDs appeared completely white in captured image which became indistinguishable from the background.

The next choice moves to four red circles having centers at the corners of a square which were detected using Hough

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

Circle Transform [5]. Although HSV color space is quite suitable for detecting colored objects in an image than RGB, identification of color became very much sensitive to light [6].

The search for a reference having minimal sensitivity to ambient lighting conditions led to chessboard pattern. Due to inherent high contrast ratio of the checkerboard pattern, four corners pixel co-ordinates can be accurately detected. Its geometry also provides the scope to estimate errors in detection through self-referencing. Therefore, a 4-by-4 chessboard with nine internal corner points has been used.

B. Finding corners of a Chessboard

Chessboard corners can be detected using Moravec Corner Detection algorithm [7]. It tests each pixel in the image to see if a corner is present by considering how similar a patch centered on the pixel is, compared to the nearby largely overlapping patches.

The similarity is measured by taking the sum of squared differences (SSD) between the two patches. A lower number indicates more similarity. If the pixel is in a region of uniform intensity, then the nearby patches will look similar. If the pixel is on an edge, then nearby patches in a direction perpendicular to the edge will look quite different, but nearby patches in a direction parallel to the edge will result only in a small change. If the pixel is on a feature with variation in all directions, none of the nearby patches will look similar. The corner strength is defined as the smallest SSD between the patch and its neighbors (horizontal, vertical and on the two diagonals). If this number is locally maximal, a feature of interest is present.

Being anisotropic is one of the main problems with this operator as pointed out by Moravec i.e. if an edge is not in the direction of the neighbors, it will not be detected as an interest point. This problem is solved by using the Harris & Stephens [8] or Shi-Tomasi [9] Corner Detection algorithm.

Sub-pixel accuracy is also achieved by using The Förstner Corner Detector. To achieve an approximate solution, the Förstner algorithm solves for the point closest to all the tangent lines of the corner in a given image. This algorithm which provides a least-square solution, works by considering the fact that the gradients of the edges intersect at the point of interest.

C. Determination of polar co-ordinates

The derivation provided uses the pinhole camera approximation and assumes the center of the reference chessboard to be located at the principle point of the captured monocular image in order to avoid Barrel Distortion.

The center point E of a 4×4 chessboard with 3×3 internal corners at A through I is taken to be the origin of the XY plane.

If the nine corners of the chessboard is identified to be A' through I' in the image and O is supposed to be the location of the camera which is located at the left side (positive X direction) from the chessboard (Fig. 2) in the horizontal plane passing though the centre of the chessboard, then $A'G' \,||\, C'I'$ and $A'G' > C'I'$.

If the view angle of the camera with pixel resolution $W \times H$ is represented by ϕ, the vision angle subtended by the sides AG and CI of the square AGIC,

$$\theta_L = A'G' \times \phi/H \qquad \text{and} \qquad \theta_R = C'I' \times \phi/H \tag{1}$$

Now, the arc AG of the circle with a radius approximately equal to OD subtends an angle θ_L at the centre O whereas the arc CI of the circle with a radius approximately equal to OF subtends an angle θ_R at the centre.

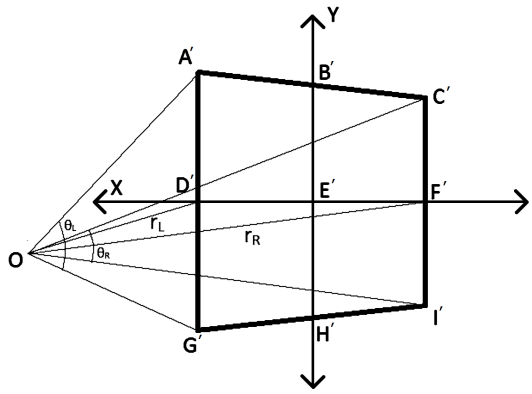


Fig. 2. Perspective view of the nine chessboard corners

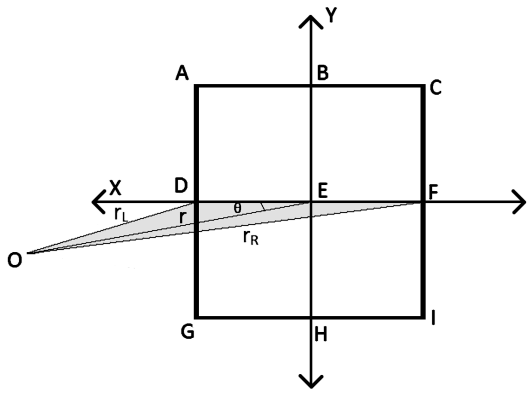


Fig. 3. Determination of polar co-ordinate using Apollonius' theorem

$$OD = AG/\theta_L \qquad \text{and} \qquad OF = CI/\theta_R \tag{2}$$

Let, the radii of the circles be OD = r_L & OF = r_R and the length of a side of the square AGIC in Fig. 3 be v, which can be measured prior to the execution i.e. $AG = CI = v$.

$$\therefore\ r_L = v/\theta_L \quad and \quad r_R = v/\theta_R \tag{3}$$

It is evident from Fig. 3, ODF forms a horizontal triangle of which OE is a median. Therefore, from Apollonius' theorem, (4) can be obtained.

$$OD^2 + OF^2 = 2(OE^2 + DE^2) \tag{4}$$

As, OE is the distance of the camera or robot from the origin (centre point of the reference) i.e. $OE = r$ and $DE = v/2$, (4) can be written as,

$$r = \sqrt{\frac{r_L^2 + r_R^2}{2} + \frac{v^2}{4}} \tag{5}$$

From the properties of a triangle, for $\triangle ODE$ in Fig. 3,

$$\cos\theta = \frac{DE^2 + OE^2 - OD^2}{2 \cdot DE \cdot OE}$$

$$\therefore\ \theta = \cos^{-1}\frac{r_L^2 + v^2/4 - r^2}{rv} \tag{6}$$

This polar co-ordinate (r, θ) can easily be converted to cartesian co-ordinate (x, y) where needed using (7).

$$x = r\cos\theta \quad and \quad y = r\sin\theta \tag{7}$$

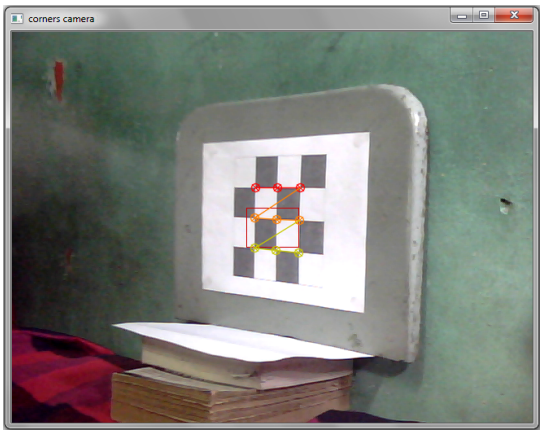


Fig. 4. Experimentation with detection of chessboard corners

TABLE I
COMPARISON BETWEEN CALCULATED AND ACTUAL MEASUREMENT

Actual reading		Calculated reading		Error percentage	
r (cm)	θ^o	r (cm)	θ^o	$\lvert\delta r\rvert$ (%)	$\lvert\delta\theta\rvert$ (%)
44.01	88.7	44.57	88.9	1.27	0.22
73.01	89.22	74.31	88.72	1.78	0.56
106.02	88.92	108.27	88.22	2.12	0.78
118.0	89.51	121.04	89.81	2.58	0.33
43.08	68.2	42.74	68.98	0.79	1.14
73.54	67.62	72.76	66.44	1.06	1.75
104.18	68.61	102.72	67.6	1.4	1.47
119.62	66.86	122.08	65.0	2.05	2.8
45.25	45.0	45.76	43.71	1.12	2.88
74.97	43.92	76.85	42.54	2.51	3.13
105.36	44.62	111.13	42.92	5.47	3.81
118.79	45.0	125.55	43.65	5.69	3.0
46.49	18.82	45.9	19.24	1.27	2.22
73.39	19.92	74.7	19.43	1.78	2.46
105.95	19.29	103.84	18.42	1.99	4.5
120.93	23.39	123.74	22.77	2.32	2.62

D. Limitations

The proposed theory assumes the projection plane of the camera and the plane of the reference pattern, both to be perpendicular to the XY plane. Inclination of either of the planes can cause the calculated co-ordinates to deviate from the actual values.

Although this method needs only four points and hence a 3-by-3 chessboard pattern should work fine, 9 internal corners are detected in this topology. However, as the shape of the chessboard is predefined, these 9 points can be used to generate a better estimation of the values of four required corners.

When the robot is in motion, the blurred image lacks the chessboard pattern. In order to avoid the motion blur, either the image is to be processed with Wiener filter [10] or a camera with much higher FPS (frames per second) should be used.

III. IMPLEMENTATION

To experimentally determine the accuracy of this localization scheme, a chessboard is printed and mounted on a vertical plane at a suitable height. An open source library, openCV 2.4.3 [11] is used in programming of the data collection program.

A. Program flow

The captured image is first checked for motion blur caused by the movement of the mobile robot using morphological filters [12] which is very fast irrespective of the presence of the chessboard in the field of view. The image is then put through chessboard corners finding algorithm [13] using adaptive threshold in order to enhance the compatibility in different lighting conditions. Although this function is a bit slower with no chessboard in the image, as the presence of the chessboard is confirmed by the previous function, the overall running time becomes quite low.

Once a chessboard pattern is detected anywhere in the image as shown in Fig. 4, the robot aligns the camera so that the center of chessboard reference is placed within a predefined region around the principle point of the image. The alignment is performed by successive rotation of a servo motor driven by the feedback from the position of the center of the chessboard in the image.

The image is then converted to gray scale and the corner positions are recalculated to sub-pixel accuracy [14] using Förstner corner detection algorithm.

The nine points are properly arranged and four corner points among them are used in calculation of the co-ordinate using (1), (3), (5) and (6) sequentially. Although the algorithm is robust enough, the corner detection exhibits some fluctuation especially in the inclined positions. Hence a moving average smoothing filter is used to stabilize the calculated values.

B. Data collection

Tests of the robot localization are performed using a chessboard with a dimension of 16.5 cm being the length of a side. Calibration procedures indicate that the field of view of the cameras is 33°. All of the experimental data are obtained using images of 640×480 pixels captured at a rate of 30 frames per second. The whole experiment is conducted in various lighting conditions and also in dynamic background.

The polar co-ordinates (r, θ) of the robot are calculated at different positions in such a way that for a certain value of θ there are 4 positions with different values of r. Thus 4 data are taken at each of the 4 different values of θ. These 16 data are analyzed to estimate the sensitivity of the system.

Table I shows the collected data which are tabulated at 4 different r with each value of θ on the horizontal plane. The calculated values are compared with the physically measured data and the percentages of errors are calculated using (8) and plotted to demonstrate the reliability of the scheme.

$$\delta r = \frac{r_{calc} - r_{actual}}{r_{actual}} \times 100\% \tag{8}$$

IV. RESULTS AND DISCUSSION

Fig. 5 shows that the error in calculation of r increases as the distance from reference increases. This restricts the upper bound of the distance. For this case, a chessboard of 16.5 cm could provide promising (more than 90% accurate) result up to a distance of 200 cm. However, a very close position (below 30 cm) cannot accommodate the full pattern in the image and hence calculation becomes impossible. These limits can be pushed further by increasing the size of the chessboard.

Another way of increasing the range of accurate measurement is to use dynamic zoom. When the size of the chessboard becomes very smaller in the image, the up sampling operation pushes the corner points further from each other and thus increases the accuracy of calculation to a great extent. Again

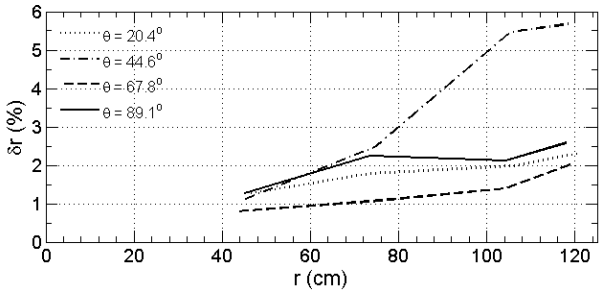


Fig. 5. Variation of δr at different θ

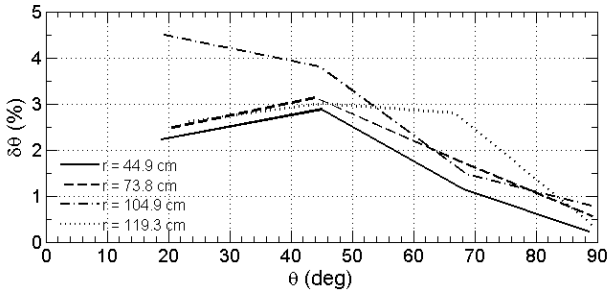


Fig. 6. Variation of $\delta\theta$ at different r

down sampling the image can help in the case when the chessboard is too large to be accommodated in the frame.

The error in calculation of $\delta\theta$ also shows a declining slope as actual θ increases in Fig. 6. As θ decreases, the increasing inclination causes the corners to be appeared very close to each other introducing an error in corner detection. Thus the method fails to measure accurate θ near the values 0° or 180°.

V. Future Development

The proposed scheme has a great possibility of further development in various aspects. As the scheme has a fairly low computational complexity, it can be implemented easily in FPGA technology to reduce the running time and to enhance hardware portability. With the advancement of autonomous vehicle technology, this topology can sufficiently enhance the robotic vision of the vehicle.

A. Three dimensional co-ordinate

As in case of the two-dimensional co-ordinate, the relative lengths of the two vertical sides in the perspective view of the square yielded the polar co-ordinate (r, θ). A more rigorous calculation including the relative lengths of all the four sides in the perspective view of the square will result in the spherical co-ordinate (r, ϕ, θ). This would find its application in localization of an aerial robot where the chessboard would be placed on the ground level.

B. Parallel processing

Parallel processing opens a huge door for further improvement. Parallel algorithm brings computational power efficiency. For parallel algorithms GPU is a very good choice, as they have quite high number of parallel processing cores e.g. 16-1536 cores as compared to CPU [15]. They also provide the opportunity for other expensive computation for decision making for a sophisticated autonomous robot with a number of objectives. Thus the technology of multipurpose robot can be advanced one step ahead by employing this developed scheme.

VI. Conclusions

The demonstrated scheme establishes a simple way of autonomous robot localization requiring minimal hardware and processing power. It provides a more pragmatic solution for autonomous localization in varying light conditions. With a fairly primitive camera, a minimum of 94% accuracy is achieved at measuring range. For terrestrial autonomous robotics application as well as outer-space applications, the proposed scheme provides sufficiently fast computation with substantial accuracy.

References

[1] D. Solanki and V. Khare, "Optical navigation system for robotics application," in *Third International Conference on Intelligent Systems, Modelling and Simulation (ISMS)*, Feb 2012, pp. 268–272.

[2] K. Al Mutib, M. Al Sulaiman, H. Ramdane, M. Emaduddin, and E. Mattar, "Active stereo vision for mobile robot localization and mapping path planning," in *14th International Conference on Computer Modelling and Simulation (UKSim)*, March 2012, pp. 293–299.

[3] R. Volpe, T. Litwin, and L. Matthies, "Mobile robot localization by remote viewing of a colored cylinder," in *IEEE/RSJ International Conference on Intelligent Robots and Systems 95. 'Human Robot Interaction and Cooperative Robots'*, vol. 1, Aug 1995, pp. 257–263 vol.1.

[4] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge: Cambridge University Press, 2004.

[5] H. K. Yuen, J. Princen, J. Illingworth, and J. Kittler, "Comparative study of hough transform methods for circle finding," *Image Vision Computing*, vol. 8, no. 1, pp. 71–77, Feb. 1990.

[6] F. Yang, H. Wang, H. Wei, Y. Yang, and Y. Wang, "Circular road sign analysis and visual localization approach based on monocular vision," in *Third Global Congress on Intelligent Systems (GCIS)*, Nov 2012, pp. 334–337.

[7] H. P. Moravec, "Obstacle avoidance and navigation in the real world by a seeing robot rover," Ph.D. dissertation, Stanford University, Stanford, CA, USA, 1980, aAI8024717.

[8] C. Harris and M. Stephens, "A combined corner and edge detector," in *In Proceedings of Fourth Alvey Vision Conference*, 1988, pp. 147–151.

[9] J. Shi and C. Tomasi, "Good features to track," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun 1994, pp. 593–600.

[10] A. Erdem, M. Sezan, and M. Ozkan, "Motion-compensated multiframe wiener restoration of blurred and noisy image sequences," in *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-92.*, vol. 3, Mar 1992, pp. 293–296 vol.3.

[11] R. Laganiere, *OpenCV 2 Computer Vision Application Programming Cookbook*. Birmingham: Packt Publishing, 2011.

[12] F. Y. Shih, C.-F. Chuang, and V. Gaddipati, "A modified regulated morphological corner detector," *Pattern Recognition Letters*, vol. 26, no. 7, pp. 931–937, 2005.

[13] (2009) Camera calibration and 3d reconstruction. [Online]. Available: http://docs.opencv.org/modules/calib3d/doc/camera_calibration_and_3d_reconstruction.html

[14] (2009) Camera calibration and 3d reconstruction. [Online]. Available: http://docs.opencv.org/doc/tutorials/features2d/feature_detection/feature_detection.html

[15] (2009) Nvidia cuda compute capability comparative table. [Online]. Available: http://www.geeks3d.com/20100606/gpu-computing-nvidia-cuda-compute-capability-comparative-table/

Control Input Converter for Robot's Leg Joint with Parallel Actuation Configuration

Addie Irawan,[1] Hoe Lih Jiun[2], Md Moktadir Alam[3,*]

Robotics and Unmanned System (RUS) group, Faculty of Electrical & Electronics Engineering,
Universiti Malaysia Pahang, Pekan, Pahang, MALAYSIA
addieirawan@ump.edu.my[1], wallancehoe@gmail.com[2], moktadir.alam@hotmail.com[3,*]

Abstract— **This paper presents the proposed control input converter (CiC) for robot leg's joint control with parallel actuation. The parallel actuation commonly used in heavy machine such as backhoe and bulldozer. This configuration performing muscle-based movement since the actuator attached becoming part of the frame. Therefore this particular switching mechanism is designed to translating a desired signal to the linear actuator from any close-loop control input. This simple algorithm is developed from the behavior study of the attached linear actuator motion on each crossed joint angle. The verification is done on actual leg of Quadruped Parallel Leg Actuation (QPAL) robot system by using a real-time target embedded system with focusing on foot motion for each leg and joint angle position error tracking from some angle inputs.**

***Keywords*—parallel actuation, control input converter, multi-joint leg.**

I. Introduction

In mechanical structure point of view various configurations have been introduced such as under-actuated leg, indirect actuated leg, flexible body and wheeled tips approaches. On the other hand, this research and development also involved on motion control, vision/image processing to increase its autonomy, dynamic and stable gait pattern, equilibrium center of body (CoB) control including walking algorithm design by mimicking legged life form living creatures. In the actuation mechanism in multi-limbs robot technology three approaches can be done either direct, parallel/indirect or hybrid mechanism in robot arm and leg design. Direct joint actuation or direct mechanism required sufficient torque to conduct both arm frames with sufficient shaft since major of attached motors contribute to the weight of the robot frame. It is different to the parallel actuation mechanism which is majorly configured as one of the frame structured with strong geared acting like a human muscle. This actuation mechanism is also configured to move leg/arm joint indirectly with limited rotational angles if compare to the direct drive mechanism.

On previous existed multi-legged robot, parallel actuation mechanism majorly applied on hydraulically driven robot design which is majorly used hydraulic cylinder/piston as its actuators such as MACANT[1], ROBOCLIMBER[2], TITAN-XI[3], BigDog[4] and COMET-IV[5]. These robots are focused on mega payload as its main characteristic to deal with the applications. Due to the current technology in electrical motor and actuation, medium and small scale high payload multi-legged robot is possible to be designed by using electrical linear actuator that function same as cylinder hydraulic actuator or servo pneumatics cylinder. Spidernaut from National Aeronautics and Space Administration (NASA) is an example of the large scale multi-legged robot that used parallel mechanism on its leg actuation. This robot has sophisticated indirect actuated leg configuration with 3 degree of freedom (DoF) which allows foot placement in space, but does not have a means of independently controlling the angle of the foot[6]. Furthermore, ADDAM from Kyushu Sangyo University, Japan also applied both indirect and direct actuation mechanism to perform human muscle-based movement for an endoskeleton leg. On its ankle, couple parallel actuation system is designed with adoption of couple parallel link mechanism with one linear actuator and 2-axis joint (cross revolute joint). Total of DoF for this robot is 12 consisting 4-DoF at ankles, 2-DoF at knees and 6-DoF at a hip[7].

Therefore this project has taken initiative to involve in this field by proposed a multipurpose medium duty Quadrupped robot system, namely Quadruped Parallel Leg Actuation (QPAL), with square shape body and 2-DoF indirect actuation and 1-DoF direct actuation of leg configuration using direct current (DC) linear actuator and DC brushed rotary motor respectively. Two revolute joints have been configured with linear actuators to realized closed-link mechanism in different way respectively. The issue will be taken here is to create a driven module to interface easily with the any close-loop control input for each leg's joint movement. The easiest configuration for error tracking on parallel actuation joint movement is to attach the encoder/potentiometer on the joint revolute point instead of estimating linear actuator's stroke movement. Another issue has come out on how to shoot the control input on the actuator driver correctly to achieve almost zero error tracking on joint angle movement. Therefore this research has proposed the control input converter (CiC) module that can be used to extract the control input in real-time legged robot system such as QPAL robot leg's configuration. The target of CiC implementation is to deal with the common and raw full-bridge DC motor driver and not for the advance driver. The research is done by verifying and validating the propose module in actual QPAL robot system with arbitrary angle inputs on each parallel actuation joints.

II. QPAL Robot System Configuration

Quadruped Parallel Leg Actuation (QPAL) is primarily concerned about the efficient structural and mechanic characteristics to deal with the medium payload application (4kg~6kg). As shown in Fig.1, the structure frame of QPAL was designed with a square shape and support polygon. The

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

research and development of this frame work are focused on several key issues that include parallel mechanism configuration design, drive and control technique. Moreover this robot also emphasized on common legged robot locomotion issues; impassable terrain, foot slippage, accidental collision, modeling errors and sensor errors [8].

QPAL was started designed and developed in July 2012 and targeted for medium payload applications such as firefighting extension system, mine detection, medium tunnel surveyor etc. With reference to the fundamental of biological insect legs such as crockroach, beatle and housefly as shown in Fig.2, the motions are made by the linear movement such as contraction and relaxation of furmur and tibia, and articulated with coxa as shoulder attached with the body [9]. Inspired with this configuration, QPAL's leg is designed by attaching linear actuators between shoulder and thigh (bended frame), and between thigh and shank frame as shown in Fig.2 respectively. The linear actuators positioned almost parallel with the links in order to make a bending movement like a muscles by performing extending and retracting as shown in Fig.3. One of the link is bended in order to achieve 90^0 bended position in offset mode as shown in Fig.3(b). This configuration allows large torque on each revolute joint from direct force on linear actuator piston[1].

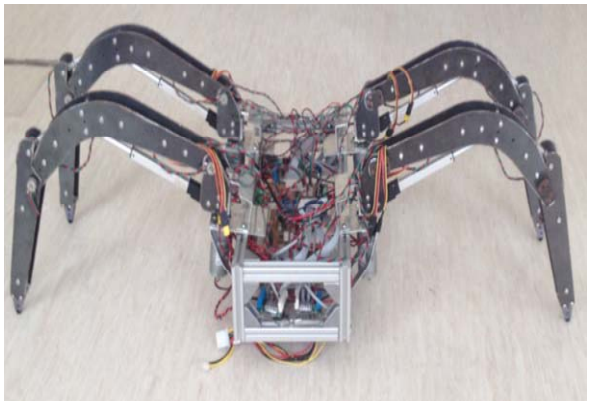


Fig. 1 QPAL robot system

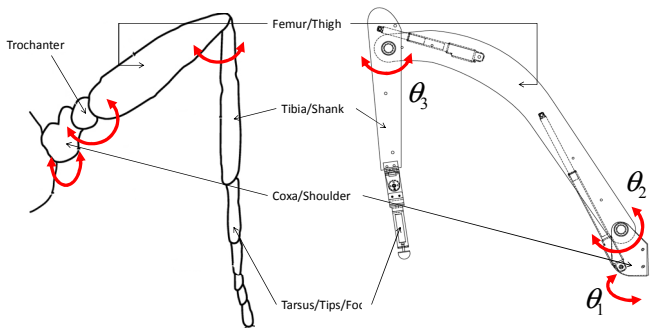


Fig. 2 QPAL's leg configuration that based on biological arthropod's leg

A customize single board named QBoard is designed and developed with extra inputs/outputs (I/Os) by applying digital multiplexer circuit to control each motor/actuator with minimizing the used of microcontroller I/Os as shown in Fig.4. This single board is specifically design for real-time control and monitoring for QPAL robot with up to 3-DoF/leg configuration. Extra detail specification about QPAL robot is tabulated in Table 1.

[1] Each linear actuator has maximum up to 23N of peak power point

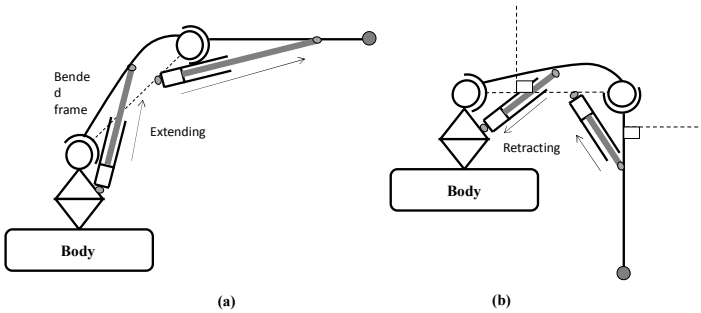


Fig. 3 A motion schematic of moving range solution for a leg, (a) Maximum extending actuation, (b) minimum retracting position (Offset position).

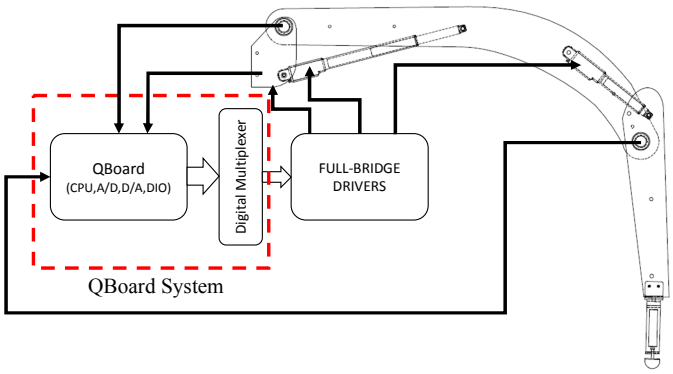


Fig. 4 QPAL electronics & mechanical Interfacing

TABLE I
SPECIFICATIONS OF THE QPAL ROBOT SYSTEM

Item	Value	Remarks
Length [m]	1.042	-
Width [m]	1.042	-
Height [m]	0.392	-
Weight [kg]	40	-
Maximum Motor Speed (Each Shoulder) [RPM]	375	4 Unit
Maximum A Linear Motor Speed [mm/s]	10~11	8 Unit
Maximum Total Torque/Leg [Nm]	80~90	4 Unit

III. CIC IN LEG'S CONTROL ARCHITECTURE

QPAL has been prepared with established and simulated control architecture as shown in Fig.5. Since parallel actuation mechanism is used on leg design, drive module becoming important part which is different from direct actuation mechanism. As shown in Fig.3 in Section II, drive module need to be designed according to the extending and retracting motion of a linear actuator to move two revolute joints θ_2 and θ_3. Therefore CiC is design with reference to the angle motion of thigh and shank joints flow mapped as shown in Table 2.

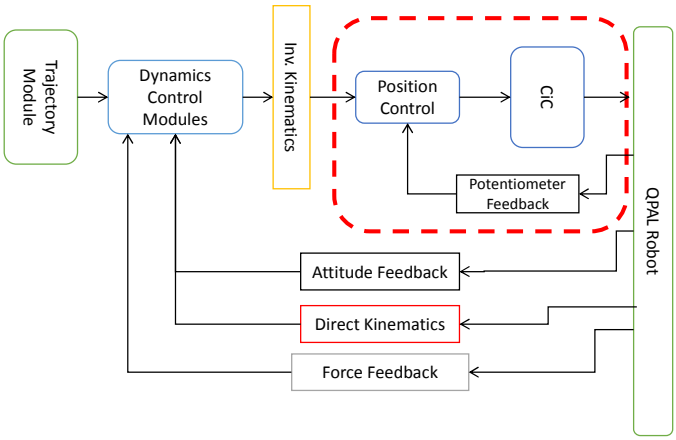


Fig. 5 Overall QPAL control architecture

TABLE II
TRUTH TABLE OF MOTIONS FOR LINEAR MOTOR LOGICAL DRIVE INPUT

θ_2	D_2	u	Actuator Function	θ_3	D_3	u	Actuator Function
0	0	0	Stop	0	0	0	Stop
↓	0	-	Retract	↑	0	+	Retract
↑	1	+	Extend	↓	1	-	Extend
0	X	0	Stop	0	X	0	Stop
↑ = angle increasing ↓ = angle decreasing X = no change							

D_2 and D_3 representing logical input (Bit 1=extending, Bit 0=retracting) for actuator poles control for θ_2 and θ_3 respectively. Moreover u is the control input from any designed control system. Thus to deal with the raw full-bridge driver, the CiC function is illustrated with the switching block as shown in Fig. 6.

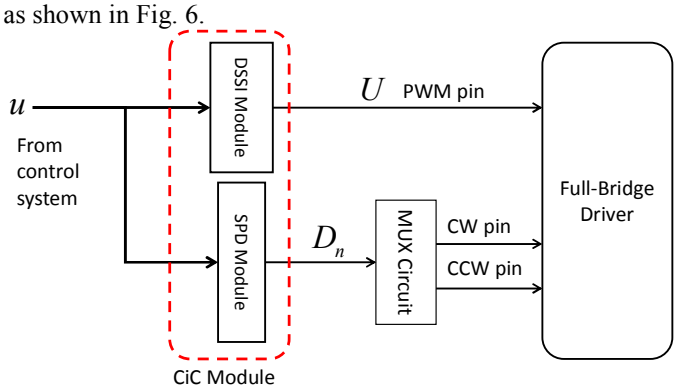


Fig. 6 CiC Module for single full-bridge driver control

The drive input pins of full-bridge driver or commonly known as pulse width modulation (PWM) pin, only accepting positive value signal, thus u need to be always in positive value. The Drive Signal Separator & Invertor (DSSI) module will play a main role here to detect the pole of u and invert this signal if u is in negative value and vice versa. On the other hand the u poles will be extracted as information for logical signal $\{1;0\}$ by Signal Pole Detector (SPD) module to the full-bridge driver via QBoard multiplexer circuit. The CiC also can be concluded in the form of equation as follows with $n = 2,3$;

$$U_n = \begin{cases} |u| & u > 0 \quad D_n = 1 \\ |u| & u < 0 \quad D_n = 0 \end{cases} \tag{1}$$

IV. EXPERIMENTAL AND RESULTS

In order to verify and validating the proposed CiC two experiments are done; analysis on the CiC output via simulation and experimental on a leg movements. Moreover this experiment is done by applying proportional, integral and derivative (PID) control on controlling the precision of joint angle movement for each leg. As shown in Fig. 7, the sample from simulation results (shank's joint angle movement) shows that CiC playing its role by inverting the negative u values from PID controller of poles and range of actuator motion is customized as different signal for retract and extend input respectively.

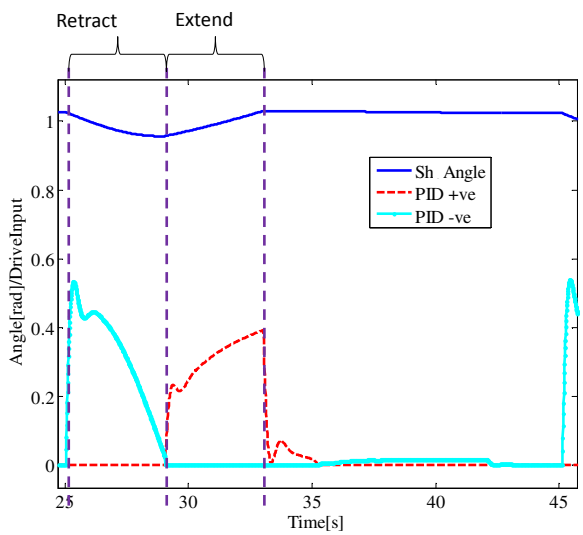


Fig. 7 Experiment results for shank angle input (sample for leg 3)

The verification works is further by moving the leg as shown in Fig.8 with some input. The QPAL robot system is prepared with real-time control structure that used hardware-in-loop concept between computer and QBoard unit in applying control architecture as shown in Fig.5. The function of CiC then is validated by showing stable movement angle as desired and error tracking can be control almost precise between 3-5% as shown in the sample results in Fig.9.

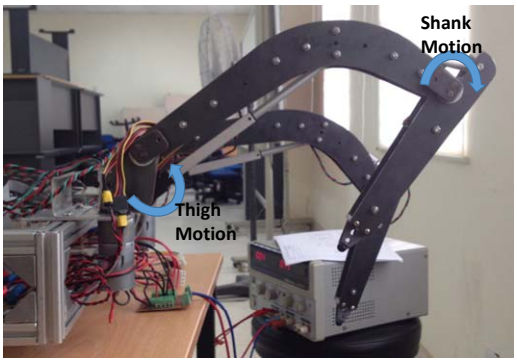


Fig. 8 Snap shot on leg movement of CiC validation (leg 2)

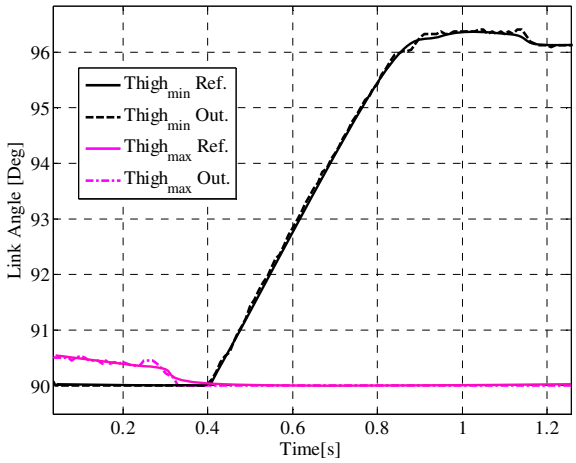


Fig. 9 Experiment results for thigh angle movement (sample for leg 2)

V. Conclusions

The simulation and experiment results show CiC able to stabilize the desired input that controlled via close-loop control such as PID controller for parallel actuation configuration of multi-DoF leg such QPAL robot. Furthermore the proposed DSSI and SPD shows stable inverting and switching operation to make PID control function such direct drive control on the targeted joint. The next progress will target on enhancing and simplifying the CiC module in some control system as adaptive element by considering the real-time hardware constraint and time response.

Acknowledgment

The research work is fully supported by Universiti Malaysia Pahang (UMP) Research Grant under the Research and Innovation Department, RDU RDU100378.

References

[1] S. M. Song, K.J.Waldron, *Machines that walk: The adaptive suspension vehicle*. DARPA and NSF. Cambridge, MA: MIT Press, 1989.

[2] H. Montes, S.Nabulasi, M.A.Armada, "Reliable, Built-in, High-Accuracy Force Sensing for Legged Robots," *International Journal of Robotics Research,* vol. 25, 2006.

[3] R. Hodoshima, T.Doi, Y.Fukuda, S.Hirose, T.Okamoto, J.Mori, "Development of a Quadruped Walking Robot TITAN XI for Steep Slope Operation - Step Over Gait to Avoid Concrete Frames on Steep Slopes," *Journal of Robotics and Mechatronics,* vol. 19, 2007.

[4] M. Raibert, K.Blankespoor, G.Nelson, R.Playter, the BigDog Team, "BigDog, the Rough Terrain Quadruped Robot," in *Proceeding of 17th World Congress, The International Federation of Automatic Control*, Seoul, Korea, 2008.

[5] A. Irawan, K.Nonami, M.R.Daud, "Optimal Impedance Control with TSK-type FLC for Hard Shaking Reduction on Hydraulically Driven Hexapod Robot," in *Proceeding of The International Conference on Intelligent Unmanned Systems 2011 (ICIUS 2011)*, Chiba, Japan, 2011.

[6] Available: http://spidernaut.jsc.nasa.gov/index.html

[7] Yong Kwun Lee, "A bio-mimetic biped robot with a coupled link mechanism and endoskeleton framework," in *The 22nd IEEE International Symposium on Robot and Human Interactive Communication*, Gyeongju, Korea, 2013, pp. 717-721.

[8] J. R. Rebula., P.D.Neuhaus., B.V.Bonnlander.,M.J.Johnson., J.E.Pratt, "A Controller for the LittleDog Quadruped Walking on Rough Terrain," in *Proceeding of IEEE International Conference on Robotics and Automation 2007 (ICRA 2007)*, Roma, Italy, 2007.

[9] V. Dürr, *et al.*, "Behaviour-based modelling of hexapod locomotion: linking biology and technical application," *Arthropod Structure & Development,* vol. 33, pp. 237-250, 2004.

Improvement of the Detection Range of a Chipless RFID for UWB system Using Receiver Diversity with Maximal Ratio Combining

Khalid Mahmud* and S.P Majumder,
Department of EEE, BUET
Dhaka, Bangladesh
*khalidarmy@dhaka.net

***Abstract*—A novel technique for increasing the detection range of a fully passive multiresonator based chipless RFID tag has been proposed using the diversity technique at the reader receiver. Analysis is carried out to find the signal to noise ratio at the output of the diversity receiver using maximal ratio combining (MRC) technique. Bit error rate (BER) is analyzed for Rayleigh fading over the channel. The analyzed approach gives a clear insight regarding the interrelationship between BER, detection ranges and reader received power with the application of receiver diversity and has given improved performance in detecting tag at a comparatively longer distance in different non line of sight (NLOS) environment.**

***Index Terms*—UWB, Chipless RFID, interrogator, tag, MRC.**

I. Introduction

Radio Frequency Identification (RFID) is a wireless data communication technology, which uses RF wave for communication between the reader and the tag [1]. The data carrying device is called a tag or a transponder. A reader or an interrogator is used to read and write the tag's information. [1]–[5]. RFID systems are classified as active and passive systems depending on whether the tags are powered up by their own power supply or by the RF energy provided by the interrogating signal. Passive tags can themselves be classified as chipped and chipless tags and are restricted to relatively small distances to the reader. Passive tags with chips are basically the conventional application specific integrated circuit (ASIC) RFID tags [3]. Passive tags without chips which is otherwise known as chipless RFID, are attractive because of their low cost and their ability to function at extreme temperature and high radiation level.

Thus far the research on chipless RFID mainly focuses on tag design, identification and short range localization [4]–[6]. However there had been numerous publications analyzing a reasonably big read range of a chipless tag [7]–[9]. In [7] the range of the ultra-wideband transmission system has been improved using rake receiver where the round trip distance is found to be less than $100m$. In [8] a general model for UWB channels has been presented that is valid for the high-frequency range (3-10 GHz) in a number of different environments as well as for the frequency. Here the maximum distance achieved is not more than $28m$.

In this paper an analytic approach is presented to improve the detection range of a chipless RFID UWB system using receive diversity in the reader with maximal ratio combining (MRC). Results are evaluated numerically for different indoor environmental conditions and improvement in increase in range is provided for different diversity conditions. The system model has been explained in section II, the analysis part has been explained in section III, in section IV the results are discussed and finally with a brief summary the concluding remarks are added in section V.

II. Proposed System Model

A. System Model of the Proposed Chipless RFID System

The system is composed of a reader (Interrogator) and a tag (Transponder) each equipped with an antenna serving the purpose of both transmitter and receiver simultaneously with the help of duplexers as shown in Fig. 1. When the interrogation signal from the reader reaches the tag, it is received using the receiving antenna of the tag and propagates further on towards the multiresonating circuit imprinted on the tag. The multiresonating circuit encodes data bits introducing attenuation and phase jumps at particular frequencies of the spectrum. After passing through the multiresonating circuit, the signal contains the unique spectral signature of the tag and is transmitted back to the reader receiver by the transmitting antenna of the tag [9].

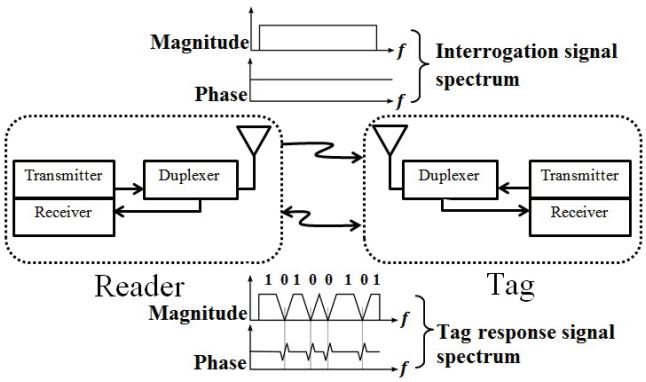


Fig. 1. System model

B. Reader Transmitter/Receiver Model Without Diversity

In the transmitter section of Fig. 2, the interrogator signal passes through a band pass filter (BPF) which acts as an input band limiting device. The band limited signal then goes through an up-converter which converts the frequency of the interrogating signal to a suitable frequency . The up-converted signal is then passed through the high power amplifier (HPA) that amplifies the modulated RF carrier to a suitable level.

On the receiver section of Fig. 2, the reader receives the low-level modulated RF carrier in the downlink frequency

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

spectrum. The low noise amplifier (LNA) is used to amplify the weak received band limited signals and improve the signal to noise ratio (SNR). RF is then reconverted to IF by the down-converter. The demodulator estimates which of the possible symbols was transmitted based on observation.

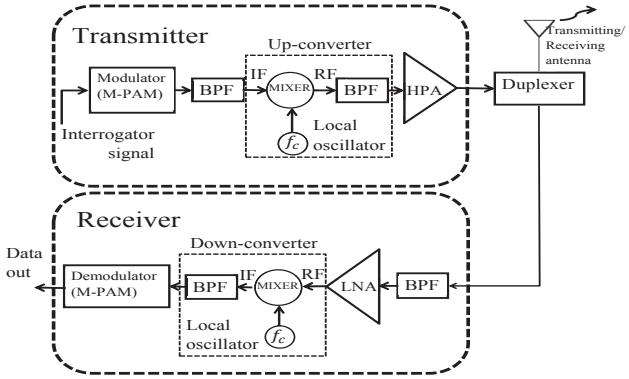


Fig. 2. Transmitter/Receiver block diagram of the proposed RFID reader

C. Reader Receiver Model Using Multiple Receiver With Maximal Ratio Combining (MRC)

In MRC, the combining technique assumes that the receiver is able to accurately estimate the amplitude fading $\alpha_l(t)$ and carrier phase distortion $\theta_l(t)$ for each diversity channel. With the knowledge of complex channel gains, $\alpha_l(t)e^{j\theta_l(t)}, l = 1, 2, ...L$, the receiver coherently demodulates the received signal from each branch. The phase distortion $\theta_l(t)$ of the received signal is removed from the lth branch by multiplying the signal component by $e^{-j\theta_l(t)}$ (Fig. 3). The coherently detected signal is then weighted by the corresponding amplitude gain, $\alpha_l(t)$. The weighted received signals from all the L branches are then summed together and applied to the decision device to achieve the best performance [10].

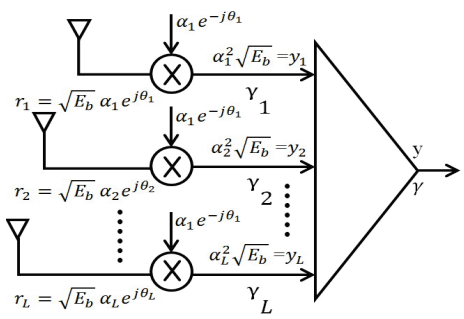


Fig. 3. Maximal ratio combining in the reader receiver end

III. Analysis

The transmitted PAM signal can be expressed as:

$$s(t) = \sqrt{2P_T}e^{j\omega_c t} \times a_i \tag{1}$$

where a_i represents the i-th symbol of the M-array PAM modulation and are given by, a_i=$\pm\frac{A}{2}$, $\pm\frac{3A}{2}$,.....$\pm\frac{(M-1)A}{2}$, M being even. The signal received by tag receiver is given by:

$$r(t) = s(t) \otimes g(t) + n(t) \tag{2}$$

The signal received by the reader receiver with $j - th$ antenna is given by:

$$\begin{aligned} y_j(t) &= r(t) \otimes h_j(t) + n_0(t) \\ &= s(t) \otimes g(t) \otimes h_j(t) + n(t) \end{aligned} \tag{3}$$

The output of the maximal ratio combiner is [11]:

$$\begin{aligned} y_0(t) &= \sum_{j=i}^{L} y_j(t) \otimes h_j^*(t) + n_0(t) \\ &= \sum_{j=i}^{L} s(t) \otimes g(t) \times |h_j(t)|^2(t) + n(t) \end{aligned} \tag{4}$$

$g(t)$ includes the effect of channel gain coefficient and $h_j(t)$ is the channel gain for tag to $j - th$ receiver. Frequency dependent path gain, $g(t)$ in a UWB environment neglecting the interdependence between the direction of the radiation and the directivity of the antennas is expressed as follows [8]:

$$G(f) = \frac{P_{RX}}{P_{TX}} = \frac{1}{2}.G_o.\eta_{TX}.\eta_{RX}.\frac{(\frac{f}{f_c})^{-2(k+1)}}{(\frac{d}{d_0})^n} \tag{5}$$

where, f_{min} = Minimum frequency, f_{max} = Maximum frequency, BW = Band Width, f_c= Reference frequency, d = Round trip distance in meter, d_0 = Reference distance, P_{TX} Transmit power, $G(f)$ = Path gain, P_{RX} = Receive power, G_0 = Path loss in the reference distance in dB, η_{TX} = Antenna transmit efficiency, η_{RX} = Antenna receive efficiency, k = Decaying factor, n = Path loss exponent. By furnishing all the necessary data in equation 5, P_{RX} is calculated. Thereafter, P_{RdBm} is calculated using the formula $P_{RdBm} = 30 + 10 \times log10(P_{RX})$

Probability Density Function in a Rayleigh fading environment using MRC technique is given as follows [10]:

$$P_r(\gamma) = \frac{(\gamma^{L-1}) \times exp^{-\left(\frac{\gamma}{\gamma_c}\right)}}{(L-1)! \times \gamma_c^L} \tag{6}$$

Here $\gamma_c = \frac{P_{RX} \times T_p}{N_0}$ (The average SNR per bit in each diversity channel), L = Receiving antenna, $T_p = \frac{1}{BW}$ (Time gap between two adjacent pulses), $N_0 = K \times T \times F$ (Noise power spectral density), $K = 1.3806488 \times 10^{-23}$ (Boltzmann Constant), T = Room temperature, F = Noise figure, γ = Overall signal to noise ratio [7]. By furnishing the data in equation 6, $P_r(\gamma)$ is calculated.

The bit error probability for an M-array PAM system is given as follows [7]:

$$P_b(\gamma) = \frac{M-1}{M} erfc\left(\sqrt{\frac{3c\gamma}{M^2-1}}\right) \tag{7}$$

Here, M = Alphabet size and c = Number of information bits carried by each symbol. By furnishing the data in equation 7, P_b is calculated.

The average bit error probability over a Rayleigh fading environment with diversity can be expressed as follows [10]:

$$BER = \int_0^\infty P_b(\gamma) \times p_r(\gamma)dx \tag{8}$$

IV. Results and Discussion

Following the analysis presented in section III, the bit error rate (BER) performance of an RFID network is evaluated numerically for different environmental condition as described bellow.

A. Indoor Residential Environment (Fixed Transmit Power)

These environments are critical for home networking, linking different appliances, as well as safety (fire, smoke) sensors

over a relatively small area. The building structures of residential environments are characterized by small units, with indoor walls of reasonable thickness [8]. Here the interrogator transmit power is assumed to be $200mW$. Other parameters are assumed as, $f_{min} = 3GHz$, $f_{max} = 10GHz$, $f_c = 5GHz$, $BW = 500MHz$, $n = 4.58$, $G_0 = -10dBm$, $d_0 = 1meter$, $\eta_{TX} = .80$, $\eta_{RX} = .20$ and $k = 1.53$. These parameters have already been used in previous analysis in a research work and have been accepted as standardized model by IEEE 802.15.4a. [8].

Fig. 4 depicts the plots of BER versus round trip distance in case of an indoor residential environment for different antenna diversity. From the plot it is clear that the BER increases with the increase of distance for a given number of antenna. It is further noticed that as the number of receiving antenna is increased, the achievable distance increases at the same BER. The achievable values of the round trip distance are plotted in Fig. 5 as a function of number of reader receiving antenna, L for several values of BER. It is noticed that, achievable distance is however smaller for lower BER values.

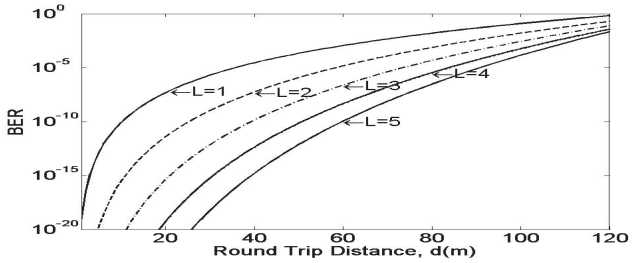


Fig. 4. Plots of BER versus round trip distance with fixed transmit power (Indoor residential NLOS environment)

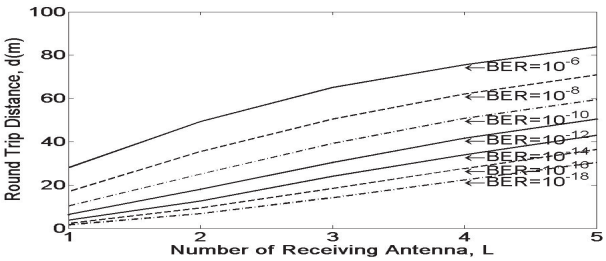


Fig. 5. Plots of round trip distance versus number of receiving antenna with fixed transmit power (Indoor residential NLOS environment)

The plots of BER versus power received by the reader receivers are shown in Fig. 6, for indoor residential environment with diversity in the receiver. It is noticed that the receiver sensitivity at a given BER improves with the increase of the number of receiving antenna at a given transmit power level. The required receiver sensitivity to attain a given BER are shown in Table I for different number of receiving antenna with a fixed transmit power. It is found that there is a significant improvement in receiver sensitivity due to receiver diversity.

B. Indoor Industrial Environment (Fixed Transmit Power)

This environment is characterized by larger enclosures (factory halls), filled with a large number of metallic reflectors and anticipated to lead to severe multipath. The parameters used here is similar to that followed in para IV-A except the value

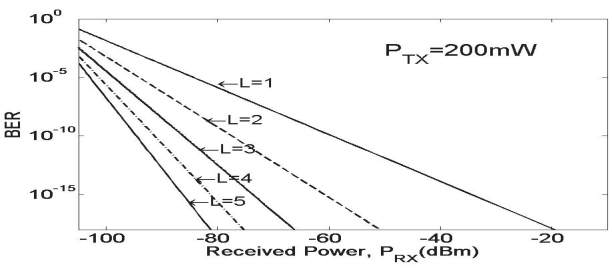


Fig. 6. Plots of BER versus received power with fixed transmit power (Indoor residential NLOS environment)

TABLE I
RECEIVER SENSITIVITY TO ATTAIN A GIVEN BER WITH FIXED TRANSMIT POWER AND DIFFERENT NUMBER OF ANTENNA (INDOOR RESIDENTIAL NLOS ENVIRONMENT)

BER	Received Power in dBm for different L				
	L=1	L=2	L=3	L=4	L=5
10^{-6}	-79.09	-90.59	-96.07	-98.85	-101.2
10^{-8}	-69.31	-83.94	-91.22	-95.26	-98.85
10^{-10}	-59.43	-77.3	-85.92	-91.31	-97.95
10^{-12}	-49.28	-71.05	-81.16	-87.27	-91.31
10^{-14}	-39.23	-63.83	-75.77	-83.14	-87.72
10^{-16}	-29.14	-57.28	-71.01	-79.09	-84.48
10^{-18}	-19.38	-50.72	-66.16	-75.37	-81.25

of path loss exponent, n and the decaying factor, k are assumed to be 2.15 and -1.427 respectively [8].

Fig. 7 and Fig. 8 shows performance similar to indoor residential environment except that the detection distance that can be achieved in case of indoor industrial environment for a particular BER is much higher compared to that achieved in case of indoor residential environment. It is due to the presence of a large number of metallic reflectors which leads to higher multipath propagation.

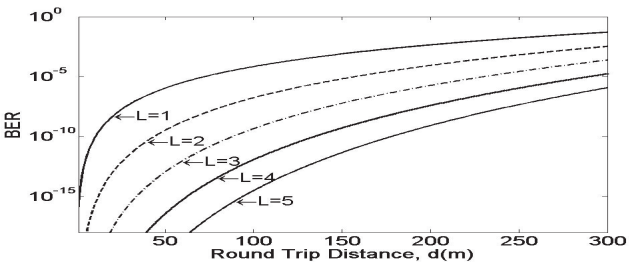


Fig. 7. Plots of BER versus round trip distance with fixed transmit power (Indoor industrial NLOS environment)

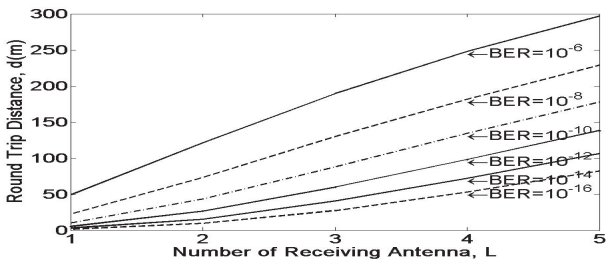


Fig. 8. Plots of round trip distance versus number of receiving antenna with fixed transmit power (Indoor industrial NLOS environment)

Fig 9 and Table II shows performance similar to indoor

residential environment except that at a given BER and at a fixed transmit power level, better improvement in receiver sensitivity is noticed with the increase of receiving antenna in case of indoor industrial environment compared to that noticed in case of indoor residential environment.

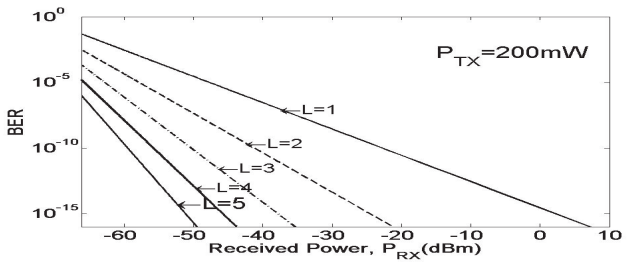


Fig. 9. Plots of BER versus received power with fixed transmit power (Indoor industrial NLOS environment)

TABLE II
RECEIVER SENSITIVITY TO ATTAIN A GIVEN BER WITH FIXED TRANSMIT POWER AND DIFFERENT NUMBER OF ANTENNA (INDOOR INDUSTRIAL NLOS ENVIRONMENT)

BER	**Received Power in dBm for different L**				
	L=1	**L=2**	**L=3**	**L=4**	**L=5**
10^{-6}	-42.71	-54.41	-60.26	-63.77	-66.12
10^{-8}	-32.7	-47.81	-55.16	-59.81	-62.76
10^{-10}	-24.94	-41.18	-50.33	-55.64	-59.58
10^{-12}	-12.63	-34.47	-45.34	-51.83	-56.11
10^{-14}	-2.76	-27.53	-39.98	-47.54	-52.98

C. Variable Power (Residential and Industrial Environment)

The plots of BER versus round trip distance for indoor industrial environment with 5 receiving antenna are shown in Fig. 10, with transmit power as a parameter. The plots indicate that for a given transmit power the BER increases as the distance is increased. It further indicates that for a given distance the value of BER decreases (BER improves) with the increase of transmitted power. The plots shown in Fig. 11 indicate the achievable round trip distance for a given BER as a function of transmit power. These plots clearly indicate that the maximum achievable distance is however lower at lower BER. The same phenomena it expected to happen in case of indoor residential model with inferior detection range. For example at $200mW$ transmit power the maximum achievable distance with 5 receiving antenna at a particular BER of 10^{-10} in case of indoor industrial environment as shown in Fig. 7 is $177.80m$ which is $59.54m$ in case of indoor residential model (Fig. 4). Similar inferiority in case of other transmit power levels $(40mW, 80mW, 120mW\ and\ 160mW)$ are expected to happen in case of indoor residential environment .

V. CONCLUSION

A low-cost novel architecture and a simple analysis method for detecting multiresonator-based chiplss RFID tags in UWB environment using MRC technique is presented where the transmit power has been kept within the limits of $40mW$ to $200mW$. The detection technique is analyzed for indoor residential and indoor industrial environment. It is found that the detection range can be increased significantly by increasing the number of receiving antenna at the reader receiver without

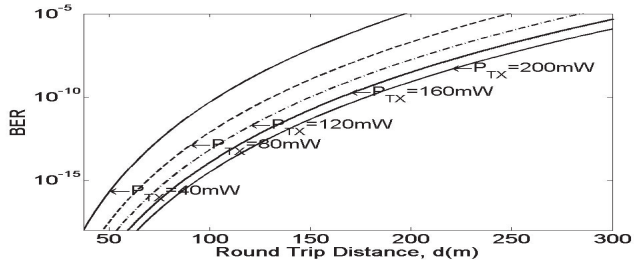


Fig. 10. Plots of distance versus BER with 5 receiving antenna (Indoor industrial NLOS environment-variable power)

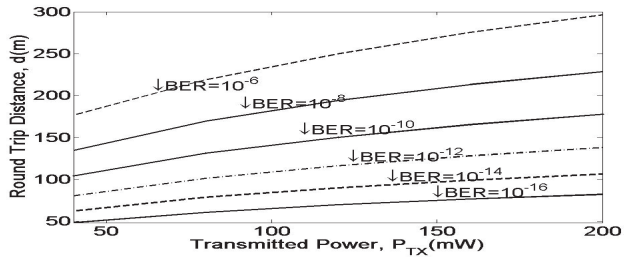


Fig. 11. Plots of transmitted power versus round trip distance with 5 receiving antenna (Indoor industrial NLOS environment-Variable power)

increasing the transmitting power. A comparative study between the two clearly indicates a better detection range for indoor industrial environment than the former one. Similar type of results can be evaluated in case of indoor office environment, outdoor environment, agricultural environment and body area network with certain changes in the parameters [8]. The proposed method not only offers an easy way to evaluate detection range but also allows the user to make certain reasonable assumptions regarding the appropriate scheme for achieving the best performance.

REFERENCES

[1] R. V. Koswatta and N. C. Karmakar, "A novel reader architecture based on UWB chirp signal interrogation for multiresonator based chipless RFID tag reading," *IEEE transcation on microwave theory and techniques*, Vol. 60, No. 9, September 2012.

[2] R. E. A. Anee and N. C. Karmakar, "Chipless RFID tag localization," *IEEE Transactions on Microwave Theory and Trchniques*, Vol. 61, No. 11, November 2013.

[3] K. Finkenzeller, *RFID Handbook: Fundamentals and Applications in Contactless Smart Cards and Identification*. Hoboken, NJ, 2nd ed., Wiley 2003.

[4] A. Ramos, D. Girbau, A. Lazaro, and R. Villarino, "Time-domain measurement of time-coded UWB chipless RFID tags," *Progress In Electromagnetics Research*, Vol. 116, 313–331, 2011.

[5] S. Preradovic, I. Balbin, N. C. Karmakar, and G. F. Swiegers, "Multiresonator-based chipless RFID system for low-cost item tracking," *IEEE Trans. Microw. Theory Techn*, Vol. 57, No. 5, pp. 1411-1419, May 2009.

[6] I. Balbin and N. C. Karmakar, "Phase-encoded chipless RFID transponder for large-scale low-cost applications," *IEEE Microw Wireless Compon. Lett.*, Vol. 19, No. 8, pp. 509-511, August 2009.

[7] S. Gaur and A. Annamalaj, "Improving the range of ultrawideband transmission using rake receivers," *IEEE 7803–7954*, 2006.

[8] A. F. Molisch, D. Cassioli, C. C. Chong, S. Emami, A. Fort, B. Kannan, J. Kunisch, H. Gregory, Schantz, K. Siwiak, and M. Z. Win, "A comprehensive standardized model for ultrawideband propagation channels," *IEEE Transaction on Antenna and Propagation*, Vol. 54, No. 11, 2006.

[9] S. Preradovic and N. C. Karmakar, *Multiresonator-Based Chipless RFID, Barcode of the Future*. Springer, 2012.

