

A Determination Process of the Number and Distance of Sea Objects using CHIRP Signal in a Three Sensors Based Underwater Network

*Shaik Asif Hossain¹, M. F. Ali¹, Md. Itmam Akif¹, Rakibul Islam¹, Alok Kumar Paul² and Amit Halder¹

¹ Department of Electronics & Telecommunication Engineering

² Department of Electrical & Electronic Engineering

Rajshahi University of Engineering & Technology

Rajshahi-6204, Bangladesh

E-mail: *asifruete@gmail.com¹, forhadete10ruet@gmail.com¹, soumikete@gmail.com¹, rakibulislammm@gmail.com¹, alokeee10b@gmail.com² and amit.halder29ete@gmail.com¹

Abstract—An inspection of signal processing approach in order to determine the number and distance of objects in the sea is conducted in this paper. In recent researches, a great emphasize is put on determining the number and distance of objects in the sea. But unfortunately, it's a difficult task due to the badly effect of high absorption, long propagation delay as well as dispersion. However, besides SONAR based techniques, some conventional techniques are also familiar to these aspects. In this research, we have proposed a statistical signal processing approach called Cross-correlation, due to its simplicity and convenience for any environment in the sea. The echo effect of CHIRP signal is used here to determine the number and distance of objects in such underwater networks where three sensors are located due to the task. We have verified this proposed theory by a simulation using MATLAB.

Keywords— sensor, cross-correlation, echo Signal, CHIRP signal, SL case, TS case.

I. INTRODUCTION

In the recent advancement of underwater communication, the underwater wireless sensor networks became one of the most effective areas due to its importance. Both acoustic waves [1], [2] and magnetic waves [3], [4] are used here. In an underwater acoustic network, a great challenge is faced at the time of determining number and distance of objects. Despite having difficulties to determine these values, it needs to perform because of its importance at the field of naval-commercial activities, naval warfare, fishery management etc. Generally, three causes arise behind the difficulties and they are harsh underwater environment, Long propagation delay in under water environment and the dynamic conditions of objects respectively. The conventional ultra sound based techniques are sometimes inefficient due to many causes like the ultra sound acts as a beacon for enemy during warfare. A deleterious effect of ultra sound is implied on the marine lives sometimes. A high cost in this purpose is another cause.

In this research a straight forward method called cross-correlation is familiarized for determination purposes. CHIRP signal, which is also called an angle modulated sweeping signal, is also used here.

Two sensors based determination process of the number and distance of sea objects is described in [5]. But, unluckily, it has some accuracy problems. However, Instead of two sensors network, we have proposed three sensors based network to achieve a better accuracy. Different cases for the sensors distribution are also taken in account.

A framework is build where the number of objects as well as their distance from the sensors is determined. A CHIRP generator is used to generate CHIRP pulses. These pulses have a great likeness to sound pulses.

II. FORMULATION OF CROSS-CORRELATION FUNCTION

For the formulation tenacities, a 3D Network is employed, that contains N number of objects which are evenly distributed. Here, three sensors are employed in two cases: Case 1 – sensors in line (SL) case; and Case 2 – triangular Sensors (TS) case; [6] as shown Fig. 1.

A CHIRP source generates pulses time to time and the objects act as echo generators. The echo signals from the objects are received by the sensors first and then, these signals are summed by each sensor and finally produce a composite signal. Lastly, however the required CCFS are created through the cross-correlation of these three composite signals.

The 3D space as a cube is considered where the dimension of the cube is equal to the sphere diameter. The sensors H_1, H_2, H_3 and an object called N_1 are located at $(x_1, y_1, z_1), (x_2, y_2, z_2), (x_3, y_3, z_3)$ and (x_4, y_4, z_4) .

Distance between sensors H_1 and H_2

$$d_{DBS_{12}} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

Distance between sensors H_2 and H_3

$$d_{DBS_{23}} = \sqrt{(x_2 - x_3)^2 + (y_2 - y_3)^2 + (z_2 - z_3)^2}$$

Distance between sensors H_3 and H_1

$$d_{DBS_{31}} = \sqrt{(x_3 - x_1)^2 + (y_3 - y_1)^2 + (z_3 - z_1)^2}$$

At SL case, $d_{DBS_{12}} = d_{DBS_{23}} = d_{DBS}$,

which implies that two CCFs are possible.

At TS case, $d_{DBS_{12}} = d_{DBS_{23}} = d_{DBS_{31}} = d_{DBS}$,

which implies three CCFs are possible.

We Consider that the echo signal that comes from N_1 is $S_1(t)$, which is finitely long. Earlier, the signals received by H_1 , H_2 and H_3 are respectively:

$$S_{r11}(t) = \alpha_{11} S_{11}(t - \tau_{11})$$

$$S_{r12}(t) = \alpha_{12} S_{12}(t - \tau_{12})$$

$$S_{r13}(t) = \alpha_{13} S_{13}(t - \tau_{13})$$

where α_{11} , α_{12} and α_{13} signifies attenuations as a result of absorption and dispersion in the medium, τ_{11} , τ_{12} and τ_{13} are the corresponding time delays for the echo signals to reach the sensors and last of all S_p is the speed of wave propagation in the water. [6]

For SL case, the CCFS are:

$$C_1(\tau) = \int_{-\infty}^{+\infty} S_{11}(t) S_{12}(t - \tau_{11}) d\tau$$

$$C_2(\tau) = \int_{-\infty}^{+\infty} S_{12}(t) S_{13}(t - \tau_{12}) d\tau$$

Now for TS case the additional CCF is:

$$C_3(\tau) = \int_{-\infty}^{+\infty} S_{13}(t) S_{11}(t - \tau_{13}) d\tau$$

We need to take the total received echo- signals so that we can find the CCFs for N number of nodes.

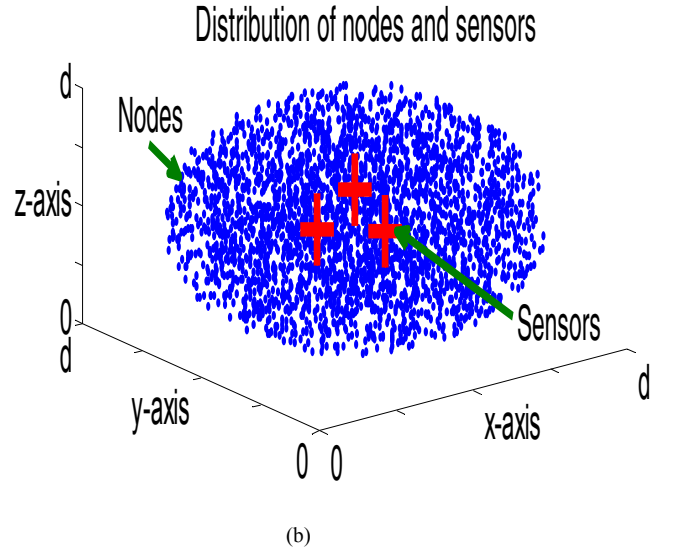
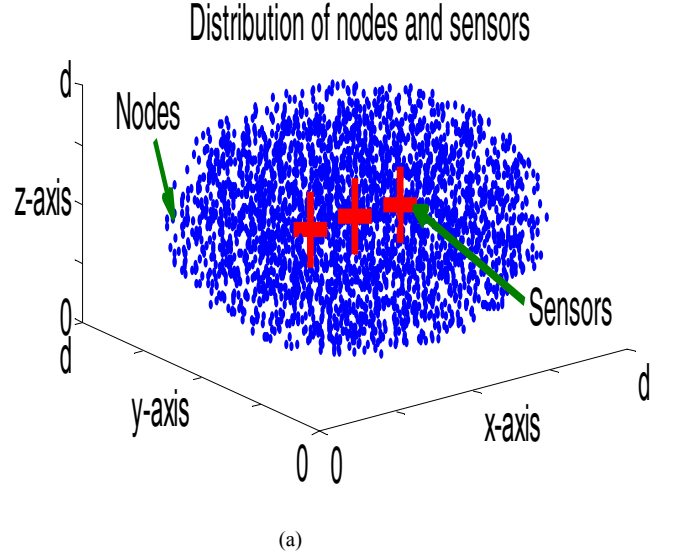


Figure. 1. An underwater network node distribution containing N number of transmitting nodes for (a) SL case; and (b) TS case.

Now, the received composite signals by H_1 , H_2 and H_3 are:

$$S_{r11} = \sum_{j=1}^N \alpha_{j1} S_j(t - \tau_{j1})$$

$$S_{r12} = \sum_{j=1}^N \alpha_{j2} S_j(t - \tau_{j2})$$

$$S_{rt3} = \sum_{j=1}^N \alpha_{j3} S_j(t - \tau_{j3})$$

And so, the total CCFs are (for SL case):

$$C_{12}(\tau) = \int_{-\infty}^{+\infty} S_{rt1}(t) S_{rt2}(t - \tau) d\tau$$

$$C_{23}(\tau) = \int_{-\infty}^{+\infty} S_{rt2}(t) S_{rt3}(t - \tau) d\tau$$

For TS case the additional CCF is:

$$C_{31}(\tau) = \int_{-\infty}^{+\infty} S_{rt3}(t) S_{rt1}(t - \tau) d\tau$$

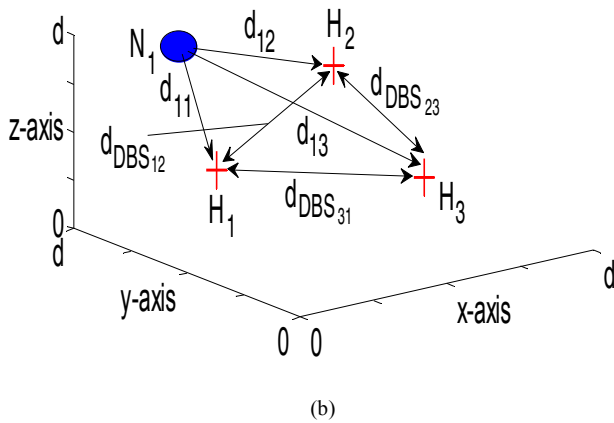
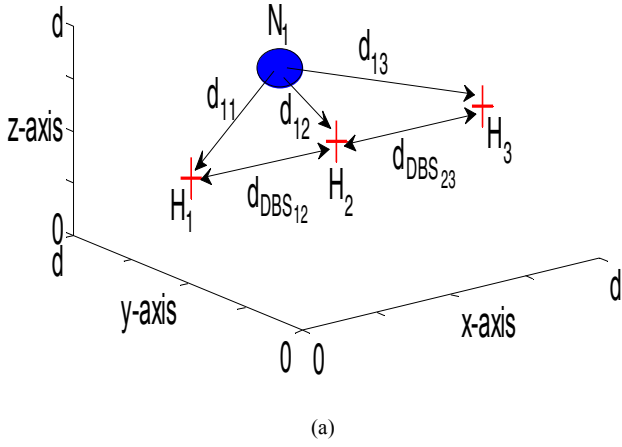


Figure 2. Three sensors (+) based underwater network and an object N_1 designed for (a) SL case; and (b) TS case.

Now, a series of delta functions are created through these.

Here, $\tau = d_{DBS} / S_p$.

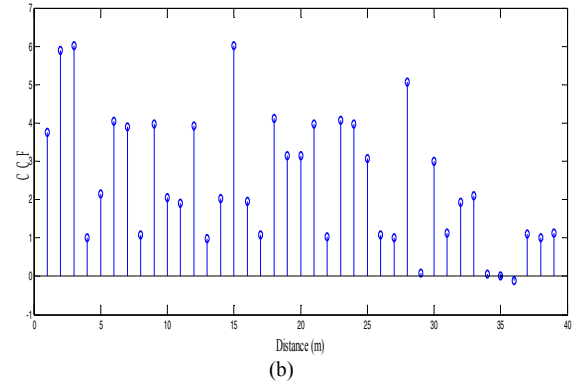
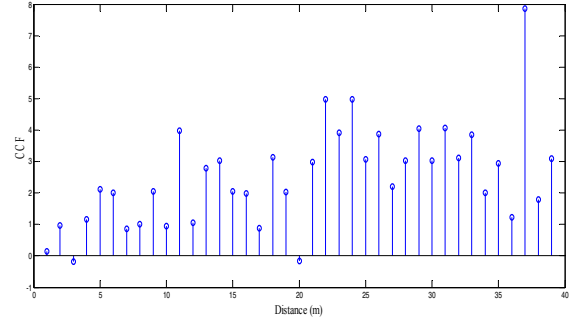


Figure 3. The cross-correlation bins, b used for (a) SL case; and (b) TS case.

III. DETERMINATION OF NUMBER AND DISTANCE OF OBJECTS USING CROSS-CORRELATION FUNCTION

In the past, the node estimation process in underwater was performed using Cross-correlation Function. [7, 8, 9]

That estimation technique was based on Gaussian signal which was used for node estimation. However, that estimation was performed for two cases:

- Equal transmitted power or ETP
- Equal receive power or ERP

We can epitomize the object determination as a certain example of estimation by means of Gaussian signal. Nevertheless, in this case, we cast-off CHIRP signal for this estimation. In this instance, a CHIRP Generator is used to transmit CHIRP Signal and echoes (CHIRP) from the objects are received by receiver (sensor).

A. CHIRP Signal

A swept-frequency signal named CHIRP Signal is a type of signal which has a time varying frequency. We can express it like:

$$X(t) = A * \cos(2 * \pi * ((f_2 - f_1)t^2 / (2 * d) + f_1 * t + P))$$

where f_1 signifies the starting frequency in Hz, f_2 signifies the ending frequency, d indicates time duration in seconds, P indicates the starting phase, and A is the amplitude. [5]

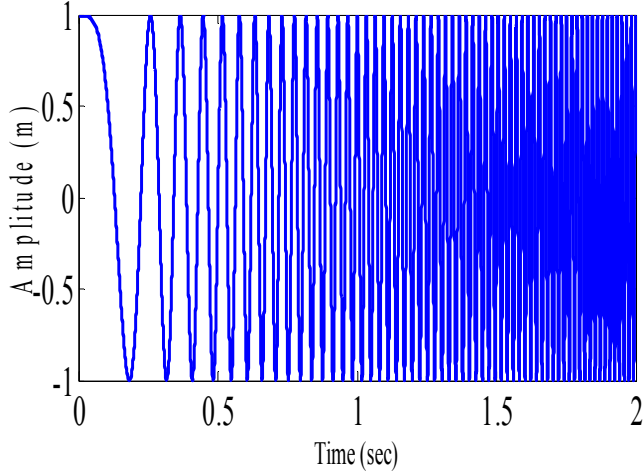


Figure 4. CHIRP signal

B. Determination of Number of Objects

Assumption:

- The echo signals are equal in power
- The objects are distributed in 3D Space

Using statistical expression, calculating CCFs to compute the standard deviation and mean is quite tough. On the other hand, cross-correlation technique is used like the problem of probability. This makes the analysis fairly easy. It is conspicuous that the mean, μ , standard deviation, σ of the CCFs and the sum are reliant on signal strength. And if the exact signal strength is known, these may be used. But this is not practical at times. On the other hand, the ratio of standard deviation to the mean of the CCFs, R , is free from reliance on signal strength. Here, it is taken as the determination tool for determining the number of objects. At this research, at first we show that R is related to N and finally we will try to show it through simulation. In Figure.5, a setup for experimental purpose is proposed for the determination of Objects using CCF. The experimental setup is similar to the presentation in Conti et al. (2006), [10]. The experiment was implemented in a tank and for a small number of Objects but the same setup can be used for determination of large number of Objects in the deep Sea. At this premises, the echo signals (CHIRP) are used.

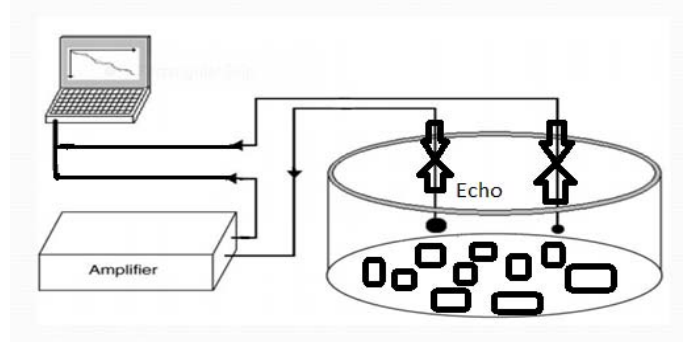


Figure 5. A setup for experimentation to determinate the number of objects using CCF.

C. Number of objects from theory

First of all, we have to determine the number of the objects. For our determination resolves, we use the ratio of standard deviation σ , to the mean μ . The ratio of Standard deviation to the mean is expressed as in the bellow: [9]

$$R = \sigma / \mu = \sqrt{\frac{(b-1)}{N}}$$

Now, b corresponds to the number of bins, can be well-defined as twice the number of samples between the sensors (NSBS), m , minus one. For SL case, the final ratio of the standard deviation to the mean can be found from the average of R_{12} , R_{23} . This indicates that two CCFs are used.

$$R_{Average}^{2CCF} = \frac{R_{12} + R_{23}}{2}$$

For TS case, the final ratio of standard deviation to the mean is obtained from the average of R_{12} , R_{23} and R_{31} . Here, three CCFs are used.

$$R_{Average}^{3CCF} = \frac{R_{12} + R_{23} + R_{31}}{3}$$

Here, we know b and at the same time we can evaluate $R_{Average}^{2CCF}$ and $R_{Average}^{3CCF}$. A point to be noted that b a function of distance between sensors, d_{DBS} , sampling rate, S_R and speed of propagation in water, S_p .

So finally we can determine the number of objects N by using the two equations above.

D. Number of objects from Simulation

As the number of bins, b is a function of distance between sensors, d_{DBS} , sampling rate, S_R and speed of propagation,

S_p , infers that b is fixed for a particular case in which the sampling rate, S_R , speed of propagation in the water, S_p , and distance between sensors, d_{DBS} are fixed. So we can extract that, the final ratio, R , shows an inversely proportional relation to the square root of the number of objects, N . So

$$R = c / \sqrt{N}$$

where, a known constant, c ($\sqrt{b-1}$) depends on occurrence between sensors, d_{DBS} , speed of propagation in the water, S_p sampling rate, S_R . As a result, the number of objects can be determined using simulation, where we have to calculate R accordingly to the simulation.

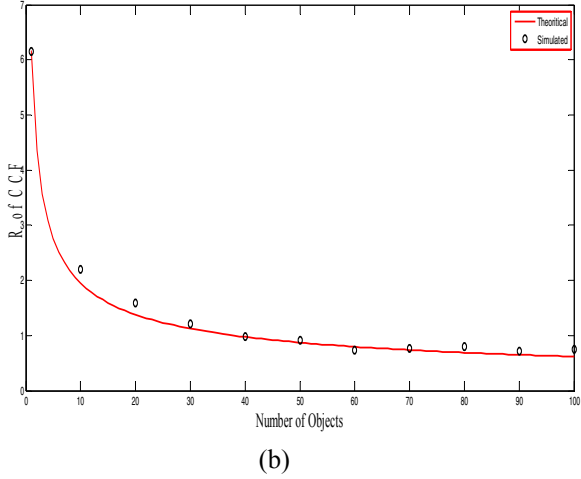
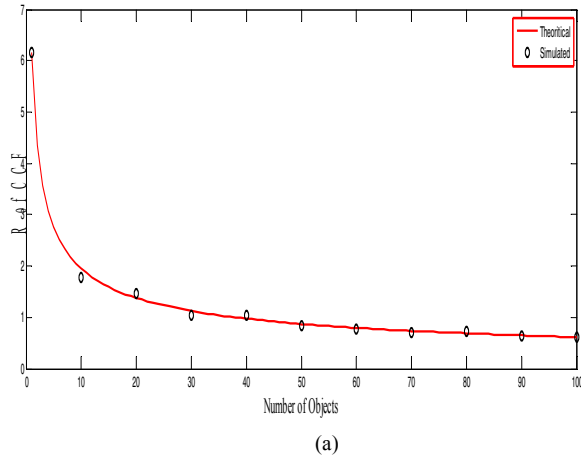
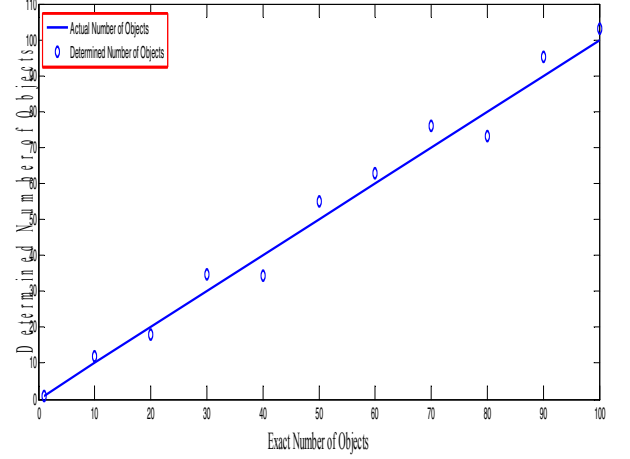


Figure 6. Ratio of Standard deviation to the Mean of CCF from theory and simulation for (a) SL case; and (b) TS case with bins, $b=39$ and $d_{DBS}=0.5m$.

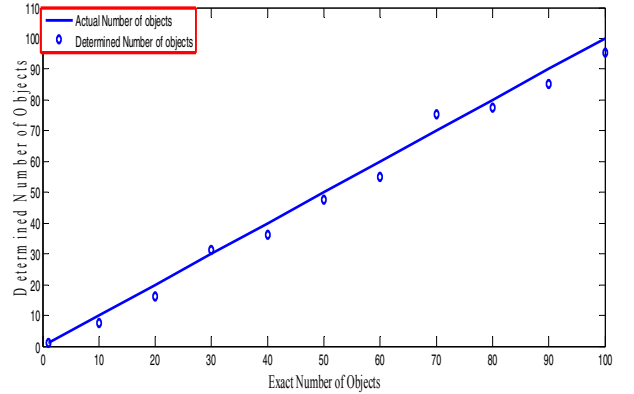
Thus the final expression can be written:

$$N = \frac{b-1}{R^2}$$

From here, as the bins, b is well-known from experimental setup, R is achieved from the simulation, the object number N is found using the equation above.



(a)



(b)

Figure 7. A variation to the determined number of objects and the actual number of objects for (a) SL case; and (b) TS case.

E. Determination of distance from simulation

The second step is for determination of distance of objects in an underwater network. In order to determine the distance, we generate CHIRP pulses using CHIRP generator to the objects. The pulses are then transmitted for several times according to our requirement. In this incident, we will know the frequency and wavelength of the pulses. When the signals come back from the objects as echoes, we can determine these according to their wavelengths and frequencies. However, a

computer is used to record the echo pulses as Figure. 5. After receiving the pulses, we will cross-correlating them and determine the cross-correlation function. The recorded pulses are good indicators of the object distances. Finally, the plot of cross-correlation functions verses distances is a direct indicator of different object position and distances.

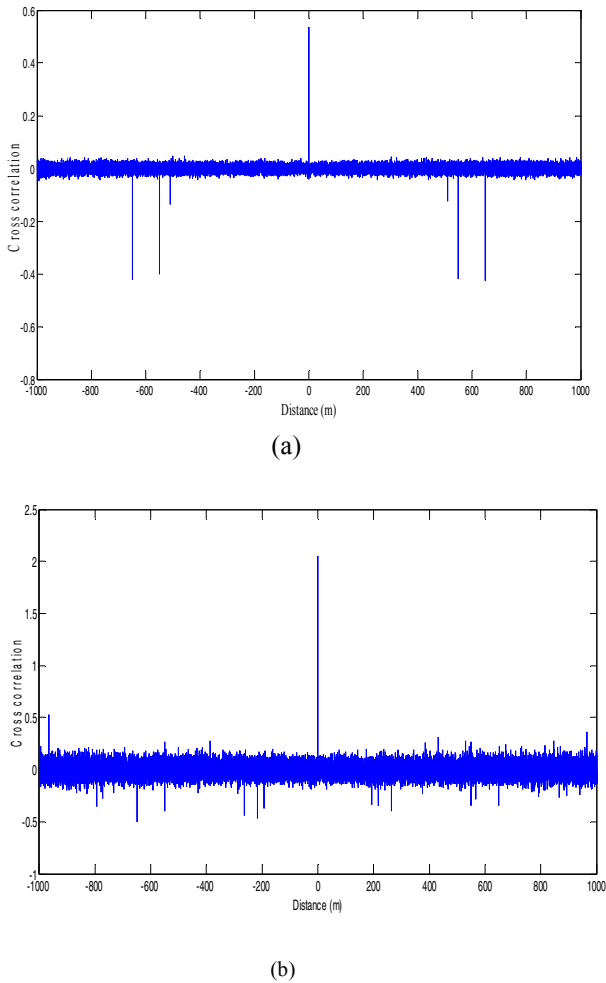


Figure 8. Different pulses represent different location of objects for (a) SL case; and (b) TS case.

From Figure. 8, it is seen that, the cross-correlation functions are found at specific positions such as (500m, 540m, 650m at sensors in line case) positions. For triangular sensors case the spike of the pulses are found at several positions such as (210m, 260m 680m etc.). These signify a good indicator of their distance parameters because the plots show a result of cross-correlation functions at different positions of objects, where both are correlated to each other.

IV. CONCLUSION

To the sum up, In the Sea, the estimation process of number of objects is difficult to use existing techniques due to some obstacles such as harsh underwater environment and mostly human interactive. So, here we proposed cross-correlation method and also used CHIRP signal which one permits high resolution on time axis. For this reason it is

greatly suitable for ranging. However, this technique has some limitations such as our consideration of receiving power is equal and the delays are presumed to be integer. But, after all this proposed method is not human interactive. That is why; it is convenient to determine the number as well as distance of objects in the sea through this process.

REFERENCES

- [1] M. Chitre, S. Shahabudeen, L. Freitag, and M. Stojanovic, "Recent advances in underwater acoustic communications & networking," IEEE Oceans'08, Quebec City, Canada, 15–18 September, 2008, pp. 1–10. doi: 10.1109/OCEANS.2008.5152045
- [2] Z. Jiang, "Underwater acoustic networks—issues and solutions," International Journal of Intelligent Control and Systems, vol. 13, no. 3, pp. 152–161, September 2008
- [3] A. A. Abdou, A. Shaw, A. Mason, A. Al-Shamma'a, J. Cullen, and S. Wylie, "Electromagnetic (EM) wave propagation for the development of an underwater Wireless Sensor Network (WSN)," in Proceedings IEEE Sensors, 2011, Limerick, Ireland, 28–31 October, 2011, pp. 1571–1574. doi: 10.1109/ICSENS.2011.6127319
- [4] A. Elrashidi, A. Elleithy, M. Albogame, and K. Elleithy, "Underwater wireless sensor network communication using electromagnetic waves at resonance frequency 2.4 GHz," in Proceedings of the 15th Communications and Networking Simulation Symposium (CNS'12), San Diego, California, USA, 2012
- [5] Hossain, S. A., Anower, M. S., & Halder, A. (2015, November). A cross-correlation based signal processing approach to determine number and distance of objects in the sea using CHIRP signal. In *2015 International Conference on Electrical & Electronic Engineering (ICEEE)* (pp. 177-180). IEEE.
- [6] Chowdhury, S. A. H., Anower, M. S., & Giti, J. E. (2014, May). Effect of sensor number and location in cross-correlation based node estimation technique for underwater communications network. In *Informatics, Electronics & Vision (ICIEV), 2014 International Conference on* (pp. 1-6). IEEE
- [7] M. S. Anower, "Estimation using cross-correlation in a communication network", PhD thesis, School of New South Wales, Australian Defense Force Academy, Australia, 2011.
- [8] M. S. Anower, M. A. Motin, A. S. M Sayem, and S. A. H Chowhury, "A node estimation technique in underwater wireless sensor network," International Conference on Information, Electronics & Vision (ICIEV), Dhaka, 17-18 May, 2013
- [9] M. S. Anower, M. R. Frater and M. J. Ryan, "Estimation by crosscorrelation of the number of nodes in underwater networks", Australasian Telecommunication Networks and Application Conference (ATNAC), 10-12 Nov. 2009
- [10] S. G. Conti, P. Roux, C. Fauvel, B. D. Maurer, and D. A. Demer, "Acoustical monitoring of fish density, behavior, and growth rate in a tank", Aquaculture, 251(2-4), pp. 314-323, February 2006.

Triple Material Double Gate TFET with Optimized Si Film Thickness

Samia Safa*

and Samantha Lubaba Noor

Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
Email: samia.safa@gmail.com

Md. Ziaur Rahman Khan

Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
Email: zrkhan@eee.buet.ac.bd

Abstract—Triple material double gate (TMDG) tunnel field-effect transistor (TFET) can be considered as a potential nominee for next generation low power high speed device fabrication. In this study, the effect of Si film thickness on performance of TMDG TFET is analyzed using two dimensional TCAD simulations. The tunneling current is substantially dependent on device thickness and the device physics regulating it is outlined. Potential and electric field distributions along film thickness reveal that double gate pairing in TMDG TFET structure can lower the tunneling barrier width at the middle of the device until an optimized silicon film thickness is achieved and provide maximum drain current.

Keywords—triple material double gate (TMDG); optimization; silicon film thickness; tunnel field-effect transistor (TFET); device physics.

I. INTRODUCTION

In recent years, Tunneling Field Effect Transistor has become a possible substitute of MOSFET devices especially in low voltage applications [1], [2]. Due to thermal injection of high energy carriers, subthreshold slope (SS) of MOSFET is limited by kT and has a minimum of $60\text{mV}/\text{dec}$. This in turn results in high leakage current and greater power consumption. The increasing power consumption is becoming a major concern in data centers due to cost rises in powering and cooling the servers. Therefore, researchers' interest has shifted to search for new device architectures geared towards low power, steep SS and high speed operation. The device physics of TFET is completely different from MOSFET. TFET is actually a gated p-i-n diode. Its working principle depends on tunneling mechanism. This makes the novel device completely unique. In general, TFET can operate at low voltage by overcoming the thermal limit of subthreshold swing voltage of $60\text{mV}/\text{decade}$ [3]. Since the device depends on tunneling, on state current is quite unsatisfactory. Various additive technologies have been introduced to the structure of the device and also different material have been proposed so far to increase the on current. Researchers have proposed vertical TFET, lateral interband TFET, carbon nanotube and silicon nanowire based TFET for improvement of on current [4]–[6]. Besides, strained material, low band gap material, high- k dielectric have been reported to increase the current [7]–[9]. Lately,

researchers studied the performance of TFET with different gate structure and varying material [10], [11] and eventually proposed triple material gate structure in TFET [12]. It has been found that the device exhibits high on state current with an improved SS and suppressed ambipolar conduction. In particular, in trimaterial gate device, the lateral energy band profile and the electric field contour are tuned by the energy band profile modulation (BPM) effect, which manipulates the carrier tunneling probability and space distribution, thus improving device performance. Besides, double gate structure has already been proved to be advantageous for MOSFETs in terms of scalability option, transconductance and Short Channel Effects (SCEs) immunity. The Trimaterial Double Gate TFET incorporates the advantages of both double gate and multiple material gate structure. Few researches have been reported on this novel device; only some analytical models have very recently been published [13], [14].

In this work, a Si based TMDG TFET has been developed in Silvaco, Atlas [15]. The device has been analyzed to find the optimized value of Si film thickness which may produce maximum possible ON current and minimum OFF current simultaneously. The findings of the work will be helpful for fabrication and further improvement of the device.

II. DEVICE STRUCTURE

Fig. 1 shows the schematic representation of the TMDG TFET device. The structure consists of a highly doped p-type source, intrinsic channel (lightly n-doped), and highly doped n-type drain region. The gate near the source region is responsible for tunneling mechanism between source/body junctions. Energy bands of the TMDG TFET during on-state ($V_{GS} = 1\text{V}$ and $V_{DS} = 1\text{V}$) and off-state ($V_{GS} = 0\text{V}$ and $V_{DS} = 0\text{V}$) are depicted in Fig. 2. At $V_{GS} = 0\text{V}$, the barrier width is very large and no tunneling occurs. At high V_{GS} , the channel conduction band goes below the source valence band. Under such scenario, the width of tunnel barrier is small enough to allow tunneling and electron tunnels from source valence band to the channel conduction band.

The TMDG TFET considered in this study has source doping of 10^{20}cm^{-3} , channel doping of 10^{15}cm^{-3} and drain doping of 10^{19}cm^{-3} . Using this type of asymmetrical doping

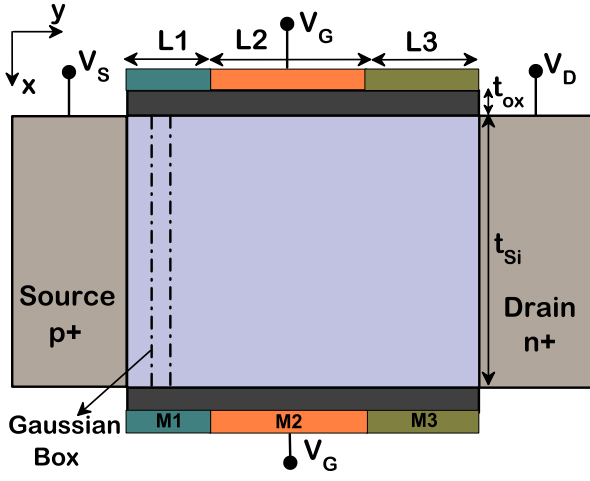


Fig. 1. Schematic representation of TMDG TFET Structure.

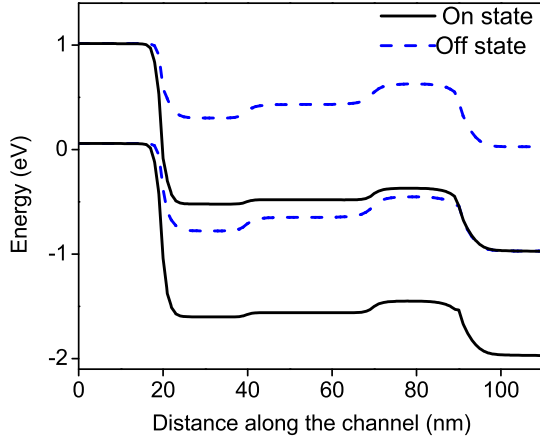


Fig. 2. Modulation of TMDG TFET energy band by gate voltage.

profile, ambipolar effect can be suppressed [8]. The gate length of the device is $L = 70\text{nm}$ with $L_1 = 20\text{nm}$, $L_2 = 30\text{nm}$, and $L_3 = 20\text{nm}$. Three metals $M1$, $M2$, and $M3$ with work-functions 4.4eV , 4.6eV and 4.8eV respectively forms the gate. Thickness of oxide region is $t_{ox} = 2\text{nm}$. Silicon film thickness, t_{Si} is varied between 3nm – 25nm .

III. RESULT AND DISCUSSION

The effects of Si film thickness on electric field, electric potential, and drain current are studied by performing simulation in Silvaco Atlas. Junctions were quasi-perfectly abrupt in all simulations. Kane's band-to-band tunneling (BTBT) model is considered for the simulations. It determines the BTBT generation rate of carrier tunneling as

$$G_{BTBT} = A_{kane} E^D \exp\left(-\frac{B_{kane}}{E}\right) \quad (1)$$

E represents local field, D depends on the tunneling process. For direct tunneling $D = 2$ and for indirect tunneling $D = 2.5$. In case of Silicon, $D = 2.5$ is used due to indirect bandgap

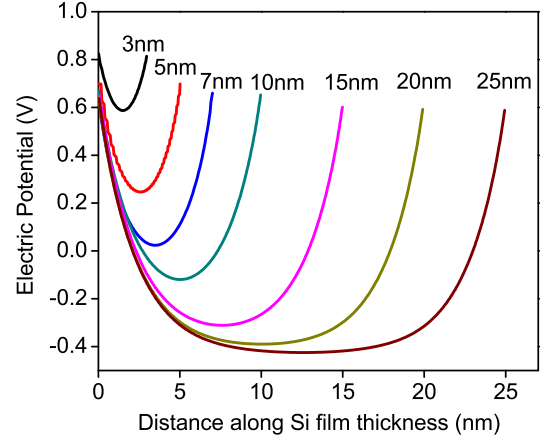


Fig. 3. Electric potential along x direction near source-channel tunneling junction.

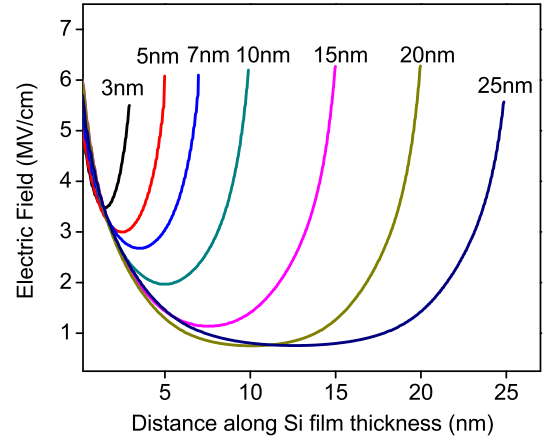


Fig. 4. Electric field along x direction near BTBT region of source-channel interface.

of the material. $A_{kane} = 4 \times 10^{19} \text{eV}^{1/2}/\text{cm-s-V}^2$ and $B_{kane} = 4.1 \times 10^7 \text{V/cm-eV}^{3/2}$ are the tunneling process dependent parameters. These parameters have been determined by precisely recreating the experimental results published in [16]. The other models considered during simulation are: electric field dependent mobility, concentration dependent mobility, auger recombination, SRH recombination, concentration dependent lifetime SRH, bandgap narrowing.

Fig. 3 and Fig. 4 show the profile of potential and electric field vertical to channel length and close to the tunneling region for different t_{Si} . Electric potential and electric field are highest at the surface and contribute a great portion of BTBT current. However, a subsurface part of Si film yields a significant fraction of the total BTBT current. From these figures, it is evident that gate-to-channel coupling at $x = t_{Si}/2$ is strong below film thickness of 7nm . Therefore, when t_{Si} is lower than 7nm , electric potential and electric field at the middle of Si film can yield substantial BTBT.

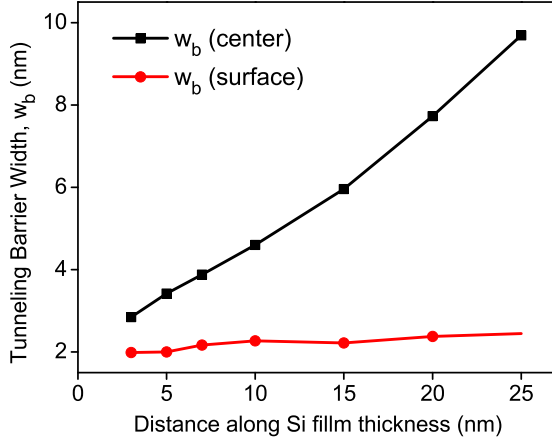


Fig. 5. Energy barrier width, w_b at the channel surface and at the middle ($x = t_{Si}/2$) of Si film for different t_{Si} .

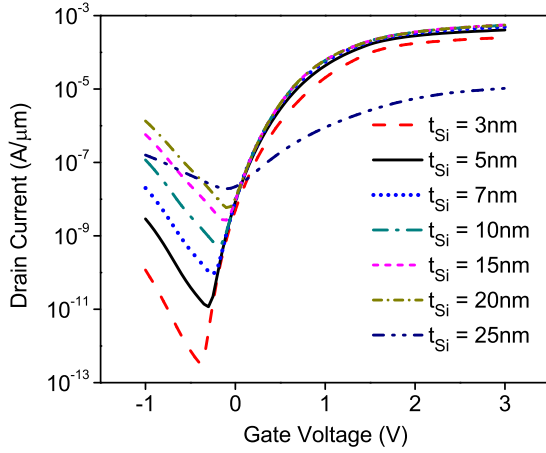


Fig. 6. Effect of t_{Si} on I_D vs. V_{GS} characteristics.

At the surface, tunneling barrier width, w_b is the smallest due to strongest gate control which is depicted in Fig. 5. The gate control becomes weaker towards the center of the film which in turn results in wider w_b and reduces the tunneling probability. From Fig. 5 it is evident that coupling of two gate electrodes is strong up to $10nm$.

Fig. 6 shows the transfer characteristics of the device for varying t_{Si} . At t_{Si} of $3nm$, potential is very high at the middle of the film but on-state drain current, I_{ON} is low due to limited availability of cross sectional area for current flow (volume limiting effect). I_{ON} increases as t_{Si} is increased due to availability of larger volume for band-to-band tunneling but at a cost of higher OFF current, I_{OFF} . I_{ON} increases slightly when t_{Si} is increased from $5nm$ to $20nm$ which is shown in Fig. 7. This happens because at t_{Si} above $7nm$ electric potential at center cannot result in enough tunneling to increase the drain current. When t_{Si} is above $20nm$, I_{ON} starts to drop for weaker double gate-to-channel coupling effect [17].

The effect of t_{Si} variation on average current density, J_{avg}

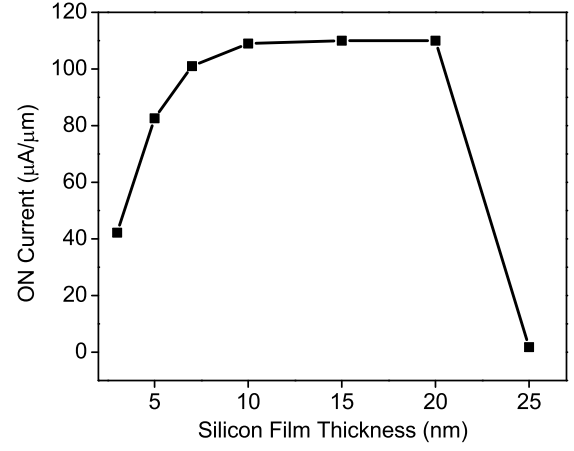


Fig. 7. Modulation of I_{ON} by Si film thickness.

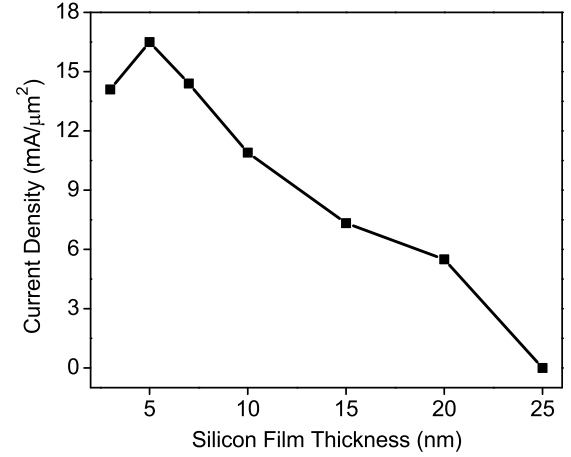


Fig. 8. Variation of average current density with Si film thickness.

is depicted in Fig. 8. J_{avg} can be calculated as

$$J_{avg} = \frac{I_{ON}}{t_{Si}} \quad (2)$$

J_{avg} increases as the film thickness is reduced and is maximum when t_{Si} is $5nm$. Below $5nm$, J_{avg} decreases as I_{ON} drops due to volume limiting effect.

Fig. 9 shows the I_{ON}/I_{OFF} for different t_{Si} . I_{ON}/I_{OFF} decreases as t_{Si} increases. This is because I_{OFF} increases with t_{Si} due to larger device volume. According to ITRS (2014), for low power devices I_{ON}/I_{OFF} of 10^5 is needed. TMDG TFETs with $t_{Si} < 15nm$ meet this requirement. Although highest I_{ON}/I_{OFF} is at $t_{Si} = 3nm$, at this t_{Si} , I_{ON} is comparatively low.