[10] J. W. Mark and W. Zhuang, *Wireless Communication and networking*. PHI Learning Private Limited, New Delhi 110001, 2008.

[11] H. Taub, D. L. Schilling, and G. Saha, *Pinciples of Communication Systems*. McGraw-Hill Education India Private Limited, third ed., 2012.

A Brief Overview of Chipless RFID Sensors with EM Transduction

Shuvashis Dey, Emran Md Amin, Jhantu Kumar Saha and Nemai Chandra Karmakar

Electrical and Computer Systems Engineering, Monash University, Clayton, Victoria-3800, Australia
shuvashis.dey@monash.edu

Abstract—**This paper provides a short overview on the state of the art in electromagnetic transduction based chipless RFID sensors. These types of sensors can be used for various applications including structural health monitoring, pressure, airflow and displacement detection, temperature, humidity, partial discharge monitoring and so on. A novel concept of high resolution crack sensing using super wideband antenna is also introduced here along with a brief review on airflow and partial discharge sensors.**

***Index Terms*—SWB Antenna, Airflow, Partial Discharge, Electromagnetic Transduction.**

I. Introduction

Radio Frequency Identification (RFID) is a wireless data capturing technology that utilizes radio frequency waves for extracting the identity of the remotely placed objects automatically. The RFID tag has the potential to replace the barcode on account of its manifold benefits, including its long reading range, non-line of sight reading as well as automated tracking and identification. The recent development of chipless RFID technology gives a new direction in the field of RF identification to address new markets such as items tracking where a unit cost comparable to those of optical barcode is an economic constraint. This technology doesn't embed any silicon ICs for the ID extraction and instead it uses the electromagnetic signature for data encoding which reduces the cost of the RFID tags to a level as low as the barcodes. [1] [2].

Sensors are a substantial part of multifaceted and sophisticated systems of contemporary technology. Sensors and associated measurement circuits pose a challenge towards the development of a low cost intelligent sensing system for which chipless RFID technology could offer an appropriate solution. Recent advancement in wireless sensor networks has paved the way to combine sensor and sensing technologies with RFID system. In addition to the automatic object identification, the RFID technology can monitor and share information about the surrounding environmental condition by using sensor enabled tags. Thus, an object which simultaneously provides its own condition and identification simplifies the total infrastructure and enhances the quality of the information. The passive chipless RFID sensors do not require any emitter integrated into the sensing unit. They are generally composed of integrated sensing components into the resonant circuits with lumped discrete elements such as inductors and capacitors as well as antennas or loops. The passive systems require energy from a reader device in order to be read, but the power levels can be very small. The reader is also utilized to analyse the backscattered signal by the sensing unit. [3] Such sensors are more adaptable to energy harvesting, has a simpler structure and more compatible with harsh environments. They are mostly suitable for applications that require sensing devices with low cost fabrication, small size and long term measurement ability. Chipless RFID tag sensors have huge potentials in retails, structural health monitoring, transportation, biotech, pharmaceuticals, security and surveillance.

Electromagnetic transduction based passive chipless RFID sensors convert the physical quantity into the variation of an EM descriptor. The modification of these descriptors by physical parameter to be measured is backscattered to the reader wirelessly where it is detected. This article presents a brief literature review on selected EM transduction based passive chipless RFID sensors as well as introduces a novel concept in crack sensing through the use of a super wide band antenna.

II. Classifications

The passive wireless sensors can be categorized into four different groups as shown in the following block diagram.

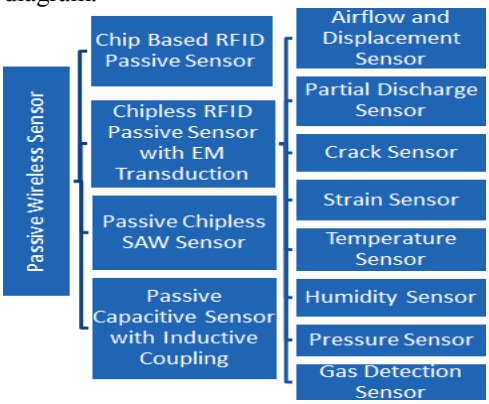


Fig. 1. Classification of Passive Wireless Sensors

The chip based RFID passive sensors can offer a very high accuracy depending upon the accuracy of the integrated transducer, however, its incompatibility with harsh environments and low interrogation distance has curtailed its application area to a very

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

limited range. The SAW sensor is the only commercially available chipless sensor but it has the disadvantage of being non-planar and having inconvenient fabrication process as well as low interrogation distance due to acoustic to EM wave conversion losses. The capacitive sensor has a high autonomy and simple technology but it also has a limited compatibility with harsh environment and requires a very low distance between the reader and sensor. The electromagnetic transduction based Chipless RFID sensors overcome these limitations and offer a pretty high interrogation distance. It can be operated at high frequencies which mean the sensors can be designed with a compact dimension. This type of sensor also meets the need for lightweight, flexible and durable wireless RFID enabled sensor nodes by enabling the use of a great choice of materials.

III. EM Transduction Principle Based Sensors

The electromagnetic transduction principle is used for the remote measurement of various kinds of physical parameters using micro-sensors requiring no active power supply and no signal processing circuitry. [4] When the sensor is placed into an environment to detect the variation or presence of a particular physical phenomenon, its RF parameter such as the resonance frequency or RF impedance is modified by that phenomenon. This modification is instigated by three key principles: the change in dimension or the variation of physical or electrical length a resonator, the modification of EM materials properties such as the variation of a resonator permittivity and the modification of RF coupling between a resonator and a microstructure.[5] Such modifications induces resistive or reactive variation of the system which influences the resonance frequency, impedance, quality factor and the radar cross section (RCS) level to provide a spectral signature containing information about the physical condition of the surrounding environment. The modified RF parameters are backscattered to a RFID reader to be analysed and measured. A wide variety of sensing properties can be addressed by Chipless RFID passive sensor with EM transduction as depicted in fig.1. This article briefly discusses about some of these properties as follows:

A. Airflow and Displacement Sensor

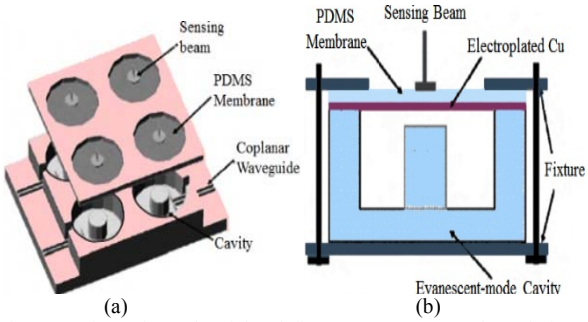


Fig. 2. Design schematic of the airflow sensor array: (a) lateral view, (b) cross-sectional view [6]

Zhao et al. has reported an airflow and displacement sensor based on RF evanescent mode cavity resonator in [6]. For a particular resonant frequency, the evanescent mode cavity resonators have a much smaller size compared to the rectangular or circular cavity resonator. Therefore, the sensitivity of the resonant frequency depending on the dimension of the cavity turns out to be more prominent in case of evanescent mode resonators. The proposed sensor has a rectangular or circular cavity with an electrically shorted centre post. When the air gap between the cavity and the post closes to zero, the parasitic capacitance increases which results in a sharp decrease in the resonance frequency. [6]

Fig.2 shows the schematic diagram of the airflow sensor where an elastic (PDMS) membrane is used to seal the cavity and a small sensing beam protrudes from the membrane surface which extends into the flow for effectively sensing the velocity of the airflow. These membrane and beam form the mechanical part of the sensor while the cavity and the coplanar waveguide (CPW) feeding structures constitute the RF part. The CPW feed is used to excite the resonator and it enables an easy integration of sensing elements with the antenna. [6]

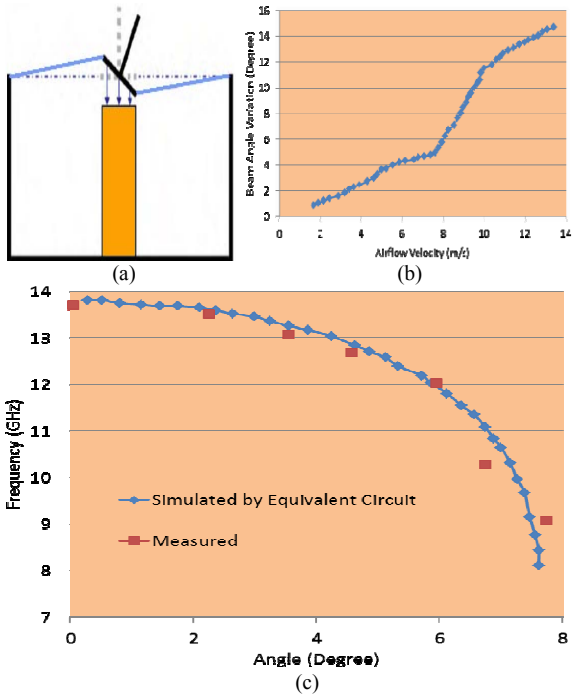


Fig.3 (a) Sensor Mechanism (b) Measured beam angle variation vs airflow velocity (c) Simulated and measured beam angle vs resonance frequency [6]

Fig. 3 (a) shows the operation principle of the reported sensor. If airflow is applied on the sensing beam, it gets deflected and hence changes the capacitance between the elastic membrane of the cavity and the top surface of the post. This in turn modifies the resonance frequency of the cavity resonator. The sensitivity of the sensor is expressed as:

$$S = S_{RF} \times S_{Mechanical}$$

$$\text{Where, } S_{RF} = \frac{\Delta f_r}{\Delta \theta_{beam}} \text{ and } S_{Mechanical} = \frac{\Delta \theta_{beam}}{\Delta v_{airflow}}$$

Fig. 3 (b) and (c) show the RF measurement plots for $S_{Mechanical}$ and S_{RF} respectively. The proposed sensor achieves an equivalent sensitivity of larger than 0.37 GHz/ (m/s) for the airflow velocity sensing. The sensitivity or dynamic range can be easily adjusted by controlling membrane thickness or sensing beam dimension in order to meet requirements of specific applications. The proposed sensor can also be used for the wireless sensing of normal directional displacement by using similar principles. [6] The fabricated displacement sensor is wirelessly interrogated by integrating an UWB antenna with the cavity resonator based sensor as well as two external reader antennas. A vertical force is applied to the sensing beam to generate a displacement. When the beam is displaced, a significant shift in the resonant frequency occurs.

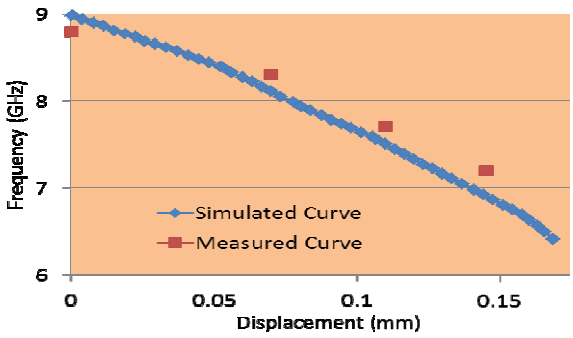


Fig. 4. Measured resonant frequency by wireless interrogation vs. beam displacement. [6]

Fig. 4 shows the resonance frequency versus vertical beam displacement and indicates a sensitivity of 16.7 MHz/µm. [6]

B. *Partial Discharge Sensor*

Electromagnetic transduction based chipless RFID sensors can be utilized for online monitoring of high voltage equipment in a switchyard without sacrificing reliability and capacity deployment. A partial discharge in HV equipment occurs when the electric field across a dielectric medium changes in such a way that a localized current is created. This current induces RF emission which is wirelessly captured and analysed for PD level and location detection [7]. A novel remote PD sensor is reported in [7], by the authors at the Monash University. It uses frequency modulation based technique for the detection of potential PD sources in a switchyard which reduces the reflection and multipath effect on source localization. This sensor consists of cascaded coplanar quarter wavelength tri section Stepped Impedance Resonators (SIR) that are coupled to a Microstrip line. The resonators act as band stop filters that modify the received UWB interrogation signal from the reader to provide with distinct frequency signatures to encode data as identification bits.

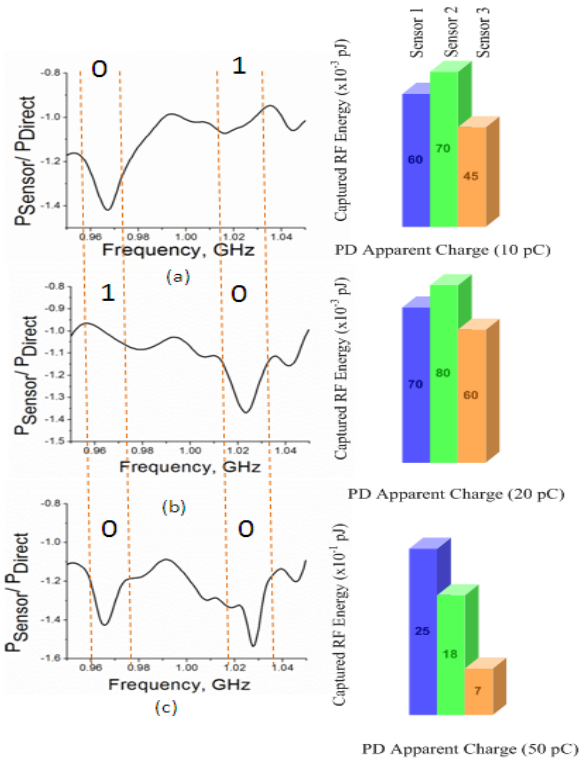


Fig. 5 Power spectrum for (a) sensor1, (b) sensor2 and (c) sensor3 having data ID 01, 10 and 00. (d) Measured RF energy for three sensors for PD apparent charge of 10, 20 and 50 pC.

Moreover, the sensor provides information about PD signal level by calculating cumulative energy. Three PD sensor operating between 950 -1050 MHz is developed using two SIR filters. The filters can encode data 0 and 1 depending on presence and absence of a resonance dip in frequency spectrum. Fig. 5 (a) - (c) shows the measured spectral response of sensor 1, sensor 2 and sensor 3 having encoded bits 01, 10 and 00. Also, Fig. 5 (d) shows measured cumulative energy for each sensor at various PD levels. For PD apparent charge 10, 20 and 50 pC, sensor 1 and sensor 2 have higher RF energy than sensor 3. This is due to the presence of both the filters in series resulting higher attenuation. However, in frequency spectral analysis the encoded data bits can be clearly identified.

C. *Crack Sensor*

In [8], it has been shown that a short duration UWB pulse which lasts only several nanoseconds can detect and localize a crack by utilizing a transmission line of few meters. The sensing system is established on the Time Domain Reflectometry (TDR) based chipless RFID technique. When an RF interrogating pulse from the reader reaches the sensor, a portion of it is immediately reflected back and another part is harnessed by the receiving antenna of the sensor which travels through the long transmission line segment until it experiences the abrupt termination of the transmission line. At this point the signal gets reflected

back towards the antenna of the sensor and ultimately gets retransmitted into the free space. These phenomena at the sensor instigate two prominent RF echoes: i.e. the initial reflection of the sensor antenna and the secondary retransmission of the harnessed energy. The time difference of arrival (TDOA) of these two echoes at the reader is determined by the length of the transmission line which in fact introduces a propagation delay to the secondary echo. In case of a structural failure, the sensor transmission line is ruptured which causes an open circuit termination at a shorter length. This premature termination of the transmission line changes the TDOA between the two echoes received at the reader and hence the TDOA can be used to determine the presence of a crack in the structure and also to localize it. [8]

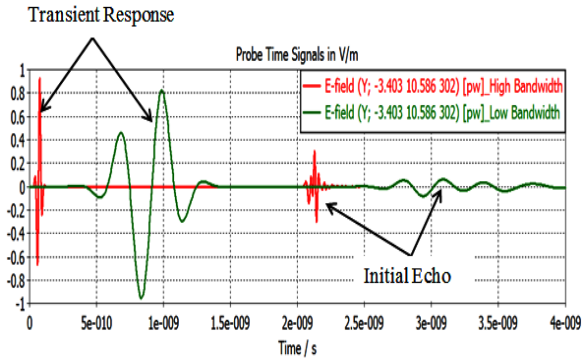


Fig. 6: Time domain response indicating potential application of the SWB antenna as high resolution sensor

A Super Wide Band (SWB) antenna based sensor can achieve more accurate localization of the crack by enabling extremely short SWB pulses. These short duration pulses also enable very high resolution estimation of distance to targets. If the crack is very near to the sensing antenna, it will not have enough time to be detected with the UWB sensors as the initial and secondary echo will be superimposed on each other. However, since the SWB sensors have shorter pulses, the initial reflection will be decayed before the arrival of secondary reflection which will give them enough time to detect the TDOA. For this type of sensors, the time domain antenna transients also decay extremely quickly to enhance the probabilities of clearly distinguishing both the echoes. Here, the SWB antenna is interrogated with a plane wave and the reflected signal is captured by a receiving probe. Fig. 6 shows the reflected time domain response at the probe. It depicts the comparison between the time domain signals of both UWB and SWB antennas. From fig. 6, it can be observed that for low bandwidth, the transient response and the initial echo is quite widespread in the time domain which indicate their longer duration for getting decayed. However, for higher bandwidth (SWB) the same antenna time domain responses get decayed very quickly and offer extremely short pulses. Thus the SWB sensor enables reflected signals to be detected more precisely and accurately since one pulse does not interfere the other.

IV. Conclusions

This paper presents a brief review on selected EM transduction based chipless RFID sensors along with the state of the art in this research field. A novel concept of crack sensing is also introduced here by using SWB technology. These sensors offer unlimited energy autonomy and consumes no power due to the absence of any embedded chip or battery. They are highly compatible with harsh environments like high temperature and high radiation level and offer a substantially higher reading distance compared to the other passive sensors. A wide range of materials can be used to design such sensors which enable them to be planar, lightweight and flexible. The printable structures and high frequency designs reduce the size and cost of these sensors significantly. Despite the enormous potential and exciting prospects, there exist a number of limitations in this technology in terms of security, networking, mass production and deployment, etc. This technology is still in its infancy and therefore, a huge room for improvement is present to bridge the gap between the transducer concept validations to actual sensor development with desired functionalities. Identifying smart materials with multi-sensing capabilities and their RF characterization as well sensor designs for various applications open a new research horizon in chipless RFID sensing technologies. Addressing these issues would be a revolutionary step towards the commercialization of EM based chipless sensors and pave the way for broadening the focus for applied research enabling the implementation of ambient intelligence.

References

[1] Islam, M.A.; Karmakar, N.C., "A Novel Compact Printable Dual-Polarized Chipless RFID System," *Microwave Theory and Techniques, IEEE Transactions on*, vol.60, no.7, pp.2142, 2151, July 2012

[2] Vena, A.; Perret, E.; Tedjini, S., "A compact chipless RFID tag using polarization diversity for encoding and sensing," *RFID (RFID), 2012 IEEE International Conference on* , vol., no., pp.191,197, 3-5 April 2012

[3] Ibrahim, A., Cumming, D.R.S. "Passive single chip wireless microwave pressure sensor" (2011) *Sensors and Actuators, A: Physical*, 165 (2), pp. 200-206.

[4] Chebila, F.; Jatlaoui, M.M.; Pons, P.; Aubert, H., "Reconfigurable multi-band scatterers for micro-sensors identification," *Antennas and Propagation Society International Symposium, 2009. APSURSI '09. IEEE*, vol., no., pp.1, 4, 1-5 June 2009

[5] Pons P, Aubert H, Menini P, Tentzeris M (2012) "Electromagnetic transduction for wireless passive sensors." Proc. Eurosensors XXVI, Procedia Engineering 47, 9–12 September 2012, Krakow, Poland, pp. 1474–1483

[6] Y. Zhao, Y. Li, B. Pan, S.H. Kim, Z. Liu, M.M. Tentzeris, J. Papapolumerou, M.G. Allen, "RF evanescent-mode cavity resonator for passive wireless sensor applications," Sensor and Actuators A, vol. 161, pp. 332-328, 2010.

[7] Amin, Emran Md; Karmakar, N., "Partial discharge monitoring of High Voltage equipment using chipless RFID sensor," *Microwave Conference Proceedings (APMC), 2011 Asia-Pacific*, vol., no., pp.1522, 1525, 5-8 Dec. 2011

[8] Kalansuriya, P.; Bhattacharyya, R.; Sarma, S.; Karmakar, N., "Towards chipless RFID-based sensing for pervasive surface crack detection," *RFID-Technologies and Applications (RFID-TA), 2012 IEEE International Conference on*, vol., no., pp.46, 51, 5-7 Nov. 2012

A Local Phase Correction Method to Mitigate Mode Conversion in Coupled Transmission Line Bends

Mohammed Ershad Ali[1,*]

[1]Inphi Corporation
112 S. Lake View Canyon Road, Westlake Village, CA 91362, USA
[*]eali@inphi.com

***Abstract*— In this paper, a local phase correction method is proposed and demonstrated to effectively mitigate mode conversion in coupled transmission line bends leading to significant improvement in their high-speed performance. Signal loss and mode conversion improvements by more than 5 dB and 16 dB respectively at 50 GHz are reported for experimental prototypes. Good correlation has been established among analytical predictions, full-wave electromagnetic simulations and measurements.

***Index Terms*—Bend discontinuity, common mode filtering, common mode noise suppression, coupled lines, differential lines, high-speed routing, intra-pair skew, microwave discontinuity, mixed mode, mode conversion, phase correction, signal integrity, transmission line bends.**

I. Introduction

Raw signalling speeds in both wired and wireless communication networks have been steadily increasing to meet unrelenting demands for higher data bandwidth and throughput. As a result, high speed transmission lines are becoming essential in all levels of interconnection in a system. Of particular importance are coupled transmission lines used as differential lines which are most common in today's high speed systems, as they provide higher bandwidth and better immunity from noise and crosstalk. In order to achieve optimal performance with this type of differential lines, careful attention needs to be paid to every details of component placement and signal routing from early on in the design cycle. In a densely populated system board or component package or integrated circuit chip, turns and bends of signal lines are unavoidable. These turns and bends would eventually limit the bandwidth performance of a system employing coupled transmission lines [1], [2].

Coupled transmission lines support two modes of propagation, namely differential (odd) mode and common (even) mode. When a uniform coupled transmission line bends, conversion from one mode to another is inevitable. A purely differential mode before the bend will be transformed into a mixed mode, partly differential and partly common mode, after the bend. In a differential system, common mode is undesirable. Not only it takes energy away from the desired differential mode increasing the insertion loss as such, it also causes crosstalk, noise and electromagnetic interference.

In order to avoid bend related signal integrity issues in a differential channel, the first preference would always be to keep critical signal paths free of bends, i.e. route them in straight lines. However this is often not possible as the signal lines have to stay clear of many keep-out zones typically found in interface, fanout and connector access areas on boards and substrates. For signals that continue in the same direction, bends occur in pairs. If the coupled lines do not have modal dispersion, the bends in a pair compensate each other producing only minimal effects on signal propagation. However if the coupled lines exhibit modal dispersion, it would be shown later in this paper that the pair would no longer be fully compensating in nature and can cause significant deterioration of signals. For signals that do not continue in the same direction, some bends would always remain uncompensated. Therefore for a method to be universally effective in mitigating bend effects, it has to locally compensate each and every bend. A number of such schemes have been recently demonstrated to improve differential bend performance [3]-[5]. However there are practical challenges in implementing these methods in real systems.

In this paper, a novel and practical local phase correction method which mitigates detrimental mode conversion in a coupled line bend is presented. The working principle and effectiveness of the method will be illustrated first based on a simplified theory of mode conversion in single and double bends. A simple implementation will be considered next for proof-of-concept demonstration and experimental validation of full wave electromagnetic simulations and analytically computed results.

II. Simplified Theory of Mode conversion in coupled line bends

In order to gain fundamental insight into how mode conversion and associated signal degradation occur in coupled line bends, a theory is presented here with some simplifying assumptions. Fig. 1 delineates a structure consisting of a length of coupled lines with two bends, encountered often in routing of high speed signal lines. The individual bends are analyzed first following a similar approach as in [6].

The entire length in Fig 1 is divided into three segments of uniform coupled lines, between reference planes R_{in} and R_1, R_1' and R_2', R_2 and R_{out}. It is assumed that signal propagates in these uniform segments either as odd mode or even mode or as a linear combination of them. In the regions between R_1 and R_1' and between R_2' and R_2, only one of the constituent lines have non-zero lengths. This asymmetry causes intra-pair skew or equivalently excess phase in the line having the extra length. Multiplying the extra length by a phase constant, the excess phase ϕ in a single bend can be expressed as:

$$\phi = \frac{4\pi f k d}{c} \tan\frac{\alpha}{2} \tag{1}$$

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

Where f is frequency, d is center-to-center distance between the lines or pitch, c is speed of light, α is the angle of the bend and k is equivalent to an effective index of refraction that is a function of the propagation constants of the coupled lines.

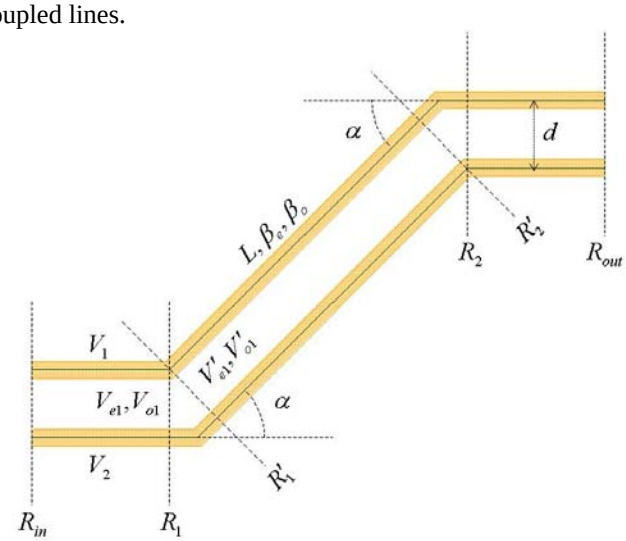


Fig. 1 Coupled lines with two bends. Reference lines are used to identify uniform segments of the lines and asymmetric regions in the bends.

Consider odd mode excitation at the input in Fig 1. The mode purity is maintained up to R_1 making terminal voltages equal and opposite in polarity there, i.e. $V_1 = -V_2 = V$. This is equivalent to the condition $V_{o1} = V$ and $V_{e1} = 0$, where V_{o1} and V_{e1} are odd and even mode voltages respectively. Although there is no even mode before the bend, the excess phase in the bend produces some even mode, V_{e1}' on the right of R_1' at the cost of reduced odd mode V_{o1}':

$$V_{e1}' = \frac{V}{2}(1 - e^{-j\phi}), \quad V_{o1}' = \frac{V}{2}\left(1 + e^{-j\phi}\right) \tag{2}$$

So the desired odd mode suffers insertion loss IL_S, and an undesired even mode of magnitude C_S gets excited due to mode conversion induced by excess phase in the bend. IL_S and C_S can be expressed in dB as:

$$IL_S = 20\log_{10}\left|\cos\frac{\phi}{2}\right|, \quad C_S = 20\log_{10}\left|\sin\frac{\phi}{2}\right| \tag{3}$$

The odd and even modes now propagate in segment 2 from R_1' to R_2' with their respective propagation constants, β_o and β_e. Each of these modes experiences insertion loss and mode conversion as it passes through the second bend. Cascading the effects of the first bend, mixed mode propagation in segment 2 and the second bend, the total insertion loss of desired odd mode, IL_T and mode conversion to even mode C_T in dB can be determined as:

$$IL_T = 10\log_{10}\left(\cos^2\frac{\Delta\beta L}{2} + \cos^2\phi\sin^2\frac{\Delta\beta L}{2}\right) \tag{4}$$

$$C_T = 20\log_{10}\left|\sin\phi\sin\frac{\Delta\beta L}{2}\right| \tag{5}$$

Where L is the length of segment 2, and $\Delta\beta = \beta_e - \beta_o$ represents modal dispersion.

Starting with even mode excitation at the input, it can be shown that (3)-(5) also hold in exact same forms. Therefore, IL_S and IL_T above represent the insertion loss of launched mode, odd or even, and C_S and C_T represent the conversion to the other mode. For mixed mode excitation, each launched mode can be treated separately and then superposed to obtain total insertion loss and mode conversion.

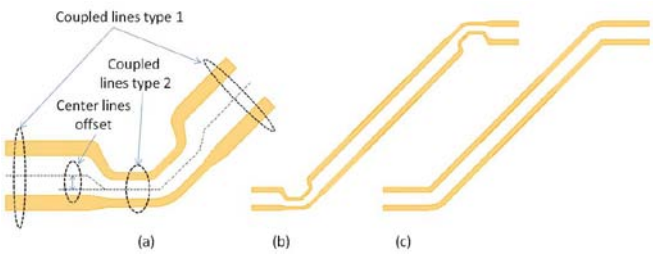


Fig. 2 (a) Proposed local phase correction method in coupled line bends, (b) coupled line with two phase corrected bends, (c) coupled line with regular bends having no phase correction

Several observations are worth noting from (3)-(5). First, for the trivial case of coupled lines with no bends, $\phi = 0$ and the launched mode suffers no insertion loss (0 dB) and no mode conversion ($-\infty$ dB). Second, for a single bend there would always be some insertion loss and mode conversion and their magnitude is a function of the pitch of the lines. Third, for a pair of bends separated by a uniform section of coupled lines as illustrated in Fig. 1, if the modal dispersion $\Delta\beta = 0$, signal suffers no insertion loss or degradation due to the bends. In this case, the bends are compensating each other. Fourth, if modal dispersion exists, even for compensating bends, there would be insertion loss and modal conversion and furthermore they would vary as a function of separation between the two bends. Last, both insertion loss and modal conversion are frequency dependent.

In the above analysis, impedance discontinuity at the bends and multiple reflections in the lines as a result of that are ignored. In addition, the odd and even modes are assumed to be perfectly terminated at both input and output ends. These are not too restrictive assumptions as a good design would always minimize impedance discontinuity at the bends and provide good terminations for all modes.

III. LOCAL PHASE CORRECTION METHOD TO MITIGATE MODE CONVERSION

Any strategy to mitigate mode conversion has to provide a means to eliminate, cancel or minimize the excess phase ϕ at the bends explained earlier. A straightforward way to minimize ϕ would be to use much smaller d, i.e. to use strongly coupled lines at the bends, since it will lead to smaller ϕ according to (1). A method that does just that is

demonstrated to improve bend performance recently [3]. However this approach requires a very small value of d and a long impedance matching taper, severely limiting its use in practical systems.

The method proposed in this paper is illustrated in Fig. 2(a). It works by modifying the construction of the coupled line in the bend area in a way such that physical lengths of the inner and outer lines are approximately equal. The modification essentially results in a narrower coupled line segment in the bend center with short transition areas on both sides connecting to incoming and outgoing lines. As can be seen in Fig. 2(a), the coupled line segment is placed slightly offset outward from the center line. This arrangement allows longer lengths for the inner line in the transition areas. By adjusting the offset, it is possible to equalize the lengths of the lines leading to a net zero excess phase. Although the reduced d in the bend area contributes to some degree to lower the excess phase, the majority of the correction comes from the introduction of the extra length of the inner line. The impedance discontinuity introduced is not too large to warrant separate impedance matching sections. The modified bend remains compact, its size being very similar to that of the original one. By judiciously choosing the parameters of the coupled lines as well as the shape of the small transition areas, good performance can be achieved.

IV. SIMULATION STUDIES

The proposed phase correction scheme lends itself to a variety of implementations based on substrate or board technology being used, dielectric and metal stack-up, material properties, fabrication capabilities and so on. In general, a good amount of design work would be involved in producing a phase corrected coupled line bend that meets often stringent specifications of a high-speed system. In order to keep this design work to a minimum, a simple yet fairly practical implementation is considered here with the goal of clearly demonstrating the effectiveness of the method. A single layer dielectric, with two metal layers, top and bottom, was used to construct the coupled line and the bends. Fig. 2(b) and 2(c) show the coupled lines with phase corrected bends and regular bends respectively. Only top metal patterns are shown. The bottom metal is a uniform pattern that serves as the ground plane for the coupled lines. Except for the bends, both types of lines have identical line widths and spacing with total lengths being approximately equal.

The coupled lines were analyzed as 4-port three dimensional structures using full wave electromagnetic simulations. A finite element based commercial software tool was used for this purpose. A frequency sweep from DC to 50 GHz was performed and results were obtained from the tool in the form of terminal scattering parameters. The scattering parameters were subsequently post-processed in a circuit simulator to first generate mixed mode scattering parameters and then compute insertion loss and mode conversion. In what follows, only results for the input differential excitation will be presented.

The differential insertion loss plots are shown in Fig 3. Significant signal degradation can be noticed at high frequencies for the coupled lines with regular bends. The loss is more than 6 dB at 50 GHz. In contrast, coupled lines with phase corrected bends have vastly improved loss performance exhibiting approximately 1 dB of loss at 50 GHz. The conversion from differential to common mode is presented in Fig. 4. As expected, a sharp reduction of mode conversion by as much as 16 dB at 50 GHz is observed for the phase corrected bends compared to regular bends.

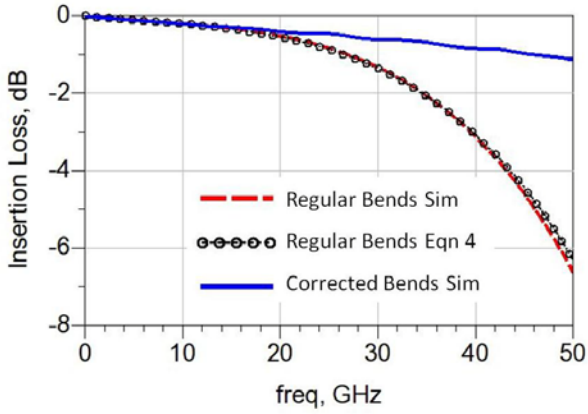


Fig. 3 Differential insertion loss for regular and phase corrected bends as a function of frequency. Results from both full wave electromagnetic simulations and analytical computations using (4) are plotted.

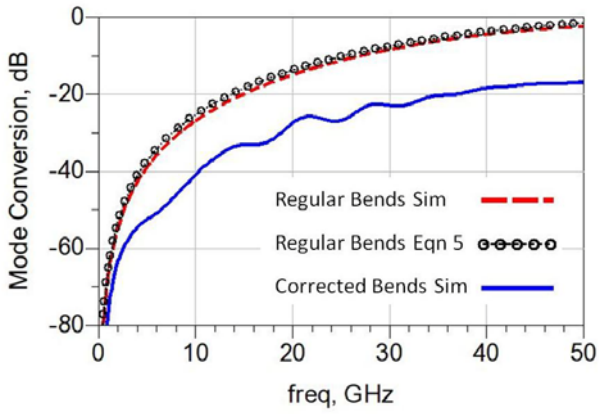


Fig. 4 Differential to common mode conversion for regular and phase corrected bends as a function of frequency. Results from both full wave electromagnetic simulations and analytical computations using (5) are plotted.

Also plotted in Fig. 3 and 4 are insertion loss and mode conversion computed analytically using (4) and (5) respectively for the coupled lines with regular bends. The analytical insertion loss also includes straight line losses, such as ohmic and dielectric losses, in addition to the loss due to bends given by (4). The straight line losses were computed from full wave electromagnetic simulation of straight coupled lines with identical line parameters and total lengths as was used in the case of the coupled lines with regular bends. The agreement between analytically computed values and those obtained from full wave electromagnetic simulations is excellent over a wide band from DC to 50 GHz.

V. EXPERIMENTAL DEMONSTRATION AND VALIDATION

Prototypes for both coupled lines with regular bends and with phase corrected bends shown in Fig. 2(b) and 2(c) were fabricated in a commercial substrate manufacturing house. In order to efficiently couple high frequency signals in and out of the coupled lines, launch structures compatible with GSGSG type on-wafer probing were integrated on both ends of the lines. Other features such as via shielding to quench substrate

mode excitations were also added on all sides of the coupled lines at a reasonable distance.

The measurements and characterization of the prototypes were performed in an on-wafer probing station equipped with a 4-port vector network analyzer (VNA). The scattering parameters obtained from the VNA were post-processed in a circuit simulator to compute differential insertion loss and mode conversion, which are plotted in Fig. 5 and 6. The corresponding plots from full electromagnetic simulation are overlaid in the same graph for comparison purposes. The agreement between measurement and simulation is almost perfect. A few spikes in the measured insertion loss data at high frequencies indicate slight coupling and resulting energy loss to substrate modes. With measured insertion loss improvement of 6 dB and mode conversion reduction of 16 dB at 50 GHz, the effectiveness of the local phase correction method is validated over a wide band.

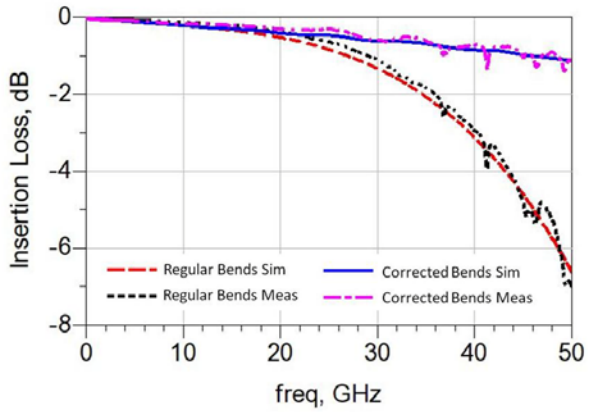


Fig. 5 Measured differential insertion loss for regular and phase corrected bends as a function of frequency. Results from full wave electromagnetic simulations are overlaid for comparison.

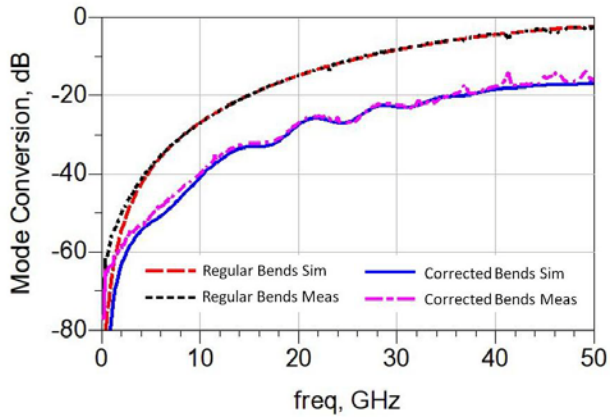


Fig. 6 Measured differential to common mode conversion for regular and phase corrected bends as a function of frequency. Results from full wave electromagnetic simulations are overlaid for comparison.

The simplified theory in section II predicted the insertion loss and mode conversion quite well over a wide band, even if it neglected one important factor, which is the impedance discontinuity caused by the bends. This vindicates the underlying assumption that the major contribution to signal degradation in coupled line bends comes from excess phase, not from impedance discontinuity. This is further confirmed by low return loss observed for the phase corrected bends as shown in Fig. 7. Although degraded somewhat compared to regular bends, the return loss of corrected bends stayed below -20 dB to 25 GHz and below -10 dB to 50 GHz, a level better than what is typically acceptable in high speed systems. The large discrepancy between simulation and measurements in Fig. 7 comes from inclusion of launch structures in experimental prototypes.

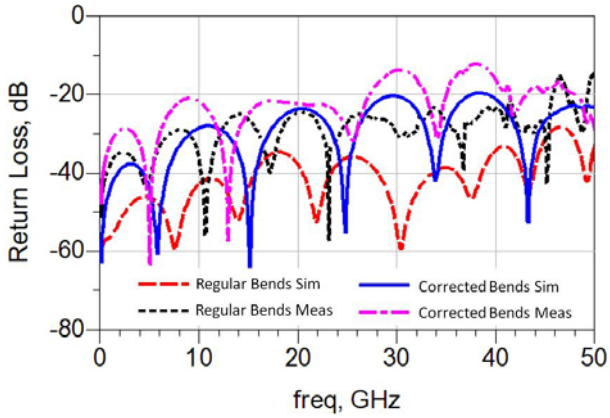


Fig. 7 Measured and simulated return loss characteristics for regular and phase corrected bends as a function of frequency.

VI. Conclusions

Signal degradation in coupled line bends is mainly attributed to mode conversion from excess phase in one of the constituent lines. A simplified analytical theory is first presented to quantify the signal degradation in a single or a pair of bends highlighting the importance of locally mitigating the bend effects. In this regard, a local phase correction method is proposed in which narrow coupled lines are offset placed inside the bends in such a way that leads to equal line lengths in the bend area. Proof-of-concept prototypes were built and characterized. The phase corrected bends yielded excellent performance over a wide frequency range from DC to 50 GHz. At 50 GHz, insertion loss and mode conversion improved by more than 5 dB and 16 dB respectively, while return loss performance remained acceptable. Measured characteristics correlated well with those obtained using full wave simulations, and the simplified analytical theory.

References

[1] Jun Fan; Xiaoning Ye; Jingook Kim; Archambeault, B.; Orlandi, A, "Signal Integrity Design for High-Speed Digital Circuits: Progress and Directions," *IEEE Transactions on Electromagnetic Compatibility*, vol.52, no.2, pp.392-400, 2010.

[2] "Practical Notes on Mixed-Mode Transformations in Differential Interconnects," *Simbeor Application Note# 200_01, January 2009*, [Online] Available http://www.simberian.com/AppNotes.php.

[3] Gazda, C.; Vande Ginste, D.; Rogier, H.; Wu, Ruey-Beei; De Zutter, D., "A Wideband Common-Mode Suppression Filter for Bend Discontinuities in Differential Signaling Using Tightly Coupled Microstrips," *IEEE Transactions on Advanced Packaging*,, vol.33, no. 4, pp. 969-978, 2010 .

[4] Chi-Fang Huang; Ching-Lung Wu; Chun-Fu Shih, "Phase redress for the bend on differential microstrip lines," , *2011 XXXth URSI General Assembly and Scientific Symposium, 2011*, pp.1-4.

[5] Guang-Hwa Shiue; Wei-Da Guo; Chien-Min Lin; Wu, Ruey-Beei, "Noise reduction using compensation capacitance for bend discontinuities of differential transmission lines," *IEEE Transactions on Advanced Packaging*, , vol.29, no.3, pp.560-569, 2006.

[6] Senguttvan, R.; Hayden, L.; Weisshaar, A, "Mode coupling in coplanar waveguide bends: a simple four-port model," *33rd European Microwave Conference, 2003*, vol.2, pp. 643-646.

Analysis of Improving the Detection Capability of RADAR in Presence of Jammer by Controlling Radar Effective Temperature

Md. Maynul Islam[1], Mubdiul Islam Rizu[2] and Jerin Tasnim Himi[3]
Department of Electrical, Electronic and Communication Engineering
Military Institute of Science and Technology, Mirpur Cantonment, Dhaka.
E-mail: [1]maynul.mist@gmail.com, [2]mubdiulislam@yahoo.com, [3]jerin.himi93@gmail.com

***Abstract*—In RADAR technology detection capability of radar is one of the most important factors. It depends on various parameters. Transmit power, wavelength, antenna gain, frequency, radar cross section area have effects on detection range of radar. Normal operation of radar may be hampered in presence of jammer. Jamming is usually referred to as an Electronic Countermeasure (ECM) which has an intention to disturb regular operation of radar. Range reduction factor of radar is decreased if there is a jammer. For this reason, the detection range of radar is also decreased. This paper has proposed a method to increase the detection capability of radar in presence of jammer by controlling radar effective temperature.**

Keywords*— *Antenna Gain, ECM, RCS, RRF, SNR

I. INTRODUCTION

Radar is an object-detection system that uses radio waves to determine the range, altitude, direction, or speed of objects. It can be used to detect aircraft, ships, spacecraft, guided missiles, motor vehicles, weather formations, and terrain. The usual noise that exists in a radar receiving system is partly of thermal origin and partly from other noise-generating processes. Most of these other processes produce noise which, within typical receiver bandwidths, has the same spectral and probabilistic nature as does thermal noise. Therefore it can all be lumped together and regarded as thermal noise. Radar jamming and deception is the intentional emission of radio frequency signals to interfere with the operation of a radar by saturating its receiver with noise or false information [1]. Jamming may be intentional, as with an electronic warfare tactic, or unintentional, as with friendly forces operating equipment that transmits using the same frequency range. Jamming is considered an active interference source, since it is initiated by elements outside the radar and in general unrelated to the radar signals. There are two types of radar jamming: Mechanical and Electronic jamming. Electronic jamming is a form of Electronic Warfare where jammers radiate interfering signals toward an enemy's radar, blocking the receiver with highly concentrated energy signals. The two main technique styles are noise techniques and repeater techniques. The three types of noise jamming are spot, sweep, and barrage. RADAR jamming decreases the range detection capability of radar. In this paper, we have discussed a method of increasing detection capability of radar in presence of jammer by controlling radar effective temperature.

II. SYSTEM MODEL

Let us consider a radar system whose detection range is 'R' in the absence of jamming. Then the signal to noise ratio at the output of the receiver will be,

$$(\mathrm{SNR})_o = \frac{P_t G^2 \lambda^2 \sigma}{(4\pi)^3 k T_e B F L R^4} \quad (1)$$

Where $'P_t'$ is called peak transmitted power, 'G' is the gain of the antenna, 'λ' is the wavelength, 'σ' is called Radar Cross Section(RCS), 'k' is the Boltzman's constant, $'T_e'$ is the effective noise temperature in degree Kelvin, 'B' is the Radar Operating Bandwidth, 'F' is the Noise Figure and 'L' is the radar losses [2, 3] .

The term Range Reduction Factor (RRF) refers to the reduction in the radar detection range due to jamming. More precisely, in the presence of jamming the effective radar detection range is,

$$Rdj = R \times RRF \quad (2)$$

Most jammers have a constant power output - that is, they always transmit the maximum available power of the transmitter (excepting desired ECM modulation). Some jammers also have a constant gain (linear) region [4, 5]. Usually these are coherent repeaters that can amplify a low level radar signal to a power that is below the level that results in maximum available (saturated) power output. At some radar to target range, the input signal is sufficiently high that the full jammer gain would exceed the maximum available power and the jammer ceases to be constant gain and becomes constant power [6].

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

In a constant gain jammer, the radar power received by the jammer,

$$10\log S_{Rj} = 10\log P_t + 10\log G_t - \alpha_{Tx} + 10\log G_{ja(Rx)}$$

(Factors in dB) (3)

The jammer constant gain power output is:

$$10\log P_{jCG} = 10\log S_{Rj} + 10\log \mathrm{G_{ja}} \quad (4)$$

By definition:

$$P_{jCG} \leq P_j \quad (5)$$

In the above equations, $'\mathrm{S_{Rj}}'$ is the Radar signal at the jammer input (receive antenna terminals), $'G_{ja(Rx)}'$ is the gain of the jammer receive antenna, $'G_t'$ *is* the gain of the jammer, $'\alpha_{Tx}'$ *is* the one-way free space loss from the radar to the target , $'P_{jCG}'$ *is* the jammer constant gain power output, $'P_j'$ *is* maximum jammer power output.

III. SIMULATION

For different effective temperature of RADAR, simulation has been carried out between range reduction factor and jammer peak power. Radar antenna gain of 45 dB, jammer antenna gain of 3 dB, frequency of 10 GHz, radar to jammer range of 750 km, radar operating bandwidth of 10 MHz have been used for simulation purpose.

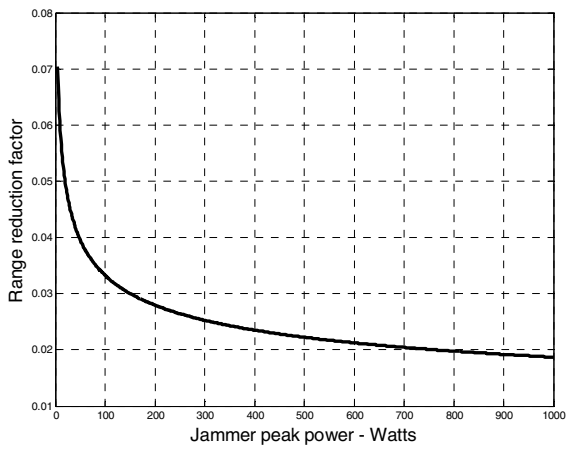
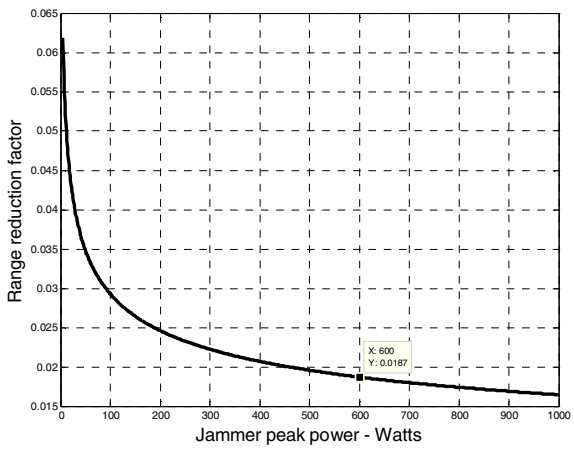


Figure 1. Relationship between range reduction factor and jammer peak power

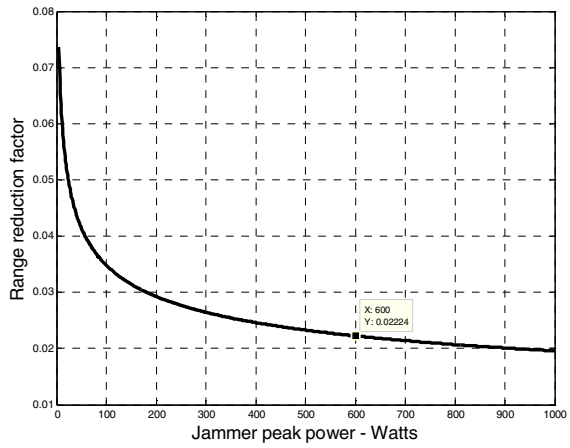
Figure 1 shows the curve which indicates the relationship between range reduction factor and jammer peak power. It is seen than increasing of jammer peak power reduces the radar range reduction factor. As a result radar detection capability decreases.

IV. RESULT AND DISCUSSION

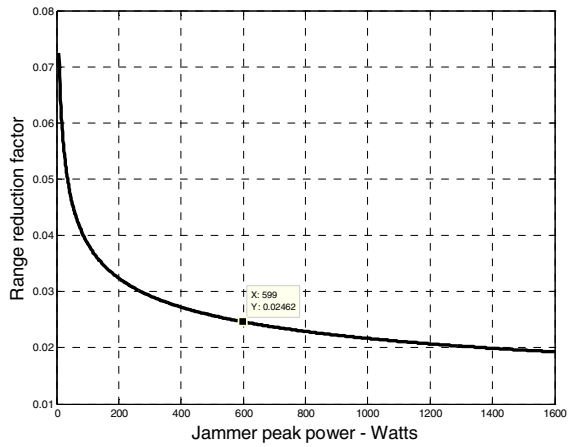
In this paper, the detection capability of radar in presence of jammer with different temperature has been compared. For different effective temperature of radar of a given jammer peak power, the range reduction factor is different. Simulation has been completed for 3 different temperatures of 300 K, 600 K, 900 K. The results have been shown in Figure 2.



(a)



(b)



(c)

Figure 2. Relationship betwwen RRF and jammer power (a) when temperature is 300K (b) temperature of 600 K (c) temperature of 900 K

Radar Effective Temperature In Kelvin	Range Reduction Factor (RRF) (Jammer peak power of 600 watt)
300 K	0.0187
600 K	0.0222
900 K	0.0246

Table 1: Value of RRF for various effective temperature of RADAR

From the above table it is clear that the effective temperature of radar has a great effect on range reduction factor in presence of jammer. When the effective temperature is lower, the range reduction factor (RRF) is lower. As a result the detection capability of radar is lower by equation (2). Increasing of temperature can improve the RRF as well as detection capability. Figure 2 shows when the effective temperature is minimum i.e. 300 K, the value of RRF is 0.0187. The value of RRF is improved to 0.0246 when the effective temperature is maximum i.e. 900 K. For all the above cases, jammer peak power is kept constant at 600 watt.

V. CONCLUSION

This paper presents a method for improvement of range reduction factor of radar in presence of jammer. Controlling the effective temperature play a very important role in this case. For a fixed jammer power the value of RRF will be improved if the effective temperature is increased. It is seen that the highest temperature can give the better value of RRF. As a result the detection range is higher though a jammer is present and performance of radar can be improved by this method. Adjustable higher effective temperature for implementation on radar may be a future work for this method.

VI. REFERENCES

[1] M. I. Skolnik, *Introduction to Radar Systems* (3rd ed.), New York: McGraw Hill, 2001, 339—369.

[2] Bassem R. Mahafza, *Radar Systems Analysis & Design using MATLAB* (2nd ed.), Chapman & Hall /CRC, 2004.

[3] S. Ariyavisitakul, N. Sollenberger, and L. Greenstein, *Introduction to Radar System*, Tata McGraw-Hill, 2001.

[4] Wang Qiang and Zhou Gu "The application of jamming analyzing technique in the radar system" Radar (Radar), 2011 IEEE CIE International Conference on; Volume: 2, Publication Year: 2011 , Page(s): 1063 – 1066.

[5] Schuerger,J. and Garmatyuk, D. "Deception jamming modeling in radar sensor networks" Military Communications Conference, 2008. MILCOM 2008. IEEE Publication Year: 2008 , Page(s): 1 – 7.

[6] Wu Shunjun ; Liao Guisheng and Zheng, E. "Advanced technology on radar signal processing" Radar, 2001 CIE International Conference on, Proceedings; Publication Year: 2001 , Page(s): 16 – 19.

Bangla License Plate Reader for Metropolitan Cities of Bangladesh Using Template Matching

Raiyan Abdul Baten[1]*, Zunaid Omair[2] and Urmita Sikder[3]
[1,2,3]Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka-1205, Bangladesh
*raiyanabdulbaten@gmail.com

***Abstract*—A simple algorithm for license plate reader for metropolitan cities in Bangladesh has been proposed in this paper. We have made use of a feature of Bangla script called "Matra" which joins the characters of a word together. Since the various words of a standard Bangladeshi license plate also have Matra, we have proposed a method where these words are segmented as single connected components and later recognized through template matching of words. There are also numbers and single letters in the number plate which are segmented and recognized as single characters. We have demonstrated that if we exploit the feature of Matra this way, the complexity of the algorithm is decreased, and the proposed system works smoothly for images taken under various conditions.**

***Index Terms*—license plate reader, template matching, connected component analysis, Bangla license plate, ANPR**

I. Introduction

Automatic Number Plate Recognition (ANPR) is an integral part of Intelligent Transport System (ITS). ANPR has become important due to its applicability in toll collection, crime prevention, traffic law implementation, travel-time measurements etc. In the metropolitan cities like Dhaka and Chittagong of Bangladesh, the importance of a fast and efficient algorithm in ANPR is explicit due to the increasing pressure of traffic.

The arrangement of different words and characters in the licence plates of metropolitan cities of Bangladesh, as regulated by BRTA (Bangladesh Road Transport Authority), was exploited in the algorithm to reduce complexity of the ANPR procedure. Again, a characteristic of Bangla script called "Matra" was also used to reduce steps of operations in the algorithm.

In this paper we approached ANPR from the standpoint of Binary Image Processing (BIP), primarily using the Connected Component Analysis (CCA) method, to segment entire words and single characters from the camera image after initial reduction of noise. Note that we could segment entire words as single connected components due to the presence of Matra. We used template matching technique, using correlation coefficient against a pre-recorded database, to recognize those segmented words and characters to provide the final output. Template matching has been named to be a suitable process for character recognition for single-font and not-rotated characters by [1] and [2]. The proposed algorithm is shown in Fig. 1.

This paper is organized as follows: section II discusses previous works done in the field. Section III gives a brief introduction to the various parts of a standard Bangladeshi license plate. Section IV describes the proposed method. Section V analyzes the results and compares them with previous works. Finally, we conclude the paper appreciating the previous researchers in the reference section.

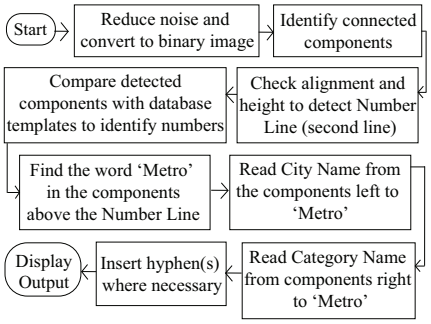


Fig. 1. Proposed Algorithm.

II. Related Works

Several methods have been proposed for Bangla ANPR, such as chain code and neural network [3], multi-layer feed forward network [4] and segmentation with adaptive threshold and artificial neural network [5]. A four layered back propagation neural network has been used in [6] and Tesseract based OCR has been employed in [7]. For license plates in other languages, [8] applied neural network for plate extraction and pattern matching for character recognition. A Gabor filter based technique has been used in [9]. Fuzzy logic for segmentation and feature based classification for character recognition has been applied by [10]. DTCNN has been shown to be applicable for license plate recognition in complex backgrounds in [11]. A multi-style recognition method has also been proposed by [12] using user-defined parameters. Connected component analysis in conjunction with probabilistic neural network has been used in [13]. Moreover, [14] has shown how correlation coefficients can be effectively exploited to dynamically vary the parameters for template matching. Word Image matching using Hausdorff distance has been shown to be effective for Chinese and English alphabets [15].

III. License Plate Properties in the Metropolitan Cities of Bangladesh

According to the regulations of BRTA, the registration number of license plates in the metropolitan cities has to be written in Bangla script painted on flat steel or aluminium plate of dimensions 524mm x 112mm for a car. The plates also must follow a specific colour code based on the type of the vehicle, the algorithm proposed here works independent of these color patterns.

The arrangement of different words and characters in a plate has a common pattern, as shown in Fig. 2:

- The first line contains the name of the vehicle registration area (which we referred to as "City Name"), the word

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

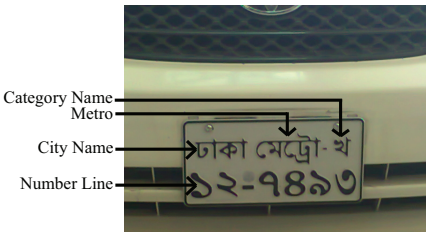


Fig. 2. A sample Bangla license plate.

"Metro" (a short form for "Metropolitan"), and after a hyphen a letter indicating vehicle category (which we referred to as "Category Name"). Here, the City Name could be "Dhaka" or "Chotto", Chotto being the short form of Chittagong. As "Category Name", thirty letters have been assigned which carry specific meanings.

- The second line (or "Number Line") contains six digits, the first two of which denote vehicle class number and the last four constitute unique vehicle registration number.

Interestingly, a characteristic of Bangla script called "Matra" comes into play here. Matra is a horizontal line at the upper part of many Bangla characters which join to make a continuous Matra when these characters combine to form a word. In the license plate, the word "Metro" and the City Name both have continuous Matra, which means if noise filtering does not damage the Matra, the words would appear as single connected components in the binary image. Therefore, we proposed detecting the word "Metro" and the City Name as whole words through template matching, which relieved the algorithm of the time consuming steps of splitting these words into characters for individual identification.

IV. Proposed System Architecture and Algorithm

The proposed system has three basic modules:

- Noise elimination and conversion to binary image.
- Segmentation of words and characters.
- Recognition of the segmented words and characters.

A. Noise Elimination and Conversion to Binary Image

The original image is resized to a height of 400 pixels with its aspect ratio unchanged, so that static filtering parameters could act upon the image successfully. The image is converted to grayscale and acted upon with median filtering to reduce small noises. Following that, a series of morphological operations, such as morphological erosion of the image, dilation of the same image and subtraction of the eroded and diluted versions result in the detection of the edges of words and single characters, as shown in Fig. 3. At this point, the characters and words in the image are contours of their original shapes.

We then perform the following 2D convolution operation to enhance the contrast of the image, resulting in Fig. 4:

$$P(x, y, z_{m+1}) = \int_{\delta D} W(x - x', y - y') P(x', y', z_m) dxdy \tag{1}$$

The image is then converted to binary and the contours of the characters and words are filled, which makes them solid again. The final binary image is shown in Fig. 5.

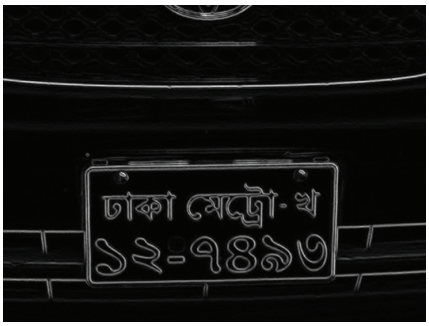


Fig. 3. Image after morphological operations.

Fig. 4. Image after convolution operation.

B. Segmentation of Characters and Words

Bounding boxes are formed around each of the connected components in the filtered binary image. Components with heights and widths below a threshold value of 10 pixels are discarded, which limits the algorithm's ability to detect too small characters but makes elimination of unwanted components easier.