From the above discussions, it can be suggested that t_{Si} should be in the range of $5nm - 7nm$. I_{ON} is low below this range. Above this range, I_{ON} does not increase significantly and larger t_{Si} reduces the I_{ON} . This range also meets the criterion for I_{OFF} ($10^{-9} A/\mu m$) mentioned in ITRS (2014).

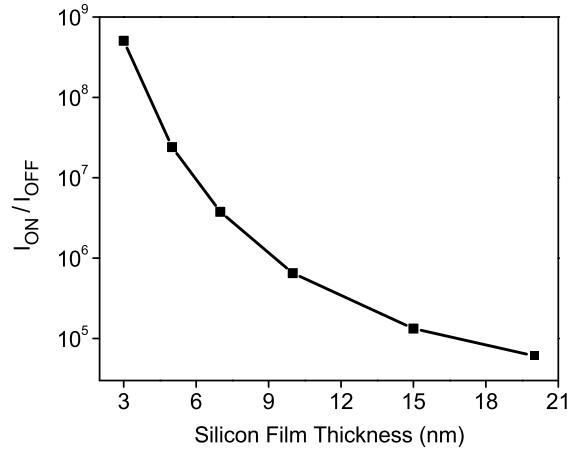


Fig. 9. I_{ON}/I_{OFF} for different Si film thickness.

IV. CONCLUSION

The underlying device physics that explains the optimization of TMDG TFET in terms of Si body thickness has been discussed. It has been found that coupling of gate is significant at the middle of Si film. Increasing Si thickness aids the tunneling process and hence increases the tunneling current upto a certain limit because of increased tunneling volume. Beyond this thickness, gate coupling reduces resulting in a reduced current. The findings of the work will be helpful for researchers and engineers to fabricate TMDG TFET device with optimized performance.

ACKNOWLEDGMENT

The authors would like to thank the Department of EEE, East West University, Bangladesh for the support with Atlas, Silvaco software.

REFERENCES

- [1] A. C. Seabaugh and Q. Zhang, "Low-Voltage Tunnel Transistors for Beyond CMOS Logic," in *Proc. IEEE*, vol. 98, no. 12, pp. 2095-2110, Dec. 2010.
- [2] A. M. Ionescu, and H. Riel, "Tunnel field-effect transistors as energy-efficient electronic switches," *Nature*, vol. 479, no. 7373, pp. 32937, Nov. 2011.

- [3] W. Y. Choi, B. -G. Park, J. Duk, and T. -J. King, "Tunneling Field Effect Transistor (TFETs) With Subthreshold Swing (SS) Less Than 60 mV/dec," *IEEE Electron Device Lett.*, vol. 28, no. 8, pp. 743-745, Aug. 2007.
- [4] C. Aydin, A. Zaslavsky, S. Luryi, S. Cristoloveanu, D. Mariolle, D. Fraboulet, and S. Deleonibus, "Lateral interband tunneling transistor in silicon-on-insulator," *Appl. Phys. Lett.*, vol. 84, no. 10, pp. 1780-1782, Mar. 2004.
- [5] J. Appenzeller, Y. -M. Lin, J. Knoch, and P. Avouris, "Band-to-band tunneling in carbon nanotube field-effect transistors," *Phys. Rev. Lett.*, vol. 93, no. 19, pp. 196805-196808, Nov. 2004.
- [6] M. T. Bjork, J. Knoch, H. Schmid, H. Riel, and W. Riess, "Silicon nanowire tunneling field-effect transistor," *Appl. Phys. Lett.*, vol. 92, no. 19, p. 193504, May 2008.
- [7] E. -H. Toh, G. H. Wang, G. Samudra, and Y. -C. Yeo, "Device physics and design of germanium tunneling field-effect transistor with source and drain engineering for low power and high performance applications," *J. Appl. Phys.*, vol. 103, no. 10, p. 104504, May 2008.
- [8] T. Krishnamohan, D. Kim, S. Raghunathan, and K. Saraswat, "Double-gate strained-Ge heterostructure tunneling FET (TFET) with record high drive currents and $<<60$ mV/dec subthreshold slope," in *IEEE IEDM*, 2008, pp. 1-3.
- [9] K. Boucart, and A. M. Ionescu, "Double-gate tunnel FET with high- κ gate dielectric," *IEEE Trans. on Electron Devices*, vol. 54, no. 7, pp. 1725-1733, Jul. 2007.
- [10] N. Cui, R. Liang, J. Wang, and J. Xu, "Lateral energy band profile modulation in tunnel field effect transistors based on gate structure engineering," *AIP Adv.* 2, p. 022111, 2012.
- [11] Upasana, R. Narang, M. Saxena, and M. Gupta, "Modeling and TCAD Assessment for Gate Material and Gate Dielectric Engineered TFET Architectures: Circuit-Level Investigation for Digital Applications," *IEEE Trans. on Electron Devices*, vol. 62, no. 10, pp. 3348-3356, Aug. 2015.
- [12] R. Liang, N. Cui, M. Zhao, J. Wang, and J. Xu, "High performance Tunnel Field Effect Transistor with a Tri-Material-Gate structure," in *Proc. ISDRS*, Dec. 2011, pp. 1-2.
- [13] D. Saraswathi, N. B. Balamurugan, G. Lakshmi Priya, and S. Manikandan, "A Compact Analytical Model for 2D Triple Material Surrounding Gate Nanowire Tunnel Field Effect Transistors," *Intelligent Computing and Applications*, vol. 343, pp. 325-332, Feb. 2015.
- [14] N. Bagga and S. K. Sarkar, "An Analytical Model for Tunnel Barrier Modulation in Triple Metal Double Gate TFET," *IEEE Trans. on Electron Devices*, vol. 62, no. 7, pp. 2136-2142, Jul. 2015.
- [15] *ATLAS Device Simulation Software*, Silvaco Int., Santa Clara, CA, USA, 2012.
- [16] J. Wan, C. L. Royer, A. Zaslavsky, and S. Cristoloveanu, "A tunneling field effect transistor model combining interband tunneling with channel transport," *J. Appl. Phys.*, vol. 110, no. 10, pp. 104503-1104503-7, 2011.
- [17] E. -H. Toh, G. H. Wang, G. Samudra, and Y. -C. Yeo, "Device physics and design of double-gate tunneling field-effect transistor by silicon film thickness optimization," *Appl. Phys. Lett.*, vol. 90, no. 26, p. 263507, 2007.

Semi Supervised Keyword Based Bengali Document Categorization

Fahim Quadery*, Abdullah Al Maruf†, Tamjid Ahmed‡, Md. Saiful Islam§

Department of Computer Science and Engineering, Shahjalal University of Science and Technology
Kumargaon, Sylhet-3114, Bangladesh

Email: *fahim@student.sust.edu, †maruf@student.sust.edu, ‡tamjid@student.sust.edu, §saiful-cse@sust.edu

Abstract—Document Categorization is an area of important research over the last couple of decades. The basic task in document categorization is classifying a given document in some predefined classes. Bengali is among the top ten most spoken languages in the world and is spoken by more than 200 million people, but the candid truth is, it still lacks significant research efforts in the area of Bengali Document Categorization. In the first phase of this paper a model has been designed that extracts keywords from a Bengali document. We crawled over 35000 news documents from popular Bengali newspapers and journals. Those documents have been stemmed and less significant words are removed using stemmer and Parts-of-Speech(POS) tagger. Statistical approach is used to extract keywords from the documents. Then probabilistic distribution and semi supervised learning with Naïve Bayes algorithm is used to approximate the category of a given Bengali document. Result and statistical data show the effectiveness of this model.

Index Terms—Chi-square, Probability distribution, Bengali Document, Word frequency, Naïve Bayes, supervised learning, semi supervised learning, Training Dataset, Pipilika, Chorki, Co-occurrence matrix, Classifier, Stemmer, POS tagger, Corpus.

I. INTRODUCTION

Document categorization has a vast application in this computer dominated time. In this paper we will discuss what document categorization is, why it is important and step by step how a model has been built that classifies a Bengali document in some predefined classes.

Document classification or document categorization is a problem in library science, information science and computer science. The task is to assign a document to one or more classes or categories. Document categorization has versatile applications. Spam filtering, sentiment analysis, language identification, finding similar articles, automatically determining the genre of a text.

II. BENGALI DOCUMENT CATEGORIZATION AND RELATED WORKS

Document categorization is now an important topic of library science. There is a healthy number of veteran researchers working on this topic all over the globe. But all these research are mostly focused on English documents categorization. Despite of being one of the most widely

spoken language, for Bengali document categorization, there is a little work and resource available sporadically on the World Wide Web.

Most of the online news portals categorize their news manually. However few research works have been done on Bengali document categorization. [1] Proposed a n-gram based technique for document categorization. In [2] Naive-Bayes classification is used to predict category for news articles. Over 7000 documents were crawled from newspapers in [2]. Documents were tokenized, stemmed and stop words were removed. Then documents were treated as a vector and using Naive Bayes Classifier documents were categorized. There are two Bengali search engines called "Pipilika" and "Chorki" that use Bengali document categorization. This paper tested and learned about [1], [2] and Pipilika search engine's approach, robustness of their models, i.e. pros and cons. And also compared their work with this paper.

III. BACKGROUND STUDIES

There are different approaches available for English and Bengali language for the keyword extraction. There is statistical approach which uses statistical models like chi-squared [3], N-gram [4] [1] etc. There is linguistic approach that mainly analyze the linguistic feature of the language to find out the important keyword of the documents [5]. Machine learning approach which is supervised or semi-supervised learning approach [6]. Noun phrase extraction approach extracts noun phrase from the document and tries to find out the important keyword from the document using statistical analysis known as Conditional Random Field (CRF) model [7]. Graph base approach is to build a graph using the linguistic analysis of the document. Then give a leap word and propagate the value to the other node of the graph [8]. Then finds out the most weighted keyword for the documents. In this paper a semi supervised machine learning based approach on top of statistical analysis is introduced which required studying several machine learning models. And finally found a good model that suits Bengali language perfectly. Study results on machine learning models and statistics is described in the paper.

A. Chi-square Distribution

Chi-Square distribution can be used to measure the difference between the actual count and the expected count of distinct elements [3]. In this paper Chi-Square was used to determine important words in a document. Taking degree of freedom one the equation of Chi-Square distribution is

$$\chi^2 = \sum_{i=0}^n \frac{\{o_i - \varepsilon_i\}^2}{\varepsilon_i} \quad (1)$$

Where,

χ^2 = Chi-Square Value

o_i = Observed frequency

ε_i = Expected frequency

B. Naïve Bayes Classifier

In machine learning, Naive Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes' theorem with strong (naïve) independence assumptions between features [9]. Naive Bayes probabilistic classification model was first used for text classification by D. Lewis [10].

Let U a Universe with all the possible outcomes (of an experiment for instance), and we are interested in some subset of them, namely some event. Say we are studying event A and B. Probability of event A and B happening individually is

$$P(A) = \frac{|A|}{|U|} \quad (2)$$

and,

$$P(B) = \frac{|B|}{|U|} \quad (3)$$

now, given B probability of event AB is

$$P(A|B) = \frac{P(AB)}{P(B)} \quad (4)$$

and vice versa, given event A, probability of event AB is

$$P(B|A) = \frac{P(AB)}{P(A)} \quad (5)$$

Putting this two equations together we get,

$$P(A|B)P(B) = P(B|A)P(A) \quad (6)$$

and finally,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (7)$$

Which is Bayes theorem. What actually the insight here is to train this algorithm with some features against some classes, and when we feed the model some features, it returns us classes.

IV. METHODOLOGY

In this paper, classifier was trained with more than 35,000 Bengali documents mostly collected from newspapers and journals. First each of these were Stemmed and less important words(i.e. অব্যয়) were omitted from the document. then the most frequent words (top 30%) were filtered out. After finding out the frequent words, the Chi-Squared distribution value of each terms with the frequent words was calculated. Those apparently sporadic terms having higher Chi-Squared value are tend to be more important in that document.

Once the important words for all documents under a category are found, then all those important words are treated like a single document for that particular category. Another Chi-square was used on important words to filter out the keywords for that particular category. This second Chi-square trimmed our training data set, streamlined the model and had a big impact on category approximation accuracy and pre-processing CPU time.

When a new document is to be classified, keywords of that documents are extracted and for that document we used Naïve Bayes classifier on those keywords with each category. And this approach captivated document categorization.

A. Extracting Keywords

According to [3] Chi-Squared value for each of the terms of the document with the most frequent terms,

$$\chi^2 = \sum_{g \in G} \frac{\{freq(w, g) - p_g n_w\}^2}{p_g n_w} \quad (8)$$

A document contains sentences of different lengths. If a term appears in a long sentence, it is more likely to co-occur with many terms. If a term appears in a short sentence, it is less likely to co-occur with other terms [3]. Here,

P_g As (the sum of the total number of terms in sentences where g appears) divided by (the total number of terms in the document).

n_w As the total number of terms in sentences where w appears.

Again $n_w P_g$ represents the expected frequency of co-occurrence. However, it's value becomes more sophisticated. A term co-occurring with a particular term $g \in G$ has a high χ^2 value. However, these terms are sometimes adjuncts of term g and not important terms. [3]

Here is the value of Chi-square on sample data-set.

Table I
CHI-SQUARE VALUES OF WORDS

No	Word	Chi-Square value
1.	পুলিশ	46901.248
2.	উপজেলা	33105.5258
3.	গতকাল	29368.5810
4.	রাত	28220.190
5.	জানা	28218.628

After the calculation of first Chi-Square on the input set, top 20% words were filtered out according to higher value. On those set of words, Chi-Square was used second time to filter out once again. And hence end up with the key words.

Table II
CHI-SQUARE VALUES AFTER 2ND PASS

No	Word	Chi-Square value
1.	পুলিশ	1324.22
2.	উপজেলা	966.98
3.	গতকাল	914.41
4.	রাত	619.34
5.	জানা	469.90

B. Classifying Documents

As mentioned previously, once the key words are extracted, Naïve Bayes classifier is used on those key words to train the model. To serve this purpose, the classifier is trained with over 35,000 documents of 12 categories.

1) *Training the Naïve Bayes Model*: The occurrence of each keyword under each class is counted. From this a table that contains the count of features against all classes is made. A sample table is shown III. This table shows some frequent key words and their corresponding occurrences under different classes [3]. From this table the Prior Probability, Probability of features and Probability of Likelihood are calculated. All these terms are essential for calculating Naïve Bayes and are described here in gist.

Table III
KEYWORD CO-OCCURRENCE TABLE

Feature Class	গতকাল	দেশ	পুলিশ	জানান	বাংলাদেশ	মোট
Accident	1064	42	761	877	71	1938
Opinion	291	3948	1049	284	2885	8457
Crime	1611	197	1679	1571	211	5269
Total	2966	4187	3489	2732	3057	15664

Now the Prior probability is guessing a category blind folded. For instance, base rate or prior probability of the above category are calculated below.

$$P(\text{accident}) = \frac{1938}{15664}$$

$$P(\text{opinion}) = \frac{8457}{15664}$$

$$P(\text{crime}) = \frac{5269}{15664}$$

Probability of Evidence or Feature Probability is the probability of getting an individual feature from all the features. Feature Probability of first 3 features is calculated.

$$P(\text{গতকাল}) = 2966/15664$$

$$P(\text{পুলিশ}) = 3489/15664$$

$$P(\text{বাংলাদেশ}) = 3057/15664$$

2) *Calculating the Probability Distribution*: Probability of Likelihood of a feature is nothing but the conditional probability of finding a feature, while a class is given. Let's calculate the likelihood of first few features on Opinion category

$$P(\text{পুলিশ}/\text{Opinion}) = 1049/8457$$

$$P(\text{জানান}/\text{Opinion}) = 284/8457$$

$$P(\text{বাংলাদেশ}/\text{Opinion}) = 2885/8457$$

Let's calculate the probability distribution of a set of features F. Which is the multiplication of likelihood of each feature from the feature set over the probability of evidence. For example, after calculating the probability of the feature set for all the classes, classes are sorted in descending order of probability value. Category with highest probability in the classifier is more likely being the category of that document.

V. TRAINING DATA AND STATISTICS

A. Classes or Categories

This paper used 12 different categories. These categories cover a major portion of Bengali news categories. The categories that were worked with are given.

Table IV
CATEGORY LIST

Accident	Art	Crime	Economics
Environment	International	Education	Entertainment
Opinion	Science-tech	Politics	Sports

The other existing Bangla news categorization model from "Pipilika" search engine worked on four categories. Categories that Pipilika used are shown in V

Table V
PIPILIKA SEARCH ENGINE'S CATEGORY LIST

Economics	Entertainment	International	Sports
-----------	---------------	---------------	--------

From the Naïve Bayes equation it is trivial to show that increasing in number of classes will reduce the performance of the classifier. Thus the number of classes and the performance is reversely related.

B. Training Model

Naïve Bayes model is trained with a total of 35,580 unique text files. Twelve categories of Bengali documents are used.

This dataset is used to train the classification model by calculating the prior probability and evidence probability. Then for every category the conditional probability is calculated using the key words of that category. Once all the conditional probabilities are calculated, then it is very straightforward to calculate the probability of a certain class for a given feature set. The number of training set has an almost linear relation with performance.

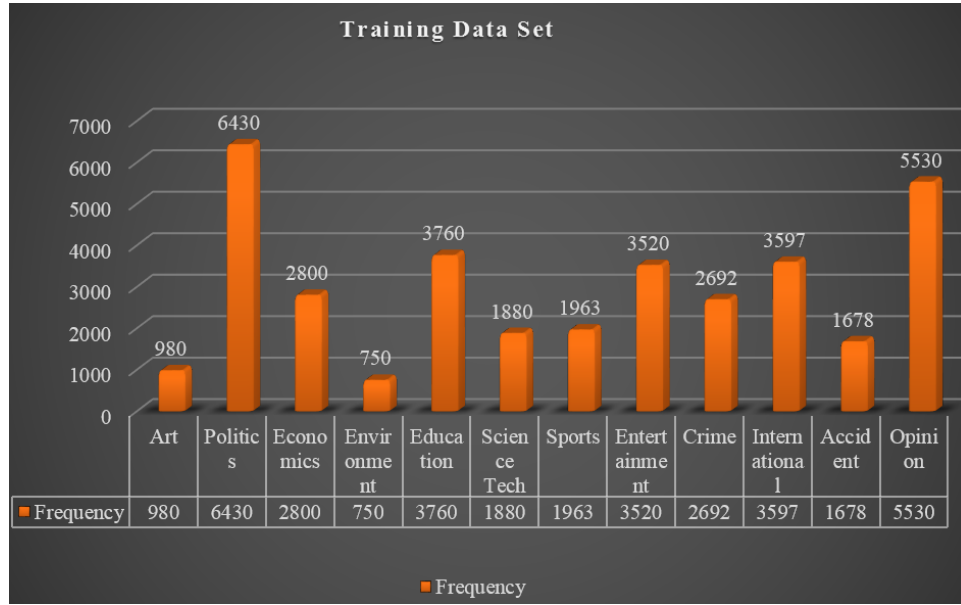


Figure 1. Training Data-set.

C. Statistical Analysis

Here some statistics of the training data is presented. For the lack of working space, the category names in English alphabets are mapped. The mapping is shown in Table VI.

Table VI
CLASS MAPPED TO ALPHABET

A	Accident	B	Art	C	Crime
D	Economics	E	Education	F	Entertainment
G	Environment	H	International	I	Opinion
J	Science-Tech	K	Politics	L	Sports

Now here in Table VIII are the most frequent words of each category. Table VII shows the top 5 words with highest Chi-Square value for each class.

Table VII
TOP FIVE WORDS WITH MAXIMUM CHI-SQUARE VALUE FOR FIVE CATEGORY

A	Word	উপজেলা	নিহত	গতকাল	পুলিশ	জানান
	χ^2 Value	20	15	14	13	12
B	Word	কথা	মানুষ	শেষ	দেখা	মন
	χ^2 Value	175	163	155	154	150
C	Word	পুলিশ	উপজেলা	গতকাল	রাত	জানা
	χ^2 Value	469	331	293	282	282
D	Word	বাংলাদেশ	টাকা	ব্যাংক	দেশ	হাজার
	χ^2 Value	50	45	42	39	34
L	Word	ম্যাচ	দল	বিশ্বকাপ	রান	বাংলাদেশ
	χ^2 Value	35	32	28	27	25

VI. PERFORMANCE ANALYSIS

A. Accuracy of The Model

From observation here it is clear that there is almost a linear relation between the size of training data-set and

Table VIII
MOST FREQUENT WORDS

A	word count	দিক	উপজেলা	গতকাল	জানান	নিহত
	Count	1220	1069	1064	877	871
B	word Count	কথা	মন	মধ্যে	মানুষ	পা
	Count	600	582	509	497	488
C	Word Count	পুলিশ	গতকাল	জানান	থানা	উপজেলা
	Count	1679	1611	1571	1510	1477
D	Word Count	বাংলাদেশ	মধ্যে	দেশ	হবে	টাকা
	Count	1552	1483	1390	1300	1264
E	Word Count	শিক্ষার্থী	হবে	আজ	ঢাকা	অংশ
	Count	2976	2560	2509	2507	2326
F	Word Count	ছবি	শুরু	হবে	করেছেন	মধ্যে
	Count	1547	1369	1352	1202	1187
G	Word Count	গেছে	দেখা	মধ্যে	এলাকা	পা
	Count	418	399	389	370	357
H	Word Count	গত	খবর	গতকাল	দেশ	মধ্যে
	Count	1963	1920	1683	1677	1592
I	Word Count	হবে	দেশ	মধ্যে	হল	কথা
	Count	4104	3948	3648	3554	3444
J	Word Count	কথা	করেন	গতকাল	হবে	সরকার
	Count	3795	3648	3433	3354	3253
K	Word Count	তথ্য	কাজ	ব্যবহার	তৈরি	নতুন
	Count	885	825	823	798	796
L	Word Count	ম্যাচ	দল	বিশ্বকাপ	শেষ	পটম
	Count	1317	1300	1000	976	880

output efficiency. Higher number of training data for a given category tends to give more accurate approximation. This data-set vs. accuracy graph is shown in figure 2.

From the linear relationship in this graph it shows that this model with a larger dataset can provide even more accurate category approximation. So a way of improving this model is increasing the training dataset.

It is easy to observe the relation between training set

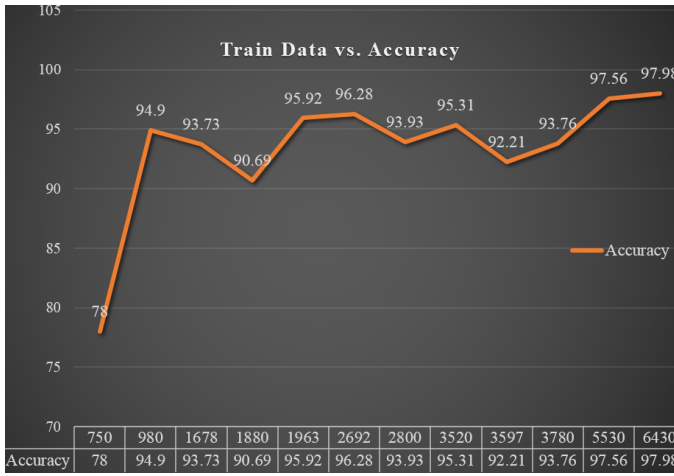


Figure 2. Relation Between Training Data-set and Efficiency.

and category approximation is not strictly linear. From observation, performance depends on availability of keywords in training set under each category. This concludes, a better stemming of dataset will increase the performance of this model.

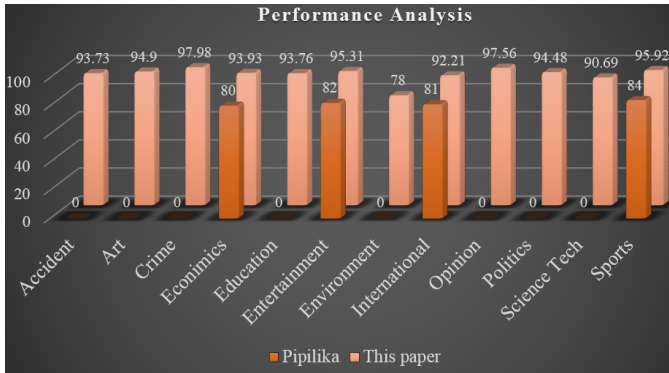


Figure 3. Performance Comparison Between this Paper and Pipilika Search Engine's model.

B. Result of Other Related Works

[1] Used one year "Daily Prothom-Alo" news corpus. It tested the model with 25 randomly selected documents. Accuracy ranges from 20% to 100% (for 3 categories using tri-gram) over 6 categories. [2] Used more than seven thousand documents extracted from "Daily Prothom-Alo" and manually tagged six hundred twenty eight news where two hundred eight items were used as training set. It achieved a highest precision rate of 0.8.

Figure 3 shows the performance analysis between this model and Pipilika search engine's model which is developed under "SUST NLP research lab". Pipilika search engine achieved a average accuracy of 82% over four category. It shows this paper outperforms Pipilika's model in every aspects.

VII. CONCLUSION

Document categorization is a dabbling topic these days. However for Bengali, a significant approach is yet to emerge. Here we developed a prolific model that approximates adequately. We achieved a average accuracy of almost 93% with maximum of 97.98% accuracy. This paper discussed the problem, what is Bengali Document categorization application, what already have done by others in this topic. Then the gist of proposed approach is given. It described out derived approach which is dedicated to describe the in-depth of this classification model and step by step described Chi-Square value of probability distribution and Naïve Bayes classifier. Next section was dedicated to depict different properties of the training data and statistical analysis of data-set. Then it showed the performance analysis of the model. Overall this paper proposed the importance of document categorization for Bengali language and developed an efficient approach for Bengali document categorization.

ACKNOWLEDGMENT

We want to thank research and development (R&D) team of Pipilika search for helping us with statistical data. We are also thankful to "SUST NLP Research Lab" for providing us Bengali stemmer.

REFERENCES

- [1] M. Mansur, N. UzZaman, and M. Khan, "Analysis of N-Gram Based Text Categorization for Bangla in a Newspaper Corpus," in *9th International Conference on Computer and Information Technology*, Dhaka, Bangladesh, 2006.
- [2] A. N. Chy, M. H. Seddiqui, and S. Das, "Bangla News Classification using Naive Bayes classifier," in *16th International Conference on Computer and Information Technology (ICCIT)*, Khulna, Bangladesh, March 2014, pp. 336–371.
- [3] Y. Matsuo and M. Ishizuka, "Keyword Extraction from a Single Document using Word Co-occurrence Statistical Information," *International Journal on Artificial Intelligence Tools*, vol. 13, no. 1, pp. 157–169, 2004.
- [4] K. Sarkar, "An N-Gram Base Method for Bengali Key phrase Extraction," in *International Conference of Information Systems for Indian Languages*, Patiala, India, March 2011, pp. 36–41.
- [5] J. Kaur and V. Gupta, "Effective Approaches For Extraction Of Keywords," *International Journal of Computer Science Issues*, vol. 7, no. 6, November 2010.
- [6] P. D. Turney, "Learning algorithm for keyphrase extraction," *Journal of Information Retrieval*, vol. 2, no. 4, pp. 303–336, 2000.
- [7] D. B. Bracewell, F. Ren, and S. Kuriowa, "Multilingual single document keyword extraction for information retrieval," in *International Conference on Natural Language Processing and Knowledge Engineering*, Wuhan, China, 2005, pp. 517–522.
- [8] Y. Ohsawa, N. E. Benson, and M. Yachida, "Keygraph: Automatic indexing by co-occurrence graph based on building construction metaphor," in *Advances in Digital Libraries Conference*, Washington DC, USA, 98, pp. 12–18.
- [9] S. Ting, W. Ip, and A. H. Tsang, "Is Na ve Bayes a Good Classifier for Document Classification," *International Journal of Software Engineering and its Applications*, vol. 5, no. 3, July 2011.
- [10] D. Lewis, "Naive (bayes) at forty: The independence assumption in information retrieval," in *European Conference on Machine Learning (ECML)*, Chemnitz, Germany, April 1998, pp. 4–15.

Estimation of Polarization Charge in Nitride Based MODFETs Using Differential Threshold Voltage Technique

Muhammad Navid Anjum Aadit¹, Sharadindu Gopal Kirtania², and Md. Kawsar Alam³

Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology
Dhaka-1205, Bangladesh

Email: ¹navidanjumaadit@gmail.com, ²sharadindugopal@gmail.com, and ³kawsaralam@eee.buet.ac.bd

Abstract—We propose a novel method to estimate net polarization charge of nitride-based Modulation-doped Field Effect Transistors (MODFETs) introducing the concept of differential threshold voltage. We simulate MODFET device using an in-house built self-consistent Poisson-Schrödinger solver and calculate net polarization charge present in the device that can be efficiently used for improving device performance. Polarization increases electron charge density and consequently improves device speed and current. In this paper, we show the simulated results of AlGaIn/GaN MODFET as an example and verify the accuracy of our charge calculation procedure by comparing our results with related works. Accurate knowledge of net polarization charge using the proposed estimation technique would provide better control of device characteristics.

Keywords—Charge Density, Differential Threshold Voltage, MODFET, Polarization, Self-consistent

I. INTRODUCTION

Modulation-doped Field Effect Transistors (MODFETs) are popular over traditional FETs for their suitability in high frequency applications [1]. Various designs have been proposed so far for MODFETs using different materials [2]. MODFETs as low noise amplifying devices are widely used in RF [3] and microwave [4] applications as well as in optoelectronics [5], [6], electronics and power applications [7]. For their versatile uses, many researches are still going on to improve MODFET device performance. Nitride-based MODFETs are most widely used because of their high electron density and mobility [8]. In particular, AlGaIn/GaN MODFETs are very much popular now-a-days as they have higher power density and efficiency than AlGaAs/GaAs MODFETs [9]. These high-performance GaN-based devices are also smaller in size than conventional devices. The most common characteristics of AlGaIn/GaN MODFETs are that they have large band gap, high saturation velocity, high breakdown electric field, large conduction-band offset and high thermal stability [10].

A various number of studies on MODFETs have been published showing its modeling techniques and different characteristics. A fully depletion based three-dimensional analytical threshold voltage model for AlGaIn/GaN MODFETs has been reported by Kumar *et al.* [11] and a quasi two-

dimensional study for the shift and variation of threshold voltage of MODFETs has also been available in literature [12]. Derivation of short-channel threshold voltage of MODFETs has also been done by De and Meindl [13]. Improved and uniform threshold voltage of MODFETs has also been discussed by Xu *et al.* [14]. Charge density is the main crucial factor that makes this device so popular. Therefore, varying charge density in AlGaIn layer we can change various characteristics and use them to further improve the performance. Spacer layer and channel charge comparison has been performed for AlGaIn/GaN MODFETs by Cevizovic *et al.* [15]. Spontaneous as well as piezoelectric polarization effects on the device performance of AlGaIn/GaN MODFETs have also been discussed in literature [16]. A two-dimensional charge-control model for MODFETs has been reported by Kim and Roblin [17]. Another charge control model for MODFETs to study subthreshold current has been developed by Tong [18]. Amount of spontaneous and piezoelectric polarization charge at the interface of AlGaIn/GaN heterostructures has been theoretically predicted [19], [20]. These predictions require different rarely known parameters like stiffness coefficients and need to be derived using rigorous first principle calculations. Moreover, influence of device structure like Schottky barrier makes these calculations erroneous [21]. Experimental measurement of polarization charge at AlGaIn/GaN interface has been reported using capacitance-voltage method by Miller *et al.* [22]. Another approach to calculate polarization charge density at the interface of AlGaIn/GaN MODFETs using both current-voltage and capacitance-voltage method has been investigated by Yuan-Jie *et al.* [23]. Both methods need to generate capacitance-voltage profile ensuring polarization charge at other interfaces and surface charges do not affect the calculation. In this paper, we present a novel and simple technique to estimate net polarization charge of nitride based MODFETs using differential threshold voltage technique that eliminates the need of generating capacitance-voltage or current-voltage profile. Moreover, without knowing stiffness coefficients, we can accurately calculate net polarization charge density which plays instrumental role in total charge density, current and speed of MODFET devices.

We demonstrate our results simulating a MODFET device using COMSOL Multiphysics linked with MATLAB. We calculate net polarization charge density at the interface using differential threshold voltage technique for different Al compositions and show that our calculation is comparable with the results obtained from capacitance-voltage method and also with theoretical predictions (prediction from stiffness coefficients). We repeat our simulation for different channel lengths and show that the simulator efficiently calculates net polarization charge density with the variation of device structure.

II. DEVICE STRUCTURE AND MODELING

A. Device Structure

Generally, AlGa_N is grown over Ga_N layer and they have different lattice constants. Both of them have polar nature. So, AlGa_N is under tensile strain and piezoelectric and spontaneous polarization both occur in AlGa_N. This polarization effect aids huge charge density at the AlGa_N/Ga_N interface [16]. As mentioned earlier, depending on the application and required operation, many designs of MODFETs have been proposed so far. We have used a suitable MODFET device for adding polarization charge effects keeping it simple in structure for faster simulation. Schematic of our simulated MODFET device structure including polarization interface is shown in Fig. 1. A thin undoped AlN spacer layer is inserted between the doped AlGa_N layer and undoped Ga_N channel layer to improve carrier confinement and conduction band offset [24]. The AlGa_N layer is intentionally doped to achieve higher sheet charge density with optimum doping concentration, $N_d = 10^{25} \text{ m}^{-3}$ [25]. Gold material is used for drain, gate and source region and sapphire is used for the bulk substrate region.

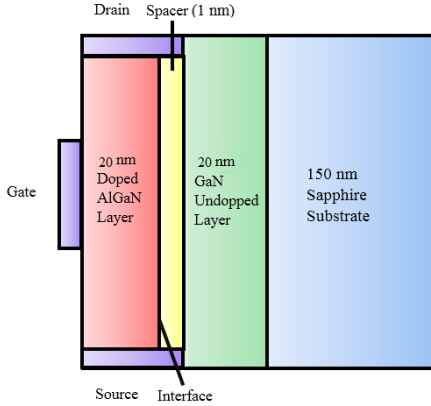


Fig. 1. Proposed schematic model of MODFET showing polarization interface

B. Modeling

In this section, we present an expression for differential threshold voltage and relate it with net polarization charge density. Then we explain our simulation technique with its flowchart.

1) *Differential Threshold Voltage*: Threshold voltage is the minimum required gate voltage to turn the device on. For MODEFTs, without considering polarization charge, it is given by [26],

$$V_{off_{init}} = \phi_B - \frac{\Delta E_c}{q} - \frac{qN_d d_{AlGaN}^2}{2\epsilon_{AlGaN}}, \quad (1)$$

where,

ϕ_B is the Schottky barrier height,

ΔE_c is conduction-band offset at the AlGa_N/Ga_N heterojunction,

N_d is doping concentration of AlGa_N layer,

q is electron charge,

$\epsilon_{AlGaN} = 8.82\epsilon_0$ is dielectric constant of AlGa_N layer,

d_{AlGaN} is thickness of AlGa_N layer,

Polarization expresses the density of permanent or induced dipole moments in dielectric materials. There are two types of polarization in wurtzite group III Nitrides namely piezoelectric and spontaneous polarizations. Since amount of polarization charge is significantly larger than that of doping charge [16], we can assume that the total electric field, $E_{AlGaN}(z)$ in AlGa_N layer is only due to polarization charge. Then the threshold voltage lowered by polarization charge predicted by our assumption is,

$$V_{off_{pol}} = \phi_B - \frac{\Delta E_c}{q} - \frac{qN_d d_{AlGaN}^2}{\epsilon_{AlGaN}} - \int_0^{d_{AlGaN}} E_{AlGaN}(z) dz, \quad (2)$$

To calculate $V_{off_{pol}}$, we can obtain $E_{AlGaN}(z)$ values from self consistent Poisson-Schrödinger solver as discussed in the next section. Now, from Gauss's law [27], net polarization charge density can be calculated as,

$$\sigma = \epsilon_{AlGaN} E_{pol}, \quad (3)$$

where, ϵ_{AlGaN} is a known quantity but polarization electric field, E_{pol} is unknown. Therefore, we need to develop a way to calculate E_{pol} .

Subtracting (2) from (1), we calculate differential threshold voltage as,

$$V_{diff} = V_{off_{init}} - V_{off_{pol}}, \quad (4)$$

$$\Rightarrow V_{diff} = \int_0^{d_{AlGaN}} E_{AlGaN}(z) dz, \quad (5)$$

If we assume that, electric field, $E_{AlGaN}(z)$ is uniform through the AlGa_N layer, then we can write,

$$V_{diff} = E_{AlGaN} \int_0^{d_{AlGaN}} dz, \quad (6)$$

$$\Rightarrow V_{diff} = E_{AlGaN} d_{AlGaN}, \quad (7)$$

We have already assumed that, E_{AlGaN} is mostly contributed by polarization charge. So we can write that,

$$E_{pol} = E_{AlGaN}, \quad (8)$$

$$\Rightarrow V_{diff} = E_{pol} d_{AlGaN}, \quad (9)$$

Although we can obtain $E_{AlGaN}(z)$ from Poisson-Schrödinger solver, we still do not know the uniform value E_{AlGaN} . Thus, E_{pol} is still unknown. To find E_{pol} , we differentiate both sides of (9) with respect to d_{AlGaN} ,

$$\frac{d(V_{diff})}{d(d_{AlGaN})} = E_{pol}, \quad (10)$$

Thus, if we know V_{diff} for different values of d_{AlGaN} , we can calculate E_{pol} . Once E_{pol} is known, we can calculate σ from (3).

2) Simulation Technique:

Step 1: We first solve 3D Poisson- 2D Schrödinger equations for the designed MODFET device self-consistently at zero gate bias, $V_{G1} = 0$ V as described in [28]. We use only doping charge density, $N_d = 10^{25} \text{ m}^{-3}$ and do not include any fixed polarization charge at the interface of the device. We obtain energy band diagram of the device after convergence of self-consistent solution [28]. The conduction band offset of GaN layer is enhanced by the AlN spacer layer to improve carrier confinement and its minimum point approximately indicates the threshold voltage of this device as shown in Fig. 2. We denote this voltage as $V_{off1} = V_{off_{init}}$.

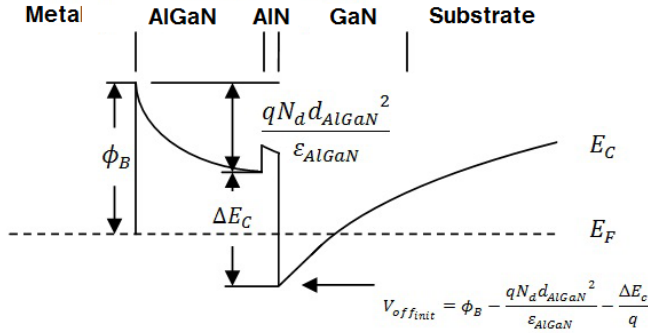


Fig. 2. Representation of threshold voltage by conduction band minimum of GaN layer in energy band diagram

Step 2: We obtain $E_{AlGaIn}(z)$ after convergence of self-consistent solution and update threshold voltage using (2). We denote this voltage as V_{off2} . However, we note that this $E_{AlGaIn}(z)$ is not accurate as it is calculated without accurate information of polarization charge, so it should be updated by iterations.

Step 3: Lowering threshold voltage by $\Delta V_{off} = V_{off1} - V_{off2}$ means the conduction band minimum in the energy band diagram is lowered by the same amount. This is the same effect if the Fermi level, E_F is lowered by ΔV_{off} . This effect is further analogous to raising gate voltage by some equivalent amount. We assume that this equivalent amount equals to ΔV_{off} . So for next iteration, we set gate voltage as, $V_{G2} = V_{G1} + \Delta V_{off}$.

Step 4: We again solve 3D Poisson- 2D Schrödinger equations for the designed MODFET device self-consistently at gate bias, V_{G2} . We still do not include any fixed polarization charge at the interface of the device to be simulated, but we incorporate the equivalent effect by applying positive gate voltage. We obtain another energy band diagram of the device after convergence of self-consistent solution and obtain the threshold voltage of this device from the conduction band minimum of GaN layer. We again denote this voltage as V_{off1} . It is to be noted that raising gate voltage by ΔV_{off} does not necessarily lower the threshold voltage by exactly ΔV_{off} as we consider self-consistent solution.

Step 5: We check if $|V_{off1} - V_{off2}| < \text{preset tolerance}$ limit is satisfied. If so, we continue to step 6; otherwise we return to step 2.

Step 6: We assign $V_{off_{pol}} = V_{off1}$ and find V_{diff} using (4). We repeat our simulation for different values of doped layer thickness, d_{AlGaIn} simply by changing this parameter in COMSOL and find V_{diff} for different values of d_{AlGaIn} .

Step 7: We plot V_{diff} vs. doped layer thickness d_{AlGaIn} in MATLAB. The curve should be a straight line and its slope gives E_{pol} according to (10). Then we can easily calculate net polarization charge density, σ using (3).

Our proposed simulation technique is shown using the flow chart given in Fig. 3.

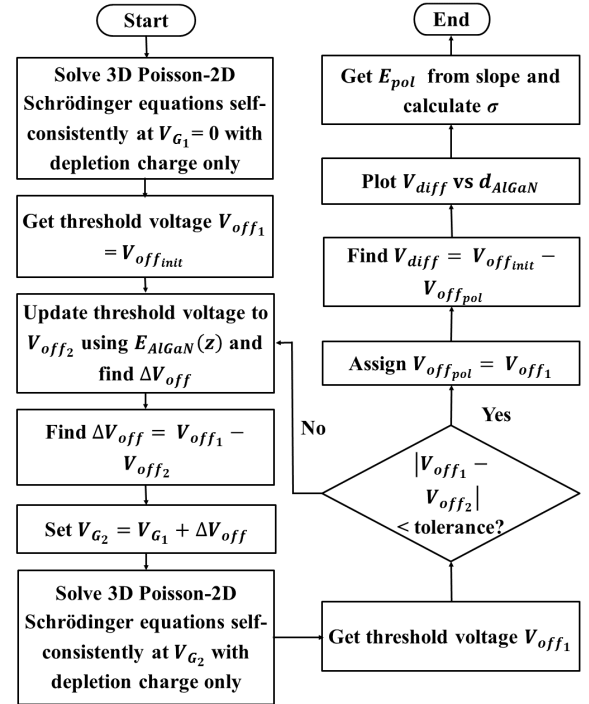


Fig. 3. Flow chart of proposed simulation technique and differential threshold voltage method

III. RESULTS AND DISCUSSIONS

At first, we run self consistent Schrödinger-Poisson solver to obtain energy band diagram of MODFET assuming doped AlGaIn layer thickness, $d_{AlGaIn} = 20$ nm and Al composition in AlGaIn layer, $x = 0.2$ at gate bias, $V_{G_1} = 0$ V. Initially, we find $V_{off_{init}}$ (about 0.5 V in the figure) from the conduction band minimum of GaN layer in the diagram. Finally, after convergence of iterations, we find $V_{off_{pol}}$ (about -3 V in the figure) from the updated diagram at gate bias, V_{G_2} . We superimpose the energy band diagrams of these two cases in Fig. 4. This figure shows electron energy values, E (eV) taking cut line along z axis at device depth, t (nm). For $V_{off_{init}}$, the polarization electric field due to polarization charge is not considered. Absence of polarization charge in the doped AlGaIn layer means electrostatic potential in this layer is relatively lower. Lower electrostatic potential means higher conduction band minimum in the energy band diagram. On the other hand, for $V_{off_{pol}}$, the polarization electric field effect is included in the simulation by applying equivalent positive gate voltage, V_{G_2} (about 2.8 V in the figure). Positive gate voltage plays similar role as polarization charge and as a result, electrostatic potential in doped AlGaIn layer is relatively higher for this case causing conduction band minimum to be lowered. As conduction band minimum represents threshold voltage of MODFET device, $V_{off_{init}}$ is greater than $V_{off_{pol}}$. It is to be noted that, for both cases AlN spacer layer enhances the conduction band offset to improve carrier confinement between AlGaIn and GaN layers. As the material of the spacer layer is AlN with only 1 nm thickness, its polarization effect is considered to be lumped with that of the AlGaIn layer.