We operate in two steps, the first of which reads the digits as single characters from the Second Line of the plate, and the second step reads the words and characters from the First Line.

To detect the 6 digits of the Second Line, we consider two criteria,

- The digits must be aligned,
- The digits must have similar heights.

A histogram is drawn with the upper-left corner ordinates of the connected components, as shown in Fig. 6. Components eliminated earlier do not appear in the histogram count. We take the average of ordinates of the components in the class with maximum elements. We add to that a small margin for error to determine a new range of acceptable ordinates

Fig. 5. Final binary image.

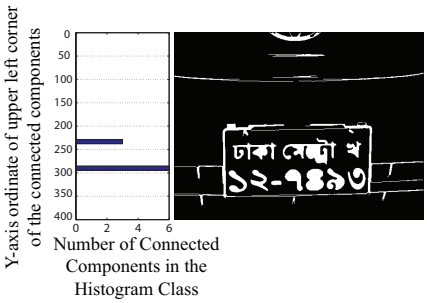


Fig. 6. Histogram showing distribution of connected objects.

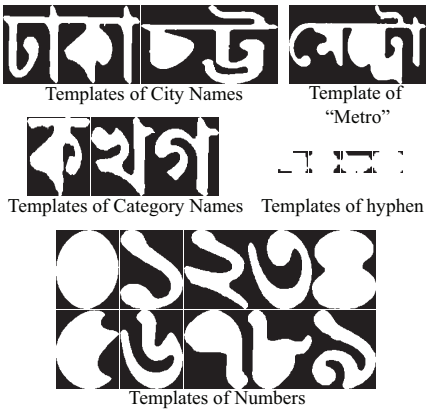


Fig. 7. Databases of City Names, Metro, Category Names, Hyphens and Numbers

and select obeying components to justify the first criterion of alignment. Then we check if these components indeed have same heights in order to justify the second criterion. This segmentation method is similar to the method proposed by [16]. We then send these selected components to the recognition module.

In the second step, we attempt to read the First Line of the plate. Here we consider the region directly above the number line and send the connected components within that region to the recognition module to locate the word "Metro". Once that has been identified, we return to the segmentation module to select the components on both sides of the word "Metro" whose upper left coordinates and heights are compatible to "Metro". The selected components to the left of "Metro" contain the City Name, and the ones to the right contain the Category Name. These segmented words and single characters are again sent to the recognition module.

C. Recognition of Segmented Characters and Words

We have formed separate and extensible databases of binary images, shown in Fig. 7, of the Bangla numbers 0-9; the words "Dhaka", "Chotto" as samples of City Name; the word "Metro"; and the letters "Ka", "Kha" and "Ga" as samples of Category Name.

From the first step of the segmentation module, segmented components of the second line are sent to the recognition module. This module first resizes the components to match the template image sizes, and then finds the 2D correlation coefficient of them against a database of the numbers 0-9, as per the following equation:

Dhaka Metro-Kha 12-7493

Fig. 8. Output for the sample in Fig. 2

$$r = \frac{\sum_m \sum_n (X_{mn} - \bar{X})(Y_{mn} - \bar{Y})}{\sqrt{(\sum_m \sum_n (X_{mn} - \bar{X})^2)(\sum_m \sum_n (Y_{mn} - \bar{Y})^2)}} \quad (2)$$

Here $\bar{X}$ and $\bar{Y}$ denotes the averages of the image matrices X_{mn} and Y_{mn} respectively. The digit which gives the highest value of r for a given image will be the digit represented by that component. This way, all the six digits from the second line are read.

From the second step of the segmentation module, the recognition module is called in two steps: first to recognize the word "Metro", then to recognize the city name and the category name. In all the cases, the module runs exactly like it did to recognize the digits, and uses appropriate database templates for recognizing different parts of the plate. The result is shown in Fig. 8.

V. Results and Analysis

The proposed system has been found to work very satisfactorily for samples that are not too noisy and not over-skewed. The results are shown in Fig. 8, Fig. 9, Fig. 10, Fig. 11, Fig. 12 and Fig.13, for images taken under various weather conditions and different spatial distances. The time performance, as measured on MATLAB r2014a software [20] running on Intel corei5 processor with 4.00GB of RAM is shown in Table. 1.

It is seen that the average recognition and execution time is comparable to generalized symmetry transform method proposed in [17] (1.3 seconds). The time and efficiency performance of the proposed system is comparable to other pattern matching algorithms, as in [18] and [19] (1.1 seconds). Moreover, it has been found that the system can work correctly for very low letter to image height ratio.

VI. Conclusions

In this paper, a method for Bangla license plate reader has been proposed, based on the segmentation algorithm for Albertan license plate proposed by [16], and it exploits features specific to the Bangla typography to reduce complexity. Template matching of single characters as well as of whole words has been applied in the connected components analysis. The database for templates can be easily extended to diversify the scope of the system. This method can deliver correct output within a reasonably short time, for images taken under various conditions.

Sample No.	Segmentation and Recognition Time(Seconds)
1(Fig.2)	0.5888
2(Fig.9)	0.6368
3(Fig.10)	0.6361
4(Fig.11)	0.6568
5(Fig.12)	0.6364
6(Fig.13)	0.6720

TABLE I
Performance analysis of the proposed system.

Output: Dhaka Metro-Ka 04-0052

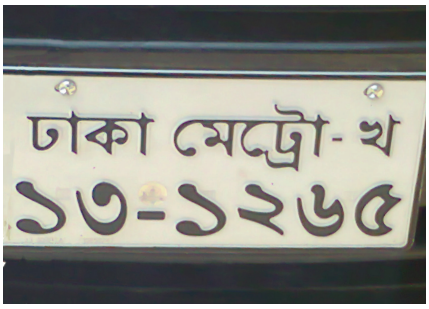
Fig. 9. A sample Bangla license plate, image taken from a distance.

Output: Dhaka Metro-Kha 11-8005

Fig. 10. A sample Bangla license plate, image taken in cloudy weather

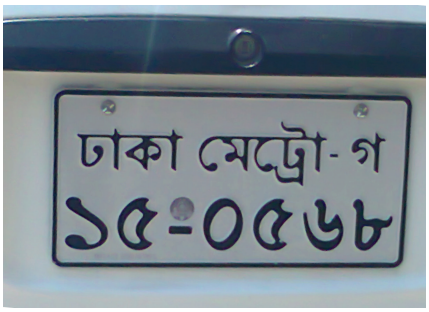
Output: Dhaka Metro-Ga 35-9425

Fig. 11. A sample Bangla license plate, image taken in sunny weather



Output: Dhaka Metro-Kha 13-1265

Fig. 12. A sample Bangla license plate, image taken from a close distance



Output: Dhaka Metro-Ga 15-0568

Fig. 13. A sample Bangla license plate, image slightly skewed

REFERENCES

[1] Christos-Nikolaos E. Anagnostopoulos, Ioannis E. Anagnostopoulos, Ioannis D. Psoroulas, Vassili Loumos and Eleftherios Kayafas, " License Plate Recognition From Still Images and Video Sequences: A Survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, Sep. 2008

[2] Shan du, Mahmoud Ibrahim, Mohamed Shehata and Wael Badawy"Automatic License Plate Recognition (ALPR): A State-of-the-Art Review," *IEEE Transactions on Circuits and Systems for Video Technology*,vol. 23, no. 2, Feb. 2013.

[3] Ashim Kumar Ghosh,Swapan K.D. Sharma, Md. Nazrul Islam, Suchitra Biswas and Sameya Akter, "Automatic License Plate Recognition(ALPR) for Bangladeshi Vehicles," *Global Journal of Computer Science and Technology*, vol. 11,Dec. 2011

[4] Md. Mahmudul Alam Joarder, Khaled Mahmud, Tasnuva Ahmed, Mohsina Kawser and Bulbul Ahmed,"Bangla Automatic Number Plate Recognition Using Artificial Neural Network,"*Asian Transactions on Science and technology*, vol. 02, Mar. 2012

[5] Md. Mahmudul Hasan, "Real Time Detection and Recognition of Vehicle License Plate in Bangla," M.Sc. thesis,Bangladesh University of Engineering and Technology, Dhaka, Bangladesh,Aug.2011.

[6] M. Shadab Mashuk, M. Arif Majid, Naimul Basher and Tarif Riyad Rahman, " Automatic Detection of Bangla Characters in Bangladeshi Car Registration Plates," in *Proc. Second International Conference on Computational Intelligence, Modelling and Simulation*, 2010.

[7] Md. Ruhul Amin, Noor Mohammad and Md. Abu Naser Bikas, " An Automatic Number Plate Recognition of Bangladeshi Vehicles," *International Journal of Computer Applications*, vol. 93, May 2014

[8] E.R. Lee, P.K. Kim, and H.J. Kim,"Automatic Recognition of a Car License Plate Using Color Image Processing," in *Proc. International Conference on Image Processing 94*, 1994.

[9] Tavsanoglu V and Saatci E,"Feature extraction for character recognition using Gabor-type filters implemented by cellular neural networks," in *Proc. 6th IEEE International Workshop on Cellular Neural Networks and their Applications, IEEE*, 2000, p.63-68.

[10] Nijhuis,J.A.G. and Brugge,M.H. ter and Helmholt, K.A., J.P.W. Pluim, L. Spaanenburg, R.S. and Venema, M.A. Westenberg," Car License Plate Recognition with Neural Networks and Fuzzy Logic."

[11] Brugge, M. H. ter and Stevens,J. H. and Nijhuis,J.A. G. and Spaanenburg,L., " License Plate Recognition Using DTCNNs ," in *Proc. Fifth IEEE lntemafiona Workshop on Cellular Neural Networks and their Applications, L.ondon, England*, 1998

[12] Jianbin Jiao, Qixiang Ye and Qingming Huang," A configurable method for multi-style license plate recognition,"*Pattern Recognition*, vol. 42,2009, p. 358-369.

[13] Christos Nikolaos E. Anagnostopoulos, Ioannis E. Anagnostopoulos, Vassili Loumos and Eleftherios Kayafas, "A License Plate-Recognition Algorithm for Intelligent Transportation System Applications,"*IEEE Transactions on Intelligent Transportation Systems*,vol. 7, no. 3, Sep. 2006.

[14] Tang Jin, Li Qing, "Fast Template matching algorithm", *Journal of Computer Applications*, no. 6, 2010.

[15] Yue Lu, Chew-Lim Tan, Weihua Huang, Liying Fan, "An approach to word image matching based on weighted Hausdorff distance," in *Proc. Sixth International Conference on Document Analysis and Recognition* , pp.921-925, 2001.

[16] Rahman, C.A. and Badawy, W. and Radmanesh, A., "A Real Time Vehicle's License Plate Recognition System," in *Proc. AVSS03*, 2003, p. 163-166.

[17] Dong-Su Kim and Sung-Il Chien," Automatic Car License Plate Extraction Using Modified Generalized Symmetry Transform and Image Warping," in *Proc. ISIE 2001, Pusan, KOREA* .

[18] Serkan Ozbay and Ergun Ercelebi ," Automatic Vehicle Identification by Plate Recognition," in *Proc. World Academy of Science, Engineering and Technology 9*, 2005.

[19] Paolo Comelli, Paolo Ferragina, Mario Notturno Granieri, and Flavio Stabil," Optical Recognition of Motor Vehicle License Plates," *IEEE Transactions on Vehicular Technology*, vol. 44, no. 4, Nov. 1995.

[20] (2014) The Mathworks website.[Online].Available: http://www.mathworks.com/products/matlab/

Complexity Measurement: A New Approach to Ensure Equal Distribution of Programming Problems for Evaluation

G. M. M. Bashir, [1,*] Samrat Kumar Dey, [2] S. S. M. Tariq, [2] and Md. Shariful Islam [2]
[1]Dept. of CCE, Patuakhali Science and Technology University
Patuakhali, Bangladesh
[2]Faculty of CSE, Patuakhali Science and Technology University
Patuakhali, Bangladesh
[*]murad98csekuet@yahoo.com

Abstract— **Distributing equivalent programming problems among the students for evaluation is an inherent problem. To remove this inherent problem, weight measurement of computer programs is necessary. Although software is the outcome of human creative activity, cognitive informatics plays an important role in understanding its fundamental characteristics. In our research, a new software complexity measure based on cognitive weight of basic control structure has been proposed that reduce the limitations of existing measures. Primarily our aim is to build a question bank with programming problems. Our designed system will access the solution codes for finding out the weight of respective programming problem from the question bank by applying the proposed method. The method will find out the complexity according to our proposed cognitive weight model. It will facilitate the instructors distributing the programming problems among the students equivalently. Thus, the automatic complexity measurement will ensure the students obtaining equal problems for programming evaluation**

***Keywords*—Cognitive Weights, Basic Control Structures, Weight Measurement, Software Complexity**

I. INTRODUCTION

An important issue encountered in computer science and engineering in the field of programming problem distribution is to distribute the problem equivalently. In the last several years, there has been a great deal of interest in defining appropriate ways to measure the complexity of software [1]. There are different facets of software complexity, some of which have been computed using widely accepted metrics like cyclomatic complexity, data/information flow metrics, but very less attempts have been made to measure the cognitive aspect of the complexity [2]. The human mind's efforts needed for the comprehension of the source code reflect a different dimension of complexity, which is being measured in this paper [2]. Another issue encountered in software complexity analysis is the consideration of software as a human creative artefact and the development of a suitable measure that recognizes this fundamental characteristic. The existing measures for software complexity can be classified into two categories: the macro and the micro measures of software complexity [3]. In this paper, we have proposed a new method for measuring software complexity, which is based on the thinking capacity of Human brain. We have used cognitive weight to measure the complexity value of a program. The complexity measures based on cognitive informatics is in developing phase. Knowing that the basic control structures (BCS) such as sequence, branch and iteration are the basic logic building blocks of any program. Here we have developed a simple method for calculating the complexity of programs in terms of cognitive weights and showed an approach for the equal distribution of the programming problems to the students during examination has been provided. Sometimes at the time of examination, programming problems may not be distributed equally to the student. In that case some students find that the problem is very easy and other students complain that the problem is very complex to solve within limited time. So it is necessary to develop a complexity model to find out complexity value that can be used to identify problem equivalence. There are ways to find out the running time of algorithms, which can be used for the equal distribution of programs by comparing two different algorithms. But, it is very difficult to measure especially when there is no defined algorithm or the algorithm is too much large and complex. In Materials and Methods section, we have conducted a survey with various programming problems; various Cognitive Weights for various Basic Control Structures have been defined. Implementation section provided with our designed algorithm which will help us to find out the complexity of programming solution code and finally in Result and Discussion section, an approach for the equal distribution of the programming problems to the students during examination has been provided.

II. METHODS AND MATERIALS

In the section of programing courses evaluation programming problems are need to be solved within a limited time are to be distributed among students .variant students will be tested with miscellaneous programming problems. According to this motive, programming problems may not be equally distributed. As a result, some can find their problems easy and others can get problems which are comparatively complex. In this context, examines will not get the proper pronouncement. But it should not happen, our aim of study is to distribute the problem equally. For this reason, to minimize these problems there should have a question bank which will

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

contain various programming problems and solution with different complexities. There are various ways to compute run time complexity of an algorithm. Different well-known algorithms have also defined complexity value based on their run time. Programming problems can be distributed consequently to these values. In a specified boundary similar and approximately similar problems will be put. Same methods will be revolved for all problems on the questions bank. Problems within a specific boundary will be deliberated as analogous kinds of problems. In this exploration, we have built a new complexity model for measure complexity values of programming problems based on their solution codes. In this case, after selecting a problem respective solution code should be developed.

A. Cognitive Weights

To perceive a given program, we naturally concentrate on other architecture and basic control structures (BCSs) of the software [4]. BCSs are a set of essential flow control mechanisms that are used for building logical software architectures [4]–[5]. Three BCS are commonly identified: the sequential, branch, and iteration structures [6]. Although it can be proven that an iteration maybe represented by the combination of sequential and branch structures, it is convenient to keep iteration as an independent BCS. In addition, two advanced BCSs in system modelling, known as recursion and parallel, have been described by Hoare et al. [6]. Wang [4]–[5], [4] extended the above set of BCSs to cover function call and interrupt. We have used 13 Basic Control Structures: SEQUENTIAL, IF-THEN-ELSE, SWITCH, BREAK, CONTINUE, RETURN, FOR, WHILE, DO-WHILE, USER DEFINED FUNCTION (UDF), RECURSION, EXCEPTION HANDLING, and REPEAT-UNTIL.

Definition: The cognitive weight of software is the measurement of difficulty or comparative time and effort required for understanding a given piece of software modelled by a number of BCSs [3].

B. Data Collection and Observation

For finding out the cognitive weight, we have selected some programs each of them represents one control structure .Every program contains nearly same number of lines. We have not used two control structures, PARALLEL and INTERRUPT individually in this observation. Because, a program that describes this control structure contains other control structures used in the observation. During the survey, we gave 20 programs containing individual BCSs to different class of students at undergraduate level. After reading the problems, every student gave the result of the selected programs. They gave the result according to their thinking capacity of brain for individual BCS. All 20 students consider the initial fact of sequential statement containing weight 1.By observing the weight of sequential statement our selected 20 students gave the other BCS respective weight. These weight will help us to find out the average weight of individual BCS and later we will use this to find the complexity of any program. Maximum number of problems is taken from C for contest by Tamnun E Mursalin. We have chosen some common programs. Reason for choosing such common program for making problems relatively in a similar level (simple or complex). Following is a sample of a sequential statement program which we have used in our survey.

```
#include<stdio.h>
int main (void) {
int first,second,result;
printf("Enters two numbers: ");
scanf("%d %d",&first,&second);
result= first+second;
printf ("The two numbers are:%d %d\n ",first, second);
printf ("The Result is: %d", result);}
```

Survey results of Level 1 and 3 students are shown in Fig 1. Survey results of Level 2 and 4 students are shown in Fig 2. The results which we found by combining Fig 1 and Fig 2 data have been used to obtain the Cognitive Weight.

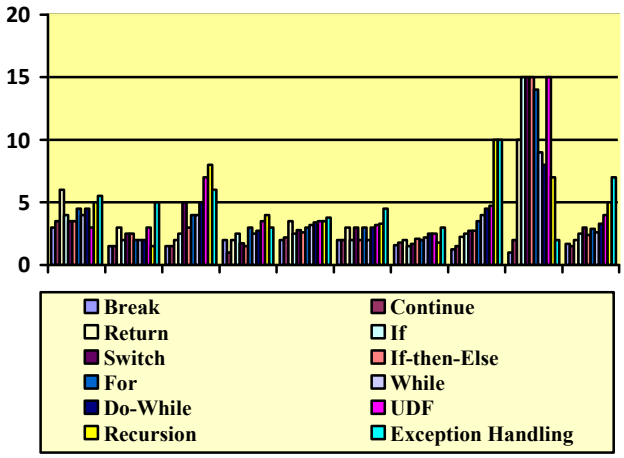


Fig 1: Graphical representation of Survey results of Level 1 and 3

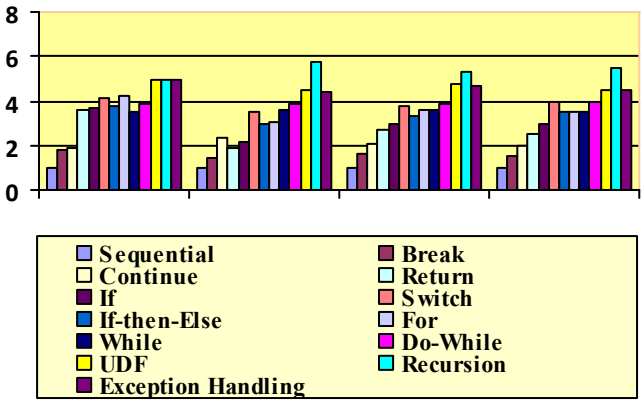


Fig 2: Graphical representation of Survey results of Level 2 and 4

We have divided the results in three Categories: a) When we get any value, x which satisfies 2 <= x <= 2.25 such as 2.24 then we consider this value as 2.b) When we get any value, y as 2.25 < y < 2.75 such as 2.41 then we consider it as 2.5 and c) When we get any value, z as 2.75 <= z <= 3.00 such as 2.76 then we consider it as 3. We have made a chart showing the comparison of our cognitive weight with others. In the chart, Cw means Cognitive Weights and Cw MSST represents our Cognitive Weight. Fig 3 shows how various cognitive weight of BCS varies with others developed cognitive weight.

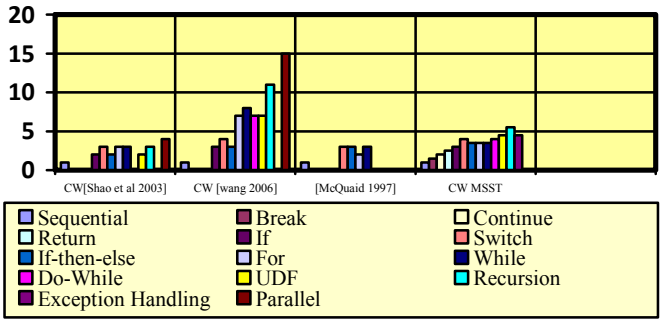


Fig 3: Comparison of Our Cognitive Weights with Others

It's quite clear that in comparison with others our survey result is unique and it do not interact with others measured value. In further work these values will help us to find out the complexity value of programs.

C. Complexity Model

In this section, we will show how our measured cognitive weights can be used to find out total weight or complexity, W_{bcs} of a given program. There are two different architectures for calculating Wbcs

- Either all the BCS's are in a linear layout. For this case, sum of the weights of all n BCS's are added

$$W_{linear}=\sum_{q=1}^{j} W_C(q) \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Here, q = 1, 2,…,j

j = Number of linear BCSs

For example, we can consider the following Java code. Numbers in the right hand side indicates the weights of the respective BCSs.

```
1. public class SST{
2. public static void main (String args []){
 a.    Integer value_1 = 20;          1 }
 b.    Integer value_2 =10;           1 }
 c.    value_1 = value_1+ 5;          1 }      += 4
 d.System.out.println (value_1);      1
3.          }
4.          }
```

In the above program, line number 1 and 2 are common in Java code. Within the brackets, there are four (a-d) linear sequential statements. They are all independent i.e. they are not embedded in other BCSs. From the previous section, we have obtained the weights of various control structures, where the weight for sequential statements is 1. Therefore, total weight of this program will be, W_{linear}=1+1+1+1=4.

- Or some BCS's are embedded in others. For this case, cognitive weights of inner BCS's are multiplied with the weights of external BCS's.

$$W_{embedded}=\prod_{i-n}^{1} W_C(i) \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

Here, i = n,……., 2, 1

n = Total number of BCSs in which the current BCS is embedded + 1

1 = Own weight

2 = Respective parent BCS of 1

3 = Respective parent BCS of 2

……..

n = Respective parent BCS of (n-1)

Parent Control Structure means the structure which encloses another control structure. For example, we can consider the following Java code:

```
1. public class SSTArray{
2. public static void main(String[]args){
A. double[]myList={2.4,4.2,5.3,6.1};        1 (W_linear)
B. for(int i=0;i<myList.length;i++){        3 (W_linear)
    a. System.out.println(myList[i]+" ");   3*1=3 (W_embedded)
C. }
D. double total =0;                         1 (W_linear)
E. for(int i=0;i<myList.length;i++){        3 (W_linear)
    b. total +=myList[i];                   3*1=3 (W_embedded)
F. }
G. System.out.println("Total is "+ total);  1 (W_linear)
H. double max =myList[0];                   1 (W_linear)
I. for(int i=1;i<myList.length;i++){        3.5 (W_linear)
    c. if(myList[i]> max)                   3*2=6 (W_embedded)
    d.  max =myList[i];                     3*2*1=6 W_embedded)
J. }
K. System.out.println("Max is "+ max);      1 (W_linear)
3. }      }
```

Total Weight, W_{bcs} = 32.00

III. IMPLEMENTATION

In this section we will show how our designed system scan the programs and then give the weight of the programs. We have built our own method to measure the weight of programming problems. Implementing this task in real life is not so much easy. It's a hard task and we have done it by our own developed program. Our initial step is to measure the complexity of programs which not contain any embedded structure. Embedded structure means one BCS contain another BCS inside his structure. We are quite successful to do this by the help of our survey result which represent the cognitive weight of the respective BCS. Our goal is to access the database which is basically a question bank containing programs and then pick the set of problems from database for measuring the complexity of respective problems. Following figure showing how our designed system works in real life from set of input problems to complexity value of problems as output.

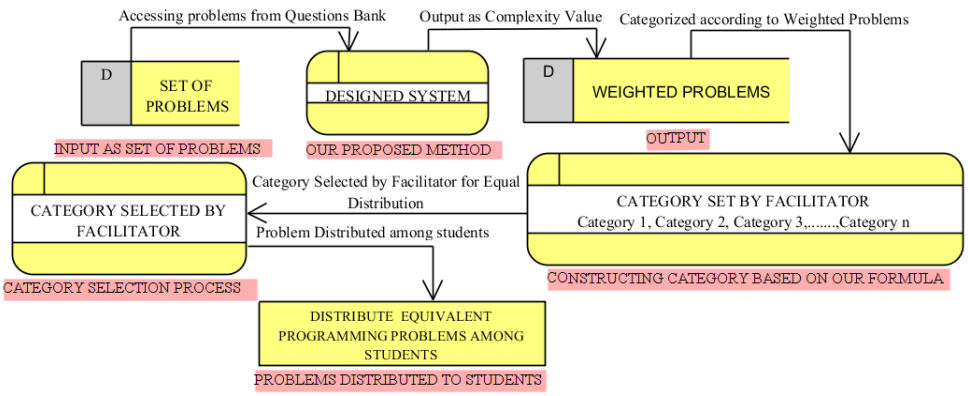


Fig 4: The Proposed System for Distributing Equivalent problems

We have built our own algorithm to find out the complexity value of programs. The following snippet of pseudo code of our program will tell how our system works for finding programs complexity values.

Input: A set of programs
Output: Calculated complexity of input programs
For each input character
IF
(Input stream==for)
THEN
Call class for ()
Complexity→ for-check ()
Index → return index

Pseudo code explained here is just only a part of our developed program and showing a few basic steps of operation.

IV. Result and Discussion

After the measurement of the complexity of some problems, we have concluded the result. We have got the complexity value of some problems. From the observation of the complexity values, we can say that, 1. For programs with increasing number of lines, the complexity increases gradually (without any exceptions). 2. If the number of linear BCSs is the same or approximately same, but one program contains BCSs of relatively higher weight, then this program will result in comparatively higher complexity value. 3. If the number of lines is the same, but one program contains BCSs which are more embedded, then this program will definitely result in high complexity value.

A. Distribution of Problems Based on Measured Complexity Value

For distributing similar types of problems to the students in the examination we have categorized the programs complexity according following formula where

L=Lowest Weight from a set of problems
H= Highest Weight from a set of problems
D=Difference between H & L
N=Number of category Facilitator Want
V=D/N
Category 1=L + V
Category 2=category 1+V
Category 3=category 2+V
……
Category N=category (N-1) +V

As a part of our thinking, now Facilitator can easily choose the category of problems according to complexity levels and can distributes the problems to the students equally.

B. Future Works

At this moment we are capable of finding out the complexity of programs only for linear structure with the help of our own designed system. Still we are working on this to improve the efficiency of our designed systems we will include the BCSs PARALLEL and INTERRUPT into our designed program in our later work.

V. Conclusion

We have designed a method and an algorithm to measure complexity of programming problems based on cognitive weights of Basic Control Structures. It is certain that this method will help a facilitator to distribute programming problems equally among the learners. It is found that measuring complexity based on cognitive weight is the most suitable measure, comparing with other existing measures, as problems are distributed to human not to the computer.

References

[1] E. J. Weyuker, "Evaluating software complexity measure," IEEE Transaction on Software Engineering, Vol. 14(9): 1357-1365,september 1988.
[2] J. K. Chhabra, "Code Cognitive Complexity: A New Measure," Proceedings of the World Congress on Engineering 2011, Vol II WCE 2011, July 6 - 8, 2011, London, U.K.
[3] J. Shao, and Y. Wang, " A new measure of software complexity based on cognitive weights," IEEE Canadian Journal of Electrical and Computer Engineering, 28(2):69–74, 2003.
[4] Y. Wang, "On cognitive informatics: Keynote lecture," In Proc. 1st IEEE Int. Conf. Cognitive Informatics (ICCI'02), Calgary, Alta., Aug. 2002, pp. 34–42.
[5] "Component-based software measurement," chap. 14 in Business Component-Based Software Engineering, ed. F. Barbier, Boston: Kluwer Academic Publishers, 2002, pp. 247–262.
[6] C. A. R. Hoare, I. J. Hayes, J. He, C.C. Morgan, A. W. Roscoe, J. W. Sanders, I. H. Sorensen, J. M. Spivey, and B. A. Sufrin, "Laws of programming," Comm. ACM, vol. 30, no. 8, Aug. 1987, pp. 672–686.
[7] McQuaid, P. A.*October* 1997. The profile metric and software quality. *International Conference on Software Quality, Montgomery*, pages 245–252.

State Space Least Mean Fourth Algorithm

Arif Ahmed,[1,*] Muhammad Moinuddin,[2] and Ubaid M. Al-Saggaf[3]
[1,2,3]Centre of Excellence in Intelligent Engineering Systems
King Abdulaziz University, Jeddah, Saudi Arabia
[1,2,3] Department of Electrical & Computer Engineering
King Abdulaziz University, Jeddah, Saudi Arabia
*arif545@gmail.com

***Abstract*—Adaptive filters generally employed for estimation purposes require high computational power when it comes to real time estimation. Therefore, in this paper we propose a computationally light yet effective estimation algorithm based on state space model. Our algorithm has been employed successfully in linear and non linear state space model based estimation problems. We investigate few examples to demonstrate the novelty of our algorithm by comparison with few existing algorithms in presence of non Gaussian noise namely uniform noise. More specifically, the state space normalized least mean squares and the Kalman filter has been compared with our algorithm.**

***Index Terms*—SSLMF, SSNLMS, State Space Least Mean Fourth, State Estimation Algorithm**

I. Introduction

Advancements in computational speed, complexity and power efficiency of digital processors have assisted in adaptive filters gaining widespread acceptance and implementations in numerous fields. Adaptive filters provide an upper hand in comparison to conventional filters because of their ability to adapt by self adjusting filter parameters according to the optimization algorithm utilized ([1],[2]). Two of the most widely used adaptive filtering algorithms are the least mean squares (LMS) algorithm ([1],[2],[3]) and the recursive least squares (RLS) algorithm ([1],[2],[3]). The state space (SS) version of these algorithms have been developed and presented by Malik et al. ([4],[5],[6],[7]) and different analyses were presented.

Adaptive filters dealing with SS model of a system yields near true results compared to non model based systems because of the availability of prior knowledge of the states via the system model. Numerous adaptive filtering algorithms can be found in the literature addressing the problem of SS based models for example the very well known Kalman filter (KF) [8], which gives the linear optimal solution by calculating the Minimum Mean Square Error (MMSE). Using observations subject to noise and other disturbances, it optimally estimates the system parameters. In the domain of non linear filtering, we have the extended Kalman filter (EKF) [8], unscented Kalman filter (UKF) [9], cubature Kalman filter (CKF) [10], quadrature Kalman filter (QKF) [11] and many other variants of KF [8]. These non linear filtering algorithms are computationally computationally complex and therefore pose difficulties in real time filtering problems due to high computational requirement. In this paper we propose and investigate the state space least mean fourth (SSLMF) algorithm by defining a cost function $\mathbf{J}[k]$ and minimizing it. We also compare our performance with the existing state space normalized least mean square (SSNLMS) algorithm [4] and KF algorithm [8]. Our proposed algorithm is superior due to the fact that it requires lesser computations per iteration compared to the existing linear and non linear model based algorithms (KF, UKF, EKF, CKF, QKF, etc.). The performance of the algorithm in presence of non Gaussian noise namely uniform noise is investigated due to the fact that least mean fourth (LMF) algorithm performs better in non Gaussian noise environments [12].

The paper is organized as such, Section II of the paper introduces the State-Space model. Then in Section III the overview of the existing SSLNMS is presented following Section IV where the SSLMF algorithm is developed. Section V presents simulation and results of the comparison of and proposed algorithm keeping in view the mean square errors. And finally we conclude the paper in Section VI.

II. State-Space Model

We begin by defining the general state-space model of a linear time varying system.

$$\mathbf{x}[k+1] = \mathbf{A}[k]\mathbf{x}[k] + \mathbf{B}[k]\mathbf{u}[k] + \mathbf{w}[k], \tag{1a}$$
$$\mathbf{y}[k] = \mathbf{C}[k]\mathbf{x}[k] + \mathbf{D}[k]\mathbf{u}[k] + \mathbf{v}[k] \tag{1b}$$

where $\mathbf{x} \in \Re^n$ are the process states, $\mathbf{y} \in \Re^m$ are the measured outputs such that $m \leq n$. $\mathbf{A}[k]$ is the state transition matrix, $\mathbf{B}[k]$ is the input matrix, $\mathbf{u}[k]$ is the input vector where $\mathbf{u} \in \Re^p$, $\mathbf{w} \in \Re^n$ is the process noise vector and $\mathbf{v} \in \Re^m$ is the measurement noise vector. The matrix $\mathbf{C}[k]$ is the output matrix where $\mathbf{dim}[\mathbf{C}[k]] = m \times n$, $\mathbf{D}[k]$ is the feed through matrix with $\mathbf{dim}[\mathbf{D}[k]] = m \times p$. It is assumed that the above system is observable. A special case is the unforced (autonomous) linear time varying system, represented as

$$\mathbf{x}[k+1] = \mathbf{A}[k]\mathbf{x}[k] + \mathbf{w}[k], \tag{2a}$$
$$\mathbf{y}[k] = \mathbf{C}[k]\mathbf{x}[k] + \mathbf{v}[k] \tag{2b}$$

The state space representation for a non linear continuous time system is

$$\dot{\mathbf{x}} = f(\mathbf{x}, \mathbf{u}, \mathbf{w}), \tag{3a}$$
$$\mathbf{y} = h(\mathbf{x}, \mathbf{u}, \mathbf{v}) \tag{3b}$$

where f and h are non linear functions and the parameters are as defined before.

III. Overview of the existing SSNLMS

Considering the system described by equation (2). A model based adaptive estimation process can be divided into the

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

following two steps. Step 1, the time update which is given by

$$\bar{\mathbf{x}}[k] = \mathbf{A}[k-1]\hat{\mathbf{x}}[k-1] \tag{4}$$

Step 2, the measurement update which is given by

$$\hat{\mathbf{x}}[k] = \bar{\mathbf{x}}[k] + \mathbf{K}[k]\varepsilon[k] \tag{5}$$

where

$$\varepsilon[k] = \mathbf{y}[k] - \bar{\mathbf{y}}[k] \tag{6}$$

here $\mathbf{y}[k]$ is as mentioned in (2), **K**[k] is the gain matrix and

$$\bar{\mathbf{y}}[k] = \mathbf{C}[k]\bar{\mathbf{x}}[k] \tag{7}$$

This is basically the structure employed in all Kalman Filtering techniques ([8],[9],[10],[11], etc.). In the case of SSNLMS [4] the gain matrix $\mathbf{K}[k]$ is derived as follows

$$\mathbf{e}[k] = \mathbf{y}[k] - \hat{\mathbf{y}}[k] = \varepsilon[k] - \mathbf{C}[k]\delta[k] \tag{8}$$

where

$$\delta[k] = \hat{\mathbf{x}}[k] - \bar{\mathbf{x}}[k] \tag{9}$$

Assuming $\mathbf{C}[k]$ is full rank, $\hat{\mathbf{x}}[k]$ is chosen such that $\mathbf{e}[k] = 0$ which implies the following

$$\varepsilon[k] = \mathbf{C}[k]\delta[k] \tag{10}$$

$\delta[k]$ is chosen as the minimum norm solution of the above equation which results in

$$\delta[k] = \mathbf{C}^T[k](\mathbf{C}[k]\mathbf{C}^T[k])^{-1}\varepsilon[k] \tag{11}$$

here

$$\mathbf{K}[k] = \mathbf{C}^T[k](\mathbf{C}[k]\mathbf{C}^T[k])^{-1} \tag{12}$$

The estimator equation from (9) and (11) is hence derived as

$$\hat{\mathbf{x}}[k] = \bar{\mathbf{x}}[k] + \mathbf{C}^T[k](\mathbf{C}[k]\mathbf{C}^T[k])^{-1}\varepsilon[k] \tag{13}$$

This is termed in ([4],[5]) as the SSNLMS algorithm. However, to avoid the non invertible scenario of $(\mathbf{C}[k]\mathbf{C}^T[k])^{-1}$, a small value γ might be added to the term and hence the estimator can be represented as

$$\hat{\mathbf{x}}[k] = \bar{\mathbf{x}}[k] + \mathbf{C}^T[k](\gamma + \mathbf{C}[k]\mathbf{C}^T[k])^{-1}\varepsilon[k] \tag{14}$$

The above estimator may be unstable and there are no known conditions for stability and hence convergence is not guaranteed. Therefore, a matrix **G** has been introduced in ([4],[5]) to overcome this problem. In ([4],[5]), it is claimed that the choice of this matrix **G** depends on the nature of the problem and one simple approach is to take **G** to be all zeros except for the first column to have non zero entries. In our investigation, we will consider the SSNLMS algorithm ([4],[5]) defined as

$$\hat{\mathbf{x}}[k] = \bar{\mathbf{x}}[k] + \mu\mathbf{G}\mathbf{C}^T[k](\gamma + \mathbf{C}[k]\mathbf{C}^T[k])^{-1}\varepsilon[k] \tag{15}$$

μ is the step size parameter for quicker achievement of convergence.

IV. Derivation of SSLMF

Considering the class of adaptive filtering based on (5) which is employed in all KF techniques, this type of update equation can be generalized with the following update rule

$$\hat{\mathbf{x}}[k] = \bar{\mathbf{x}}[k] - \mu\nabla\mathbf{J}[k] \tag{16}$$

where $\mathbf{J}[k]$ is the cost function to be minimized, $\nabla\mathbf{J}[k]$ is the gradient and μ is the step size parameter. To derive the SSLMF, we start by defining the cost function as

$$\mathbf{J}[k] = \mathbf{E}\left[\|\varepsilon[k]\|^4\right] \tag{17}$$

Minimizing the cost function $\mathbf{J}[k]$ with respect to the predicted states $\bar{\mathbf{x}}[k]$ result in

$$\nabla\mathbf{J}[k] = -4\|\varepsilon[k]\|^2\mathbf{C}^T[k]\varepsilon[k] \tag{18}$$

Equation (16) can therefore be written as

$$\hat{\mathbf{x}}[k] = \bar{\mathbf{x}}[k] + \mu\mathbf{G}\|\varepsilon[k]\|^2\mathbf{C}^T[k]\varepsilon[k] \tag{19}$$

which is our general estimator algorithm.Comparing equations (5) and (19) yields

$$\mathbf{K}[k] = \mu\mathbf{G}\|\varepsilon[k]\|^2\mathbf{C}^T[k] \tag{20}$$

Matrix **G** was imposed for the condition of controllability ([4],[5]) which is required due to the dynamics of the system where the algorithm is being applied.

V. Simulation Results & Discussion

Simulation results are presented here to validate the performance of the algorithm. We perform the estimation of the state parameters of a noisy sinusoid tracking problem, a real time trajectory tracking problem and the state parameters of a Van der poll oscillator ([4],[13]). These problems were investigated in the presence of uniform noise. The process and observation error covariance matrices were chosen to be similar to the noise covariances. An overview of the root mean square error (RMSE) of the output observations can be referred to in Table I. SSNLMS, KF, and SSLMF were compared. 100 simulations were performed while the estimation algorithms ran simultaneously.

A. Example 1. Tracking Sinusoids

In the first example we consider the system reported in ([4],[5]). More specifically we investigate a second order transversal filter with known frequency and unknown phase and amplitude of sinusoids which produces a 4th order system given by

$$\mathbf{A}[k] = \mathbf{diag}\left\{\begin{bmatrix}\cos(\omega_i T) & \sin(\omega_i T)\\ -\sin(\omega_i T) & \cos(\omega_i T)\end{bmatrix}\right\},\ i = 1,2 \tag{21a}$$

$$\mathbf{B}[k] = \begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix}^T \tag{21b}$$

$$\mathbf{C}[k] = \begin{bmatrix}1 & 0 & 1 & 0\end{bmatrix} \tag{21c}$$

The ω_i's for this system are known and kept constant. For the purpose of our study, we have set the values of the frequencies ω_1, ω_2 to 0.5 and 0.25 respectively. The sampling time is considered as $T = 0.1s$ and μ for SSNLMS was taken as 0.05 whereas, μ for SSLMF was chosen as 300. The

observation is subject to uniform noise of covariance $\sigma^2 = 0.01^2$. The process noise was considered to be of covariance $\sigma^2 = 0.001^2$. The actual initial system states were considered to be $\mathbf{x}[0] = [0.1\ 0.1\ 0.1\ 0.1]^T$, and the initial estimate for tracking were chosen to be $\hat{\mathbf{x}}[0] = [0.15\ 0.2\ 0.05\ 0.16]^T$. The plot of observation and the mean square observation error are presented in Figure 1 and Figure 2 respectively. It is observed from Figure 1 that SSLMF performs better in terms of observation. A closer look at Figure 2 reveals that SSLMF maintains a lower level of observation error, while KF has the lowest. However, KF is computationally complex and thus state space least mean algorithms have an advantage due to their light computational requirement. Table I shows that the root mean square observation error for SSLMF is lower than that of SSNLMS and hence performed better than SSNLMS.

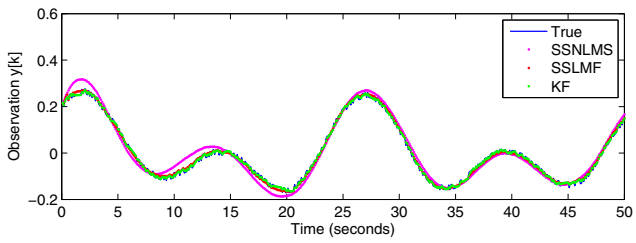


Fig. 1: Observation y[k].

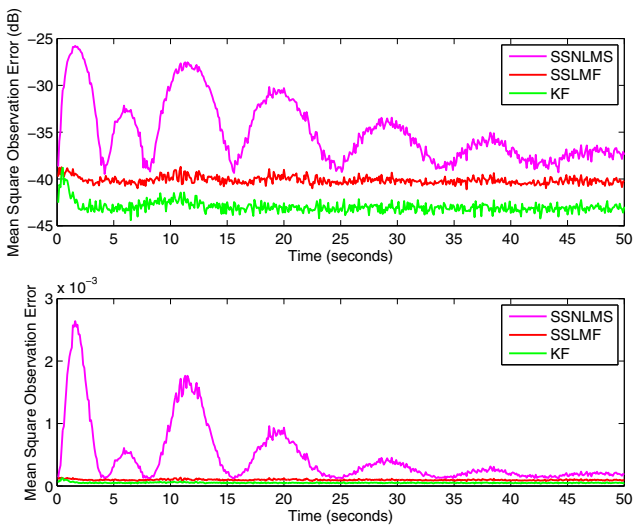


Fig. 2: Mean Square Observation Error y[k]

B. Example 2. Real time tracking

Considering the example of real time tracking ([8],[5]), the state space representation is as follows

$$\mathbf{x}[k+1] = \begin{bmatrix} 1 & T & T^2/2 \\ 0 & 1 & T \\ 0 & 0 & 1 \end{bmatrix} \mathbf{x}[k] \tag{22a}$$

$$y[k] = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \mathbf{x}[k] \tag{22b}$$

The sampling time T is taken as 0.01s. The initial true state was considered as $\mathbf{x}[0] = [0.4\ 0.3\ 0.4]^T$ and the initial state for the estimation algorithms were chosen as $\hat{\mathbf{x}}[0] = [0.2\ 0.1\ 0.4]^T$. The observation was subject to uniform noise of covariance $\sigma^2 = 0.1^2$ and the process noise was considered to be of covariance $\sigma^2 = 0.01^2$. Figure 3 presents the observation plot while Figure 4 presents the mean square observation error. It was observed that both algorithms effectively estimate the states. However, SSLMF converges fast than SSNLMS which can be observed in 4. Referring to I, it can be confirmed that SSLMF performed better than SSNLMS.

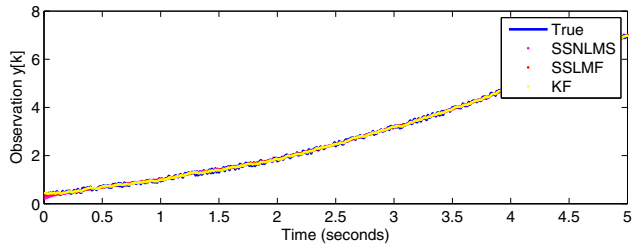


Fig. 3: Observation y[k]

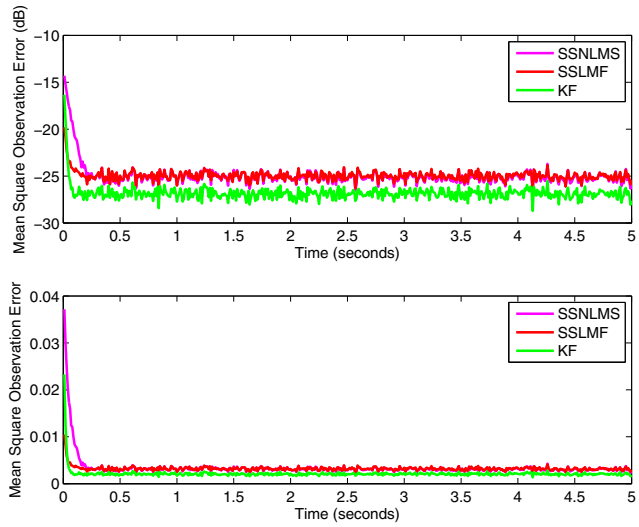


Fig. 4: Mean Square Observation Error y[k]

C. Example 3. Van der poll oscillator

In this example we consider the commonly used Van der poll oscillator which is a highly nonlinear system exhibiting both stable and unstable limit cycles [14]. We consider the case of stable limit cycle therefore, as time proceeds the system states converge to zero. The system is represented by the following differential equations ([13],[14])

$$\dot{x}_1 = -x_2 \tag{23a}$$

$$\dot{x}_2 = x_1 - \alpha(1 - {x_1}^2)x_2 \tag{23b}$$

here $\alpha = 0.2$ and the system has been discretized with a sampling time of 0.01s. The state space representation is as follows

$$\mathbf{x}'(t) = \begin{bmatrix} 0 & -1 \\ 1 & -\alpha(1 - x_1(t)^2) \end{bmatrix} \mathbf{x}(t) \tag{24a}$$

$$\mathbf{y}(t) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{x}(t) \tag{24b}$$

μ for the SSNLMS was chosen as 0.1 and μ for SSLMF as 0.7. The observation was subject to uniform noise of covariance $\sigma^2 = 0.01^2$ and the process noise was considered to be of covariance $\sigma^2 = 0.001^2$. The system was simulated with initial system states to be $\mathbf{x}[0] = [1.4\ 0]^T$, and the initial estimate for the algorithms were chosen to be $\hat{\mathbf{x}}[0] = [1.3\ 0.2]^T$. The observations and the mean square observation error is presented in Figure 5 and Figure 6 respectively. It can be observed from I that the observation error remains lower

for SSLMF in comparison to SSNLMS. Moreover, SSLMF converges faster compared to SSNLMS.

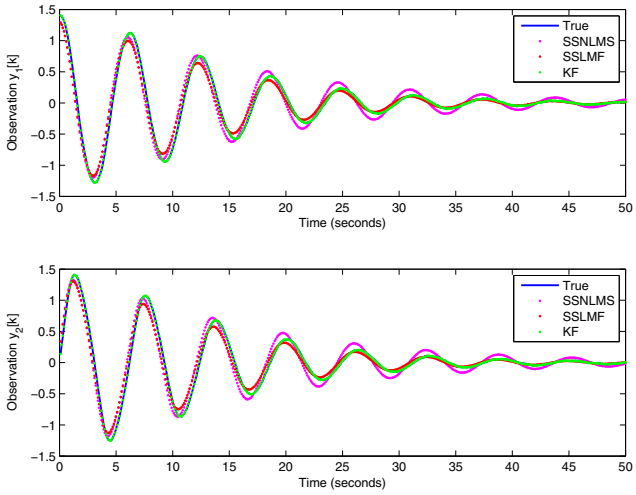


Fig. 5: Observation $y_1[k]$ and $y_2[k]$.

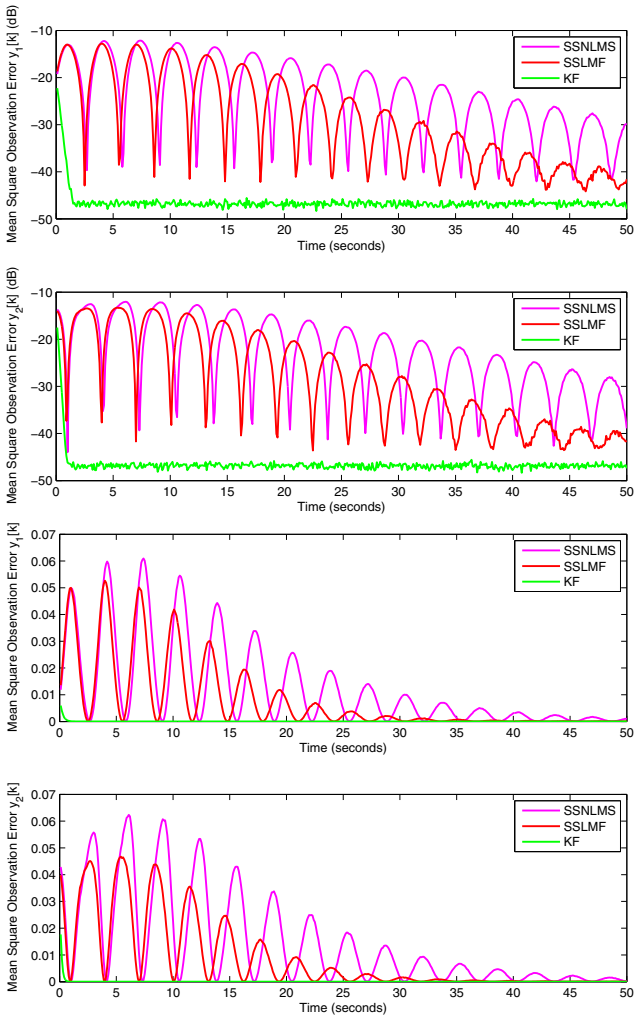


Fig. 6: Mean Square Observation Error for $y_1[k]$ and $y_2[k]$.

It is clear from the study that SSLMF indeed is a novel technique for estimation. Moreover, it is computationally very light when compared to other model based techniques like KF, EKF, UKF, etc. Our study reveals that SSLMF can be effectively utilized in non Gaussian noise environments.

TABLE I: Root Mean Square Error

		Root Mean Square Error		
		SSNLMS	SSLMF	KF
Example 1	Observation y	0.0215	0.0099	0.0071
Example 2	Observation y	0.0587	0.0566	0.0462
Example 3	Observation y_1	0.1108	0.0881	0.0073
	Observation y_2	0.1174	0.0911	0.0093

VI. Conclusion

In this paper the novel SSLMF algorithm has been proposed and investigated. The proposed algorithm is efficient and computationally light compared to existing model based linear and non linear estimation techniques (KF, EKF, UKF, etc.). Three different examples were investigated where SSLMF performed better than the SSNLMS and showed acceptable performance in comparison to KF keeping in mind the fact that KF is highly intensive in computation. For further studies SSLMF for Gussian noise environments can be proposed and investigated along with time varying step size parameters or hybrid algorithms for yet better results.

Acknowledgment

The authors acknowledge the support of King Abdulaziz University and Centre of Excellence in Intelligent Engineering Systems, King Abdulaziz University for carrying out this work.

References

[1] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, 3rd Edition, Upper-Saddle River, NJ, 1996.
[2] A. H. Sayed, *Fundamentals of Adaptive Filtering*, New York: Wiley-Interscience, 2003.
[3] B. Widrow, J. M. McCool, M. G. Larimore, and C. R. Johnson, “Stationary and Nonstationary Learning Characteristics of the LMS Adaptive Filter,” *IEEE Proceedings*, Vol. 64, pp. 1151–1162, 1976.
[4] M.B. Malik, M. Salman, ”State-space least mean square”, Digital Signal Processing, Vol. 18, Issue 3, pp. 334-345, May 2008.
[5] Malik, M.B, Bhatti, R.A, ”Tracking of linear time-varying systems using state-space least mean square,” IEEE International Symposium on Communications and Information Technology, Vol. 1, pp. 582-585, Oct. 2004.
[6] Mohammad Bilal Malik ”State-Space Recursive Least Squares: part I,” Signal Processing, vol. 84/9, 2004, pp 1709-1718.
[7] Mohammad Bilal Malik ”State-Space Recursive Least Squares: part II,” Signal Processing, vol. 84/9, 2004, pp 1719-1728.
[8] C. K. Chui and G. Chen, *Kalman Filtering With Real-Time Applications*, Springer, 4th Edition, 2009.
[9] Wan, E. A., and van der Merwe, R., ”The unscented Kalman filter for nonlinear estimation,” IEEE Proceedings of Symposium on Adaptive Systems for Signal Processing, Communication and Control, Lake Louise, Alberta, Canada, Oct. 2000.
[10] Arasaratnam, I., Haykin, Simon ”Cubature Kalman Filters”, Automatic Control, IEEE Transactions on, On page(s): 1254 - 1269 Volume: 54, Issue: 6, June 2009.
[11] I. Arasaratnam , S. Haykin and R. J. Elliott ”Discrete-time nonlinear filteringalgorithms using Gauss-Hermite quadrature”, Proc.IEEE, vol. 95, no. 5, pp.953 - 977, 2007.
[12] E. Walach and B. Widrow, “The least mean fourth (LMF) adaptive algorithm and its family,” IEEE Trans. Inf. Theory , vol. IT-30, pp. 275283, Feb. 1984.
[13] Rambabu Kandepu, Bjarne Foss, Lars Imsland, “Applying the unscented Kalman filter for nonlinear state estimation,” Journal of Process Control, vol. 18, issues 7-8, pp. 753-768, Sep. 2008.
[14] H.K. Khalil, Nonlinear Systems Third Edition, Prentice Hall, 2000.

Particle filter based attitude matching algorithm for in–flight transfer alignment

Suvendu Chattaraj,[1*] and Abhik Mukherjee,[1†]
[1]Department of Computer Science and Technology, IIEST, Shibpur (Formerly BESU, Shibpur)
PO: Botanic Garden, Dist: Howrah, West Bengal, India - 711103
*suvendu.chattaraj@gmail.com
†abhik@cs.becs.ac.in

***Abstract*—Attitude plus velocity matching transfer alignment (TA) (Rapid Alignment Prototype RAP) algorithm has the advantage of faster convergence and it does not require pre - planned lengthy manoeuvre like velocity matching algorithm, which makes it best suited for tactical missions. For large initial misalignment angles, TA problem becomes nonlinear, for which, a conventional particle filter (CPF) based TA algorithm can be used to estimate misalignment. Time varying nature as well as dependencies on external parameters like sensor measurements of state transition matrix of TA problem, makes the system behavior unpredictable and hard to model. A CPF fails in this situation, due to its inability to capture complex nonlinearity associated with the system through system dynamics, due to sample impoverishment problem. Current work addresses this scenario and proposes an evolutionary strategy based algorithm, which performs effectively in such condition, by simulating varied system dynamics through generation of multiple support points. These algorithms are designed and tested for both perfectly modeled and perturbed systems. Simulation results are presented which shows the effectiveness of the proposed algorithm.**

***Index Terms*—particle filter, evolutionary strategy, attitude matching, system dynamics**

I. Introduction

Accurate initialization of navigation information of deliverable guided munitions (called daughters) from a moving platform (called mother), is a prerequisite for delivery of guided munitions. Transfer alignment (TA) is a process of such initialization and calibration of daughter inertial navigation system (INS). Measurements of mother and daughter INS differ due to the misalignment present between them and associated noise of daughter INS measurement. TA algorithm uses theis difference in measurements of mother and daughter INS to estimate their misalignment. A Kalman Filter (KF) is used as an estimator of misalignment and filter of noises associated with daughter INS measurement. Various TA algorithms have been studied in [1], [2].

Velocity matching TA algorithm uses the differences in velocity measurements provided by mother and daughter INS, and used extensively in case of in-flight TA algorithm. Although velocity matching provides optimal estimates of misalignment, it requires a preplanned large manoeuvre to excite inertial sensors of both mother and daughter, and thus, avoided in tactical missions, where faster convergence is of primary concern. Velocity plus attitude matching TA algorithm, also called Rapid Alignment Prototype (RAP) [3], considers differences of velocity as well as attitude in both mother and daughter and converges fast without requiring any large manoeuvre [4], [5].

Due to the failure of small angle approximation employed by the KF in the classical state space formulation, TA problem becomes nonlinear, if the initial misalignment is large. In such case, KF fails to estimate misalignment and some variant like unscented KF (UKF), extended KF (EKF) are used. Both UKF and EKF depend heavily on system models and also suffer from assumption of Gaussian white system as well as measurement noise statistics, which may not be the case in real time. Another approach of handling nonlinearity is particle filter (PF) which does not require these assumptions. Performance of PF based TA algorithms has been studied extensively [6], [7].

One problem associated with PF is sample impoverishment problem in which, only a few particles remain in the system after a few iterations, others being thrown away by the process of re-sampling. This loss of samples becomes a significant problem when the system model is not perfect. Factors such as external input like GPS measurements in the computation of time varying state transition matrix of TA problem, makes the system model erroneous. When CPF uses this erroneous system model to propagate its initially generated particles, it yields poor samples in the system. Upon resampling, some of these poor samples are retained in the system which in turn, affects overall performance of the filter.

To remedy this problem, evolutionary strategy based particle filter algorithm can be applied, which avoids the complex task like resampling of a CPF altogether. Description of such algorithm can be found in [8]. Initially generated N samples of this filter are supported by multiple support points in each pass and among these particle support combination, best N samples closest to measurements, are chosen. This redundant sample helps in avoiding re-sampling by retaining N particles in each pass.

Performance of such a filter named residual evolutionary particle filter (REPF) (named because it brings residual value closer to zero in all situations) has been studied for velocity matching TA algorithm [9]. For tactical flights where faster time of convergence with least possible manoeuvre is of primary concern, velocity plus attitude matching algorithm is preferred. Objective of current work is to investigate the performance of REPF based velocity plus attitude matching algorithm for in flight TA problem. A REPF based RAP algorithm is designed and its performance is compared with that of CPF. In normal situation, performance of the proposed algorithm is found at par with CPF both in terms of time and accuracy. However, simulation results shows that, REPF outperforms CPF in situation, when, system modeling is imperfect.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

Organization of the paper is as follows. After introduction, some basics of state space formulation of TA problem and the problem definition is given in the Section II. Proposed REPF based RAP algorithm is described in Section III. Computer simulation results are given in Section IV. Analysis of effectiveness of the proposed algorithm is discussed in Section V which is followed by conclusion in Section VI.

II. Background

RAP algorithm, has the advantage of faster convergence over only velocity matching TA, without using any large manoeuvre. For these reasons, RAP is preferred in tactical missions. Factor incorporating nonlinearity in the TA system is large initial misalignment, which entails nonlinear formulation of TA problem to be fit with some RAP based nonlinear filtering technique. Nonlinear error propagation equations for large misalignment angles involving large azimuth misalignment has been studied in [10], [11]. For the sake of simplicity in implementation avoiding complex computational loads, the linearized form of nonlinear TA problem described in [12], [13] can be used, which in discretized form, can be represented as follows:

$$\dot{X} = FX + G_w w \tag{1}$$

$$Y = HX + v \tag{2}$$

where, $X = [\delta V^T \ \psi_m^T \ \psi_a^T \ \bigtriangledown_s^{sT} \ \epsilon_s^{sT}]$ is the system state vector comprising of variables as described in Table I.