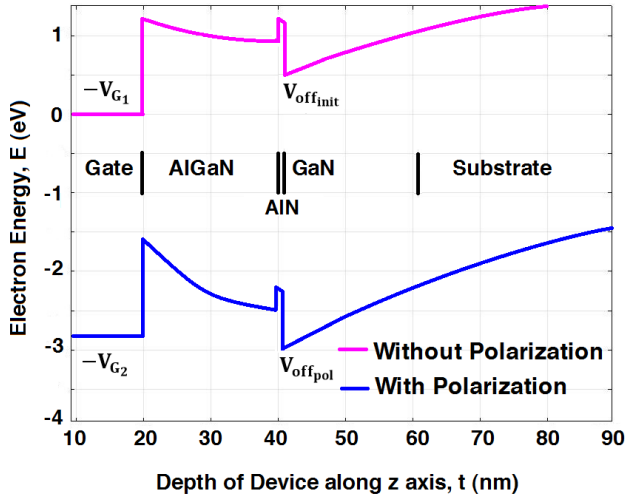


Fig. 4. Superimposed energy band diagrams of MODFET after self-consistent solution, assuming doped layer thickness, $d_{AlGaIn} = 20$ nm and Al composition, $x = 0.2$

We have also investigated $V_{off_{init}}$ and $V_{off_{pol}}$ vs. doped layer thickness, d_{AlGaIn} simulating the device for different doped layer thicknesses which is shown in Fig. 5. From the figure, we find that threshold voltage is lowered when polarization effect is included and the difference between initial and polarized threshold voltages vary with doped layer

thickness. Increasing thickness of the doped AlGaIn layer includes more polarization charge at the interface and so threshold voltage of the device is lowered more. From (9), it is also obvious that, increasing doped AlGaIn layer thickness reduces threshold voltage linearly.

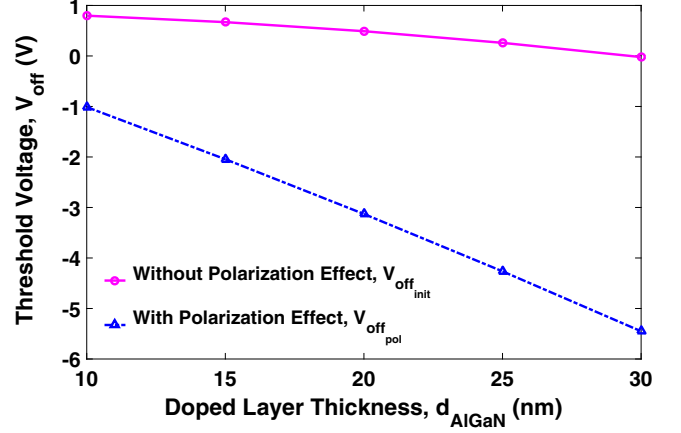


Fig. 5. Variation of threshold voltage with doped layer thickness (legend: circle refers to solution without polarization effect and triangle refers to solution including polarization effect).

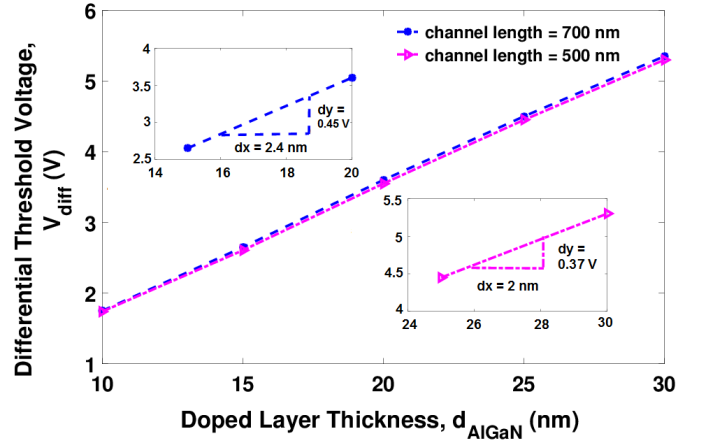


Fig. 6. Linear relationship between differential threshold voltage and doped layer thickness, calculations for slope $= \frac{dy}{dx} = E_{pol}$ are shown in insets for two different channel lengths.

Now, we present the differential threshold voltage, V_{diff} vs. doped layer thickness, d_{AlGaIn} and find that the relation is linear. This is shown in Fig. 6 for two different channel lengths. As polarization charge does not depend on channel length, these two curves almost overlap with each other. From slope of such linear curve (shown as $\frac{dy}{dx}$ in insets of Fig. 6), we calculate net polarization charge, σ simply by multiplying the slope by permittivity constant, ϵ_{AlGaIn} according to (3). We compare our result with theoretical predictions [19] and also with the result by Miller *et al.* [29]. The results of net polarization charge calculation are summarized in Table I where Al composition in AlGaIn layer, x is assumed to be

TABLE I
NET POLARIZATION CHARGE CALCULATION

Channel Length (nm)	dy (V)	dx (nm)	slope = $\frac{dy}{dx}$ = E_{pol} (V/nm)	Estimation of σ using (3) (cm^{-2})	Theoretical prediction of σ from [19] (cm^{-2})	Measured Value of σ by Miller <i>et al.</i> [29] (cm^{-2})
500	0.37	2.0	0.1850	0.9029611×10^{13}	0.81325×10^{13}	0.98×10^{13}
700	0.45	2.4	0.1875	0.9151633×10^{13}	0.81325×10^{13}	0.98×10^{13}

20%. As channel length variation is not shown in literature, we simply assume same values for comparison at different channel lengths in cases of related works.

Finally, we calculate net polarization charge density σ with variation of Al composition, x as shown in Fig. 7. Values of σ calculated using differential threshold voltage technique fit midway between those obtained by theoretical predictions [19] and by Miller *et al.* [29] that shows the efficacy of the proposed technique. We have calculated σ for 6 different x compositions and assumed $\sigma = 0$ for $x = 0$. The resultant curve is interpolated so that values of σ at any practical composition can be estimated. Thus, differential threshold voltage technique can estimate net polarization charge density effectively that can be helpful to study MODFET device characteristics.

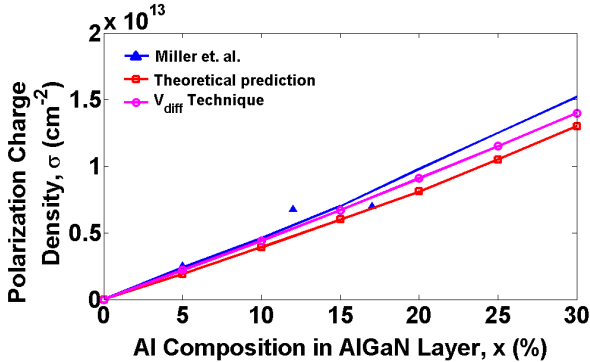


Fig. 7. Net polarization charge density σ with variation of Al composition, x (legend indicates the method used)

IV. CONCLUSION

We have presented polarization effect at the interface of AlGaIn/GaN MODFETs and shown variation of threshold voltage with it. We have also shown how polarization effect depends on doped layer thickness. We have discussed a new method to calculate net polarization charge density present in the device that does not require any capacitance-voltage profile or rarely available stiffness coefficients. Finally, we have calculated net polarization charge density for different Al compositions. The result is comparable with both theoretical predictions and related works and it defines the acceptability of our proposed technique. This method can also be applied to other MODFETs by changing layer materials in the simulator

schematic model. Thus, this proposition of net polarization charge estimation technique will be helpful to study MODFET device behavior and improve different characteristics.

REFERENCES

- [1] S. Sen, M. Pandey, and R. Gupta, "Two-dimensional C-V model of AlGaAs/GaAs modulation doped field effect transistor (MODFET) for high frequency applications," *Electron Devices, IEEE Transactions on*, vol. 46, no. 9, pp. 1818–1823, Sep 1999.
- [2] A. Fathimulla, D. Gutierrez, and H. Hier, "A novel insulated-gate InP/InAlAs MODFET," in *Indium Phosphide and Related Materials, 1993. Conference Proceedings., Fifth International Conference on*, Apr 1993, pp. 428–431.
- [3] H.-C. Yuan, N. Jiang, E. Croke, and Z. Ma, "Si-based multi-finger n-MODFET RF power devices," in *Silicon Monolithic Integrated Circuits in RF Systems, 2006. Digest of Papers. 2006 Topical Meeting on*, Jan 2006.
- [4] P. Roblin, S. Kang, A. Ketterson, and H. Morkoc, "Analysis of MODFET microwave characteristics," *Electron Devices, IEEE Transactions on*, vol. 34, no. 9, pp. 1919–1928, Sep 1987.
- [5] A. Ketterson, J. Seo, M. Tong, K. Nummala, J. Morikuni, K.-Y. Cheng, S.-M. Kang, and I. Adesida, "A MODFET-based optoelectronic integrated circuit receiver for optical interconnects," *Electron Devices, IEEE Transactions on*, vol. 40, no. 8, pp. 1406–1416, Aug 1993.
- [6] L. Adesida, A. Ketterson, J.-W. Seo, M. Tong, K. Nummala, D. Ballegeer, K. Cheng, and S. Kang, "A pseudomorphic AlGaAs/InGaAs/GaAs MODFET-based optoelectronic receiver with a 10 GHz bandwidth," in *Broadband Analog and Digital Optoelectronics, Optical Multiple Access Networks, Integrated Optoelectronics, Smart Pixels, LEOS 1992 Summer Topical Meeting Digest on*, July 1992, pp. B29–B30.
- [7] H. Huq, M. Alam, and S. Islam, "Analysis of temperature model on device characteristics for AlGaIn/GaN MODFET for high power electronics," in *Semiconductor Device Research Symposium, 2005 International*, Dec 2005, pp. 250–251.
- [8] P. Romanini, M. Peroni, C. Lanzieri, A. Cetrionio, M. Calori, A. Passaseo, B. Poti, A. Chini, L. Mariucci, A. Di Gaspare, V. Teppati, and V. Camarchia, "Very high performance GaN HEMT devices by optimized buffer and field plate technology," in *European Microwave Integrated Circuits Conference, 2006. The 1st*, Sept 2006, pp. 61–64.
- [9] W. Lu, J. Yang, M. Khan, and I. Adesida, "Wideband AlGaIn-GaN HEMTs on SiC for low noise applications," in *Device Research Conference*, June 2000, pp. 39–40.
- [10] J. Shealy, "Progress in Si-based AlGaIn HEMTs for RF power amplifiers," in *Silicon Monolithic Integrated Circuits in RF Systems*, Sept 2001, pp. 166–169.
- [11] S. Kumar, A. Agrawal, R. Chaujar, S. Kabra, M. Gupta, and R. Gupta, "3-dimensional analytical modeling and simulation of fully depleted AlGaIn/GaN modulation doped field effect transistor," in *Physics of Semiconductor Devices, 2007. IWPSD 2007. International Workshop on*, Dec 2007, pp. 373–376.
- [12] S. Sen, M. Pandey, M. Khanna, and R. Gupta, "A quasi two-dimensional analytical model for threshold voltage of a modulation doped field effect transistor," in *Microwave Conference Proceedings, 1997. APMC '97, 1997 Asia-Pacific*, Dec 1997, pp. 705–708 vol.2.
- [13] V. De and J. Meindl, "An analytical threshold voltage and subthreshold current model for short-channel AlGaAs/GaAs MODFETs," *Solid-State Circuits, IEEE Journal of*, vol. 28, no. 2, pp. 180–183, Feb 1993.

- [14] D. Xu, T. Enoki, T. Suemitsu, Y. Umeda, Y. Yamane, and Y. Ishii, "Improving threshold-voltage uniformity of 0.1 μm InP-based MODFETs with different gate layouts," *Electronics Letters*, vol. 34, no. 16, pp. 1614–1615, Aug 1998.
- [15] D. Cevizovic, R. Sasic, and R. Ramovic, "Sheet density of electrons in spacer layer of AlGaIn/GaN MODFET: calculating and analysing," in *Microelectronics, 2002. MIEL 2002. 23rd International Conference on*, vol. 2, 2002, pp. 475–478.
- [16] F. Sacconi, A. Di Carlo, P. Lugli, and H. Morkoc, "Spontaneous and piezoelectric polarization effects on the output characteristics of AlGaIn/GaN heterojunction modulation doped FETs," *Electron Devices, IEEE Transactions on*, vol. 48, no. 3, pp. 450–457, Mar 2001.
- [17] Y. M. Kim and P. Roblin, "Two-dimensional charge-control model for MODFETs," *Electron Devices, IEEE Transactions on*, vol. 33, no. 11, pp. 1644–1651, Nov 1986.
- [18] K. Tong, "Accurate charge control model for MODFET including subthreshold region," *Electronics Letters*, vol. 27, no. 8, pp. 668–669, April 1991.
- [19] H. Morkoc, *Field Effect Transistors and Heterojunction Bipolar Transistors*. Wiley-VCH Verlag GmbH and Co. KGaA, 2009, pp. 349–707.
- [20] F. Bernardini, V. Fiorentini, and D. Vanderbilt, "Accurate calculation of polarization-related quantities in semiconductors," *Phys. Rev. B*, vol. 63, p. 193201, Apr 2001.
- [21] L. Yuan-Jie, F. Zhi-Hong, G. Guo-Dong, Y. Jia-Yun, F. Yu-Long, W. Yuan-Gang, T. Xin, Z. Xing-Ye, L. Zhao-Jun, J. Zi-Wu, and C. Shu-Jun, "Influence of the AlGaIn barrier thickness on polarization Coulomb field scattering in an AlGaIn/AlN/GaN heterostructure field-effect transistor," *Chinese Physics B*, vol. 24, no. 8, p. 087306, 2015.
- [22] E. Miller, E. Yu, C. Poblenz, C. Elsass, and J. Speck, "Direct measurement of the polarization charge in AlGaIn/GaN heterostructures using capacitance-voltage carrier profiling," *Applied Physics Letters*, vol. 80, no. 19, pp. 3551–3553, 2002.
- [23] L. Yuan-Jie, L. Zhao-Jun, Y. Ying-Xia, M. Ling-Guo, C. Zhi-Fang, L. Chong-Biao, and W. Zhan-Guo, "A simple method of extracting the polarization charge density in the AlGaIn/GaN heterostructure from current-voltage and capacitance-voltage characteristics," *Chinese Physics B*, vol. 21, no. 9, p. 097104, 2012.
- [24] M. Wosko, B. Paszkiewicz, R. Paszkiewicz, and M. Tlaczala, "Influence of AlN spacer on the properties of AlGaIn/AlN/GaN heterostructures," *Optica Applicata*, vol. 43, no. 1, 2013.
- [25] O. Ambacher, B. Foutz, J. Smart, J. R. Shealy, N. G. Weimann, K. Chu, M. Murphy, A. J. Sierakowski, W. J. Schaff, L. F. Eastman, R. Dimitrov, A. Mitchell, and M. Stutzmann, "Two dimensional electron gases induced by spontaneous and piezoelectric polarization in undoped and doped AlGaIn/GaN heterostructures," *Journal of Applied Physics*, vol. 87, no. 1, pp. 334–344, 2000.
- [26] M. Z. M. Charfeddine, H. Belmabrouk and H. Maaref, "2-D theoretical model for current-voltage characteristics in AlGaIn/GaN HEMTs," vol. 3, no. 8, 2012, pp. 881–886.
- [27] Wikipedia. (2016) Gauss's law. [Online]. Available: https://en.wikipedia.org/wiki/Gauss%27s_law
- [28] J. Wang, E. Polizzi, and M. 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.
- [29] H. Zhang, E. Miller, E. Yu, C. Poblenz, and J. Speck, "Measurement of polarization charge and conduction-band offset at InGaIn- xN/GaN heterojunction interfaces," *Applied Physics Letters*, vol. 84, no. 23, pp. 4644–4646, 2004.

Switching Optimisation in Huffman Code for Power Efficient Data Transmission

Sohag Kabir*, Youcef Gheraibia[†] and Tanzima Azad[‡]

*Department of Computer Science, University of Hull, UK, E: shgkabar@gmail.com

[†]Department of Computer Science, University of Mohammed Cherif Messaadia, Algeria, E: youcef.gheraibia@gmail.com

[‡]Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology, Bangladesh
E: azad.tanzima@dbbl.com.bd

Abstract—Different technologies have been emerged to address the issues of power consumption in digital communication. In CMOS technology, dynamic power accounts for 70%-90% of the total power dissipation and it largely depends on the representation of data and increases linearly with the switching activities (transition from logic level High to Low and vice versa). An efficient representation of data can minimise power consumption by reducing switching activities. In this paper, we have proposed an approach using genetic algorithm to optimise switching activities in the Huffman code for biological data compression. The performance of the approach has been evaluated by applying it to a set of biological datasets. The experiments yield that the proposed approach reduces the switching activity by 45.47% in the best case, by 36.45% in the average case, and by 16.42% in the worst case.

I. INTRODUCTION

In recent years, application of battery-powered portable devices, e.g., laptop computers, personal digital assistant (PDA), and mobile phones has increased significantly. The reliability and performance of such devices are primarily dependent on their power consumption, i.e., effective battery life. Bigger battery size could help to improve power efficiency, however, in a portable device the size of the battery is restricted by the size and weight of the device itself. The cost of providing power to both portable and non-portable devices has resulted in significant interest in power consumption reduction. CMOS technologies were developed in order to reduce the power consumption both in data processing and transmission. In [1], an approach was proposed to minimise power consumption in CMOS circuits. The method considered optimising the technology used to implement the digital circuits, circuits style and topology, the architecture for implementing the circuits and the algorithms that are being implemented in the devices.

As data transmission is a common operation in digital communication, reducing power dissipation in transmission systems is important. Authors in [2], [3] have investigated different criteria for low power data transmission. Power consumption for data transmission in IEEE 802.11 multi-hop networks and 3G mobile wireless networks were analysed in [4] and [5] respectively. In order to increase transmission speed and reduce power dissipation, parallel data transmission methods are widely used. However, parallel transmission is limited to short distance communications, e.g. locally connected devices, internal buses. Ruling out the possible availability of

parallel transmission links over long distance, we are left with its serial alternative only. If we attempt to transfer big files (e.g. DNA sequence), over a serial transmission link then it would take a significant amount of time and obviously a significant amount of power would be dissipated in the process. However, data compression techniques are widely used to reduce size of data before transmitting over serial medium to reduce transmission time and cost.

Huffman code [6] is an entropy encoding algorithm widely used for lossless data compression and its applications are pervasive throughout computer science [7], [8]. For any given set of symbols and associated occurrence probabilities, Huffman algorithm generates a binary tree (Huffman tree) to encode a message with an optimal number of bits. The codeword for a distinct symbol is generated by traversing the tree from the root to the leaf node representing the symbol. In this traversal process, a move towards a left and a right node is represented by 0 and 1 respectively. A major drawback of the classical Huffman code is that the whole stream must be read prior to encoding.

Power dissipation to transmit a binary string is proportional to the number of switching, i.e., transition from 0 to 1 and vice versa, present in the transmitted string. Although a number of approaches like [9], [10] have extended the classical Huffman code by treating binary bits differently (with unequal letter cost), a little effort has been made to optimise the switching activities in the Huffman code to develop a power efficient Huffman code. To the knowledge of the authors, the only effort to generate Huffman code considering switching activities is seen in [11]. This approach follows a greedy strategy and works in two steps to reduce total number of switching in the Huffman code. In the first step, the switching activity between each pair of symbol (inter-switch) is reduced. In the second step, switching activity in the codeword for each symbol (intra-switch) is reduced. As the switching optimisation is done independently in two different steps, it is highly likely that the gain obtained in one step may be compromised in other step.

In this paper, a genetic algorithm based approach is proposed to reduce the total number of switching in the Huffman code where inter-switch and intra-switch are optimised in a single step, i.e, a balance is made between two different types of switching. The approach generates a set of Huffman trees,

each tree represents a set of codewords (solution). After that, a set of genetic operators such as selection, crossover and mutation are used to evolve (create new trees) the population to minimise number of total switches. The evolution process is controlled by the total number of inter- and intra-switches. Inter-switch between two adjacent symbols depend on their codewords, i.e., how the codeword of the preceding symbol finished (with 0 or 1) and how the codeword for the descendent symbol started. If the end bit of the preceding symbol and the starting bit of the successor symbol differs then there exists a switch and the total number of switch is the number of times the two symbols in consideration occur one after another. To obtain the total number inter-switch between pair of symbols, we used an adjacency matrix which is a two dimensional matrix represents the number of occurrence of each symbol after another. This matrix is formed when the message to be transmitted is scanned to obtain the frequency of distinct symbols. The number of intra-switch for a symbol is calculated by multiplying the frequency of that symbol by the number of logic level transition (0 to 1 or 1 to 0) in the codeword of that symbol. The efficiency of the approach is evaluated by applying it to a set of biological datasets. The experiments yield that the proposed approach reduces the total number of switches by 45.47% in the best case, by 36.45% in the average case and by 16.42% in the worst case.

II. BACKGROUND

A. Effect of Switching in Power Consumption

CMOS is the most widely used technology implemented in VLSI chips because of their power efficiency. Power dissipation in CMOS circuits has three different components: dynamic, static, and leakage [12].

$$P_{avg} = P_{dynamic} + P_{static} + P_{leakage}$$

$$= 0.5 \times C_L \times V_{dd}^2 \times E(sw) \times f_{clk} + I_{sc} \times V_{dd} + I_{leakage} \times V_{dd} \quad (1)$$

where

P_{avg} is the average power dissipation

C_L is the load capacitance,

V_{dd} is power supply voltage,

$E(sw)$ is average transition (switching) number per clock cycle,

f_{clk} is the clock frequency,

I_{sc} is the short circuit current, and

$I_{leakage}$ is the leakage current.

Out of the three components, the dynamic power dissipation is the dominant and it counts for 70%-90% of the total power consumption [13]. For this reason, optimisation of dynamic power consumption is considered in this paper. In CMOS circuit, dynamic power consumption arises when the load capacitance (C_L) is charged to address a transition from the low to the high voltage level. The parameters C_L , V_{dd} , and f_{clk} in dynamic power dissipation are primarily determined by the circuit layout and the fabrication technology. However, amount of switching, $E(sw)$ completely depends on the digital representation of the data. As the dynamic power dissipation is

proportional to the average number of switching, a reduction in the total switching number will eventually results in the reduction of the dynamic power consumption. Therefore, we focus on reducing total number of switching in Huffman code to improve power efficiency of digital data compression and communication.

B. Huffman Code and Its Relationship with Switching Activity

In computer science and information theory, Huffman code is an entropy encoding algorithm used for lossless data compression. The classical Huffman code requires to pass the input data two times to complete the encoding operation. In the first pass, it reads the whole stream to determine the occurrence frequencies of distinct symbols. After that, symbols along with their frequencies are used to generate a binary tree known as Huffman tree. The process of Huffman tree generation is shown in algorithm 1 [14].

Algorithm 1 Huffman(C)

```

1:  $n \leftarrow |C|$ 
2:  $Q \leftarrow C$ 
3: for  $i = 1$  to  $n - 1$  do
4:   allocate a new node  $z$ 
5:    $left[z] \leftarrow x \leftarrow EXTRACT\_MIN(Q)$ 
6:    $right[z] \leftarrow y \leftarrow EXTRACT\_MIN(Q)$ 
7:    $f[z] \leftarrow f[x] + f[y]$ 
8:    $INSERT(Q, z)$ 
9: end for
10: return  $EXTRACT\_MIN(Q)$ 
```

In the algorithm, C is the set of n distinct symbols, and each symbol $c \in C$ has a frequency denoted as $f[c]$. The algorithm starts its operation with a set of $|C|$ leaf nodes where each node corresponds to a distinct symbol. After that, a sequence of $|C| - 1$ merging operation is performed to obtain the final tree. The merging is done by extracting two least frequent symbols from the minimum priority queue, Q , and creating and inserting a new node (parent) to the queue for that two symbols. These two symbols are considered as the left and right child of their parent node respectively. The frequency of the new node is calculated as the sum of the frequency of its child nodes. The resultant tree ensures the optimal encoding of symbols to obtain maximal compression performance. To form a codeword for a symbol, the tree is traversed from the root to the leaf node representing the symbol. In the traversal process, any move made towards the left direction is replaced by '0' and any move towards the right direction is replaced by '1'. As a result, codeword for each symbol consist of a sequence of 0's and/or 1's. Once the codewords are obtained, in the second pass, the Huffman algorithm encodes the symbols of the input message with the codewords, thus reduce the size of the whole message. For a set of symbols more than one Huffman tree is possible, i.e., optimal solution can be obtained from different structure of the tree. As the codeword of a symbol depends on its position in the tree, changing the structure of the tree will change the codeword for

the symbol. A codeword can contain zero or more switching activity within itself referred to as intra-switch. For example, if a symbol 'A' is assigned a codeword '000' then there exists no intra-switch, on the other hand, if the codeword '101' is assigned to a symbol 'B' then there exists two intra-switches. When the codewords of different symbols are concatenated together during transmission they form switching based on the head and tail bit of the codewords. For example, if we want to transmit 'AB' then according to the above codewords we will transmit '000101' which will result in an inter-switch because the tail bit (0) of 'A' and the head bit (1) of 'B' are different. That means, in this case, we have three total switches – two intra-switches and one inter-switch. Now if 'B' is assigned the codeword '010', the new transmitted string is '000010'. In this case, we still have two intra-switches for the codeword of 'B' but the inter-switch no more exists. So, changing tree structure, i.e., codewords, will change the total number of switching. Hence, the aim of this paper is to find an optimal Huffman tree that will minimise the total number of switching by making a balance between intra- and inter-switches.

C. Issues in Biological data transmission

The size of biological databases are continuously increasing. In the last decade the access to these databases has increased significantly. The exponential growth of these databases pose a big problem to all biological data processing methods. Data compression methods are widely used to reduce the overheads of treating, e.g., transmitting, the high volume of data. The main objective of data compression methods is minimising the number of bits in the data representation. Authors in [15] proposed a new approach to genomic data compression based in suffix and common substring and code the difference with Huffman code. In [16], authors propose a general scheme based on Huffman code for coding only the difference between the target genome and the referential one. Recently, a method has been proposed to deal with the transmission of genomic data as an email attachment [17]. All the methods used for biological data compression aims only at reducing the size of the data. However, due to the huge size of the input data, the compressed data contains a huge number of switching in them which has a big impact on power consumption. As no consideration is paid to the reduction of switching, the compression process may be size efficient but not power efficient. Therefore, it is believed that a more efficient compression of biological data can be obtained by optimising the size of the data as well as the total number of switching in the transmitted data.

III. THE PROPOSED APPROACH FOR SWITCHING OPTIMISATION

The main objective of the proposed approach is to minimise the total number of switching in the representation of data produced by the Huffman algorithm. To do so, it is required to minimise both intra- and inter-switches among the codewords. The total number of intra-switch depends on the codewords

themselves and the frequency of the symbols. As a result, for a given set of codewords an optimal number of intra-switches can be obtained by assigning the codewords with minimum number of switching to the most frequent symbol and so on. However, it will not ensure the optimal number of inter-switches. Because, inter-switches not only depend on the codewords of the symbols but also the association between the symbols. By association, we mean how many times a symbol occurs after another symbol in the whole message, i.e., all possible pairing combination among the symbols. As the number of distinct symbols increases in the message, the number of combination increases. So improving total number of inter-switching is a combinatorial problem.

Genetic algorithm (GA) has widely been applied to solve a range of combinatorial problem in complex VLSI design [18], [19]. For this reason, GA is identified as a potential solution to the above mentioned problem. GA is one of the most popular bio-inspired meta-heuristic algorithm inspired by the natural evolution of species [20]. It is a population based algorithm starts with the generation of a random initial population. The population contain a set of feasible solutions called individuals, that are usually far from the optimal. The GA optimisation process uses a set of natural genetic operators such as selection, crossover and mutation to converge to the optimal solution. The proposed approach is shown in algorithm 2.

Algorithm 2 Switches optimising Huffman codes

Require: Textual representation of a Genome Dataset

Ensure: Huffman codes with optimal number of switching

- 1: Calculate symbol frequencies and generate adjacency matrix
 - 2: Generate the initial population of trees
 - 3: **repeat**
 - 4: Select part of the population
 - 5: Generate children candidates via crossover
 - 6: Mutate children
 - 7: Add children to population
 - 8: Rank by fitness
 - 9: Remove worst solutions until population limit reached
 - 10: **until** the switching number stop decreasing
-

A. Objective function

The main objective of the proposed approach is to generate a set of codewords for the input symbols to minimise the overall number of switching. The objective function of the switching optimisation algorithm can be formulated as follows:

Definition: Let $F = \{F_1, F_2, \dots, F_n\}$ and $C = \{C_1, C_2, \dots, C_n\}$ be the set of frequencies and codewords for the input symbols $S = \{S_1, S_2, \dots, S_n\}$ respectively. $M(n, n)$ be the adjacency matrix where each entry $M[i][j]$ represents how many times the input symbol S_j appears after the symbol

S_i in the whole sequence.
The Objective Function is to:

$$\text{Minimise } \left(\sum_{i=1}^n (Sw(C_i) \times F_i) + \sum_{i=1}^n \sum_{j=1}^n (\mu(C_i, C_j) \times M[i][j]) \right) \quad (2)$$

where,

$Sw(C_i)$ is the number of intra-switches in the codewords C_i
 $\mu(C_i, C_j) = 1$ if the head (starting) bit of the codeword C_j and the tail (ending) bit of the codeword C_i are different.
 $\mu(C_i, C_j) = 0$ if the head bit of the codeword C_j and the tail bit of the codeword C_i are same.

B. Genetic operations

1) *Selection*: The first operator of the genetic algorithm is the selection. The main objective of the selection is to choose the part of the current population to be a candidate for the different genetic operators in order to breed the next generation of population. Many selection techniques have been proposed in the literature [21]. In this approach the process of natural selection is maintained. At first, we generate a random even number *rand* between 2 and the population size. This number represents how many parents will be processed. After that, an unbiased random selection of *rand* individuals (solutions) from the population is made.

2) *Crossover*: The main operator of the GA is the crossover which allows to construct new solutions from the selected part of the population [22]. The selected solutions are ranked by fitness and sorted in descending order. After that, at a time, a pair of solution is taken from the sorted population and then crossed to generate two new solutions. This process continues until all the pairs of solutions are crossed. The main objective of the crossover is to benefit from two solutions to generate better solutions. Two parent Huffman trees are crossed to find two new child trees in the following way:

Step 1: Select an internal node (a sub-tree) from each parent tree with equal number of nodes (the dashed selection in Fig. 1(a) and (b)).

Step 2: Put the selected sub-tree of the first tree in the place of the selected sub-tree of the second tree and vice versa (see Fig. 1(c) and (d)).

In each of the trees, some symbols are repeated and some are missed, but number of repeated and missed symbols are equal. This is because, we replace a sub-tree by other sub-tree without any modification. As a result each new (child) tree contains a part from its parent tree and a part from the other tree. If the remaining part from the parent tree and the newly added part has one or more symbols in common then these symbols will be repeated in the tree. On the other hand, the uncommon symbols between the newly added sub-tree and the replaced sub-tree will be missed. We solved this conflict by putting the missed symbols one by one in the place of the

repeated symbols in the remaining part from the parent tree. That means, the newly added sub-tree is kept unchanged and all the replacements are done on the old part of the tree.

3) *Mutation*: The mutation operator changes the positions of two leaf nodes of the generated children (result of the crossover) [23]. It introduces the diversity in the search process. This diversification strategy allows the algorithm to converge to the optimal solution. The algorithm for this operator goes through the leaf nodes and change the positions of two leaf nodes from their current positions to another positions in the tree. The selection of the nodes to be mutated is done following a fixed mutation rate.

C. Population update

The crossover operation generates new solutions from the current solutions to build the second generation of the population. Furthermore, the mutation process provides the diversity in the search on the solution space by changing the position of the nodes in the same tree. After these two operations, the next step of the genetic algorithm is to breed the new population. Firstly, the new solutions (children) are added to the current population and ranked by fitness. The algorithm then removes the worst solutions until the initial size of the population is achieved. The genetic algorithm repeats the selection, crossover, mutation, and population update operations until the stopping criteria is achieved. In the proposed approach the stopping criteria is when the objective function (number of switches) stop decreasing.

IV. EXPERIMENTS AND COMPARISON

The proposed approach is used to reduce the total number of switching activities during data transmission. The approach optimises the total number of switching by finding the best arrangement of nodes by rearranging their positions in the codewords tree with the best allocation of these codewords to the different frequencies. It continues swapping between the individual population until a balance is found between inter-switching and intra-switching. The effectiveness of the approach has been evaluated by applying is to compress different real genomic biological data. Different genomes size have been used for the evaluation, from small genome to large human chromosome. The datasets were downloaded as FASTA text file from a recent version of The National Center for Biotechnology Information (NCBI) available on ([http : //www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) [24]. The used genomes are described in table 1 with their reference in the NCBI database, the size of all the files are given in Megabytes (MBs).

In the experiments, we compare the performance of the classical Huffman code with the performance of the proposed method in terms of the total number of switching. The size of the initial population is fixed to 12 trees and the mutation rate is set to 2%. These values are obtained by varying the two parameters and after these two values the approach continuing without any improvements. Table 2 presents the results obtained by the proposed approach and the classical Huffman code. The results are presented as the total number

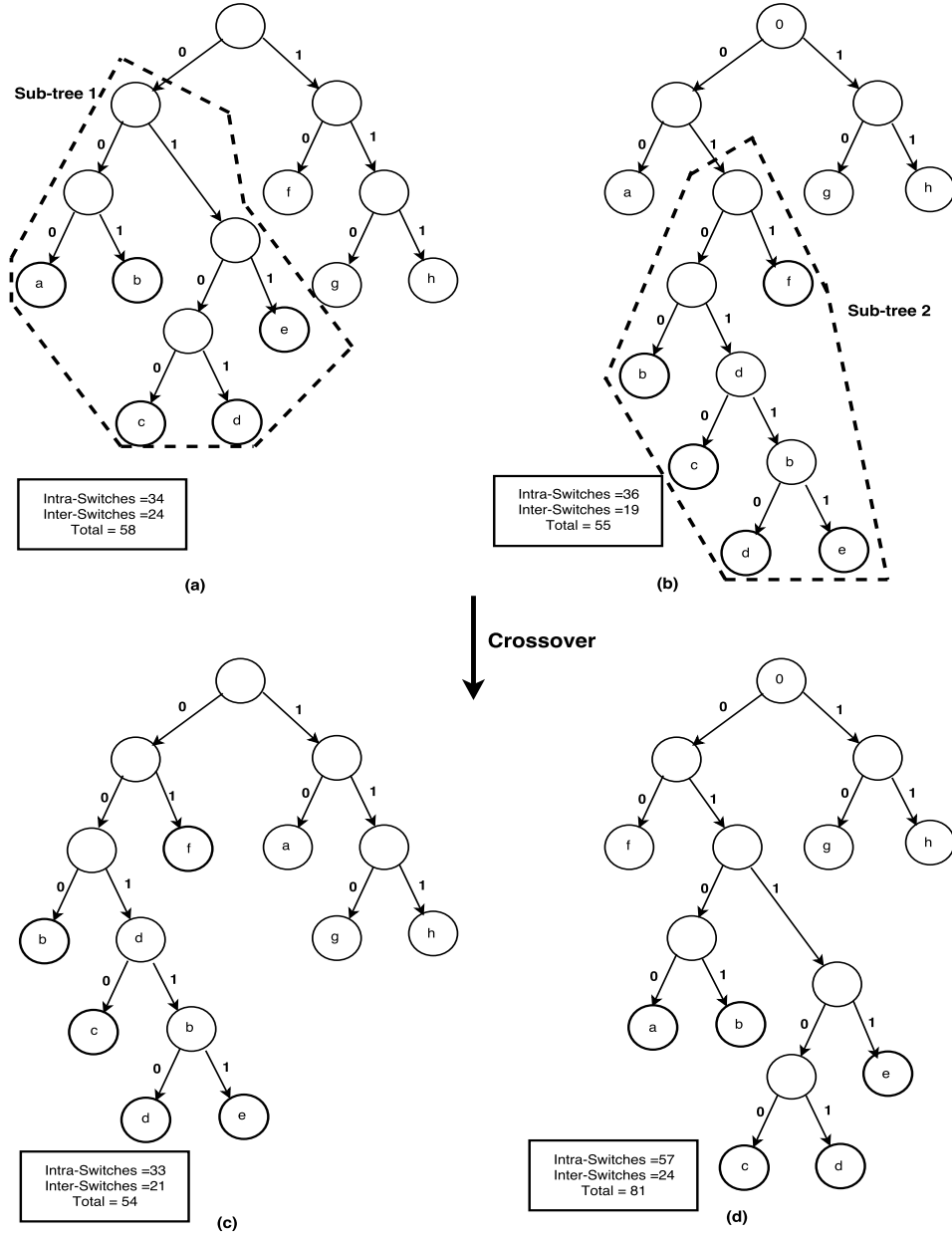


Fig. 1. The Crossover operation

of switching activities in both the approaches and the improvement rate. The results show that the proposed approach optimises the total number of switching of the whole dataset by reducing the total number of inter- and intra-switching. In the best case the proposed approach reduces the total number of switching over the classical Huffman code by 45.47%, in the worst case by 16.42%, and on an average by 36.45%.

V. CONCLUSION

In this paper, we described how switching activities affect the power consumption in digital communication. After that,

the relationship between the Huffman code and the switching activities has been explained. Genetic algorithm is identified as a potential method to optimise the total number of switching: inter- and intra-switching in Huffman code. Subsequently, an approach is proposed to show how the genetic algorithm can be used to reduce the power consumption in biological data transmission by reducing total number of switching in the transmission process. The advantage of this approach is that it minimises the total number of switching considerably, thus increases the power efficiency of data transmission. The effectiveness of the proposed approach is evaluated by applying it

TABLE I
DATASETS DESCRIPTION

Data sets	Size (MB)	Reference
Dataset 1	6.66	CP009496
Dataset 2	8.30	NZ_KB912942
Dataset 3	6.11	CP003169
Dataset 4	8.54	NZ_KB911581
Dataset 5	6.66	CP009494
Dataset 6	6.76	NZ_KI421511
Dataset 7	7.42	NZ_CM001442
Dataset 8	8.27	NZ_AXBH00000000.1
Dataset 9	5.74	NZ_CM000747
Dataset 10	6.03	NZ_CM000748
Dataset 11	6.48	NZ_AFXI01000001
Dataset 12	5.97	NZ_CM000753
Dataset 13	5.75	NZ_CM000750
Dataset 14	6.28	CP006982
Dataset 15	25.58	GL456087
Dataset 16	56.14	CM002885
Dataset 17	76.64	CM000680
Dataset 18	99.94	CM000684

TABLE II
COMPARISON OF THE TOTAL NUMBER OF SWITCHES (IN MILLIONS)
BETWEEN THE HUFFMAN CODE AND THE PROPOSED APPROACH

Datasets	Huffman Code	The proposed approach	Improvement (%)
Dataset 1	18.16	10.05	44.65
Dataset 2	24.66	15.63	36.61
Dataset 3	18.46	10.66	42.25
Dataset 4	25.18	16.26	35.42
Dataset 5	19.24	16.08	16.42
Dataset 6	19.61	14.96	23.71
Dataset 7	22.40	13.21	41.02
Dataset 8	23.87	17.70	25.84
Dataset 9	17.66	10.04	43.14
Dataset 10	18.14	9.89	45.47
Dataset 11	19.63	11.48	41.51
Dataset 12	17.59	11.86	32.57
Dataset 13	17.39	10.93	37.14
Dataset 14	18.53	11.64	37.18
Dataset 15	78.19	50.60	35.28
Dataset 16	168.98	102.13	39.56
Dataset 17	217.68	130.93	39.85
Dataset 18	243.70	149.80	38.53

to several biological datasets. The performance of the approach has been compared with the performance of the classical Huffman code. The results show that the proposed approach reduces the total number of switching significantly to improve the power efficiency of data transmission. In future, we hope to improve the allocation method of codewords to different symbols to reduce further the total number of switching.

REFERENCES

- [1] A. P. Chandrakasan and R. W. Brodersen, "Minimizing power consumption in digital cmos circuits," *Proceedings of the IEEE*, vol. 83, no. 4, pp. 498–523, 1995.
- [2] S. Gregori, Y. Li, H. Li, J. Liu, and F. Maloberti, "2.45 GHz power and data transmission for a low-power autonomous sensors platform," in *Proceedings of the 2004 International Symposium on Low Power Electronics and Design (ISLPED '04)*, Aug 2004, pp. 269–273.
- [3] D. Yates, A. Holmes, and A. Burdett, "Optimal transmission frequency for ultralow-power short-range radio links," *IEEE Transactions on Circuits and Systems*, vol. 51, no. 7, pp. 1405–1413, July 2004.
- [4] L. Wang, A. Ukhanova, and E. Belyaev, "Power consumption analysis of constant bit rate data transmission over 3g mobile wireless networks," in *11th International Conference on ITS Telecommunications (ITST)*, Aug 2011, pp. 217–223.
- [5] W. Toorisaka, G. Hasegawa, and M. Murata, "Power consumption analysis of data transmission in iee 802.11 multi-hop networks," in *The Eighth International Conference on Networking and Services (ICNS)*, 2012, pp. 75–80.
- [6] D. Huffman, "A method for the construction of minimum-redundancy codes," *Proceedings of the Institute of Radio Engineers*, vol. 40, no. 9, pp. 1098–1101, Sept 1952.
- [7] S. Kabir, "A compressed representation of mid-crack code with huffman code," *International Journal of Image, Graphics and Signal Processing*, vol. 7, no. 10, pp. 11–18, 2015.
- [8] S. Kabir, T. Azad, and A. A. Alam, "Freeman chain code with digits of unequal cost," in *8th International Conference on Software, Knowledge, Information Management and Applications (SKIMA)*. IEEE, 2014, pp. 1–6.
- [9] M. J. Golin, C. Mathieu, and N. E. Young, "Huffman Coding with Letter Costs: A Linear-Time Approximation Scheme," *SIAM Journal on Computing*, vol. 41, no. 3, pp. 684–713, 2012.
- [10] S. Kabir, T. Azad, A. A. Alam, and M. Kaykobad, "Effects of unequal bit costs on classical huffman codes," in *17th International Conference on Computer and Information Technology*, December 2014, pp. 96–101.
- [11] C.-Y. Chen, Y.-T. Pai, and S.-J. Ruan, "Low power huffman coding for high performance data transmission," in *International Conference on Hybrid Information Technology (ICHIT)*, Nov 2006, pp. 71–77.
- [12] N. H. E. Weste and K. Eshraghian, *Principles of CMOS VLSI design: a systems perspective*. Addison-Wesley, 1988.
- [13] M. Pedram, "Power minimization in IC design: principles and applications," *ACM Transactions on Design Automation of Electronic Systems (TODAES)*, vol. 1, no. 1, pp. 3–56, 1996.
- [14] T. H. Cormen, C. Stein, R. L. Rivest, and C. E. Leiserson, *Introduction to Algorithms*, 2nd ed. McGraw-Hill Higher Education, 2001.
- [15] M. C. Brandon, D. C. Wallace, and P. Baldi, "Data structures and compression algorithms for genomic sequence data," *Bioinformatics*, vol. 25, no. 14, pp. 1731–1738, 2009.
- [16] C. Scott, L. Yiming, L. Chen, and X. Xiaohui, "Human genomes as email attachments," *Bioinformatics*, vol. 25, no. 2, pp. 274–275, 2009.
- [17] C. Wang and D. Zhang, "A novel compression tool for efficient storage of genome resequencing data," *Nucleic acids research*, vol. 39, no. 7, pp. e45–e45, 2011.
- [18] H. Chan, P. Mazumder, and K. Shahookar, "Macro-cell and module placement by genetic adaptive search with bitmap-represented chromosome," *Integration, the VLSI journal*, vol. 12, no. 1, pp. 49–77, 1991.
- [19] M. Bright and T. Arslan, "Genetic framework for the high level optimisation of low power vlsi dsp systems," *Electronics Letters*, vol. 32, no. 13, pp. 1150–1151, 1996.
- [20] M. Mitchell, *An introduction to genetic algorithms*. MIT press, 1998.
- [21] T. Blicke and L. Thiele, "A comparison of selection schemes used in genetic algorithms," 1995.
- [22] E. Osaba, R. Carballedo, F. Diaz, E. Onieva, I. de la Iglesia, and A. Perallos, "Crossover versus mutation: A comparative analysis of the evolutionary strategy of genetic algorithms applied to combinatorial optimization problems," *The Scientific World Journal*, vol. 2014, 2014.
- [23] H. Mühlenbein, "How genetic algorithms really work: Mutation and hillclimbing," in *PPSN*, vol. 92, 1992, pp. 15–25.
- [24] K. D. Pruitt, T. Tatusova, W. Klimke, and D. R. Maglott, "NCBI reference sequences: current status, policy and new initiatives," *Nucleic acids research*, vol. 37, no. suppl 1, pp. D32–D36, 2009.

Modeling and Simulation of Stand-alone Fuel Cell System for Distributed Generation Application

Himadry Shekhar Das, Chee Wei Tan, A.H.M. Yatim, Nik Din bin Muhamad

Department of Electrical Power Engineering, Faculty of Electrical Engineering

Universiti Teknologi Malaysia

Johor, Malaysia

Himadry.s.das@gmail.com, cheewei@utm.my, halim@utm.my, nikd@fke.utm.my

Abstract— Fuel cell (FC) is one of the most suitable candidates for distributed generation system either in standalone or grid connected mode. Due to the cleanness, modularity and higher potentiality, FC systems can be used in both single and hybrid systems for power generation. In this study a 26 kW standalone proton exchange membrane fuel cell (PEMFC) based energy system is developed using MATLAB/Simulink®. The PEMFC stack is modeled based on the mathematical equations and implemented using the Simulink blocks. Similarly the power converters are also modeled using the Simulink block sets. The PEMFC output DC voltage is stepped up using a boost DC-DC converter then fed to a PWM based inverter. The controller for DC-DC boost converter is a PI based single loop voltage controller. The inverter controller uses RMS feedback controller and LC filter. The AC output power is tested with resistive (R) and inductive (RL) loads. The dynamic performance of FC stack is also verified by applying step changes in both resistive and inductive loads. Simulation results shows that the PEMFC stack is capable of delivering the maximum output power with voltage and current ripple within tolerable range. Moreover, the AC output voltage, frequency are within the acceptable range.

Keywords—fuel cell; distributed generation; system modeling; PEMFC; MATLAB/Simulink

I. INTRODUCTION

The search of alternative energy sources for power generation has resulted in photovoltaic (PV), wind, hydropower, biomass, and fuel cell (FC) and so on [1]. Among these resources fuel cell is one of the choices which is free from environmental effects. Also FC systems are capable of operating at higher efficiency (around 60%) than the other traditional energy sources [2]. Thus, the research of fuel cell power generation has been dramatically increased in recent years. FC is an electrochemical device which transforms the chemical energy of hydrogen and oxygen into electricity by producing water and heat as output [3]. For small scale power generation, proton exchange membrane fuel cell (PEMFC) or solid oxide fuel cell (SOFC) are appropriate [4]. Among these two types, PEMFC works at low temperature (60°C -100°C) and SOFC works at high temperature (500°C-650°C) [5]. Thus for small scale power generation PEMFC is more appropriate.

Fuel cell power generation system can operate in two modes: standalone and grid connected. Also, it can operate as single source system or hybrid system with any other energy

sources or energy storage devices [6]. In this study a PEMFC power generation system is designed which operates as single source in standalone mode. The standalone single sourced fuel cell system is considered to supply electricity without being connected to the utility grid. This type of systems can be used as power supply to remote areas, back-up power supply or transportation [7]. The complete system development contains a PEMFC stack, boost DC-DC converter, inverter, controller for both DC-DC and DC-AC converters and electrical load [8]. A storage device can be incorporated into the system to support the system during critical load conditions such as abrupt load change or during start-up because FC system response for such cases is slow.

This study focuses on modeling and simulation of standalone PEMFC based system without battery backup using MATLAB/Simulink®. The PEMFC model is constructed using a set of mathematical equations considering both the steady state and dynamic characteristics [9, 10]. Similarly the Simulink model of DC-DC converter and inverter are also modelled. The converter controller model ensures the FC operation at its maximum power point and PWM inverter controller ensures acceptable single phase AC operating voltage, frequency and THD (total harmonic distortion). The complete system block diagram is shown in the Figure 1.

This paper starts with the mathematical modeling of the system, then the simulation settings are elaborated, next the simulation results are presented and finally the discussions on simulation results are presented.

II. MATHEMATICAL MODELING

The fuel cell modeling is done based on mathematical equations derived in several literatures [1, 9, 11-14]. The system mathematical model consists of steady state characteristics, dynamic characteristics and also thermodynamic characteristics.

A. PEMFC Model

The PEMFC model parameters chosen for this study are adapted from Ballard Mark V FC module. The specification of modelled stack is given in the Table 1 [11].

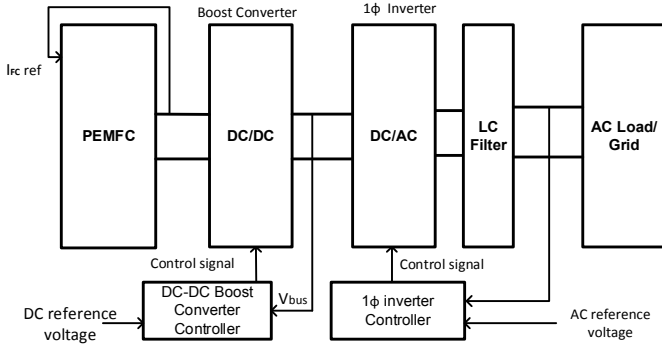


Fig. 1. Block diagram of the PEMFC system.

1) Steady State model

The steady state model consists of the reversible potential (E_{Nernst}), activation voltage loss (V_{act}), ohmic voltage loss (V_{ohm}), and concentration voltage loss (V_{con}). The equation of Nernst potential is given in equation 1 [11].

$$E_{\text{Nernst}} = 1.229 + 0.85 \times 10^{-3}(T - 298.15) + 4.3085 \times 10^{-5} \cdot T \cdot [\ln(P_{H_2}) + 1/2 \ln(P_{O_2})] \quad (1)$$

The activation voltage equation is given in Equation 2 [11, 12].

$$V_{\text{act}} = -[\xi_1 + \xi_2 \cdot T + \xi_3 \cdot T \cdot \ln(C_{O_2}) + \xi_4 \cdot T \cdot \ln(i)] \quad (2)$$

Here, $\xi_1 = -0.948$, $\xi_2 = [286 + 20 \ln(A) + 4.3 \ln(C_{H_2})] \times 10^{-5}$, $\xi_3 = 7.6 \times 10^{-5}$ and $\xi_4 = -1.93 \times 10^{-4}$ are the parametric coefficients.

The ohmic voltage drop equation is shown in the Equation 3.

$$V_{\text{ohm}} = -i(R_c + \frac{\rho_M \cdot \ell}{A}) \quad (3)$$

$$\text{Where, } \rho_M = \frac{181.6 \left[1 + 0.03 \cdot \left(\frac{i}{A} \right) + 0.062 \cdot \left(\frac{T}{303} \right) \left(\frac{i}{A} \right)^{2.5} \right]}{\left[\lambda - 0.634 - 3 \cdot \left(\frac{i}{A} \right) \right] \cdot \exp \left[4.18 \cdot \left(\frac{T - 303}{T} \right) \right]}$$

The concentration voltage drop is calculated by the equation 4.

$$V_{\text{con}} = -B \cdot \ln \left(1 - \frac{j}{j_{\text{max}}} \right) \quad (4)$$

2) Dynamic Model

The dynamic characteristics of the FC performance is because of the double-layer charging effect. The dynamic relationship can be expressed as equation 5 [9, 11]:

$$\frac{dV_d}{dt} = \frac{I}{C_{dl}} - \frac{V_d}{R_a C_{dl}} \quad (5)$$

Where, C_{dl} is the double layer charge, V_d and R_a is expressed as $V_d = V_{\text{act}} + V_{\text{con}}$, and $R_a = \frac{(V_{\text{act}} + V_{\text{con}})}{i}$.

3) Thermodynamic Model

The thermodynamic model provides the transient temperature of the fuel cell which is used to estimate the reactant gas diffusion, Nernst potential and all the voltage drop equations. Equation 6 shows the mathematical expression of the thermodynamic model.

TABLE I. FUEL CELL MODEL SPECIFICATION

Parameters	Value
Rated Power	25400W
Voltage at maximum power	126V
No of Cell	310
Unit Cell output voltage	0.57V
Current (A)	0 to 340 A
Operating Temperature (K)	323.5 K
Pressure at Anode (Pa)	2.3816 atm
Pressure at Cathode (Pc)	2.3816 atm
Thickness, L	178e-4 cm
FC Area	232 cm ²

$$mC \frac{dT}{dt} = i(E_{\text{Nernst}} - V_{\text{cell}}) - hA(T - T_0) \quad (6)$$

Where, m is the cell mass and C is the average specific heat (J/K) of the cell body, hA is the total heat transfer coefficient for the surface of fuel cell (W/K), T is the operating temperature (323.5 K) and T_0 is the standard temperature (296.5K).

B. DC-DC Converter Model

The DC-DC boost converter is a unidirectional converter which converts the low output voltage of FC into high voltage DC. In this system the converter converts the FC output voltage from 100-200V to 400V. The inductor (L_{dd}) and capacitor (C_{dd}) of the converter is designed using the equation (7) and (8). The detail specification of the boost converter is displayed in the Table 2.

$$\frac{D(1-D)^2 R}{f_s L_{dd}} \leq 5\% \quad (7)$$

$$\frac{D}{R f_s C_{dd}} \leq 5\% \quad (8)$$

Where, R is the equivalent resistance of the boost converter output, D is the duty cycle, and f_s is the switching frequency.

To control the boost converter a proportional integral (PI) controller is designed. The controller measures the output voltage and compares with a supplied reference voltage and using the error signal, the PI controller produces the duty ratio. The PWM uses the duty ratio to produce the switching pulses. The controller configuration is shown in the figure 2.

TABLE II. SPECIFICATION OF DC-DC BOOST CONVERTER

Specification	Values
Rated input voltage	180 V (150V-220V)
Rated output voltage	400 V
Rated power	26 kW
Input current ripple	$\leq 5\%$
Output voltage ripple	$\leq 5\%$
Switching frequency, f_s	25 kHz
Inductor, L_{dd}	1 mH
Capacitor, C_{dd}	100 μ F
Duty cycle, D	0.685
Integral gain, K_i	0.001
Proportional gain, K_p	25

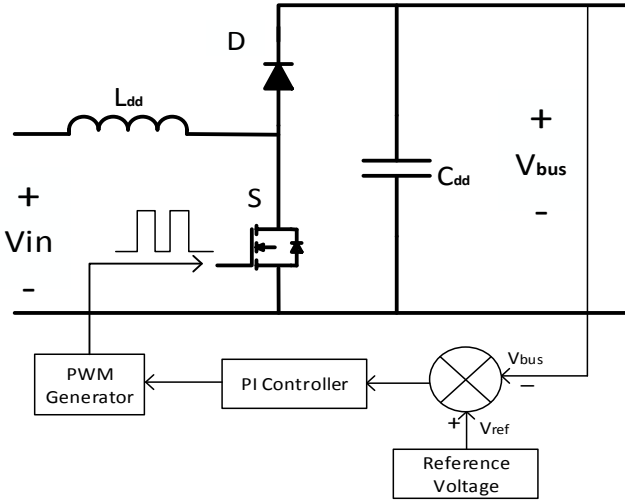


Fig. 2. DC-DC converter controller.

C. DC-AC Converter Model

The DC-AC converter or inverter produces AC output from DC input voltage. In this system a single phase PWM based voltage source inverter (VSI) is applied. The inverter converts 400V DC voltage into 230Vrms sinusoidal AC voltage. The Main components of the inverter are the input Capacitor and output filter. The input capacitor or the DC link capacitor suppresses the boost converter output voltage ripple more effectively, and the output filter converts pulsating output AC voltage into pure sinusoidal AC voltage. The VSI configuration with load voltage RMS feedback controller and LC filter is shown in the Figure 3.

Here, a PI based Voltage and frequency controller is designed for the inverter. The actual load voltage in per unit signal is first converted to root mean square (RMS) value and then compared with RMS reference (0.7071 V). The resultant error signal is fed to the PI controller. The PI controller provides the duty ratio in the range of 0 V to 0.85 V and it is then multiplied with a sinusoidal reference signal of fundamental frequency 50Hz to provide the main command signal for the inverter PWM. The AC duty signal is supplied to the PWM generator to produce switching pulses for the IGBTs. The inverter feeds a stand-alone AC load of 25 A, 50 Hz, @ 240 V_{rms} ±5% [16]. The inverter circuit parameters are given in the Table 3.

The inverter produces pulsating AC voltage at the output. Thus, an effective output filter is mandatory to design. A second order LC filter is designed using the equation (13) [1].

$$\frac{X_L}{X_C} \geq \frac{23.22}{n^2} \quad (13)$$

Where n is the ratio of the switching frequency to the fundamental frequency f. n is given by:

$$n = \frac{f_s}{f} \quad (14)$$

D. Load Model

To test the FC system two types of load are used. Firstly, the system is tested with a resistive load of 8.82Ω. Next the

load is selected as resistive inductive load of 19500W and 1950 Var. AC load variation is tested with step change for both R and RL loads.

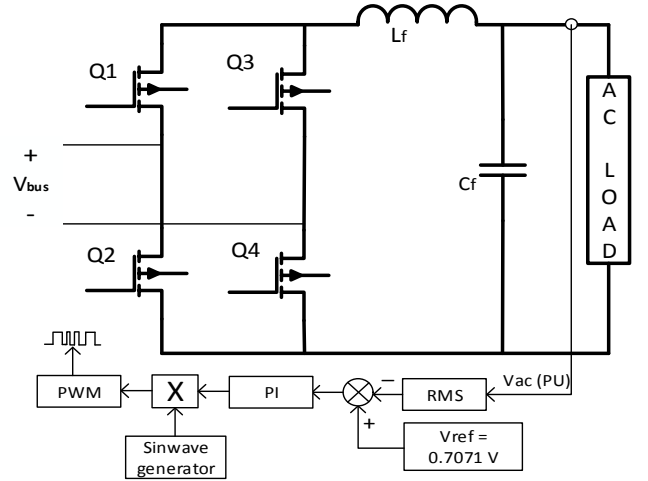


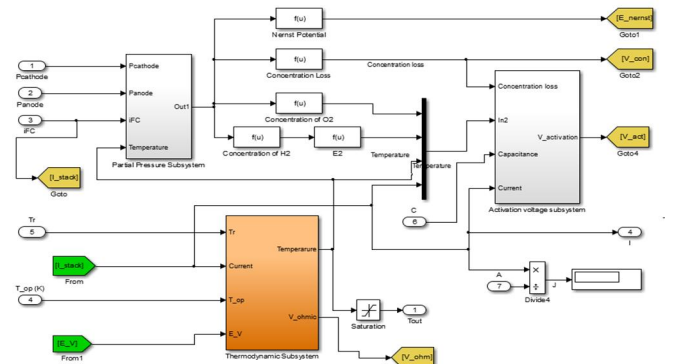
Fig. 3. dc-ac single-phase inverter with load voltage RMS feedback controller and LC filter.

TABLE III. SPECIFICATION OF DC-AC CONVERTER

Specification	Values
Rated input voltage	400V DC
Rated output voltage	230Vrms AC
Rated power	26 kW
Career frequency, f_s	1.950 kHz
DC link Capacitor, C_{dc}	12000 μ F
Integral gain, K_i	0.5
Proportional gain, K_p	20
Filter inductance, L_f	2 mH
Filter Capacitance, C_f	100 μ F

III. SIMULATION

The mathematically modelled PEMFC is simulated in MATLAB Simulink. Here in this model both steady state and dynamic characteristics are adopted. To plot the characteristics curves the PEMFC model is simulated without any load and with a ramp current from 0 to 330A. Other parameters of the fuel cell are kept constant. The simulation diagram is shown in the Figure 4. The Figure 5 shows the V-I characteristics and P-I characteristics curve achieved from the no load test.



(a)

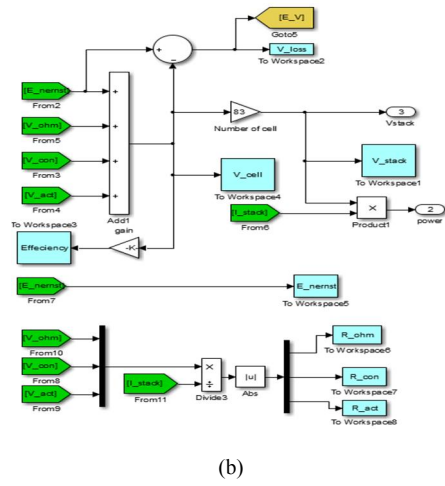


Fig. 4. PEMFC MATLAB/Simulink model.

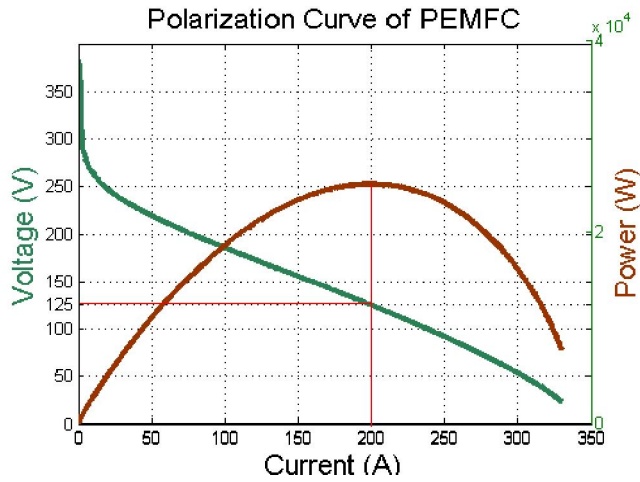


Fig. 5. PEMFC model characteristics

It is clear that, the maximum power is 25400W at the stack voltage 126V while approximately 200A current is drawn from the stack. Figure 6 displays the MATLAB/Simulink model of complete single phase PEMFC system with the DC-DC converter and the inverter.

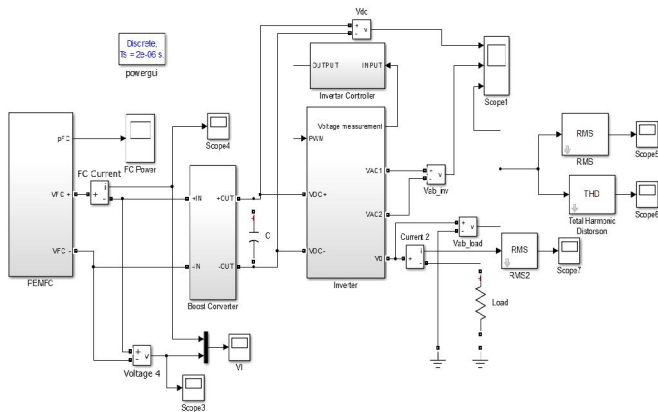


Fig. 6. Fuel Cell system MATLAB/Simulink model.

IV. RESULTS AND DISCUSSION

In this section the simulation results with resistive and inductive load will be presented. Also the dynamic characteristics of the fuel cell will be analyzed by observing the output power and voltage change with step change in load resistance. The simulation is performed on $2\mu s$ sampling time. Figure 7 presents the output RMS voltage and current with 2.3Ω resistive load. It is clear that the output RMS voltage is constant at 240V with the load current 105A. Figure 8 shows the PEMFC stack voltage and current characteristics under resistive load condition. It is clear that, PEMFC voltage becomes constant at 3seconds approximately. This is due to the slow nature of PEMFC itself. Similar characteristics is shown in Figure 9 and 10 where, the load is inductive (RL). The load used is of 25000 kW and 2500 kVar. The response of the system under both resistive and inductive condition are indifferent. The Figure 11 to 14 shows the power and voltage characteristics of load and PEMFC stack under different resistive and inductive conditions. It can be seen that, till the maximum power point the voltage is stable. After the maximum power point the voltage drops drastically. The reason is, the PEMFC cannot support beyond that power and terminal voltage is dropped.

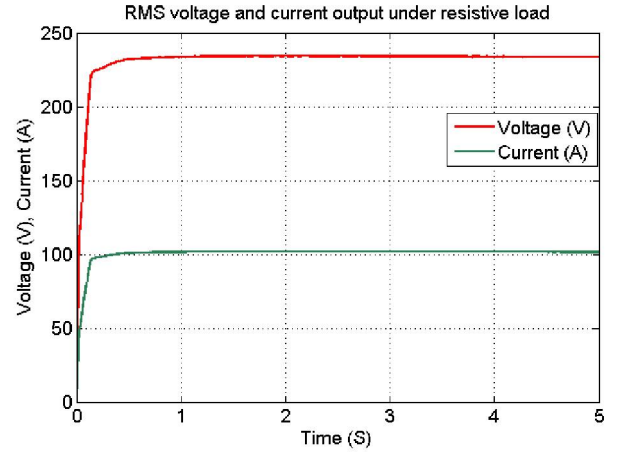


Fig. 7. RMS voltage and current output with 2.3Ω resistive load.

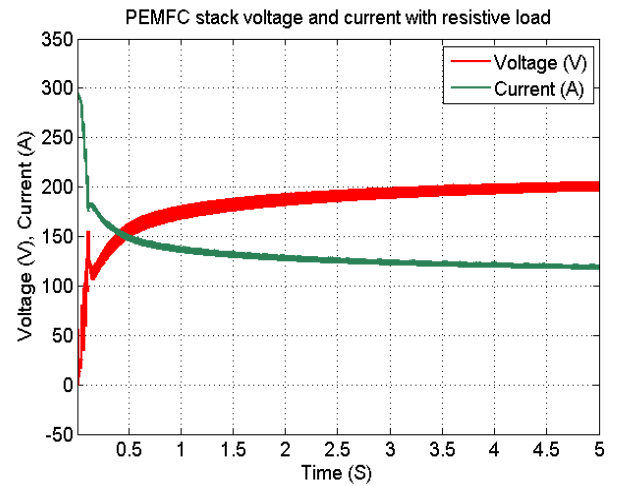


Fig. 8. PEMFC stack voltage and current with 2.3Ω resistive load

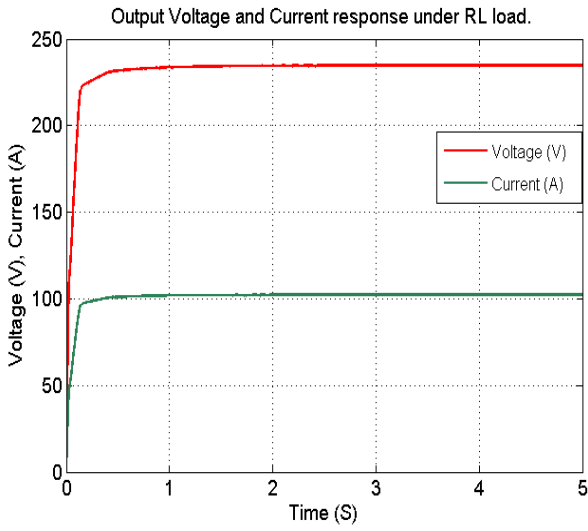


Fig. 9. Output RMS voltage and current with inductive load.

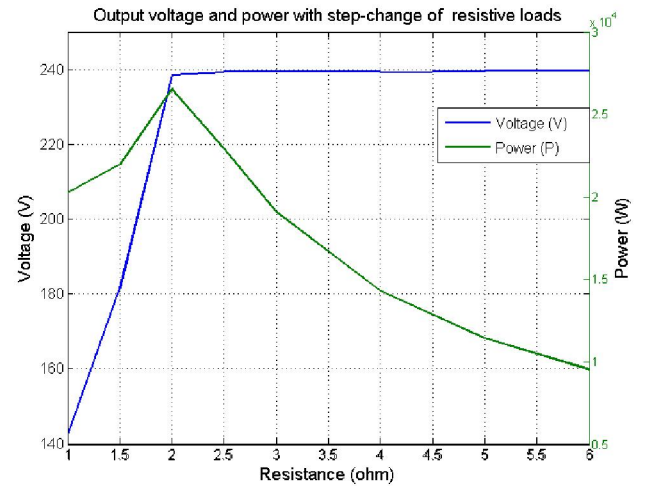


Fig. 12. AC output voltage and power with step-change of resistive loads in Ohm

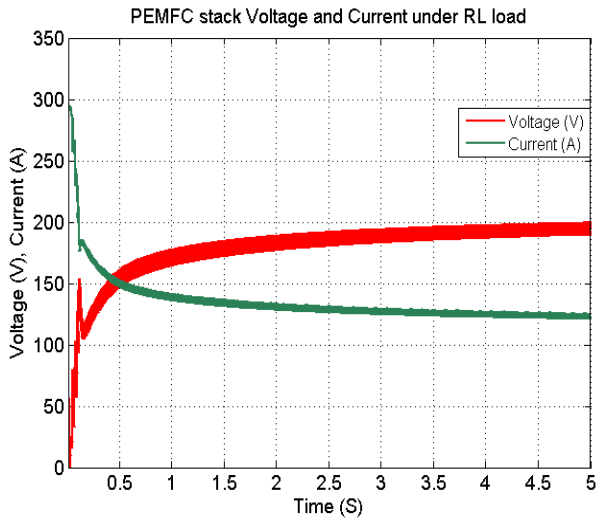


Fig. 10. PEMFC stack voltage and current with inductive load.

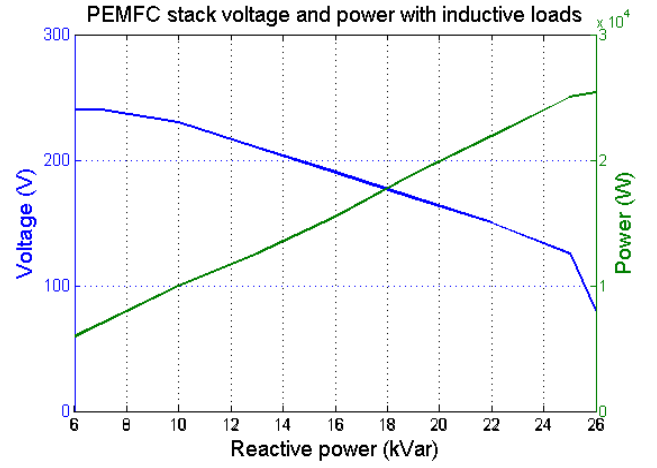


Fig. 13. PEMFC stack output voltage and power with step-change of inductive loads in kW and kVar

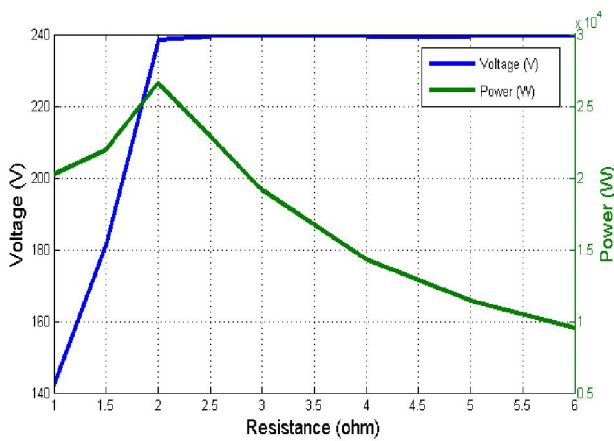


Fig. 11. PEMFC stack output voltage and power with step-change of resistive loads in Ohm

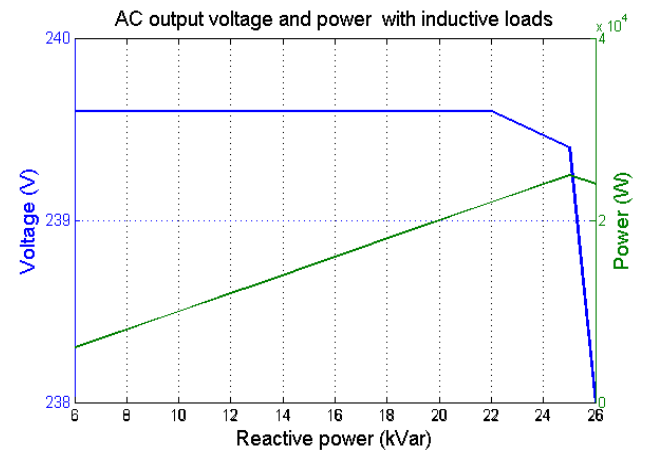


Fig. 14. AC output voltage and power with step-change of resistive loads in kW and kVar

The pure sinusoidal AC output waveform at maximum power point is presented in Figure 15.

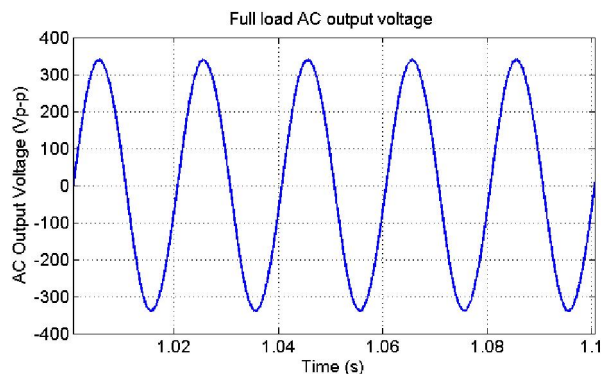


Fig. 15. The sinusoidal AC output waveform (Vp-p) at MPP (26 kW) output power

V. CONCLUSION

In this paper a mathematical approach in modeling the PEMFC stack and a complete energy system is shown using MATLAB/Simulink. The developed PEMFC model consists of steady state model, dynamic model and thermodynamic model. The V-I and P-I characteristics are presented in the simulation section. Next, The PEMFC model is simulated with DC-DC converter and inverter under resistive and inductive loads. PI based controller is designed for both DC-DC and DC-AC converters. The voltage and current response of FC stack and load is presented and analyzed. Furthermore, the system is interfaced with step change in resistive load and inductive loads to check the dynamic characteristics of the FC model. The results illustrate DC output responses of the FC stack model maximum power output capability. Similarly, AC output elements are justified within the acceptable tolerance. Although with the power electronic arrangements the FC slow dynamic problem has been resolved, the system cannot handle sudden high volume load change. To handle this situation different alternative energy sources and storage devices can be integrated. It will not only increase the reliability of fuel cells applications, but also improve the performances of the integrated system as the energy is clean and environmental friendly.

ACKNOWLEDGMENT

The authors would like to pay gratitude to Universiti Teknologi Malaysia (UTM) for supporting with lab and library facilities. The appreciation also goes to the Ministry of Higher Education, Malaysia (MOHE). The funding provided by fundamental research grant scheme (FRGS) under vote 4F596, Universiti Teknologi Malaysia (UTM) is also acknowledged. Lastly thanks to those colleagues who have either directly or indirectly contributed to the completion of this work.

REFERENCES

- [1] E. G. Services, P. Inc., and S. A. I. Corporation, *Fuel Cell Handbook*: U.S Department of Energy, 2002.
- [2] H. Gregor, "Fuel cell technology handbook," ed: CRC Press, New York, 2003.
- [3] G. Hoogers, *Fuel cell technology handbook*: CRC press, 2002.
- [4] A. Kirubakaran, S. Jain, and R. Nema, "A review on fuel cell technologies and power electronic interface," *Renewable and Sustainable Energy Reviews*, vol. 13, pp. 2430-2440, 2009.
- [5] S. Mekhilef, R. Saidur, and A. Safari, "Comparative study of different fuel cell technologies," *Renewable and Sustainable Energy Reviews*, vol. 16, pp. 981-989, 2012.
- [6] N. M. Isa, H. S. Das, C. W. Tan, A. H. M. Yatim, and K. Y. Lau, "A techno-economic assessment of a combined heat and power photovoltaic/fuel cell/battery energy system in Malaysia hospital," *Energy*, vol. 112, pp. 75-90, 2016.
- [7] S. F. Tie and C. W. Tan, "A review of energy sources and energy management system in electric vehicles," *Renewable and Sustainable Energy Reviews*, vol. 20, pp. 82-102, 2013.
- [8] A. Kirubakaran, S. Jain, and R. K. Nema, "A review on fuel cell technologies and power electronic interface," *Renewable and Sustainable Energy Reviews*, vol. 13, pp. 2430-2440, 2009.
- [9] Y. Xiao and K. Agbossou, "Interface design and software development for PEM fuel cell modeling based on Matlab/Simulink environment," in *Software Engineering, 2009. WCSE'09. WRI World Congress on*, 2009, pp. 318-322.
- [10] D. M. Ali, "A simplified dynamic simulation model (prototype) for a stand-alone Polymer Electrolyte Membrane (PEM) fuel cell stack," in *Power System Conference, 2008. MEPCON 2008. 12th International Middle-East*, 2008, pp. 480-485.
- [11] C.-H. Lee and J.-T. Yang, "Modeling of the Ballard-Mark-V proton exchange membrane fuel cell with power converters for applications in autonomous underwater vehicles," *Journal of Power Sources*, vol. 196, pp. 3810-3823, 2011.
- [12] H.-C. Chen, S.-y. Tzeng, and P.-h. Chen, "Optimization design of PID controllers for PEMFC with reformer using genetic algorithm," in *Machine Learning and Cybernetics (ICMLC), 2010 International Conference on*, 2010, pp. 2990-2995.
- [13] M. Uzunoglu and M. Alam, "Dynamic modeling, design and simulation of a PEM fuel cell/ultra-capacitor hybrid system for vehicular applications," *Energy Conversion and Management*, vol. 48, pp. 1544-1553, 2007.
- [14] A. Zaidy, P. Pokharkar, R. Krishnan, and D. Sonawane, "Dynamic modeling and simulation of a PEM fuel cell: MATLAB and LabVIEW modeling approach," in *Non Conventional Energy (ICONCE), 2014 1st International Conference on*, 2014, pp. 272-276.
- [15] M. H. Nehrir and C. Wang, *Modeling and control of fuel cells: distributed generation applications* vol. 41: John Wiley & Sons, 2009.
- [16] M. Trigg and C. Nayar, "Matlab Simulink modelling of a single-phase voltage controlled voltage source inverter," *Electrical Engineering*, pp. 1-6, 2006.

Randomized Energy-Based AODV Protocol For Wireless Ad-Hoc Network

Syeda Nyma Ferdous and Md. Shohrab Hossain

Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh

Email: nyma.ferdous@gmail.com, mshohrabhossain@cse.buet.ac.bd

Abstract—MANET is one kind of non-infrastructure network which is composed of various mobile nodes. Due to the battery energy constraints in MANET, energy efficiency is an important factor while designing ad hoc wireless networks. Previous works tried to optimize the AODV protocol by proposing different schemes. In this paper, two solutions have been proposed to control the congestion problem of wireless ad-hoc network in MANET. Some changes have been made in the RREQ packet to calculate the drop factor. The energy-based proposed scheme has been simulated with network simulator 2.35. It is seen from the result that the solutions proposed has performed comparatively better than the traditional AODV with respect of packet delivery ratio, throughput and delay.

Index Terms—MANET, Mobile Ad-hoc Network, Wireless Ad-hoc Network, Routing Protocol, On Demand Routing Protocol.

I. INTRODUCTION

Mobile Ad-hoc Network (MANET) is formed by a group of mobile nodes that are connected via wireless links with no pre-established network infrastructure [1]–[3]. As nodes move randomly in the network, it causes rapid changes and the network topology becomes unpredictable. The frequent movement of nodes in MANET makes routing in the network a challenging problem. MANETs can be applied in different fields, such as collaborative and distributive computing, emergency operations, disaster management etc.

MANET comprises of many mobile nodes that operate with limited battery power. So, energy efficiency is an important criteria to take into account for designing routing protocols in MANET.

There exist many solutions [4]–[6] for efficient AODV routing. Some of the proposals [7]–[9] are related to energy efficient AODV protocol design. Shahin Tajik proposed WAODV [10] where stable path is selected by sending *Hello* packets. Chonggun Kin proposed R-AODV [11] which deals with the node mobility problem by sending multiple RREPs. K. R. Shobha proposed named E-AODV [12] where blind flooding of RREQ message is avoided.

A. Objective

Our *objective* of this paper is to propose new solutions to ensure efficient routing. We have presented two solutions of AODV protocol efficient routing.



Fig. 1: Wireless ad-hoc network

B. Contributions

Contribution of this work contains: (i) proposing a new energy based AODV protocol. (ii) proposing a new hello message based AODV protocol for better route stability.

The organization of rest of this paper is as follows. In Section IV, new proposed AODV modified algorithm is explained. Section V outlines the network topology, simulation parameters and performance metrics. Section VI presents an analysis of all the results that we simulated. Finally, the paper comes to conclusion in Section VII.

II. BACKGROUND

A. Routing Protocol Types

According to route discovery, routing protocols are basically classified into three categories.

- Re- active (On Demand) Routing Protocols
- Pro- active (Table Driven) Routing Protocols
- Hybrid Routing Protocols

Re-active or On demand routing protocols follow on demand route requesting mechanism. These protocols store only active routes. [13], [14] So, memory requirement is reduced. Here, source node invokes route discovery mechanism when it

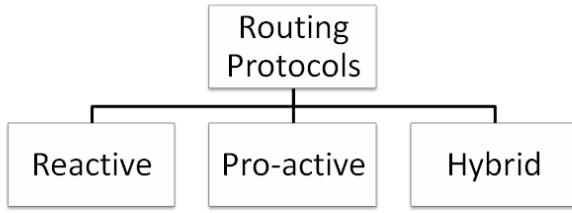


Fig. 2: Different types of Routing Protocols

needs to communicate with destination. In pro-active or table driven protocols, nodes maintain routes in routing tables and these tables are shared periodically. Nodes search the table to find routes from source to destination. Hybrid protocols combine the advantages of both on demand and table driven routing protocols.

III. AD HOC ON DEMAND DISTANCE VECTOR ROUTING PROTOCOL

Ad-hoc On Demand Distance Vector Routing Protocol, in short AODV [15] is a widely used routing protocol that is bi-directional reactive in nature. It is an enhancement over Destination Sequenced Distance Vector (DSDV) routing protocol. As it is a re-active protocol, the path of communication is established when it is needed. In AODV, nodes maintain routes only if they are used by source node. This protocol uses sequence number to keep up fresh routing information and to prevent formation of routing loops. There are mainly three kinds of packets in AODV protocol.

- **Route Request (RREQ) Packet:** To send out a packet from source to destination, originating node starts a path discovery process and sends RREQ packet containing destination address to its neighbor nodes to find a path to the corresponding destination node. When a node receives RREQ, a reverse data transmission path from destination node to source node is created in its routing table. Whenever a RREQ packet with identical source address with broadcast Id is received by a node, the redundant RREQ packet is dropped by the corresponding node. A reverse routing path from source node to destination node is formed by the above mentioned process.
- **Route Reply (RREP) Packet:** When the RREQ packet reaches to the destination node or to an intermediate node that knows route to the destination node, a reply packet named RREP is sent back to the corresponding source node via the nodes which were used to forward the RREQ packet.
- **Route Error (RERR) Packet:** Many link breakages can occur because of the dynamic movement of nodes in ad-hoc network. These breakages result in packet drop and a

message named RERR is generated. These messages are sent to those neighbors or a local repair will take place.

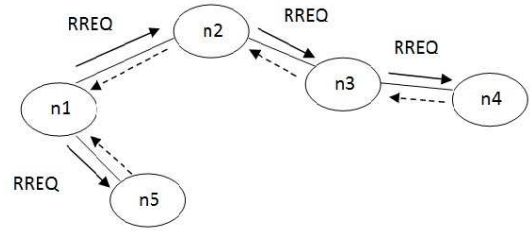


Fig. 3: RREQ packet traversal

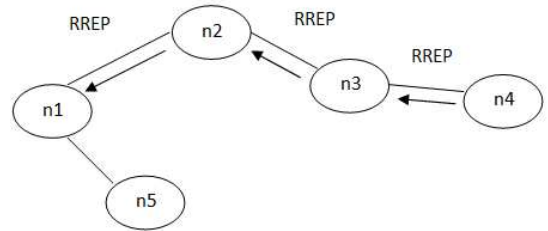


Fig. 4: RREP packet traversal

The working of AODV protocol is illustrated using Fig. 3. Suppose n1 wants to send data to n4. So, n1 needs a path for communication with n2. To discover the path, n1 will send RREQ to its neighbors n2 and n5. n2 and n5 will build a reverse path by making their destination n1, next hop n1, hop count 1. n2 does not know any path to n4. So, it will forward the RREQ packet to n3. n3 knows the path to destination n4.

RREP works in the following manner as shown in Fig. 4. n5 sends a RREP packet to n3, n3 then forwards the packet to n2 and n1.

IV. PROPOSED MODIFICATION

A. Using Hello Packet

Nodes in MANET may change their position from time to time. Hello packets are used for determining connectivity information of neighbor nodes. By default, this connectivity information is achieved in AODV by Link Layer Detection mechanism. Failure of reception of hello message means the node is not in the range of communication. According to our proposed scheme, all nodes will listen some predefined time intervals say c to count the number of Hello packets. The size of each interval will be say d seconds. Getting more hello packets from the i^{th} node for $(i+1)^{\text{th}}$ node means it will stay longer in transmission range of $(i+1)^{\text{th}}$ node. So, the node is stable for communication. After c intervals, $(i+1)^{\text{th}}$ node divides number of total received Hello packets by c . This value will be the node in range factor which will be a number between 0 and d . Nodes in MANET are highly unstable.