TABLE I
STATE VARIABLES

Parameters	Description
δV^T (3×1)	Velocity differences between mother and daughter INS
ψ_m^T (3×1)	Measurable time varying misalignment between mother and daughter INS
ψ_a^T (3×1)	Physical mislaignment between mother and daughter which need to be estimated
$\bigtriangledown_s^{sT}$ (3×1)	Accelerometer bias
ϵ_s^{sT} (3×1)	Gyroscopic drifts

F is the state transition matrix, constituted from nonlinear formulation of error propagation equations of sate variables like velocity, attitude as described in [13], [14] and given in Equation No. 3, each entry in the matrix having dimension 3×3. w and v are zero means Gaussian white noise for system and measurement respectively.

$$F = \begin{bmatrix} -(2\omega_{ie}^n + \omega_{en}^n)\times & -C_{s^*}^n f_s^s \times & C_{s^*}^n f_s^s \times & C_{s^*}^n & 0 \\ 0 & -\omega_{ns}^s \times & \omega_{ns}^s \times & I & I \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{3}$$

Measurements used by the filter in this problem is constituted by the differences in velocities and attitudes of both mother and daughter INS, which is a linear function of the differences in measurements of these quantities from mother and daughter INS. Thus, the measurement sensitivity matrix H becomes time invariant, and measurement model can be represented in discretized form by

$$Y = HX + v$$

where, the measurement sensitivity matrix H of dimension 6×15, can be expressed as follows:

$$H = \begin{bmatrix} I & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 \end{bmatrix}$$

A conventional KF applies small angle approximations in the computation of direction cosine matrix (DCM) for ease of computation, which fails in the case of large initial misalignment. A conventional particle filter (CPF) can handle nonlinearity effectively, but suffers from sample impoverishment problem (retention of a few initially generated samples, while omitting others). Moreover, the time varying nature of state transition matrix F due to use of measurements of mother and daughter INS in its computation, specifically in the computation of various DCMs and specific force (f_s^s), makes the system behavior unpredictable and hard to model perfectly, in real time. Such a nonlinear and time varying system cannot be handled by a CPF properly and calls for use of some variant.

III. Proposed Algorithm

The REPF algorithm is a evolutionary strategy based algorithm. The algorithm initially generates N number of particles with a known density and associated weights (equal to $1/N$ for each particles). These particles are then propagated through the system dynamics like CPF. Next, n ($n << N$) number of support points along with associated weights are generated for each particles, based on a known distribution. The system now have nXN particles and support points combination. From this, N best samples, closest to measurements are chosen for further computation, by applying some constant time sorting technique. This essentially avoids the complex and time consuming re-sampling technique required for CPF, and solves the sample impoverishment problem of CPF by retaining all N samples in every iteration.

Generation of n support points for N particles is similar to propagation of N particles through n different systems. This is comparable to multiple model adaptive estimation technique [15], and performs similar to CPF in normal condition (when system is modeled perfectly). Advantage of such filter can be found in situation when the knowledge about the system is inadequate. By generating n support points, REPF encompasses n different systems, among which, the one which provides samples closest to measurement, is selected for next iteration. Effectiveness of this approach has been established by applying it to the classical $2D$ tracking problem, and then applied to velocity matching TA problem in [9].

IV. Simulation Results

Results of computer based simulation is presented in this section in order to prove the effectiveness of the proposed algorithm.

A. Simulation Criteria

Parameters used in this work for the simulation purpose are shown in Table II

For both the filters, particles are initially generated from a known distribution with $\mathcal{N}(0, 0.1)$. System and measurement noises w and v are assumed to be white Gaussian with $\mathcal{N}(0, Q^2)$ and $\mathcal{N}(0, R^2)$ respectively. A regulator is designed which computationally corrects the misalignment and converts the estimation problem to zero tracking problem.

TABLE II
PARAMETERS USED IN SIMULATION RUNS

Parameters	Value
Initial Misalignment	$\psi_\alpha = 7^o$ $\psi_\beta = 8^o$ $\psi_\gamma = 9^o$
Gyro Drift	$3^o/hr$
Accelerometer Bias	$500\mu g$
System noise covariance $Q(15 \times 15)$	$diag([(2.2500e-0016)_{3\times3}\ (2.3504e-8)_{6\times6}\ (0.06e-16)])$
Measurement noise covariance $R(6 \times 6)$	$diag([(3e-4)_{3\times1}\ (2.3504e-7)]$

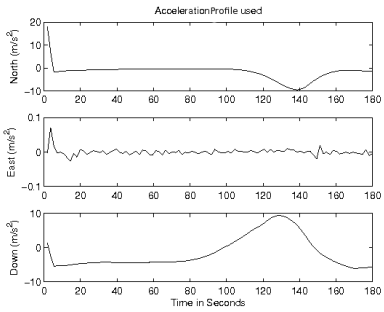


Fig. 1. Less than $1g$ manoeuvre used for simulation

B. Results

Figure 1 shows the acceleration profile used for the simulation purpose. Less than $1g$ manoeuvre is used to justify the fact that, RAP does not require high g manoeuvre to estimate the misalignment. Availability of mother measurement is taken in the interval of 1.8 seconds. Figures 2, 3 and 4 depicts the execution of both CPF and REPF based RAP algorithm, when the system is perfectly modelled. For the purpose of convenience in representation, CPF based RAP algorithm is designated as AMPF and REPF based RAP algorithm is designated by AMEPF in every plots of results. From these figures it can be seen that, less than 1 arc minute accuracy is achieved by both the filters over a time span of 10 seconds (from 3rd to 13th seconds) which is considered adequate for TA assuming 7°, 8° and 9° initial misalignment and both the filters converge in less than 5 seconds time which is typical for RAP.

To simulate the scenario of imperfect system model, parameters of the state transition matrix F as described in Equation 3 are perturbed by injecting about 10% errors in its constituent elements for the second run. For perturbation purpose, elements constituting ψ_a are chosen specifically, keeping others, like $\bigtriangledown_s^{sT}, \epsilon_s^{sT}$ (which are independent of and can not affect any parameters) in the system states, unperturbed. These parameters basically act as an additive noise to the system. Although similar results can be obtained by adding noise in the system, this approach is chosen to simulate imperfect system dynamics rather than noisy system. Performance of these two filters in this situation is recorded and shown in Figures 5, 6 and 7. It can be observed that, CPF is estimating with error in the range of 40 arc minutes, sometimes even showing nonconvergence. For the purpose of estimation of initial misalignment in the range of $[7^o - 9^o]$, this performance is considered poor. Contrary to the performance of CPF, REPF still manages to estimate the misalignment with the accuracy of < 5 arc minutes and converges in less than 6 seconds time which is at par with the standard of any RAP filter.

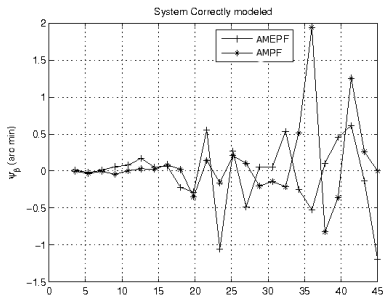


Fig. 3. Performance of two filters in the estimation of misalignment angle ψ_β when system is perfectly modeled

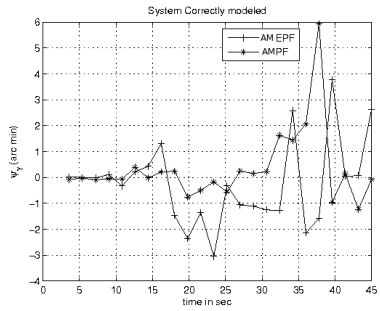


Fig. 4. Performance of two filters in the estimation of misalignment angle ψ_γ when system is perfectly modeled

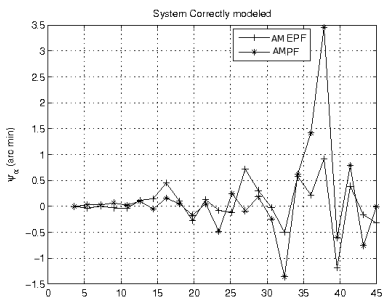


Fig. 2. Performance of two filters in the estimation of misalignment angle ψ_α when system is perfectly modeled

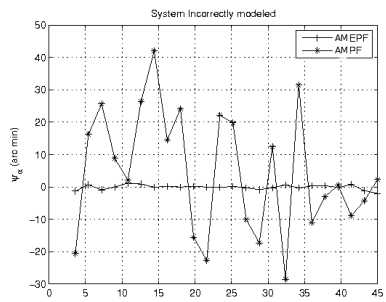


Fig. 5. Performance of two filters in the estimation of misalignment angle ψ_α when system is NOT perfectly modeled

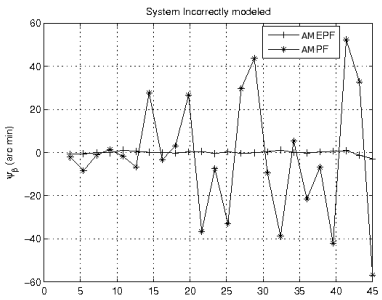


Fig. 6. Performance of two filters in the estimation of misalignment angle ψ_β when system is NOT perfectly modeled

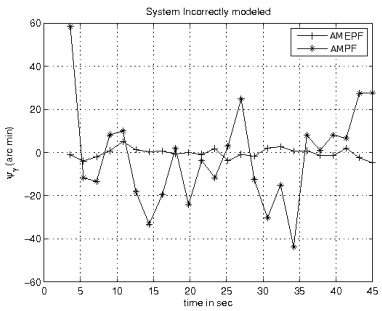


Fig. 7. Performance of two filters in the estimation of misalignment angle ψ_γ when system is NOT perfectly modeled

V. Discussion

To simulate imperfect system model, elements of state transition matrix was perturbed in the second run. This perturbation affected the initially generated particles by propagating through wrong system dynamics. Few of these corrupted particles, comparatively closer to measurement than others, were retained in the system by the method of resampling while others were dropped. CPF continues with these particles, which affects its performance altogether. On the other hand, the REPF based algorithm can use one of many choices of system dynamics in every iteration. By this way, it can get samples closer to measurement by covering more search space and thus, handling of poorly modelled systems becomes possible.

From the results presented in previous section and discussion presented above, it can be concluded that, REPF based RAP algorithm is best suited for situations when:

- knowledge of the system is not proper
- initial misalignment is large
- faster and accurate convergence is of interest
- large manoeuvre can not be performed as in tactical mission

Time complexity of both CPF and REPF algorithm are linear in terms of particle count. One factor affecting the time as well as space complexity of overall algorithm is the size of the state vector used and the size of the state transition matrix, which, in this case, is a 15×1 vector and 15×15 matrix. Reduction of the size of state vector can lead to better performance of the filter.

VI. Conclusion

An evolutionary strategy based in flight TA algorithm has been developed. Its performance has been compared with similar type of CPF algorithm. It is shown by computer simulation that, both the proposed REPF and CPF algorithms perform similar in normal situation, when knowledge of the system dynamics is good. In real time situation, factors such as interrupted supply of mother measurement, external aiding in the computation of state transition matrix makes the system behavior unpredictable and such systems are hard to model. It is shown by simulation that, a CPF fails to handle such nonlinear and time varying systems mostly due to the sample impoverishment problem. Proposed algorithm uses one among many system dynamics to select best particles and can perform satisfactorily even in such adverse situation.

Although comparable in terms of time as well as space complexity with a CPF, some implementation issues such as clever and efficient handling of state transition matrix, scheduling of some tasks of REPF in multiple processors, parallel execution of more than one algorithm in a single processor for more accurate and timely convergence of TA algorithm are some topics that need to be explored for further research.

References

[1] S. Chattaraj, A. Mukherjee and S. K. Chaudhuri, "Transfer Alignment Problem: Algorithms and Design Issues", in *Gyroscopy and Navigation*, Vol. 4, pages:130–146, 2013

[2] P.D Groves.,"Optimising the transfer alignment of weapon INS.", in *Journal of Navigation*, 56(2), pages:323–335, 2003.

[3] K.J. Shortelle, W.R. Graham, and C. Rabourn.,"F-16 flight tests of a rapid transfer alignment procedure", in *Position Location and Navigation Symposium, 1996.*, IEEE 1998, April, pages:379–386

[4] J. E. Kain and J. R. Cloutier, "Rapid transfer alignment for tactical weapon applications", in *Proceedings of AIAA Guidance, Navigation and Control Conference*, pages:1290–1300, 1989.

[5] KEVIN SPALDING.,"An efficient rapid transfer alignment filter", in *AIAA Guidance, Navigation and Control Conference, Hilton Head Island, SC; UNITED STATES;*, pages:1276–1286, August 1992.

[6] Cheng, Y. and J. L. C., "Particle Filtering for Sequential Spacecraft Attitude Estimation", in *Collection of Technical Papers AIAA Guidance, Navigation, and Control Conference*, August 2004.

[7] Guoqiang Ding, Xiaoyu An, Weidong Zhou, and Guangzhao Cui.,"A New Numerical Approximation CDD Method Based for Particle Filtering Algorithm Research and Its Applications to TA System", in *Third International Conference on Measuring Technology and Mechatronics Automation*, Vol. 1, pages: 178–182, January. 2011.

[8] Uosaki, K. and Kimura, Y. and Hatanaka, T.,"Nonlinear state estimation by evolution strategies based particle filters", in *The 2003 Congress on Evolutionary Computation, 2003. CEC '03.*, Dec. 2003, pages: 2102–2109, Vol.3.

[9] Suvendu Chattaraj and Abhik Mukherjee, "Efficient in-flight transfer alignment using evolutionary strategy based particle filter algorithm", accepted in *11th International Conference on Informatics in Control, Automation and Robotics (ICINCO)*, September, 2014.

[10] H. S. Hong, J. G. Lee, and C. G. Park, "In-flight alignment of SDINS under large initial heading error", in *AIAA Guidance, Navigation, and Control Conference and Exhibit*, pp. 1–6, 2001.

[11] Mechanical Kong, Mario Nebot and Hugh Durrant-whyte, "The Ins. Development of a non-linear psi-angle model for large misalignment errors and its application in ins alignment and calibration", in *Proceedings 1999, IEEE International Conference on Robotics and Automation Cat No 99CH36288C,*.

[12] S.P. Dmitriyev, O.A. Stepanov, and S.V. Shepel., "Nonlinear filtering methods application in ins alignment", in *IEEE Transactions on Aerospace and Electronic Systems*, 33(1), pages: 260–272, January. 1997.

[13] Zhilan Xiong, Liang Guo, Genbao Feng, Yufeng Qu and Feng Sun,"Take-off Alignment Algorithm of Carrier Aircraft INS", in *2nd IEEE Conference on Industrial Electronics and Applications, 2007. ICIEA 2007.*,May,2007, pages: 2333–2339.

[14] B.M. Scherzinger.,"Inertial navigator error models for large heading uncertainty", in *Position Location and Navigation Symposium, 1996.*, IEEE 1996, pages: 477–484, April 1996.

[15] P.D. Hanlon and P.S. Maybeck.,"Multiple-model adaptive estimation using a residual correlation Kalman filter bank", in *IEEE Transactions on Aerospace and Electronic Systems*, 36(2): pages:393–406, April 2000.

p-i-n Solar Cell Modeling with Graphene as Electrode

M. W. Khan, T. D. Gupta and S. Subrina*, Member, *IEEE*

Department of Electrical & Electronic Engineering, Bangladesh University of Engineering & Technology
Dhaka- 1000, Bangladesh

*samiasubrina@eee.buet.ac.bd

Abstract— **Graphene is a 2-D atomic layer of carbon atoms with unique electronic properties like outstanding carrier mobility, high carrier saturation velocity, excellent thermal conductivity, high mechanical strength, transparency, thinness, and flexibility which make graphene an excellent choice of material for advanced applications in future solar cell design. We modeled a solar cell using graphene as the front electrode to study its performance and compare the performance with that of other possible contenders– indium tin oxide (ITO), widely used material at present and carbon nanotube (CNT), another promising material in this regard. Numerical solutions of the electrostatic and transport equations were obtained using the finite-element method. It was found that solar cell with graphene electrode can outperform the others. We also studied its performance as a function of various parameters. The developed model and obtained results are important for the design of solar cell with graphene as electrode.**

Index Terms— **CNT, Electrode, Graphene, ITO, Solar Cell.**

I. Introduction

The electrode of the solar cell module which is on the side of illumination has a significant impact on the energy conversion efficiency. Reflection, absorption and resistive losses of the electrode play an important role on the efficiency of the solar cell. So this front electrode must be highly transparent and conductive as well as must have very low resistance. Tin-doped indium oxide or ITO is the most widely used transparent-conducting material or Transparent Conductive Oxide (TCO). Such transparent conducting thin films are in widespread use as electrodes in a variety of optoelectronic devices. Currently it is used in most of the solar cells as its front electrode because of its optical transparency and electrical conductivity. But due to the supply constraint of ITO and increasing price, we must consider other materials which can replace ITO and give even better solar cell performance.

Graphene is a revolutionary material which can be considered as the primary replacement of ITO. It is a 2D carbon sheet with hexagonal crystal structure. It shows very high transparency of upto 97.7% whereas ITO has a transparency of 85% in the visible spectrum [1]–[4]. Graphene also has a superior electron mobility of 20,000 $cm^2/V.s$ as compared to 70 $cm^2/V.s$ for ITO [5]–[6].The high electron and hole transport mobility, large specific area and the inertness against oxygen and water vapour make graphene a promising candidate in photovoltaic application. Flexibility is one of the prime demand for upcoming electronic devices and solar cells. Graphene is simultaneously flexible and robust whereas ITO is rigid and fragile.

We also compare another possible candidate carbon nanotube (CNT), one dimensional carbon allotrope, which also has very favourable properties like high electron mobility conductivity and transparency to act as solar cell front electrode [7]–[9]. Unlike graphene, CNT presents higher contact resistance. Two dimensionality and structural flatness make it easier to combine graphene with other semiconductor material such as silicon.

In this paper, we observe the current density (J)-voltage (V) characteristics of our model with graphene as electrode and compare the performance with that obtained from solar cells with CNT and ITO electrodes. We also studied the effect of different parameters on the performance of our proposed model.

II. Solar Cell Model

With a view to study the performance of solar cell with graphene as front electrode, we solved electrostatic and transport equations in our proposed structure numerically. The base model, as shown in Fig. 1, consists of an intrinsic layer of silicon between n and p type silicon. At the top there is a thin layer of graphene which acts as the front electrode of the solar cell model.

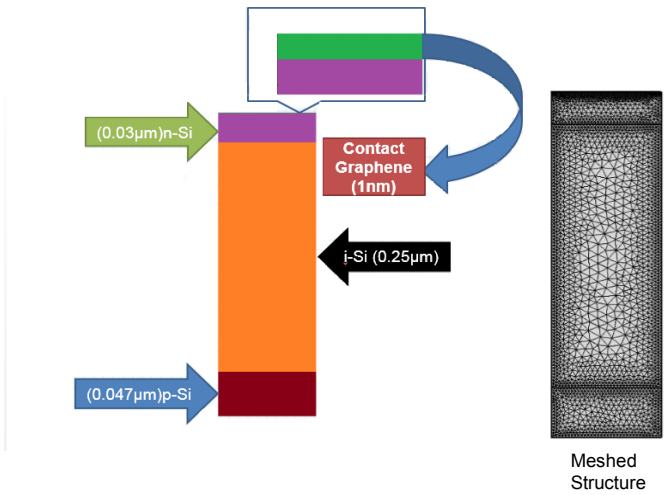


Fig. 1. Schematic cross section of p-i-n solar cell with graphene as electrode

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

The thicknesses of intrinsic, n and p type silicon and front electrode are assigned to be 0.25µm, 0.03µm, 0.047µm and 1nm respectively. The donor and acceptor concentration for silicon is considered to be 10^{17}cm^{-3}. The sheet density of graphene is 10^{-12} cm^{-2}[10]. Ali *et al.* shows that the concentration for ITO decreases with its thickness [11]. Here we use a thin electrode so the effect of thickness in the carrier concentration of ITO is considered.

We have solved Poisson's Equation and Drift-Diffusion Equation using finite element method with the help of COMSOL [12].

In electrostatics, Poisson's equation has the form of

$$\nabla.\mathbf{D} = \rho_v \tag{1}$$

Where $\mathbf{D}$ = electric displacement field = $\varepsilon\mathbf{E} = \varepsilon_o\varepsilon_r\mathbf{E}$ and ρ_v = free charge density.

The convection–diffusion equation is a combination of the diffusion and convection (advection) equations [13].

$$\frac{\partial c}{\partial t} = \nabla.(D\nabla c) - \nabla.(\mathbf{u}c) + \mathrm{R} \tag{2a}$$

Where c is the variable of interest (species i.e., electron or hole concentration for mass transfer), D is the diffusion coefficient $\mathbf{u}$ is the average velocity of the species. R describes sources or sinks of the quantity c. At steady state (i.e., $\partial c / \partial t = 0$) the equation takes the form as:

$$\nabla.(-\mathrm{D}\nabla \mathrm{c} + \mathbf{u}c) = \mathrm{R} \tag{2b}$$

This leads to drift-diffusion equation in semiconductor,

$$\frac{\mathbf{J_n}}{-q} = -D_n\nabla n - n\mu_n\mathbf{E} \tag{3a}$$

$$\frac{\mathbf{J_p}}{q} = -D_p\nabla p + p\mu_p\mathbf{E} \tag{3b}$$

Where n and p are the concentrations of electrons and holes, respectively, q is the charge of an electron, $\mathbf{J_n}$ and $\mathbf{J_p}$ are the electric current density due to electrons and holes respectively, $\mathbf{J_n}/-q$ and $\mathbf{J_p}/q$ are the corresponding particle currents of electrons and holes respectively.R in equation 2b represents carrier generation ($R>0$) and recombination ($R<0$), D_n and D_p are electron and hole diffusion coefficient, μ_n and μ_p are electron and hole mobility, respectively and $\mathbf{E}$ is the electric field vector.

According to the Beer-Lambert law we get generation, G which varies as light passes in the y direction as

$$G = G_o \exp(-\alpha\, \mathrm{y}) \tag{4}$$

The current density of the solar cell is obtained from

$$J = -J_{ph} + J_d \tag{5}$$

Where J_d is the diode current density where

$$J_d = J_n + J_p \tag{6}$$

Here J_n and J_p are the electron and hole current density respectively and J_{ph} is the photo-current density which is expressed as

$$J_{ph} = \frac{qG_o}{\alpha}\left\{1 - exp\left[-\alpha\left(\ell n + W + \ell p\right)\right]\right\} \tag{7}$$

Here α is the absorption coefficient of the medium through which light passes, G_o is the carrier generation rate at the top surface of solar cell, q is the charge of electron, W is depletion width and ℓn & ℓp are minority carrier diffusion length of electron and hole respectively.

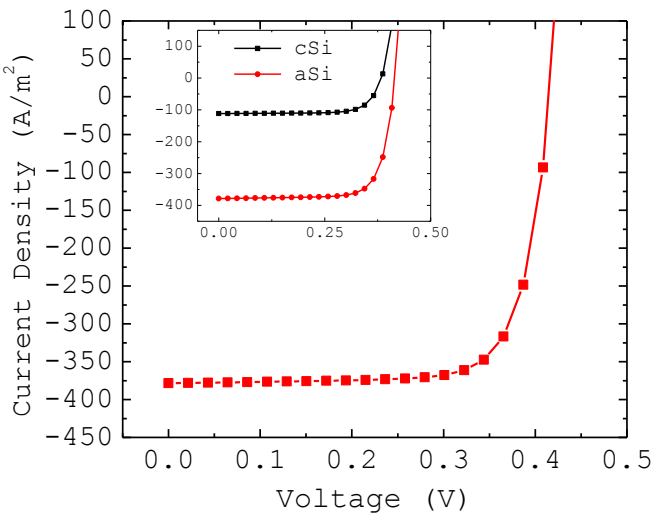


Fig. 2. J-V Characteristic for p-i-n graphene solar cell; Inset: Comparison between amorphous (Red) and crystalline (Black) silicon solar cell.

III. Results

We obtain different J-V characteristics as a result of our simulation. The input optical power is considered to be 1000W/m^2 of a wavelength of 530nm if not otherwise stated. Photocurrent density or short circuit current density (J_{ph}), open circuit voltage (V_{oc}), fill factor (FF) and efficiency (Eff) are the main indicator of the performance of a solar cell. We have used current density rather than only current. It is more convenient while comparing the characteristics of different modules or devices with different sizes.

Fig. 2 shows the J-V characteristics of p-i-n amorphous silicon solar cell with graphene as electrode. It shows that at 1000W/m^2 input optical power of wavelength 530nm, the photocurrent density, open circuit voltage and efficiency is 378.5 A/m^2, 0.4147 V and 12%, respectively.

According to theoretical calculation, photo-current densities for crystalline and amorphous silicon solar cell are 118A/m^2 and 415A/m^2, respectively.

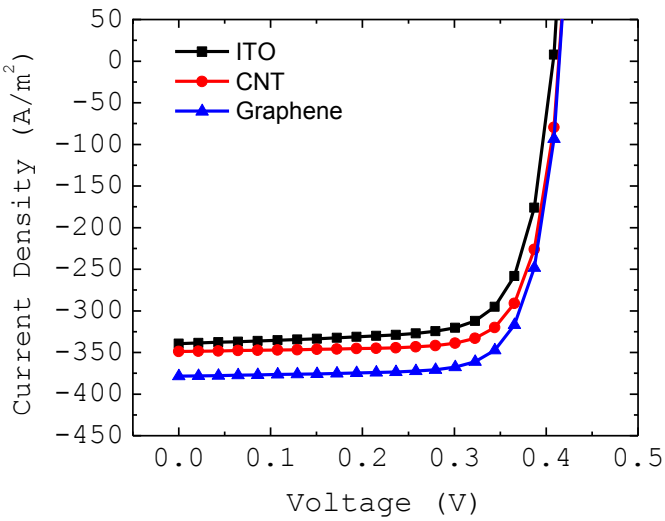


Fig. 3. Performance comparison among the solar cells with three different electrodes.

TABLE I
SOLAR CELL PERFORMANCE USING THREE DIFFERENT ELECTRODES

Performance indices	Graphene	CNT	ITO
J_{ph} (A/m^2)	378.5	349.5	339.3
V_{oc} (V)	0.4147	0.4135	0.4076
FF	76.45%	75.72%	68.23%
Eff	12%	10.9%	9.4%

The photo-currents from simulation are comparable with these theoretical values under same condition since in our simulation we obtain a photocurrent density of 111.25A/m^2 for crystalline silicon and 378.5A/m^2 for amorphous silicon at 1000W/m^2 input power. Amorphous silicon solar cell gives higher current due to higher absorption coefficient. The value of absorption coefficient is 1.0206x10^6 m^{-1} for crystalline and 1.58x10^7 m^{-1} for amorphous silicon [14].

In Fig. 3 we show the J-V characteristics obtained for the amorphous silicon solar cells with three different electrodes. The performances of solar cells with these electrodes are also presented in TABLE I for better comparison. The photocurrent density, open circuit voltage, fill factor and most importantly the efficiency is given for each solar cell with graphene, carbon nanotube and ITO electrode.

The best performance is obtained from graphene with a photocurrent of 378.5 A/m^2 and efficiency of 12% at 1000W/m^2 input power with a wavelength of 530nm. It also shows that carbon nanotube gives better characteristics than indium tin oxide.

Inclusion of intrinsic layer into the structure creates larger depletion region and thus the efficiency of solar cell is increased. This matches with the result obtained from our simulation as shown in Fig. 4 such that with the decrease of intrinsic layer length from 0.25µm to 0.09µm the photocurrent density as well as the efficiency decreases. It also shows that the photocurrent is significantly reduced when p-n junction is used instead of p-i-n.

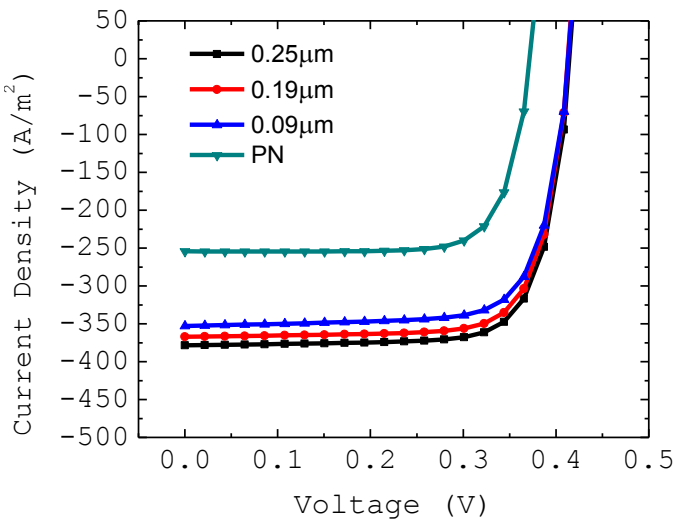


Fig. 4 J-V Characteristic for p-n solar cell and p-i-n solar cell with variation of intrinsic layer length (µm)

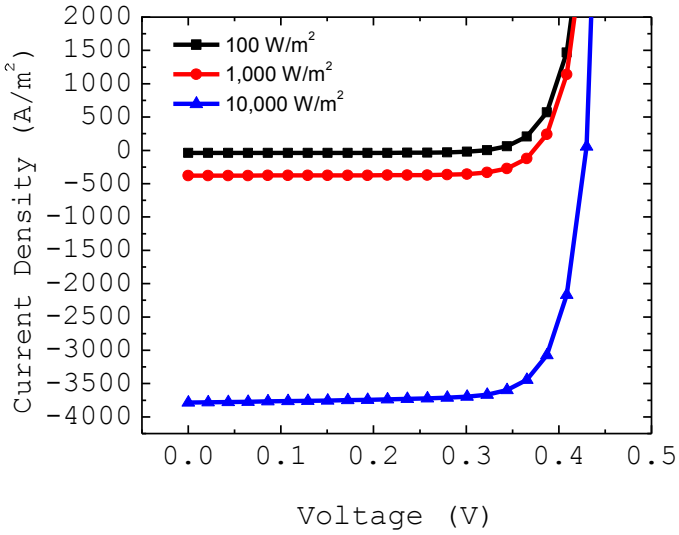


Fig. 5. J-V Characteristic for p-i-n graphene solar cell with varying input optical power intensity

At this stage, we vary the optical input power and obtain the result as shown in Fig 5. Since the current density is proportional to the input power intensity, if we make the input optical power ten times larger, the photocurrent also increases by ten times. For 100, 1000 and 10000 W/m^2 incident optical intensity the photocurrent density is 37.85, 378.5 and 3784.65 A/m^2, respectively. The open circuit voltage is a logarithmic function of input optical power. So it varies slightly with the variation of input power intensity.

As we increase the wavelength, keeping the input power intensity fixed, photon flux is increased due to decrease of photon energy and so does the generation up to a certain level. Hence the photo-current density increases for our proposed model as shown in Fig 6. Dimension of the model is a concerned issue in this regard.

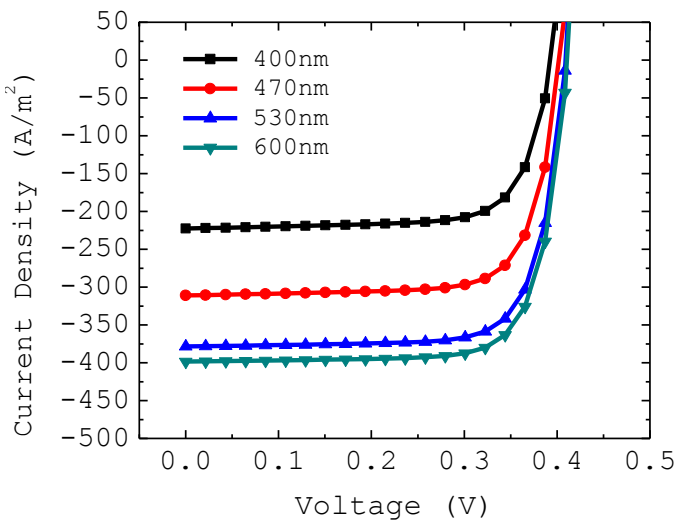


Fig. 6. Variation in current density with voltage for p-i-n graphene solar cell for different incident wavelength.

IV. Conclusion

In this work we developed a model of solar cell with graphene as electrode and studied the device performance. From the simulation we observed that the revolutionary material graphene gives better performance in comparison to CNT and ITO, two other solar cell front electrode candidates. Numerical results showed that the photocurrent is a strong function of input light intensity, wavelength of the incident light as well as the dimension of the structure. The properties of graphene make the material extremely suitable for using it as the electrode in solar cell and other electronic devices. Its flexibility and high mechanical strength can also pave the way to flexible and strong solar panels.

References

[1] D. E. Sheehy and J. Schmalian, "Optical transparency of graphene as determined by the fine-structure constant," *Physical Review B*, vol. 80, Nov. 2009

[2] F. Bonaccorso, Z. Sun, T. Hasan and A. C. Ferrari, "Graphene photonics and optoelectronics," *Nature Photonics 4*, Aug. 2010

[3] H. Kim, C. M. Gilmore, A. Pique´, J. S. Horwitz, H. Mattoussi, H. Murata, Z. H. Kafafi, and D. B. Chrisey, "Electrical, optical, and structural properties of indium–tin–oxide thin films for organic light-emitting devices," *Journal of Applied Physics*, vol. 86, Dec. 1999

[4] Plasma Quest Ltd ITO Thin Film Datasheet

[5] S.V. Morozov, K.S. Novoselov, M.I. Katsnelson, F. Schedin, D.C. Elias, J.A. Jaszczak and A.K. Geim, "Giant Intrinsic Carrier Mobilities in Graphene and Its Bilayer," *Physical Review Letters*, vol. 100, Jan 2008

[6] S. H. Keshmiri, M. Rezaee-Roknabadi and S. Ashok, "A novel technique for increasing electron mobility of indium-tin-oxide transparent conducting films," *Thin Solid Films*, 2002

[7] T. Dürkop , S. A. Getty , Enrique Cobas , and M. S. Fuhrer, "Extraordinary Mobility in Semiconducting Carbon Nanotubes," *Nano Letters*, vol. 4, pp 35-39, 2004

[8] Y. Ryu, L. Yin and C. Yu, "Dramatic electrical conductivity improvement of carbon nanotube networks by simultaneous de-bundling and hole-doping with chlorosulfonic acid**,"** *Journal of Materials Chemistry*, vol. 22, Jan. 2012

[9] H.-Z. Geng, D. Lee, K. K. Kim, S. J. Kim, J. J. Bae and Y. H. Lee, "Effect of Carbon Nanotube Types in Fabricating Flexible Transparent Conducting Films," *Journal of the Korean Physical Society*, vol. 53, pp 979-985, Aug. 2008

[10]J.-H. Chen, C. Jang, S. Xiao, M. Ishigami and M. S. Fhurer, "Intrinsic and extrinsic performance limits of Graphene devices on SiO_2," *Nature Nanotechnology*, vol. 3, Mar. 2008

[11] M. K. M. Ali, K. Ibrahim, O. S. Hamad, M. H. Eisa, M. G. Faraj and F. Azhari, "Deposited Indium Tin Oxide (ITO) Thin Films by DC-Magnetron Sputtering on Polyethylene Terephthalate Substrate (PET)," *Rom. Journ. Phys.*, vol. 56, 2011

[12] [Online]. Available: http://www.comsol.com

[13] Thomas Stocker, *Introduction to Climate Modelling*, p57

[14] Refractive Index Database, Available: refractiveindex.info

Effects of physical parameters on subthreshold characteristics of nitride and antimonide-based double material gate (DMG) HEMTs

Arman-Ur-Rashid*, Md. Aynal Hossain, Tanvir Rahman and Farseem M. Mohammedy
Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology (BUET)
Dhaka-1000, Bangladesh
*Email: arman0806161@yahoo.com

Abstract— **In this papar we have varied the physical parameters of double material gate (DMG) HEMT and noted the change in subthreshold characteristics. The physical parameters are: control-gate length, barrier layer thickness and the workfunction difference between two gate materials. A semi-classical analytical model has been used to determine the channel potential and electric field. One nitride-based and three antimonide-based formations have been used in this work and their results have been compared. We have found that nitride-based DMG HEMTs are less sensitive to process variation of gate fabrication compared to antimonide-based DMG HEMTs. It has been seen that variation in barrier layer thickness affects nitride-based HEMTs more strongly than antimonide-based DMG HEMTs, while the effect of change in workfunction difference is more prominent in antimonide-based HEMTs. Among the antimonide-based HEMTs, it have been seen that the formation with deep conduction band discontinuity is more susceptible to change in control-gate length than formations with shallow conduction band discontinuity. But effects of barrier layer thickness change and change in workfunction difference are more noticeable in shallow conduction band formations, making these formations more effective as double material gate (DMG) structures.**

Index Terms— **HEMT, antimonide-based, nitride-based, physical parameters.**

I. Introduction

With the continuous trend of scaling, the silicon-based MOSFET technology is rapidly approaching to its limit [1]. As a result new structures are continuously being investigated to produce faster, smaller and less power consuming device. HEMT (high electron mobility transistor) is one such device which is being extensively researched. HEMTs have inherent higher mobility and are easy to fabricate compared to same size MOSFETs [2]. In field of HEMT, nitride-based materials have already made their mark. Nitride-based HEMTs are well known for their ability to work in high voltage, high power environments [3]. They are also suitable to work in integrated circuits with ultraviolet photo detectors. On the other hand, researchers are interested in antimonide-based HEMTs for their low on-voltage, ultra high speed and very low power consumption characteristics [4]. Suitable for the integration of third generation infrared photo-detector with ultra-high speed transistor, antimonide-based materials are opening a promising window of research through HEMT structure [5].

To make faster devices, scaling is a must. Like other device-structures, HEMT is also going through scaling [6]. With the advancement of fabrication techniques and material science, even sub-micrometer range HEMTs with excellent speed performance are now reality [7]. But as the gate length decreases below 300nm, different short channel effects and carrier transport inefficiency become the prime concern. Much research is being done to solve these problems of HEMT. Various structures like double gate HEMT, split gate HEMT etc. were proposed , but none can solve the problems of short channel effect and carrier transport inefficiency simultaneously [8]. The ultimate solution was given by Long et al. [9] in his proposed Dual Material Gate (DMG) FET structure.

In DMG FET, two gates made of different materials of different workfunctions are joined laterally to create a single gate. The workfunction of one gate (M1) is chosen to be to be greater than the other (M2). Gate (M1) works as control-gate and gate M2 works as screening gate for drain voltage. It enhances the average electric field and inserts a step function in channel potential at the interface of two gates. The enhanced average electric field improves the carrier transport efficiency and the step function reduces the short channel effects. But as the scaling goes into deep sub-micron level, need for an analytical model of subthreshold behavior of DMG HEMT emerges. S.P.Kumar et al. [8] first proposed an analytical model for the subthreshold behavior of nitride-based DMG HEMT. In one of our previous paper [10] the model of [8] was modified and an analytical model for antimonide-based HEMTs was developed.

We have worked with three formations of antimonide-based HEMTs and a nitride-based HEMT. In [10] we used the barrier/channel formation of $Al_{0.7}Ga_{0.3}Sb/InAs$. Among the antimony-based material AlSb and GaSb are approximately lattice matched to InAs and these semiconductors are jointly known as 6.1A^o family. In this paper we also study the $Al_{0.7}Ga_{0.3}Sb/GaSb$ and $Al_{0.7}Ga_{0.3}Sb/In_{0.15}Ga_{0.85}Sb$ barrier/channel combinations and compare their results. We also derived the results of $Al_{0.2}Ga_{0.8}N/GaN$ DMG HEMT from the proposed model of [8] and compared those results with antimonide-based DMG HEMTs.

In this paper, we have varied three key physical parameters of different antimonide-based and nitride-based DMG HEMTs and compared the results. Our focus is on the change in minimum channel potential (φ_{limin}) due to change in physical parameters. This parameter is the key to understand the effect of DMG structure over the HEMT. So by varying the physical parameters we can not only understand the effects of possible

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

scaling of DMG HEMT but also the effectiveness of DMG structure over various formations.

II. MODEL FORMULATION.

Fig. 1 shows the cross-section of proposed DMG HEMT structure. Here, L1 and L2 are the lengths of gates with metal M1 and M2 respectively. L is the total length. The drain and source regions are doped uniformly with donor concentration N_d+=$10^{25}m^{-3}$. The workfunctions of metal M1 is chosen φ_1=5.3 V, which is greater than the workfunctions of metal M2 φ_2=4.1 V (if not stated otherwise). φ_1 is chosen greater than φ_2 to make the HEMT n-channel [12]. The gate voltage and drain voltage were chosen -0.2 V and 0.2 V respectively to make the barrier layers completely depleted and confined electrons in the heterointerface. The typical values of other parameters are listed in Table. I. If not stated otherwise these are values used in the model.

The details of the analytical model and simulation are given in [10] for antimonide-based DMG HEMTs and we followed the nitride-based DMG HEMT model from [8].

Table I. Lists of parameters

Parameter	Name of the parameter	Value
N_d	Doping in the barrier layer	10^{23} m^{-3}
L	Effective channel length	120nm
d_d	thickness of doped barrier layer	25nm
d_i	Thickness of spacer layer	5nm
ε_s	Dielectric constant of $Al_{0.7}Ga_{0.3}Sb$	13.1
T	Operating temperature	300K

In this paper we have chosen three different formations of antimonide-based DMG HEMTs to compare the effect of DMG structure within different antimonide-based HEMTs. The barrier layer is chosen to be $Al_{0.7}Ga_{0.3}Sb$ which gives us the freedom of bandgap engineering with ternary AlGaSb (Al=1~0.7) where minimum mole fraction of Al required to make type I-straddling heterojunction is 0.7. Moreover it gives better hole blocking performance in valence band [13]. The first channel layer is InAs which will provide large conduction band offset, high peak electron velocity and high electron concentration in InAs channel [4]. This will be the formation with the deepest conduction band discontinuity at the heterojunction used in our analysis. For our second HEMT's channel layer, we selected the $Ga_{0.85}In_{0.15}Sb$. It gives a shallower conduction band offset and its growth on GaSb substrate is well studied [10]. The third structure has GaSb as channel layer. It gives the shallowest band-offset in the heterointerface among the chosen formations. In case of nitride-based DMG HEMT we followed the structure used in [8].Table II. gives the lists of various formations of DMG HEMTs used in our paper.

Table II. Formations of HEMT

Barrier layer	Channel layer	Substrate
$Al_{0.7}Ga_{0.3}Sb$	InAs	GaSb
$Al_{0.7}Ga_{0.3}Sb$	$Ga_{0.85}In_{0.15}Sb$	GaSb
$Al_{0.7}Ga_{0.3}Sb$	GaSb	GaSb
$Al_{0.2}Ga_{0.8}N$	GaN	GaN

While forming HEMT structures, it is important to consider the strain in the heterojunctions [4]. Too much strain in forming the heterostructure degrades the mobility and thus hampers the main objective of using HEMT. Table III gives a percentage of strain for the various HEMTs in this paper.

Table III. Percentage strain in different heterojunctions

Heterostructure	Percentage of strain
InAs/ GaSb	0.6173
$Ga_{0.85}In_{0.15}Sb$/ GaSb	0.9426
$Al_{0.7}Ga_{0.3}Sb$/ GaSb	0.4544
$Al_{0.2}Ga_{0.8}N$ /GaN	0.4829

III. RESULTS AND DISCUSSION

We have used ATLAS device simulator to simulate channel potential and electric field distribution along the channel to verify the model of [10]. Fig.2 shows electric field variation vs normalized channel length and the channel potential vs. normalized channel length for DMG $Al_{0.7}Ga_{0.3}Sb$/GaSb HEMT. Close match between proposed model in [10] and ATLAS's simulation is evident from the figure.

In fig.3 we have shown the effects of varying the gate length

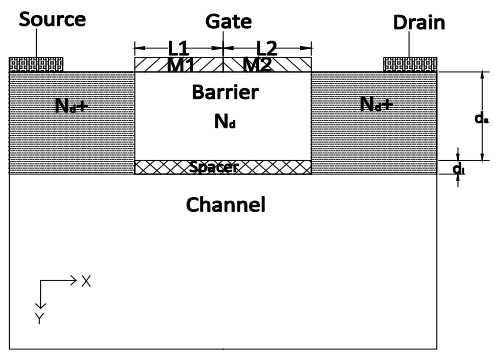


Fig.1: Cross sectional view of a DMG HEMT.

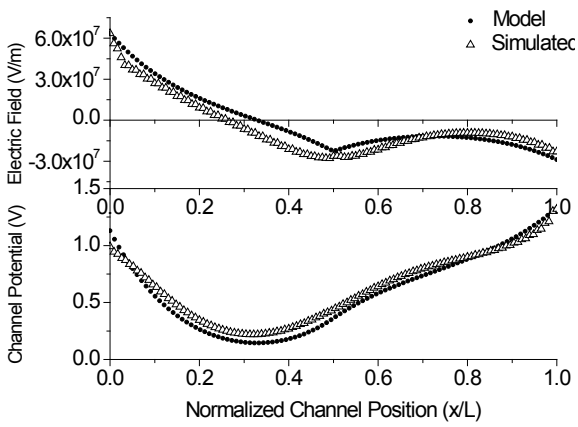


Fig.2: variation of channel potential and electric field vs the normalized channel length

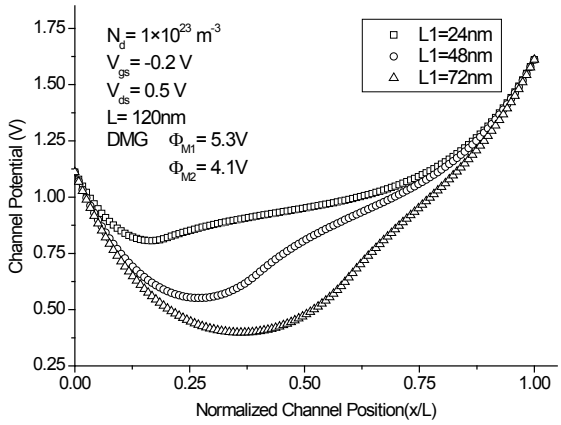


Fig.3: variation of channel potential with normalized channel position for different combinations of gate length L1 and L2, keeping the L constant for $Al_{0.2}Ga_{0.8}N$ /GaN DMG HEMT.

of first metal gate (L1) keeping the total gate length L constant. As we can see in the figure, the minimum potential (φ_{limin}) increases as the L1 decreases. It reduces the potential barrier and thus enhances the drive current (I_{on}). Moreover it increases the control of gate over the channel potential [14]. In table IV we have listed the change in φ_{limin} with change in L1 for various formations of HEMTs.

Table IV. Change in φ_{limin} for change in L1

HEMT composition	Change in φ_{limin}(volt) for change in L1 between 24-48 nm	Change in φ_{limin}(volt) for change in L1 between 48-72 nm
$Al_{0.2}Ga_{0.8}N$ /GaN	0.2548	0.1537
$Al_{0.7}Ga_{0.3}Sb$ /InAs	0.4279	0.2756
$Al_{0.7}Ga_{0.3}Sb$ /$Ga_{0.85}In_{0.15}Sb$	0.3668	0.2469
$Al_{0.7}Ga_{0.3}Sb$ /GaSb	0.3546	0.2197

From the table IV we see that the change of φ_{limin} in range (24-48) nm is always greater than the change in the range (48-72) nm. So while designing DMG HEMT L1 should be as small as possible with available fabrication process. From the data in the table IV it is evident that for nitride-based DMG HEMT the change of φ_{limin} is smaller compared to various antimonide-based DMG HEMT. It reveals that nitride-based DMG HEMTs are less dependent on process variation when comes to fabricate DMG structure. Among the antimonide-based DMG HEMTs, the $Al_{0.7}Ga_{0.3}Sb$ /InAs formation has the highest change in minimum-potential φ_{limin} for the change in control gate length, while the $Al_{0.7}Ga_{0.3}Sb$ /GaSb formation has the lowest. The deep conduction band discontinuity at the heterojunction produces more 2-D electron gas for $Al_{0.7}Ga_{0.3}Sb$/InAs enabling sharp change in potential barrier with the structural variation.

In fig. 4 we can see the effect of varying the barrier layer thickness (d) on DMG HEMT. We can see that as the barrier layer decreases the φ_{limin} increases. So reduction of barrier layer thickness reduces the channel potential barrier. It increases the drive current. It also ensures better gate control and improved carrier transport efficiency [15]. Table V lists the rate of change in φ_{limin} for change in barrier layer thickness (d). From the table we can see that rate of change in nitride-based DMG HEMT is higher compared to antimonide-based DMG HEMTs. It is because barrier layer $Al_{0.2}Ga_{0.8}N$ has lower dielectric constant compared to $Al_{0.7}Ga_{0.3}Sb$. As a result charge concentration is thinner and control of gate over the charge is larger. On the other hand the charge concentration in deep conduction band discontinuity at the heterojunction of $Al_{0.7}Ga_{0.3}Sb$/InAs is largest among the proposed concentration. As a result the control of gate is smaller. So change in thickness of barrier layer does not produce as strong effect as in the shallow heterojunction $Al_{0.7}Ga_{0.3}Sb$ /GaSb formations where concentration of carriers is comparatively thinner.

Table V. rate of change in φ_{limin} for change in barrier thickness (d)

HEMT composition	Rate of change in φ_{limin}(volt/nm)
$Al_{0.2}Ga_{0.8}N$ /GaN	0.04364
$Al_{0.7}Ga_{0.3}Sb$ /InAs	0.021920
$Al_{0.7}Ga_{0.3}Sb$ /$Ga_{0.85}In_{0.15}Sb$	0.02622
$Al_{0.7}Ga_{0.3}Sb$ /GaSb	0.034

In fig. 5 we see the effects of changing the workfunction

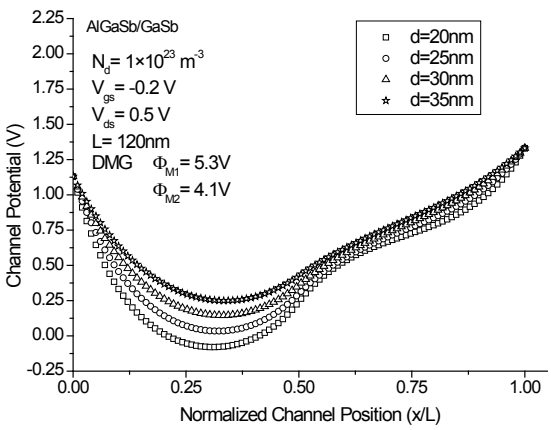


Fig.4: variation of the channel potential with the normalized channel position for different values of barrier layer thickness for $Al_{0.7}Ga_{0.3}Sb$/GaSb DMG HEMT.

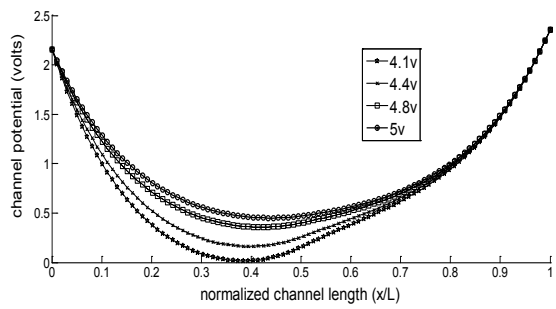


Fig.5: variation of channel potential with normalized channel position for DMG $Al_{0.7}Ga_{0.3}Sb$ /$Ga_{0.85}In_{0.15}Sb$ HEMT for different workfunction difference in between the gate metals.

difference between the two metal of the gate. The workfunction of the M1 (φ_1=5.3 V) is kept fixed, while M2 (φ_2=4.1, 4.4, 4.8, 5 V) is varied to vary the workfunction difference between two metal of the gate. As the workfunction difference increases, the φ_{limin} decreases. This means lower drive current. But it reduces the short channel effects and improves carrier transport efficiency [8, 10]. As a result, the overall (I_{on}/I_{off}) improves. In table VI the rate of change in φ_{limin} for change in workfunction difference of different formations of DMG HEMTs are listed.

Table VI. Rate of change in φ_{limin} for change in workfunction difference

HEMT composition	Rate of change in φ_{limin}(volt/volt)
$Al_{0.2}Ga_{0.8}N$ /GaN	0.0152
$Al_{0.7}Ga_{0.3}Sb$ /InAs	0.4803
$Al_{0.7}Ga_{0.3}Sb$ /$Ga_{0.85}In_{0.15}Sb$	0.554
$Al_{0.7}Ga_{0.3}Sb$ /GaSb	0.5754

From the table VI we can see that rate of change of φ_{limin} is lower in nitride-based HEMTs while higher in antimonide-based HEMTs. DMG structure is all about the difference in workfunctions between two metals of the gate. So DMG structure is more effective in antimonide-based HEMTs compared to their nitride-based counterpart. Among the antimonide-based DMG formations the swallow heterojunction $Al_{0.7}Ga_{0.3}Sb$ /GaSb gives the highest rate of change of φ_{limin} and thus more effective than deep heterojunction formations of $Al_{0.7}Ga_{0.3}Sb$ /InAs. This conclusion was also made in our previous paper [15] comparing other set of characteristics of DMG HEMT.

IV. Conclusions

To solve the problem of short channel effect and carrier transport inefficiency of the submicrometer range HEMTs, a new structure called double material gate (DMG) was proposed. But with the advent of fabrication technology, even the submicron range HEMT has become a field of active research. Thus it is extremely necessary to investigate the effect of variation of physical parameters on DMG HEMT. In this paper we varied three key physical parameters of DMG HEMT namely control gate length (L1), barrier layer thickness (d) and workfunction difference between two gate materials. We have used four different formations of DMG HEMT and compared the results to determine the effects. We have seen that nitride-based DMG HEMTs are less sensitive to the variation of control-gate length compared to antimonide-based DMG HEMTs. We have also seen that as the conduction band discontinuity at the heterojunction increases, the effect of control-gate length change also increases due to increase in 2-D electron gas. It was observed that the effect of changing barrier layer is more prominent in nitride-based DMG HEMT due to its low dielectric constant of the barrier layer. In case of varying the workfunction difference between two gate metals, we observed that increase in the workfunction difference has more prominent effect on antimonide-based HEMTs than nitride-based HEMT which proves that antimonide-based HEMTs will react more strongly to double material gate (DMG) structure. Also the effect of workfunction difference increases as conduction band discontinuity at the heterojunction between barrier and channel layer reduces. Thus DMG structure is more effective in the shallow heterojunction formations like $Al_{0.7}Ga_{0.3}Sb$/GaSb compared to deep heterojunction formations of $Al_{0.7}Ga_{0.3}Sb$ /InAs. We only considered semi-classical model in this work. An improved version of this work can be produced using Quantum Electrodynamics Model.

Acknowledgment

We would like to take this opportunity to thank Mr. Tonmoy Kumar Bhowmick for for his useful advices. This work was supported under BUET grant no. CASR#218, agenda 97, meeting held on 06/09/2009.

References

[1] D. A. Muller, T. Sorsch, S. Moccio, F. H. Baumann and G. Timp, “The electronic structure at the atomic scale of ultra-thin gate oxides”, *Nature*, volume **399**, 758-761 (1999).

[2] Rashmi, Abhinav Kranti, S. Haldar, R.S. Gupta, “An accurate charge control model for spontaneous and piezoelectric polarization dependent two-dimensional electron gas sheet charge density of lattice-mismatched AlGaN/GaN HEMTs”, *Solid-State Electronics*, Volume 46, Issue 5, Pages 621-630, May 2002.

[3] I. P. Smorchkova, C. R. Elsass, J. P. Ibbetson, R. Vetury, B. Heying, P. Fini, E. Haus, S. P. DenBaars, J. S. Speck and U. K. Mishra, “Polarization-induced charge and electron mobility in AlGaN/GaN heterostructures grown by plasma-assisted molecular-beam epitaxy” , *Journal of Applied Physics*, vol.86, pp. 4520-4526, 1999

[4] Brian R. Bennett, Richard Magno, J. Brad Boos, Walter Kruppa, Mario G. Ancona, “Antimonide-based compound semiconductors for electronic devices: A review”, *Solid-State Electronics*, Volume 49, Issue 12, Pages 1875-1895,December 2005.

[5] C. Liu, Y. Li and Y. Zeng, "Progress in Antimonide Based III-V Compound Semiconductors and Devices," *Engineering*, vol. 2 no. 8, pp. 617-624, 2010.

[6] Awano, Yuji; Kosugi, M.; Kosemura, Kinjiro; Mimura, T.; Abe, Masayuki, "Short-channel effects in subquarter-micrometer-gate HEMTs: simulation and experiment," *Electron Devices, IEEE Transactionson* , vol.36, no.10, pp.2260-2266, Oct 1989.

[7] Palacios, T.; Chakraborty, A.; Heikman, S.; Keller, S.; DenBaars, S.P.; Mishra, U.K., “AlGaN/GaN high electron mobility transistors with InGaN back-barriers,” Electron Device Letters, IEEE , vol.27, no.1, pp.13-15, Jan. 2006.

[8] Sona P. Kumar, Anju Agrawal, Rishu Chaujar, Mridula Gupta, R.S. Gupta, “ Analytical modeling and simulation of subthreshold behavior in nanoscale dual material gate AlGaN/GaN HEMT,”*Superlattices and Microstructures*, Volume 44, Issue 1, Pages 37-53,July 2008.

[9] Wei Long; Ou, H.; Kuo, J.-M.; Chin, K.K., "Dual-material gate (DMG) field effect transistor," *Electron Devices, IEEE Transactions on* , vol.46, no.5, pp.865-870, May 1999.

[10] Arman-Ur-Rashid; Md.Aynal Hossain; Rahman, Tanvir; Mohammedy, Farseem Mannan, "Analytical modeling and simulation of subthreshold behavior of dual material gate (DMG) Al0.7Ga0.3Sb/InAs HEMT," *Informatics, Electronics & Vision (ICIEV), 2014 International Conference on* , vol., no., pp.1,4, 23-24 May 2014.

[11] A. Godoy, J.A. López-Villanueva, J.A. Jiménez-Tejada, A. Palma, F. Gámiz, “A simple subthreshold swing model for short channel MOSFETs”, *Solid-State Electronics*, Volume 45, Issue 3, Pages 391-397, March 2001.

[12] Chaudhry, A.; Kumar, M.J., “Controlling short-channel effects in deep-submicron SOI MOSFETs for improved reliability: a review,” *Device and Materials Reliability, IEEE Transactions on* , vol.4, no.1, pp.99,109, March 2004.

[13] A.F.M Anwar, RT Webster, “Energy bandgap of $Al_xG_{1-x}As_{1-y}Sb_y$ and conduction band discontinuity of $Al_xGa_{1-x}As_{1-y}Sb_y$/InAs and $Al_xGa_{1-x}As_{1-y}Sb_y$/InGaAs heterostructures,”*Solid-State Electronics*, Volume 42, Issue 11, Pages 2101-2104, November 1998.

[14] X. Zhou and W. Long, “A novel hetero-material gate (HMG) MOSFET for deep-submicron ULSI technology,” *IEEE Trans. Electron Devices*, vol. 45, pp. 2546–2548, Dec., 1998.

[15] Arman-Ur-Rashid, Md. Aynal Hossain, Tanvir Rahman, Farseem M. Mohammedy, “Comparative study of subthreshold characteristics of different antimonide-based and nitride-based Dual Material Gate (DMG) HEMTs,” accepted at the *11th Int. Conf. on Semiconductor Electronics (ICSE), Kuala Lumpur, Malaysia, 2014*.

Estimating the Transit Time of an $Al_yGa_{1-y}As$ HBT in the Gaussian Doped Base Region for Low to Moderately High Level of Injection

Nusrat Jahan[1], Rowshon Ara Mannan[1], Priyanka Biswas[1], and Yeasir Arafat[2,*]

[1]Dept. of Electrical, Electronic & Communication Engineering, Military Institute of Science and Technology
[2]Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
*arafat@eee.buet.ac.bd

***Abstract*—For a Gaussian doped graded base *AlGaAs* Hetero -junction Bipolar Transistor (HBT), base transit time (BTT) has been determined using a suitable analytical model. Here the authors have investigated the dependency of base transit time on different physical and input parameters such as the collector current density, base emitter voltage etc. Electric field dependent carrier mobility also has been included into the analytic model and found that the base transit time worsens compared to the value found without taken the field dependent mobility into account.**

***Keywords*—Base transit time of AlGaAs HBT; Graded base; Gaussian doping; Field dependent mobility.**

I. Introduction

In this work, $AlGaAs$ has been chosen for its higher speed compared to SiGe HBT under normal operating condition. $SiGe$ HBTs had been introduced as the technology of choice for RF integrated circuits due to its high speed performance and mature silicon process [1]. This technology is also being used to develop electronics for space applications because of its excellent analog and radio frequency (RF) performance over an extremely wide range of temperatures [2]. On the other hand, $AlGaAs$ HBT process technology is reported to be used in the fabrication of large number of complex analog integrated circuits for commercial and military applications [3].

The transit time of a junction transistor has three components: emitter transit time, base transit time and collector transit time. Since the base transit time contributes to more than 70% of the junction transistor's total transit time [4], its speed is mainly determined by the factors which affects the base transit time of the device. The base transit time of a heterojunction bipolar transistor is influenced by many physical and electrical parameters and various other factors. These factors include base width, doping concentration profiles in the base region (uniform, exponential, Gaussian), various alloy material profiles in the base region (box, triangular, trapezoidal), velocity saturation effect, collector current density, electric field and doping dependent carrier mobility, bandgap narrowing and base resistance. In this paper the authors have investigated trend of some physical parameter, the effect of some of these parameters on the base transit time of an $AlGaAs$ HBT.