So, this node in range factor will determine, which nodes to send RREQ so that unnecessary RREQ packet delivery will be reduced. As some RREQ packet is dropped, network congestion will be lower. The algorithm is given below.

```

1.Calculate node in range factor.
2.Generate a random number between 0 and d.
3.If (random_value > node in range factor)
4.then broadcast/forward RREQ_packet,
5.else drop RREQ_packet.

```

B. Using Energy Based Parameter

Nodes in MANET rely on batteries. So, designing a routing protocol based on energy criteria can be an optimized solution for MANET. During route discovery process, every node decides either to forward RREQ packet or drop it. All nodes contain a certain energy. A new field will be added in the RREQ for calculating transmission energy. The format of the new RREQ will be

```

<source_address, dest_sequence_id, source_sequence_id,
broadcast_id, dest_address, hop_count, transmission_energy>

```

After each transmission, nodes will lose some energy. A transmission loss value will be deducted after each transmission. Prior to forward a packet, each and every node calculates a drop factor. This factor is calculated by dividing the remaining energy and total energy. This **drop factor** is in the range 0 to 1. Every node also generates a random number. If generated random number is greater than calculated drop factor, then RREQ packet is forwarded. Otherwise this packet is dropped. When the velocity of node is increased, then number of redundant RREQ packet is increased. So, the goal of our algorithm is to decrease number of redundant RREQ packets as much as possible. As some RREQ packet is dropped, so network congestion will be lower and throughput will be higher in this protocol. The modified algorithm is given below.

```

1.Calculate drop factor d for every transmitting node.
2.Generate a random number between 0 and 1.
3.If (random_value > drop factor)
4.then broadcast/forward RREQ_packet,
5.else drop RREQ_packet.

```

V. SIMULATION

In this section, we have explained the topology, parameters used in our simulation and the performance metrics that were measured for our proposed schemes and traditional AODV protocol.

A. Simulation Topology

The topology of 10 nodes in our simulation is shown in Fig. 5.

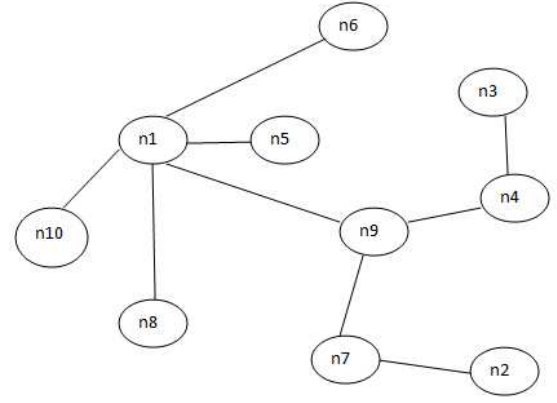


Fig. 5: Simulation topology.

B. Simulation parameters

For measurement of the performance of the proposed AODV protocol, we have used NS-2.35 [16] is used to simulate the idea with AODV protocol. We defined a scenario with $900m \times 400m$. Every node has an omnidirectional antenna and propagation model is used TwoRayGround. The average velocity of node is 20 m/s. We have used CBR (Constant Bit Rate) traffic at every node and UDP protocol in the transport layer. The simulation time is 200s. Table describes the parameters for 10 nodes of a scenario.

TABLE I: Simulation Parameters

Simulation Parameters	Values
Ns-2 Version	ns-2.35
Channel	WirelessChannel
Propagation	TwoRayGround
Model	Random Waypoint Model
Network Interface Type	Phy/WirelessPhy
Mac Type	Mac/802_11
Interface queue type	Queue/DropTail/PriQueue
link layer type	LL
Antenna Model	Antenna/OmniAntenna
Number of Nodes	10
Routing Protocol	AODV
Interface Queue Length	50
Simulation Time	200s

C. Performance Metrics

The energy-based solution is evaluated. The source-destination is selected randomly over the network. Each node starts from an arbitrary position and selects a destination. Simulation result is obtained by averaging several trial runs. Several performance metrics have been considered to evaluate the algorithm [17].

- *Packet Delivery Fraction (PDF)*: The rate of data packet conveyed to destination generated by CBR source.
- *Throughput*: The ratio of amount of packets successfully received by the nodes to the amount of packets sent by the nodes in a communication network.
- *Delay*: This includes all possible delays comprising re-transmission delay, route discovery latency, etc.

VI. RESULTS

We have compared conventional AODV with the proposed modified AODV protocols in terms of packet delivery fraction, throughput and delay. We varied number of nodes and mobility rate while comparing traditional AODV with the proposed schemes.

A. Varying Mobility

It can be observed that the delay is reduced in the proposed modified protocol. When higher mobility is added, the protocol shows better performance. As speed of the node increases, RREQ packet is frequently generated, causing network congestion but the new protocol drops some additional redundant RREQ packets. Packet delivery fraction decreases as the amount of packets successfully reaching to destination decreases (if mobility rate is increased). Fig. 6 shows packet delivery fraction of the traditional AODV and the proposed schemes.

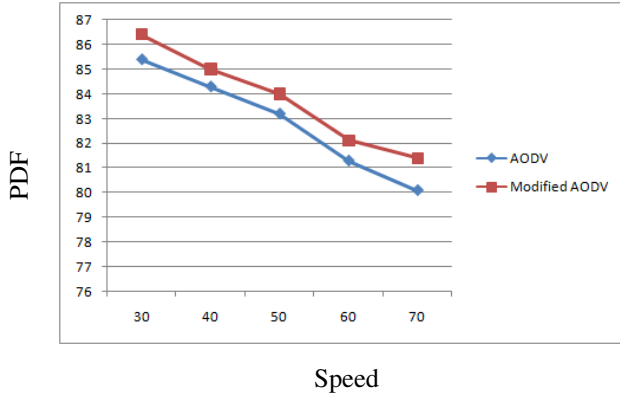


Fig. 6: Packet delivery ratio vs. speed.

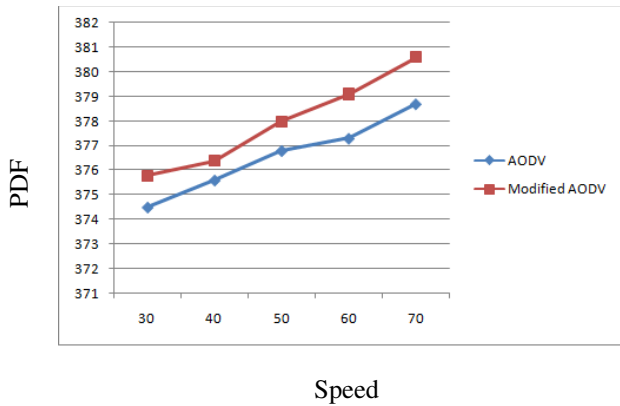


Fig. 7: Throughput vs. speed.

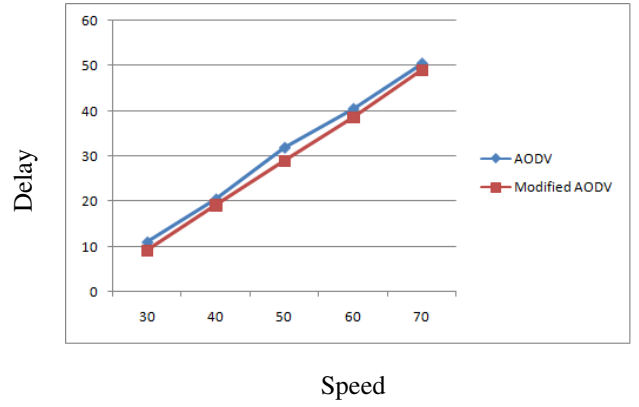


Fig. 8: Average end-to-end delay vs. speed.

Fig. 7 and Fig. 8 compare AODV and the proposed scheme in terms of throughput and delay, respectively. The modified protocol shows better performance with respect of these two performance metrics.

B. Varying Number of Nodes

Obviously, as number of node increases, delay also increases. The proposed scheme has higher packet delivery ratio than traditional AODV.

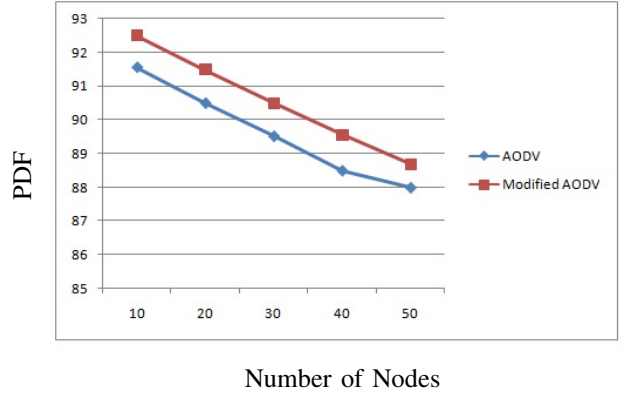


Fig. 9: Packet delivery ratio vs. number of nodes.

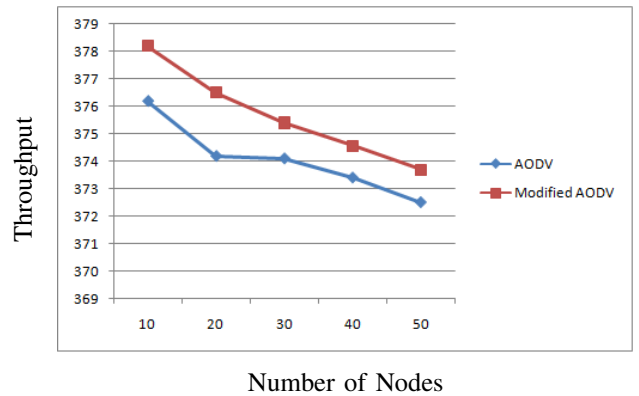


Fig. 10: Throughput vs. number of nodes

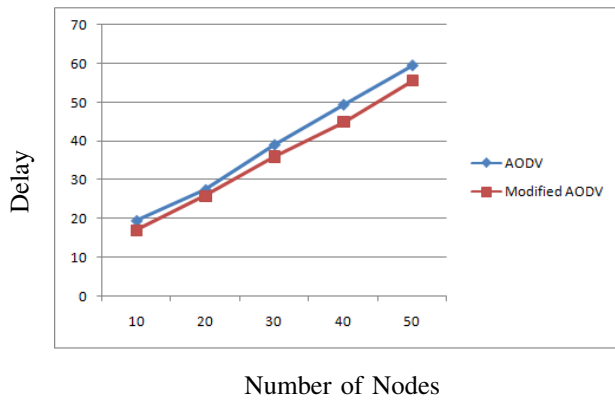


Fig. 11: Delay vs Number of Nodes

We can see that the modified energy based solution gives us better throughput than AODV since this protocol reduces network congestion and ensures successful message delivery by dropping redundant RREQ packets. We can see from Fig. 10 that as the network size grows, throughput also declines.

VII. CONCLUSION

In this paper, several new schemes for effective AODV routing protocol are proposed. But these solutions suffer from increasing network lifetime problem. We have simulated the energy based proposed scheme using ns-2.35 [16]. According to the simulation, our proposed scheme showed better performance than traditional AODV [15] in terms of throughput, average end-to-end delay and packet delivery fraction.

REFERENCES

- [1] A. S. Tanenbaum, *Computer Networks*, 2003.
- [2] T. W. M. Abolhasan and E. Dutkiewicz, "A review of routing protocols for mobile ad hoc networks," pp. 1–22, Nov 28-Dec 02, 2004.
- [3] T. P. Bahuguna Renu, Mandoria Hardwari lal, *Routing Protocols in Mobile Ad-Hoc Network: A Review*. New York, NY: Springer Berlin Heidelberg, 2013.
- [4] X. M. Zonghua Mou, "A modified aodv routing protocol based on route stability in manet," in *Wireless, Mobile Multimedia Networks (ICWMMN 2011), 4th IET International Conference on*, Beijing, Nov. 27–30 2011, pp. 63–67.
- [5] S. R. K. Vijayendra Chaudhary, Utkarsh Patel, "Aodv-mbr: A new routing protocol with modified backup approach," in *Wireless Networks and Computational Intelligence*, Bangalore, August 2012, pp. 66–73.
- [6] A. Boukerche, "Performance evaluation of two congestion control mechanisms with on-demand distance vector (aodv) routing protocol for mobile and wireless networks," presented at the 9th International Euro-Par Conference Klagenfurt, Austria, August 26–29, 2003 Proceedings, 2003.
- [7] B. C. Hannan Xiao, Dashti M. Ibrahim, "Energy consumption in mobile ad hoc networks," presented at the 2014 IEEE Wireless Communications and Networking Conference (WCNC), 2014.
- [8] N. N. . D. E. O. Abdellah Jamali, "An enhanced on ad hoc wireless network routing based on aodv," in *Wireless Days (WD), 2013 IFIP*, Valencia, Nov. 13–15, 2013, pp. 1 – 3.
- [9] M. G. Ranjan Kumar, "Route stability and energy aware based aodv in manet," in *High Performance Computing and Applications (ICHPCA), 2014 International Conference*, Bhubaneswar, Dec. 22–24 2014, pp. 1 – 5.
- [10] G. F. . S. Z. Shahin Tajik, "Performance of modified aodv (waiting aodv) protocol in mobile ad-hoc networks," pp. 160 – 164, Jun. 2010.
- [11] B. A. Chonggun Kim, Elmurod Talipov, "A reverse aodv routing protocol in ad hoc mobile networks," pp. 522–531, 2006.

- [12] K. R. K. R. Shobha, "Enhanced aodv routing protocol for mobile adhoc networks," pp. 265–276, 2011.
- [13] E. M. Royer and C. K. Toh, "A review of current routing protocols for ad hoc mobile wireless networks," Apr. 1999.
- [14] S. K. S. Charu Wahi, "Mobile ad hoc network routing protocols: A comparative study," *International Journal of Ad hoc, Sensor Ubiquitous Computing (IJASUC)*, vol. 3, no. 2, 2012.
- [15] S. D. C.E. Perkins, E. Belding-Royer, "Ad hoc on demand distance vector (aodv) routing," IETF Draft, Nov. 2003.
- [16] K. Fall, "The ns manual (formerly ns notes and documentation)," 2011.
- [17] S. Corson and J. Macker, "Mobile Ad hoc Networking (MANET):Routing Protocol Performance Issues and Evaluation Considerations," RFC2501, Jan 1999.

An Improved Universal Re-encryption Based on Okamoto-Uchiyama Cryptosystem

Nazmul Islam¹, Kazi Md. Rokibul Alam², Yasuhiko Morimoto³

¹Institute of Information and Communication Technology,

²Department of Computer Science and Engineering,

Khulna University of Engineering & Technology, Bangladesh

³Graduate School of Engineering, Hiroshima University, Higashi-Hiroshima 739-8521, Japan

Email: nazmul.kuet@gmail.com, rokibcse@yahoo.com, morimoto@mis.hiroshima-u.ac.jp

Abstract— Universal re-encryption technique is an important tool that enables an intermediary to re-encrypt a ciphertext without any knowledge of corresponding public key. This technique, as shown by Golle et. al., conventionally utilizes ElGamal cryptosystem for its operations and generates two pairs of ciphertexts. With ElGamal cryptosystem, efficiency of this technique decreases as resources in terms of time and memory are consumed significantly by the amount of generated ciphertexts. The purpose of the paper is to increase the efficiency of universal re-encryption technique by reducing the ciphertext quantity and to make the technique suitable for resource constraint devices while keeping the facilities of conventional universal re-encryption technique in full concern. To show the effectiveness of proposed technique, this paper includes an analysis on experimental data from the simulation of mix-networks.

Keywords— Okamoto-Uchiyama cryptosystem; Universal reencryption, ElGamal based universal re-encryption, Mix-network

I. INTRODUCTION

Universal re-encryption, established by Golle et. al. [1], focuses on giving the ability of re-encryption to a party without knowing the key under which the original encryption was performed but still permits an intended party to retrieve the plaintext. Though Golle et. al. proposed universal re-encryption with an intention to facilitate anonymous bulletin boards and mix-networks (mixnets) [2][3], it has potentiality to be used in many other applications such as untraceable messaging, anonymous communications [4][5] over different networks [6][7], cloud computing [8]. Covering steganographic filing systems [9] against observation-based attacks may also be a potential application of universal re-encryption.

In spite of being effective and secure, the construction of Golle et. al.'s technique along with ElGamal cryptosystem shows some significant disadvantages. With ElGamal cryptosystem, their proposed technique generates two pairs of ciphertext from a single plaintext i.e. increases the amount of ciphertexts four times than the plaintext. While in re-encryption, each pair of ciphertext requires individual operations which make the process computationally expensive and time consuming; hence infeasible for resource constraint devices.

To increase the efficiency of universal re-encryption technique, this paper, proposes a technique to re-encrypt

ciphertexts universally based on Okamoto-Uchiyama cryptosystem that reduces the computational requirements in terms of time and memory by generating only one pair of ciphertexts. Hence the proposed technique enables universal re-encryption to be executed in memory constraint inexpensive devices and makes the operation faster than before.

II. PRELIMINARIES

A brief description of ElGamal based universal re-encryption and, as a practical example, the working principle of this technique while implemented in a mix-network are presented in this section. Moreover, the Okamoto-Uchiyama (OU) cryptosystem [10][11], which is the base of the proposed re-encryption technique here, is also presented.

A. P. Golle et. al's ElGamal based Universal re-encryption

P. Golle et. al. have presented the universal re-encryption technique which can be applied over any proper algebraic group. They have added to a ciphertext of standard probabilistic cryptosystem e.g. ElGamal a second ciphertext on the corresponding identity element. Utilizing the second ciphertext it is possible to change the encryption factor in the first ciphertext because of algebraic homomorphism of the underlying cryptosystem. This technique has relaxed a party from having the knowledge of public key used in the encryption while performing the re-encryption operation.

If ElGamal encryption on a plaintext m is loosely denoted as $E(m)$, a ciphertext on message m in universal cryptosystem comprises with the pair $\langle E(m); E(1) \rangle$. Because of homomorphic property of ElGamal [12] i.e. $E[a] \times E[b] = E[ab]$ for group operator $\times$, it is possible to use the second constituent to re-encrypt the first without the associated public key. Let G stands for the underlying group of the ElGamal cryptosystem, q denote the order of G and g be a publicly known generator for G . The universal cryptosystem goes as follows.

- **Key generation (UKG):** Public and private key pair $(PK; SK) = (y=g^x; x)$ for $x \in Z_q$. Where $Z_q = \{1, \dots, q-1\}$.
- **Encryption (UE):** Inputs are— a message m , a public key y , and a random encryption factor $r=(k_0, k_1) \in Z_q^2$.

Output is– ciphertext $C = \langle (\alpha_0, \beta_0); (\alpha_1, \beta_1) \rangle = \langle (m^{k_0}, g^{k_0}); (y^{k_1}, g^{k_1}) \rangle$

- *Decryption (UD)*: Input– $C = \langle (\alpha_0, \beta_0); (\alpha_1, \beta_1) \rangle$. Verify $\alpha_0, \beta_0, \alpha_1, \beta_1 \in G$ and $\alpha_1/\beta_1^x = 1$ then accept $\alpha_0/\beta_0^x = m$.
- *Re-encryption (URE)*: Input– $C = \langle (\alpha_0, \beta_0); (\alpha_1, \beta_1) \rangle$ and a random re-encryption factor $r' = (k_0', k_1') \in \mathbb{Z}_q^2$. Output is a ciphertext $C' = \langle (\alpha_0', \beta_0'); (\alpha_1', \beta_1') \rangle = \langle (\alpha_0 \alpha_1^{k_0'}, \beta_0 \beta_1^{k_0'}); (\alpha_1^{k_1'}, \beta_1^{k_1'}) \rangle$

B. Mix-networks using ElGamal based Universal re-encryption technique

In ElGamal based universal re-encryption mixnet, a sender i broadcasts two ciphertext pairs generated by ElGamal cryptosystem, one of which contains the encrypted message m and other contains the encryption of identity element *i.e.* $E_{K_S}(m, R_i) \| E_{K_S}(1, R_i) = (g^{R_i} \| mK_S^{R_i}) \| (g^{R_i} \| K_S^{R_i})$ where $(R_i, R_i') \in \mathbb{Z}_q^2$ are random numbers chosen by the sender. After receiving a batch of such ciphertexts, a random mix-node j performs its mixing operation *i.e.* re-encrypts each entity of the batch and shuffles, namely $E_{K_S}(m, R_i + R_i' r_j) \| E_{K_S}(1, R_i' r_j) = (g^{R_i + R_i' r_j} \| mK_S^{R_i + R_i' r_j}) \| (g^{R_i' r_j} \| K_S^{R_i' r_j})$ with a random number $(r_j, r_j') \in \mathbb{Z}_q^2$ and transfers the batch to another mix-node l selecting randomly from the remaining mix-nodes for further mixing operation. The process continues until all N mix-nodes complete their mixing and thus sends $E_{K_S}^N(m, R_i^*) \| E_{K_S}^N(1, r_i^*) = (g^{R_i^*} \| mK_S^{R_i^*}) \| (g^{r_i^*} \| K_S^{r_i^*}) = (g^{R_i + R_i' r_j' + \dots + R_i' r_j' \dots r_N'} \| mK_S^{R_i + R_i' r_j' + \dots + R_i' r_j' \dots r_N'}) \| (g^{R_i' r_j' \dots r_N'} \| K_S^{R_i' r_j' \dots r_N'})$ to the receiver. Finally, the receiver gets the message $m = mK_S^{R_i^*} / (g^{R_i^*})^x$ by his private key x if $K_S^{R_i^*} / (g^{R_i^*})^x = 1$ is valid. Here, it is clear that the amount of ciphertexts and the cost of computation are double than those of the basic ElGamal based re-encryption mixnet. So, efficiency of mixnet with ElGamal based universal re-encryption becomes significantly lower than that of the ElGamal based re-encryption which, though requires keying information available at intermediary mix-nodes, produces only one pair of ciphertexts.

C. The Okamoto–Uchiyama(OU) cryptosystem

Let, p and q are large primes and n be the multiplication of p^2 and q *i.e.* $n = p^2 q$. The OU [10][11] scheme is constructed over the ring $\mathbb{Z}/n\mathbb{Z}$ *i.e.* multiplicative group of integers with modulo n . The public key of the OU scheme is the modulus n , an random element $g \in (\mathbb{Z}/n\mathbb{Z})^*$ such that the order of $(g^{p-1} \bmod p^2)$ is p and $h = (g^n \bmod n)$. The secret keys are the primes p and q . the ciphertext C of the message $m (0 < m < 2^{k-1})$, where k is the bit length of p is calculated as $E(m, r, \{g, n, h\}) = \{g^m h^r \bmod n\}$ where random number $r \in \mathbb{Z}/n\mathbb{Z}$. The ciphertext C is decrypted as $D_{OU}(C)$ *i.e.* $D(C, \{p, q\}) = \{L(C^{p-1} \bmod p^2) / L(g^{p-1} \bmod p^2)\} \bmod p = \{L(g^{m(p-1)} \bmod p^2) / L(g^{p-1} \bmod p^2)\} \bmod p = \{L(g^{m(p-1)} \bmod p^2) / L(g^{p-1} \bmod p^2)\} \bmod p = m$ where L is a logarithmic function and defined as $L(x) = (x-1)/p$. Here, the one-wayness and the semantic security of the OU scheme are as intractable as factoring the modulus n . For simplicity modulo operation is omitted in the remaining sections.

III. UNIVERSAL RE-ENCRYPTION BASED ON OKAMOTO–UCHIYAMA(OU) CRYPTOSYSTEM

This section describes the proposed re-encryption technique along with analyzing the security. Moreover, the working principle of the proposed technique while implemented in mix-network is also presented for a practical example.

A. Proposed technique for re-encryption

Efficiency of universal re-encryption technique can be improved by utilizing Okamoto–Uchiyama(OU) cryptosystem which will be shown experimentally in the next section. In this section the proposed universal re-encryption technique is presented. Unlike ElGamal based universal re-encryption the proposed technique generates only one pair of ciphertext; second constituent of which is a ciphertext on $t(t = 0)$ instead of the identity element 1. Because of algebraic homomorphism of Okamoto–Uchiyama(OU) cryptosystem the second constituent of the pair can be used to re-encrypt the first constituent. Hence, with Okamoto–Uchiyama(OU) based re-encryption technique the mix nodes are relieved from the hassle of weighty algebraic operations as two pairs of ciphertexts are replaced with only one while maintaining the facilities of ElGamal based universal re-encryption cryptosystem.

Let $(\mathbb{Z}/n\mathbb{Z})^*$ denote the multiplicative group of integer modulo n and $n = p^2 q$ where p, q are large primes. The proposed technique of universal re-encryption based on OU cryptosystem proceeds as follows.

- *Key generation (KG)*: Output is public key and secret key sets $\langle PK, SK \rangle = \langle \{n, g, h\}, \{p, q\} \rangle$ where $g \in (\mathbb{Z}/n\mathbb{Z})^*$ such that $g^p \neq 1 \bmod p^2$ and $h = g^n \bmod n$.
- *Encryption (E)*: Input consists of a message $m \in \mathbb{Z}/n\mathbb{Z}$, public key set $\{n, g, h\}$, and a random encryption factor $r_0 = (r, s) \in \mathbb{Z}/n\mathbb{Z}$. The ciphertext $C_0 = (\alpha_0, \beta_0) = (g^m h^r, h^s)$ becomes the output of this phase.
- *Decryption (D)*: Input is a ciphertext $C_0 = (\alpha_0, \beta_0)$. Calculate $m_0 = D_{OU}(\alpha_0)$ and $m_1 = (\beta_0)^{p-1} \bmod p^2$. If $m_1 = 1$, then the output is $m = m_0$ or else, the decryption fails and a special symbol $\perp$ is returned. Note that this verification guarantees the correlation between public keys and ciphertexts; particularly, decryption of a given ciphertext can be carried out only under one given key.
- *Re-encryption (RE)*: Input is a ciphertext $C_0' = (\alpha_0', \beta_0')$ and a random number s' where $s' \in \mathbb{Z}/n\mathbb{Z}$. Output is a ciphertext $C_1' = (\alpha_1', \beta_1') = (\alpha_0' \beta_1', (\beta_0')^{s'})$.

Here, similar to the ElGamal based re-encryption, the proposed re-encryption technique generates ciphertexts twice in number than encryption with underlying cryptosystem. But the proposed technique becomes more efficient as it generates comparatively less ciphertexts and requires operation with only one random number while re-encryption. Also, similar to [1], the proposed re-encryption technique is probabilistic and commutative[13].

B. Security analysis

The proposed technique is semantically secure as it is not feasible to reveal any information about a plain text given the corresponding cipher text. Let's denote the proposed technique as $CS = (KG, E, D, RE)$ and system parameters— message space ($\mathbf{M}$), encryption factors set ($\mathbf{R}$) and ciphertext space ($\mathbf{C}$) are generated according to a common security parameter k' . If an adversary is given the public/private key sets $(PK, SK) \leftarrow KG$ and computes $(C_1, C_2) = (E(m_1, r_1, PK), E(m_2, r_2, PK))$ from a chosen plain text pair (m_1, m_2) where $(r_1, r_2) \in \mathbf{R}$, the adversary can guess the random bit x for the pair (C_x, C_{1-x}) with probability at most negligibly larger than $1/2$. Similarly, in order to prove the re-encryption (RE) of the proposed cryptosystem semantically secure an experiment RE_{SS} is considered for an adversarial algorithm V which terminates on issuing an output of single bit. The adversary is allowed to compute ciphertexts using randomly generated key sets PK_1 and PK_2 which are then re-encrypted using RE . The adversary should not be able to distinguish between the two re-encryptions with non-negligible advantage.

Experiment: $RE_{SS_V}(CS, k')$

```

 $PK_1 \leftarrow KG; PK_2 \leftarrow KG;$ 
 $(m_1, m_2, r_1, r_2) \leftarrow V(PK_1, PK_2, \text{"ciphertexts specification"});$ 
if  $m_1, m_2 \notin \mathbf{M}$  or  $r_1, r_2 \notin \mathbf{R}$  then
    return NULL;
 $C_1 \leftarrow E(m_1, r_1, PK_1);$ 
 $C_2 \leftarrow E(m_2, r_2, PK_2);$ 
 $r'_1, r'_2 \in \mathbf{R};$ 
 $C'_1 \leftarrow RE(C_1, r'_1); C'_2 \leftarrow RE(C_2, r'_2);$ 
 $x \in \{0, 1\};$ 
 $x' \leftarrow V(C'_x, C'_{1-x}, \text{"guess"});$ 
if  $x = x'$  then
    return "1";
else
    return "0";

```

Note that RE is semantically secure for V with resources polynomial in k' if $P[RE_{SS_V}(CS, k') = "1"] - 1/2$ is negligible in k' . As the keys associated with ciphertext remain obscured in re-encryption process, the adversary fails to defeat the semantic security of re-encryption process in the proposed technique.

C. Mix-networks using proposed Universal re-encryption

With the proposed universal re-encryption cryptosystem sender i , for her message m , broadcasts a single cipher text pair *i.e.* $\{g^m h^{R_i}, h^{R'_i}\}$ where R_i and R'_i is a random number chosen by the sender and $(R_i, R'_i) \in \mathbb{Z}/n\mathbb{Z}$. After receiving a batch of such ciphertexts from senders, a random mix-node j , before transferring the batch to another random node, performs mixing operation on the batch *i.e.* re-encrypts each input of the batch, namely $\{g^m h^{R_i + R'_i r_j}, h^{R'_i r_j}\}$ with a random number $r_j \in \mathbb{Z}/n\mathbb{Z}$ and shuffles the resultant re-encrypted ciphers within the batch. The operation continues until all N mix-nodes complete their mixing and, thus, sends $\{g^m h^{R_i + R'_i r_j + \dots + R'_i r_j \dots r_N}, h^{R'_i r_j \dots r_N}\} = \{g^m h^{R_i^*}, h^{R_i'^*}\}$ to the receiver. Finally, the receiver gets the message decrypting as $D_{OU}(g^m h^{R_i^*}) = m$ by his private key p if $\{(h^{R_i'^*})^{p-1} \bmod p^2\} = 1$ is valid.

IV. EXPERIMENTAL ANALYSIS

This section presents the experimental results and their analysis to prove the efficiency of Okamoto–Uchiyama(OU) based universal re-encryption technique. The experiments involve simulation of ElGamal universal re-encryption mixnet and OU cryptosystem based universal re-encryption mixnet to show the performance improvement with the proposed re-encryption technique. Here, the performance of the mixnets is evaluated while varying the number of messages and the number of mix-nodes involved in the mixnet. Moreover, an experiment is also presented to prove the memory efficiency of OU based universal re-encryption technique while keeping the number of messages and mix-nodes stable.

A. Environment setup

To perform the experiments, an environment consisting of 64-bit Windows machine with 3.06 GHz CPU(Core i3) and 2048MB of RAM is used in which technical computing language Matlab is utilized for simulation purpose. For simplicity, to visualize their actual performance of the mixnets *i.e.* the re-encryption techniques, the addresses of senders and receivers are not considered while simulation and node-to-node travelling time is assumed to be negligible *i.e.*, computation times do not include the communication time. Moreover the different operations of privacy-related applications that are not related to cryptography are not considered. The mixnets are simulated using the key length of 1024 bits and keeping the message length fixed at 40 digits.

B. Principle of simulated mixnet

Let's consider a number of senders who wish to send messages through an externally anonymous channel to conceal the transferring information from everyone but the senders and receivers themselves. It is assumed that, all senders and receivers have distinct public/private key pairs and a sender knows only the public key of the intended receiver of her message. The simulated mixnets of the experiments here goes with the following principle-

Submitting the inputs; a sender, with the public key of intended receiver, encrypts her message and submits the ciphertext to the mixnet. Similar to ElGamal, the OU cryptosystem also provides semantic security to the submitted ciphertexts *i.e.* makes it unfeasible to differentiate from each other according to their plaintexts without the corresponding private keys. After receiving a pre-defined number of ciphertexts from the senders, mixnet proceeds to the next stage after forwarding the ciphertext batch to an arbitrary mix-node.

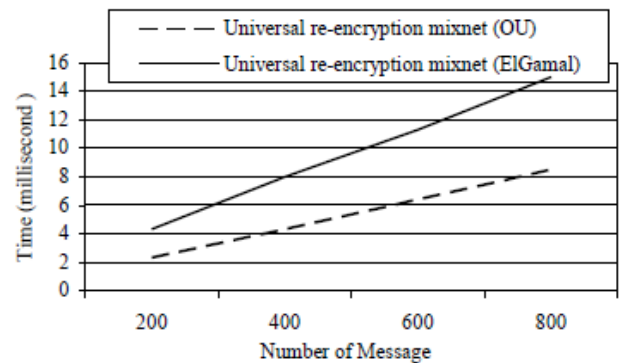


Figure 1. Time efficiency of OU based universal re-encryption technique while number of message varies (Y axis displays units in thousand).

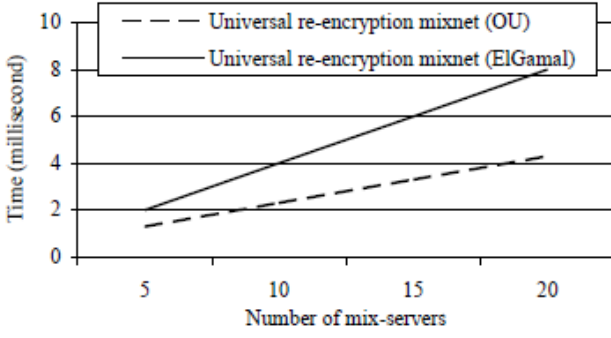


Figure 2. Time efficiency of OU based universal re-encryption technique while number of mix nodes varies (Y axis displays units in thousand)

Mixing; a complete batch of ciphertexts is transferred to an arbitrary node which performs mixing operation on its input *i.e.* re-encrypts and shuffles the ciphertexts of the batch. The shuffling operation here ensures unlinkability between senders and receivers.

Retrieving the outputs; each potential recipient tries to decrypt the mixed ciphertexts using his private key upon receiving the batch. Note that, key to ciphertext binding while encryption ensures a cipher text to be decrypted only under the corresponding secret key.

C. Experimental results and discussion

The experiment, result of which is shown in Fig. 1, is conducted to prove the time efficiency of the OU based universal re-encryption technique. Throughout the experiment, the number of messages varies while the size of the mixnet remains fixed with 5 nodes. It is seen from the result that time efficiency of mixnet with conventional ElGamal based universal re-encryption is almost half than that of OU based universal re-encryption technique. The reason of such result can be understood by analyzing the underlying cryptosystems utilized for corresponding mixnets. Encryption with ElGamal cryptosystem generates two ciphertexts while the OU cryptosystem generates only one. Hence in re-encryption, if ElGamal is exploited, a node has to perform individual operations on two ciphertexts. Whereas, OU cryptosystem relieves a node from the hassle of immense operations by reducing the number of ciphertexts. Moreover, the proposed technique requires operations with only one random number rather than two while re-encryption which, as a consequent, reduces the time requirement. Thus the overall operation of OU based universal re-encryption mixnet becomes simpler and faster than the conventional technique. The same reason stands for the second experiment that aims to observe the time efficiency of OU based universal re-encryption technique while varying the number of interim mix-nodes with fixed number of messages. The result of the experiment is shown in Fig. 2 where the batch size is kept fixed with 200 messages. Each message in the experiments described here, is a number of 40 digits.

The third experiment aims at revealing the memory efficiency of proposed OU based universal re-encryption technique. The experiment is performed by simulating a

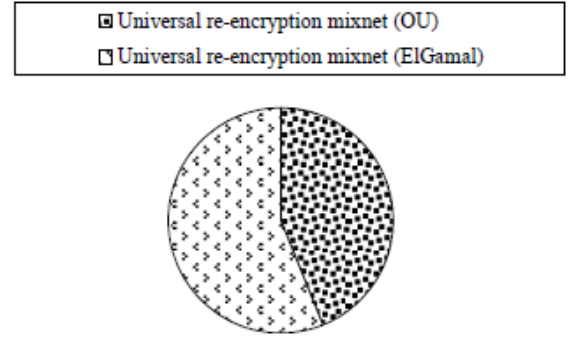


Figure 3. Memory efficiency of OU based universal re-encryption technique.

network of 5 mix-nodes with 200 messages. From the result of the experiment, as shown in Fig. 3, it is clear that the OU based universal re-encryption technique is memory efficient than ElGamal based universal re-encryption technique which is a consequence of reduced ciphertexts. The result shows that a batch while re-encrypted by the proposed OU based universal re-encryption technique, consumes only 70B of a node's memory whereas consumes about 90MB while re-encrypted by ElGamal based universal re-encryption technique. Note that, initial encryption in the proposed technique seems memory consuming as its operation includes public key $h = g^x$ ($n = p^2q$) but in ElGamal based technique the public key is $y = g^x$. Though p and q are large primes, x also need to be large to increase security. Moreover, a modification in [10] relieves p from being prime. Besides, the proposed OU based universal re-encryption technique achieves memory efficiency generating reduced ciphertext pair while re-encryption. Thus the proposed technique discharges mix-nodes from managing memory space for additional ciphertexts and allows them to be more light and inexpensive.

V. CONCLUSION AND FUTURE WORK

This paper proposes a universal re-encryption technique based on Okamoto-Uchiyama cryptosystem. The universal re-encryption technique with a public key cryptosystem *e.g.* ElGamal becomes less effective for resource constraint devices due to large amount of generated ciphertexts and non-negligible computational requirements. The proposed technique focuses on reducing the number of ciphertexts in order to decrease the resource consumption for computations. Hence, the technique adopts Okamoto-Uchiyama cryptosystem based on which it proposes a universal re-encryption technique. Similar to the conventional universal re-encryption, the proposed technique is half as efficient as its underlying cryptosystem and obscures the need of any keying information for re-encryption; hence, the keys are kept secret at sender and recipient. However, it is always possible for a receiver to identify any ciphertext heading to him throughout the mixing operation which sacrifices the privacy of the mix nodes and is an open research issue to solve. Moreover, efficient hybrid construction can also be a future research topic.

References

- [1] P. Golle, M. Jakobsson, A. Juels, and P. Syverson. "Universal reencryption for mixnets," in *2004 Cryptographers' Track RSA Conf.(CT-RSA)*, California, pp. 163–178.
- [2] N. Islam, K.M.R. Alam and A. Rahman, "The effectiveness of mixnets –an empirical study," *Comput. Fraud Security*, vol. 2013, no. 12, pp. 9–14, 2013.
- [3] S. Tamura, H. A. Haddad, N. Islam, and K. M. R. Alam, "An Incoercible E-Voting Scheme Based on Revised Simplified Verifiable Re-encryption Mix-nets." *Inform. Security Comput. Fraud*, vol. 3, no. 2, pp 32-38, 2015.
- [4] M. Klonowski, M. Kutyłowski and F. Zagórski, "Anonymous Communication with On-line and Off-line Onion Encoding," in *SOFSEM 2005: Theory and Practice of Computer Science*, Slovakia, 2005. pp 229-238.
- [5] C. Q. Huang, L. H. Li, and S. F. Fu. "An Improved Construction for Reusable Anonymous Return Channels Based on Universal Reencryption," in *2012 IEEE Int. Conf. Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*, Sanya, pp. 155-159.
- [6] N. Islam, "Certificate revocation in vehicular Ad Hoc networks: a novel approach," in *2nd Int. Conf. on Networking Syst. Security (NSysS)*, Dhaka, 2016, pp. 1-5.
- [7] N. Islam, M. K. Hossain, G. G. M. N. Ali, P. H. J. Chong "An Expedite Group Key Establishment Protocol for Flying Ad-Hoc Network(FANET)," in *5th Int. Conf. on Informatics, Electronics & Visionon*, Dhaka, 2016.
- [8] Q. Liu, C. C. Tan, J. Wu, and G. Wang. "Reliable re-encryption in unreliable clouds," in *IEEE Global Telecomm. Conf. (GLOBECOM)*, Texas, 2011, pp. 1-5.
- [9] E. Zielińska, W. Mazurezyk and K. Szczypiorski, "Trends in steganography," *Commun. ACM*, vol. 57, no. 3, pp. 86-95, Mar, 2014.
- [10] K. Nuida, N. Itakura and K. Kurosawa. "A Simple and Improved Algorithm for Integer Factorization with Implicit Hints," in *Topics in Cryptology-CT-RSA*, California, 2015, pp. 258-269.
- [11] T. Okamoto and S. Uchiyama. "A new public-key cryptosystem as secure as factoring," in *Advances in Cryptology-EUROCRYPT'98*, Espoo, 1998, pp. 308-318.
- [12] X. Guangli, Y. Benzhi, Z. Ping, "An Algorithm of Fully Homomorphic Encryption," in *9th IEEE Int. Conf. Fuzzy Systems and Knowledge Discovery(FSKD)*, Sichuan, 2012, pp. 2030–2033.
- [13] N. Islam, K. M. R. Alam, and S. S. R., "Commutative Re-encryption Techniques: Significance and Analysis," *Inform. Security J.: A Global Perspective*, vol. 24, no. 4-6, pp. 185-193, 2015.

A Survey on Use of Hybrid Fillers in Conductive Polymer Composites

Al Basir¹, Md. Mamunur Rahman², Md. Imran Khan³ and Md. Billal Hossain⁴

¹School of Engineering and Materials Science, Queen Mary University of London, United Kingdom.

^{2,4}Dept. of ECE, Khulna University of Engineering & Technology (KUET), Khulna-9203, Bangladesh.

³IICT, Bangladesh University of Engineering and Technology (BUET), Dhaka-1000, Bangladesh.

E-mail: ¹al.basir005@yahoo.com, ²mr.kuet@yahoo.com, ³imrankhankuet@gmail.com, ⁴billal.0709018@gmail.com

Abstract—Conductive polymer composite (CPC) is a very significant research topic in the area of self-regulating heating. It acts in self-regulated heating mode due to its positive temperature coefficient (PTC) effect. PTC effect is the consequence of breakup of conductive paths formed in CPC due to heat up of polymer composite. Currently a good number of researchers are going on conductive polymer composites. In this paper, the important aspects of conductive polymer composites; such as: conduction process, positive temperature co-efficient effect, negative temperature co-efficient etc has been analyzed. Moreover, a summary of the researches conducted by researchers in the fields of conductive polymer composites by employing hybrid fillers has also been discussed.

Keywords- Conductive polymer composite, positive temperature coefficient, negative temperature co-efficient, self-regulating heating, hybrid fillers.

I. INTRODUCTION

The uses of Conductive Polymer Composites (CPC) are increasing day by day because of their suitability to use in self regulated heaters. In Conductive Polymer composites conductive fillers dispersed in an insulating polymers matrix. Various types of metal and carbon powders may be used as conductive fillers. The interactions between the conductive fillers to form conductive paths play a vital role for the conductivity of composites. When the temperature of CPC reaches close to the melting temperature of polymer matrix then an immediate change in resistivity is observable due to disruption of conductive path formed by conductive fillers. This phenomenon is called Positive Temperature Co-efficient (PTC) effect which is an excellent property of CPC in order to control the temperature. The uses of CPCs in self regulated heaters have a number of advantages. These include: low energy consumption, no danger of overheating, high reliability and easy to design. Moreover, self regulated heaters are widely using in variety of applications, such as: melting of snow, heating of drum or cylinder, process temperature maintenance, de-icing of roof, soil warming, heating of long pipeline, hot water maintenance, prevention of freezing of fluids in pipes etc. In this paper, the author will give a summary of understanding on conductive polymer and the recent researches that are going on by using hybrid fillers in conductive composites.