II. Basis of Analysis

For a trapezoidal Al profile inside the *npn* base of an $Al_yGa_{1-y}As$ HBT, the mole fraction y will be $m_{Al}x + y_E$ where, W_B is the base width, x is the distance of a point in the base from the base side of the *B-E* junction, $m_{Al} = (y_C - y_E)/W_B$; y_C and y_E are Al fraction at the collector and emitter end respectively. For y_E=0, and y_E=y_C, the profile becomes triangular and box shape respectively [5]. Total Al content within the base is called Al dose and can be calculated as $y_D = y_{av}W_B$ [6], where, $y_{av} = (y_C + y_E)/2$. The base transit time of an arbitrary HBT can be expressed as [7]

$$\tau_B = q\int_0^{W_B} \frac{n(x)}{J_n(x)}dx \tag{1}$$

where, $n(x)$ is the minority carrier concentration and $J_n(x)$ is the electron current density within the base, q is the charge of an electron. For the base region, $n(x)$ can be determined from the transport equations (for electron & hole) and electric field equation [8]. As the base width of today's high-speed HBT is even less than 30nm [6], so the carrier recombination in the base region can safely be neglected, resulting in J_p (hole current density) approximately zero and eventually J_n equals to J_c (collector current density). Using Einstein's relation $D_n(x) = V_T\mu_n(x)$, the current density becomes

$$-J_n = qD_n(x)\frac{n_{ie}^2(x)}{n(x)+N_B(x)}\frac{d}{dx}\left[\frac{n(x)\{n(x)+N_B(x)\}}{n_{ie}^2(x)}\right] \tag{2}$$

where, $D_n(x)$ and $\mu_n(x)$ are the electron diffusivity and mobility respectively, $n_{ie}(x)$ is the effective intrinsic carrier concentration. $N_B(x) = N_B(0)\exp(-mx^2)$ is the base doping concentration assumed to follow Gaussian distribution [9], $N_B(0)$ is the peak base doping concentration; $m = \frac{1}{2\sigma^2}$ and $\sigma = W_B/\sqrt{2\ln\frac{N_B(0)}{N_B(W_B)}}$. Under low electric field, the doping density dependent diffusivity in $GaAs$ can be written in the light of [10] as $D_{nGaAs}(x) = D_n\exp(m_1x^2)$, where, $D_{nl} = D_{n0}(N_B(0)/N_r)^{-\gamma_1}$; $D_{n0} = 207\ cm^2/s$[11]; $m_1 = m\gamma_1$; $\gamma_1 = 0.42$ and $N_r = 10^{17}cm^{-3}$. In presence of Al, the electron diffusivity is assumed to be changed following [12] as $D_{nAlGaAs}(x) = bD_{nGaAs}(x)$, where, $b = 1 + 3y_{av}$. The saturation velocity inside the $AlGaAs$ alloy also will differ from that in $GaAs$ and can be written in the light of [13] as $v_{sA} = cv_s$, where, $c = 0.342/[0.342 + y_{av}(1 + y_{av})]$; and $v_s = 7\times10^6\ cm/s$[14].

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

The effective intrinsic concentration in $AlGaAs$ in the light of [12] is $n_{ieAlGaAs}^2(x) = n_{ioGaAs}^2 \gamma_r e^{\frac{\Delta E_{geff}(x)}{kT}}$, where γ_r is the ratio of the effective DOS in $AlGaAs$ base to the effective DOS in $GaAs$ and can be written as $\gamma_r = \exp(-\sqrt{5y_{av}})$ [15], [16]. $n_{ioGaAs} = 2.1 \times 10^6 cm^{-3}$, is the intrinsic carrier concentration in undoped $GaAs$ [14]; k is the Boltzmann constant; T is the temperature in degrees Kelvin; $E_{geff}(x)$ is the effective bandgap reduction in the $GaAs$ base that can be expressed as $E_{geff}(x) = \Delta E_{gHD}(x) + \Delta E_{gAl}(x)$, where $\Delta E_{gAl}(x)$ is the BGN due to the presence of Al which is assumed to have a linear dependency on mole fraction; $\Delta E_{gHD}(x)$ is the BGN due to heavy doping effects and is identical to that of $GaAs$ [15]. An approximation of the Slotboom–de Graff BGN model [17] is used for this term [18] resulting $\Delta E_{gHD}(x) = qV_{gHD} \ln\left(\frac{N_B(x)}{N_r}\right)$, with, V_{gHD}= 0.0229V and the BGN due to the presence of Al can be written as [19] $\Delta E_{gAl}(x) = qV_{gAl}y(x)$, with, V_{gAl}= 10E-10V. Finally, the effective intrinsic concentration in $AlGaAs$ can be expressed as

$$n_{ieAlGaAs}^2(x) = n_{ieAlGaAs}^2(0)\exp\,(m_3x - m_2x^2) \quad (3)$$

where, $n_{ieAlGaAs}^2(0) = n_{ioSi}^2\, \gamma_r \exp\,(\gamma_3 y_E) \left(\frac{N_B(0)}{N_r}\right)^{\gamma_2}$;$\gamma_2 = V_{gHD}/V_T$; $\gamma_3 = V_{gAl}/V_T$; $V_T = kT/q$; $m_2 = m\gamma_2$; $m_3 = m_{Al}\gamma_3$. Under low level of injection, $n(x) \ll N_B(x)$ and the quasi-neutral condition in base becomes $p(x) = N_B(x)$ [20]. Putting this condition, the electric field equation for low injection becomes $E_{lAlGaAs}(x) = -V_T(m_3 + m_4x)$, where $m_4 = 2(m - m_2)$. Using the equations for $n_{ieAlGaAs}^2(x)$, $N_B(x)$ and $D_{nAlGaAs}(x)$, differential equation (2) for minority carrier concentration $n_l(x)$ can be solved for $n_l(x)$ as [21]

$$n_l(x) = exp\left(m_4\frac{x^2}{2} + m_3x\right)\left(n_l(0) - J_{nl}(B_1(x) + B_2(x) + B_3(x))\right) \quad (4)$$

At $x = 0$ (E-B boundary), electron concentration can be found by using [22] with necessary correction for $AlGaAs$ as-

$$n_l(0) = \frac{n_{ieAlGaAs}^2(0)}{N_B(0)} e^{\frac{V_{be}}{V_T}} \times f_w \quad (5)$$

$$\text{where, } f_w = \frac{1}{0.5 + \sqrt{0.25 + \frac{n_{ieAlGaAs}^2(0)}{N_B(0)} e^{\frac{V_{be}}{V_T}}}} \quad (5a)$$

Usage of this initial condition ensures the inclusion of Webster effect [23]. Putting $x = W_B$ in (4), the electron concentration at B-C boundary becomes

$$n_l(W_B) = \frac{n_l(0)exp\left(m_4\frac{W_B^2}{2} + m_3W_B\right)}{1 + qv_{sA}exp\left(m_4\frac{W_B^2}{2} + m_3W_B\right)(B_1(W_B) + B_2(W_B) + B_3(W_B))} \quad (6)$$

Assuming velocity saturation at the B-C depletion region, the low injection electron current density at $x = W_B$ will be $J_{nl} = qv_{sA}n_l(W_B)$, the base stored charge per unit area $Q_{Bnl} = q\int_0^{W_B} n_l\,(x)dx$, and the base transit time $\tau_{Bl} = \frac{Q_{Bnl}}{J_{nl}}$.

This value of low injection can be extended to incorporate moderate level of injection with slight modification using perturbation theory. The modified carrier concentration and current density for moderate injection are then, $n_m(x) = f_W n_l(x)$ and $J_{nm} = f_W J_{nl}$ respectively. Assuming that the injected electron concentration is perturbed by the modulated electric field from $n_m(x)$ only a little and also using quasi-neutrality of charge equation, minority electron concentration for moderate injection will be $n(x) = n_m(x) + \delta n(x)$ and hole concentration $p(x) = n(x) + N_B(x)$ becomes-

$$p(x) = p_m(x) + \delta n(x) \quad (7)$$

Now, (2) can be re-solved to find $n(x)$ for moderate injection, [9]

$$n(x) = \frac{\exp\,(m_3x - m_2x^2)}{p_m(x)}\left(n(0)p_m(0) + \frac{J_n}{qv_{sA}}N(x)\right) \quad (8)$$

where, $N(x) = \frac{a}{b}N_1(x) - \frac{v_{sA}}{bD_n}N_2(x)$; $N_1(x) = n_m(0)\{\exp(mx^2) - 1\} - J_{nm}\{B(x)\exp(mx^2) - B(0)\} + N_B(0)\left\{exp\left(-\frac{m_4}{2}x^2 - m_3x\right) - 1\right\}$; $N_2(x) = n_m(0)\{E_1(x) - E_1(0)\} - J_{nm}\{E_2(x) - E_2(0)\} + N_B(0)\{E_3(x) - E_3(0)\}$.

Boundary value of injected minority carrier at the collector edge, $n(W_B)$ can be found by putting $x = W_B$ in equation (8), and hence J_n for moderate injection can be calculated. To estimate base transit time, equation (1) has been utilized.

III. Result

To demonstrate the dependency of base transit time on various physical and input parameters, some typical relations are plotted in this section of the paper. Normalized injected electron concentration depends on the slope of base doping and is illustrated in Fig. 1. As the value of $N_B(W_B)$ increases, slope of base doping decreases. With the increase in slope of the base doping, $n(x)/n(0)$ decreases more sharply. The injected electron concentration profiles shown in Fig. 1 is not linear due to the built in electric field caused by the Gaussian base doping and the gradient in Al profile. It is also found that the minority carrier has maximum value at the base-emitter junction and minimum value at the collector-base junction. It decreases from its peak value as the distance from the emitter increases. At the collector-base junction the structure having the highest slope of base doping has the maximum normalized carrier concentration. In this work velocity saturation of electron at the collector-base junction is considered. So at collector-base junction, the minority carrier concentrations for low and high levels of injection are not zero.

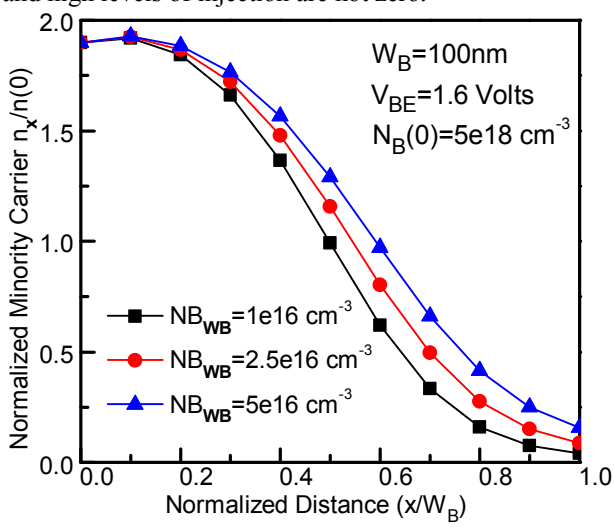


Fig. 1 Normalized minority carrier electron distribution within the base for different slopes of base doping with trapezoidal Al profile $y_C = 0.25$, $y_E = 0.1$.

The variation of collector current density J_C with base-emitter voltage V_{be} for three different shapes of Al profiles is

shown in Fig. 2 for low injection. Changing the value of y_E, triangular, trapezoidal or box Al profile effects are incorporated. J_C increases with V_{be} maintaining a constant logarithmic slope; with the increase in V_{be}, the slope of the curve decreases. Collector current density for J_C triangular shape Al profile is higher than that of box profile and for trapezoidal profile it remains in the middle. This is expected as the injected electron concentration at the collector base boundary ($n(W_B)$) is the highest for triangular Al profile and higher $n(W_B)$ contributes to higher J_C, which agrees with [24].

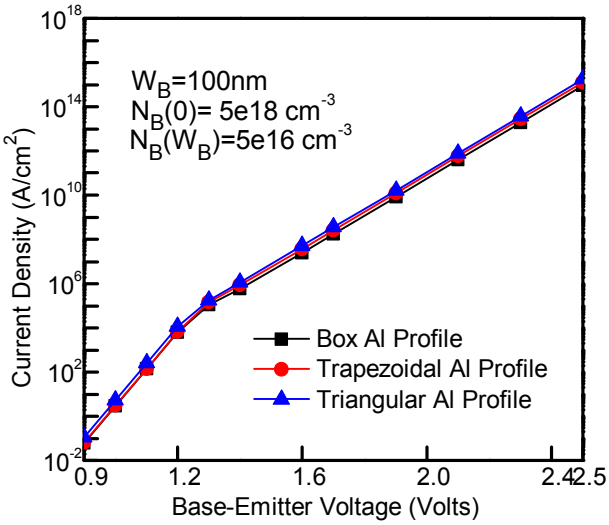


Fig. 2 Collector current densities for different shape of Al profiles (Box: $y_C = 0.25, y_E = 0.2499$, Trapezoidal: $y_C = 0.25, y_E = 0.1$, Triangular: $y_C = 0.25, y_E = 0.01$) as a function of base-emitter voltage.

The variation of base transit time with base-emitter voltage for various Al profiles is shown in Fig. 3. It is observed that for both low and moderate levels of injection the base transit time is independent of base-emitter voltage but the value of base transit time is larger for moderate injection region than that of low injection region. This is due to reduction of aiding field in the Gaussian doped base when the level of injection increases. For reduction in aiding field, electrons slow down and hence transit time increases. As V_{be} increases, base transit time increases in the intermediate level of injection.

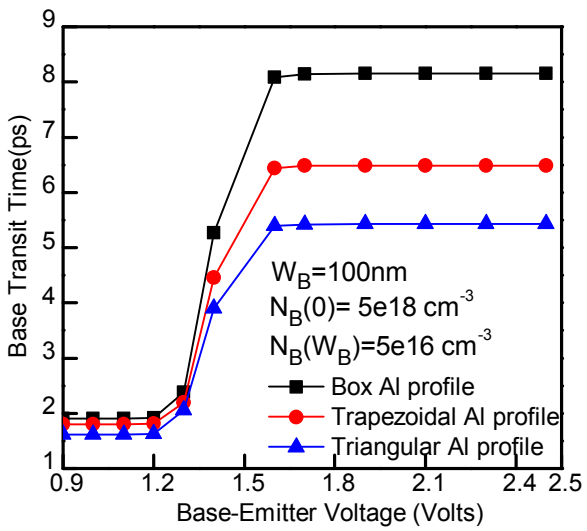


Fig. 3 Base transit time as a function of base-emitter voltage for different shape of Al profiles (Box: $y_C = 0.25, y_E = 0.2499$, Trapezoidal: $y_C = 0.25, y_E = 0.1$, Triangular: $y_C = 0.25,\ y_E = 0.01$).

From Fig. 3, it also can be observed that box shape Al profile takes the highest time to pass the electron to the collector, triangular Al profile takes the lowest time and the trapezoidal Al profile takes in-between time. This happens because the transport of electron through the base is assisted by the gradient in mole fraction profiles [25], which is highest for triangular profile, zero for box profile and in-between for trapezoidal profile. So for triangular Al profile, electron passes slower through the base to enter the collector.

One of the main objectives of this paper is to obtain base transit time of a Gaussian doped base $Al_yGa_{1-y}As$ HBT considering electric field dependence of carrier mobility in excess to doping dependent mobility. The effect of electric field dependence of mobility on base transit time was investigated and the result is shown in Fig. 4. The transit time is found to be higher if the field dependence of mobility is considered. This is because minority carrier mobility is inversely related to the absolute value of electric field [20]. With the increase in electric field, minority carrier mobility in the base region decreases. So, with the increases in electric field, it decreases the minority carrier transportation through the base region and hence contributes to the increase in base transit time.

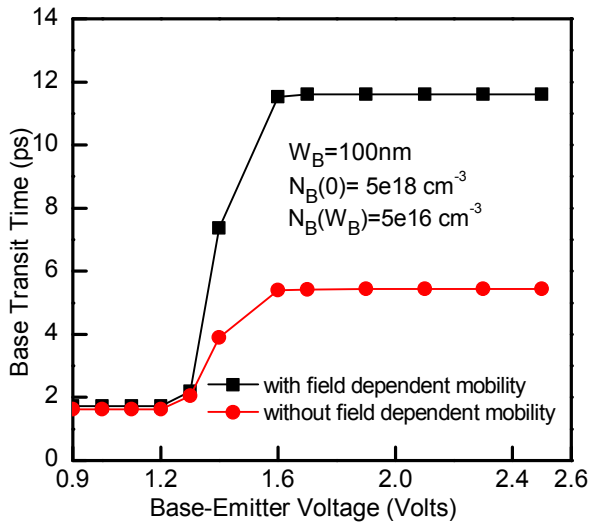


Fig. 4 Base transit time as a function of base-emitter voltage with and without field dependent electron mobility considering triangular Al profile ($y_C = 0.25, y_E = 0.01$).

The variation of base transit time (τ_B) with collector current density for different Al profiles is shown in Fig. 5. From this figure, it is understandable that base transit time is independent of collector current density for both low and moderately high level of injection but the value of base transit time is higher for high injection region than the base transit time for low injection region. This is due to the reduction of aiding electric field in the Gaussian doped base when the level of injection increases. Due to this effect, electron slow down and transit time increases. In the intermediate level of injection, base transit time increases as J_C increases. From the Fig.5, it is also perceptible that box shape Al profile takes higher transit time than triangular Al profile to pass the electron to the collector and the trapezoidal Al profile takes in-between time. This happens due to the difference in gradients of Al profiles. Fig. 5 follows Fig. 3 ($\tau_B\ vs. V_{be}$) as collector current is a logarithmic function of V_{be} (Fig. 2).

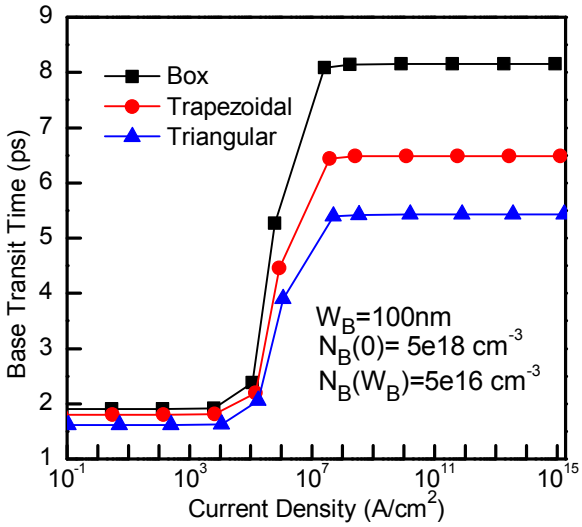


Fig. 5. Base transit time as a function of collector current density for different shape of Al profiles (Box: $y_C = 0.25, y_E = 0.2499$, Trapezoidal: $y_C = 0.25, y_E = 0.1$, Triangular: $y_C = 0.25, y_E = 0.01$).

IV. Conclusion

How the base-emitter voltage, collector current density, and electric field dependent mobility influence the base transit time of a $AlGaAs$ HBT with Gaussian doped base has been investigated for all types (box, trapezoidal and triangular) of Al profiles ranging from low level of injection to moderately high level of injection. It is found that the transit time decreases as the profiles change from box to triangular types. It is also observed that collector current density shows an increasing trend with raising base emitter voltage for similar trend of profile change as previous. If electric field dependent mobility is taken into account, the model gives higher transit time as predicted.

Acknowledgment

The authors of this work would like to express gratitude to the corresponding Departments of their respective Institute (MIST)/University (BUET) for various supports during the preparation of this manuscript.

References

[1] K. Lee, D. H. Cho, K. W. Park, and B. Kim, "Improved VBIC Model for SiGe HBTs With a Unified Model of Heterojunction Barrier Effects," IEEE Tran. on Electron Devices, vol. 53(4), p. 743, 2006.

[2] Z. Xu, G. Niu, L. Luo, P. S. Chakraborty, P. Cheng, D. Thomas, and J. D. Cressler, "Cryogenic RF Small-Signal Modeling and Parameter Extraction of SiGe HBTs," IEEE Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems, SiRF'09, pp. 1-4, Jan 2009.

[3] F. M. Yamada, A. K. Oki, D. C. Streit, D. K. Umemoto, L. T. Tran, et al., "Reliable ICs fabricated using a production GaAs HBT process for military and commercial applications" Military Communications Conference, MILCOM '95, Conference Record, IEEE, vol. 2, pp 760 – 764, Nov 1995.

[4] S. K. Mandal, G. K. Marskole, K. S. Chari, and C. K. Maiti, "Transit time components of a SiGe-HBT at low temperature," Proc. 24th International Conference on Microelectronics, vol. 1, pp. 315-318, May 2004.

[5] Z. R. Tang, T. Kamins, and C. A. T. Salama, "Analytical and experimental characteristics of SiGe HBT with thin α-Si : H emitters," Solid-State Electron., vol. 38, pp. 1829-1834, 1995.

[6] K. H. Kwok and C. R. Selvakumar, "Profile design considerations for minimizing base transit time in SiGe HBTs for all levels of Injection before onset of Kirk effect," IEEE Tran. on Electron Devices, vol. 48, pp. 1540-1549, 2001.

[7] H. Kroemer, "Two integral relations pertaining to electron transport through a bipolar transistor with a nonuniform energy gap in the base region," Solid-State Electron., vol. 28, pp. 1101–1103, 1985.

[8] J. S. Yuan, "Effect of base profile on the base transit time of the bipolar transistor for all levels of injection," IEEE Tran. on Electron Devices, vol. 41, pp. 212-216, Feb 1994.

[9] S. M. M. Islam, M. I. B. Chowdhury, Y. Arafat, and M. Z. R. Khan, "Base Transit Time of a Heterojunction Bipolar Transistor (HBT) with Gaussian Doped Base Under High-Level of Injection," in Proc. of International Conference on Devices, Circuits and Systems (ICDCS), Mar 2012.

[10] K. Suzuki, "Optimum base doping profile for minimum base transit time considering velocity saturation at base-collector junction and dependence of mobility and bandgap narrowing on doping concentration," IEEE Tran. on Electron Devices, vol. 48, pp. 2102-2107, 2001.

[11] M. Shur, Physics Of Semiconductor devices, A Prentice-Hall. Inc, 1990.

[12] S. Basu, "Analytical modelling of base transit time of SiGe HBTs including effect of temperature," International Semiconductor Conference, CAS'08, vol. 2, pp. 339-342, Oct 2008.

[13] S. T. Chang, C. W. Liu, and S. C. Lu, "Base transit time of gradedbase Si/SiGe HBTs considering recombination lifetime and velocity saturation," Solid-State Electron., vol. 48, pp. 207–215, 2004.

[14] S. M. Sze, Physics Of Semiconductor devices, A Wiley Inter science publication, 1981.

[15] V. S. Patri and M. J. Kumar, "Profile Design Considerations for minimizing base transit time in SiGe HBT's," IEEE Tran. on Electron Devices, vol. 45, pp. 1725-1731, Aug 1998.

[16] A. Zareba, L. Lukasiak, and A. Jakubowski, "Modeling of SiGe-base heterojunction bipolar transistor with gaussian doping distribution," Solid-State Electron., vol. 45, pp. 2029-2032, 2001.

[17] J. W. Slotboom and H. C. d. Graaff, "Measurement of bandgap narrowing in Si bipolar transistors," Solid-State Electron., vol. 19, pp. 857–862, 1976.

[18] T. C. Lu and J. B. Kuo, "A closed form analytical BJT forward transit time model considering bandgap narrowing effects and concentration dependent diffusion coefficients," Solid-State Electron., vol. 35, pp. 1374-1377, 1992.

[19] H. Kroemer, "Two integral relations pertaining to electron transport through a bipolar transistor with a nonuniform energy gap in the base region," Solid-State Electron., vol. 28, pp. 1101–1103, 1985.

[20] M. M. S. Hassan and M. W. K. Nomani, "Base transit time model considering field dependent mobility for BJTs operating at high-level injection," IEEE Tran. on Electron Devices, vol. 53, pp. 2532-2539, Oct 2006.

[21] S. M. M. Islam, Y. Arafat, M. I. B. Chowdhury, M. Z. R. Khan, and M. M. S. Hassan "Base Transit Time of a Heterojunction Bipolar Transistor with Gaussian Doped Base", in Proc. of 6th International Conference on Electrical and Computer Engineering ICECE 2010, Dhaka, Bangladesh, December 2010.

[22] K. Suzuki, "Analytical base transit time model of uniformly-doped base bipolar transistors for high-injection regions," Solid-State Electron., vol. 36, pp. 109-110, 1993.

[23] W. Webster, "On the variation of junction-transistor current amplification factor with emitter current," Proc IRE, vol. 42(6), pp. 914-920, 1954.

[24] M. K. Das, N. R. Das, and P. K. Basu, "Effect of Ge content and profile in the SiGe base on the performance of a SiGe/Si HBT," Microwave and Optical Technology Letters, vol. 47, pp. 247-254, Nov 2005.

[25] J. D. Cressler and G. Niu, Silicon-Germaneum Heterojunction Bipolar Transistor. Norwood, MA: Artech House, 2003.

An Analytical Model of Minority Carrier in Exponentially Doped Solar Cell under Illumination

Bishwajit Debnath, Topojit Debnath, Mokter Mahmud Chowdhury, M. Ziaur Rahman Khan*

Department of Electrical and Electronic Engineering,
Bangladesh University of Engineering and Technology
Dhaka-1205, Bangladesh
*zrkhan@eee.buet.ac.bd

Abstract— **A compact analytical model of minority carrier concentration is presented for an exponentially-doped quasi-neutral region of Silicon solar cell. It is an iteration-free and integral-free mathematical expression that addresses the non-uniformity in doping and contribution from terrestrial solar spectrum. Due to the position dependency of transport parameters, the differential equation becomes intractable. The proposed model incorporates an elegant approximated carrier generation rate to overcome the problem and presents a new solution for minority carrier concentration by using the modified Bessel function and Hypergeometric function. The compact analytical solution is in good agreement with COMSOL drift-diffusion model and TCAD device simulator for a wide variation of doping and surface recombination velocity.**

Index Terms— **Analytical model, Exponentially doped emitter, Approximated generation rate, Compact solution.**

I. Introduction

An extensive study of minority carrier profile is essential for successful understanding of the performance and efficiency of a solar cell under illumination [1]. An obvious step in modelling solar cell is to analyse the transport phenomena of minority carrier as a function of doping profile and surface recombination velocity. The difficulty in developing analytical model for non-uniformly doped solar cell arises from the fact that the minority carrier transport parameters become dependent on position. Incorporation of these non-linear dependencies in the continuity equation and the minority carrier current equation makes the differential equation complicated. Hence, numerical analysis has become a common approach for describing the behaviour of minority carriers in solar cells. Although numerical model of solar cell, TCAD-based models [2, 3] and efficient computer codes like PC-1D, AMPS [4] can address the photovoltaic physics reliably, a complete analytical model helps to gain insight into the physics. Unlike numerical simulation, the mathematical analysis offers an excellent way to study the effects from specific contributions, which can be invaluable in computer aided design and device optimization. Besides, the analytical modelling has added advantages of faster calculation time and no convergence issues.

Under dark condition (no sunlight), several authors [5, 6] have analytically solved the minority carrier transport, taking into account the position dependencies. Verhoef *et al.* [5] has developed a closed-form expression for minority carrier transport by proposing new empirical expression of carrier mobility and lifetime; but the expression is developed excluding the carrier generation rate, rather focusing only on the dark saturation current. Analytical approach for illuminated solar cell also exists in literature [7-10], but these works contain multiple integral asymptotic series, which are not in closed-form and only accurate after certain number of iterations [11]. In this work, an analytical model is presented, which does not need any iteration or numerical integration. For developing this compact model, the carrier generation rate from solar excitation will be approximated by position-dependent exponential terms [12]. Along with that, the effects of doping-dependent lifetime and mobility are included in the transport equations by using the existing empirical equations [5]. To validate the developed analytical solution, a drift-diffusion based Finite Element Model (FEM) is implemented in COMSOL multi-physics. Besides, this work is compared with a 2-D numerical solar cell model created in Silvaco/Atlas device simulator, which incorporates Auger recombination, concentration-dependent Shockley-Read-Hall recombination and doping-dependent mobility.

II. General Solution for Minority Carrier Density

An exponentially doped n-type emitter region for an n(x)-p silicon solar cell, as shown in Fig. 1, is considered for analysis. This drift-field (DF) solar cell structure will be used to establish this model. Unlike previous works in [1, 7], this work will avoid any iterative or numerical scheme. The main focus of this work is to develop the analytical model for certain doping profile using advance mathematical functions, but the solution will be in a closed-form expression of position and spectrum. Contrary to existing literature, both lifetime and mobility variation with external excitation from solar illumination will be included, resulting in an all-in-one compact model. Since all transport parameters are functions of depth [9, 10], the analysis is considered for one dimension only. Besides, the minority carrier continuity and current equations are solved considering the quasi-neutrality and low-injection condition.

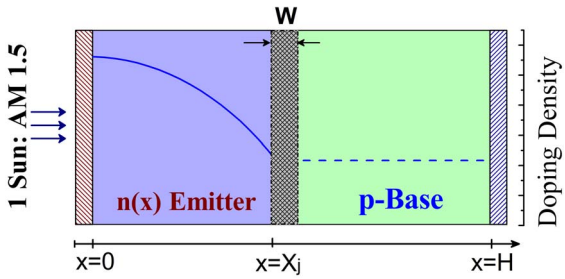


Fig. 1 A DF (Drift Field) solar cell structure

In steady state, the transport equations for excess minority carriers in the n-type emitter region are expressed by (1)-(2),

$$J_p(x) = J_{\text{Diffusion}}(x) + J_{\text{Drift}}(x) = -qD_p \frac{dp}{dx} + q\mu pE \quad (1)$$

$$\frac{dJ_p}{dx} = -q\frac{p}{\tau_p} + qG \quad (2)$$

Here, J_P denotes hole-current density, τ_P is the minority carrier recombination lifetime, μ is carrier mobility, D_P is hole diffusivity, p is the minority carrier (Hole) density and E is the electric field introduced due to non-uniform doping in emitter. Another important parameter is carrier generation rate (G), which is the result of exposed solar illumination. Generation rate is a function of depth into the material (x) and wavelength (λ) of incident solar spectrum. G(x) is generally considered as a sum of individual contributions over the whole solar spectrum, but it can be well approximated by a series of three to five exponential terms [13]. The approximated generation rate of minority carrier can be expressed by,

$$G(x) = \sum_{i=1}^{n} a_i e^{-b_i x} \tag{3}$$

The accuracy and validity of this approximation is already established in literature [12, 13] for different photovoltaic materials and different solar spectrums. For this model, standard terrestrial spectrum AM1.5G will be considered. The values of a_i and b_i recalculated for this work are mentioned in fig. 2. As seen in fig. 2, the approximated G(x) has good match with Silvaco/Atlas TCAD model.

Since the emitter is non-uniformly doped, lifetime (τ) and mobility (μ) is a function of doping (and hence, depth). The dependence of τ and μ on doping N is accounted by the following empirical relations [5] –

$$\mu(N) = MN^{-m} \tag{4}$$

$$\tau(N) = KN^{-k} \tag{5}$$

The above empirical equations predict the Experimental data [14] as a function of N quite well: k=1, K=1.3x10^{12} cm^{-3}/s for $10^{15}<N<10^{18}$ cm^{-3} and k=2, K=5.5x10^{30}cm^{-6}/s for $5\times10^{18} \le N<10^{20}$ cm^{-3}. This form of expression is common in literature. The model here uses these general parameters K, k, M and m to address the position dependency of mobility and lifetime in the final solution. Hence, the same model can be used to analyse a wide range of photovoltaic materials if the parameters are tailored appropriately for other doping profile and material properties.

The electric field associated with the exponential doping is taken into account by equation (6) for quasi-neutrality condition. The effect from doping gradient is included in the first term, but the band gap narrowing is ignored, i.e. $\Delta E_G \cong 0$.

$$E = -V_T \frac{d\ln N}{dx} + \frac{1}{q}\frac{d\Delta E_G}{dx} \tag{6}$$

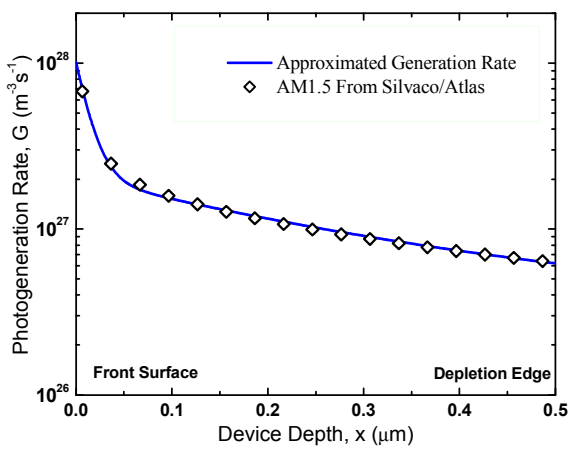


Fig. 2 Carrier generation rate (AM1.5G) along the depth. Solid line represents the approximated carrier generation in (3). Generation parameters – a_i ($\times10^{20}$ s^{-1}cm^{-3}): 78.1, 15, 4.8, 0.8, and 0.0011; b_i ($\times10^4$ cm^{-1}): 75.44, 4.19, 0.6, 0.052 and 0.0015; Symbol represents data from TCAD Device simulator.

When the emitter region is fabricated by implantation followed by a drive-in anneal, the doping profile follows an exponential dependence with depth. We are considering such an exponential doping in equation (7).

$$N(x) = N_{d0} e^{-\alpha x} \tag{7}$$

The peak doping concentration (N_{d0}) and exponential decay (α) are positive quantities. As mentioned, the lifetime and mobility will also vary according to equation (4)-(5). Both lifetime and mobility decrease along x and unlike previous analytical models, the variation of G(x) (fig. 2), N(x), μ(x), τ(x) will be considered together for developing a compact expression.

By placing equation (3)-(6) into equation (1)-(2) and after some mathematical manipulations, the minority carrier hole concentration p(x) can be expressed in terms of N.

$$\frac{d^2p}{dx^2} - (m-1)\frac{dp}{dx}\frac{d\ln N}{dx} - \left[m\left(\frac{d\ln N}{dx}\right)^2 - \frac{d^2\ln x}{dx^2} + \frac{N^{m+k}}{KMV_T}\right]p = -\left(\frac{N^m}{MV_T}\right)\sum_{i=1}^{5} a_i e^{-b_i x} \tag{8}$$

Equation (8) is a second-order non-homogeneous differential equation. By inserting N(x) from (7) into (8), the differential equation can be expressed as a function of device depth, x.

$$\frac{d^2p}{dx^2} + \alpha(m-1)\frac{dp}{dx} - \left[m\alpha^2 + \frac{N_{d0}{}^{m+k} e^{-\alpha(m+k)x}}{KMV_T}\right]p = -\left(\frac{N_{d0}{}^{m} e^{-\alpha m x}}{MV_T}\right)\sum_{i=1}^{5} a_i e^{-b_i x} \tag{9}$$

Substituting $z = \exp\left[-\alpha(m+k)x/2\right]$ will result in (10).

$$z^2\frac{d^2p}{dz^2} + (1-2A)z\frac{dp}{dz} - \left(\beta^2 z^2 + \kappa^2\right)p = -\sum_{i=1}^{5} C_i z^{D_i} \tag{10}$$

Here,

$$A = \left(\frac{m-1}{m+k}\right),\quad \beta = \frac{2}{\alpha(m+k)}\sqrt{\frac{N_{d0}{}^{m+k}}{KMV_T}},\quad \kappa = \left(\frac{2\sqrt{m}}{m+k}\right)$$

$$C_i = \frac{4N_{d0}{}^{m}}{MV_T\alpha^2(m+k)^2}a_i,\qquad D_i = \frac{2(b_i+\alpha m)}{\alpha(m+k)} \tag{11}$$

Finally introducing $p(z) = z^A u(z)$, equation (10) can be reduced to a non-homogeneous Bessel equation of order ν.

$$z^2\frac{d^2u}{dz^2} + z\frac{du}{dz} - \left(\beta^2 z^2 + \nu^2\right)u = -\sum_{i=1}^{5} C_i z^{\gamma} \tag{12}$$

Here, $\nu = \sqrt{A^2+\kappa^2} = (m+1)/(m+k)$, which is a non-integer number and $\gamma = (D_i - A)$. The homogeneous solution of (12) can be easily obtained in terms of modified Bessel function of first kind.

$$u(z) = C_1 I_\nu(\beta z) + C_2 I_{-\nu}(\beta z)$$

To obtain the particular solution of u(z) considering the non-homogeneous term in the right side of equation (12), the variation of parameter technique [15] is employed. Eventually substituting $p(z) = z^A u(z)$ gives the desired minority carrier profile as a function of z.

$$p(z) = P_1(z)\,C_1 + P_2(z)\,C_2 + P_3(z) \tag{13}$$

With,

$$P_1(z) = z^A I_\nu(\beta z),\ P_2(z) = z^A I_{-\nu}(\beta z)$$

$$P_3(z) = -\sum_{i=1}^{5} \frac{E_i \pi}{2\sin(\pi\nu)} \times z^A$$

$$\left[I_\nu(\beta z) \frac{2^\nu (\beta z)^{\gamma-\nu}}{(\gamma-\nu)\Gamma(-\nu+1)} {}_1F_2\left(\frac{\gamma-\nu}{2}; -\nu+1, \frac{\gamma-\nu+2}{2}; \frac{\beta^2 z^2}{4}\right) \right.$$

$$\left. - I_{-\nu}(\beta z) \frac{2^{-\nu} (\beta z)^{\gamma+\nu}}{(\gamma+\nu)\Gamma(\nu+1)} {}_1F_2\left(\frac{\gamma+\nu}{2}; \nu+1, \frac{\gamma+\nu+2}{2}; \frac{\beta^2 z^2}{4}\right) \right]$$

Here, I_ν and $I_{-\nu}$ are the modified Bessel functions of the first kind and ${}_1F_2$ is the generalized Hypergeometric function. $\Gamma\nu$ denotes the Gamma function of ν. C_1 and C_2 are constants which will be determined by using proper boundary conditions.

III. Results & Analysis

To analyse and understand the proposed mathematical model, an emitter with exponentially doping profile is considered under AM1.5 solar spectrum. The emitter doping at the front surface (x=0) is higher (10^{20} cm^{-3}) and then gradually falls to 5×10^{18} cm^{-3} at the depletion edge ($x=x_j$). Due to the range and non-uniformity in doping, lifetime and mobility inside the quasi-neutral emitter will vary intensely along depth. Besides, the quasi-neutral region has doping gradient; hence, carrier transport here will be also dominated by drift component from induced Electric field E(x). The boundary conditions in emitter for the DF cell [5] (fig. 1) is,

$$p(x_j) = \frac{n_i^2}{N(x_j)} \exp\left(\frac{V}{V_T}\right) \tag{14}$$

$$\left[\mu(x) E p(x) - D_p(x) \frac{dp(x)}{dx} \right]_{x=0} = S\, p(0) \tag{15}$$

The boundary conditions can easily be modified for z domain. The 1st boundary condition is evaluated at $z_j = \exp\left[-\alpha(m+k)x_j/2\right]$. The 2nd boundary condition describes front-surface characteristics through surface recombination velocity (S). Equation (15) involves the derivatives of p(z), which means the derivatives of Bessel function and Hypergeometric function are to be calculated first. Using the identities of differentiating Bessel and Hypergeometric function, $dp(z)/dz$ can be obtained as follows,

$$\frac{dp(z)}{dz} = X_1(z)\, C_1 + X_2(z)\, C_2 + X_3(z) \tag{16}$$

Here, $X_1(z) = Az^{A-1} I_\nu(\beta z) + z^A \frac{\beta}{2}\left[I_{\nu-1}(\beta z) + I_{\nu+1}(\beta z)\right]$

$$X_2(z) = Az^{A-1} I_{-\nu}(\beta z) + z^A \frac{\beta}{2}\left[I_{-\nu-1}(\beta z) + I_{-\nu+1}(\beta z)\right]$$

$$X_3(z) = -\sum_{i=1}^{5} \frac{E_i \pi}{2\sin(\pi\nu)} \times \left[F_{+\nu}(z) - F_{-\nu}(z)\right]$$

$F_{\pm\nu}(z)$ is defined in (17) which strongly depends on transport constants (K, k, M, m) and generation coefficients (a_i, b_i).

$$F_{+\nu}(z) = \frac{2^\nu \beta^{\gamma-\nu} z^{\gamma-\nu+A}}{(\gamma-\nu)\Gamma(-\nu+1)} \times$$

$$\left[{}_1F_2\left(\frac{\gamma-\nu}{2}; -\nu+1, \frac{\gamma-\nu+2}{2}; \frac{\beta^2 z^2}{4}\right) \left\{ \beta I_{\nu+1}(\beta z) + \frac{\gamma+A}{z} I_\nu(\beta z) \right\} \right. \tag{17}$$

$$\left. + {}_1F_2\left(\frac{\gamma-\nu+2}{2}; -\nu+2, \frac{\gamma-\nu+4}{2}; \frac{\beta^2 z^2}{4}\right) \frac{\beta^2 z}{2} \frac{(\gamma-\nu)(-\nu+1)}{(\gamma-\nu+2)} I_\nu(\beta z) \right]$$

Replacing the derivative in (15) and after some rigorous mathematical operations, the following equation is obtained:

$$C_1 T_{21} + C_2 T_{22} = T_{23} \tag{18}$$

Where,

$$T_{21}(z) = \left[MN_{d0}^{-m} V_T \alpha z^{\left(\frac{-2m}{m+k}\right)} \left\{ P_1(z) + \left(\frac{m+k}{2}\right) z X_1(z) \right\} - SP_1(z) \right]$$

$$T_{22}(z) = \left[MN_{d0}^{-m} V_T \alpha z^{\left(\frac{-2m}{m+k}\right)} \left\{ P_2(z) + \left(\frac{m+k}{2}\right) z X_2(z) \right\} - SP_2(z) \right]$$

$$T_{23}(z) = \left[SP_3(z) - MN_{d0}^{-m} V_T \alpha z^{\left(\frac{-2m}{m+k}\right)} \left\{ P_3(z) + \left(\frac{m+k}{2}\right) z X_3(z) \right\} \right]$$

These simplified equations obtained from the boundary conditions give the coefficient C_1 & C_2. Replacing C_1 & C_2 back in equation (13) results in a complete analytical expression of Hole profile p(z) for non-uniformly doped emitter of DF solar cell.

$$p(z) = \overbrace{\left[\underbrace{\frac{\frac{n_i^2}{N_d(z_j)} \exp(V/V_T)}{P_1(z_j) - \frac{T_{21}(z_0)}{T_{22}(z_0)} P_2(z_j)}}_{\text{Contribution from Biasing}} + \underbrace{\frac{T_{23}(z_0) + \frac{T_{22}(z_0)}{P_2(z_j)} P_3(z_j)}{T_{21}(z_0) - \frac{T_{22}(z_0)}{P_2(z_j)} P_1(z_j)}}_{\text{Contribution from Spectrum}} \right]}^{C_1} z^A I_\nu(\beta z)$$

$$+ \overbrace{\left[\underbrace{\frac{\frac{n_i^2}{N_d(z_j)} \exp(V/V_T)}{P_2(z_j) - \frac{T_{22}(z_0)}{T_{21}(z_0)} P_1(z_j)}}_{\text{Contribution from Biasing}} + \underbrace{\frac{P_1(z_j)\frac{T_{23}(z_0)}{T_{21}(z_0)} + P_3(z_j)}{\frac{T_{22}(z_0)}{T_{21}(z_0)} P_1(z_j) - P_2(z_j)}}_{\text{Contribution from Spectrum}} \right]}^{C_2} z^A I_{-\nu}(\beta z) \tag{19}$$

$$+ \underbrace{P_3(z)}_{\text{Illumination term}}$$

As seen from the compact expression above, the contribution from biasing, variable doping, surface states, dark component and spectrum response can be identified and separately analysed for optimizing cell design. The success of this work is that together all these contributions, as well as their inter-dependence can be explored effectively using one unified expression of minority carrier.

If the spectral contribution of solar illumination is completely removed, equation (19) simply represents the solution for dark condition, which is identical to equation (11) of Verhoef and Sinke [5]. Once the carrier profile is known, minority carrier dark and illumination current density (with corresponding drift and diffusion parts) is quite straightforward to obtain by simply placing (19) into (1).

The minority carrier profile for a thin (0.5μm) emitter, calculated using the developed model, is shown in fig. 3. As expected, the hole concentration strongly depends on the surface recombination velocity at the front interface. At the edge of emitter-base space-charge region, both the doping concentration and carrier injection for applied bias will determine the hole profile. The developed solution in equation (13) is compared with a Drift-Diffusion based COMSOL Multi-physics Model. The good match with COMSOL model guarantees the soundness of our mathematical solution. The proposed solution of (13) also shows good agreement with the TCAD model developed in Silvaco/Atlas, as depicted in fig 3. The reason behind the slight mismatch near the depletion edge

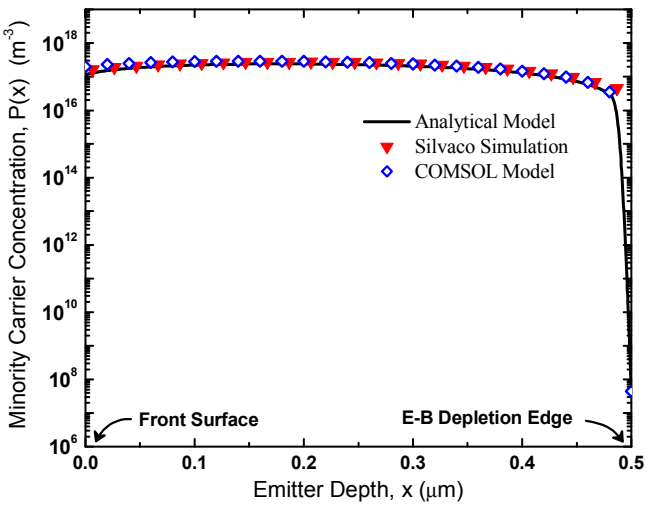


Fig. 3 Minority carrier Hole concentration for n-type Emitter (Depletion edge is at 0.5μm). The front surface recombination velocity (S) is 5×10^4 cm/s. The symbols represent the solution from numerical solar cell models (COMSOL drift-diffusion model and TCAD model) and line indicates analytical solution. Concentration dependent mobility and lifetime have been included in Silvaco/Atlas simulation.

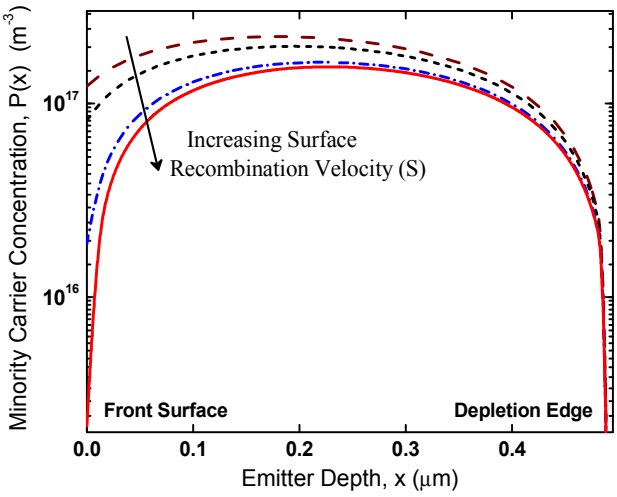


Fig. 4 Effect of surface recombination on Hole concentration. The front surface recombination velocity is varied from 5×10^3cm/s to 5×10^6cm/s.

is that, unlike Silvaco/Atlas, the carrier concentration at junction boundary is solely determined by $p_0 \approx n_i^2/N_D$ under short circuit condition (see equation (14) for applied bias V=0). The model also predicts the reduction of carrier near the top surface when the surface recombination velocity is increasing (fig. 4). In the figure, the minority carrier variation is presented for surface recombination velocities of 5×10^3, 5×10^4, 5×10^5 and 5×10^6 cm/s. While some of the previous analytical models only work well for either high S or low S [9], this analytical approach covers a wide variation of surface recombination velocity. As evident in the fig. 4, by varying S, the effect of using different front contact materials with different work-functions can be studied through this analytical model.

The complexity of the developed solution can be reduced by approximating the Hypergeometric function for thin-film emitters and by reducing Bessel function when its argument (βz) is in smaller range [15]. This model is easily expandable to other single & multi-layer bipolar photovoltaic devices, p-i-n solar cell and CIGS cell.

IV. Conclusions

This work proposes a new analytical model of minority carrier concentration for exponentially doped quasi-neutral emitter which incorporates the drift field, variable lifetime & mobility and solar illumination from standard spectrum. It has been found that this theoretical model predicts the carrier profile successfully over a wide range of doping and surface recombination velocity, at an added complexity of Bessel and Hypergeometric function. The solution is mathematically consistent with numerical physics-based model developed in COMSOL Multi-physics and shows good match with TCAD device simulator. This model can be a useful tool for analysing and optimizing simultaneous contributions from device dimensions, biasing, doping, surface condition and spectrum.

References

[1] S. Daliento and L. Mele, "Approximate Closed-Form Analytical Solution for Minority Carrier Transport in Opaque Heavily Doped Regions Under Illuminated Conditions," *Electron Devices, IEEE Transactions on,* vol. 53, pp. 2837-2839, 2006.

[2] M. H. Sanders, "Modeling of Operating Temperature Performance of Triple Junction Solar Cells Using Silvaco's Atlas," M. Sc. Thesis, Department of ECE, Naval Postgraduate School Monterey, CA, 2007.

[3] P. Bommareddy, "Effects of Front and Bottom Texturings On Gallium Arsenide Solar Cell Characteristics," M. Sc. Thesis, Dept. of ECE, University of Missouri-Columbia, 2012.

[4] A. B. Sproul and M. A. Green, "Parameters of laser grooved silicon solar cells determined from matching theoretical and experimental results," *Proc. 4th Int. Photovolt. Sci. Eng. Conf. (Sydney, Australia)* 1989.

[5] L. A. Verhoef and W. C. Sinke, "Minority-carrier transport in nonuniformly doped silicon-an analytical approach," *Electron Devices, IEEE Transactions on,* vol. 37, pp. 210-221, 1990.

[6] A. R. Burgers, C. Leguijt, P. Lolgen, and W. C. Sinke, "New analytical expressions for dark current calculations of highly doped regions in semiconductor devices," *Electron Devices, IEEE Transactions on,* vol. 44, pp. 171-179, 1997.

[7] J. Park, x, Sik, A. Neugroschel, and F. A. Lindholm, "Systematic analytical solutions for minority-carrier transport in semiconductors with position-dependent composition, with application to heavily doped silicon," *Electron Devices, IEEE Transactions on,* vol. 33, pp. 240-249, 1986.

[8] N. Rinaldi, "Modeling of minority-carrier transport in nonuniformly doped silicon regions with asymptotic expansions," *Electron Devices, IEEE Transactions on,* vol. 40, pp. 2307-2317, 1993.

[9] J. A. del Alamo and R. M. Swanson, "Validity of the quasi-transparent model of the current injected into heavily doped emitters of bipolar devices," *Electron Devices, IEEE Transactions on,* vol. 34, pp. 455-456, 1987.

[10] S. Daliento, L. Mele, E. Bobeico, L. Lancellotti, and P. Morvillo, "Analytical modelling and minority current measurements for the determination of the emitter surface recombination velocity in silicon solar cells," *Solar Energy Materials and Solar Cells,* vol. 91, pp. 707-713, 5/4/ 2007.

[11] N. F. Rinaldi, "Modeling of small-signal minority-carrier transport in bipolar devices at arbitrary injection levels," *Electron Devices, IEEE Transactions on,* vol. 45, pp. 1501-1510, 1998.

[12] M. M. Chowdhury and B. Debnath, "Approximation of carrier generation rate in common solar cells and studies for optimization of n+p silicon solar cell for AM1.5G and AM1.5D," in *7th International Conference of Electrical & Computer Engineering (ICECE)*, 2012, pp. 327-330.

[13] J. Furlan and S. Amon, "Approximation of the carrier generation rate in illuminated silicon," *Solid-State Electronics,* vol. 28, pp. 1241-1243, 1985.

[14] J. del Alamo, S. Swirhun, and R. M. Swanson, "Measuring and modeling minority carrier transport in heavily doped silicon," *Solid-State Electronics,* vol. 28, pp. 47-54, 1985.

[15] W. W. Bell, *Special Functions for Scientists and Engineers*: Dover Publication, 2004.

Capacitance-Voltage Characterization and Semi-classical Transport Analysis of $In_xGa_{1-x}As$ Surface Channel Quantum Well MOSFET

*Ehsanur Rahman, Abir Shadman, Kanak Datta, Sudipta Romen Biswas and Quazi D.M. Khosru

Department of Electrical & Electronic Engineering,
Bangladesh University of Engineering & Technology
Dhaka-1000, Bangladesh
*EhsanEEEBuet@gmail.com

Abstract—**In this paper, the Capacitance-Voltage (C-V) and Ballistic Current characteristics of arsenide based surface channel Quantum Well (QW) MOSFET were investigated. Self-consistent simulation was performed by solving coupled Schrödinger-Poisson equation incorporating wave function penetration into oxide. Although Experimental C-V and I-V characteristics of the Surface Channel QW MOSFET are available in recent literature, a self-consistent simulation based C-V and I-V characterization is yet to be reported. We studied some important parameters variation like oxide material, doping concentration and their impacts on C-V characteristics. We have also simulated carrier transport in the Ballistic limit using top of the barrier approach.**

Index Terms **—Surface Channel, Quantum Well MOSFET, Self-Consistent Analysis, Gate Capacitance**

I. Introduction

Although successive device scaling needs to be continued for sustaining the evolutionary technological improvements, power consumption has been the bottleneck for future silicon CMOS technology scaling [1]. III-V semiconductors have emerged as potential candidates for future high frequency, low power logic applications [2]-[3]. With better carrier mobility, options for possible bandgap modulation, better carrier confinement in thin body structures; III-V materials could meet the demand of high performance solid state electronics in the future. III-V surface channel quantum well device structures incorporating high-k dielectric materials are considered to be promising device technologies. Surface channel device, though mobility is more affected by the interface, has stronger gate controllability over channel which is more desirable as devices are being scaled down aggressively [4]. Despite extensive experimental research, a thorough self-consistent simulation based study on C-V characteristics is yet to be done in literature. In this work, we have performed a simulation based study of C-V characteristics solving self-consistent Schrödinger-Poisson equation for $In_xGa_{1-x}As$ n type surface channel MOSFET. Also I-V characteristics of the n type surface channel device have been simulated semiclassically.This semi classical treatment of transport characteristics has successfully captured short channel effects like DIBL and it provides us with a fair upper level estimate of current in the surface channel MOSFET.

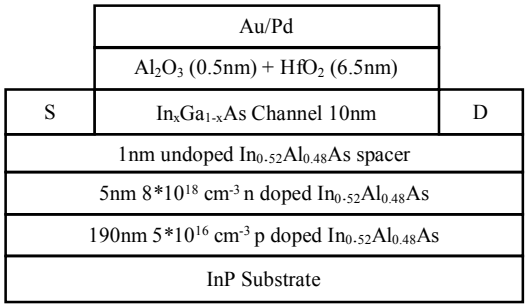


Fig. 1. Basic device structure of $In_xGa_{1-x}As$ surface channel QW MOSFET [5]

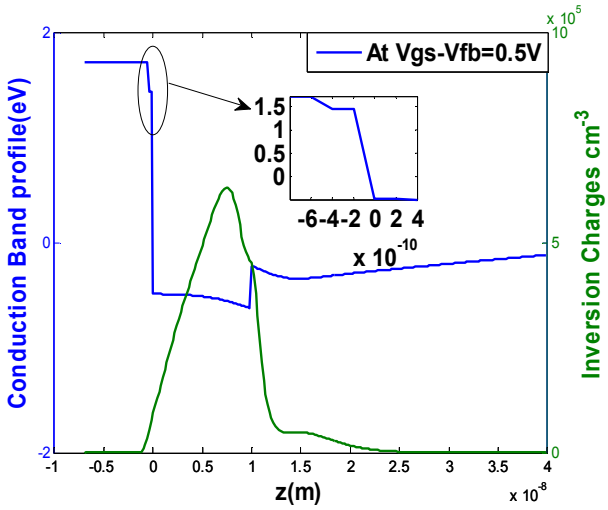


Fig. 2. Conduction band profile along with carrier profile (Inset, transition from 0.5nm Al_2O_3 to 6.5nm HfO_2 is shown (composite gate oxide))

II. Self-consistent C-V Modelling

We have developed a device simulator based on self-consistent solution of Schrödinger-Poisson equation [6]-[7] to characterize the device [5] in Fig. 1. The device has 5nm delta-doped n type $In_{0.52}Al_{0.48}As$ layer separated from channel by 1nm $In_{0.52}Al_{0.48}As$ spacer. In this work we have used Finite Element Method (**FEM**) for solving 1D Schrödinger-Poisson equation using MATLAB coupled with COMSOL. The conduction band profile and carrier profile for the device are illustrated in Fig. 2.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

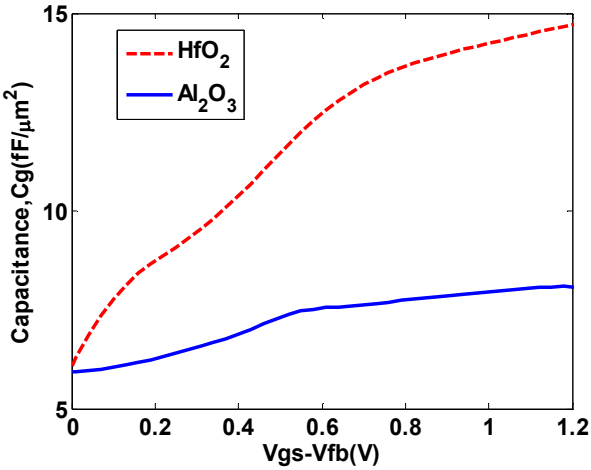


Fig. 3. Gate capacitance variation with different dielectric materials

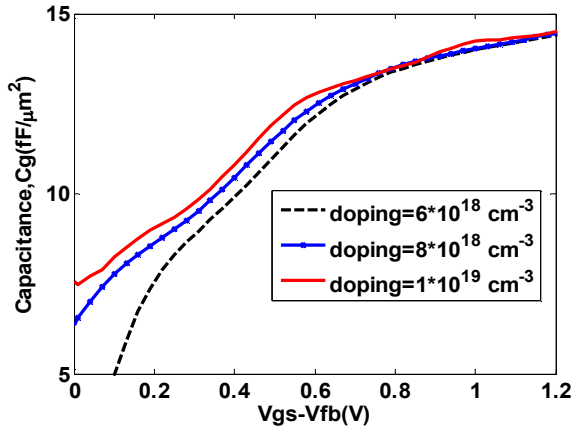


Fig. 5. Gate capacitance variation with various delta doping concentrations.

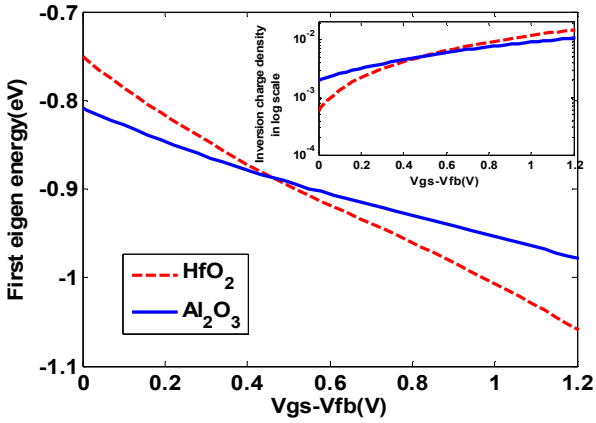


Fig. 4. First eigen energies and sheet carrier density with different dielectric materials for the same device.

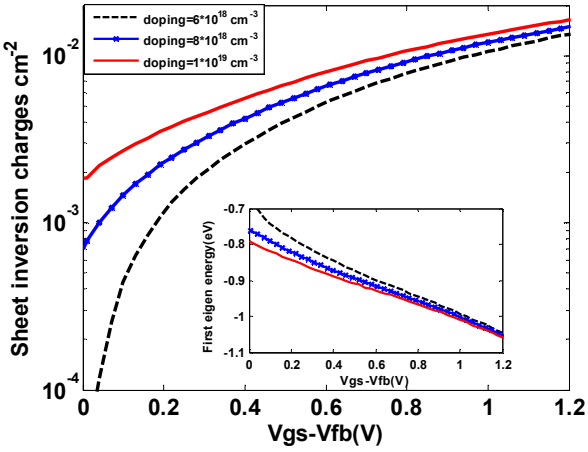


Fig. 6. Sheet carrier density and First eigen energies for different delta doping concentrations

From Fig. 2, we observe that the carriers are well confined within the quantum well channel. Delta doping concentration and dielectric materials are varied to find out its impact on C-V profile. All simulations are done at room temperature (T=300k.)