This paper is described as follows. Section II describes self regulated heaters and section III shows conduction mechanism and percolation behavior of CPC. Effect of PTC, polymer

matrix, hybrid fillers, filler's size and shape and NTC are described in section III part a, part b, part c, part d and part e respectively. Finally section IV is for conclusion.

II. SELF-REGULATED HEATERS

Self-regulated heaters consist of two parallel bus wires embedded in a semi-conductive polymer composite forming a heating core with an insulating polymer jacket and at the same time an additional outer jacket may be applied in order to protect the self-regulated heater from hazard and corrosion. The variation of power output with change in temperature is a common phenomenon for self regulated heaters or cables. At high temperature, the conductive networks formed by the conductive fillers in heating core get disrupted due to volume expansion of polymer matrix which causes the reduction in heat output of self regulated heater. As soon as the temperature reduces, the distance between the conductive fillers decrease due to contraction of polymer matrix and consequently disrupted conductive networks form again which leads an increment of heat output. The mechanism of self-regulated heater is shown in figure 1.

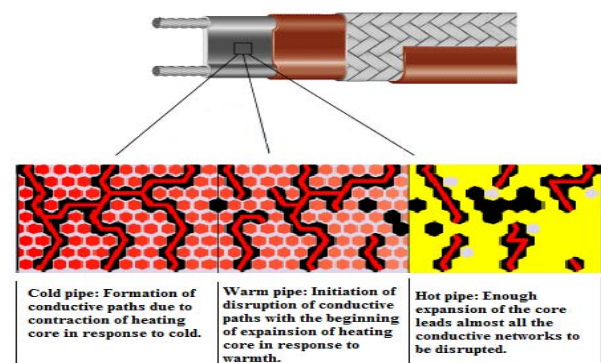


Fig. 1. Schematic of mechanism of self-regulated heater.

III. CONDUCTION MECHANISM AND PERCOLATION BEHAVIOR OF CPC

In order to explain conduction mechanism, Struimpler and Glatz-Reichenbach [1] stated that the average distance between the conductive fillers is large when the filler fraction is very low and thus the conductivity throughout the composite is negligible. When the amount of filler is increased sufficiently then the mean distance between the conductive fillers decreases and the consequence is the linkages between conductive fillers

and formation of initial conductive path through the entire material [1]. The corresponding filler content required to form the initial conductive path is designated as percolation threshold [1]. After the percolation threshold, conductivity of polymer composite can change dramatically for small variation of filler content [1]. At high filler concentration, the number of conductive paths increases significantly which forms a three dimensional network and in this state the conductivity of the material is remarkably high and minimum change in the conductance of polymer composite is observed for further increase in conductive filler [1]. The shape and the kind of spatial distribution of conductive filler in polymer matrix have certain effects on percolation behavior [2]. The method of preparation of composite also influences the percolation threshold behavior notably [3]. Mamunya et al. illustrated that when conductive filler having conductivity σ_f dispersed in polymer matrix having conductivity σ_p then the composite prepared in this way gains the conductivity value σ [2]. According to their explanation, the composite becomes conductive when volume filler fraction ϕ reaches the percolation threshold or critical value ϕ_c [2]. The change in conductivity below the percolation threshold is very small and in this state the conductivity of composite is almost equal to the conductivity of polymer σ_p [2]. A dramatic increase in the value of σ occurs typically from the value σ_c at percolation threshold to maximum σ_m when filler concentration reaches from critical value ϕ_c to filling limit F [2]. The dependence of conductivity on filler volume content is shown in figure 2.

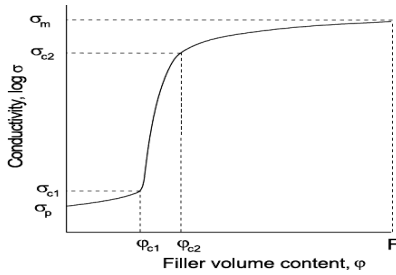


Fig. 2. Dependence of electrical conductivity on filler content [2].

Percolation theory gives phenomenological explanation while experimenting with conductive polymer composites and is employed to characterize the properties of materials. This theory is broadly utilized by researchers in order to investigate the conductive paths formed by conductive fillers within CPCs. The following equation is used to describe the dependency of conductivity of CPC on filler concentration [4]:

$$\sigma = \sigma_0 (P - P_c)^t \quad (1)$$

Where $P > P_c$

In Eq.1 σ_0 and P_c is known as scaling factor and percolation threshold respectively [4]. Basically at percolation threshold, the initial conductive path is formed by conductive fillers within CPC and after this point conductivity increases greatly due to add comparatively small amount of filler. Percolation threshold is an important aspect while experimenting with conductive filler because it gives the indication of initial conductive path throughout the polymer composite. Moreover, critical exponent is denoted by t in Eq.1 which is used to explain the criteria of conductive network within CPCs [4].

Basically t rely upon the dimension of conductive network and it is found that a number of studies have been made to determine a universal value for critical exponent t [1, 5, 6, 7]. For three dimensional network, Fisch and Harris calculated the value of t was 1.95 while Kirkpatrick [6] reported that $t = 1.60$ [6,7,8]. Finally, it is experimentally agreed that for two dimensional lattices the value of t is 1-1.3 whereas for three dimensional lattices the calculated value of t varies from 1.6 to 2 [4, 9, 10, 11]. In conductive polymer composites, conductive fillers are dispersed in polymer matrices and often obstructions exist within fillers. Therefore, it is experimentally evident that in CPCs, electron tunneling happens through barriers [1]. Conductivity in CPCs occurs due to direct contact between conductive fillers and electron tunneling [1]. Thus the amount of conductive fillers and interaction between them play imperative role in CPC. Adequate amount of filling of conductive fillers within CPC leads the fillers to be in contact with each other and in this situation conduction of charge carriers occurs through the conductive network formed by the conductive fillers within polymer matrix [1]. Stru'mpler and Glatz-Reichenbach described that when the amount of conductive fillers within polymer matrix is below the percolation threshold then the mean distance between conductive particles are high and therefore there is no possibility of formation of conductive path [1]. Electron tunneling does not occur within conductive fillers while the mean separation distance is more than 10 nm [1]. It is also reported that electron tunneling is possible while the mean separation distance between filler particles are less than 10nm [1, 12]. Tunneling resistance is another important factor that arises while tunneling of electron occurs because of having barriers within conductive fillers in CPCs and the distinctive nature of tunneling resistance is its resistivity temperature dependence [8,13]. Sheng et al. [14] introduced Fluctuation Induced Tunneling (FIT) model to explain conduction mechanism in carbon filled polyvinylchloride (PVC) composites. It is observed that this FIT model established by Sheng et al. and it is supported by many researchers such as Zhang [8]; Kilbrid et al. [15].

A. Positive Temperature Coefficient (PTC) Effect

The most important characteristic of conductive polymer composite is positive temperature coefficient effect. Frydman invented the PTC effect of CPC successfully in 1945 [16]. Kohler [17] proposed that PTC effect is the consequence of thermal expansion. Kohler also suggested that conductive particles spread within polymer matrix form chain like conductive network and as the material is heated, the disruption of conductive network due to expansion of polymer matrix increases the resistivity of the composite which defines the positive temperature coefficient (PTC) effect [17]. Ohe and Naito [18] explained that small gap width within conductive fillers allow them for large-scale tunneling when the temperature is relatively low but this situation does not remain same when the temperature is comparatively high. At high temperature, the inter-particle gap width increases due to expansion of polymer matrix and the consequence is the difficulty in electron tunneling between conductive particles. As a result, resistivity increases distinctly though the gap width within conductive fillers does not change notably.

B. Effect of Polymer Matrix

Polymer matrix is responsible for occurring positive temperature coefficient effect in CPC. The expansion of polymer matrix at its melting temperature plays a significant role for the happening of PTC effect. A lot of researches have been done by using polyethylene as polymer matrix [3, 19-25]. Usually polymer molecules are partially crystalline (or semi crystalline), having crystalline region dispersed in the amorphous region. In crystalline region molecules are arranged in regular order where in amorphous region molecules are randomly organized and crystallinity indicates the amount of crystalline region in polymer in correspondence with amorphous region [26]. Often various polymer properties such as: melting point, stiffness, hardness etc. are influenced by crystallinity [26]. A schematic diagram of crystalline-amorphous mixed macromolecular structure of polymer has been given in figure 3.

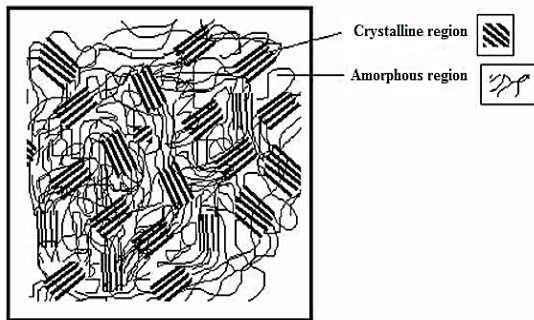


Fig. 3. Schematic of crystalline-amorphous mixed structure of polymer [26].

Kono et al. [27] used Nickel (Ni) particle dispersed poly (vinylidene fluoride) (PVDF) to find out the positive temperature coefficient effect of resistivity. Experimentally they found the percolation threshold of Ni content was 19 vol% when composites with Ni content respectively 20, 30, 40 and 50 vol% were used in experiment [27]. Kono et al. [27] proved that for composites with Ni content 20 and 30 vol % PTC effect can be observed below the melting temperature of PVDF matrix whereas composites with Ni content 40 and 50 vol% shows PTC effect close to the melting temperature of matrix. Rybak et al. [28] showed an analysis on new conductive polymer composite where silver nanoparticles (Ag) were dispersed in homo and heterogeneous polymer blends of high density polyethylene (HDPE), polybutylene terephthalate (PBT) and poly (m-xylene adipamide) (PAMXD). They found that the use of immiscible polymer blends leads the reduction of percolation threshold twice that not only reduce the cost of preparation of CPC but keep the mechanical properties better as well [28]. Rybak et al. [28] also observed a sharp and strong increase of resistivity in this type of conductive polymer composite. Thus they comment that low cost and better PTC intensity make such composites promising for using as temperature sensors and current limits devices. Feng and Chan [29] employed polypropylene (PP) and ultra-high molecular weight polyethylene (UHMWPE) as polymer matrix and carbon black (CB) as conductive filler to prepare polymer composites. Feng and Chan [29] experimentally observed two PTC amplitudes for UHMWPE/PP/CB composites which were due to thermal volume expansion related with PP and

UHMWPE. Feng and Chan also conducted their investigation to observe the NTC behavior of UHMWPE/PP/CB composites and they claimed that the NTC effect of such composites was removed by employing a high viscosity polymer as one of the components [29]. Xi et al. [25] used ultra-high molecular weight polyethylene (UHMWPE) and low molecular weight polyethylene (LMWPE) in their experiment. Xi et al. [25] showed that it is possible to achieve a deservng PTC amplitude by preparing composites consist of UHMWPE, LMWPE and carbon fiber. Experimentally they found that it is possible to eliminate NTC effect by implementing a blend of UHMWPE and LMWPE as polymer matrix [25].

C. Effect of Hybrid Fillers in Conductive Polymer Composite

Conductive filler is one of the major constituent of conductive polymer composites and a huge number of researches have been carried out by utilizing single filler in CPCs. Here single filler denotes the use of one particular type of conductive filler in CPCs. Recently a good number researchers are employing hybrid fillers in conductive polymer composites to investigate their performance and it has been reported that the use of hybrid fillers displays a huge influence in the conductivity and the PTC effect of polymer composites. Liang et al. [30] explained that addition of multi-wall carbon nanotubes increase the conductivity of polypropylene (PP) reinforced silver (Ag) nanoparticles. They described that PP/Ag composites show poor conductivity due to insufficient conductive paths and an addition of low amount of MWNT increases the conductivity of the composite gradually [30]. They also found that as soon as MWNT reaches 2 wt. % then the conductivity enhances by almost four orders of magnitude [30]. Liang et al. [30] claimed that MWNTs successfully form bridges among Ag nanoparticles and therefore increments of numbers of conductive paths ultimately increase the conductivity of PP/Ag/MWNT composites.

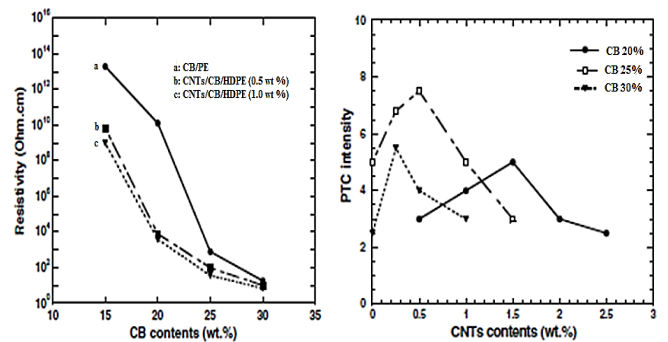


Fig. 4. Resistivity of nanocomposites with and without adding MWNTs at room temperature and (left) and PTC intensities of hybrid composites with different CB and MWNT contents(right) [22].

The Lee et al. [22] examined the effect of addition of MWNTs in CB filled HDPE composites. Lee et al. [22] prepared composites by incorporating and without incorporating nanotubes in order to investigate their resistivity as a function of carbon black content which is shown in figure 4(a). In the case of CB filled HDPE composites, CPCs behave as an insulator when CB filler content is below 15 wt. % and a dramatic decrease in resistivity occurs when the range of CB

filler content is between 20 wt. % and 25 wt. % [22]. These phenomena illustrate that for conduction of charge carries within CPCs, a minimum loading of conductive fillers are needed in order to make them capable to tunnel. It is also explained by Lee et al. [22] that because of adding of small amount of MWNTs, the domain of sharp decline of resistivity is shifted from previous and the range is between 15-20 wt. % which is observable in figure 4(a). In Fig 4(b), Lee et al. [22] not only showed the switching of peak intensity to higher CNT content with reduction of CB content but also showed the deterioration of intensity after those peak points. Lee et al. [22] stated that incorporation of excess amount of MWNTs in HDPE/CB composites does not lead them to take part in network and the consequence is the reduction of final resistivity. Li et al. [31] implemented similar hybrid conductive fillers like Lee et al. [22] during their experimental investigation on PTC behavior of conductive polymer composites. They used carbon black (CB) and carboxylated multi-walled carbon nanotubes (c-MWNTs) in the polymer blends of polypropylene (PP) and ultra-high molecular weight polyethylene (UHMWPE) in their study [31]. Li et al. [31] experimentally observed that addition of 0.5 wt. % c-MWNT within PP/UHMWPE/CB not only lower the room temperature resistivity of composites but also it was found that PP/UHMWPE/CB/c-MWNT composites exhibit higher PTC intensity than PP/UHMWPE/CB. According to Li et al. [31], the reinforcement of MWNTs within PP/UHMWPE/CB composites can bridge the CB particles which are not connected and the consequence is the continuity of conductive paths and the reduction of percolation threshold. Dang et al. [32] showed that the use of hybrid fillers has substantial influence on PTC behavior and they proved it by employing high-aspect-ratio carbon fiber as second filler.

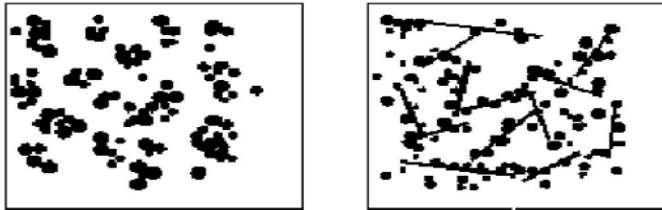


Fig. 5. Schematic of formation of bridge network for Only MNCB(left) and for conductive paths in composites connected by MNCB and CF(right) [32].

In their experiment composites were prepared by adding modified nanosclad carbon black (MNCB) and carbon fiber (CF) with high-density polyethylene (HDPE) and polypropylene (PP) [32]. Dang et al. [32] described that in the experiment 1.0 wt. % titanate coupling agent (NDZ-102) was employed to modify the surface of CB and to do so at first this titanate coupling agent was dissolved in the solvent of isopropanol. Dang et al. [32] found out that CF added MNCB/HDPE and MNCB/PP composites has displayed higher conductivity compared to composites filled with MNCB and they explained that the reason of exhibiting high conductivity is the capability of CF to bridge the unconnected MNCB conductive fillers and thus build a strong conductive network within composites. The schematic of formation of bridge network is shown in figure 5. Shen et al. [33] conducted almost similar experiment like Dang et al. [32]. Shen et al. [33] used

carbon black (CB) and carbon fiber (CF) as conductive fillers while high density polyethylene (HDPE) and polypropylene (PP) as polymer. In their experiment, conductive polymer composites were made by HDPE with CB and CF and at the same time polymer blend of PP and HDPE filled with CB and CF [33]. These two types of conductive fillers can be designated as HDPE/CB/CF and HDPE/PP/CB/CF. Shen et al. [33] stated that for composites with hybrid fillers (CB/CF), the content of CF was kept unchanged and that was 2 wt %. It is also reported by Shen et al. [33] that for polymer composites where CB and CF conductive fillers were mixed with polymer blends (HDPE/PP), the weight ratio of HDPE to PP was 60/40. Shen et al. [33] experimentally found that the CB/CF filled HDPE and HDPE/PP composites exhibit higher PTC intensities than composites filled with only CB conductive fillers.

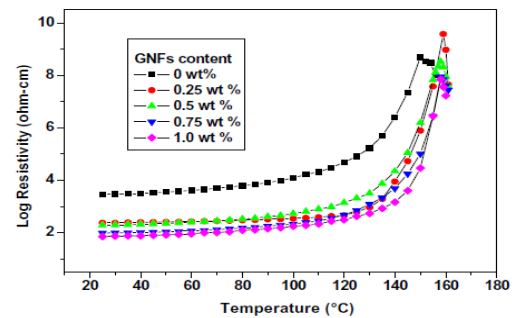


Fig. 6. The resistivity as a function of temperature for CB filled HDPE composites with various wt. % of GNFs contents [23].

Zheming et al. [34] carried out their experiment on conductive polymer composites which was based on hybrid fillers named graphite oxide (GO) and multi-walled carbon nanotubes (MWNTs). In their experiment polymer composites were made via melt blending polypropylene (PP) with hybrid fillers GO-MWNTs and only MWNTs which can be designated as PP/GO-MWNTs and PP/MWNTs respectively [34]. Zheming et al. [34] observed lower electrical resistivity and percolation threshold for PP/GO-MWNTs composites while they compare it with PP/MWNTs composites. Zheming et al. [34] claimed that the reason of improved electrical properties for PP/GO-MWNTs composites is the implementation of hybrid conductive fillers that has made effective conductive network within composites. Li et al. [23] used composites of graphite nanofibers (GNF) and carbon black (CB) incorporated in HDPE to analysis the PTC behavior and they claimed that reinforcement of GNF within composites successfully increase the PTC intensity. Li et al. [23] experimentally observed that CB filled HDPE composite without the incorporation of GNFs exhibited PTC intensity 5.229 while 0.25 wt% GNF reinforced HDPE/CB composite showed the maximum PTC intensity with the value 7.212. Li et al. [23] explained that the increment of GNFs within polymer composites reduce the room temperature resistivity as shown in figure 6. Kar and Khatua [35] investigated the influence of coefficient of thermal expansion (CTE) of polymer matrix and conductive filler on PTC behavior of poly (methyl methacrylate) PMMA/Silver-coated glass bead composites. Kar and Khatua [35] observed a quick increase in resistivity at a temperature of 115 °C for PMMA/CB (10 wt %) composites

whereas the glass transition temperature (T_g) of PMMA is around 113 °C. For PMMA/CB composites, the difference in resistivity with temperature at various heating cycles is shown in figure 7(a). On the other hand, Kar and Khatua [35] also observed that for composites of PMMA/Ag-coated glass bead (40 wt %), the trip temperature of PTC was approximately 92 °C which was far below the glass transition temperature of PMMA which is shown in figure 7(b). Stability in room temperature resistivity and PTC trip temperature was also observed for PMMA/Ag-coated glass bead composites at different heating cycles which are shown in Fig 7(b). Kar and Khatua [35] explained that the difference in PTC transition between the PMMA/CB and PMMA/Ag-coated glass bead composites signifies that PTC effect within PMMA/Ag-coated glass composite is not dominated by the T_g of PMMA. Kar and Khatua [35] experimentally found that addition of 1 phr of nanoclay within PMMA/CB composites has enhanced the PTC trip temperature to around 120 °C while almost same phenomena is shown by PMMA/Ag-coated glass composites where the PTC trip temperature is raised to 98 °C.

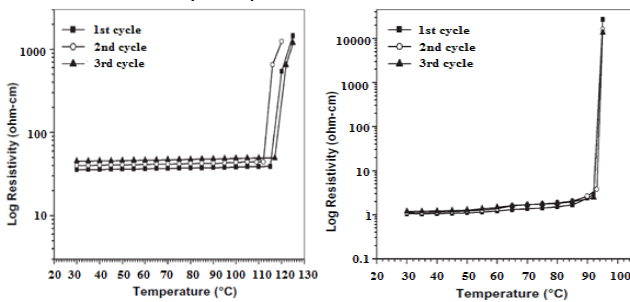


Fig. 7. Resistivity as a function of temperature at three consecutive heating cycles for CB (10 wt %) filled PMMA composites (left) and for Ag-coated glass bead (40 wt %) filled PMMA composites(right) [35].

D. Effect of Size and Shape of Conductive Filler

It has been showed by various studies that the size and shape of conductive filler has huge influence on percolation behavior and PTC effect of conductive polymer composites. Therefore, the effect of size and shape of conductive fillers have got adequate importance in the study of CPC. Luo and Wong [36] showed that size of conductive filler has immense influence on PTC behavior. In experiment they used composites made of five different carbon blacks named N660, Conductex 975, BP 2000, Culcan XC72 and Conductex SC filled in high density polyethylene (HDPE) and they observed that larger conductive filler N660 filled HDPE exhibited higher PTC intensity in comparison with smaller conductive filler BP 2000 filled HDPE [36]. Luo and Wong [36] explained that smaller particle size BP 2000 carbon blacks formed strong conductive network within composite because of higher in number of carbon black particles in contact with each other and due to thermal expansion of polymer matrix very few conductive paths were disrupted. They also explained that for the composite filled with N600 carbon blacks the arrangement of conductive paths within larger sized conductive fillers were such that conductive paths broke easily due to very small expansion of polymer matrix and the resultant was the higher PTC amplitude [36]. Mamunya et al. [2] considered irregular, dendritic shaped Copper (Cu) and sphere shaped Nickel (Ni) fillers in order to study the electrical and thermal properties of

composites. Epoxy resin (ER) and poly (vinyl chloride) (PVC) filled with these metal powders were used in their experiment. They described packing factor F as a very important parameter as this parameter fluctuate with the variation of the shape of conductive filler and experimentally it was found that the packing factor of Cu was 0.30 where the value was 0.51 for Ni [2]. In the experiment, they observed that the critical volume content or percolation threshold value of sphere shaped Ni filled ER-Ni system is higher than Cu filler filled ER-Cu and PVC-Cu systems and therefore they claimed that percolation threshold value of CPC depends on shape of conductive fillers and considerably related with packing factor F [2]. Hence it can be said that shape of conductive fillers has huge impact on percolation threshold value as well as the conductivity of polymer composite.

E. Negative Temperature Coefficient (NTC) Effect

The applications of CPCs are increasing day by day because of its PTC effect but polymer composites show another effect which is designated as Negative temperature coefficient (NTC) effect. NTC effect is an unacceptable property of CPC which is usually occurred after the effect of PTC and various approaches have been taken by researches to eliminate this. NTC effect is basically observed after the melting point of polymer matrix due to reagglomeration of filler particles within CPC and the consequence is the decrease of resistivity within conductive polymer composite with increasing temperature which is highly undesirable. An indication of NTC effect is shown in figure 8 where a sharp reduction of resistivity can be contemplated just after the melting point of high density polyethylene. One of the well-known methods to remove NTC effect is cross-linking of polymer matrix. It is found that there are several physical and chemical procedures which have been implemented to achieve this cross-linking [37, 38].

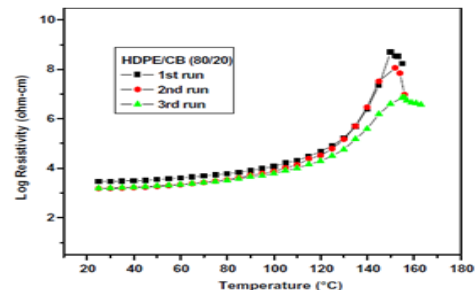


Fig. 8. NTC effect of HDPE/CB (80/20) composites for three heating cycles [23].

IV. CONCLUSION

In this paper, an overview on conductive polymer composites and its various phenomena has been presented. Conductive polymer composites (CPC) are extensively using in self regulated heaters and manufacturers want to improve the quality and reliability of self regulated heaters by implementing suitable PTC intensity because of their increasing demand in industrial and residential applications. A good number of researches have already been carried out to find out convenient PTC intensities of composites by applying different conductive fillers and polymers. Recently it has been found that there are huge scopes of researches on conductive

polymer composites by employing hybrid fillers in them and some reports have confirmed that the use of hybrid fillers displays a huge influence in the conductivity and the PTC effect of polymer composites.

REFERENCES

- [1] Stru"mpler, R. and Glatz-Reichenbach, J., "FEATURE ARTICLE Conducting Polymer Composites," *Journal of Electroceramics*, vol. 3, no. 4, pp. 329-346, 1999.
- [2] Mamunya, Y. P., Davydenko, V. V., Pissis, P. and Lebedev, E. V., "Electrical and thermal conductivity of polymers filled with metal powders," *European Polymer Journal*, vol. 38, no. 9, pp. 1887-1897, 2002.
- [3] Lisunova, M. O., Mamunya, Y. P., Lebovka, N. I. and Melezhyk, A. V., "Percolation behaviour of ultrahigh molecular weight polyethylene/multi-walled carbon nanotubes composites," *European Polymer Journal*, vol. 43, no. 3, pp. 949-958, 2007.
- [4] Bauhofer, W. and Kovacs, J. Z., "A review and analysis of electrical percolation in carbon nanotube polymer composites," *Composites Science and Technology*, vol. 69, no. 10, pp. 1486-1498, 2009.
- [5] Zallen, R., *The physics of Amorphous Solids*, John Wiley & Sons, Inc., 1998.
- [6] Kirkpatrick, S., "Percolation and Conduction," *Reviews of Modern Physics*, vol. 45, no. 4, pp. 574-588, 1973.
- [7] Fisch, R. and Harris, A. B., "Critical behavior of random resistor networks near the percolation threshold," *Physics Review B (Condensed matter)*, vol. 18, no. 1, pp. 416-420, 1978.
- [8] Zhang, R., "Conductive TPU/CNT composites for strain sensing," PhD. dissertation, School of Engineering and Materials Science, Queen Mary, University of London, UK, 2009.
- [9] Liu, F., Zhang, X., Li, W., Cheng, J., Tao, X., Li, Y. and Sheng, L., "Investigation of the electrical conductivity of HDPE composites filled with bundle-like MWNTs," *Composites Part A: Applied Science and Manufacturing*, vol. 40, no. 11, pp. 1717-1721, 2009.
- [10] Zhang, R., Dowden, A., Deng, H., Baxendale, M. and Peijs, T., "Conductive network formation in the melt of carbon nanotube/thermoplastic polyurethane composite," *Composites Science and Technology*, vol. 69, no. 10, pp. 1499-1504, 2009.
- [11] Vionnet-Menot, S., Grimaldi, C., Maeder, T., Str ssler, S. and Ryser, P., "Tunneling-percolation origin of nonuniversality: Theory and experiments," *Physical Review B*, 71:064201, 2005.
- [12] Van Beek, L. K. H. and Van Paul, B. I. C. F., "Internal field emission in carbon black-loaded natural rubber vulcanizates," *Journal of Applied Polymer Science*, vol. 6, no. 24, 1692.
- [13] Sichel, E. K., Carbon black-polymer composites: The physics of electrically conducting composites (Plastics engineering). New York : M. Dekker, 1982.
- [14] Sheng, P., Sichel, E. K. and Gittleman, J. I., "Fluctuation-Induced Tunneling Conduction in Carbon-Polyvinylchloride Composites," *Physical Review Letters*, vol. 40, pp. 1197-1200, 1987.
- [15] Kilbride, B. E., Coleman, J. N., Fraysse, J., Fournet, P., Cadek, M., Drury, A., Hutzler, S., Roth, S. and Blau, W. J., "Experimental observation of scaling laws for alternating current and direct current conductivity in polymer-carbon nanotube composite thin films," *Journal of Applied Physics*, vol. 92, no. 7, pp. 4024, 2002.
- [16] Frydman, E., U.K. patent specification 604 695 I 718 14S, 1948.
- [17] Kohler, F., Resistance element, U.S. Patent 3,243,753, 1966.
- [18] Ohe, K. and Naito, Y., "New resistor having an anomalously large positive temperature coefficient," *Japanese Journal of Applied Physics*, vol. 10, no. 1, pp. 99-108, 1971.
- [19] Bai, B. C., Kang, S. C., Im, J. S., Lee, S. H. and Lee, Young-Seak., "Effect of oxyfluorinated multi-walled carbon nanotube additives on positive temperature coefficient/negative temperature coefficient behavior in high-density polyethylene polymer switches," *Materials Research Bulletin*, vol. 46, no. 9, pp. 1391-1397, 2011.
- [20] Isaji, S., Bin, Y. and Matsuo, M., "Electrical conductivity and self-temperature-control heating properties of carbon nanotubes filled polyethylene films," *Polymer*, vol. 50, no. 4, pp. 1046-1053, 2009.
- [21] Krupa, I., et al. (2007). "Electrically conductive composites of polyethylene filled with polyamide particles coated with silver," *European Polymer Journal*, vol. 43, no. 6, pp. 2401-2413, 2007.
- [22] Lee Joong-Hee, Kim, S. K. and Kim, N. H., "Effects of the addition of multi-walled carbon nanotubes on the positive temperature co-efficient characteristics of carbon-black-filled high-density polyethylene nanocomposites," *Scripta Materialia*, vol. 55, no. 12, pp. 1119-1122, 2006.
- [23] Li, Q., Siddaramaiah, Kim, N. H., Yoo Gye-Hyoung and Lee, J. H., "Positive temperature coefficient characteristic and structure of graphite nanofibers reinforced high density polyethylene/carbon black nanocomposites," *Composites: Part B*, vol. 40, no. 3, pp. 218-224, 2009.
- [24] Thongruang, W., Spontak, R. J. and Balik, C. M., "Bridged double percolation in conductive polymer composites: an electrical conductivity, morphology and mechanical property study," *Polymer*, vol. 43, no. 13, pp. 3717-3725, 2009.
- [25] Xi, Y., Ishikawa, H., Bin, Y. and Matsuo, M., "Positive temperature coefficient effect of LMWPE-UHMWPE blends filled with short carbon fibers," *Carbon*, vol. 42, no. 8-9, pp. 1699-1706, 2004.
- [26] Hegde, R. R., Kamath, M. G. and Dahiya, A. (2004). *Polymer Crystallinity*. [Online] Available FTP: <http://web.utk.edu/~mse/Textiles/Polymer%20Crystallinity.htm> [Accessed: 15 November 2012].
- [27] Kono, A., Shimizu, K., Nakano, H., Goto, Y., Kobayashi, Y., Ougizawa, T. and Horibe, H., "Positive temperature-coefficient effect of electrical resistivity below melting point of poly(vinylidene fluoride) (PVDF) in Ni particle-dispersed PVDF composites," *Polymer*, vol. 53, no. 8, pp. 1760-1764, 2012.
- [28] Rybak, A., Boiteux, G., Melis, F. and Seytre, G., "Conductive polymer composites based on metallic nanofiller as smart materials for current limiting devices," *Composites Science and Technology*, vol. 70, no. 2, pp. 410-416, 2010.
- [29] Feng, J. and Chan Chi-Ming., "Double positive temperature coefficient effects of carbon black-filled polymer blends containing two semicrystalline polymers," *Polymer*, vol. 41, no. 12, pp. 4559-4565, 2000.
- [30] Liang, G. D., Bao, S. P. and Tjong, S. C., "Microstructure and properties of polypropylene composites filled with silver and carbon nanotube nanoparticles prepared by melt-compounding," *Materials Science and Engineering B*, vol. 14, no. 2-3, pp. 55-61, 2007.
- [31] Li, Q., Basavarajaiah, S., Kim, N. M., Heo, Seok-Bong and Lee, J. H. (2010). "Synergy effect of hybrid fillers on the positive temperature coefficient behavior of polypropylene/ultra-high molecular weight polyethylene composites," [Online] Available from: <http://dx.doi.org/10.1002/app.31478> [Accessed: 26 July 2013].
- [32] Dang Zhi-Min, Li Wei-Kang and Xu Hai-Ping. (2009). "Origin of remarkable positive temperature coefficient effect in the modified carbon black and carbon fiber cofilled polymer composites," *Journal of Applied Physics*, [Online] Available: <http://link.aip.org/link/?JAP/106/024913/1> [Accessed: 30th November 2012].
- [33] Shen, L., Wang, F. Q., Yang, H. and Meng, Q. R., "The combined effects of carbon black and carbon fiber on the electrical properties of composites based on polyethylene or polyethylene/polypropylene blend," *Polymer testing*, vol. 30, no. 4, pp. 442-448, 2011.
- [34] Zheming, G., Chunzhong, L., Gengchao, W., Ling, Z., Qilin, C. and Xiaohui, L., "Electrical properties and morphology of highly conductive composites based on polypropylene and hybrid fillers," *Journal of Industrial and Chemistry*, vol. 16, no. 1, pp. 10-14, 2010.
- [35] Kar, P. and Khatua, B. B. (2011). "Highly reversible and repeatable PTCR characteristics of PMMA/Ag-coated glass bead composites based on CTE mismatch phenomena," *Polymer Engineering and Science* [Online] Available: <http://dx.doi.org/10.1002/pen.21966> [Accessed: 30th June 2013].
- [36] Luo, S. and Wong, C. P., "Conductive Polymer Composites with Positive Temperature Coefficient," 1999 *International Symposium on Advanced Packaging Materials*, Braselton, GA, 14-17 March 1999.
- [37] Seo Min-Kang, Rhee Kyong-Yop and Park Soo-Jin., "Influence of electro-beam irradiation on PTC/NTC behaviors of carbon blacks/HDPE conducting polymer composites," *Current Applied Physics*, vol. 11, no. 3, pp. 428-433, 2011.
- [38] Tsao Keng-Yu, Tsai Ching-Shan and Huang Chi-Yuan., "Effect of argon plasma treatment on the PTC and NTC behaviors of HDPE/carbon black/aluminum hydroxide nanocomposites for over-voltage resistance positive temperature coefficient (PTC)," *Surface and Coatings Technology 205, Supplement*, vol. 1, no. 0, pp. S279-S285, 2010.

A Low Cost MEMS and Complementary Filter Based Attitude Heading Reference System (AHRS) for Low Speed Aircraft

Md.Saiful Islam¹, Md.Shajid-UI-Mahmud¹

¹Aeronautical Engineering Department

⁵Electrical, Electronic and Communication Engineering
Department

Military Institute of Science and Technology
Dhaka, Bangladesh

Tariqul Islam¹, Md. Syedul Amin², Md Hossam-E-Haider³

²Bangladesh Air Force
Dhaka, Bangladesh

Abstract— AHRS is an important system that provides the 3-dimensional orientation (roll, pitch and yaw or heading) of an aircraft from which the flying performance can be evaluated. The avionics of aircraft are very sophisticated and majority of them do not allow any external interface for logging. As such, a standalone external AHRS is needed which should be comprised of low cost devices so that it becomes affordable for other aircrafts. This paper proposes a low cost Micro-Electrical-Mechanical system (MEMS) based AHRS for low speed aircraft to evaluate the flying performance. MEMS based 3-axis accelerometer, gyroscope and magnetometer have been used to build the AHRS. Low cost MEMS based sensors suffer from bias, drift and sudden spike. Low pass filter has been utilized to reduce the spikes. Complementary filter has been used to pin down the drift of the gyroscope from accelerometer reading. The system has been implemented in Arduino environment. Simulation and real time results have shown correct functioning of the system. Besides using as an AHRS in low speed aircraft, the system can be used for control and stabilization of other real time system.

Keywords— AHRS; Euler angle; Gyroscope; Accelerometer; Complementary filter Introduction.

I. INTRODUCTION

AHRS is a vital part of an aircraft. It provides the flight dynamics data which can contribute to estimate the aircraft performance during a flight [1-2]. However, getting the data from the AHRS of the aircraft is very difficult as the avionics system of an aircraft is very sophisticated and normally does not allow external device to be connected. As such, an external AHRS is a better way without interfering the aircraft's existing avionics [3].

Inertial Measurement Unit (IMU) is the building block of an AHRS. The price of the AHRS largely depends on the type and the quality of the IMU and varies from few hundred dollars to millions according to their accuracy and reliability. The recent advances in MEMS lead to very inexpensive INS. But accumulation of the bias, drift and other errors reduces the accuracy of the AHRS built from the MEMS based IMU [4].

An IMU generally comprises of three axis gyroscopes, accelerometers and magnetometers for measuring the 3D orientation (roll, pitch and yaw angle). Generally the gyroscopes measure the orientation from changing angular

velocity. But the low cost MEMS based gyroscopes are prone to drift as it cannot detect the rotation of the earth [5]. On the other hand, low cost MEMS based accelerometers suffer from bias and sudden spikes which need to be filtered appropriately [6].

The limitation of the MEMS based sensors can be overcome by sensor fusion. Accelerometer can be utilized to pin down the drift of the gyroscope. On the other hand, bias and sudden spike of the accelerometer can be reduced from gyroscope as it does not suffer from bias. However, appropriate filters are needed for the sensor fusion. The complementary filter is a widely used and less complex fusion technique where the complement is the desired result [7].

This paper proposes a standalone AHRS utilizing low cost MEMS based three axis accelerometers, gyroscopes and magnetometers using a complementary filter to determine the flying performance of a low speed aircraft.

II. METHODOLOGY

The attitude of a flying object can be expressed in the inertial coordinate frame in various ways which includes direction cosine matrix, Euler angles and quaternions. Direction cosine matrix is simple but requires nine parameters. Quaternion needs less parameters but needs more hardware resource. Compared to quaternion, Euler Angles are simple and intrinsic to analysis and control. As such, Euler angles are used due to its intrinsic nature, simple sensor fusion algorithm and nonlinearity. However, it suffers from Gimbal lock [8]. But due to the low maneuverability of low speed air craft near ± 90 degrees pitch angle, the Gimbal lock is not likely to occur. The rotational sequence for representing an orientation is yaw, pitch and lastly roll angle.

Euler angle provides a 3D orientation by using combination from three axes rotations as shown in Fig 1. For describing the orientation several coordinate frames exists such as Inertial frame, Body frame, Navigational frame. The axes of inertial frames are earth-fixed and axes of body frames are aligned with sensors as shown in Fig. 1.

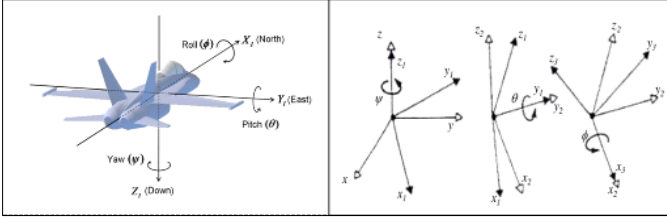


Fig. 1. Roll, pitch, yaw angle of a rotating object

An inertial frame is the only frame in which Newton's law of motion is valid. It is the frame of absolute zero motion fixed relative to star and consist of a set of mutually perpendicular axes that neither accelerate nor rotate with respect to inertial space. It is used as fixed reference which comprises a set of earth fixed axes. The x, y, z axis points refer as north, east and down accordingly. The body frame is a type of coordinate reference frame whose origin is at the center of mass of a vehicle and aligned with the sensor body. The x-axis of the body frame represents the nose direction; y-axis gives the direction of right side of fuselage and z-axis for the bottom of fuselage.

A. Rotation Matrix

To obtain a proper rotational matrix, at first the yaw rotation is changed. The z-axis is aligned with the inertial frame and the x and y axes are rotated according to the yaw angle P. The rotation matrix is given in (1).

$$P = \begin{bmatrix} \cos(x) & \sin(x) & 0 \\ -\sin(x) & \cos(x) & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

The rotation matrix with pitch and yaw is given by (2).

$$Q = \begin{bmatrix} \cos(y) & 0 & -\sin(y) \\ 0 & 1 & 0 \\ \sin(y) & 0 & \cos(y) \end{bmatrix} \quad (2)$$

The rotation matrix for moving from the inertial frame to the vehicle frame consists of the yaw matrix and multiplied by the pitch matrix as given in (3).

$$R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(z) & \sin(z) \\ 0 & -\sin(z) & \cos(z) \end{bmatrix} \quad (3)$$

The complete rotation matrix is given by (4).

$$M(z, y, x) = R(z) Q(y) P(x) \quad (4)$$

where

$$M = \begin{bmatrix} c(x)c(y) & c(y)s(x) & -s(y) \\ c(x)s(z)s(y) - c(z)s(x) & c(z)c(x) + s(y)s(x)s(z) & c(y)c(x) \\ s(z)s(x) + c(z)c(x)s(y) & c(z)s(x)s(y) - c(x)s(z) & c(z)c(y) \end{bmatrix} \quad (4)$$

And c denotes as cosine, and s denotes as sine.

If the rotation is opposite direction to the inertial frame, then the rotation matrix is expressed by (5).

$$M_{Opposite}(z, y, x) = R(-z) Q(-y) P(-x) \quad (5)$$

B. Gyroscope

Gyroscope computes the orientation by the integration of the kinematic equations. It calculates the roll, pitch, yaw angular values ($\omega_x, \omega_y, \omega_z$). It is then transformed with Euler angles rates given in (6). For getting the angular rate, the output of gyroscope need to be converted to body frame. The transformation matrix resulting for Euler angle from conversion of body frame is given in (7).

$$\begin{bmatrix} z' \\ y' \\ x' \end{bmatrix} = \begin{bmatrix} 1 & \sin(x)\tan(y) & \cos(x)\tan(y) \\ 0 & \cos(z) & -\sin(z) \\ 0 & \sin(z)/\cos(y) & \cos(z)/\cos(y) \end{bmatrix} \begin{bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{bmatrix} \quad (6)$$

$$A(z, y, x) = \begin{bmatrix} 1 & \sin(x)\tan(y) & \cos(x)\tan(y) \\ 0 & \cos(z) & -\sin(z) \\ 0 & \sin(z)/\cos(y) & \cos(z)/\cos(y) \end{bmatrix} \quad (7)$$

C. Accelerometer

The accelerometer measures the gravitational acceleration of earth pointing at known direction. This information is used to obtain angular position relative to the earth. But it is accurate only when magnitude of a 3-dimensional force vector is the same as earth's gravity and points to the same direction. A magnitude of the force vector can be calculated by (8a), following equation.

$$|a| = \sqrt{(a_x^2 + a_y^2 + a_z^2)} \quad (8a)$$

If accelerometer is not moving then gravitational force only affect the accelerometer triad. Then pitch and yaw can be measured using (8b) and (8c).

$$\phi_a = \tan^{-1}(a_y / (\sqrt{a_x^2 + a_z^2})) \quad (8b)$$

$$\theta_a = \tan^{-1}(a_x / (\sqrt{a_y^2 + a_z^2})) \quad (8c)$$

Yaw is not calculated with the accelerometer as the vertical axis of the accelerometer is directly parallel to earth's gravity. This can be overcome by using the magnetometer to measure yaw.