III. Results and Discussions

Due to higher gate control over channel sheet carrier, device with HfO_2 dielectric shows a higher inversion capacitance (Fig. 3). Fig. 4 shows change of eigen energy and sheet carrier density with gate bias for two different dielectric materials. Al_2O_3 provides slightly better confinement (not shown) at low gate bias. So we get lower eigen energy for it.But at higher bias the HfO_2 dielectric provides a better carrier confinement than Al_2O_3. So eigen energy is lower and confined charge is higher for HfO_2. Fig. 5 shows Gate capacitance variation with various doping concentrations. In Fig. 6, carrier concentration appears to be much higher at low gate bias in heavily doped device compared to lightly doped one. For low gate bias, the well (not shown) in the heavily doped device gets more skewed. Consequently carrier concentration at low gate bias is substantially higher for heavily doped device. At high gate bias however, the shape of the well remains almost same (not shown) for all doping concentrations. As a result, eigen energies become almost similar too (Fig. 6) and sheet carrier densities approach closer value.

However, the device with higher doping density still shows higher carrier concentration. Fig. 7 shows the simulated gate capacitance and experimental gate capacitance data [5] on the same plot. By taking the metal semiconductor work function difference into account, we find good agreement between the two C-V profiles. The simulation is performed without taking any non-ideal effects like fixed oxide charges, interface trapped charges or any other oxide defects which may explain the discrepancy between simulation and experimental observation.

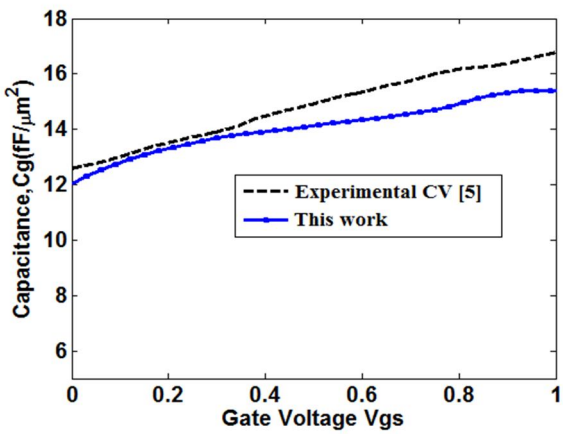


Fig. 7. Simulated and experimental gate capacitance profile.

IV. SEMI CLASSICAL BALLISTIC TRANSPORT MODEL

We have also developed a self-consistent simulator to characterize carrier transport through the device. In our treatment, the quantum effect in surface channel QW MOSFETs is accounted for by solving the Schrödinger equation in the gate confinement direction. From this solution we obtain discrete subbands within which carriers (electrons in our case) are constrained. To account for carrier transport a 1D solution to the Boltzmann Transport Equation (BTE) in the ballistic limit is calculated [8]. Assuming device widths are infinite and translational invariance holds, subband potential energies can only vary in one direction in response to different bias conditions. Therefore, a 1D solution will be sufficient to describe the electron transport within the MOSFETs.More specifically, this 1D solution provides a charge-sheet description.

The approach we have described above is called semi classical approach as the physics in the direction normal to the channel are treated quantum mechanically but carrier transport along the channel is treated classically.Semiclassical ballistic transport only considers the top of potential barrier in channel with everything above transmitted and below reflected. On the other hand, full quantum solution accounts for quantum effects in both the confinement and the transport directions. Semi classical approach can capture the thermionic emission current but it cannot account for any quantum related effects, such as electron tunnelling through the source-channel barrier and quantum interference within ballistic devices. Despite these limitations the semi classical approach gives a fair upper limit approximation of carrier transport across the channel.

V. TRANSPORT CHARACTERIZATION

The conduction band profile and normalized wavefunction obtained from quantum mechanical treatment along confinement direction are shown in Fig. 8. From the figure, we observe that the carriers are well confined within the quantum well channel even in transport condition.

Fig. 9 (both a & b) shows the computed self-consistent conduction subband profiles versus position under various Vgs & Vds bias points keeping one constant at a time.Here we plot the bottom of the first subband versus position, not the conduction band edge.Form Fig. 9 (a) we observe that for a fixed gate bias, increasing drain bias causes the energy of the sub-band portion at the drain end to decrease gradually.
Fig. 9 (b) shows the computed self-consistent potentials within the model nanoscale MOSFET under low and high drain bias with gate voltage as a parameter. At low gate voltages, the energy barrier between the source and drain is high. So the source and drain injected carriers are reflected off the channel potential barrier and the device is off. A high gate voltage lowers the potential energy barrier and electrons flow from source to drain. So the height of the energy barrier is controlled by the voltage on the gate.

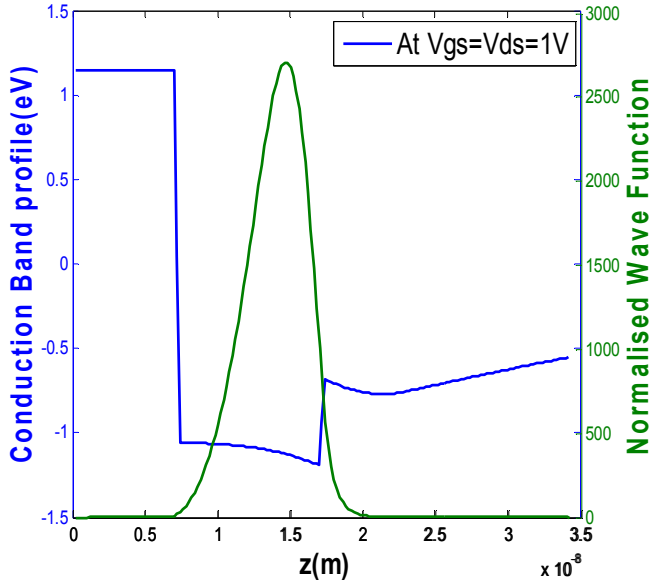
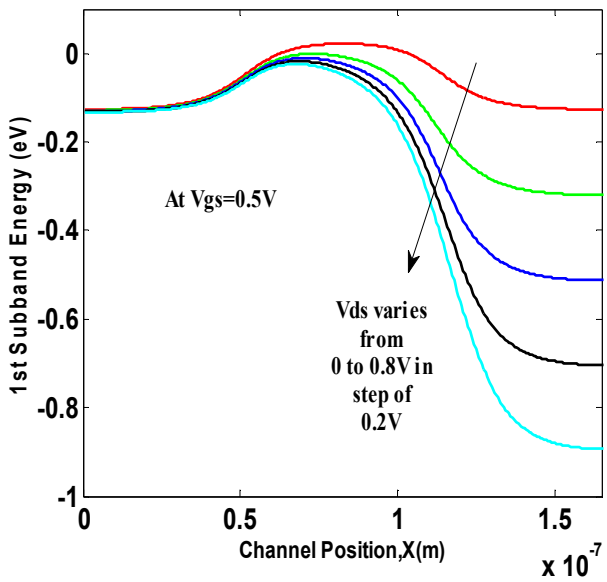


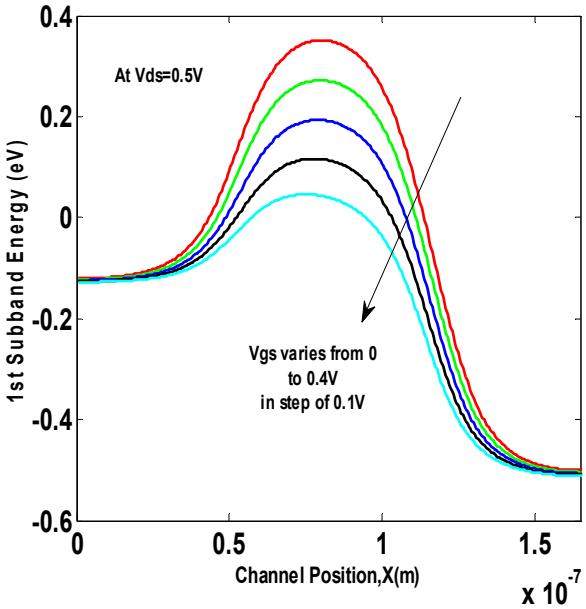
Fig. 8. Conduction band profile along with normalized wavefunction

Drain Induced Barrier Lowering (DIBL) is an important short channel effect for nanoscale transistors. MOSFETs usually operate by modulating the potential barrier in the channel between source and drain. For long channel devices, this potential barrier is generally modulated by gate voltage.

However, as the channel length is being scaled down, voltage applied at the drain terminal starts to modulate this energy barrier in the channel near the source end. This phenomenon also leads to the lowering of threshold voltage for short channel devices. This effect has been captured by our simulator as observed from lowering of channel potential barrier for the 1st subband near the source end with increasing drain bias voltage (Fig. 9(a)).



(a)



(b)

Fig. 9. First subband profile versus channel position for (a) Vgs = 0.5 V and Vds varying from 0 V to 0.8 V (b) Vds = 0.5 V and Vgs varying from 0 V to 0.4 V.

Fig. 10 shows simulation results for intrinsic gate bias voltage ranging from 0.4V to 1V with 0.1V step and drain bias voltage ranging from 0V to 1.0 V. The ballistic current is higher than the reported experimental value for the same device. This mismatch between experimental and simulated data can be attributed to the absence of any inelastic scattering in our simulation. We have considered ideal contact without any contact resistance which also contributes to this discrepancy.

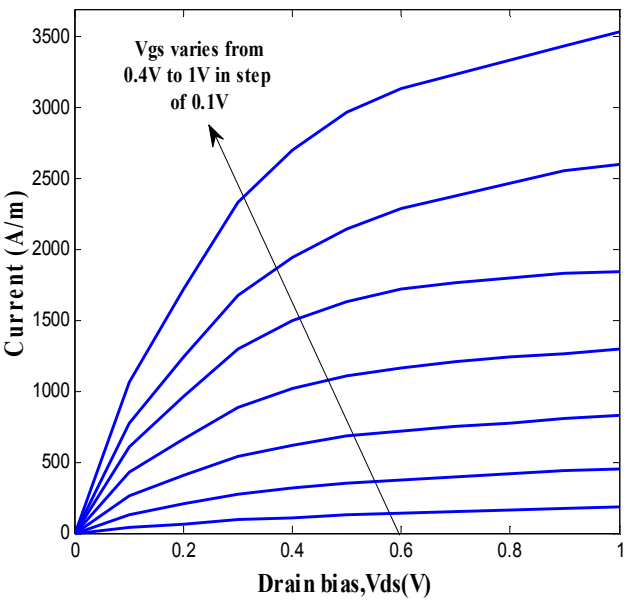


Fig. 10. The intrinsic I_d-V_{ds} characteristics for the simulated device structure.

VI. CONCLUSION

We have performed a self-consistent quantum mechanical C-V characterization of 55 nm Quantum Well device reported in [5].We observed that variations of doping concentration and oxide material have strong effects on C-V characteristics. HfO_2, as a high k insulator, provides much better control over channel than Al_2O_3.Delta doping variation causes large change in sheet charge density at low gate bias although the effect is less pronounced at higher bias. So higher delta doping can be a useful mean of increasing sheet charge density in the channel without altering the device structure. By keeping the length of side spacer region equal to gate length of the structure, the carrier transport through the channel has been analysed semi classically. We have determined subband profile in the $In_{0.53}Ga_{0.47}As$ channel quantum mechanically by solving coupled Schrödinger Poisson equation. Current in ballistic regime has been calculated using top of the barrier approach. The simulated ballistic current is higher than experimental measurements [5].This discrepancy can be attributed to the avoidance of scattering phenomenon in the QW channel. Also, in this approach, the effect of doped heterostructure stack along with contact resistance has been neglected. To get a more realistic picture of transport characteristics, quantum effect should be incorporated in transport direction using NEGF formalism [9]-[11].

REFERENCES

[1] E. Pop, “Energy Dissipation and Transport in Nanoscale Devices”,Nano. Res. 3, 147 (2010).

[2] D. A. Antoniadis and A. Khakifirooz, "MOSFET Performance Scaling: Limitations and Future Options," IEEE International Electron Devices Meeting 2008, Technical Digest, pp. 253-256, 2008.

[3] R. Chau, "Challenges and opportunities of emerging nanotechnology for VLSI nanoelectronics," 2007 International Semiconductor Device Research Symposium, Vols 1 and 2, pp. 449-449, 2007.

[4] Fei Xue, et al ,“Investigation of Surface Channel InGaAs MOSFETs with ALD Gate Dielectric” on 218th ECS Meeting.

[5] M. Egard et. al., “High Transconductance Self-Aligned Gate-Last Surface Channel $In_{0.53}Ga_{0.47}As$ MOSFET”,IEEE IEDM, p.303(2011).

[6] F. Stern, “Self-consistent results for n-type si inversion layers,” P*hys. Rev. B*, vol. 5, pp. 4891 –4899,1972.

[7] C. Moglestue, “Self-consistent calculation of electron and holeinversion charges at silicon-silicon dioxide interface,” *J. Appl.Phys*., vol. 59, no. 5, pp. 3175 –3183,1986.

[8] Z. Ren, Nanoscale MOSFETs: Physics, Simulation, and Design, Ph.D.Thesis, Purdue University, West Lafayette, IN (2001).

[9] S. Datta, ”Nanoscale device modeling: the Green’s functionmethod.” Superlattices and Microstructures 28.4 (2000): 253-278.

[10] Z. Ren, R. Venugopal, S. Goasguen, S. Datta, M. Lundstorm, "nanoMOS 2.5: A two-dimensional simulator for quantum transport in double-gate MOSFETs," Electron Devices, IEEE Transactions on, vol 50, pp. 1914-1925, Sept. 2003.

[11] Ehsanur Rahman, Abir Shadman, Sudipta Romen Biswas, Kanak datta, Quazi D. M. Khosru “$In_xGa_{1-x}As$ Surface Channel Quantum Well MOSFET: Quantum Ballistic Simulation Using Mode Space Approach,” 14th International Conference on Semiconductor Electronics, 2014, Kuala Lumpur, Malaysia (Accepted).

Control of Reflection through Epsilon near Zero Graphene based Anisotropic Metamaterial

Ayed Al Sayem, [1*] Arif Shahriar[1], M.R.C. Mahdy, [2**] and Md. Saifur Rahman[1]

[1]Dept. of EEE, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh

[2] Department of Electrical and Computer Engineering, National University of Singapore, 4 Engineering Drive 3, Singapore

[*]ayedalsayem143@gmail.com

[**]elemrc@nus.edu.sg

Abstract— **In graphene dielectric multilayer structure, a very interesting phenomena such as epsilon near zero property rises naturally. However, this epsilon near zero property (along with the proper control of chemical potential and gate voltage) of graphene multi-layer stack has not been used so far to make novel photonic switches .In this article, we have shown theoretically that in graphene-dielectric stack, which acts as an anisotropic metamaterial, full control of reflection and so transmission of light at a specific wavelength or frequency can be controlled very simply by external gate voltage. This external voltage tunes the chemical potential and so the parallel permittivity of the anisotropic structure. Our theoretical prediction may pave the way to novel voltage control photonic logic switches, which may play as an intermediate solution before entering into all optical commercial switches (as current silicon and electron based CMOS technology is facing a dead end).**

***Index Terms*— Graphene, anisotropic metamaterial, reflection, chemical potential, tunability.**

I. Introduction

In condensed-matter physics, graphene can be considered as the latest and most valuable and promising material. One of very few two-dimensional materials, graphene has already shown its fascinating electrical properties due to its unconventional and amazing crystal and electronic properties [1]. Graphene transistors at first were thought to replace silicon logic transistors. However, due to the absence of band-gap they failed to do so though they have been successful in application of terahertz transistors [2]. Beside electronic applications, graphene has fascinating photonic applications such as transparent electrodes in displays and photovoltaic modules [3], photo-detectors [4] etc. Recently graphene based anisotropic metamaterials have gained enormous attention to researchers due to their exciting tunable electromagnetic properties [5], [6], [7], [8]. Graphene based multilayered metamaterials can show both elliptical and hyperbolic phenomena and they can be characterized as uniaxial anisotropic metamaterials. Hyperbolic metamaterials with hyperbolic dispersion have shown their extreme ability and applications exhibiting negative refraction [9], negative index [10], sub-wavelength imaging [11], light confinement [12], [13], nonlinearity enhancement [14], bio-sensing [15] etc. Graphene-dielectric multilayer structure can also behave as hyperbolic medium and has already been mentioned in literature [5], [6], [7], [8] for its applications in many occasions. But in this article, we are more interested in elliptical behaviour of graphene dielectric anisotropic metamaterials.

In this article, we have considered a graphene-dielectric multilayered anisotropic structure. The structure can both work as hyperbolic and elliptic metamaterial. We describe the permittivity behaviour of the structure at different frequency and also its dependence on chemical potential, which is the most exciting phenomena in graphene dielectric multilayered structure (as in case of metal-deielctric multilayered structure, it is not possible to change the electromagnetic characteristic of the structure after fabrication). This limitation can also be seen from our previous work [16] where Relation between Perfect tunneling and Band Gaps for SNG Metamaterials has been briefly discussed. But both perfect tunnelling phenomena and band-gaps are not actively tunable and they require both negative permeability and permittivity. As graphene is an actively tunable material, our multilayered structure is also tunable.By means of this tunability option, we theoretically show that reflection form the structure can be fully controlled very simply by external voltage using exciting behaviour of the graphene in THz range (where intra-band transitions dominate).

II. Theory

Uniaxial anisotropic metamaterial can be identified by the permittivity tensor.

$$\varepsilon = \begin{bmatrix} \varepsilon_{||} & 0 & 0 \\ 0 & \varepsilon_{||} & 0 \\ 0 & 0 & \varepsilon_{\perp} \end{bmatrix}$$

Where, $\varepsilon_{||}$ is the parallel permittivity and $\varepsilon_{\perp}$ is the longitudinal permittivity of the structure. There are two ways to engineer hyperbolic metamaterial structure:

a) Multilayer realization consisting of alternating sub-wavelength metal and dielectric layers.
b) Metal nanowires in a dielectric host.

These uniaxial materials can show both elliptic and hyperbolic dispersion if they satisfy the condition $(\varepsilon_{||}\varepsilon_{\perp}) > 0$ and $(\varepsilon_{||}\varepsilon_{\perp}) < 0$ respectively.

For metal dielectric hyperbolic metamaterial structure, as long as the length of one unit cell consisting one metal and one dielectric layer is less than the wavelength of light used, we can use effective medium theory.

By effective medium theory we get,

$$\varepsilon_{\parallel} = f\,\varepsilon_m + (1-f)\varepsilon_d$$
$$\varepsilon_{\perp} = \left(\frac{f}{\varepsilon_m} + \frac{(1-f)}{\varepsilon_d}\right)^{-1}$$
$$f = \frac{t_m}{t_d + t_m}$$

Where, t_m is the thickness of metal layer, t_d is the thickness of dielectric layer, ε_d is the permittivity of

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

dielectric layer, ε_m is the permittivity of metal layer and f is the filling ratio.

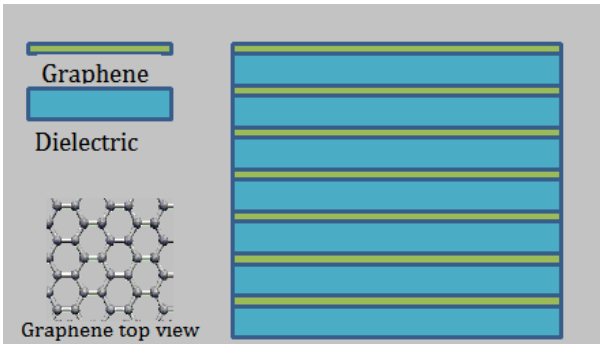


Fig. 1 Graphene-dielectric multilayer structure considered in this work.

Graphene is a single atom thick two dimensional plane of carbon atoms forming a hexagonal lattice layer of carbon atoms with a lattice constant of 0.264 nm. Transverse magnetic (TM) surface plasmons are supported by graphene. Recently it has been shown that graphene-dielectric alternating stack can work as a uniaxial anisotropic metamaterial similar to those of metal-dielectric anisotropic metamaterials in the mid infrared and THz frequencies. In case of graphene, in THz frequency range, intra-band transitions dominate. Frequency and chemical potential dependent conductivity of a monolayer graphene can be given by Kubo formula,

$$\sigma(\omega,\mu_c) = -\frac{j4\pi e^2 k_B T}{h^2(\omega - j2\mathcal{T})}\left(\frac{\mu_c}{k_B T} + 2\ln\left(e^{\frac{-\mu_c}{k_B T}} + 1\right)\right) - \frac{j4\pi e^2(\omega - j2\mathcal{T})}{h^2}\int_0^\infty \frac{f_D(-\xi) - f_D(\xi)}{(\omega - j2\mathcal{T})^2 - 16\left(\frac{\pi\xi}{h}\right)^2} d\xi \quad (1)$$

The effective permittivity can be derived in following form,

$$\varepsilon_g = 1 - j\frac{\sigma(\omega,\mu_c)}{\omega\varepsilon_o t_g} \quad (2)$$

$$\varepsilon_\parallel = \frac{t_g\varepsilon_g + t_d\varepsilon_d}{t_g + t_d} \quad (3)$$

$$\varepsilon_\perp = \frac{\varepsilon_g\varepsilon_d(t_g + t_d)}{t_g\varepsilon_d + t_d\varepsilon_g} \cong \varepsilon_d\ ;\ t_g \ll t_d \quad (4)$$

Where k_B is Boltzmann constant, T is temperature, h is plank's constant, μ_c is chemical potential of graphene, ω is angular frequency, $\mathcal{T}$ is phenomenological scattering rate (.1meV), t_g is the thickness of graphene layer, t_d is the thickness of dielectric layer and ε_d is the permittivity of dielectric layer and Fermi Dirac function $f_D(\xi) = \left(e^{\frac{\xi - \mu_c}{k_B T}} + 1\right)^{-1}$. In the THz range we can expect graphene-dielectric stack to act both as an elliptic media or hyperbolic media depending on the chemical potential.

Anisotropic epsilon near zero metamaterails have been demonstrated to show exciting full transmission even with impedance mismatch. Here at first, we demonstrate the physical conditions for full transmission and full reflection of an anisotropic metamaterial slab. We neglect absorption because in our selected frequency region loss is negligible and another mechanism for absorption in this kind of metamaterial, 'Slow light' [17], [18], [19] is absent as we are working on elliptical region. The slab will be fully transmitting if its permittivity along the transverse plane equals to 1 and its permittivity along propagation direction also equals to 1. This simply means that the slab has to be made of graphene and air stack though the dependence on longitudinal permittivity is much less than the parallel permittivity. If we can make a graphene structure spaced by air instead of dielectric then the longitudinal permittivity is always equal to 1 as long as the spacing of air $t_{air} \gg t_g$ (eqn. 4). Transverse permittivity of graphene–air/dielectric multilayered structure can be tuned at different frequency by tuning temperature and also chemical potential. Even with graphene-dielectric structure full transmission can be achieved because transmission through a uniaxial metamaterial slab is not critically dependent on longitudinal permittivity. And here instead of a garphene-air structure we consider a graphene-dielectric structure which will also show similar phenomena though the longitudinal permittivity is not equal to 1.

To obtain full reflection, the transverse permittivity has to be zero to achieve complete impedance mismatch. This is physically possible, as we can change the permittivity of graphene (varying graphene's chemical potential) by controlling external voltage or by changing the temperature.

The theoretical formula for reflection for a uniaxial anisotropic metamaterial can be given by following formula [20]:

$$R = |r|^2 = \left|\frac{\varepsilon_\parallel \tan\theta_t - \tan\theta}{\varepsilon_\parallel \tan\theta_t - \tan\theta}\right|^2 \quad (5)$$

$$\theta_t = \tan^{-1}\sqrt{\frac{\varepsilon_\perp \sin^2\theta}{\varepsilon_\parallel\varepsilon_\perp - \varepsilon_\parallel \sin^2\theta}}\ ;\text{TM polarisation}$$

Full impedance matching is achieved when $\varepsilon_\parallel = 1$ and when $\varepsilon_\parallel = 0$ full impedance mismatch is achieved for an anisotropic uniaxial metamaterial which can be seen from eqn. (5)

Graphene anisotropic metamaterial can be tuned by temperature or by chemical potential which can be achieved by applying gate voltage [5]. So for a specific dielectric thickness t_d, dielectric permittivity ε_d and specific chemical potential μ_c, we can obtain the frequency at which parallel permittivity of the structure is closely equal to 1 so higher transmission is achieved.

Relationship between gate voltage and chemical potential can be given by following formula [5], [21]

$$|\mu_c| = \hbar v_F\sqrt{\pi|a_o(V_g - V_{dirac})|} \quad (6)$$

Where, V_{dirac} is the voltage offset which is zero if there is no impurity or doping in graphene, v_F is the Fermi velocity (~1e6 ms^{-1}) [1] of Dirac fermions in graphene and $a_o = 9 \times 10^{16} m^{-1}V^{-1}$. Bias voltage can change the chemical potential though it is also possible to tune chemical potential by magnetic field and chemical doping.

Here we want to demonstrate an opto-electronic structure, which is basically a voltage controlled device. In this device, the carrier is photon and control mechanism is voltage. For 0 gate voltage, the anisotropic graphene-dielectric metamaterial is tuned so that it has transverse permittivity 1 at a specific frequency ~44.78 THz. And so we have almost full transmission or zero reflection. For high gate voltage (which

is calculated as 30.5 mV), transverse permittivity value decreases from 1 to 0. Thus we have almost full reflection.

III. Results and Discussion

In this section we discuss our theoretical results introduced in section II.

Fig.2 shows both the real permittivity (in parallel direction of the structure) and longitudinal permittivity of graphene-dielectric uniaxial anisotropic metamaterial. It can be seen that effective permittivity (in parallel direction) is a function of frequency. But the longitudinal permittivity is almost fixed at 2.2 as thickness of dielectric layer is much greater than the graphene layer thickness. This is also a basic difference in metal-dielectric multilayered structure as we can't make ultra-thin metal on dielectric so longitudinal permittivity of metal-dielectric metamaterial also changes with frequency. Again parallel permittivity can take both positive and negative values showing their capability of both elliptic and hyperbolic dispersion profile though in this paper we are only interested in elliptical region.

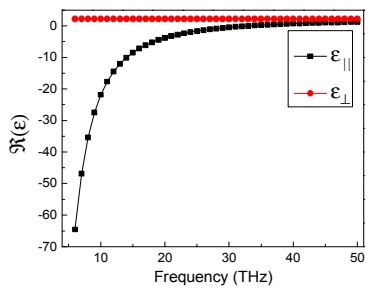


Fig. 2 Permittivity in parallel and longitudinal direction of graphene-dielectric metamaterial slab.

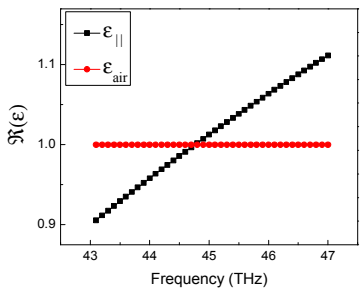


Fig. 3 Parallel permittivity of graphene-dielectric slab crossing air permittivity at a critical frequency

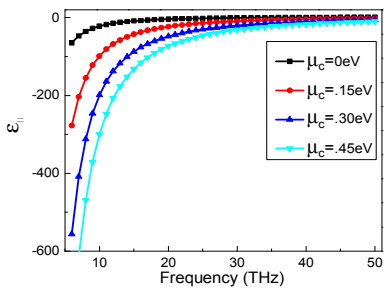


Fig. 4 Parallel Permittivity of graphene-dielectric metamaterial slab at different chemical potential.

Fig. 3 shows the parallel permittivity in the frequency range 43 THz to 47 THz. As in this range, graphene-dieletric multilayered metamaterial crosses the air permittivity. Permittivity of air is also shown in the graph to indicate the intersecting point of metamaterial permittivity and air permittivity. We name the intersecting frequency point as critical frequency as at this critical frequency the parallel permittivity of metamaterial is equal to one which is a necessary condition for almost zero refection. Also this critical frequency is obtained when the structure is unbiased that means at zero chemical potential or at zero gate voltage considering $\boldsymbol{V_{dirac} = 0V}$

Fig.4 shows both the real permittivity (in parallel direction of the structure) at different frequency with different chemical potential. For higher chemical potential or higher gate voltage, parallel permittivity obtains more negative value making the anisotropic structure to epsilon near zero to hyperbolic structure though in this paper we are not interested in the hyperbolic region as we want to vary the parallel permittivity between one to zero.

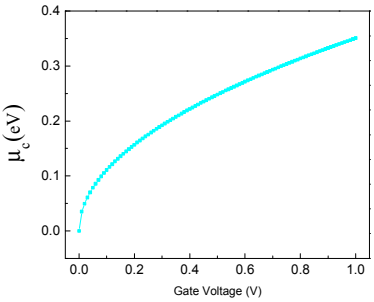


Fig.5 Chemical potential of graphene-dielectric metamaterial slab at different gate voltage.

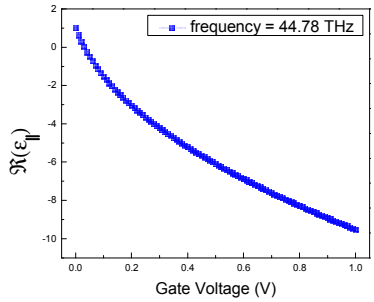


Fig. 6 Parallel Permittivity of graphene-dielctric metamaterial slab at different gate voltage at critical frequency 44.78 THz where the structure is tuned at parallel permittivity equaling one.

Fig.5 shows the relation between chemical potential and gate voltage. Chemical potential can be changed from 0eV to ~.35eV by applying gate voltage from 0 to 1V. Parallel permittivity is dependent on chemical potential and chemical potential is dependent on gate voltage. Parallel permittivity can be changed by applying gate voltage, which we can see from fig. 6. Here critical frequency is fixed at 44.78 THz. At this frequency, at zero biased condition, parallel permittivity of the structure is equal to ~1. This permittivity value (1) is the required condition for very low negligible reflection. Also from the figure it can be observed that gate voltage can tune the structure from elliptic to hyperbolic. Here we want to mention another point that the graphene can be chemically doped and by chemically doping, it is possible to obtain higher chemical potential value which will give the parallel permittivity condition of negligible reflection at a higher frequency. And also the thickness and permittivity of the dielectric layer can be changed (reduced) to obtain critical frequency at a high value than what we have used here.

Gate voltage induced (equivalently changing chemical potential (fig. 7)) zero to full reflection is clearly observed in

Fig.8. As our structure was tuned at critical frequency 44.78 THz, so that the real part of permittivity in parallel direction

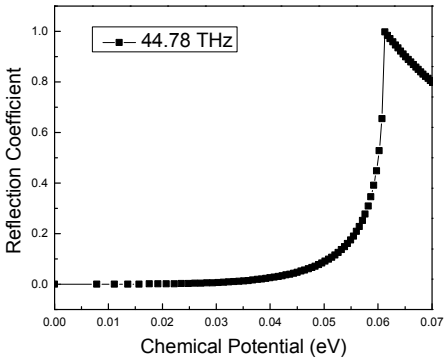


Fig. 7 Chemical Potential dependent reflection from the graphene dielectric metamaterial structure. As chemical potential is tuned by gate voltage, parallel permittivity of the structure changes and as a result reflection from the structure goes from zero to unity.

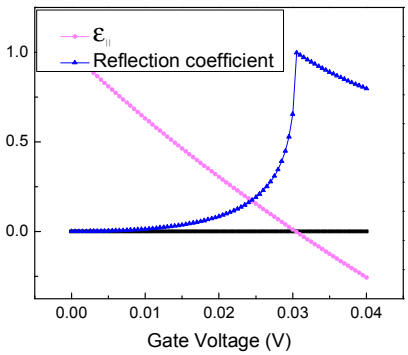


Fig. 8 Zero reflection from the structure at zero gate voltage, reflection increases as gate voltage is applied and reaches unity at 35mV, pink line shows the parallel permittivity of the structure as it changes from one to zero because of gate voltage, the structure changes its behavior from zero reflection to full reflection as it is expected.

meets the desired value one, we can observe negligible reflection at zero gate voltage or equivalently at zero chemical potential or for an unbiased structure. When the voltage is ramped, chemical potential of each graphene layer is changed and so as a result the parallel permittivity of the whole structure decreases and at a specific voltage parallel permittivity reaches zero which is the satisfactory condition for full reflection. And we obtain the voltage required for full reflection is 30.5mV. After this voltage permittivity crosses zero value to negative value letting the structure work as a hyperbolic metamaterial. So from above discussion if we apply a pulse from 0 to 35mV, we should observe zero to full reflection. In our work, we didn't mention the imaginary part of permittivity of graphene layer. This is because the imaginary part of the permittivity of graphene layer is relatively low compared to the real part at our critical frequency. Permittivities of the whole structure (parallel and perpendicular) are dependent on graphene's permittivity. So at our critical frequency, the imaginary part of the permittivity of the structure (parallel and perpendicular) is also relatively low compared to the real part. As a result, it should not hamper our theoretical predictions.

IV. Conclusion

In this article, we have at first briefly described the graphene-dielectric multilayered anisotropic metamterial, its permittivity properties in THz frequency range, dependence on chemical potential and equivalently on gate voltage. From simple theoretical calculation, it has been shown that the multilayer structure can be tuned to control the reflection properties of the device, which may act as a photonic switch. Our work may pave the way to novel manipulation of graphene based metamaterials in logic device applications for faster operation replacing current CMOS technology.

References

[1] A. H. Castro Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, and A. K. Geim, "The electronic properties of graphene", *Rev. Mod. Phys.* 81, 109 (2009)

[2] Jiaxin Zheng, Lu Wang, Ruge Quhe, Qihang Liu, Hong Li, Dapeng Yu, Wai-Ning Mei, Junjie Shi, Zhengxiang Gao and Jing Lu, "Sub-10 nm Gate Length Graphene Transistors: Operating at Terahertz Frequencies with Current Saturation*", Scientific Reports* 3,1314 (2013).

[3] Bonaccorso, "Graphene photonics and optoelectronics*", Nature Photon*, 4, 611–622 (2010).

[4] F. Xia, T. Mueller, Y. M. Lin, A. Valdes-Garcia & P. Avouris, "Ultrafast graphene photodetector", *Nature Nanotech.* 4, 839–843 (2009)

[5] Bofeng Zhu, Guobin Ren, Siwen Zheng, Zhen Lin, and Shuisheng Jian, "Nanoscale dielectric-graphene-dielectric tunable infrared waveguide with ultrahigh refractive indices**,"** *Optics Express*, Vol. 21, Issue 14, pp. 17089-17096 (2013)

[6] K. V. Sreekanth, A. De Luca, and G. Strangi, "Negative refraction in graphene-based hyperbolic metamaterials," *Applied Physics Letters* 103, 023107 (2013).

[7] Mohamed A. K. Othman, Caner Guclu, and Filippo Capolino, "Graphene-based tunable hyperbolic metamaterials and enhanced near-field absorption," *Optics Express*, Vol. 21, No. 6.

[8] Ivan V. Iorsh, Ivan S. Mukhin, Ilya V. Shadrivov, Pavel A. Belov, and Yuri S. Kivshar, "Hyperbolic metamaterials based on multilayer graphene structures**,"** *Phys. Rev. B* 87, 075416 (2013)

[9] D.R. Smith, P. Kolinko, and D. Schurig, "Negative refraction in indefinite media," *JOSA B*, 21(5):1032-1043, 2004.

[10] V.A. Podolskiy and E.E. Narimanov, "Strongly anisotropic waveguide as a nonmagnetic left-handed system," *Physical Review B*, 71(20):201101, 2005.

[11] Z. Jacob, L.V. Alekseyev, and E. Narimanov, "Optical hyperlens: Far-field imaging beyond the diffraction limit," *Optics Express*, 14(18):8247-8256, 2006.

[12] A.A. Govyadinov and V.A. Podolskiy, "Metamaterial photonic funnels for sub-diffraction light compression and propagation," *Physical Review B*, 73(15):155108, 2006.

[13] J. Yao, X. Yang, X. Yin, G. Bartal, and X. Zhang, "Three-dimensional nanometer-scale optical cavities of indefinite medium. *Proceedings of the National Academy of Sciences*," 108(28):11327, 2011.

[14] GA Wurtz, R. Pollard, W. Hendren, GP Wiederrecht, DJ Gosztola, VA Podolskiy, and AV Za-yats, "Designed ultrafast optical nonlinearity in a plasmonic nanorod metamaterial enhanced by nonlocality," *Nature nanotechnology,* 6(2):107-111, 2011

[15] AV Kabashin, P. Evans, S. Pastkovsky, W. Hendren, GA Wurtz, R. Atkinson, R. Pollard, VA Podol-skiy, and AV Zayats, "Plasmonic nanorod metamaterials for biosensing," *Nature materials*, 8(11):867-871, 2009.

[16] Arif Shahriar, M.R.C. Mahdy, Jubayer Shawon, Golam Dastegir Al-Quaderi, M. A. Matin, "On the Relation between Perfect Tunneling and Band Gaps for SNG Metamaterial Structures," arXiv:1308.326

[17] Tian Jiang, Junming Zhao and Yijun Feng, "Stopping light by an air waveguide with anisotropic metamaterial cladding," *Optics Express*, Vol. 17, Issue 1, pp. 170-177 (2009).

[18] Yanxia Cui,Kin Hung Fung,Jun Xu,Hyungjin Ma,Yi Jin,Sailing He and Nicholas X. Fang, "Ultrabroadband Light Absorption by a Sawtooth Anisotropic Metamaterial Slab," *Nano Lett.*, 2012, 12 (3), pp 1443–1447

[19] Haifeng Hu, Dengxin Ji, Xie Zeng, Kai Liu Qiaoqiang Gan, "Rainbow Trapping in Hyperbolic Metamaterial Waveguide," *Scientific Reports* 3, Article number: 1249.

[20] M. A. Noginov, Yu. A. Barnakov, , G. Zhu, T. Tumkur, H. Li and E. E. Narimanov, "Bulk photonic metamaterial with hyperbolic dispersion," *Appl. Phys. Lett*. 94, 151105 (2009)

[21] A. Vakil and N. Engheta, "Transformation Optics Using Graphene," *Science* 332(6035), 1291–1294 (2011).

Data logging and energy consumption analysis of two houses in St. John's, Newfoundland

Tasnova Alam Jesha, [1,*] and M. Tariq Iqbal, [2]
[1,2]Faculty of Engineering, Memorial University
St. John's, NL A1B3X5, Canada
[*]taj275@mun.ca

Abstract— **In this paper a design of a low cost data logger to log temperature and power consumption in a house is presented. The designed data logger is capable of logging four AC current and two temperature measurements. It is based on a low cost microcontroller PIC18F4550. Microcontroller is programmed using Mikrobasic. The circuit displays temperature measurements on an LCD display and all measurements are also stored in separate files on a sd card. Power consumption of two houses was logged for one year using a sampling time of 2 minutes. The data stored in a sd card is used for further analysis. This paper also presents design of data logger and a detailed analysis of energy consumption of two houses based on collected data. Data analysis indicates power consumption pattern and its variation over a year. Analysis indicates that the logged data and monthly data collected by the utility company give similar energy consumption.**

Index Terms— **House energy consumption, data logger design, microcontrollers, net-zero houses, energy analysis.**

I. Introduction

According to statistical agency of Canada [1] the total population of Newfoundland & Labrador is 526,896 till January 2014. Many houses are electrically heated. Energy consumption in each house is logged and recorded once a month by the utility company [2]. The total electricity consumption of each house varies with type and size of the house, no of occupants etc. A much frequent data logging is required to determine energy consumption pattern in a house. A data logger design is presented in this paper that is capable of logging house input currents and temperatures at a much higher rate. The designed data logger can also sense and record the inside and outside temperature of the house and total current coming into the house.

A number of data logger designs have been published in literature. For example in [3] multiple parameters (temperature, pH) of green house are logged using sensors which is transferred to a PC using NI-DAQ cards for analysis in LabVIEW. Rosiek [4] proposed a microcontroller based data acquisition system which collects data from different meteorological stations using eight sensors. To test the functionality of the designed system results are compared with two commercial data acquisition systems. Ameur designed [5] a data acquisition system to monitor growth of bananas in a greenhouse. They also used a temperature regulation card to keep the greenhouse temperature around 25 degree C for a better production of banana. Commercial data loggers do not reveal their design and provide a limited functionality. Data loggers have been used for energy analysis of houses. Shuqin [6] investigated residential house energy consumption patterns in summer in seven cities in China. It was observed that Hong Kong uses highest amount of household electricity compared to other cities. Lindberg [7] analysed 5 year long building energy consumption data, collected from a data acquisition system which used almost 520 sensors in each of 6 buildings. For different exterior wall material the data was collected for analysis. They showed for reducing building heating and cooling energy load, use of thermal mass of a house has significant effect. Authors [8, 9] presented an energy and economic analysis of a zero energy house with solar application and a conventional house. Among several conclusions it was observed that an air conditioner with a water-cooled air condenser saves noticeable electricity. Full design of a data logger and energy analysis of electrically heated houses is missing in the literature. A design of a low cost data logger and energy analysis of two houses is presented below.

II. Data logger Design

A data logger is an electronic instrument that records different physical or electrical parameters such as temperature, humidity, wind speed, current through an appliance over time. Data loggers are based on, a computer connected to a number of sensors and data storage. Power consumption in computers is high and data is lost in case of a power failure. Recent availability of low cost and low power microcontrollers allowed new design of data loggers. In this section a data logger design is introduced which records current consumption in a house and both inside and outside temperature of a house.

Four current sensors and two temperature sensors have been used in the design. Data logger circuit diagram is shown in figure 1. Figure 2 shows a photo of designed data logger. A 2x16 LCD is used to display required measurements. A microchip microcontroller PIC18F4550 is used for recording the data from the sensors. Data is stored in separate files on a sd card. Data taken from the temperature sensor is shown on a 2x16 LCD. And data taken from current sensors are directly saved in separate files on the sd card. For further analysis the data from sd card can easily be transferred to a computer.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

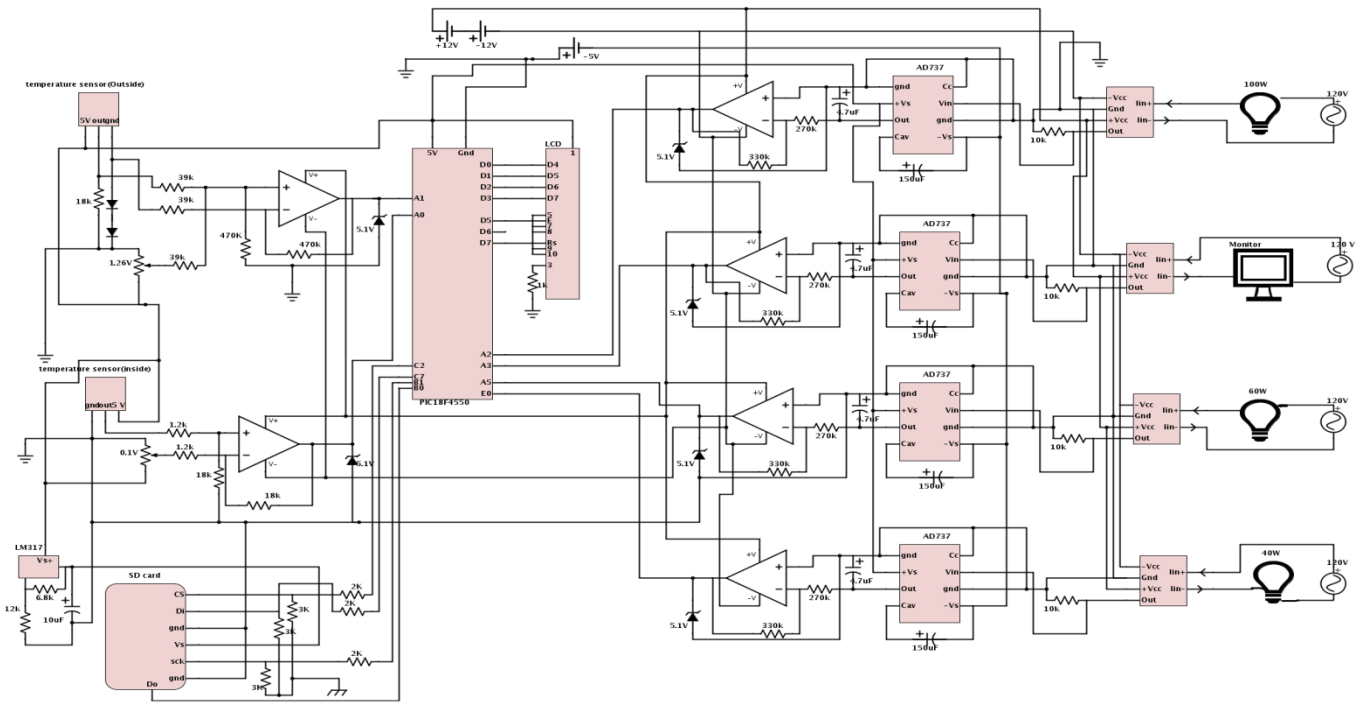


Fig.1 Data logger circuit diagram

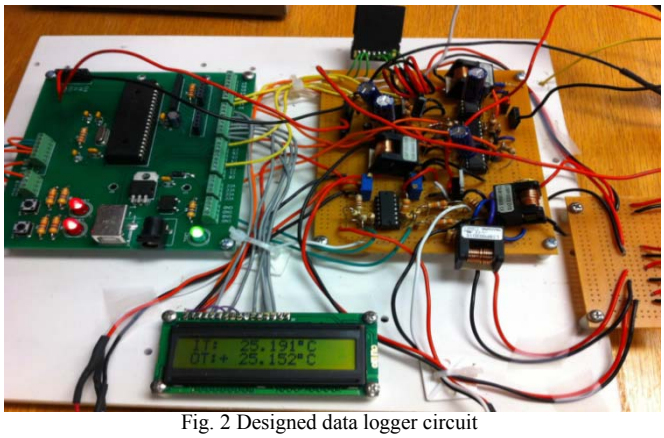


Fig. 2 Designed data logger circuit

A. Temperature sensor

LM35 series temperature sensor has been used whose output voltage is linearly proportional to the Centigrade temperature. Scaling factor of LM35 is +10mv/◦C, which means for every degree Celsius changes in temperature it gives 10mV analog output voltage. Two temperature sensors are used, one for the inside and another for the outside house temperature. Temperature range selected for the inside house temperature is (+10 to+40) C and for the outside temperature is (-30 to+40) C. Outputs of LM35 are amplified to 0-5V for input to microcontroller. Zener diodes (5.1V) are used for circuit protection.

B. Current sensors

Current sensors are connected in series with incoming current to a house in the main panel. Hall sensor based current sensors generate small ac voltage proportional to input current, Rated current for which is ±3A and maximum current is ±9A. Current sensor output voltage is converted to DC using arms to dc converters (AD737). Amplifiers are used to amplify the output of rms to dc converter to 0-5V for PIC microcontroller. Four identical circuits are used for four current sensors.

C. Microcontroller & Sd card

A PIC18F4550 microcontroller is used in the data logger design, which is the most important element in the whole circuit and needs a 5V to operate. This microcontroller uses a 10 bit resolution ADC converter. Both the current and temperature sensors give analog output which goes to the ADC(analog to digital conversion) port of the microcontroller. Thus analog value is converted to digital value to show on the LCD correctly. Conversion from analog to digital value is done using equation (1)

$$\text{Voltage}=(5 * \textit{value read by microcontroller})/1024 \quad (1)$$

A 2GB micro sd card of class 2 is used for the system design. The card is formatted with FAT16 file system. Operating voltage of Sd card is 2.7V-3.6V. A voltage regulator LM317 is used to supply 3.16 V to the sd card in the design. In basic connection 2K and 3K resistors are used on pin 1, 2 & 5. Resistors are used to lower the input voltage of the sd card.

D. Circuit diagram

Full data logger circuit is given in figure1. In the test circuit, different range of bulbs and a computer monitor was connected as the load to measure the currents using current sensors. The AC output voltage of the current sensor is converted to DC voltage using a rms to DC converter. The output of the converter is negative that is why inverting op-amp is used. For programming of PIC18F4550, mikrobasic compiler is used. Once a programme is written and compiled it generates a hex file which is loaded to the PIC using a device programmer. In figure2 the constructed data logger circuit is shown. When the picture was taken both the sensors were inside of the house. So in the LCD it is showing same temperature (25 degree C) for inside and outside of the house. And the current readings are saved in the sd card with both the temperatures. Some actual value readings from the sensors and saved on the sd card are shown in Table-1. Both

the inside and outside temperature are shown and both the sensors were inside of the house while the data was being logged. Outside temperature value is higher because that sensor was close to a light bulb. Value of only one current sensor is shown using a 100w bulb as the load.

TABLE I
ACTUAL SENSOR READINGS

Date	inside temperature C	outside temperature C	current(A)
13/6/14	25.64	26.99	0.834
13/6/14	25.77	27.07	0.834
13/6/14	25.81	27.38	0.834
13/6/14	25.84	29.53	0.834

III. HOUSE ENERGY ANALYSIS

Two typical electrically heated houses in St. John's are selected for the energy analysis. First one (house-1) is a two storey building with a semi-finished basement. Other house is a two storey house with a garage attached to it. From the utility company [2] and using a data logger the total energy consumption data of one year for both the houses was collected.

According to utility company total electricity consumption for house-2 was 27300 kwh in the year 2012(figure-3) and for house-1 total consumption was 15747 kwh in year 2010. Again using a data logger data of a whole year has been logged for both the houses. Sampling was done once every 2 minutes. Figure4 & 6 shows the monthly logged data value in the chart, from Jan (2012-2013) and from September 2003 to August 2004 for house-2 and house-1 respectively. For both the houses collected and logged power consumption data follows a similar pattern. Again in figure 3 and 4 we can see there is a mismatch between the measured and collected data by the utility company. It is because some data logger values were missing but the main reason of mismatch is that utility company do not measure energy consumption for 3 to 4 winter months in a year. Utility send an estimated bill and correct that later when all snow has melted away. Also the year long data is adjusted using year

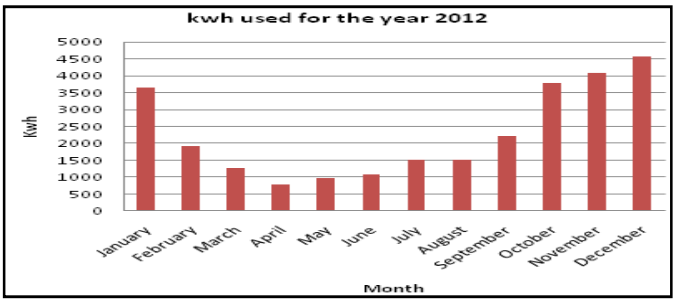


Fig. 3 Energy consumption by the utility in the year 2012 of house-2.

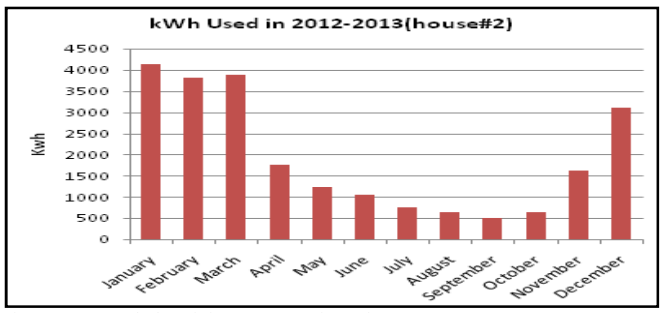


Fig. 4 Measured electricity consumption of year 2012-2013

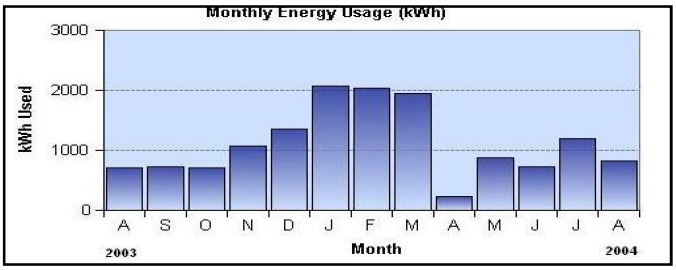


Fig. 5 Energy usage provided by the utility (2003-2004) for house-1.

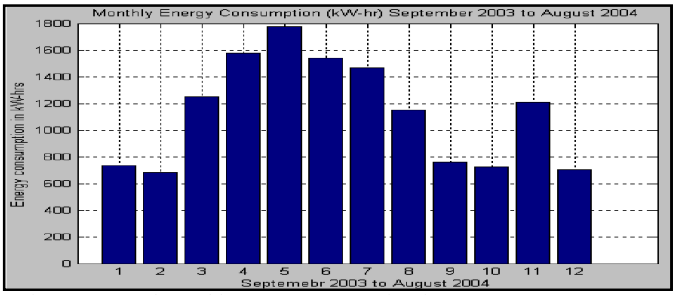


Fig. 6 Measured monthly energy consumption from Sept. 2003 to August 2004 for house-1

2012 and 2013 power consumption data. Figure 5 shows monthly energy consumption of house-1 from August 2003 to August 2004 as reported by the utility company. On the other hand figure 6 shows the measured energy consumption for year 2003-2004.Energy consumption in the year 2003-2004 was maximum in January 2004 (2084kWhr) while it was minimum in April 2004 (only 237 kWhr) [9]. It is clear from this data that measurements were not done at a regular interval and it also includes estimated values. In other words utility reported measurements are not a good indication of actual energy consumed in the house during winter months when their meter is not accessible. If we compare data in figure 5 and figure 6 we find similarity in pattern but energy consumption value reported by the Newfoundland power (fig. 5) and measured value (fig .6) for the month of April 2004 do not match. This mismatch is because of adjustment done by the Newfoundland power during that month. Figure 7 shows power consumption on a typical winter day in house 2. The start time of the data is

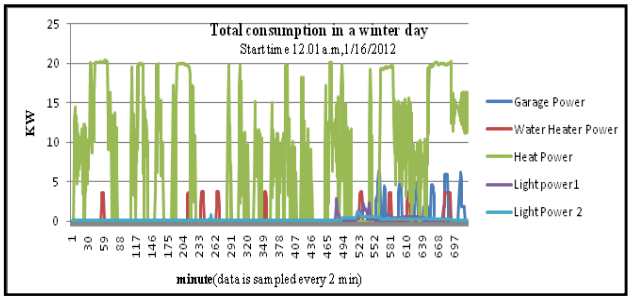


Fig. 7 Power consumption in typical winter day of year 2012

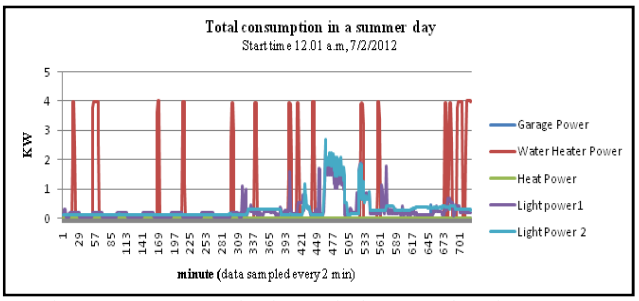


Fig. 8 Power consumption in a typical summer day of year 2012.

12.01 am at night. We can see that the main power consumption is due to heating. It indicated that most heating was done during night and late evening. Similarly Power consumption for the same house in a typical summer day is shown in figure 8. We can see there is no house heating power for the day. The major electrical load was electric water heater. In figure 9 hourly averages for a day in winter

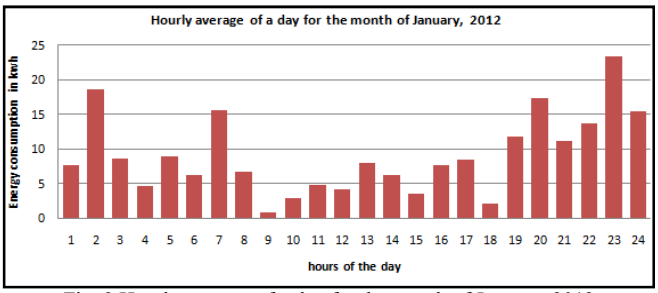


Fig. 9 Hourly average of a day for the month of January, 2012.

is shown using bar chart. We can conclude from this chart that consumption is high at night, late evening and in early morning. Similarly hourly average of a day in summer is

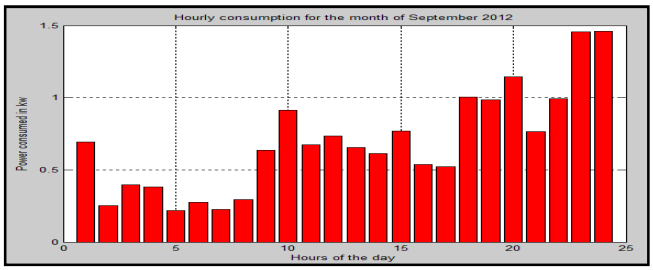


Fig. 10 Hourly average power consumption in September of house-2

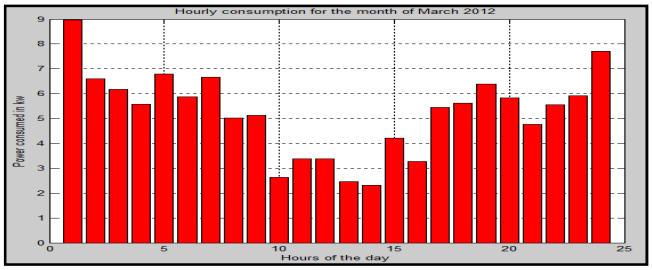


Fig. 11 Hourly average power consumption in March of house-2

observed where it can be noticed that the consumption is high in the afternoon. Again for house 2, from figure 10 & 11 we observe minimum hourly power consumption was below 1.5 kW in September whereas the maximum consumption was 9kW in the beginning of March in year 2012. For house-1, it is observed from figure12 that the maximum average hourly power consumption was in January 2004, which exceeded 5000W level. Similarly from figure13 the minimum hourly power consumption was 250W in early hours of the day in

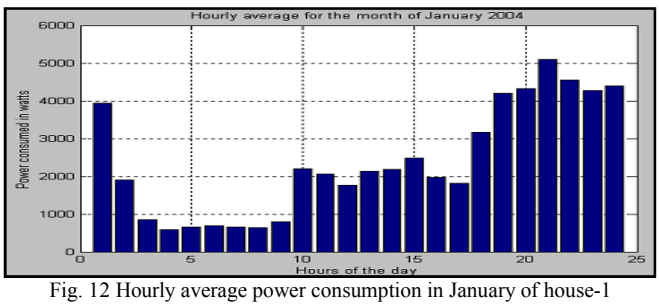


Fig. 12 Hourly average power consumption in January of house-1

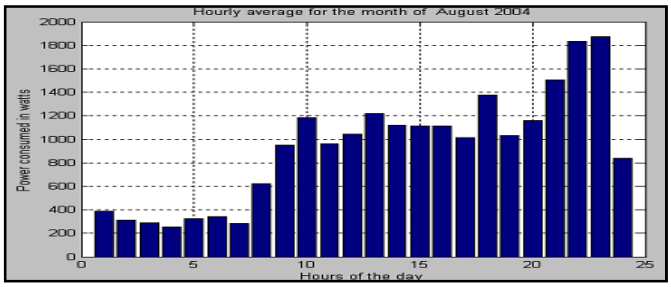


Fig. 13 Hourly average power consumption in August of house-1

August 2004. The hourly power consumption graphs are generated using matlab code using all the collected data, sampled every 2 min from the data loggers.

IV. Conclusions

Two houses in St. John's have been considered to analyse the power consumption for one year. Data of power consumption for both the houses is collected from utility company and then compared with the measured value from data logger. Data acquisition system was installed at home to give a true picture of energy consumption. A data logger that is used in the house is designed using two temperature sensors and four current sensors. This type of data logger can be used in almost every household to get different parameters of a house within a certain interval. Daily, monthly and yearly data can be analysed to understand the exact power consumption nature of the houses. Hourly average of each month and energy consumption in a winter day and in a summer day is shown by a time series plot for the house. Such plots are necessary to design a renewable energy system for a house and also to study human behaviour and energy usage.

Acknowledgment

Authors would like to thank National Science and Engineering Research Council (NSERC) for providing supports for this research.

.

References

[1] http://www.stats.gov.nl.ca/
[2] http://www.newfoundlandpower.com
[3] NilimamayeeSamal, and Umesh Chandra Pati.,'Multi-Channel Data Acquisition and Data Logging for Green House Application,' 2014 IEEE Students' Conference on Electrical, Electronics and Computer Science.
[4] S. Rosiek, F.J. Batlles., 'A microcontroller-based data-acquisition system for meteorological station monitoring,' Energy Conversion and Management 49 (2008) 3746–3754
[5] S. Ameur, M. Laghrouche, A. Adane.,'Monitoring a greenhouse using a microcontroller-based meteorological data acquisition system,'Renewable Energy 24 (2001) 19–30
[6] Shuqin Chena, Hiroshi Yoshino, NianpingLi ., 'Statistical analyses on summer energy consumption characteristics of residential buildings in some cities of China,' Energy and Buildings 42 (2010) 136–146.
[7] R. Lindberg, A. Binamu, M. Teikari.,'Five-year data of measured weather, energy consumption, and time-dependent temperature variations within different exterior wall structures,' Energy and Buildings 36 (2004) 495–501
[8] L. Zhu, R. Hurt, D. Correa, R. Boehm.,'Comprehensive energy and economic analyses on a zero energy house versus a conventional house,'Energy 34 (2009) 1043–1053
[9] M. T. Iqbal.,'Year long power consumption and metrological data analysis of a home in St. John's, Newfoundland,' NECEC 2004, St. John's, NL

Switched Capacitor Based Soft-Switching DC-DC Boost Converter for High Voltage Gain

Md. Rifat Alam Siddique[1], Md. Jannatul Ferdous[1], and Istiaque Islam[2]

[1]Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology (BUET)
Dhaka - 1205, Bangladesh

[2]Department of Electrical and Electronic Engineering, Islamic University of Technology
Gazipur- 1704, Bangladesh

E-mail: rifatalam99@gmail.com, j.ferdous63@yahoo.com, isti9065@yahoo.com

***Abstract*— A switched capacitor based step up converter with high voltage gain is proposed in this paper. The proposed novel topology of the boost converter can provide almost threefold output voltage relative to its input at 50% duty cycle. In order to guarantee small switching losses and to eliminate the current spikes problem across the switches (which usually exists in classical converter), an inductor is used in the circuit which creates a resonant tank and assist in zero current switching (ZCS). In addition, the efficiency curve of the new converter circuit shows high value for a wide range of output power which is rare in literature. Besides, this proposed topology is free from the circuitry design complexity. In renewable energy systems like photovoltaic panels or fuel cells, where a high step up voltage is required, the proposed converter can bear its suitability.**

Keywords*— Switched capacitor, high voltage gain, zero current switching (ZCS), high efficiency, DC-DC boost converter.

I. Introduction

Boost converters have been widely used in many applications such as battery backup devices, uninterruptible power supply (UPS), wind turbine system, photovoltaic (PV) system etc. In order to satisfy the high DC voltage requirement as well as power quality, many researches have been conducted so far on how to design the DC-DC boost converter [1]-[2]. A conventional single phase single switch boost converter is shown in Fig. 1 where V_{gs} is the gate signal of the switch; V_s, I_s are the voltage and current through the switch and V_0, I_0 are the voltage and current at the output. The conventional boost converters are usually employed in high voltage and power factor correction (PFC) application due to its simple circuit structure. In theory, the voltage gain of the boost converter can be infinite when the duty cycle is close to one. However, the switch turn off period becomes short when the duty cycle increases towards one. As a consequence, the current ripples of the power devices are large with high step-up conversion which increases the power device conduction losses. Moreover, the voltage stress of the switch and the diode are proportional to the output voltage which is large in high output voltage applications [3].

To overcome the aforementioned problem and to increase the current rating, interleaved boost converters have been proposed in the literature. Interleaved boost converter in [3]-[8] can increase the power level and can minimize the current ripple but the efficiency of these converters are still limited by the output diode reverse recovery problem in high voltage applications. Besides, the use of bulky inductor as a boosting element increases the overall size, weight and losses of the interleaved boost converter. In order to remove the bulky inductor and to get high voltage conversion ratio, the switched capacitor concept based boost converters have been proposed in [9]-[11]. But the problem with these topologies is that, some switches remain floating in the structure and thereby it needs additional isolated gate driving circuits to operate which eventually makes the system more complex. Moreover, the converter still operates in hard switching mode resulting in large switching losses as well as high current spikes across MOSFETs. To minimize the switching losses and high current spikes problem across the switches, a new soft switching DC-DC boost converter circuit based on capacitor switching topology is proposed in this paper. The proposed converter can produce threefold voltage at its output compared to its input voltage at a fixed duty cycle (50%). A 1μH small inductor is placed in series with the input source to assist the zero current switching (ZCS) in the circuit. This ensures low switching losses and eliminates current spikes across the switch. Besides, the circuit shows fairly high efficiency over a broad range of output power. In order to prove the aforementioned features of the proposed circuit, simulations have been carried out which demonstrates the effectiveness of the circuit more clearly.

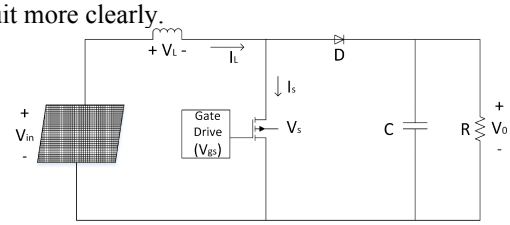


Fig. 1. Single phase single switch boost converter.

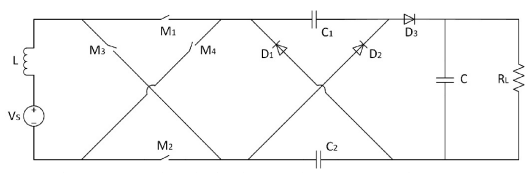


Fig. 2. The proposed switched capacitor DC-DC boost converter

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

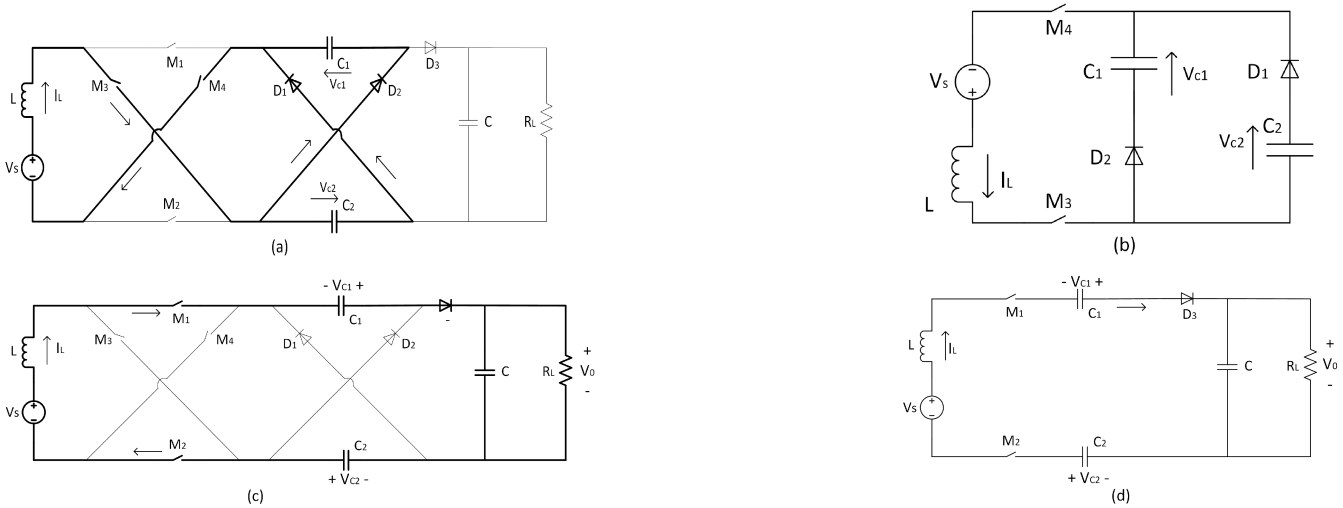


Fig. 3. (a) Current path of the proposed boost converter circuit for first state [t_0-t_1], (b) Equivalent circuit of the first state, (c) Current path for the third state [t_2-t_3], (d) Equivalent circuit of the third state.

II. Analysis of the Proposed Circuit Operation

The operating principle of the proposed boost converter with commutation is discussed step by step. But before analysing the operation of the circuit, several assumptions are made. It is considered that all the devices are ideal and the capacitance of the output capacitor is large enough so that the output voltage is fairly constant. All the simulated curves of current at different switching states are shown in Fig. 4. Here V_{g1} and V_{g2} are gate pulse of Switch M_1, M_2 & M_3, M_4 respectively.

1) First State [t_0-t_1]: At t_0, M_3 and M_4 switches are turned on. This creates a current path between the voltage source V_s and the capacitors C_1 & C_2. Fig. 3(a) shows the current path of the proposed converter and Fig. 3(b) is the simplified circuit at this state. Here the two diodes D_1 and D_2 are in a conducting state. Due to this close current path, capacitor C_1 and C_2 both get the opportunity to become charged to a voltage of V_s. Hence the inductor L and both capacitor C_1 and C_2 resonate together with voltage source V_s. They all start resonating at t_0 from the current equal to zero in sinusoidal manner. Since the current increases gradually at t_0, so both of the switches M_3 and M_4 are turned on under zero current switching (ZCS) condition. They stop resonating when the current reaches zero again at t_1 and both of the diodes (D_1 and D_2) will be reverse biased.

2) Second State [t_1-t_2]: At t_1, resonance stops and the inductor current becomes zero. The currents through the switches are also zero. The switches M_3 and M_4 are turned off at this state under zero current switching (ZCS). The voltage of both the capacitors (C_1 and C_2) remains unchanged.

3) Third State [t_2-t_3]: At t_2, switch M_1 and M_2 are turned on and diode D_3 is in a conducting mode. Both of the switches are turned on under ZCS condition. Total circuit configuration is illustrated at Fig. 3(c) and its equivalent circuit is displayed at Fig. 3(d). Now it is evident from Fig. 3(d) that we can get the output voltage from the capacitor C which discharges to the load R_L. To perform the ZCS in the circuit, a small valued inductor (1μH) is added in series with the voltage source. The voltage induced across the inductor is too small due to its small value and we can neglect it. Now neglecting the inductor voltage we can write-

$$V_0 \cong V_s + V_{C1} + V_{C2} \tag{1}$$

Where V_{c1} and V_{c2} are the voltage of capacitors C_1 and C_2 respectively which have same value and it is almost equal to the input voltage V_s (50V). From the above expression it is vivid that the proposed converter circuit provides almost three times larger voltage at the output relative to its input. In addition, the current across the inductor changes in a sinusoidal manner at this state. Mathematically, we can express-

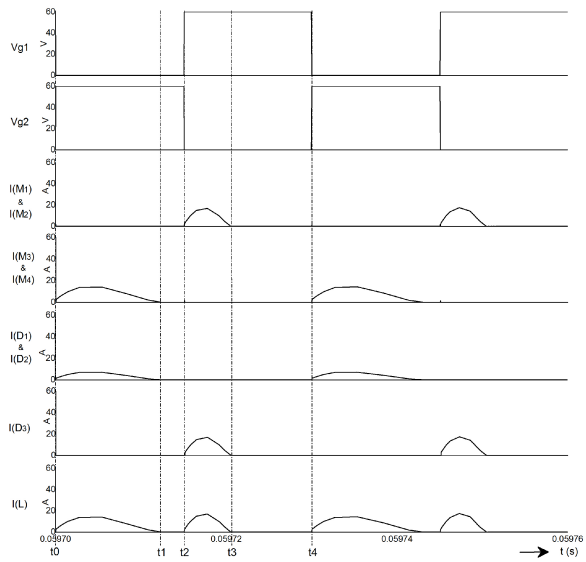


Fig. 4. Analysis waveforms of the proposed converter.

$$I_L = I_M \sin \omega_0 (t - t_2) \quad (2)$$

Here, I_M=Maximum amplitude of the inductor current.

$$\omega_0 = \frac{1}{\sqrt{2LC_{eq}}}$$

C_{eq} is the equivalent capacitance of the circuit.

4) Fourth State [t_3-t_4]: At t_3, the current through the inductor (L) as well as the switches M_1 and M_2 becomes zero. So, at this state, both switches M_1 and M_2 are turned off under ZCS condition.

III. Performance Appraisal

The proposed converter can manipulate the high efficiency power conversion and provide a wide range constant efficiency curve. We have compared the conversion performance result for hard switching and soft switching for our proposed converter circuit. The simulations are carried out in PSPICE software and the parameters are given in TABLE I.

TABLE I. KEY PARAMETER CONSIDERATIONS

PARAMETERS	VALUE
V_{in}	50V
V_{out}	More than 140V
Switching frequency f_{sw}	33.3KHz
Duty Cycle	50%
Charge pump capacitor C_1,C_2	6 µF
Output Capacitor, C	1500 µF
Inductor, L	1 µH
MOSFETs	IRF530
Diode	MBR10100

In Fig. 5, the output voltage curves of the capacitors are depicted when hard switching is occurred. Fig. 5(a) and Fig. 5(b) exhibit the output voltage of C_1 and C_2 capacitors while Fig. 5(c) displays the total output voltage across the capacitor C. The voltage profile of the capacitors (C_1 & C_2) is totally undesirable as the charging and discharging state across them is erratic in nature. So the performance of the converter deteriorates during hard switching as the switching loss goes high and we get lower output voltage (almost 130V) across the capacitor C.

On the other hand, Fig. 6 provides the simulated curves during the soft switching (under ZCS condition). It is evident from the Fig. 6(a) and Fig. 6(b) that the voltage profile of the capacitors C_1 and C_2 has improved. These two curves are smooth and unruffled compared to the previous case. Due to soft switching, low switching loss occurs and the converter provides high output voltage (more than 140V which is almost three times relative to the input).

The switching loss, which reduces the output voltage of the converter, occurs due to high current across the MOSFET. Fig. 7 depicts the phenomenon more lucidly. Fig. 7(a) and Fig. 7(b) are the current curves of M_1 and M_3 during the hard switching where it can be observed that a large current spike (more than 200A) generates across the M_1 and M_3. But Fig. 7(c) and Fig. 7(d) represent the current curves during soft switching which indicate the lower magnitude of MOSFET current (around 10A). Due to larger current across the switch, larger switching loss occurs which reduces the output voltage as well as the overall efficiency of the converter.

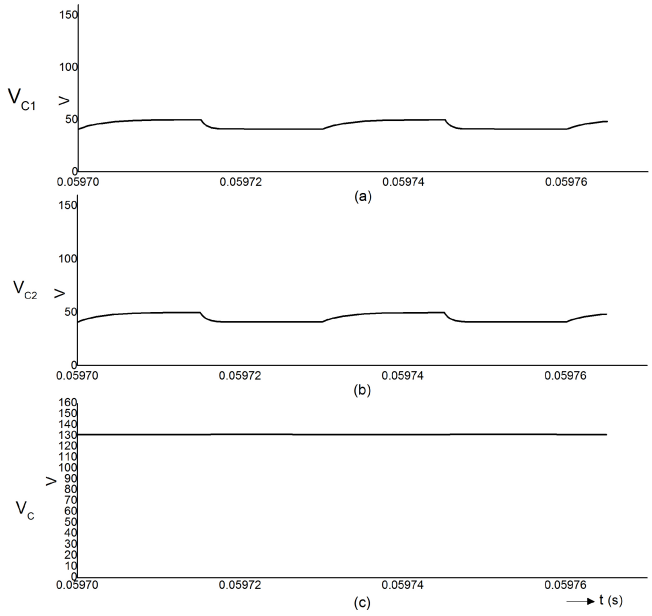


Fig. 5. Voltage without ZCS (a) across capacitor C_1 (b) across capacitor C_2 and (c) across capacitor C

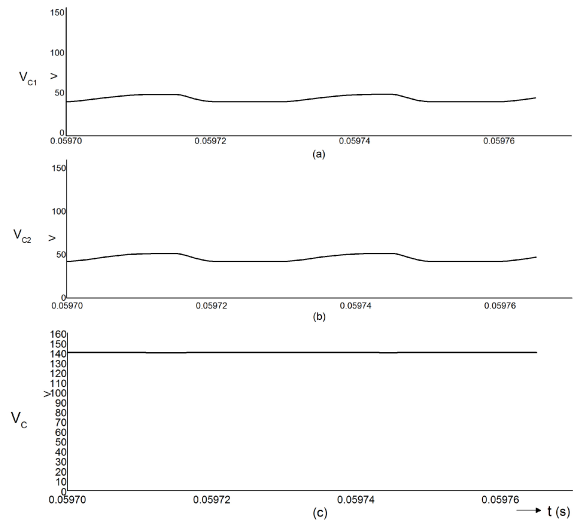


Fig. 6. Voltage with ZCS (a) across capacitor C_1 (b) across capacitor C_2 and (c) across capacitor C

IV. Efficiency Curves

It has previously been mentioned that the converter circuit with hard switching (without ZCS) provides lower efficiency than with soft switching (with ZCS). In order to verify the statement, we have simulated the efficiency curve. Fig. 8 shows the efficiency curve without ZCS where it indicates that highest 82% to 85% efficiency at a certain range of output power. But the Fig. 9 shows both with and without ZCS efficiency curves where it is clearly observed that the efficiency of the converter becomes 92% to 97% at a certain range of output power with ZCS. All of these statements have

been summed up in TABLE II. In addition, the efficiency curve with ZCS also remains at a high value for a wide range of output power which is one of the major features of the proposed converter.

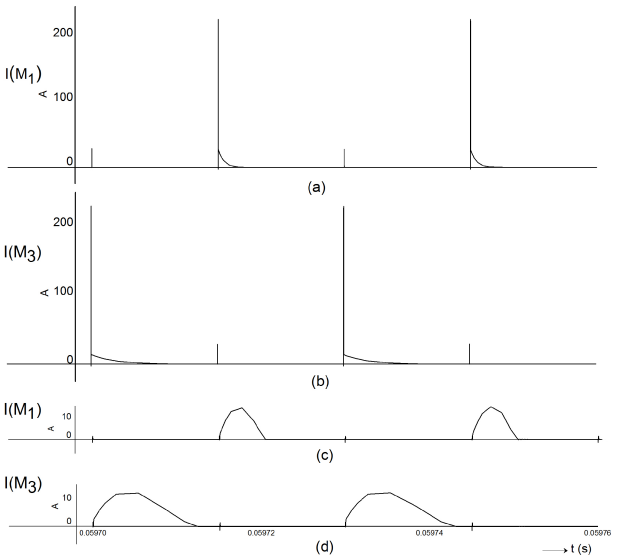


Fig. 7. Output current without ZCS (a) across MOSFET M_1 (b) across MOSFET M_3 and with ZCS (c) across MOSFET M_1 (d) across MOSFET M_3

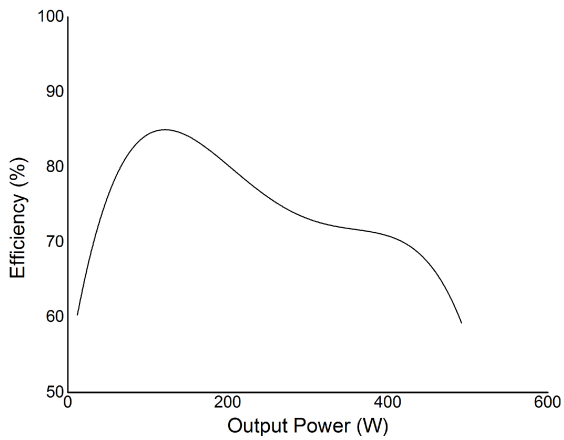


Fig. 8. Efficiency curve for the proposed Boost Converter with Hard switching (without ZCS)

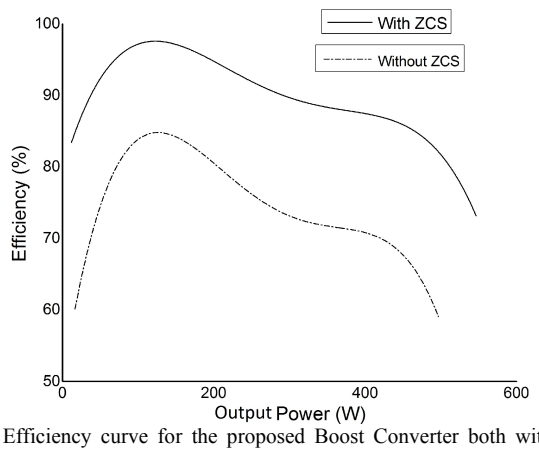


Fig. 9. Efficiency curve for the proposed Boost Converter both with Hard switching and Soft switching (without and with ZCS)

TABLE II. COMPARISON BETWEEN HARD & SOFT SWITCHING

Input Power for Hard Switching (W)	Input Power for Soft Switching (W)	Same Output power for both switching (W)	Efficiency at Hard Switching (%)	Efficiency at Soft Switching (%)
66.85	54.019	50	74.79	92.56
119.24	102.88	100	83.86	97.20
178.31	154.44	150	84.12	97.12
248.57	210.83	200	80.46	94.86
328.68	271.97	250	76.06	91.92
410.73	334.70	300	73.04	89.63
488.28	396.24	350	71.68	88.33
565.61	457.19	400	70.72	87.49
665.77	524.04	450	67.59	85.87

V. CONCLUSION

In this paper, a novel topology of boost converter is proposed and analyzed which can provide a high voltage gain at its output. Moreover, a method of soft switching with ZCS is presented that ensures a minimum switching loss and eliminates the current spikes which usually exists in classical converter. The proposed circuit is simulated and the results are presented in this paper. The efficiency curve exhibits that, for a wide range of output power, the efficiency of the circuit remains at a high value. So, for a system where large voltage step-up is necessary, this proposed boost converter suits properly.

REFERENCES

[1] S. C. Smith, P. K. Sen, and B. Kroposki, "Advancement of energy storage devices and applications in electrical power system," *inProc. IEEE Power Energy Soc. General Meeting*, Jul. 2008, pp. 1–8.

[2] R. D. Middlebrook and S. Cuk, "A general unified approach to modeling switching-converter power stages," in *Advances in Switched-Mode Power Conversion*. Pasadena, CA: TESLAco, 1983, vol. 1, pp. 73–89.

[3] Wuhua Li, Xiangning He, "Review of Nonisolated High-Step-Up DC/DC Converters in Photovoltaic Grid-Connected Applications", *IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS,* vol. 58, no. 4, April, 2011.

[4] P. W. Lee, Y. S. Lee, D. K. W. Cheng, and X. C. Liu,"Steady-state analysis of an interleaved boost converter with coupled inductors," *IEEE Trans. Ind. Electron.*, vol. 47, no. 4, pp. 787–795, Aug. 2000.

[5] L. Balogh and R. Redl, "Power-factor correction with interleaved boost converters in continuous inductor-current mode," in *Proc. IEEE APEC'93 Conf., San Diego, CA*, Mar. 1993, pp. 168–174.

[6] H. A. C. Braga and I. Barbi,"A 3-kW unity-power-factor rectifier based on a two-cell boost converter using a new parallel-connection technique," *IEEE Trans. Power Electron*., vol. 14, no. 1, pp. 209–217, Jan.1999.

[7] M. S. Elmore,"Input current ripple cancellation in synchronized, parallel connected critically continuous boost converters," in *Proc. IEEE APEC'96 Conf., San Jose, CA*, Mar. 1996, pp. 152–158.

[8] C. H. Chan and M. H. Pong,"Interleaved boost power factor corrector operating in discontinuous-inductor-current mode," in *Proc. IEEE PCC'97 Conf.*, Nagaoka, Japan, Aug. 1997, vol. 1, pp. 405–410.

[9] O. Abutbul, A. Gherlitz, Y. Berkovich, and A. Ioinovici, "Step-up switching-mode converter with high voltage gain using a switched capacitor circuit," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl*., vol. 50, no. 8, pp. 1098–1102, Aug. 2003.

[10] B. Axelrod, Y. Berkovich, and A. Ioinovici, "Transformerless DC–DC converters with a very high DC line-to-load voltage ratio," inProc. *IEEE ISCAS*, May 2003, vol. 3, pp. 435–438.

[11] F. L. Luo and H. Ye, "Ultra-lift Luo-converter," *IEE Proc. Electr. Power Appl*., vol. 152, no. 1, pp. 27–32, 2005.

Design of a charge controller based on SEPIC and Buck topology using modified Incremental Conductance MPPT

Taohid Latif[1] and Syed R. Hussain[2]

[1]Department of Electrical and Computer Engineering, North South University, Bangladesh
[2]Department of Electrical and Electronics Engineering, American International University-Bangladesh, Bangladesh
[1]taohidlatif@gmail.com, [2]rezwanur.hussain@gmail.com

Abstract—**Design and simulation of a fast, efficient and cost effective charge controller for Photovoltaic (PV) System is presented. To harness maximum power from the PV panel, the charge controller is equipped with Maximum Power Point Tracking (MPPT) feature. An efficient algorithm for MPPT is developed combining the Constant Voltage (CV) MPPT and the Incremental Conductance (IncCond) MPPT method. The new method, unlike CV, can obtain the actual power from PV and work faster than IncCond. The open circuit voltage, V_{oc}, needs to be taken only once in order to minimize the power loss. The performance of this charge controller is demonstrated via the simulation in Powersim (PSIM) using the parameters of a practical solar panel (Solarex MSX-60). Furthermore, this paper delineates elementary concepts of PV system so as to help beginners get a strong basic idea about PV systems and MPPT.**

Index Terms—**charge controller; dc-dc converter; MPPT; Constant Voltage MPPT; Incremental Conductance MPPT; SEPIC converter; Buck Converter.**

I. Introduction

Even today when global warming is an important issue of discussion not only among environmentalists, but also among commoners and industrialists alike, we still rely on fossil fuels for much of our energy needs. Only 10% of global energy supplied comes from the renewable energy sources [1-13]. Although Sun is the primary source of all energy, solar energy plays such a miniscule role that it is listed behind the other sources of energy in terms of the contribution for meeting the world's energy demand. Photovoltaic (PV) energy is one of the leading sources of renewable energy, which gets more preference in comparison to other forms of alternate energy due to its availability, simplicity, lower maintenance, environmental friendliness, reliability and many other benefits.

A PV array (which is comprised of several solar cells) uses photons of light to generate electricity through photovoltaic effect – the creation of voltage or electric current in a material upon exposure to light. The amount of solar energy that strikes a given area over a specific time is called insolation. Insolation per unit time is defined as irradiance. As the irradiance increases, more power is generated.

The temperature also has an effect on the solar cells in that the higher the temperature, the lower the power generated. Even at fixed irradiance and temperature, a solar cell does not produce steady voltage or current. I-V characteristic of the solar cell changes as the load is varied across it. The I-V and subsequent P-V characteristics of a typical solar cell are provided in Fig. 1 and Fig. 2 [13].

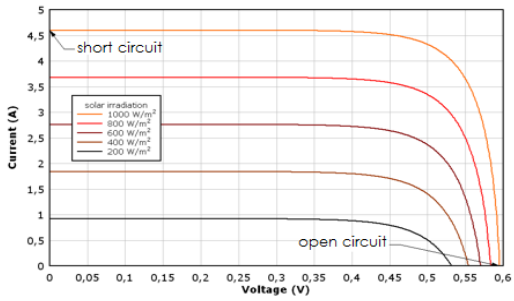


Fig. 1 I-V characteristic of solar cell for different irradiance levels

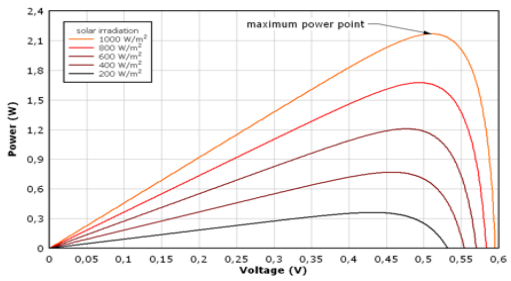


Fig. 2 P-V characteristic of solar panel for different irradiance level

A PV cell has non-linear current-voltage qualities. The power delivered by the PV array increase, up to a certain point, as more and more current is drawn from it. In order to ensure that a PV array is producing the maximum output power, a technique called Maximum Power Point Tracking (MPPT) is used. The voltage of the PV array, at which the maximum power is delivered, can be defined as V_{mpp}.

There are many MPPT algorithms [11-20]. Some of the more commonly used and effective MPPT techniques are the Constant Voltage (CV) method [3,4], β method [5], System Oscillation method [6,7], Ripple Correlation method [8], Incremental Conductance method (IncCond) [9] and the Perturb and Observe method [10-12]. Among the MPPTs CV has the easiest algorithm, hence easiest to implement. The power delivered to the load is momentarily interrupted and the open-circuit voltage (V_{oc}) is measured. The controller then resumes operation with the voltage controlled at a fixed ratio, such as 0.76, of V_{oc}. A particular constant, k, is defined as:

$$k = V_{mpp}/V_{oc} \quad (1)$$

The value of the duty cycle (D) is then calculated from the equation:

$$D = (k \times V_{mpp})/V_{oc} \quad (2)$$

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

This method however, yields an approximate result and the variation in V_{mpp} due to climate change is not taken into account.

The theory of the incremental conductance method is to determine the variation in direction of the terminal voltage for PV modules by measuring and comparing the incremental conductance (I/V) and the instantaneous conductance (ΔI/ΔV) of PV modules. If the value of the incremental conductance is equal to that of the instantaneous conductance, then the MPP is obtained; the MPPT algorithm calculates the optimal duty cycle D in order to maximize the power flow. IncCond needs less iteration than the Perturb and Observe (P&O) method and consequently needs a lower time to reach the MPP. Moreover, unlike the P&O, the IncCond technique allows the MPP to be reached without any fluctuation around the MPP.

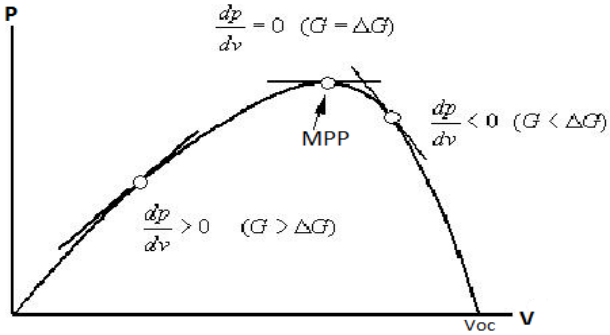


Fig. 3 Incremental conductance at different position of P-V curve

II. Proposed MPPT Technique

A MPPT method combining CV and the IncCond methods has been proposed. The new MPPT takes open circuit voltage by momentarily disconnecting load. During that time a capacitor is used to supply power to load. By switching gate of the SEPIC converter, which is used for performing the MPPT operation, at a fixed rate, a fraction of V_{oc} is supplied to the load. Next by varying duty cycle according to IncCond algorithm, MPP is reached. The algorithm is depicted in Fig. 4.

III. Proposed Design

The design and simulation of the proposed charge controller was performed using the PSIM simulation software. The block diagram of the proposed charge controller is shown in Fig. 4. The input was taken from the built-in PV model of the software. Irradiance and temperature were required as input for the PV model. It was then connected to a SEPIC converter in which PWM switching signals where provided by the MPPT algorithm. The output of the sepic converter was then connected to a buck converter in order to ensure that the voltage output of the buck converter is kept at 14 V which is required in order to effectively charge a 12 V battery.

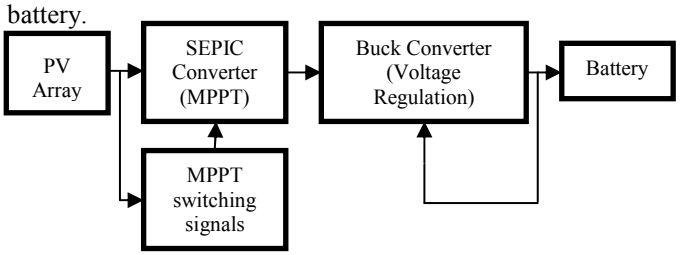


Fig. 5 Block Diagram of the Charge Controller System

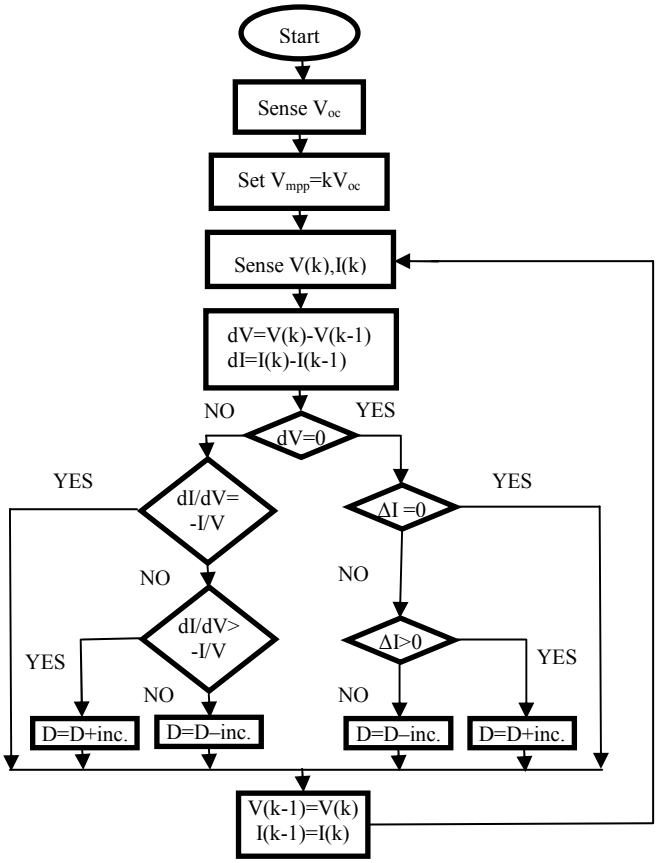


Fig. 4 Algorithm of the proposed MPPT technique

A. SEPIC Converter

The PV array is connected to the SEPIC Converter in which the gating signals are provided via the MPPT Algorithm discussed earlier. A SEPIC is a DC-DC converter similar to a buck-boost converter [14] which is helpful in applications where the input voltage is above and/or below the intended output. In comparison to the buck-boost converter, it has an output with the same polarity as the input voltage and via the use of the series and additional capacitors can respond more quickly to the changes in input and output. The voltage and current values of the PV array are measured by a voltage and a current sensor respectively and the gates of the SEPIC converter are thus switched accordingly in order to ensure that the PV array is in the ideal operating condition. The parameters of the SEPIC converter are defined by the equations given below

$$D = \frac{V_o + V_f}{V_i + V_o + V_f} \tag{3}$$

$$\Delta I_L = 30\% \times \frac{I_i}{n} \tag{4}$$

$$L1a(\min) = L1b(\min) = \frac{1}{2} \times \frac{V_{i(\min)} \times D_{\max}}{\Delta I_L \times f_{sw(\min)}} \tag{5}$$

$$C_o \geq \frac{I_o \times D_{\max}}{\Delta V_{rpl} \times f_{sw(\min)}} \tag{6}$$

Where D = duty of the converter, V_f = forward voltage drop of the diode, V_i = input voltage, V_o = output voltage, ΔI_L = inductor ripple current, η = worst-case efficiency, L1a = L1b = inductor values, f_{sw} = switching frequency, C_o = Capacitor value, I_o = output current

B. Buck Converter

The output of the SEPIC converter is connected to a buck converter which is in turn used for voltage regulation. In a buck converter, the output voltage is less than the input voltage [15, 16]. The output of the buck converter was compared with a reference voltage and the error signal was sent to the PI controller, the output of the PI controller was then compared with a triangular wave and the resultant pulses were then used for switching of the MOSFET buck converter. The parameters of the buck converter are defined by the equations given

$$V_o = D \times V_i \tag{7}$$

$$\Delta I_L = \frac{V_o(V_i - V_o)}{f_{sw} \times L \times V_i} \tag{8}$$

$$\Delta V_{rpl} = \frac{\Delta I_L}{8 \times f_{sw} \times C} \tag{9}$$

Where D = duty of the converter, V_i = input voltage, V_o = output voltage, ΔI_L = inductor ripple current, L = inductor value, f_{sw} = switching frequency, C = Capacitor value, I_o = output current, ΔV_{rpl} = output ripple voltage

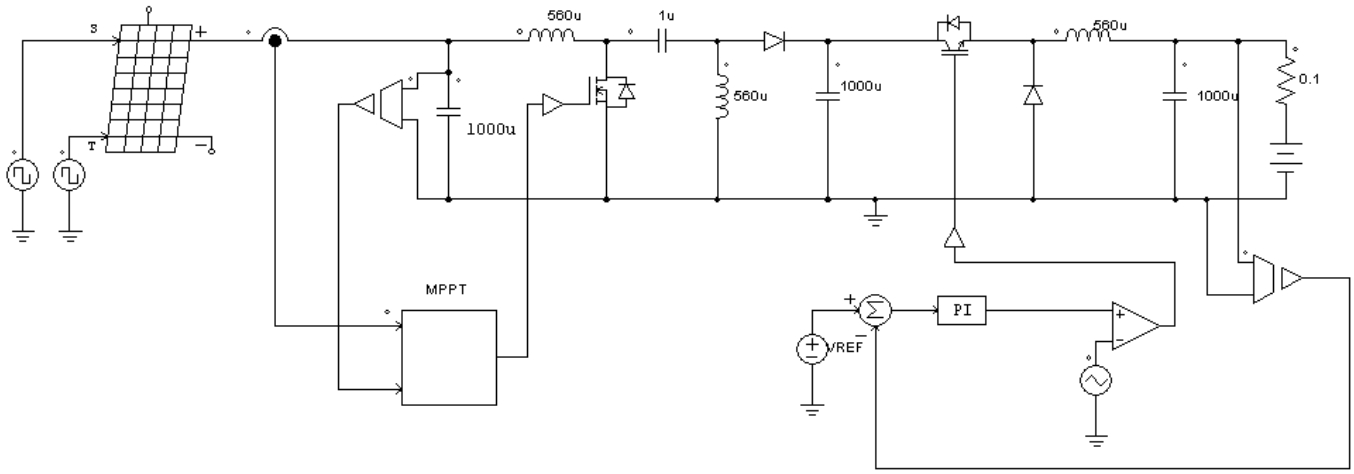


Fig. 5 Full Circuit Diagram

IV. Simulation Results

The PSIM simulation results are provided in this section. The irradiance was varied from 600 to 1000 W/m^2.

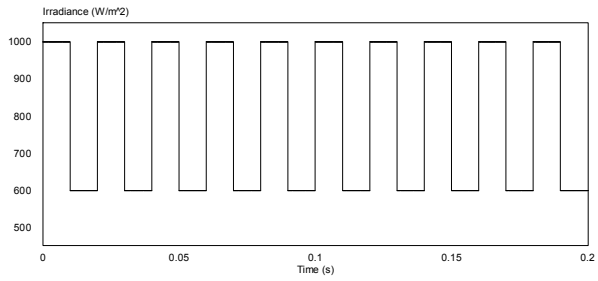


Fig. 6 Irradiance variation

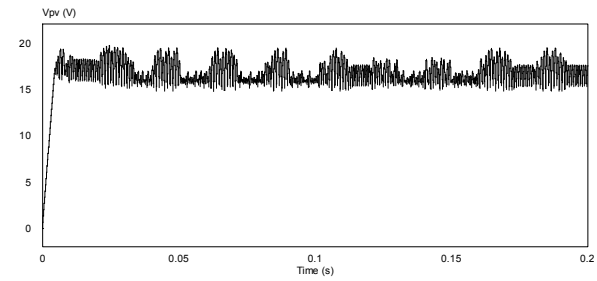


Fig. 7 Voltage from the PV module

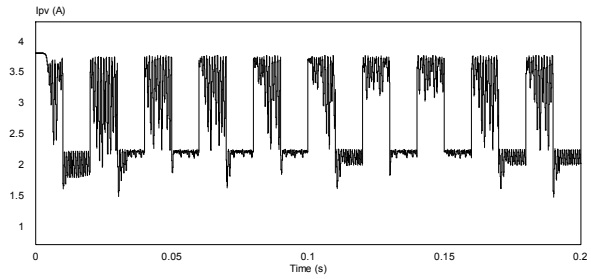


Fig. 8 Current from the PV module

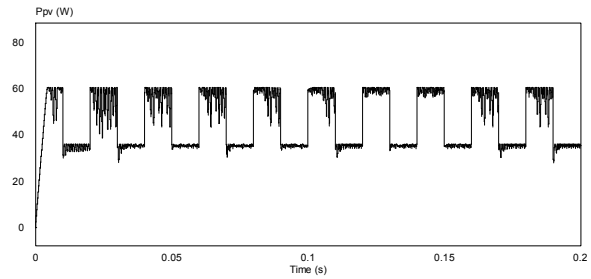


Fig. 9 Power from the PV module

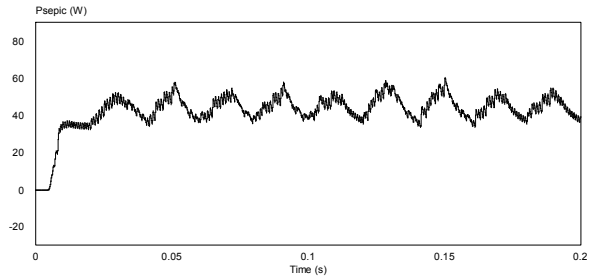


Fig. 10 Power output from the SEPIC converter

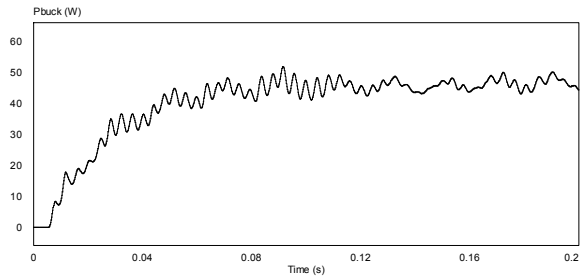


Fig. 11 Power output from the Buck converter

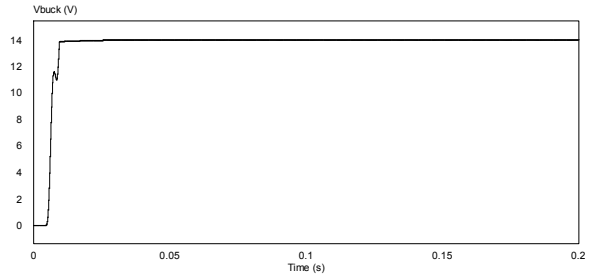


Fig. 12 Voltage output from the Buck converter

V. Conclusions

Design of a charge controller using SEPIC and buck converters along with modified Incremental Conductance has been presented. The simulation results depicting the behaviour of the charge controller have been presented in the previous section. In the future, the simulation results would be expanded and the hardware of the charge controller would also be constructed.

References

[1] S. M. Metev IEC 2013 key world energy statistics and V. P. Veiko, *Laser Assisted Micro technology*, 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.

[2] J. Breckling, Ed., *The Analysis of Directional Time Series: Applications to Wind Speed and Direction*, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.

[3] S. Zhang, C. Zhu, J. K. O. Sin, and P. K. T. Mok, "A novel ultrathin elevated channel low-temperature poly-Si TFT," *IEEE Electron Device Lett.*, vol. 20, pp. 569–571, Nov. 1999.

[4] M. Wegmuller, J. P. von der Weid, P. Oberson, and N. Gisin, "High resolution fiber distributed measurements with coherent OFDR," in *Proc. ECOC'00*, 2000, paper 11.3.4, p. 109.

[5] R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, "High-speed digital-to-RF converter," U.S. Patent 5 668 842, Sept. 16, 1997.

[6] (2002) The IEEE website. [Online]. Available: http://www.ieee.org/

[7] M. Shell. (2002) IEEEtran homepage on CTAN. [Online]. Available: http://www.ctan.org/tex-archive/macros/latex/contrib/supported/IEEEtran/

[8] *FLEXChip Signal Processor (MC68175/D)*, Motorola, 1996.

[9] "PDCA12-70 data sheet," Opto Speed SA, Mezzovico, Switzerland.

[10] A. Karnik, "Performance of TCP congestion control with rate feedback: TCP/ABR and rate adaptive TCP/IP," M. Eng. thesis, Indian Institute of Science, Bangalore, India, Jan. 1999.

[11] J. Padhye, V. Firoiu, and D. Towsley, "A stochastic model of TCP Reno congestion avoidance and control," Univ. of Massachusetts, Amherst, MA, CMPSCI Tech. Rep. 99-02, 1999.

[12] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification*, IEEE Std. 802.11, 1997.

[13] PV Resources. [Online]. Available: http://www.pvresources.com/introduction/solarcells.asp/

[14] Jeff Falin, *Designing DC/DC converter based on SEPIC topology,* Texas Instrument Inc., 2008.

[15] Muhammad H. Rashid*, Power Electronics Handbook,* 2nd ed., Ed. Academic Press, 2010.

[16] *PSIM User Manual,* Powersim Technologies Inc., 2012.

Reliable Fault Diagnosis of Induction Motors using 2D Signal Processing Techniques

Jia Uddin,[1,*] Md. Rashedul Islam[1], and Jong-Myon Kim[1]

[1]Department of Electrical, Electronics, and Computer Engineering, University of Ulsan

Bldg. #7, Room #308, 93 Daehak-ro Mugeo-dong, Nam-gu, Ulsan 680-749, South Korea

[*]jia@mail.ulsan.ac.kr

Abstract—**This paper proposes a reliable fault diagnosis approach of induction motors using two dimensional (2D) signal processing techniques with acoustic emission (AE) signal. The proposed approach first converts time domain AE signals into 2D gray-level images to exploit texture information from the converted images. Then, a 2D Gabor filter with various scales and orientation angles is applied to each gray-level image to reduce abnormalities and enhance texture information, and the singular valued decomposition (SVD) algorithm is utilized to extract a feature vector, where the diagonal singular values of SVD are used as the feature vector of each fault. Finally, the texture feature vector is used as an input of multi-class support vector machines (MCSVMs) to classify each fault, where the Gaussian radial basis function kernel is utilized with SVMs to deal with non-linear fault features of the AE signal. To evaluate the performance of the proposed fault diagnosis approach, this paper uses multiple bearing defects under different noisy conditions. The experimental results indicate that the proposed approach achieves an average classification accuracy of 99.9% and outperforms conventional fault diagnosis algorithms in classification accuracy.**

Index Terms—**Texture Features, Fault Diagnosis, Gabor Filter, Singular Valued Decomposition, Induction Motors.**

I. Introduction

Texture feature extraction has been playing a significant role in fault diagnosis. Therefore, a number of researchers have worked on extracting different texture features for fault diagnosis [1]-[8]. Do *et al.* [1] proposed a texture feature extraction based fault diagnosis model, where a scale invariant feature transform (SIFT) algorithm was used to extract texture information. As the SIFT is invariant to image rotation, translation and scale variation and partially invariant to illumination changes, coefficient diagnosis of motor faults was expected even in the presence of background noise. However, the major drawbacks of the SIFT include: (i) uncertainty in the number of keypoints for different images, and (ii) high computational cost for the processing of 128-dimensional feature descriptors. Shahriar *et al.* [4] proposed local binary patterns (LBP) based texture fault diagnosis model, where LBP provides a binary code for each pixel of an image that is calculated by thresolding the local neighbours at the gray level of the pixel of interest. However, due to the non-adaptive selection of the number of neighbour pixels and radius of the LBP operator, it is not robust to extract features from the images. Dinh *et al.* [5] proposed a 2D fault diagnosis model, where the authors used dominant pixels using the Gaussian distribution function to extract the texture features. Although this algorithm is very simple and has low computational complexity, the non-adaptive selection of derivatives order and threshold value in the Gaussian radial function effects on the fault classification degradation. Jia *et al.* [7] proposed a fault classification model using a global neighbourhood structure (GNS) map, where GNS maps were calculated by averaging the local neighbourhood structure (LNS) map of central pixels. The non-adaptive selection gap between central pixels, size of searching window, and neighbour window are the major drawbacks in this model, which may decrease the classification accuracy.

To overcome of these limitations inherent in the conventional algorithms, this paper proposes a reliable texture feature based fault diagnosis model, where a 2D Gabor filter is utilized to remove abnormalities in the converted gray-level images and SVD is used to extract the significant information of the Gabor filtered image. Finally, an MCSVM is used to classify each AE fault of an induction motor.

The rest of this paper is organized as follows. Section II describes more details about the proposed fault diagnosis approach, and Section III presents experimental results and analysis. Finally, Section IV concludes this paper.

II. Proposed Fault Diagnosis Approach

The proposed fault diagnosis approach for multiple faults of bearing defects is composed of an optimal band selection, a gray-level image conversion, texture feature extraction, and classification. Fig. 1 shows a block diagram of the proposed approach. In this study, we use three different AE fault signals: inner, outer and roller. These AE signals were directly collected from the three different axial AE sensors attached to the surface of an induction motor. In addition, a normal bearing is used as a baseline in this study.

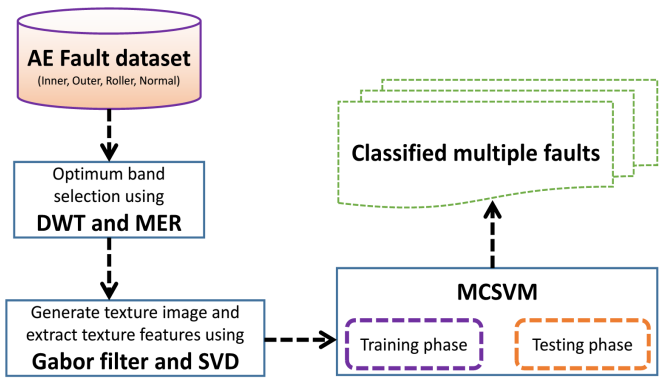


Fig. 1 Propose fault diagnosis approach based on 2D texture feature extraction.

Corresponding author: Jong-Myon Kim (jmkim07@ulsan.ac.kr).
978-1-4799-4166-7/14/$31.00 ©2014 IEEE

In general, most of the AE information lies in the high frequency components, where maximum energy is located. Therefore, we apply the discrete wavelet transform (DWT) and maximum energy ratio (MER) to select the significant frequency bands. There are four different combinations of coefficients in the DWT: approximation-approximation (A-A), approximation-detail (A-D), detail-approximation (D-A), and detail-detail (D-D) coefficients [3]. In this study, we use A-D combination to select the significant frequency components because most of the AE fault information is located at the high frequency and it has the symmetric property of energy distribution. In addition, we set to 8 for the maximum level of DWT because most of the energy is located at up to 8th level. Finally, a time domain AE signal is reconstructed by selecting the optimal energy band using the MER algorithm which is expressed as:

$$Ratio_E = \frac{E_i}{\sum_{i=1}^{l} E_i} * 100 \tag{1}$$

where E_i is the energy of each level of wavelet decomposition and l is the maximum number of decomposition level.

In contrast to the conversional 1D fault diagnosis algorithms [9]-[18], the proposed 2D fault diagnosis approach can reduce noise effects and allow to use small size of input signal by converting time domain 1D AE signals to 2D gray-level images, where texture information of each fault is visible. Fig. 2 presents an example of the conversion process, where amplitude of each sample is normalized as an intensity value of pixel.

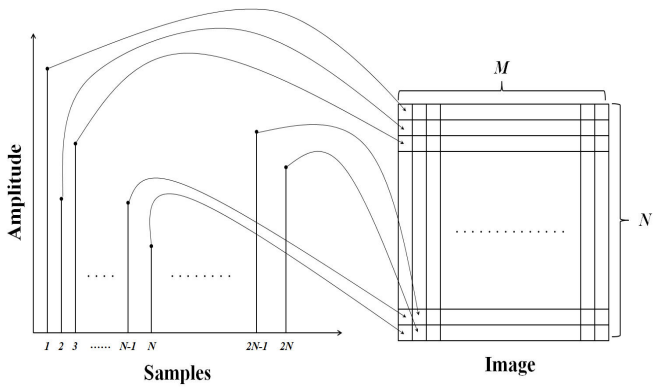


Fig. 2 An example of 1D time domain to 2D gray-level image.

Then, we apply a 2D Gabor filter with various scales and orientation angles to each gray-level image to extract the global information for each fault image. The Gabor filter has the following mathematical properties: (i) it has a smooth and differentiable shape, (ii) it does not have side lobes neither in space nor frequency domain, and (iii) it generates high-dimensional representation. Fig. 3 shows resultant texture images of different faults with and without using the Gabor filter, where the texture information is more visible by using the Gabor filter.

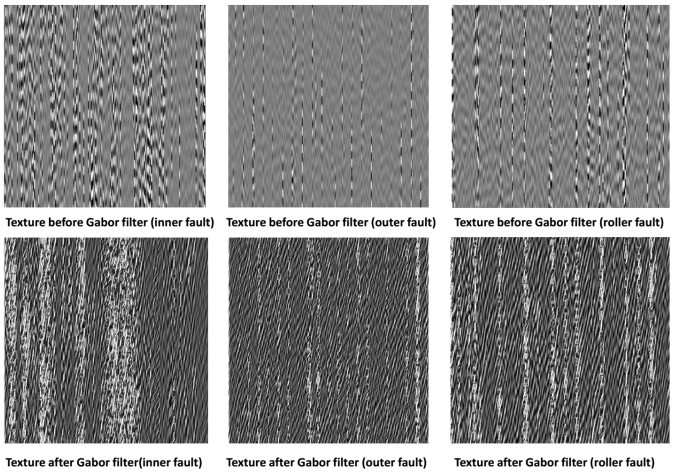


Fig. 3 Resultant texture images with and without using the Gabor filter.

We then utilize singular valued decomposition (SVD) [9] to replace a large dimensional data points with a small dimensional space and expose substructure of the original data more clearly ordering it from most variation to the least. A rectangular matrix of size mxn can be decomposed with the SVD using the following equation [9]:

$$A_{mn} = U_{mm} S_{mm} V_{nn}^T \tag{2}$$

where $U^T U = I$, $V^T V = I$; the columns of U are orthonormal eigenvectors of AA^T, the columns of V are orthogonal eigenvectors of $A^T A$, and S is a diagonal matrix containing the square roots of eigenvalues, $\sigma_1, \sigma_2, \sigma_3, ...\sigma_m$ from U or V in descending order. The matrix S is represented as follows:

$$S = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 \\ 0 & 0 & . & 0 & 0 \\ 0 & 0 & 0 & . & 0 \\ 0 & 0 & 0 & 0 & \sigma_m \end{bmatrix}.$$

Since singular values of SVD contains meaningful information of the 2D space, we decompose SVD with each Gabor filtered image and collect a feature vector considering the singular values only. In this study, the size of each feature vector is 512, where the converted image size is 512x512.

III. Experimental Results

The use of a real dataset is important in the performance evaluation of the fault diagnosis approach. In the experiment, we set a test rig, as depicted in Fig. 4. We made physical cracks in different locations of the induction motor bearing and recorded AE signals using AE sensors. To make cracks on the surface of the bearing, a diamond cutter bit and air speed grinding tool were used. In our experimental dataset, each fault has 90 different signals. The length of each signal is 0.5 second and sampling frequency of each signal is 1MHz. Each SVM was trained with 50% of total dataset and the rest of the dataset were utilized for testing. Table I presents detailed statistical information of the AE fault signals.

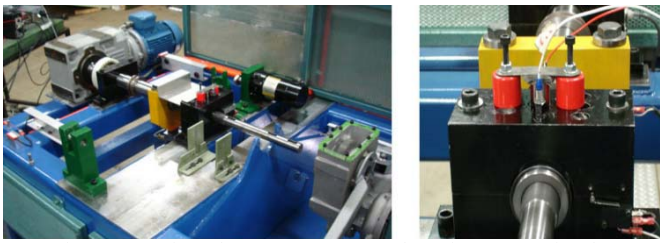


Fig. 4 A test rig and AE sensors to record continuous AE signals.

TABLE I
STATISTICAL INFORMATION OF AE FAULT SIGNALS USED IN THIS STUDY

Parameters	Values
Operating frequency	15
Slim ratio	1
Rotation per minute	900
Maximum display frequency	900
Outer defect frequency	78.6285
Inner defect frequency	116.3715
Roller defect frequency	74.5980

In this study, we evaluate the performance of the proposed fault diagnosis approach in terms of classification accuracy, where classification accuracy is measured with respect to true positive (TP). TP is the number of faults in class i that are correctly classified into the class i. In addition, we utilize a multi-class support vector machine (MCSVM) as a classifier, which consists of four distinct SVMs, as presented in Fig. 5.

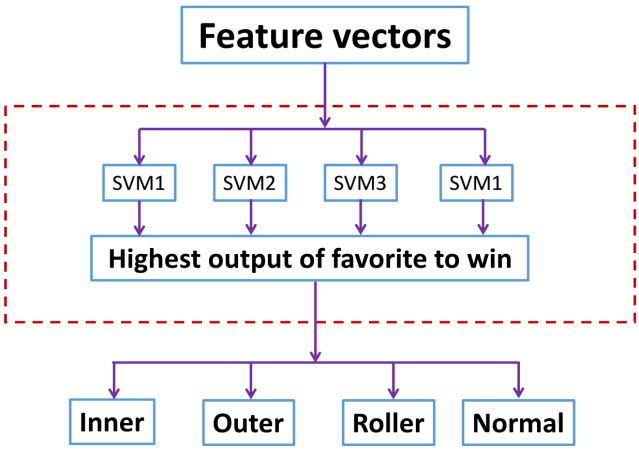


Fig. 5 One-against-all MCSVM for classifying multi-faults into the appropriate fault category.

SVM can work as a non-linear classifier with numerous kernel functions, such as polynomial, Gaussian radial basis function, fisher, graph and string kernels. In this study, we uses the Gaussian radial basis function kernel in each SVM due to its simplicity and dynamic non-linear classification capacity since numerous statistical information of AE signals overlaps with each other and we need to use a non-linear classifier to separate the numerous faults based on the statistical parameters, as shown in Fig. 6.

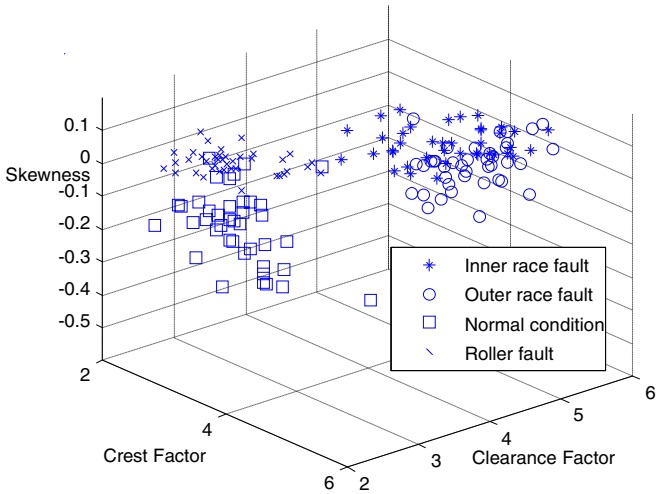


Fig. 6 Nonlinear properties of different AE fault signals.

Table II shows the results of the proposed fault diagnosis approach. The proposed approach achieves 100% classification accuracy with the experimental AE fault dataset except for the inner case in which 99.3% classification accuracy is achieved with the proposed approach.

TABLE II
CLASSIFICATION ACCURACY OF PROPOSED MODEL WITH DIFFERENT SNRS

SNR	Inner	Outer	Roller	Normal
No noise	100	100	100	100
20dB	99.3333	100	100	100
15dB	99.3333	100	100	100
10dB	99.2444	100	100	100

In addition, we compare the performance of the proposed model and other conventional fault diagnosis algorithms with respect to classification accuracy. Table III presents classification accuracy of the proposed model and two other conventional models.

TABLE III
CLASSIFICATION ACCURACY OF THE PROPOSED APPROACH AND OTHER STATE-OF-ARTS ALGORITHMS

Models	Inner	Outer	Normal	Roller	Average
Nguyen et al. [5]	84.72	93.06	100	100	94.445
Kang et al. [6]	99.41	100	100	93.89	98.325
Proposed approach	99.56	100	100	100	99.89

Nguyen *et al.* [5] proposed a fault diagnosis algorithm using a texture feature extraction technique, where texture feature information was extracted by selecting dominant pixels only. The authors used *n*th-order partial derivatives of the Gaussian function and the number of significant pixels depends on the order of partial derivative and threshold value. In that model, optimal selection of the parameters is significantly important to control the classification accuracy. Kang et al. [6] proposed a fault diagnosis algorithm using short-term energy (STE) with SVD to classify the fault signals of an induction motor. The authors divided each time domain signal into a number of subparts and measured STE of each subpart STEs. Then, feature vectors were extracted from the decomposed image with the SVD. Although the method exhibited improved performance compared to other conventional methods, it still misclassified for a few fault signals that the energy distributions of these fault signals are

quite similar. Overall, the proposed approach outperforms other fault diagnosis algorithms for all the cases in classification accuracy.

IV. CONCLUSION

This paper presented a reliable fault diagnosis approach based on the 2D signal processing techniques. In this study, time-domain AE signals were first converted into 2D gray images to exploit texture features from the converted images of each faulty AE signal. Feature vectors were then extracted by decomposing SVD with a 2D Gabor filer. As singular values of SVD contain significant information, we calculated feature vectors of numerous faults by considering the diagonal single values of SVD. Finally, using the distinctive features as inputs, an MCSVM identified each fault of the induction motor. The experimental results indicated that the proposed approach exhibited average 99.89% classification accuracy for our AE dataset and outperforms other conventional algorithms.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2012R1A1A2043644), and this research was supported by the Leading Industry Development for Economic Region (LeadER) grant funded the MOTIE (Ministry of Trade, Industry and Energy), Korea in 2014. (No. R0001220).