D. Magnetometer

A three-axis magnetometer measures strength of the magnetic field. So, the yaw or heading can be measured from three axis magnetic fields (m_x, m_y, m_z) and using roll and pitch angle from accelerometer and gyroscope from the following (9a) to (9c).

$$Y = mx \times \cos(z) + my \times \sin(z) \quad (9a)$$

$$X = mx \times \cos(y) + my \times \sin(z) \times \sin(y) - mz \times \cos(z) \times \sin(y) \quad (9b)$$

$$Yaw, P = \arctan(Y/X) \quad (9c)$$

If magnetometer is parallel to earth's magnetic field, heading can be calculated accurately using just magnetic field strength values Y and X. Here, the yaw measurement totally depends on roll and pitch angles.

E. Complementary Filter

The complementary filter is used to fuse the sensors for getting reliable attitude information. The basic principle of complementary filter is to weight more on the trusted sensor. To remove the initial data error, the accelerometer data are passed through the low pass filter and the gyroscope data are passed through the high pass filter. The transfer function for the input passing through the low pass filter is $\frac{1}{1+Ts}$ and the

high pass filter is $\frac{Ts}{1+Ts}$. Then the data are being integrated. The weighted average is measured by (10a). The total gain of the two filters is 1.

$$G1(s) + G2(s) = 1 \quad (10a)$$

In terms of transfer function, the equation of the complementary filter can be written as (10b).

$$\phi(final) = \frac{a}{(1+Ts)} + \left[\frac{T}{s(1+Ts)} * \frac{1}{s} \right] b = \frac{a+Tb}{1+Ts} \quad (10b)$$

Where, T = Filter cutoff frequency and a = Accelerometer data input.

Calculated angle of gyroscope after passing through an integrator is $\frac{b}{s}$

Using backward difference method, the final equation is given in (10c). So, the complementary filter corrects the error using the same function.

$$Q_n = c [Q_{(n-1)} + b_n(t)]_+ + (1-c) a_n \quad (10c)$$

Where, $c = \frac{T/t}{1+(T/t)}$ and t = time interval.

F. Architecture and Hardware

The project uses a low cost 10DOF GY-88 IMU which comprises of 3-axis MPU6050 gyroscopes and accelerometers and HMC5883L magnetometers. It also contains a barometric pressure sensor. But the pressure sensor has not been utilized in this project. Arduino Uno has been used to capture the sensor data. Arduino is a low cost electronic platform that provides easy way for researchers and professionals to create devices that can work with environment using various sensors.

The system architecture of the proposed methodology is shown in Fig. 2. Data from the GY-88 3 axis sensors are captured through the Arduino Uno. After filtering, the data are integrated. Then the data are then fused using the complementary filter. Finally, the system provides roll, pitch and yaw.

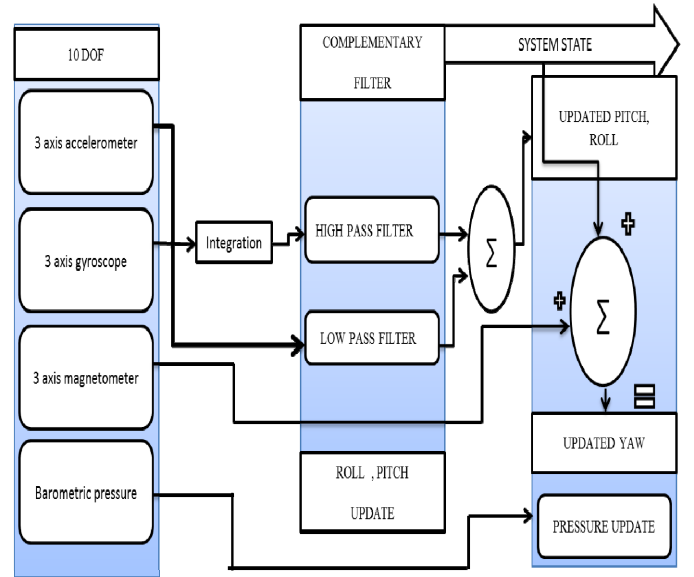


Fig. 2. Proposed architecture

III. RESULT AND DISCUSSION

The GY-88 was connected to the Arduino. The Arduino was programmed for getting data from the GY-88 sensors. The sensor data were then captured in the laptop through the USB cable of the Arduino as shown in Fig. 3.

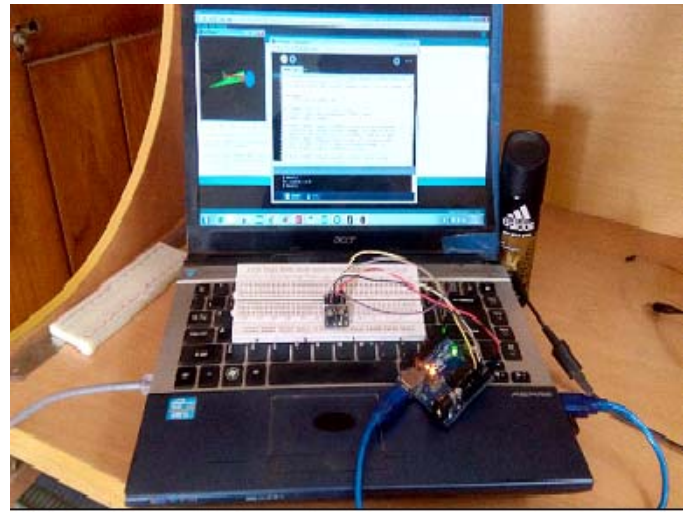


Fig. 3. Test setup for simulation

The roll, pitch and yaw are shown in Fig. 4, Fig. 5 and Fig. 6. The gyroscope suffered from drift due to earth rotation. Accelerometer and magnetometer show error due to gravity and earth's magnetic field. For the translational motion, the accelerometer measurement was not accurate as it showed spikes.

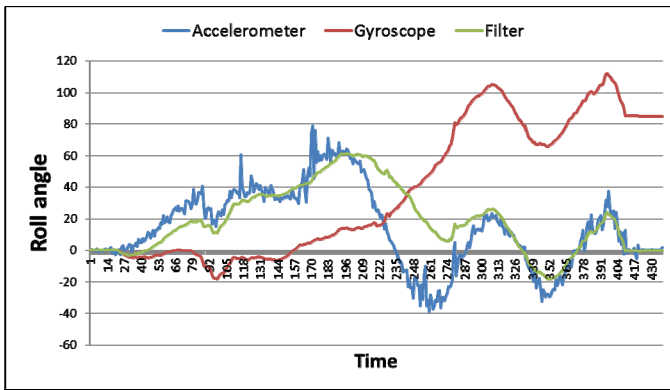


Fig. 4. Roll vs. Time graph

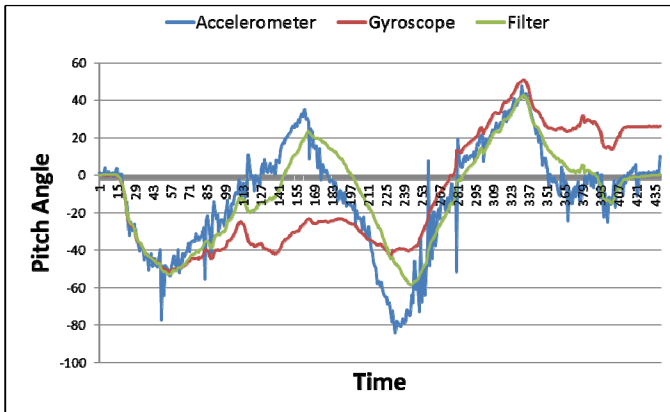


Fig. 5. Pitch vs. Time graph

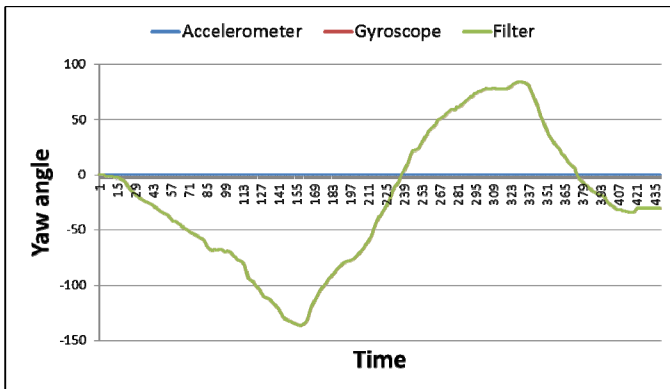


Fig. 6. Yaw vs. Time graph

The three dimensional (3D) representation of the roll, pitch and yaw is shown in Fig. 7 and Fig. 8 after applying the Complimentary Filter. The error is reduced to an acceptable level. It has been observed that when pitch angle is 90 degrees the sensors do not give the result accurately. This happened due to Gimbal lock.

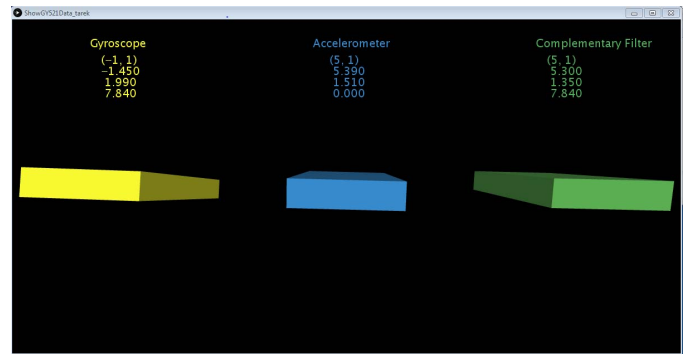
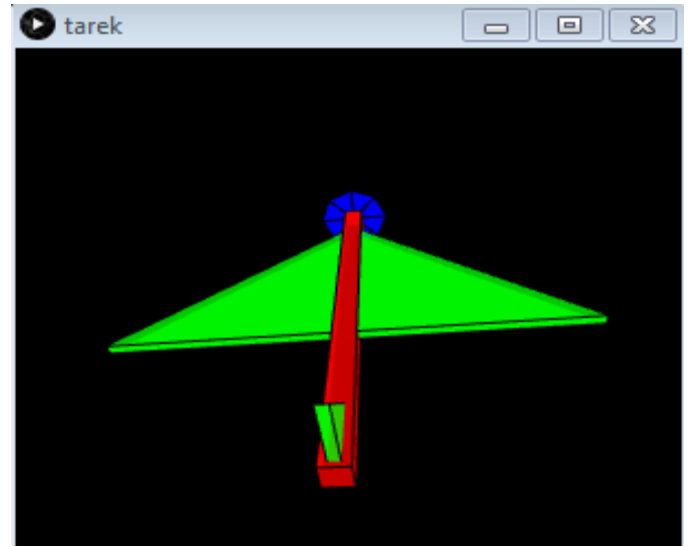


Fig. 7. 3D representation of Roll, Pitch and Yaw



S

Fig. 8. 3D representation of roll, pitch and yaw in aircraft model

IV. CONCLUSION

A low cost standalone MEMS based AHRS has been proposed in this paper for recording the attitude of a low speed aircraft for training evaluation. The low cost MEMS based accelerometers, gyroscope and magnetometers data were captured by Arduino Uno and the errors were corrected. Complimentary Filter was used to overcome the limitation of each sensor. The drift of the gyroscope was corrected with the value from accelerometers. The final output were at acceptable level. Besides using it in low speed aircraft, the system can also be used in robot navigation, camera stabilization etc.

REFERENCES

- [1] J. Lu, L. Xie, and B. Li, "Applied Quaternion Optimization Method in Transfer Alignment for Airborne AHRS under Large Misalignment Angle," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, pp. 346-354, 2016.
- [2] M. S. Amin, M. B. I. Reaz, S. S. Nasir, and M. A. S. Bhuiyan, "Attitude Heading Reference System based vehicle stationary state detection," in *International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT)*, Dhaka, Bangladesh, 2014, pp. 1-4.

- [3] L. Sherry, C. Brown, B. Motazed, and D. Vos, "Performance of automotive-grade MEMS sensors in low cost AHRS for general aviation," in *The 22nd Digital Avionics Systems Conference, 2003*, Indianapolis, USA, 2003, pp. 12.C.2-12.1-5.
- [4] M. S. Amin, M. B. I. Reaz, S. S. Nasir, M. A. S. Bhuiyan, and M. A. M. Ali, "A Novel Vehicle Stationary Detection Utilizing Map Matching and IMU Sensors," *The Scientific World Journal*, vol. 2014, pp. 1-13, 2014.
- [5] G. K. Balachandran, V. P. Petkov, T. Mayer, and T. Balslink, "A s3-Axis Gyroscope for Electronic Stability Control With Continuous Self-Test," *IEEE Journal of Solid-State Circuits*, vol. 51, pp. 177-186, 2016.
- [6] M. S. Amin, S. S. Nasir, M. M. B. I. Reaz, M. A. M. Ali, and T.-G. Chang, "Preference And Placement of Vehicle Crash Sensors," *Technical Gazette*, vol. 21, pp. 889-896, August 2014 2014.
- [7] Z. Wu, Z. Sun, W. Zhang, and Q. Chen, "A Novel Approach for Attitude Estimation Based on MEMS Inertial Sensors Using Nonlinear Complementary Filters," *IEEE Sensors Journal*, vol. 16, pp. 3856-3864, 2016.
- [8] Y. Zhao, "GPS/IMU Integrated System for Land Vehicle Navigation based on MEMS," Masters, Division of Geodesy and Geoinformatics, Royal Institute of Technology, 2011.

Machine-to-machine Communication Based Smart Home Security System by NFC, Fingerprint, and PIR Sensor with Mobile Android Application

Sayidul Morsalin, A. M. Jahirul Islam

Department of Electrical and Electronic Engineering
Chittagong University of Engineering & Technology
Chittagong-4349, Bangladesh

Email: {sayidul.morsalin, a.m.jahirulislam}@gmail.com

Golam R. Rahat, Syed R. H. Pidim,

Abdur Rahman, Md A. B. Siddiqe

Chittagong University of Engineering & Technology
Email: {grahmanrahat, abubakar.khadachara}@yahoo.com
{pidim.srh, kakoncu9}@gmail.com

Abstract—The rapid growth of burglary and theft over the last few years has been threatening to the vulnerability of traditional home security systems. In this research, this hazardous circumstance persuades the authors to develop a machine-to-machine (M2M) communication based smart home security system. Here a multi-layer home security system (HSS) has been developed which maintains six levels of security and necessary control responses against unwanted burglars and intruders. As a part of this development, the first security level uses Near Field Communication (NFC) tag, the second level uses a secured password system and the third level uses fingerprint authentication. After that, a GSM module embedded with the proposed HSS sends the logged password to a remote server via M2M communication. The server encrypts the password and notifies the homeowner via an android based mobile application whether the person is an authenticated person or not. In the final stage of security, if an intruder wants to enter the room by breaking the password without NFC tag and fingerprint encryption, a Passive Infrared (PIR) motion sensor will work and sounds a buzzer for alarm. Therefore, the automatic and embedded operation of the proposed HSS makes it more secure, reliable and robust than other traditional systems.

Index Terms—Smart home; NFC tag; PIR Sensor; M2M Communication; GSM System

I. INTRODUCTION

Smart maintenance of security has emerged as a cardinal concern for any personnel systems, especially for an individuals dwelling place. As well, security systems of personal home place are emphasized to be automated and robust for better convenience and safety issues [1]. The reason is that, over the last decade, the rapid rise of burglary and theft all over the world is threatening due to the vulnerability of traditional home security systems [2]–[4]. In addition, the necessity of smart home security systems (HSS) emerges inevitable for people of developing or developed countries [5]. According to the United Nations Office on Drugs and Crime (UNODC), the burglary rate of the developed countries such as the United States of America (USA), Australia, Switzerland was 892.8, 602.15, and 850.8 respectively in 2013 [6]. Besides, the statistics of the European Institute for Crime Prevention and Control, affiliated with the United Nations (HEUNI) in

2011 reports that a developing country like Bangladesh rated 2.2 points and ranked in 53rd position for burglaries, and rated 0.7 points and ranked in 71st position for auto theft around the world in 2006 [5]. For the USA, the position is sixth for auto theft and ninth for burglary. Both of these statistics are based on the number of burglaries/auto thefts recorded by police per one million population of the respective country. Hence the burglary and theft rate all over the world including Bangladesh motivated the authors to work on this research.

A lot of research has been conducted to maintain home security [7]–[11]. All of them neglected the integrated and embedded form of HSS, with near field communication, PIR sensor, machine-to-machine (M2M) communication-based GSM system with a smart android mobile application (i.e. apps) for robust operation. The term *M2M* implies that the communication between two electronic systems occurs autonomously without any human intervention. All the previous work maintains a maximum of 3-4 layers of security systems. In [12], Kumar et al. develop an Infrared (IR) based HSS with a GSM module. In [13], the authors present wireless monitoring of a home security system with GSM/GPRS systems. Both the later two neglect the M2M operation with smart mobile apps for user convenience. This proposed research overcomes the limitation of all previous work. Here the authors describe the implementation of a six-layer smart HSS. The layers are implemented by a near field communication (NFC) tag, secured password protection, fingerprint, M2M system, an android application and a passive infrared (PIR) sensor (i.e. motion sensor). The assembly of all the components together in a single cheap system makes the proposed HSS more robust, cost-effective and user-friendly. At any kind of emergency or power cutoff, a rechargeable battery will give backup power to the system. Moreover, the remote access via M2M communication and mobile application enables the homeowner when they are away to set an emergency alarm if the inside of the home atmosphere diver from its normal condition and exceeds a set threshold. So, analyzing complexity, efficiency and economy, the proposed system is found to be smart choice for the security protection of a dwelling place.

II. SYSTEM DESCRIPTION

Previously a CMOS- based switch is used for door locking, which has a problem of password storage when the main power is turned off. Also, many security systems were designed for door opening but not for door closing. Here the six-layer security system is designed into a simple integrated and embedded form. As well, an electro-mechanical lock is designed which can be opened by a valid key when the power is turned off. This design also includes the options

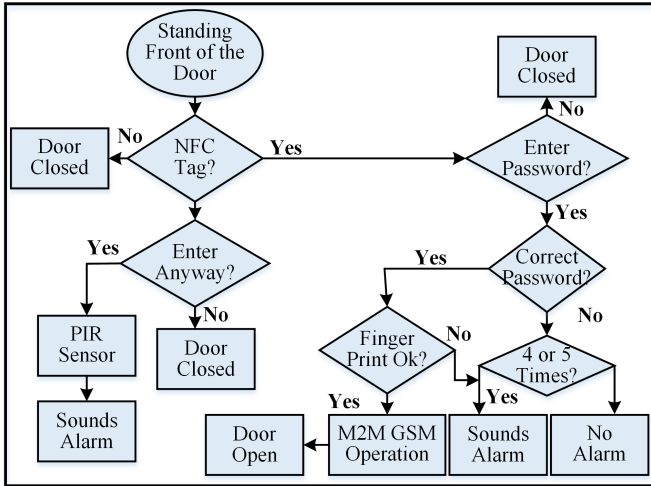


Figure 1: Flowchart of the System

of password changing and triggering the alarm circuit when the wrong password is provided 4-5 times. For a valid NFC Tag, password and fingerprint, the home door will be opened automatically by a servomotor coupled with a lock in its shaft. If a burglar wants to access a room he must have an NFC tag. If he shows the NFC tag, then the system requires the password and fingerprint. For any invalid user with 4-5 consecutive wrong trials for the password in the keypad, the system will identify the user as malignant and generates the emergency alarm. However if he enters the room without these three protections, e.g. he breaks the password and fingerprint, then the PIR sensor works and sounds the alarm. If tag, password and fingerprint match, then a notification from the GSM module of the HSS is transmitted to a remote server via M2M communication. After that, the server sends the login alert message to the cellphone of the homeowner via an android application. After receiving the message, the owner can easily identify who is trying to access the home, and whether they are an authenticated user or not. Therefore, maximum security is ensured by implementing the proposed HSS. The system is implemented with cheap electronic devices and automated, so no physical labor is required to monitor the home condition. Since an NFC shield works within a small distance, it is very difficult to break NFC security, unlike RFID. The total working process of the system is illustrated by the flowchart of Figure 1.

III. MACHINE-TO-MACHINE COMMUNICATION

Machine-to-machine (M2M) communication takes place between embedded devices at one end and a network server at the other end via an internet network using either wired, wireless or a combination of both, using little or no user intervention [14]. The functions of an M2M communication network for HSS can be divided into three types. Firstly, **Tracking** which is necessary for real-time applications e.g. sensing the data (i.e. password, fingerprint). Secondly, **Logging** which is necessary for scheduled data communication and medium data density for remote applications. Finally, **Notifications** which is an exception-based application to notify and alert users in real time to take decisions. The M2M network architecture suggested by the European Telecommunications Standards Institute (ETSI) for home security is shown in Figure 2 [15]. This architecture has basically three parts. These are: 1. Access network, 2. Gateway, 3. Core network.

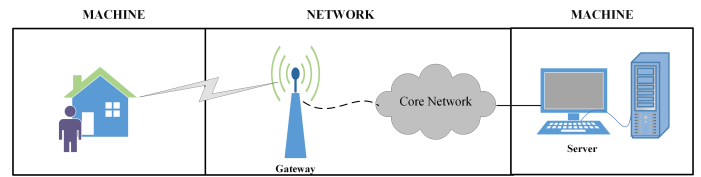


Figure 2: Simplified HSS M2M Communication Configuration

For this research, the notification alert from the HSS is sent to a dedicated server by using the radio access network of a GSM cellular system. After that, the server manipulates the data and sends a message to the homeowner.

IV. BLOCK DIAGRAM AND CIRCUIT OPERATION

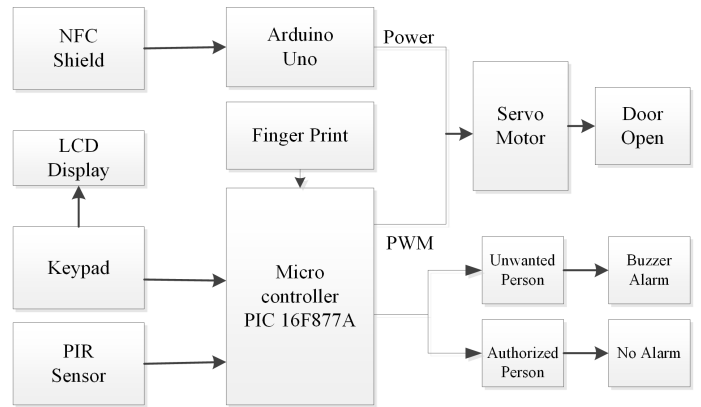


Figure 3: Block Diagram of the System

The block diagram shown in Figure 3 represents an operational view of the system. Here an NFC (Near Field Communication) tag card, password from the keypad, finger print and infrared detector activated by the motion of a person (i.e. from PIR sensor) are taken as inputs. The output of the system is information visualization in an LCD display, a Buzzer alarm, and a servo motor. To implement the system, a PIC (Peripheral

Interface Controller) microcontroller 16F877A, 32×2 LCD display, a 4×4 keypad to enter a password, a fingerprint sensor, a servo motor and a PIR sensor are used. The servo motor is used for door locking, and also incorporates a power supply unit and the PIR sensor. Arduino Uno as shown in the Figure 3 provides power supply and the microcontroller provides a pulse width modulated (PWM) signal. The servo motor needs both the power signal and the PWM signal to operate. Only if the correct password is pressed on the keypad, and the fingerprint is matched from the scanner, does the servo get the PWM signal. So the motor will not operate due to the absence of the required power signal from the Arduino board. Besides, only if an NFC tag card is provided in front of the shield does the servo get power, but in the absence of the PWM signal, it will not operate. Here the microcontroller also takes the PIR sensor signal as input. When the sensor gets this signal it sends a signal to the Buzzer and sounds the alarm.

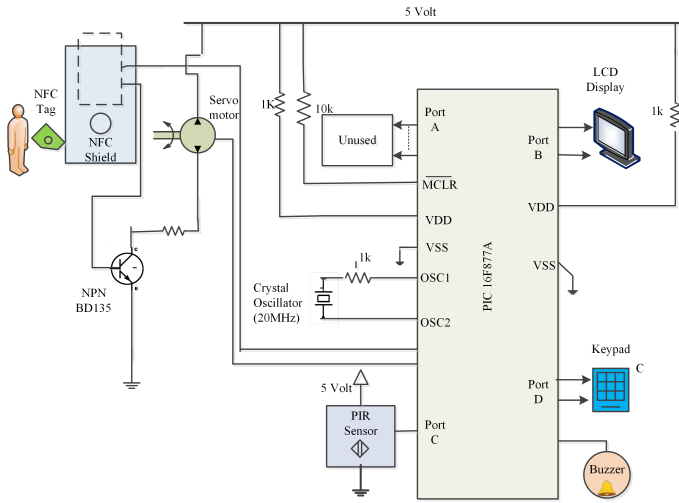


Figure 4: Circuit Diagram of the System

Figure 4 shows the complete circuit diagram of the system. Here a crystal oscillator is used to generate a clock signal of frequency 20 MHz. The EEPROM of the PIC stores the passwords set by the user. When any password is entered through the keypad, the microcontroller decodes the input and verifies it with the stored password. If the password and fingerprint match with the correct one set by the authorized person, the microcontroller sends a signal to the servo motor. At the same time, if the NFC tag is placed in front of the shield, the coil inside the tag is energized, and gets power from the shield. After that, the signal from the Arduino board is sent to the base of an NPN transistor. The transistor acts as a switch when it gets a signal at its base and is turned on. Then it connects the servo to the power supply unit. The servo runs its shaft to operate, and finally, the lock of the door is turned off.

V. HARDWARE IMPLEMENTATION

A. NFC Shield plus Tag

Near Field Communication (NFC) is a wireless non-contact short-distance communication of radio-frequency electromagnetic waves to transfer data automatically for identifying and tracking tags attached near the shield [16]. The tag stores information electronically, in a non-volatile memory. Tags are usually powered by electromagnetic induction within a small distance or by a local power source such as a battery. The NFC Shield maintains SPI communication with the Arduino board. The NFC operates in the ISM band at 13.56 MHz, but this operating range can be limited by the antenna and power radiation design [17], [18]. The NFC tag consists of two main parts. The main body is an integrated circuit which is used for storing and processing information, modulating and demodulating a radio-frequency (RF) signal and collecting DC power from a signal transmitted from the shield. The second part is an antenna capable of receiving and transmitting the signal. The NFC reader, i.e. the shield transmits an encoded radio frequency signal to interact with the tag. The tag receives the signal and responds with its unique identification and other information such as a serial number or may be product-related information such as a stock number or batch number, production date or other specific information set by the manufacturer [19]. The radio performance of the NFC offers some unique benefits over Radio frequency identification (RFID) communication. Literally, RFID operates by one-way communication whereas NFC is a two-way communication. RFID works in a certain radio frequency range, but NFC works at 13.56 MHz frequency. NFC follows close proximity communication, 5 cm or less; on the other hand, RFID covers typically a distance of 1 m [20].

B. Interfacing of Arduino with NFC Shield

The NFC shield communicates with the Arduino board through a serial peripheral interface (SPI) protocol as shown in Figure 5. The data exchange rate in this mode is higher than an I²C (Inter-Integrated Circuit) interface although interfacing multiple devices is easier in I²C mode. But in this work only one NFC device is necessary, but if several NFC devices are needed at several checking points, only one Arduino or microcontroller can monitor them through the I²C protocol using the addressing mode [18]. So, the steps of checking the tag are:

Step 1: Read the unique tag number of the tag.

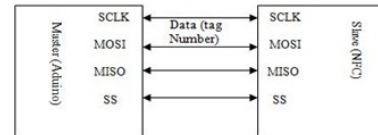


Figure 5: Interfacing of Arduino with NFC Shield

Step 2: Transmit of the tag number to Arduino by SPI protocol.
Step 3: Check the data received in Arduino.

Step 4: Arduino respond as per the instruction code based on the tag number.

Here SCLK= serial clock; MOSI= master out slave in; MISO= master in slave out; SS= slave select.

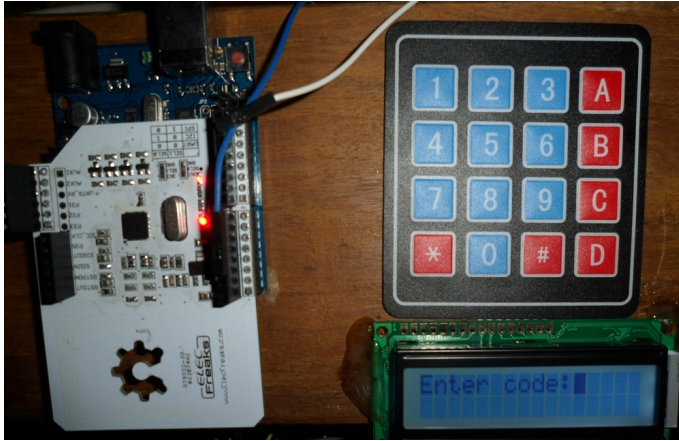


Figure 6: NFC Shield Interfacing in Practical Circuit

On every cycle of the working frequency the control unit (Arduino here) sends a signal on the MOSI line to check for data from the NFC device and on every cycle the NFC device transmits data to the controller through the MISO line. The slave select pin is active always. Interfacing of several NFC devices in SPI mode through a Daisy Chain would increase the complexity of connection. In that case, I²C will be preferable. The practical implementation of the NFC shield and Arduino with Keypad, LCD display in this research is shown in Figure 6.

C. Interfacing of GSM module with Microcontroller

Table I: AT Commands for GSM Module

AT Command	Description
+CMGS	To send message/information
+CMGC	Sending the command
+CMSS	Send message from storage
+CMMS	Additional SMS to send
+CMGI	Module testing
+CMGW	To write message/information
+CMGD	To delete message/information

In this research, an SIM908 module is used as a GSM module for M2M communication. The module is a complete Quad-Band GSM/GPRS module which includes GPS technology for satellite navigation. It works by AT commands. Moreover, when this power module is in sleep mode, it can save power consumption by up to 40% over the current industry average. Its voltage range for GPRS is about 3.2-4.8 V, while for GPS it ranges from 3.0 to 4.5 V, with very low power consumption at operating temperature ranges from 40 to 85°C. The AT commands are used as the control instructions of the modem and so several AT commands are used to control the dial-up modem or the GSM/GPRS modem [21], [22]. The AT

commands used for interfacing with the microcontroller are listed in Table I.

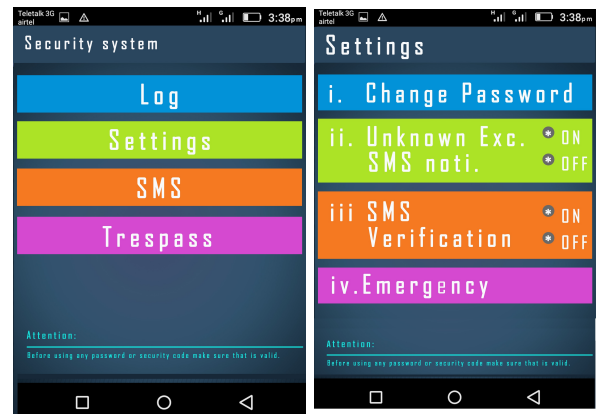
D. PIR Sensor

The passive infrared (PIR) sensor is made of a pyroelectric sensing coating which can detect levels of infrared radiation. The sensor in a motion detector is actually split into two halves [23]. The reason for that is to detect a change of motion, not to detect average IR levels. The sensor remains idle when both slots receive the same amount of IR. However, when a warm body like a human or animal passes through its protected region, and affects just one-half of the PIR sensor, this causes a positive differential change between the two halves. When the warm body leaves the sensing area, the reverse happens, whereby the sensor generates a negative differential change. These changes generate pulses which are detected by the microcontroller, which generates a high output and sounds the buzzer [23].

VI. EMERGENCY ELECTROMECHANICAL LOCK

In this research, a lock has been developed which is operated both electrically and mechanically. The lock is coupled with the shaft of the servo motor, used for precise angular position. The shaft is connected to a top portion of the lock mechanically. The inner portion of the lock is operated by a compression spring. When the servo's shaft rotates clockwise, the spring is compressed and the door opens. Conversely, when the servo releases power, the spring is decompressed and the door closes. The lock also carries the advantage of opening the door with a mechanical key which can be designed by the manufacturer.

VII. ANDROID MOBILE APPLICATION



(a) Main Window

(b) Sub Window

Figure 7: Screenshots of User's Android Application

For the purposes of research, an Android-based mobile application (app) is developed in Android studio software by JAVA coding. The pop up window of the app is divided into four sub-windows. These are access log, settings, SMS panel, and trespass as shown in Figure 7. All types of access are notified in an access log, showing whether the person is an

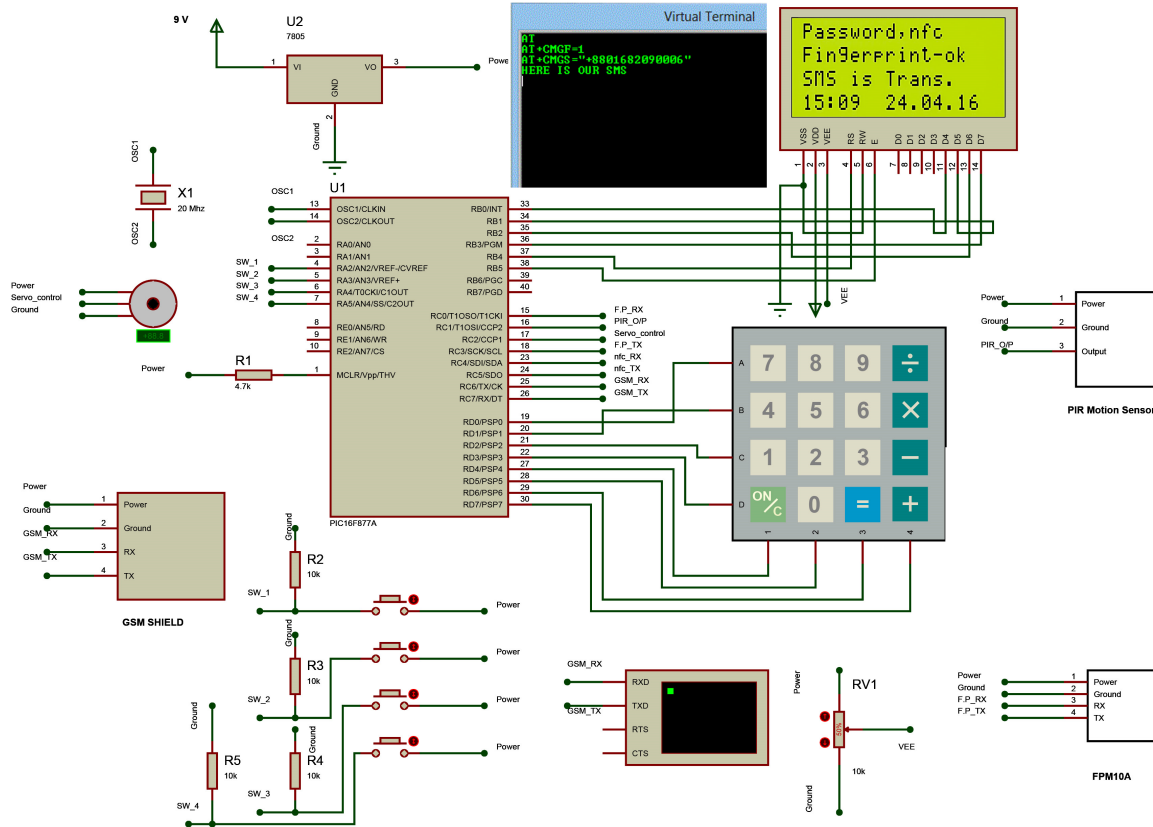


Figure 8: Simulation of the Embedded System

authorized person or an unknown person. The SMS panel shows the communication between the GSM module and the server. When a trespass is on and if anyone is trying to break the lock, a notification SMS is sent to the nearest police station to take necessary steps.

VIII. SOFTWARE SIMULATION

A complete circuit diagram to demonstrate the performance and evaluate efficiency has been simulated in Proteus simulating software as shown in Figure 8. Proteus software is generally used for microprocessor simulation, schematic capture and printed circuit board (PCB) design [24]–[26]. According to the system algorithm, the code is written in mikro C. After writing, the HEX file generated from mikro C Pro is loaded into the microcontroller to generate the appropriate output.

IX. IMPLEMENTED SYSTEM VISUALIZATION



Figure 9: Various LCD Notifications for the Proposed HSS

Before permitting entry to the door, there is a text in the LCD display, *Show Your Tag* will be displayed as in Figure 9. When an NFC tag is shown in front of the NFC shield then the system will ask for a password. If the correct password is entered, then the system will ask for a finger print. For a valid tag, correct password and finger print, the display will show *Door Opened*. Otherwise it will display door closed. For a

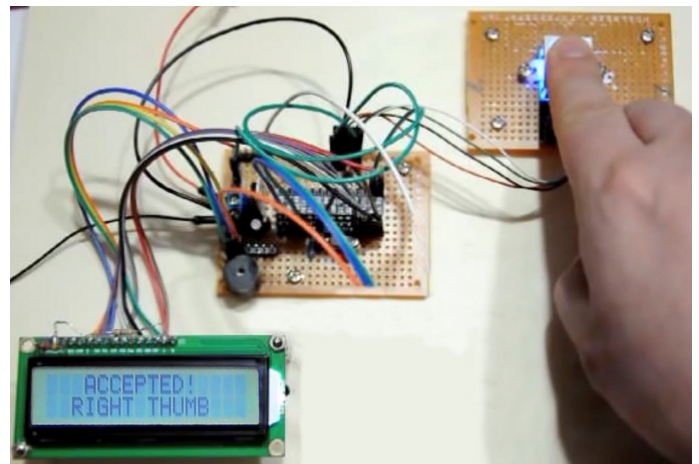


Figure 10: Accepted Notification of Fingerprint

wrong password, the display will show like *Wrong Password*.

For an authenticated user, the display will show that the finger print is matched with this notification *Accepted! Right Thumb* as shown in Figure 10.

The home owner can know who is trying to access this room and when by the mobile application indicating that the person is an unwanted intruder or a verified person. The mobile application window is shown in Figure 11.

SMS Panel				Log			
No.	Number	Details	Time	No.	Name	Date	Time
01	+8801947357817	verified...	3.08p.m.	01	Morsalin	24.4.2016	3.00p.m.
02	+8801675383098	verified...	3.09p.m.	02	Rahat	24.4.2016	3.06p.m.
03	+8801682090006	verified...	3.10p.m.	03	Tusher	24.4.2016	3.10p.m.
04	+8801675481252	verified...	3.11p.m.	04	Khizir	24.4.2016	3.14p.m.
05	+8801675935647	unknown access...	3.11p.m.	05	Unknown	24.4.2016	3.19p.m.

(a) SMS Panel Window

(b) Log Window

Figure 11: Login Information of the Proposed HSS

X. CONCLUSION

In this research, the developed system works very well without any basic error unless it is physically damaged; as well it is very smart, reliable and robust. So the system can be implemented in any residential, commercial, industrial or military area, as it is very user-friendly. Moreover, it is more economical than any existing system. Since the system ensures maximum security, it is quite impossible for any burglar to enter the room without warning the owner. Some features can be added in future to enhance the robustness of the proposed system such as retinal scanning, iris scanning, voiceprint identification etc.

REFERENCES

- [1] G. Demiris, B. K. Hensel *et al.*, "Technologies for an Aging Society: A Systematic Review of Smart Home Applications," *Yearb Med Inform.*, vol. 3, pp. 33–40, 2008.
- [2] T. Hope, "Conceptualising the Trend in Burglary in England and Wales," 2007.
- [3] L. R. Tabrizi and A. Madanipour, "Crime and the city: Domestic burglary and the built environment in Tehran," *Habitat International*, vol. 30, no. 4, pp. 932–944, 2006.
- [4] D. Deadman, "Forecasting residential burglary," *International Journal of Forecasting*, vol. 19, no. 4, pp. 567–578, 2003.
- [5] S. Harrendorf, M. Heiskanen, and S. Malby, *International statistics on crime and justice*. European Institute for Crime Prevention and Control, affiliated with the United Nations (HEUNI), 2010.
- [6] UNODC, "International Burglary, Car Theft and Housebreaking Statistics," United Nations Office on Drugs and Crime (UNODC), Tech. Rep., 2015.
- [7] R. J. Robles, T.-h. Kim, D. Cook, and S. Das, "A review on security in smart home development," *International Journal of Advanced Science and Technology*, vol. 15, 2010.
- [8] D. M. Konidala, D.-Y. Kim, C.-Y. Yeun, and B.-C. Lee, "Security framework for RFID-based applications in smart home environment," *Journal of Information Processing Systems*, vol. 7, no. 1, pp. 111–120, 2011.
- [9] H. H. Kim, K. N. Ha, S. Lee, and K. C. Lee, "Resident location-recognition algorithm using a Bayesian classifier in the PIR sensor-based indoor location-aware system," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 39, no. 2, pp. 240–245, 2009.
- [10] C. Suh and Y.-B. Ko, "Design and implementation of intelligent home control systems based on active sensor networks," *IEEE Transactions on Consumer Electronics*, vol. 54, no. 3, pp. 1177–1184, 2008.
- [11] I. Kaur, "Microcontroller based home automation system with security," *International journal of advanced computer science and applications*, vol. 1, no. 6, pp. 60–65, 2010.
- [12] P. Kumar and P. Kumar, "Arduino based wireless intrusion detection using IR sensor and GSM," *International Journal of Computer Science and Mobile Computing*, vol. 2, no. 5, pp. 417–424, 2013.
- [13] Y. Zhao and Z. Ye, "A low cost GSM/GPRS based wireless home security system," *Consumer Electronics, IEEE Transactions on*, vol. 54, no. 2, pp. 567–572, 2008.
- [14] D. Niyato, L. Xiao, and P. Wang, "Machine-to-machine communications for home energy management system in smart grid," *Communications Magazine, IEEE*, vol. 49, no. 4, pp. 53–59, 2011.
- [15] S. M. Besen, "The European Telecommunications Standards Institute (ETSI): A preliminary analysis," *Telecommunications policy*, vol. 14, no. 6, pp. 521–530, 1990.
- [16] T. G. Zimmerman, "Personal area networks: near-field intrabody communication," *IBM systems Journal*, vol. 35, no. 3.4, pp. 609–617, 1996.
- [17] Z. Antoniou and D. N. Kalofonos, "NFC-based mobile middleware for intuitive user interaction with security in smart homes," in *Communication Systems and Networks*, 2006, pp. 97–102.
- [18] C. A. Opperman and G. P. Hancke, "A generic NFC-enabled measurement system for remote monitoring and control of client-side equipment," in *Near Field Communication (NFC), 2011 3rd International Workshop on*. IEEE, 2011, pp. 44–49.
- [19] P. Agrawal and S. Bhuraria, "Near field communication," *SETLabs Bridfings*, vol. 10, no. 1, pp. 67–74, 2012.
- [20] A. Lahtela, M. Hassinen, and V. Jylha, "RFID and NFC in healthcare: Safety of hospitals medication care," in *Pervasive Computing Technologies for Healthcare, 2008. PervasiveHealth 2008. Second International Conference on*. IEEE, 2008, pp. 241–244.
- [21] S. Morsalin, A. Rahman, A. B. Siddiqe, P. Saha, and R. Halim, "Password protected multiuser wireless electronic noticing system by GSM with robust algorithm," in *Electrical Information and Communication Technology (EICT), 2015 2nd International Conference on*. IEEE, 2015, pp. 249–253.
- [22] S. Morsalin, K. Mahmud, M. R. Halim, P. Saha, and I. A. Jamil, "Freighter fuel level detection and overload alarming system with safety notification via GSM," in *Informatics, Electronics & Vision (ICIEV), 2014 International Conference on*. IEEE, 2014, pp. 1–5.
- [23] K. N. Ha, K. C. Lee, and S. Lee, "Development of PIR sensor based indoor location detection system for smart home," in *SICE-ICASE, 2006. International Joint Conference*. IEEE, 2006, pp. 2162–2167.
- [24] K. Mahmud, M. S. Alam, and R. Ghosh, "Design of digital thermometer based on PIC16F77A single chip microcontroller," in *Consumer Electronics, Communications and Networks (CECNet), 2013 3rd International Conference on*. IEEE, 2013, pp. 246–249.
- [25] K. Mahmud, M. S. Alam, and A. Jobayer, "Design and performance evaluation of PIC16F77A microcontroller based fluid velocity meter," in *Global High Tech Congress on Electronics (GHTCE), 2013 IEEE*. IEEE, 2013, pp. 21–24.
- [26] M. Azim, K. Mahmud, and C. Das, "Automatic Train Track Switching System with Computerized Control from the Central Monitoring Unit," *International Journal of u-and e-Service, Science and Technology*, pp. 201–212, 2014.