REFERENCES

[1] V. T. Do, and U. Chong, "Signal Model-Based Fault Detection and Diagnosis for Induction Motors Using Features of Vibration Signal in Two Dimension Domain," *Journal of Mechanical Engineering,* vol. 57, no. 9, pp. 655–666, 2011.

[2] W. C. Jang, Y. H. Park, M. Kang, and J. M. Kim, "Mechanical Fault Classification of an Induction Motor using Texture Analysis," *Journal of the Korea society of computer and information*, vol. 18, no. 12, pp. 11–19, 2013.

[3] R. Islam, J. Uddin, and J. M. Kim, "Reliable Fault Diagnosis of Induction Motors using an Acoustic Emission Sensor and Signal Processing Techniques," in *Proc. 2nd FTRA International Conference on Ubiquitous Computing Application and Wireless Sensor Network (UCAWSN-14)*, 2014, pp. 1–7.

[4] R. Shahriar, T. Ahsan, and U. Chong, "Fault diagnosis of induction motors utilizing local binary pattern-based texture analysis," *EURASIP Journal on Image and Video Processing*, vol. 2013, no. 29, pp. 1–11, 2013.

[5] D. N. Van, M. Kang, C. H. Kim, and J. M. Kim, "Highly Reliable State Monitoring System for Induction Motors Using Dominant Features in a 2-Dimension Vibration Signal," *New review of Hypermedia and Multimedia*, vol. 19, no. 3, pp. 248–258, 2013.

[6] M. Kang, and J. M. Kim, "Reliable fault Diagnosis of Induction Motor Defects Using a Two-Dimensional Representation of Shannon Wavelets," *IEEE Transactions on Magnetics*, pp. 1–13, 2014

[7] J. Uddin, M. Kang, D. V. Nguyen, and J. M. Kim, "Reliable Fault Classification of Induction Motors Using Texture Feature Extraction and a Multiclass Support Vector Machine," *Mathematical Problems in Engineering*, vol. 2014, pp. 1–9, 2014.

[8] W. C. Jang, M. Kang, and J. M. Kim, "Fault Classification of an Induction Motor Using Texture Features of Vibration Signals," *LNEE, Springer,* pp. 177–183, 2014.

[9] J. Wang, and A. K. Asundi, "A computer vision system for wineglass defect inspection via Gabor-filter-based texture features," *International Journal of Information Science*, pp. 157–171, 2000.

[10] F. Liu, "Texture identification using Gabor wavelet based signal processing," *Institute of Technology and Innovation*, *University of Southern Denmark*, 2011, pp. 1–5.

[11] S. Lahmiri, and M. Boukadoum, "Hybrid Discrete Wavelet Transform and Gabor Filter Banks Processing for Features Extraction from Biomedical Images," *Journal of medical Engineering*, vol. 2013, pp. 1–13, 2013.

[12] M. Kang, and J. Kim, "Singular Value Decomposition Based Feature Extraction Approaches for Classifying Faults of Induction Motors," *Mechanical Systems and Signal Processing*, vol. 41, no. 1, pp. 348–356, 2013.

[13] H. Li, H. Zheng, and L. Tang, "Gear Fault Detection Based on Teager-Huang Transform," *International Journal of Rotating Machinery*, vol. 2010, pp. 1–9, 2010.

[14] Z. Xu, J. Xuan, T. Shi, B. Wu, and Y. Hu, "Application of a modified fuzzy ARTMAP with feature-weight learning for the fault diagnosis of bearing," *Expert Systems with Applications*, vol. 36, pp. 9961–9968, 2009.

[15] H. Keskes, A. Braham, and Z. Lachiri, "Broken rotor bar diagnosis in induction machines through stationary wavelet packet transform and multiclass wavelet SVM," *Electric Power Systems Research*, vol. 97, pp. 151–157, 2013.

[16] G. F. Bin, J. J. Gao, X. J. Li, and B. S. Dhillon, "Early fault diagnosis of rotating machinery based on wavelet packets- Empirical mode decomposition feature extraction and neural network," *Mechanical Systems and Signal Processing*, vol. 27, pp. 696–711, 2012.

[17] A. Bouzida, O. Touhami, R. Ibtiouen, A. Belouchrani, M. Fadel, and A. Rezzoung, "Fault Diagnosis in Industrial Induction Machines Through Discrete Wavelet Transform," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 9, pp. 4385–4395, 2011.

[18] S. Lahmiri, and M. Boukadoum, "Hybrid Discrete Wavelet Transform and Gabor Filter Banks Processing for Features Extraction from Biomedical Images," *Journal of Medical Engineering*, vol. 2013, pp. 1–13, 2013.

Laboratory-Based Project-Oriented Introductory Course for Electrical Engineering

Wayne Blanding[1,*] and Kala Meah[1,*]

[1]Electrical and Computer Engineering, York College of Pennsylvania
York, PA, USA
[*]wblandin@ycp.edu; [*]kmeah@ycp.edu

Abstract—**Based on evidence from assessment data and in discussions with engineering employers, the Electrical and Computer Engineering program at York College of Pennsylvania saw the need to revise its freshman engineering experience. This paper describes the need for this new course and discusses the architecture and pedagogy of the course. Of particular interest were the goals of ensuring all students developed proficiency in operating basic electric/electronic test equipment, skills in basic DC circuit analysis techniques, exposure to a broad range of electric/electronic devices, and introduction to fundamentals of engineering design. Consequently the course was primarily laboratory based and project oriented.**

Index Terms—**engineering education, freshman engineering experience, electrical engineering, engineering design.**

I. Introduction

York College of Pennsylvania is a small, private college in south-central Pennsylvania, USA. It is primarily an undergraduate school with a wide variety of majors. Its engineering programs include Mechanical, Electrical, and Computer Engineering all of which are ABET [1] accredited.

Prior to the course revision described in this paper, all engineering programs had identical course requirements in the freshman year, which included a two-course sequence introducing students to engineering as a discipline and exposing students to all three engineering disciplines offered at the college. These courses were primarily project-based and hands-on.

As part of this curriculum revision, the essential elements of the original two-course sequence were repackaged into a single course offered in the fall of the freshman year and incorporated the essential learning objectives of these courses (computer-aided drawing, introduction to each engineering discipline, introduction to the machine shop). The Mechanical Engineering students then split from the Electrical and Computer Engineering students in the second semester of the freshman year to their own introductory courses, Introduction to Mechanical Engineering and Introduction to Electrical Engineering.

This paper describes the driving factors that led the Electrical and Computer Engineering (ECE) majors to make this curriculum change, the organization and structure of the new Introduction to Electrical Engineering course as a laboratory-based and project oriented class, and provides initial information regarding the instructor assessment and student feedback from the first offering of the class to students.

II. Course Requirements and Structure

This section describes the needs which motivated the development of this new course, the course outcomes that were defined as the curriculum was developed, and the vision for the organization of the course.

A. The Need for a New Course

The Electrical and Computer Engineering program had maintained the original two-semester common curriculum with the Mechanical Engineering program since its inception in 2006. By 2011 sufficient assessment data was obtained to reveal areas where the curriculum could be strengthened. Specifically:

1) *Test Equipment Proficiency*: The introductory electrical engineering course which teaches DC and AC circuit analysis techniques occurs in the spring sophomore year in the curriculum. Course instructors in their assessment reports noted that lack of student proficiency with basic test equipment (multimeter, power supply, function generator, oscilloscope, breadboarding) hindered the laboratory component of the course and resulted in less productive activity being accomplished during laboratory periods. If students had this proficiency entering into this class, student learning of the essential concepts of DC and AC circuit analysis could be enhanced.

2) *Engineering Design Experience*: As originally structured, the Electrical and Computer Engineering programs contained an introduction to engineering design in the freshman year but had limited opportunity for any significant engineering design projects until the major Capstone design experience in the senior year. The Capstone design experience is the culminating design experience for students in which they spend a year working as part of an interdisciplinary project team of 12-20 working on an engineering design and construction project. Course instructors noted in their assessment reports that students entered into the first Capstone design course underprepared and needed several weeks to gain an understanding and appreciation of the engineering design process. Loss of this time at the start of the course limited the team's progress in that first semester, which pushed activities into the second semester leading to a crisis management mentality by the end of the second semester. If students had opportunity to engage in a shorter design project early in the junior year to learn about the design process, then the Capstone design course could more effectively achieve its instructional goals.

3) *Engineering Can Be Fun*: A key factor in retaining engineering students after the first year is a sense of connection with the discipline, a sense of enthusiasm about a career in engineering, and a sense that engineering can be fun. Fundamentally students need to have some external motivation to be able to push through the challenges and difficulties they will face in pursuing an engineering degree. This requirement was developed in an effort to improve first-year retention within the major, defined as the ratio of the

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

number of students who returned for the start of their sophomore year compared to the number who initially enrolled as freshman. For the four years prior to this new course coming on-line, the first-year retention averaged 65%.

B. Course Outcomes

In order to address the issues discussed in section *A* above, the new Introduction to Electrical Engineering course needed to have a heavy hands-on focus, where each student in the class got nearly daily hands-on practice in building and testing circuits using test equipment. Consequently the new course was structured as a two-credit hour laboratory course meeting for six hours each week.

The new course also brought forward some of the basic DC circuit analysis techniques previously taught in the sophomore-level circuits class. The effect of bringing forward this material opened up room in a junior-level circuits class to engage in a 4-week design project to introduce students to engineering design in a more formal and rigorous manner than that introduced in the freshman year.

The new course, being about 70% hands-on with laboratory activities and projects, was also intended to positively motivate students to remain in the major and to give students the sense that engineering work can be fun.

With these overarching goals laid out for the new course, a set of course outcomes were established. These course outcomes support the student outcomes established for the Electrical and Computer Engineering majors which are the (a) through (k) student outcomes from Criterion 3 of the Criteria for Accrediting Engineering Programs, published by the Engineering Accreditation Commission of ABET.

Following completion of the Introduction to Electrical Engineering course, students should (letters in parenthesis indicate student outcomes (a) – (k)):

- Be proficient in the use of electronic test equipment (multimeter, power supply, function generator, oscilloscope) (k)
- Conduct, analyse, and interpret results from an experiment (b)
- Be proficient at drawing, reading, and interpreting a circuit schematic (g)
- Be familiar with the content and use of component data sheets and is able to extract needed information from data sheets (g)
- Be proficient in applying circuit analysis techniques to solve circuit models containing voltage sources, current sources, and resistors (a)
- Be familiar with the functions, use, and typical applications of other circuit elements (a)
- Be familiar with the engineering design process and is able to perform an engineering analysis and design of a simple engineering design problem (c)

C. Course Organization

The course was organized to meet three times a week, for two hours at each class session in one of the electrical engineering laboratories. During each class session a variety of methods were used to deliver educational content to the students and involved some combination of instructor delivery of information to the students, student individual or group problem-solving work, individual hands-on laboratory work, and group (two-person teams) project-based work. Typically from two to four activities were planned for each class session.

For the initial offering of the course in Spring 2014, out of the 41 class days scheduled, the class engaged in laboratory work for 18 of the first 22 class sessions. During this time students also developed skills in DC circuit analysis techniques. The final 15 class sessions (5 weeks) of the course were broken up into a two-week project illustrating the engineering design process and a second three-week project where students in teams of two designed and built a device to customer specifications.

Evaluation of student knowledge consisted of a combination of in-class and out-of-class worksheets, laboratory reports, four hour-long exams, the two projects, and a hands-on practical exercise as the “final exam” for the course. The hands-on practical exercise consisted of a timed laboratory experiment requiring students to breadboard a simple circuit and demonstrate proficiency in the use of each piece of test equipment used by students in the course.

III. Laboratory Exercises

The laboratory work was a key focus of the course in terms of ensuring that students developed proficiency in use of test equipment. For this reason, the course was designed such that students worked individually in the laboratory exercises.

Laboratory exercises were written for a variety of intents and purposes including:

- Learning the use and operation of a particular piece of test equipment
- Exploratory labs investigating a theoretical topic before the theory is presented (e.g., series and parallel equivalent resistance rules)
- Confirmation labs designed to illustrate a theoretical topic after the theory is presented (e.g., Kirchoff’s Voltage and Current laws)

This mixture of laboratory types was intended to provide variation in how topics were developed in the class and to retain student interest/motivation.

The time at which a lab activity was performed was also varied—sometimes the lab exercise occurred at the beginning of class, at other times the laboratory exercise occurred mid-class or near the end of class.

As this course is the student’s first exposure to engineering laboratory activities, the laboratory handouts were highly structured. The philosophy of the Electrical and Computer Engineering program is to develop student laboratory writing skills over a number of courses, with each course providing a scaffold for subsequent courses in terms of expectations.

For this course, the laboratory handouts included a set of analysis questions leading students through the thought process of analysing the data that they have recorded and noting potential sources of experimental or measurement errors. The intent of these structured questions was to expose the student to the thought process needed to independently and without prompting analyse the results of an experiment in a subsequent course.

The final question for each laboratory exercise included a reflection prompt forcing students to engage in critically thinking about what they learned from the experiment and how the experiment relates to real-life phenomena.

IV. Projects

This course includes two projects to facilitate students’ experience on brainstorming, component selection, subsystem

development, system integration, packaging, and documentation while solving real world engineering problems. Project I and II durations were two weeks and three weeks, respectively.

A. *Project I—Assistive Technology for Disabled Individual*

This project idea was adopted from IEEE real world engineering project (IRWEP) library [2] and modified to provide a real world experience in the classroom. Students were asked to fabricate a device for a person with severe apraxia. This person is unable to interact independently with outside world. Occupational therapist examination reveals that only reliable movement of the patient is the hip which lifts the left knee. The objective of this device is to provide the patient an ability to turn on/off a portable fan. The original IRWEP design was on breadboard and uses a relay with 3 V signal power supply. This project packaged the product as a standalone battery operated device with an on board voltage regulator to satisfy the need of various voltage requirements. The relay was replaced with a semiconductor switch. This was a group project and each group consists of two students. A functional block diagram of this project is shown in Fig. 1. ADXL335 accelerometer was used to sense the movement of the knee. Expected inputs and outputs of each block were discussed in the class and students were provided with circuit diagrams. The objective was to teach students how electronic component processes input information to provide the required output. System integration and user interface were discussed to emphasize user friendliness of the device. Students' responsibility was to build and test each subsystem on a breadboard separately to ensure the expected input and output relationships. Once all subsystems were tested appropriately, students were asked to integrate all subsystems and to test for overall system performance.

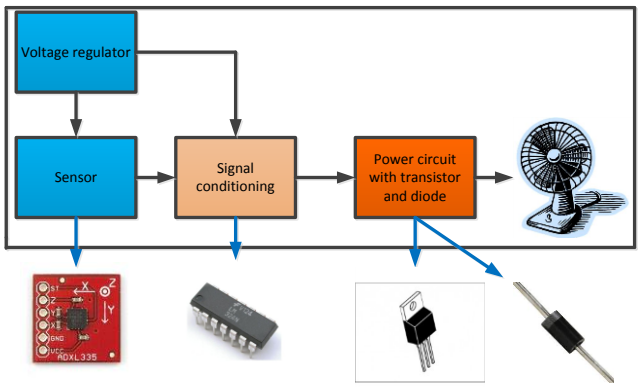


Fig. 1 Functional block diagram of project I.

After the successful testing on the breadboard, packaging was the final step for project I. Students were provided with packaging materials. The machine shop was also available for use. The device was required to have an on/off switch, an adjustment knob, and connecting terminals for the fan and power supply. A finished product of project I is shown in Fig. 2. Ten of the 15 devices produced by students worked perfectly. The remaining devices worked on the breadboard, but failed to work correctly when packaged.

B. *Project II—Let There Be Light*

Project II was an open ended problem with specific customer requirements. Students were asked to develop an inexpensive, reliable, clean, safe, and efficient source of light for people in under-developed countries who currently do not possess it. This project was adopted from IEEE real world engineering project (IRWEP) library [3] and modified to provide a project design experience with real world constraints to freshman engineering students. The original IRWEP design narrowed the component selection by asking students to use LED, photoresistor, and battery. This design project asked students to select components based on the problem statement, constraints, and their analysis. This project packaged the product with a simple on/off switch to make it a turn-key device.

1) *Problem Statement*: A remote village of 500 people has limited access to electrical power and, as a result, the village homes are lit with candles and kerosene lamps after dark. The villagers share a common solar collector in the village square. This centralized power source can be used to charge portable electronic devices during daylight hours. The objective of this project is to develop a prototype portable lighting system that is cleaner and safer than kerosene lamps. The device will be charged during the day in the village square and is to provide enough light for a villager to read for a minimum of two hours at night. Each team represents a company who will bid on this project. As part of the bidding process, each team must design and build a prototype according to the project specifications below. Each team is in a competition with other teams, and so should consider design ideas to be proprietary. The device should be: low cost, reliable, energy efficient, easy to recharge, easy to operate, safe to use, environment friendly, and durable. Each team had a budget of $50.00 and in house resources such as a stockroom containing standard parts, the machine shop, and a 3D printer.

Fig. 2 A finished product of project I.

2) *Project Planning and Execution*: This was a three-week long project and students were provided with a day-by-day schedule starting with brainstorming, design concept, and finishing with final project presentation, report, and device. Fig. 3 shows some of the devices that students developed as a part of project II. There were 14 devices and all of them were functional and met the requirement of two hours reading time.

V. Outcome Assessments

Proficiency in electronic test equipment use, conducting experiment and gathering data, and engineering design were the driving factors leading to development of this course. Throughout the semester instructors observed students and helped them to achieve the above skills. The end of semester practicum test was used to gauge students' ability to use

equipment/component to build a circuit on breadboard and to collect appropriate data. More than 80% students satisfied this requirement.

Fig. 3 Efficient light sources developed by students.

ABET requires continuous assessment and improvement of engineering courses for a successful accreditation. Electrical and Computer Engineering program at York College of Pennsylvania developed a strategy to complete an outcome assessment rotation in every three years. As a part of this process student outcome (c)—an ability to design a system, component, or process to meet desired needs within realistic constraints such as economic, environmental, social, political, ethical, health and safety, manufacturability, and sustainability use assessed on project II using performance indicator c.1—Designs a system, component, or process to meet desired needs within quantitative constraints. Table I shows the assessment rubric with student achievement.

TABLE I: Student Outcome Assessment

Assessment Rubric			
Attribute	**Exceeds Expectations**	**Meets Expectations**	**Below Expectations**
Project design quantitative specifications: Design a lighting source to operate for at least 2 hours and provides sufficient light for a person to read at night	• Final design met the 2-hour lifetime specification • Final design met the illumination specification • Engineering analysis was correct and supported actual performance	• Final design met the 2-hour lifetime specification • Final design met the illumination specification • Engineering analysis had flaws resulting in an incorrect expected source lifetime	• Final design did not meet one or both specificati ons.
Student Achievement			
Number (%) of teams:	6/14 (43%)	8/14 (57%)	0/14 (0%)

VI. Student Feedback

Student feedback is an important part of the process to develop and improve an engineering course. During this first time offering instructors regularly sought student feedback to adjust and improve the course material to benefit students' learning. Students were surveyed anonymously twice to gauge their interest in course materials and learning abilities. The final survey included three parts: overall course, project I and project II. 23 students out of 28 returned the final survey. Fig. 4 shows survey questions with student's feedbacks for the overall course. Survey results show that majority of students enjoyed and learned the materials while having fun doing so.

Students were also asked to list 3 things they like about the class/course and to list 3 things they do not like about the class/course. Here are some highlighted responses from students:

A) 3 Things Students Like About the Class/Course

- Projects were very fun and rewarding,
- Frequent labs and homework helped me remember the materials,
- Real-world applications through projects,
- Many hands on opportunities.

B) 3 Things Students Do Like About the Class/Course

- Homework problems were hard,
- Could not pick partner for the second project,
- Sometime lab felt rushed. Should be a 3-credit course,
- Too early in the morning.

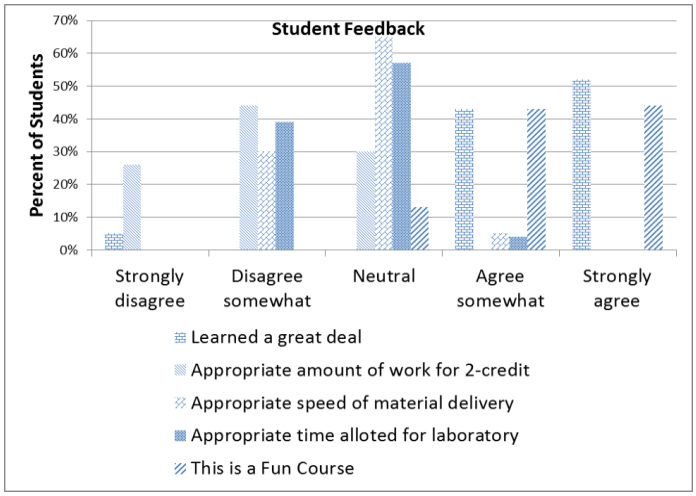


Fig. 4 Students' feedback on the course.

VII. Conclusion

This introductory course for electrical engineering was successfully developed and taught for the first time. The objective of the course was achieved through hands on laboratory work and solving real-world problems with projects. Survey results showed that students were engaged and learned the course materials while having fun. The greatest challenge of this course was to show the freshman electrical and computer engineering students the link between theory and application. Students' outcome assessment and feedback show that this course successfully created that link and kept the students' interest high in engineering education.

References

[1] (2014) The ABET website [Online]. Available: http://www.abet.org

[2] T. Ward, *An adjustable movement-triggered switch access system for people with disabilities*, IEEE Real World Engineering Problems website [Online]. Available: http://www.realworldengineering.org/

[3] L. Wyard-Scott and J. Smith, *Solid state lighting for the developing world*, IEEE Real World Engineering Problems website [Online]. Available: http://www.realworldengineering.org/

Mobile Interactive Problem Solving for Active Teaching and Learning

M. Muztaba Fuad, Debzani Deb and Shequi'lla Whitaker

Department of Computer Science
Winston-Salem State University
Winston-Salem, NC 27110, USA.
{fuadmo, debd}@wssu.edu

Abstract—**In traditional classroom setting, lecturing is the prominent method of transferring knowledge to students. However this passive mode of instruction does not keep student engaged and interested longer in the class. Although mobile devices are used extensively in everyday life and have a bad reputation of creating distraction during class, it can be used to make student's class participation more interactive and engaging. This paper presents one such approach, which is fundamentally different than existing ones by allowing students to participate in active learning during class session by using their mobile devices. This research envision interactive problem applications which will run in student mobile devices, which faculty can initiate remotely and where student can actively participate to solve problem in a hands-on manner. This paper presents the development of a Karnaugh map interactive mobile application and its incorporation into Mobile Response System.**

Index Terms—**Active learning, Interactive problems, Hands-on activities.**

I. Introduction

In a traditional face-to-face classroom setting, if faculty is only using lecturing to deliver course material, they will loose attention of many students, minutes after the class has begun. In recent years, classes in Science, Technology, Engineering and Mathematics (STEM) courses faced a significant learning barrier, which is the problem of students' inability to keep engaged and interested in classroom. Particularly the proliferation of mobile devices in daily life and its use during the class makes this problem exacerbated. In this research, we explore the applicability of using mobile devices in the classroom and incorporation of interactive problem solving using those devices to increase class engagement and active learning for students. By allowing the students to solve hands-on problems in their preferred devices will create a friendly learning environment where students want to remain, be active and be skill full. This paper presents one such interactive problem (IP) application (app) designed to be deployed using the Mobile Response System (MRS) [1]. This app facilitates Karnaugh map (k-map) simplification for students by allowing them to interact with the problem at hand. Faculty can define the k-map problem using a server side tool and use the MRS backbone to initiate a hands-on session during class time in student's mobile devices. This paper presents ways to develop IP app and ways to incorporate that IP in the MRS software.

II. Background

There has been a plethora of research work performed to address the lack of student engagement and motivation during class time. However, the research presented in this paper is different because along with providing a mobile interactive problem-solving platform it also provides a teaching methodology that goes hand-in-hand with this software tool.

A. Teaching Methodology

To improve student learning in the STEM disciplines, we need evidence based teaching and learning methods [2]-[4], which is based on constant monitoring of outcomes from assessment and resulting continuous improvement of the applied method. Study shows that in order to improve student learning in the STEM disciplines, traditional pedagogical approaches are not enough to transfer critical knowledge to students. A feedback driven, evidence based teaching and learning technique has to be devised and implemented in our classes to transfer knowledge to students. One of the key factors of such evidence based teaching methodology is for students to actively participate in class and receive instant feedback about their performance in such hands on activities. In that regard, evidence based teaching methodology was incorporated and assessed for couple of semesters. The findings [5] from that study and student's use of mobile devices during class time has encouraged us to developed this interactive problem solving teaching methodology. As several other [6]-[7] research works have shown that using technology during class can improve student's learning abilities, we are confident that the presented teaching tool will also have positive impact on student in-class engagement and problem solving skills.

B. Mobile Response System

To facilitate interactive problem solving in real-time manner, a comprehensive backend system is necessary. To satisfy requirements for interactive problem solving, a Mobile Response System is developed, which is independent of any interactive problem or its domain and transparent to the end user. MRS is a client-server software that allows the faculty to dynamically prompt the students with interactive problems synchronized with the lecture material in their mobile computing devices (Clients). Students are then able to actively interact with the problem through multiple steps and send their work back to the faculty computer (Server). MRS then performs grading of the exercises, by using the provided rubric. The other important features supported by MRS are immediate and context-sensitive feedback, anonymous question, polling, summarized grading etc. More information about MRS can be found in [1] and [8].

C. Interactive Problem (IP)

An IP is one, where students have to devise the answer following a set of steps and by following a particular algorithm/process. In each step, students have to make key choices that will have impact on the next step of the interaction. During these interaction steps, students can go back and forth and change their answer. This will allow them to see what is the affect of different selection on the result and

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

how every piece fits together. Problems can be started bottom up or at the middle to give students different perspective on the problem and assess their problem solving skills. Only after the student traverse each of the steps or the allotted time to answer a problem runs out, the result of their interaction is sent back to the server as the answer of that student. Each problem has a rubric that not only grade correct answers but also partial answers to gauge student's problem solving skills and thinking models. Each IP has two different components. One at the server side to define the problem and other at the client-side to allow students to interact with the problem. In the server side, the intended problem and the rubric have to be defined by using the MRS standard protocol. On the client side, an app needs to be developed to allow corresponding interactive-ness. MRS provides corresponding Application Programmer Interface methods to convert user provided definition into the corresponding MRS format. Different parts of the format are as follows:

- *Question type*: It must match the Android package name of the client side interactive app and the package name of the server side program for that IP.
- *Question number*: Any number or word, as long as there is no other definition with the same question number.
- *Time to Answer*: Have to be expressed in seconds (which in turn is converted to milli-seconds for better time keeping) with the lower limit of 0 and with the upper limit of the maximum value allowed in a long variable.
- *Instruction/Description*: Any instruction or description associated with this IP. This is optional.
- *Problem Components*: Any number of components needed to render the corresponding IP app on the client.
- *Rubric Components*: Any number of components needed to process the answer of the student and what background information from the IP app should be monitored. The number of problem component and the number of rubric component can be different and has no upper value.

D. Related Work

Different studies [9]-[10] have found the benefits of using mobile devices in classrooms and in recent years, there has been a plethora of work [11]-[12] performed to incorporate mobile devices in classrooms. However, these works are traditional in nature and lack the interactive-ness needed in todays engineering classrooms. In Science and Engineering courses, students need to actively solve problems by interacting with the problem in a hands-on approach. Students cannot develop skills such as synthesizing a problem and developing critical thinking only by using multiple-choice or true-false questions. It is our belief that by presenting the problems as interactive entities, where students can actively play with the problem to solve it; we will be able to improve student's critical thinking and problem solving skills. Although there has been a lot of work on algorithm animation and visualization, there are hardly any [13] interactive tools for supporting k-map simplification as presented in this paper.

III. Interactive Problem App

To utilize MRS for facilitating interactive problem solving, one needs to follow several sequences of steps and some simple rules as shown in Figure 1 and described in the following sub-sections. Although the K-Map problem is used as an example to describe the process, the MRS software is developed independent of any problem domain. This allows it to be used by any one as long as they follow the described process and develop their own interactive problem and its rubric.

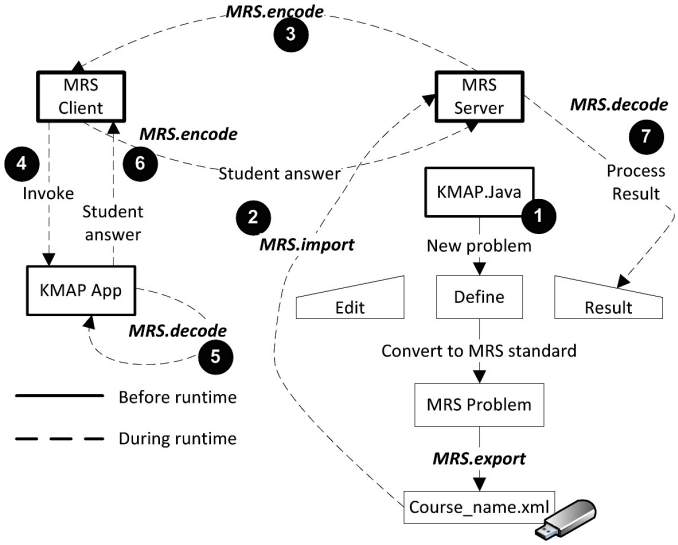
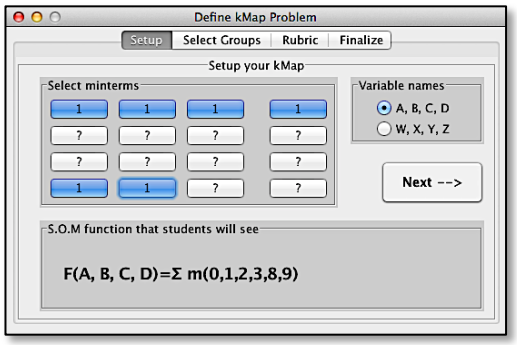


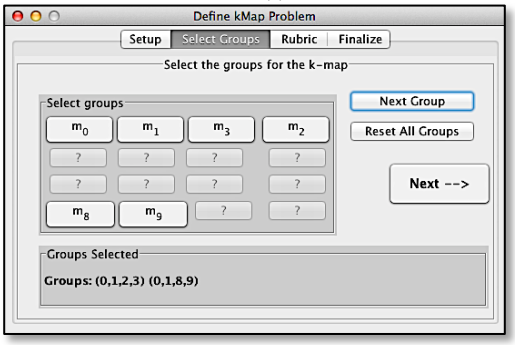
Figure 1. Component organization and workflow.

- *Step 1*: In this step, the IP has to be defined following the MRS protocol as described in Section II.C. If a server side program is written by the user to define the IP instead of defining it by hand, following rules needs to be followed:
  - The program should be in a package with the same name as the package of the IP app and the corresponding question type.
  - The program should have three methods, *define*, *edit* and *result* so that MRS can dynamically call them for corresponding functionality; i.e. *define* to define a new problem, *edit* to edit an existing definition and *result* to process student answers and produce corresponding graphical representation.
- *Step 2*: Once the IP is defined and saved in the storage as a XML representation, it can be imported to MRS by using the MRS GUI, which uses ***MRS.import*** to read the IP definition on the file. One file can have multiple types of IPs in it with unique question numbers.
- *Step 3*: Faculty uses the broadcast button in MRS to broadcast the IP definition to clients and MRS uses the ***MRS.encode*** to broadcast it to all available clients. In encoding and broadcasting, MRS does not include the rubric components of the definition.
- *Step 4*: Once the MRS client receives the encoded definition, it finds the corresponding IP app (by matching the question type with the app package name) and invokes it with the encoded definition as an Android Intent [14] with the key "QUESTION".
- *Step 5*: Once the IP app is invoked, it needs to get the encoded question with the key "QUESTION" and then use the ***MRS.decode*** method to convert it to the workable definition format.
- *Step 6*: Once student submit an answer or the time to answer the problem runs out, the IP app will quit and return the student response/result as Android Intent to MRS with the key "ANSWER". Once MRS client receives that Intent, it forwards that contained message back to the server using ***MRS.encode.***
- *Step 7*: MRS server locates the server side program to process the result and invokes the *result* method.

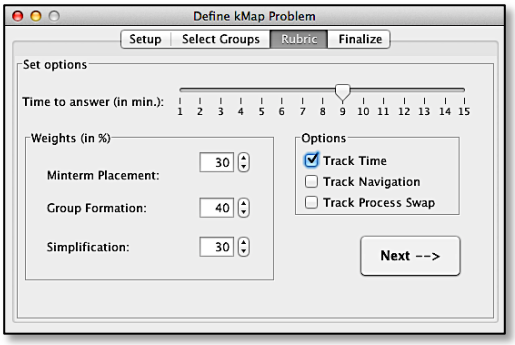
A. Server Side



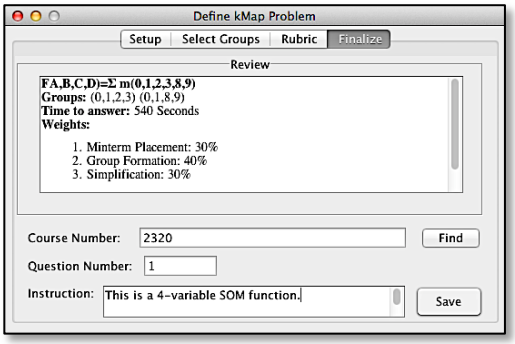
(a)



(b)



(c)



(d)

Figure 2. Defining the k-map IP.

For the k-map IP, at the server side a GUI-based program is designed for easy design of a problem. It follows a sequence of intuitive steps for the user to define a K-map IP. As shown in Figure 2 (a), faculty can choose their min-terms for the expression to be simplified and then choose the valid groups (Figure 2 (b)), set different parameters for problem solving and how it will be graded (Figure 2 (c)) and finally can review (Figure 2 (d)) the problem before saving it to the file. In each step, corresponding validity checks are made to make sure faculty is producing a correct k-map problem. For instance, when selecting groups, faculty will be prompted if the grouping does not follow any allowed rules of the algorithm.

Once the faculty saves the IP definition, it looks something like Figure 3. MRS uses a simple XML schema, which is space saving and faster to decode and manipulate. If faculty do not want to design a GUI, they can use any text editor to define their IP following the definition scheme and the XML schema.

```
<q_type>edu.wssu.kmap</q_type>
<q_num>2</q_num>
<q_time>540000</q_time>
<q_inst>Sample kmap problem</q_inst>
<comp_num>2</comp_num>
    <comp>SOM 0, 1, 2, 3, 8, 9</comp>
    <comp>Variable Names: A,B,C,D</comp>
<comp_num>7</comp_num>
    <comp>Groups |(0,1,2,3) (0,1,8,9)</comp>
    <comp>Placement weight: 30%</comp>
    <comp>Grouping weight: 40%</comp>
    <comp>Simplification weight: 30%</comp>
    <comp>Track student time: true</comp>
    <comp>Track student navigation: false</comp>
    <comp>Track process swap: false</comp>
```

Figure 3. IP definition as saved in the file.

Now, faculty can use the MRS server side GUI to import that problem definition and broadcast it to the available clients as shown in Figure 4. Once all student responses is received back from the clients, either a set of standard statistical tools (average, standard deviation, frequency etc.) within MRS will be used to produce the graphical representation of the student responses or the *result* method of the user definition program will be called to produce corresponding representation.

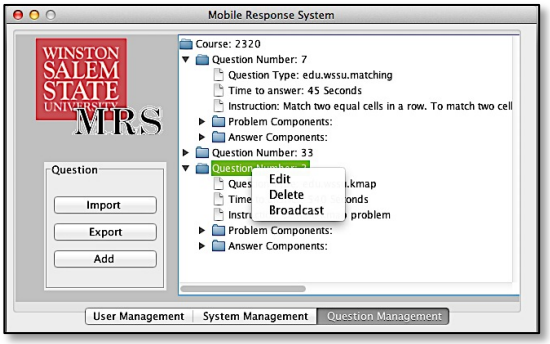


Figure 4. MRS GUI to manipulate IP definition.

B. Client Side

On the client side, students run the MRS client and use their credentials to log into the system. Once they do that, they are presented with the screen as shown in Figure 5. Students can also send anonymous questions or feedback using the client. Once a problem definition is received by the client, it finds the corresponding IP app and pass the definition to it. As shown in Figure 6, student is prompted with the given problem and the app allows the student to traverse back and forth to interact with the problem and solve it within the given time. At first, students can select cells corresponding to the equation (Figure 6(a)). This tests, whether students learned how the map is laid out and what every cell points to. Once this step is finished, students make groups with the cells following the minimization algorithm (Figure 6(b)). Students can select any cell and each group made will take a different colour (with the group number as superscript) to distinguish it from other groups. This will test whether student understands different characteristics of making groups. Next, students interact with

each group and select variables from each group (Figure 6(c)) following the minimization algorithm. This tests whether students fully comprehend the algorithm. The app checks for algorithm validity (such as grouping) with student responses but since the rubric is not sent to the clients, can never tell the student whether his/her answer is correct or not. That way, students cannot figure out the correct answer by iterations. Only after all students submitted their answer, the faculty will be able to show them the correct answer. It is completely up to the developer of such IP app on how much interaction and how many screens they want the student's to encounter during the allotted time.

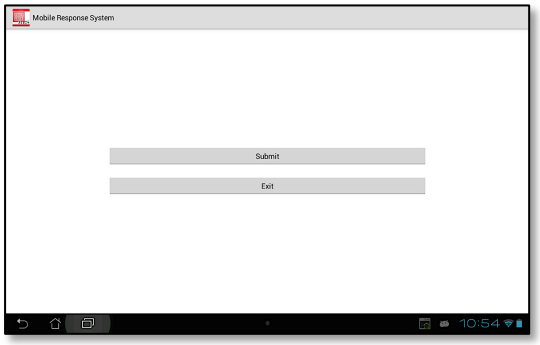
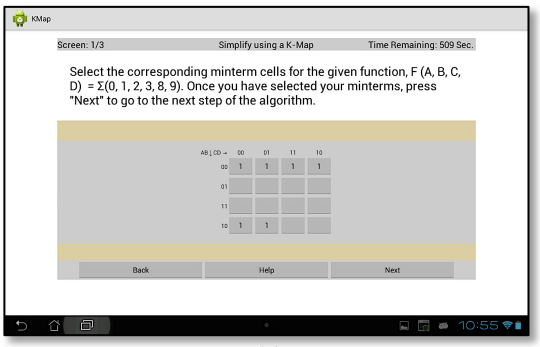
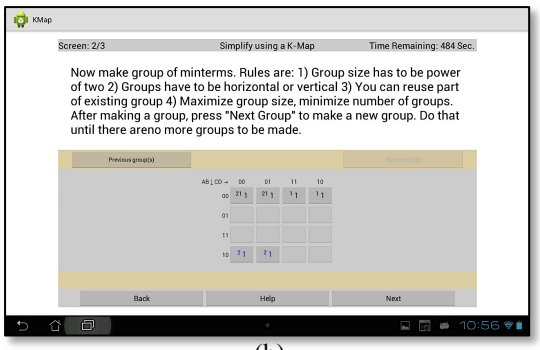


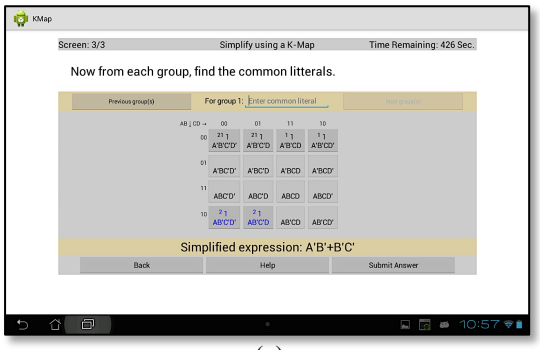
Figure 5. MRS Client.



(a)



(b)



(c)

Figure 6. IP interactions.

C. *Evaluation Plan*

One sophomore level hardware organization course (offered in both semesters) is selected to assess the system in Fall of 2014. This course is a good candidate for evaluation of this approach because it assess mathematics and problem solving skills and we can incorporate several kinds of interactions during the class to enhance these skills. Also, student performance and course evaluation data for last couple of years are available to compare with the new data after deploying the new methodology. Both formative and summative evaluation is planned with the help of data collection, pre and post tests, student surveys, performance data comparisons and focus groups.

IV. Conclusions

Interactive problem solving during class time will allow students to engage and participate more on the class activities and retain more of the lecture content. This paper presents an interactive problem app and its associated components and its incorporation into the Mobile Response System software. Although not presented, software quality matrices for the developed interactive problem app are measured and they fall within acceptable range. Evaluation plan is presented which will be implemented in the future.

Acknowledgment

This research was supported by National Science Foundation grant #1332531.

References

[1] M. M. Fuad and D. Deb, "Design and Development of a Mobile Classroom Response System for Interactive Problem Solving", Accepted for publication at the 26th International Conference on Software Engineering and Knowledge Engineering, Canada, 2014.

[2] D. Perry, "How The Brain Learns Best", *Instructor Magazine*, 2000.

[3] B. K. Saville, "Using Evidence-Based Teaching Methods to Improve Education", *Excellence in Teaching*, vol. IX, 2010, pp. 48-54.

[4] J. Fairweather, "Linking Evidence and Promising Practices in Science, Technology, Engineering, and Mathematics (STEM) Undergraduate Education: A Status Report for The National Academies Research Council Board of Science Education", Washington, DC, 2010.

[5] M. M. Fuad, D. Deb and J. Etim, "An Evidence Based Learning and Teaching Strategy for Computer Science Classrooms and its Extension into a Mobile Classroom Response System", Accepted for publication at the 14th IEEE International Conference on Advanced Learning Technologies, Athens, Greece, 2014.

[6] Mockus, L. and Edel-Malizia, S., "The Impact of Mobile Access on Motivation: Distance Education Student Perceptions", 17th Annual Sloan Consortium International Conference on Online Learning, 2011.

[7] Jones, A., & Issroff, K., "Motivation and mobile devices: exploring the role of appropriation and coping strategies", Research in Learning Technology, *Vol. 5*, No. 3, 2007.

[8] D. Deb and M. M. Fuad, "Use of mobile application to improve active learning and student participation in the computer science classroom", 45th ACM technical symposium on Computer science education, 2014.

[9] Roberts, J., Harvesting fragments of time. Mobile learning pilot project. Technical report, McGraw-Hill, 2003.

[10] Jeffrey R Young, Mobile College App: Turning iPhones Into 'Super-Clickers' for Classroom Feedback, Chronicle of higher education, 2008.

[11] Whattananarong, K., "An Experiment in the Use of Mobile Phones for Testing at King Mongkut's Institute of Technology North Bangkok", Intl. Conference on Making Education Reform Happen: Learning from the Asian Experience & Comparative Perspectives, 2004.

[12] Meisenberger M., Nischelwitzer A., "The Mobile Learning Engine (MLE) – a mobile computer-aided, multimedia based learning application", Multimedia Applications in Education Conference, 2004.

[13] Action Based Learning for Switching and Automata Theory, http://vlsi.colorado.edu/~mooni/N_ABLE/N_ABLE.html.

[14] Android Intent, http://developer.android.com/reference/android/content/Intent.html.

University-Industry Research Collaboration

Nemai Chandra Karmakar
Electrical and Computer Systems Engineering, Monash University
Clayton, VIC 3150, Australia
Email:nemai.karmakar@monash.edu

***Abstract*—University and industry research collaboration hugely impacts an academic's research career and curriculum development. The collaboration brings new prospects of research funds, real world problems and research challenges, and new ingredients in curriculum development. The collaboration also creates innovation and provides national economic benefits. The paper presents the mechanism to university-industry research collaboration.**

***Index Terms*—University-industry relationship (UIR), engagement model, intellectual property, research excellence, research impact, research collaboration, absorptive capacity**

I. Introduction

Universities make significant contributions to the economic development of a nation. A direct benefit comes from their pedagogical operations to educate and train young-adult student cohorts. In publicly-funded tertiary institutions like research universities, a major activity beyond the conventional pedagogy is research. In both developed and developing countries, governments fund research activities and support infrastructure, staff and research projects. This funding amounts to about ten billion dollars per year in the Australian context [1, 2]. OECD countries such USA, the European Union and Canada are bigger contributors of research funding as the population and the mass of economic activities are larger compared to those for Australia. In those countries, vibrant university-industry collaborations and research cultures exist [3]. In Asian developed and developing countries such as China, India, Japan, Korea, Singapore and Thailand, significant strides in research funding are observed [4]. Universities generate, share and disseminate new knowledge, train students for the betterment of national economic benefits and growth, and finally, obtain recognition. The commercial aspects of pedagogical practices such as higher degree research student (HDR) supervision and income generation from the outcomes of HDR research projects are not usually perceived as the main activity of universities. In most universities, the main income comes from students' tuition fees for teaching courses. A very small fraction of total university income comes from funded research and revenue from licensing fees for innovations. In Australian universities, most research funding comes from the federal government's nationally competitive research grant schemes. The success rates of these grant applications are as low as less as 20% [5]. Academics and researchers under the schemes focus on high quality publications to fulfil the benchmarks for excellence of research of Australia (ERA). Thus, the research outcomes of these funded projects generate publications, citations and prestige in the form of university world rankings. However, ERA also acts to deter university-industry interaction and innovation. Industry wants to protect intellectual property (IP) and products instead of producing high quality publications. IP protection works against publication and academics suffer from publication barriers when they are heavily engaged in industry-based research and innovation. The philosophies of university and industry are disparate. While the university focuses on pedagogy, the gathering and sharing of new knowledge and its recognition, industry focuses on return on investment. This gap can be reduced with innovation through collaborative research projects. However, to bring these two entities closer for the greater benefit of the national economy requires long term strategic planning.

This paper addresses these barriers in university-industry collaboration and considers how university and industry can overcome these barriers with mutual trust, by sharing cultural values and the successful execution of collaborative research projects.

II. University-Industry Collaboration

Integration between industry's needs and university research is needed for the smooth translation of research outcomes to the economic benefit of a country [6]. For research negotiation and collaboration between educational institutions and industry, the guidelines are as follows:

(i) There must be industry participation in the development of the research project to ensure its relevance. It must be updated regularly to keep abreast of the needs of the industry, particularly in areas experiencing rapid change.

(ii) The research project should provide HDR students/researchers the opportunity to acquire industrial experience via internships and design projects conducted by professional engineers and faculty members with industrial experience. There should be an industrial attachment unit to facilitate this aspect of the program [7, 8, 9].

(iii) There must be a communication channel between the educational institution and the industry, especially so that the industry can give feedback to the faculty concerned on: a) the quality of the research outcome, b) the absorptive capacity of the research outcome by the industry, and c) the relevance of the research outcomes to the global market place.

Universities suffer from insufficient links with real-world problems and narrowly focused learning objectives. Therefore, industry links are significant in the following ways:

(i) university-industry links create a real-world research atmosphere. HDR students and researchers acquire a sense of belonging to real-world problems. They see applications of their research outcomes in commercial settings, and a result, develop confidence in addressing the challenges which arise.

978-1-4799-4166-7/14/$31.00 ©2014 IEEE

(ii) The link generates IP for all parties concerned. IP generation is imperative for the sustainable development of the national economy and provides extra income for the university. The revenue from university inventions is used in activities such as scholarships for HDR students to conduct further research, upgrading equipment and facilities, and expanding university infrastructure.

(iii) Industry collaboration also brings brand names and prestige to institutions.

(iv) Industry obtains employment-ready graduates from collaborative links with universities.

Fig. 1 University offerings to businesses in collaboration

Therefore, the facilitation of linkages between industry and university with government support is crucial. An effective and sound framework similar to SBIR can enhance this drive significantly [6]. Substantial streams of external funding, enhanced opportunities for professors and graduates to work on groundbreaking research projects, vital inputs to keep teaching and learning on the cutting edge of a discipline, and the impact of delivering solutions for pressing global challenges are the key benefits of university-industry collaboration [3, 10].

Universities are huge places where teaching and research run concurrently. Universities are huge depositories of knowledge, have skilled workforces, and aim for excellence in research and state-of-the-art laboratory facilities. Therefore, universities can offer a wide range of skilled workers, resources and ideas to conduct industry-sponsored research projects. As shown in Fig. 1, universities can offer a range of facilities in university-industry interaction: (i) create talent pipeline; (ii) provide ready access to affordable and rapid resources for SMEs such as skilled workforce, equipment, lab facilities and new and fresh ideas. This leads to improved productivity of SMEs. (iii) Strong brands and international networks, (iv) problem-solving services to businesses such as desk-top reviews and initiation of collaboration, (v) increased funding support for research collaboration through government funding and internal research grants such as faculty strategic research funds, (vi) improved pitching processes to attract venture capitalists and innovators, (vii) excellence of research that creates impacts in productivity in businesses, and finally (viii) recruitment services in particular disciplines for SMEs. Therefore, SMEs greatly benefit from the collaboration. University-industry collaboration brings the following benefits to SMEs: (i) Easy access to best science at cheapest price. (ii) Easy gateway to network. (iii) Development of research culture and innovation. (iv) Joint supervision and mentoring of research students and making them employment-ready for industry. (v) Leverage of research funding and resources via university research.

From the above analysis, it is obvious that university-industry collaboration brings many significant benefits to both entities. However, there are certain barriers that hinder access to university expertise and resources by SMEs. The main barrier is the difference of perceptions of SMEs and universities. The philosophies of these two entities are very different. Industry's main priority is doing business and therefore, they focus on "*return on investment*". SMEs invite universities to collaborate in research for the following two reasons: (i) to facilitate their new product/service development; and (ii) improved productivity for cost-effectiveness and increased profit margins. Time is money for industry. The timeliness of delivery of productivity and services is imperative in the interaction. This rigidity of timeliness requires sound planning of collaboration. Usually university-industry interaction goes through a range of activities before any concrete collaboration can be initiated.

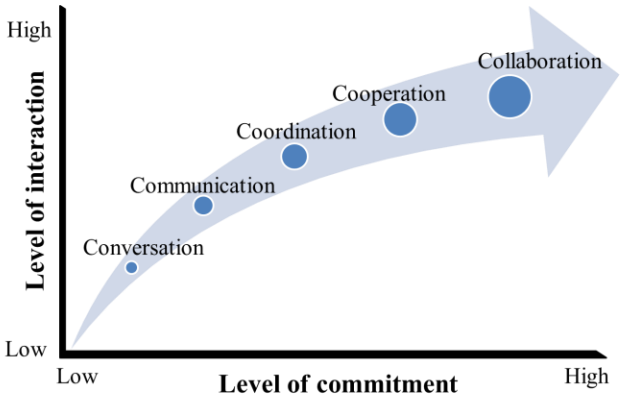


Fig. 2 Interaction and commitment levels for successful university-industry interaction

Fig. 2 shows these interactions and their levels of commitment [11]. As Fig. 2 shows, interaction between university and industry starts with very low level of negotiation in the form of conversations between key personnel. This facilitation may happen through the commercialization arm of the university or individual researchers and the marketing/engineering managers of SMEs. In this initial phase of negotiation, they come to know each other and try to find common ground and interests that may further establish communication between higher-level management and streamline discussion. After successful communication, coordination of specific thematic topics that have potential for investigation is initiated by coordinators on both sides. Usually business development managers (BDMs) and specific team members of SMEs engage in further discussions to set the topic and its specifics. A framework is established in the coordination process. The engagement model is established through this coordination process. Cooperation between the two entities facilitates the development of an engagement model with two vital ingredients of collaboration: financial commitment and research agreements. Once this phase is finalized and

successful outcomes are achieved through secured funding, the actual collaboration happens. The whole process may take a few months to a few years for university and industry to reach high-level commitment to focused and effective collaboration. The ingredients for successful collaboration are: (i) **P**atience, (ii) **P**erseverance; (iii) **P**erformance; (iv) **P**ro-activeness, (v) **P**re-condition; (vi) **P**erception; (vii) **P**articipation; (viii) **P**romotion; and (ix) **P**oliteness. These attributes are necessary for both parties concerned to initiate and execute successful collaboration.

III. REMOVING HURDLES IN COLLABORATION

Although all parties concerned—university, industry and government—understand the significance of university-industry collaboration for sustainable economic growth and improved productivity of the country, many systemic barriers hinder successful collaboration [6]. Two important reports present the impacts of university-industry collaboration in both the developed and developing worlds [3, 12]. The impacts are: financial insolvency; cultural gaps; lack of absorptive capacity of research outcomes by SMES, lack of links/knowledge/network, and finally, different time scale.

SMEs may lack understanding of the university research culture and operation, and may not know the appropriate university research group for collaboration on certain projects. New SMEs perceive universities as sub-contractors of projects, and set stringent guidelines and time-scales to deliver the outcomes of the project. These cultural mismatches in needs, goals and priorities hinder research collaboration.

The most effective approach to bridge the gap is to facilitate relationship-building through many negotiations, demonstrations of research outcomes, and inviting SME personnel to research seminars and university open days on a regular basis. Executives may offer to teach courses and develop interaction. Work placements of university graduates to SMEs and the development of research programs aligned with the needs of industry help bridge the cultural gap [7, 8]. Work placements and industrial research projects without any financial commitment encourage SMEs to bridge the cultural gap and create win-win situations for both parties. Both government and universities can offer HDR scholarships for industry-sponsored projects.

From the above discussion on overcoming cultural barriers to improve university-industry collaboration, it is apparent that both universities and industry need to work closely to understand each other's cultures and strive to remove the barriers. The driving forces for collaboration are: (i) willingness to collaborate; (ii) improved standards in SMEs and development of a sense of belonging to a larger community to build the nation; (iii) good quality assurance and benchmarking in project deliverables; and finally, (iv) preparedness for international competitiveness, which is important for the current open market economy. In the following sections, effective engagement models are described.

IV. GOVERNMENT SUPPORTED ENGAGEMENT MODEL

Fig. 3 shows the various steps of negotiations and collaboration between university and industry. Once negotiations at various levels succeed though observation, reflection, planning and action (action research), which could be motivated by *market push* or *market pull*, an engagement model as shown in Fig. 3 can be implemented [13]. After comprehensive negotiations with industry and understanding their needs and market opportunities, an academic researcher can identify a research domain.

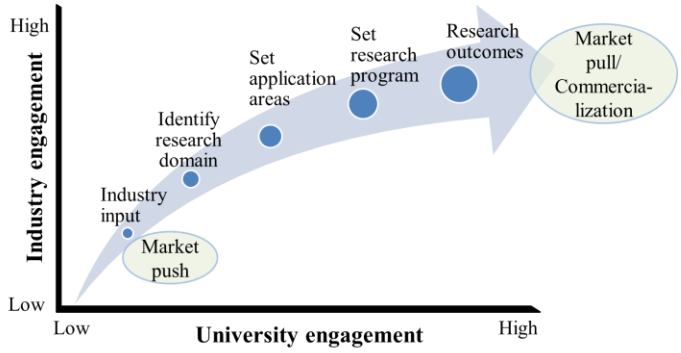


Fig. 3 University-industry engagement model

Fig. 4 shows the government-supported industry-university engagement model. Government-supported research programs often demand that university researchers secure cash and in-kind commitments from industry before lodgment of applications to funding bodies. Once these industry supports are obtained after comprehensive negotiations following the stringent engagement model depicted in Figure 4, university researchers apply for grants from the government. These grant schemes are highly competitive and internationally recognised experts review the applications. If successful, government departments such as the ARC invite researchers to execute research agreements with industry before starting the project. Industry has the right to withdraw support at any time and/or prolong negotiations to come to an agreement. This depends on the market opportunities, cash flow, changes in priorities and restructuring of innovation drives. If the research agreement is not finalised by the deadline set by the funding body, the government funding is returned to the funding body. Once the project commences, the university researchers report progress according to the milestones set by the industry partners. The researchers are required to submit annual progress reports to the funding body. If progress is not satisfactory, both industry and the funding body may withdraw support. A final report is prepared and submitted to the department and industry partner once all PhD students graduate and the funding is exhausted.

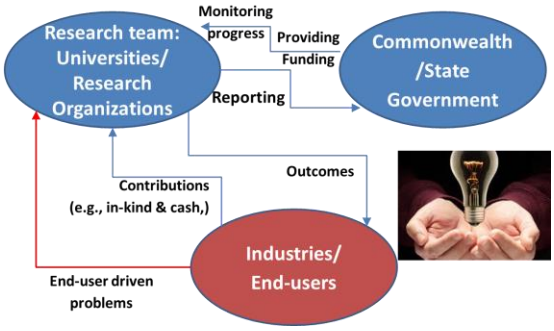


Fig. 4 Government supported university-industry engagement model

V. Conclusion and Recomendations

Improved productivity and sustainable economic growth are two key priority areas for a country to maintain her citizens' living standard and competitive advantages in the global economy. The Australian economy depends mainly on primary industries such as mining and agriculture. High technology manufacturing is leaving Australia due to the cheaper production costs in the developing world. Job losses, training people with new skills and the creation of new jobs to maintain living standards of families are the biggest challenges for Australia. Sustainable growth can only be maintained with innovation and knowledge-based economic activities. Innovations through university-industry collaboration can greatly enhance Australia's competitive advantages. SMEs make a major contribution to Australia's GDP and private sector employment. However, Australian SMEs work mainly in service sectors [14]. Developing synergies in the services and manufacturing sectors with high technology components underpins improved productivity and sustainable growth. The university-industry collaboration culture is developing in Australia but needs significant encouragement to translate inventions to innovations for business success. Some nationally-competitive grant schemes such as ARC Linkage Projects, Commercialization Australia and Cooperative Research Centres are the major contributors to research collaboration. However, the requirement of matching funds from industry, financial limitations of SMEs for R&D, cultural gaps, lack of absorptive capacity and networks, and different time-scales are key barriers to collaboration. Understanding various engagement models, ownership of IP and removing barriers in IP protection and publication through negotiation can assist the development of a conducive culture of university-industry collaboration. Irrespective of the scale of success of such research collaboration, university-industry interaction brings many tangible benefits to pedagogical practice, awareness of real-world issues, and strategic partnerships for the economic benefit of Australia. A large-scale centre for research excellence with long-term vision in a collaborative atmosphere can bring more benefits for Australian innovations. The government should develop a more comprehensive and easy-to-access research collaboration culture and support for SMEs and universities. Large profitable organizations can also play a bigger role in developing university-industry synergies. They should work collaboratively with SMEs, universities and government to facilitate university-industry engagement for the greater economic benefit of Australia.

University-industry collaboration is vital for the economic growth and competitive advantage of any country. This collaboration can be fostered and enhanced with cooperation, sound vision, planning, and execution with support from government and large profitable industries.

VI. References

[1] ABS (2013), www.abs.gov.au (accessed 12 May 2014)
[2] ABS (2013), http://www.abs.gov.au/ausstats/abs@.nsf/mf/8104.0/ (accessed 12 May 2014)
[3] Sciencebusiness (2012) Making Industry-University Partnerships Work Lessons from successful collaborations (http://www.sciencebusiness.net/assets/94fe6d15-5432-4cf9-a656-633248e63541.pdf, accessed 04/05/2014)
[4] WIPO (2007). Technology transfer, intellectual property and effective university-industry partnership. The experience of China, India, Japan, Philippines, the Republic of Korea, Singapore and Thailand. http://www.wipo.int/export/sites/www/freepublications/en/intproperty/928/wipo_pub_928.pdf. (accessed 12 March 2014)
[5] ARC (2013), http://arc.gov.au/media/releases/media_28June2013.htm (accessed 12 May 2014)
ARC (2014). http://www.arc.gov.au/ncgp/default.htm (accessed 12 May 2014)
[6] ATSE. (2013) Translating researcher into economic benefit for Australia, Position paper, Oct. 13
[7] Arora, V.K. & Chua, S.J. (1995) Electrical Engineering Education Relevant to Industrial Needs: Singapore Experience, ASEE'95.
[8] Karmakar, N.C. (2005). Developing engineering students' transferable skills through work placement at NTU, Singapore, MHEd Thesis, Griffith University, Australia.
[9] Strunz, K.; Yokoyama, A. and Behnke, R.P. (2003). Collaboration Is Key Internationally, IEEE Power & Energy Magazine, July/Aug. 2013: 50-55.
[10] Plewa, C. (2005) Key Drivers of University-Industry Relationships and the Impact of Organisational Culture Difference; A Dyadic Study, PhD Thesis, Adelaide University.
[11] Piematrix.com (2014) http://www.piematrix.com (accessed 30 April 2014)
[12] Phan, P.H. (2006). The effectiveness of university technology transfer: Lessons learned from quantitative and qualitative research in the US and the UK. Department of Economics, Rensselaer Polytechnic Institute, NY, US.
[13] Zimmermann, M; Engel, S.; Baccarella, C. and Voigt, K.I. (2012). Contract Research as a Part of New Product Development, PICMET12:.2524-2532.
[14] Daley, J. (2013). Australian government spending on innovation, Grattan Institute (grattan.edu.au, accessed 12 May 2014).