Hybrid Neural Network for Efficient Training

Md. Billal Hossain^{1*}, Sayeed Islam², Noor-e-Hafsa Zhumur³, and Najmoon Nahar Khanam⁴

[1, 2, 3, 4][#]Electrical and Electronic Engineering Department, MIU, Dhaka-1216, Bangladesh.

*billal.0709018@gmail.com

Abstract—Neural Network is a platform to implement artificial intelligence and Backpropagation (BP) is one of the most widely used neural networks. Local minimum, slower convergence, premature saturation, training pattern overspecialization limits the performance of Backpropagation algorithm. To withstand those problems several modified algorithms was proposed. In this work, a faster supervised algorithm named Hybrid Backpropagation algorithm (HBP) for the neural network training is found from Back-propagation with Chaotic Learning (BPCL) with different types of chaos, Maximization of gradient function (BPfast) and Error back-propagation (EBP) to mitigate the limitations of BP. HBP is tested on five real world benchmark categorize problems such as breast cancer, diabetes, glass, Australian credit card, and horse. HBP outperforms BP, BPfast, BPCL and EBP in terms of generalization ability and also convergence rate then any single algorithm.

Keywords—Backpropagation, Artificial Intelligence, Neural Network.

I. INTRODUCTION

Neural network models come from neuroscience, and the network model exhibits a number of the brain's characteristic. Brain is made by simple processing elements called neuron which resides in brain as a massively interconnected complex network. According to the principal of operation of biological neural network, artificial neural network model is proposed. Unlike biological neural network an artificial neural network is a massively parallel distributed processor consists of some input, output vectors, functions that have a natural propensity for storing experiential knowledge which is making it available for use. It has similarity with the brain in two respects: Knowledge is gathered by the network through a learning process and inter neurons are connected so that robustness's are used to store the knowledge [1]. The learning mechanisms of neural networks are broadly classified as Supervised, Unsupervised and Reinforced learning methods. Several models of artificial neuron have been proposed [1-8].

Backpropagation (BP) [7] is a systematized method of training multilayer artificial neural network. It is built on high mathematical and has very good application potential in last few decades in many fields of science and technology. BP is applied to a wide range of practical problems such as classification forecasting, pattern recognition signal processing, image compression, speech recognition etc [9]. BP suffers from a number of problems [10], [11]. Among them major problems are –i) it may converge to a local minimum, ii) BP learning is very slow also it is not guaranteed and iii) BP does not require normalization of input vectors iv) training pattern overspecialization. When BP states in a local minima, it becomes unable to find optimal solution of a problem [12] and it leads the training to premature saturation or ‘flat spot’ [13]. As results, the error remains unchanged due to the small

gradient of sigmoid activation function. There are two main reason responsible for ‘flat spot’ problem—i) derivative of sigmoid activation function that usher the training to premature saturation ii) constant values of learning rate (LR). Nonlinearity of error surface in real world problem, BP training often falls into premature saturation that leads the training to a non updating zone. When some output nodes of Multi-layer perceptrons (MLP) like BP are incorrectly steeped, the strong error signal of the output nodes updates the correlated weights so that they can escape the incorrectly saturated state hence accelerate the learning speed. For the appropriately saturated output nodes, only the proper error signal balks the overspecialization of learning for training patterns [14].

There are many enhancement and variations of BP with goals of the network's performance. BP trains NNs with constant values of learning rate (LR) where dynamic learning rate of BP accelerate the learning speed. In BPCL, chaotic learning rate proposed for solving premature saturation with local minima problem [15]. To generate chaotic LR, chaotic maps are used such as logistic map, tent map, gaussian map, tinkerbelle etc. Maps may be specified by a discrete-time or a continuous-time parameter [16]. Another proposed algorithm BPfast focuses on the derivative information of sigmoid activation function which escapes the NN training from embryonic saturation [17]. The conventional Error Backpropagation Algorithm (EBP) is proposed in [18] which is slightly modified version of standard Backpropagation (BP). A modified error function from the proposal of EBP [14] used and as a results, the appropriate error signal of the output nodes updates the associated.

In this paper, a modified BP algorithm, called ‘Hybrid Backpropagation for efficient training’ (HBP) is proposed which is the combined form of modification algorithms [16], [17] of BP. In HBP chaotic learning rate with two different types of chaotic series is used instead of fixed learning which helps to escapes the NN training from premature saturation, the slope of sigmoid activation function is used according to proposal of BP fast [16] with as to resolve the issue premature saturation and learning speed along with all modification from those helps the HBP to find optimal global solution. HBP is applied on five benchmark problems—breast cancer, heart disease, Australian credit card, glass, diabetes and horse. For all the problems, HBP is capable to train NNs with better abstraction ability and also faster convergence rate than that of BPfast, BPCL (Tent), BPCL (Logistic), EBP and standard BP.

The rest of the paper is organized as follows: The proposed algorithm HBP is described in Section II. In Section III

includes the experimental studies. Discussion is presented in Section IV and Section V stands for conclusion of the paper.

II. HBP

Let us consider the three-layer network with the input layer having 'l' nodes, hidden layer 'm' nodes, and an output layer with 'n' nodes, where the subscripts I, H, O denote input, hidden and output neurons respectively in Fig.1. We consider slope of sigmoid functions for activation function for the hidden and output layer and linear activation function for input layer. The number of training example is $\{I_{I1}, I_{I2}, \dots, I_{Il}, y_{O1}, y_{O2}, \dots, y_{On}\}$, where I_I is input vector and y_o is the target vector $\{I\}_H$ and $\{I\}_O$ are the outputs of hidden unit and outputs of output unit for the n-th training examples. Δ represents the difference between the current and new value of the network weights.

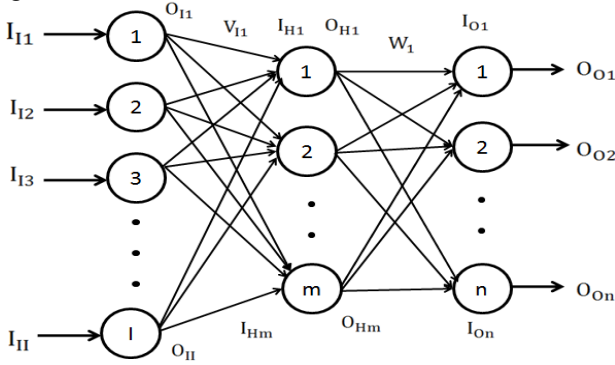


Figure 1. A feed-forward neural network with single hidden layer.

Step 1) Normalize the inputs and outputs with respect to their maximum values. It is proved that the neural networks work better if inputs and outputs lie between 0-1. For each training pair, assume there are ' T ' inputs given by $\frac{\{I\}_I}{\max}$ and ' n ' outputs in $\frac{\{O\}_O}{n \times 1}$ a normalized form.

Step 2) $[v]$ Represents the weight of synapses connecting input of neurons and hidden neurons and $[w]$ represents weight of synapses connecting hidden neurons and output neurons. Set the iteration number, $ITE = 0$. Compute the Rescaled LR by using following equation:

$$RLR(ITE) = M * C(ITE) - \beta$$

Step 3) Compute the inputs to the hidden layer $\{I\}_H$ and hidden to output layer $\{I\}_O$ by multiplying corresponding weights of synapses as:

$$\{I\}_H = f(\sum_{l=1}^l v_l \{O\}_I)$$

$$\{I\}_O = f(\sum_{H=1}^m w_H \{O\}_H)$$

Where, the slope of sigmoid function $f(x) = \frac{2}{1+e^{-x}} - 1$ is used as the stimulation function. Compute the changes of the weights, Δw_H and Δv_I for the n-th example using the following:

$$\Delta w_H = -RLR(ITE) \frac{\partial E}{\partial w}$$

$$= -RLR(ITE) \{O\}_H(d) + RLR(ITE) o_{ok}(1 - o_{ok})(1 - 2o_{ok}) \{O\}_H$$

$$\Delta v_I = -RLR(ITE) \{I\}_I(d^*)$$

Where,

$$E_n = \frac{1}{2} \sum_{k=1}^k [T_k - O_{Ok}]$$

$$d(t) = (t_k - o_{ok}) \frac{(1-o_{ok})(1+o_{ok})}{2}$$

$$d^*(t) = \frac{(1-o_{Hi})(1+o_{Hi})}{2} [w] \{d(t)\}$$

Update w_H and v_I by adding Δw_H and Δv_I with them correspondingly. When this process is repeated for all the training examples, iteration is completed and ITE is increased by one.

Step 4) Compute the mean square error for the iteration of step 2 as follows:

$$E(t) = \frac{\sum E_n}{T * n \text{ set}}$$

Check the termination condition. If the $E(t)$ is greater then the error threshold, the training is stopped and the trained NN is ready for test otherwise compute the rescaled LR (RLR) for each iteration using following chaotic time series and equation.

For tent map, we use:

$$ITE(t) = \begin{cases} \alpha \times ITE(t-1), & \text{if } ITE(t-1) < 0.5 \\ \alpha \times (1 - ITE(t-1)), & \text{if } 0.5 \leq ITE(t-1) \end{cases}$$

$$RLR(ITE) = M \times C(ITE) - \beta$$

Where, α is any value within 0 and 2.

And go to step 3.

III. EXPERIMENTAL STUDIES.

A. Introduction to Benchmark datasets

HBP is applied on five benchmark classification problems – breast cancer, heart disease, Australian credit card, glass, diabetes and horse. The datasets are collected from the University of California at Irvine (UCI) Repository of the machine learning database [18] and PROBEN1 [19] which contains 349 Data Sets. Some characteristics of the datasets are listed in Table I. For example, glass is a classification problem having total examples of 214 with 9 aspect and 6 classes. Other problems are arranged in an identical model. The total examples of a dataset are divided into – training examples, assertion examples, and testing examples. The training examples are used to train the NN and the trained NN is tested with the testing examples.

TABLE I
CHARACTERISTICS OF BENCHMARK DATASETS.

Datasets	Number of Examples		
	Total Examples	Input Attributes	Output Classes
Breast Cancer	699	9	2
Diabetes	768	8	2
Glass	214	9	6
Credit Card	690	51	2
Horse	364	58	3

B. Experimental process and Superiority Analysis

A single hidden layered feed-forward NN is taken for each problem. The numbers of input and output units of the NN are equal to the number of attributes and number of classes of the dataset correspondingly. The number of hidden units is taken same as previous work for easy of comparison for several datasets. The weights of the NN are initially randomized in the interval $(-1, +1)$. The training is stopped when the mean square error on the validation set steady in consecutive five iterations. The experimental results are reported in terms of LR and Mean Square Error (MSE). Mean indicate the average values of 20 independent trials. HBP compared with BPfast, BPCL (Tent), BPCL (Logistic), Error Backpropagation (EBP) and standard BP. The LR of BP is selected in such a way that, this fixed value of learning rate is always an intermediate point in the range $[-\beta, M-\beta]$. Training MSE after 100 iterations tabulated for the performance comparison of each method for different testing data set which are tabulated in Table I.

1) *Breast Cancer*: This problem has 699 examples. We take first 150 examples for training and last 150 examples for testing. A NN of 9-4-2 (nine input, four hidden, and two output units) is taken and it is trained. The experimental results are listed in Table II for modified algorithms with respect to different learning rates (η). Fig. 2 shows the training error of BPfast, BPCL (Tent), BPCL (Logistic), Standard BP, EBP and HBP algorithms with respect to 100 iterations. It is shown that HBP overcomes the premature saturation and decreases the training error where SBP and another modified algorithms are saturated and also it's clear from the Fig. 2 that it goes to convergence in less the 10 iteration which is faster than the other methods.

TABLE II
RLR AND MSE OF BPFast, BPCL (Tent), BPCL (Logistic), EBP, SBP AND HBP FOR CANCER PROBLEM.

Algorithm	RLR/LR	MSE
BPfast	0.10	0.0174
	0.15	0.0179
	0.20	0.0176
BPCL (Tent)	-0.03 to 0.17	0.0113
	-0.03 to 0.27	0.0111
	-0.03 to 0.37	0.0098
BPCL (Logistic)	-0.03 to 0.17	0.0106
	-0.03 to 0.27	0.0107
	-0.03 to 0.37	0.0103
EBP	0.10	0.0075
	0.15	0.0073
	0.20	0.0072
Standard BP	0.10	0.0116
	0.15	0.0104
	0.20	0.0110
HBP	-0.03 to 0.17	0.0038
	-0.03 to 0.27	0.0034
	-0.03 to 0.37	0.0037

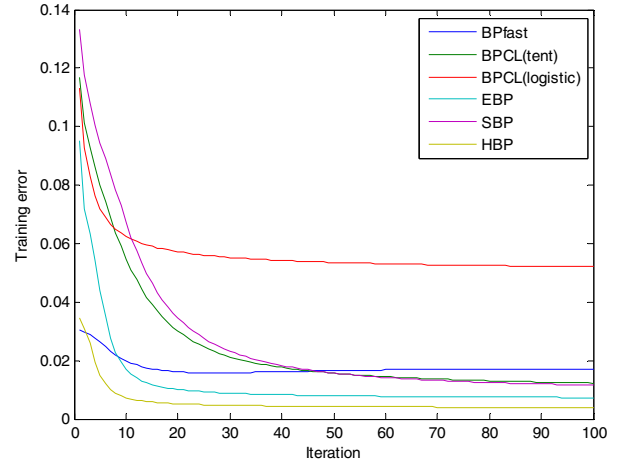


Figure 2. Training error vs. iteration comparison for cancer problem.

2) *Glass*: The size of NN is considered as 9-6-6. This problem has 214 examples. The first 107 examples are used to train the network, 50 examples for validation and the trained network is tested with last 50 examples. The obtained average results are reported in Table III. For example, when RLR is -0.03 to 0.17, MSE of HBP is 0.0035, while these are 0.0443, 0.0481, 0.0448, 0.0235 and 0.0609 for BPfast, BPCL (Tent), BPCL (Logistic), EBP and standard BP respectively. HBP outperforms SBP and all other modified algorithms in all the cases. The convergence curve of HBP is also good as shown in Fig. 3. From the Fig. 3 which stands for the glass problem, it clearly indicates that the rate of convergence of HBP is faster than any methods. It needs 20 iteration to reach optimum level of the training error.

TABLE III
RLR AND MSE OF BPFast, BPCL (Tent), BPCL (Logistic), EBP, SBP AND HBP FOR GLASS PROBLEM.

Algorithm	RLR/LR	MSE
BPfast	0.10	0.0443
	0.15	0.0412
	0.20	0.0400
BPCL (Tent)	-0.03 to 0.17	0.0481
	-0.03 to 0.27	0.0427
	-0.03 to 0.37	0.0423
BPCL (Logistic)	-0.03 to 0.17	0.0448
	-0.03 to 0.27	0.0420
	-0.03 to 0.37	0.0396
EBP	0.10	0.0235
	0.15	0.0199
	0.20	0.0224
Standard BP	0.10	0.0609
	0.15	0.0463
	0.20	0.0450
HBP	-0.03 to 0.17	0.0035
	-0.03 to 0.27	0.0036
	-0.03 to 0.37	0.0033

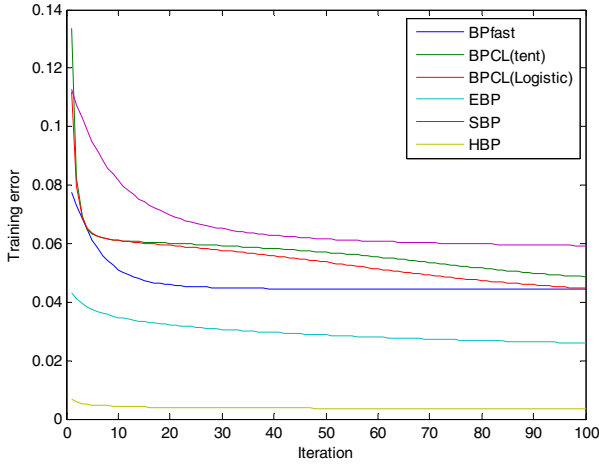


Figure 3. Training error vs. iteration comparison for glass problem.

3) *Credit Card*: A 51-4-2 NN is trained with first 345 examples. The training is stopped with 150 validation examples and the NN is tested with last 150 testing examples. Table IV shows the comparative results between these modified algorithms with HBP algorithm. When RLR is -0.03 to 0.37, MSE of HBP is 0.0071, while these are 0.0376, 0.0247, 0.0171, 0.0146, 0.0264 for BPfast, BPCL (Tent), BPCL (Logistic), EBP and standard BP respectively. Fig. 4 shows the competitive convergence curves of HBP, BPfast, BPCL (Tent), BPCL (Logistic), EBP.

TABLE IV
RLR AND MSE OF BPFast, BPCL (Tent), BPCL (LOGISTIC), EBP, SBP AND HBP FOR CARD PROBLEM.

Algorithm	RLR/LR	MSE
BPfast	0.10	0.0384
	0.15	0.0379
	0.20	0.0376
BPCL (Tent)	-0.03 to 0.17	0.0304
	-0.03 to 0.27	0.0256
	-0.03 to 0.37	0.0247
BPCL (Logistic)	-0.03 to 0.17	0.0289
	-0.03 to 0.27	0.0247
	-0.03 to 0.37	0.0171
EBP	0.10	0.0161
	0.15	0.0164
	0.20	0.0146
Standard BP	0.10	0.0305
	0.15	0.0272
	0.20	0.0264
HBP	-0.03 to 0.17	0.0084
	-0.03 to 0.27	0.0066
	-0.03 to 0.37	0.0071

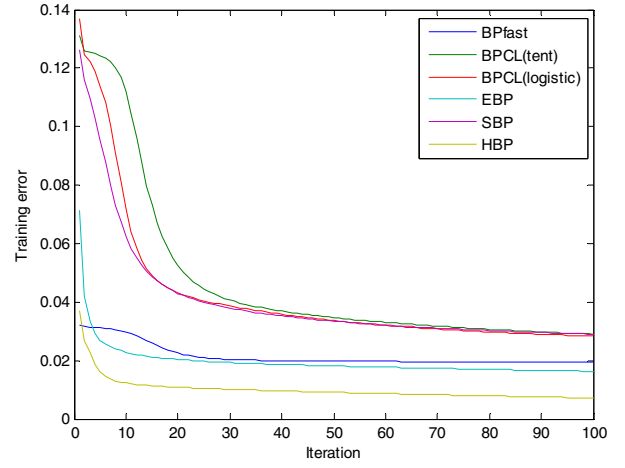


Figure 4. Training error vs. iteration comparison for credit card problem.

4) *Diabetes*: This problem has 768 examples. The first 384 examples are used for training, 192 examples for validation and the endwise 192 examples are used for testing. Here the size of NN is 8-4-2. The results for different LR's are reported in Table V. It is shown that HBP has good generalization ability and also fast convergence rate than all other modified function and SBP also. For example, when RLR is -0.03 to 0.17, MSE of HBP are 0.0018, while these are higher for all other function. Fig. 5 shows the convergence curves.

TABLE V
RLR AND MSE OF BPFast, BPCL (Tent), BPCL (LOGISTIC), EBP, SBP AND HBP FOR DIABETES PROBLEM.

Algorithm	RLR/LR	MSE
BPfast	0.01	0.0286
	0.02	0.0285
	0.03	0.0285
BPCL (Tent)	-0.03 to 0.17	0.0734
	-0.03 to 0.27	0.0694
	-0.03 to 0.37	0.0683
BPCL (Logistic)	-0.03 to 0.17	0.0826
	-0.03 to 0.27	0.0661
	-0.03 to 0.37	0.0777
EBP	0.01	0.1039
	0.02	0.0988
	0.03	0.0977
Standard BP	0.01	0.0510
	0.02	0.0487
	0.03	0.0462
HBP	-0.03 to 0.17	0.0018
	-0.03 to 0.27	0.0016
	-0.03 to 0.37	0.0017

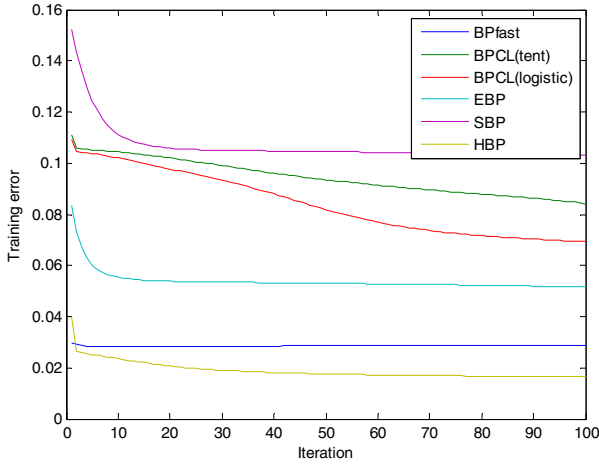


Figure 5. Training error vs. iteration comparison for diabetes problem.

5) *Horse*: The horse dataset has overall 364 examples. The first 182 examples are used for training, 91 examples for verification and last 91 examples for testing. A 58-7-3 NN is trained. The numerical results are reported in Table VI and the error curves are shown in Fig. 6. The outcome of HBP is also superior then other single algorithm.

TABLE VI
RLR AND MSE OF BPFast, BPCL (TENT), BPCL (LOGISTIC), EBP, SBP AND HBP FOR HORSE PROBLEM.

Algorithm	RLR/LR	MSE
BPfast	0.01	0.0133
	0.02	0.0117
	0.03	0.0113
BPCL (Tent)	-0.03 to 0.17	0.0151
	-0.03 to 0.27	0.0060
	-0.03 to 0.37	0.0028
BPCL (Logistic)	-0.03 to 0.17	0.0074
	-0.03 to 0.27	0.0066
	-0.03 to 0.37	0.0014
EBP	0.01	0.0051
	0.02	0.0015
	0.03	0.0023
Standard BP	0.01	0.0167
	0.02	0.0070
	0.03	0.0056
HBP	-0.03 to 0.17	0.0167
	-0.03 to 0.27	0.0167
	-0.03 to 0.37	0.0169

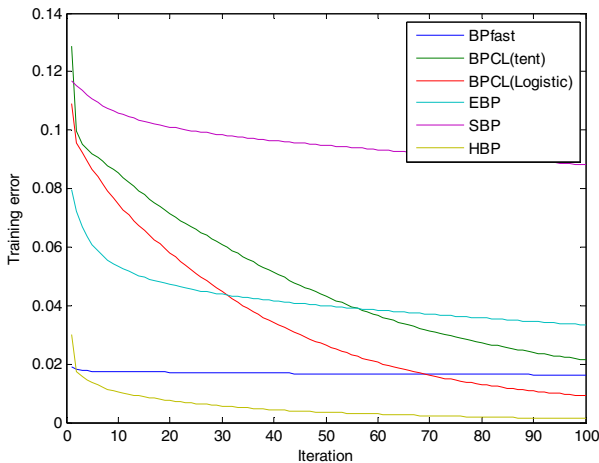


Figure 6. Training error vs. iteration comparison for horse problem.

IV. CONCLUSIONS

In this paper, we studied over several versions of BP and found a combined solution in HBP by using Chaotic BPCL, BPfast, and EBP. We have compared HBP with all other single algorithm with same database and same number of attributes so that we can compare our findings easily. The experimental studies clearly represents that the HBP is better convergence and faster learning than any other single algorithm which makes better generalization capacity for the test patterns. This algorithm could be used for future application of artificial intelligence.

REFERENCES

- [1] Simon Haykin "Neural Networks – A Comprehensive Foundation".
- [2] McCulloch, W. and Pitts, W. "A logical calculus of the ideas immanent in nervous activity." *Bulletin of Mathematical Biophysics*, 7: 115 – 133 (1943).
- [3] "Adeline (Adaptive Linear)", "Introduction to Artificial Neural Networks." *Department of Computer Science, University of Texas at San Antonio. CS 4793: M.*
- [4] Rosenblatt, Frank, "The Perceptron--a perceiving and recognizing automaton". *Report 85-460-1, Cornell Aeronautical Laboratory (1957).*
- [5] Bertsekas, D.P., Tsitsiklis, J.N. "Neuro-dynamic programming, Athena Scientific". pp. 512. ISBN 1 886529-10-8 (1996).
- [6] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation", in *Parallel Distributed Processing: Exploration in the Microstructure of Cognition, vol. 1*, Cambridge, MA:MIT Press, 1986
- [7] Arthur Earl Bryson, Yu-Chi Ho (1969). *Applied optimal control: optimization, estimation, and control.* Blaisdell Publishing Company or Xerox College Publishing. pp. 481.
- [8] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities", *Proceedings of the National Academy of Sciences of the USA*, vol. 79 no. 8 pp. 2554–2558, April 1982.
- [9] X. Zhou, J. Zheng, and K. Mao, "Application of Modified Backpropagation Algorithm to the Prediction of the Chloride Ion Concentration in Cracked Concrete", *In the Proc. of 3rd International Conference on Natural Computation*, pp. 257-261, 2007.
- [10] S. Haykin, "Neural Networks, A Comprehensive Foundation", *New York 10022: IEEE Society Press, Macmillan College Publishing*, 1994.
- [11] S. E. Fahlman, "An empirical study of learning speed in back propagation networks," *tech. rep., CMU-CS-88-162, Carnegie Mellon University, Pittsburgh, PA., 1988.*
- [12] M. Gori and A. Tesi, "On the problem of local minima in backpropagation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 76–86, 1992.
- [13] J. E. Vitela and J. Reifman, "Premature saturation in Backpropagation networks: Mechanism and necessary conditions," *Neural Netw.*, vol. 10, no. 4, pp. 721–735, 1997.
- [14] Sang-Hoon Oh and Youngjik Lee, "A Modified Error Function to Improve the Error Back-Propagation Algorithm for Multi-Layer Perceptrons" *ETRI Journal*, volume 17, number 1, April 1995.
- [15] Mobarakol Islam, Md. Rihab Rana, Sultan Uddin Ahmed, A. N. M. Enamul Kabir, Md. Shahjahan, "Training Neural Network with Chaotic Learning Rate". *JCETECT*, pp. 781 – 785, 2011.
- [16] Olivares, E. I. Vazquez-Medina, R. Cruz-Irisson, and M. Del-Rio-Correa, "Numerical calculation of the lyapunov exponent for the logistic MAP", *12th International Conference on Mathematical Methods in Electromagnetic Theory*, 2008.
- [17] Sultan Uddin Ahmed, Md. Shahjahan, Kazuyuki Murase, "Maximization of the Gradient Function for Efficient Neural Network Training". *ICCIT (AUST)*, pp. 424 – 429, 2010.
- [18] D. E. Rumelhart and J. L. McClelland, *Parallel Distributed Processing*. Cambridge, MA: MIT Press, 1986.
- [19] L. L. Prechelt, "PROBEN1-A set of neural network benchmark problems and benchmarking rules", *Technical Report 21/94, Faculty of Informatics, University of Karlsruhe*, 1994.
- [20] JunshanGao, Baiyu Sun and Jiaxiang Yang, "The Construction of the Cubic Logistic Chaotic Function Family", *In Proc. of Sixth World Congress on Intelligent Control and Automation*, 2006.

A Meta Study of Reduction of Speckle Noise Adopting Different Filtering Techniques

Prajoy Podder

Bangladesh University of Engineering & Technology
prajoypodder@gmail.com

Md. Mehedi Hasan

Khulna University of Engineering & Technology
mehedihasanece@gmail.com

Abstract—Reduction of noise in digital images is a challenging task for the researchers in digital image processing. Speckle noise creates most critical disturbances and affects the quality of synthetic aperture radar, ultrasound and medical images in the field of biomedical imaging. SAR imagery uses the microwave radiation, which receives coherent summation of reflected radiation. Speckle noise in SAR produces degraded, low quality image which is causing misdirection of active radar in automatic target detection and recognition due to image interpretation. Ultrasound images are affected by speckle noise because the images are obtained with a simple linear scan, ultrasound probe which are produced granular appearance noise. Medical images are also corrupted through the speckle noise during its acquisition and transmission. Image with speckle reduce the image contrast and resolution, thereby decreasing the diagnostic values of the ultrasound image. There are several filtering techniques available to reduce this speckle noise. So, this framework includes several filtering techniques like mean, median, Weiner, log and Gaussian to evaluate the filter performance through the parameters like mean square error and peak signal to noise ratio to attenuate the speckle noise without degrading the image information from the synthetic aperture radar, ultrasound and medical imagery. Experimental analysis has been done under MATLAB environment. MSE, PSNR have been calculated and compared for different value of variance for better performance measurement.

Keywords— *De-noising filters; Mean square error (MSE); Peak signal to noise ratio (PSNR); Speckle noise; Synthetic Aperture Radar image; Ultrasound image;*

I. INTRODUCTION

Speckle noise is a common phenomenon in synthetic aperture radar (SAR), ultrasound, laser, acoustic, medical imaging system. Image de-noising has become a very essential tool for modern digital image processing, including medical image, satellite image, radar image and ultrasound image processing. To extract important diagnostic information, de-noising must be done in medical images. Ultrasound imaging is practically harmless and non-invasive to the human beings [1]. But significantly poorer with respect to other medical imaging system. Unfortunately, US images are affected by speckle noise consisting of multiplicative noise appears as a random granular pattern [1]. Radar images are mainly formed by receiving return signal from multiple distributed objects [2]. But SAR images are composed of speckle noise due to random fluctuations of return signal that results, increasing the mean gray level [3]. When a SAR image is formed by coherently processing of the return signal from successive radar impulses,

this effect results a pixel-to-pixel variation in intensity, which establishes itself as a granular pattern termed as speckle or fading [4]. Variation in dark and bright picture element consequence in a SAR image and that unable to control a constant mean radiometric level in consistent areas [5]. There are two main approaches to de-speckle the image using digital image processing. The first method is to apply digital filtering in the frequency domain. In the second methodology, the noise is removed by statistically manipulating the values of nearby pixel. Those frequency domain (wavelet)-based filter and spatial domain (statistics)-based filters are capable of de-speckle the image efficiently. A good comparison of those filters is provided by [6]. The main objective of Image de-speckle techniques is necessary to remove such noises while retaining as much as possible the important signal features. Iris images, Satellite images are generally degraded by noise during image acquisition and transmission process. Various filtering techniques to remove undesirable noise have been discussed in this paper. Efficient method enhances the Signal to Noise ratio and perceives the original features of the images. Filtered iris images from various databases are considered to apply mean, median, Gaussian, Weiner, log, disk filters and have been observed from various aspects by quantitative measuring.

The paper can be organized as follows: section II describes the overview of the working process, section III describes noise model, which are further processed by adopting the mentioned filters. Section IV designates the Mean, Median, Gaussian, Weiner, Log filters which are used for noise reduction, experimental results and discussions followed by section V, where mean square error, peak signal to noise ratio are determined precisely by varying variance and section VI conclude the paper.

II. OVERVIEW OF THE PROPOSED METHOD

1. The original colored (SAR, Medical and Iris) images are given as input images and converted the images to the gray scale from the RGB images.
2. Special noise has been added in the gray scale images with appropriate noise density.
3. Variance of the speckle noise has been calculated from the following equation, where is threshold value, ψ is scale factor, is standard deviation and is variance. Scale factor has been calculated by,
4. The noisy images have been observed.

5. To remove the unwanted noise from the (SAR, Medical and Iris) images de noising filter like (Median, Gaussian, Mean, Log, Disk, Weiner) filters have been applied.

6. The output filtered images have been observed.

7. For observing the performance of the filters, the values of the mean square error (MSE), peak signal to noise ratio (PSNR) have been calculated by varying the variance.

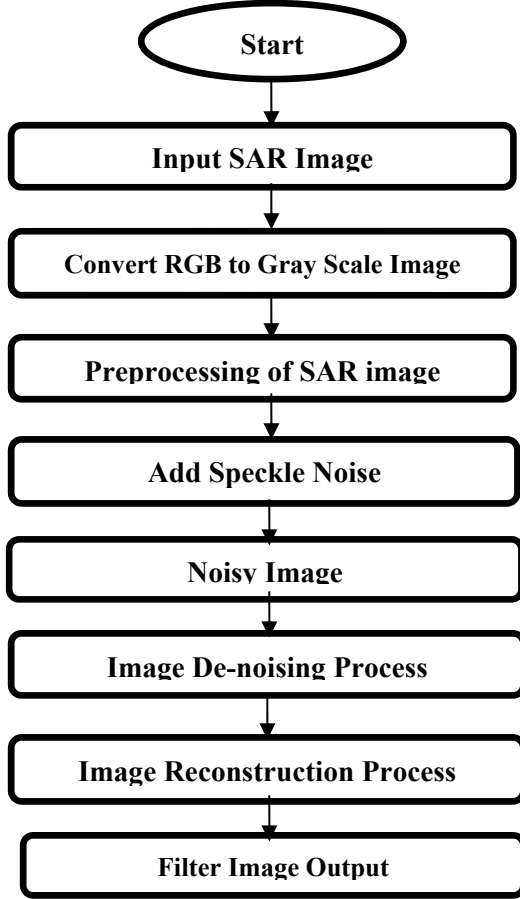


Figure 1: Flow chart of the working process.

Fig. 1 describes the total evaluation process. Medical images or SAR images are used as input images or test images. Speckle noises are added with the gray scale image, after converting the RGB image to the gray scale images. Then the noisy images are observed by choosing the variance of speckle noise. In order to perform the image restoration, process the images de noising filters likes (Wiener, mean, median, LOG, Gaussian) filter have been applied. At last the change of MSE and PSNR have been calculated and observed for different speckle noise variances.

III. NOISE MODEL

Speckle noise is haphazard and deterministic in nature. Speckle has negative impact on ultrasound imaging, SAR imaging, etc. Ultrasound imaging technique is a popular method for the modern medicine. It is used for internal organ of the body, visualization of muscles and injuries. Degradation of speckle noise from the ultrasound imaging system can be expressed by the following equation,

$$g(x,y) = f(x,y) * n(x,y) \quad (1)$$

The characteristics of the medical ultrasound imaging system can be categorized into three classes:

a) The arbitrary appropriated diffusing exists inside the resolution cell of the imaging framework makes the speckle pattern. Such as, blood cells.

b) The second layer of cell tissue diffuses is no haphazardly conveyed with the long-extend order [7]. For instants, lobules in liver parenchyma tissue.

c) A spatially invariant rational structure including inside the irregular scattered region. For example, organ surfaces and blood vessels.

IV. IMAGE DE-NOISING FILTERS

Image de-noising filters have a wide range of applications such as medical image analysis, signal (audio, voice, video) analysis, data mining, radio astronomy and many more. Image de-noising (smoothing) filters are playing an important role in suppressing the high frequency components of images and the sharpening filters are used to suppress the low frequencies in the image. Image restoration and enhancement approaches are classified into spatial domain and frequency domain. The spatial domain noise removal requires very less processing time than frequency domain. Depend on time there are two types of filters: time varying (linear) and time invariant (non-linear). The linear filters are faster processing, but unable to preserve edges. The non-linear filters are preserving edges, but slower processing. Now, six types of filters have been discussed below, which will be used further to remove the noisy information from the original images.

A. Median Filter

The median filter is a simpler technique and it can also remove pulse or spike noise [1] in which the central pixel of a moving window is replaced by its median value. The median filter is better than the mean filter but has limited success when applied to SAR data. Due to replace with the neighbor median value median filter moving all the pixels in the image. The value is calculated by first sorting all the pixels in numerical order and replacing the pixel with considered pixel value. Median filters are effective for the bipolar and unipolar impulse noise.

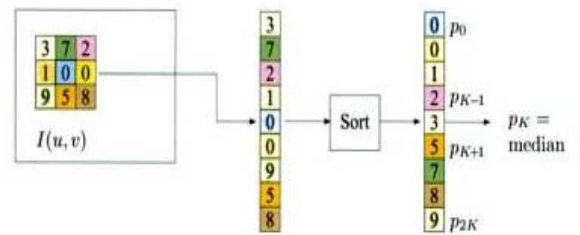


Figure 2: Computation of a 3x3 pixel median filter

Fig. 2 demonstrates the computation of the median filter of size 3x3 pixels.

B. Gaussian Filter

Gaussian filter is a linear smoothing filter, where the weights chosen for the smoothing purpose according to the shape of Gaussian function. Gaussian filter in the continuous space and can be expressed by the following equation,

$$h(m,n) = \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{m^2}{2\sigma^2}} \right) \times \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{n^2}{2\sigma^2}} \right) \quad (2)$$

The impulse response of Gaussian filter has been expressed with the standard deviation as parameter is given below,

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \quad (3)$$

C. Mean Filter

The mean filter also known as averaging filter, which calculate the average of all pixel values within the window. Its mathematical expression can be given as follows,

$$\hat{f}(x,y) = \frac{1}{mn} \sum_{(r,c) \in W} g(r,c) \quad (4)$$

Here, g is the corrupted image, r and c are the row and column co-ordinates respectively within a window size of $m \times n$ and $\hat{f}(x,y)$ is the filtered image. The geometric mean filter mathematical expression can be given as follows,

$$f'(x,y) = \left[\prod_{(r,c) \in W} g(r,c) \right]^{1/mn} \quad (5)$$

D. Wiener Filter

Wiener filter is applied to an image adaptively depending on the variance while smoothing the image. The adaptive filters produce better results than the linear filter, preserving more image features (e.g. edge, high frequency part). However, wiener filter requires more computation time than linear filtering. The transfer function of the wiener filter is given by,

$$H(u,v) = \frac{P_f(u,v)}{P_f(u,v) + \sigma^2} \quad (6)$$

E. Log Filter

The Laplacian of Gaussian is a smoothing filter that can operate with a 2nd spatial derivative of an image. The LOG is often applied to an image to reduce its sensitivity to noise for smoothing the images and make easier for edge detection of the images. The laplacian $L(x,y)$ with pixel intensity $I(x,y)$ can be expressed by,

$$L(x,y) = \frac{d^2 I}{dx^2} + \frac{d^2 I}{dy^2} \quad (7)$$

V. RESULTS & DISCUSSION

Matrix laboratory software (MATLAB) has been used to perform the remove the speckle noise by removal filters like, wiener filter, mean filter, median filter, Gaussian filter, disk filter, log filter.

A. Iris images

Fig. (3,4) describes the effect of filtering sequentially for removing the speckle noise through different filters in CASIA database at variance 0.04 and 0.06. Where, fig. (3(a), 4(a)) represents the original images, fig. (3(b), 4(b)) represents

speckle noise images, fig. (3(c, d, e), 4(c, e, f)) are represents sequentially wiener, mean, gaussian filters output for removing the speckle noise from the CASIA database images, fig. (3(f), 4(g)) represents the disk filter and log filter output for removing the speckle noise from the CASIA database images.

Fig. 5 describes the comparison curves of wiener, mean, median, gaussian, log filters for observing the effect in case of MSE (fig. 5(a)) and PSNR (fig. 5(b)) by varying the variance of images in CASIA database.

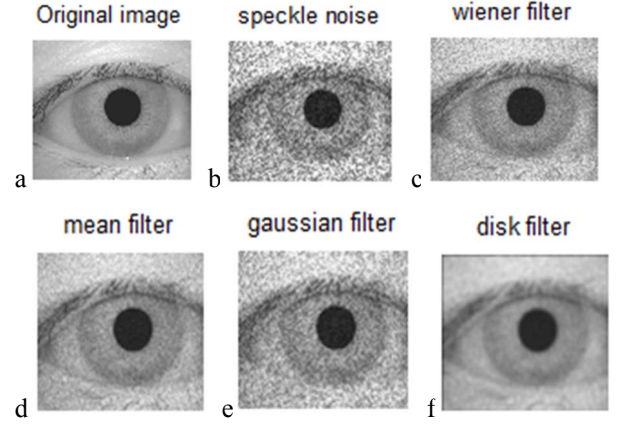


Figure 3: performance analysis of various filters to remove speckle noise at Variance=0.04 in CASIA database.

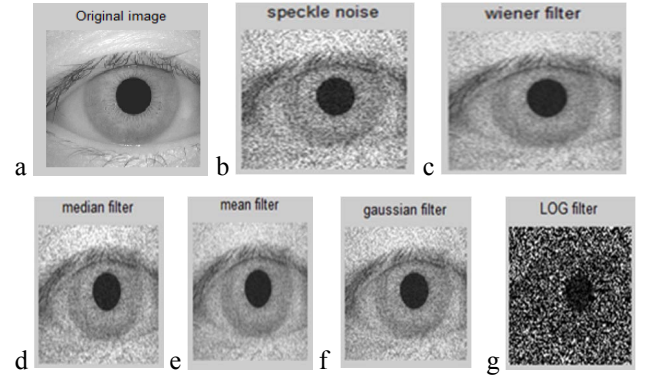
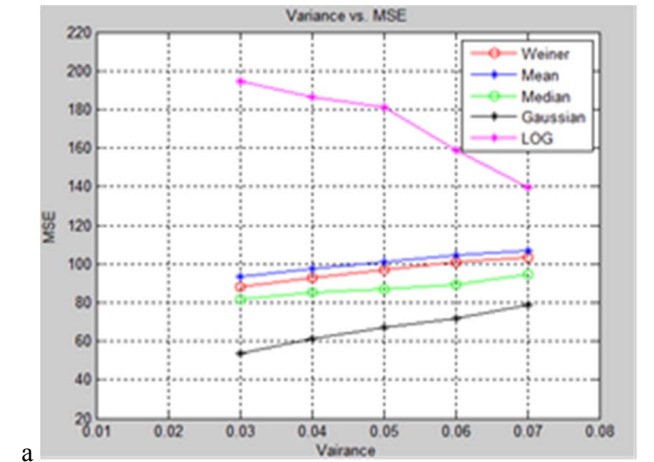


Figure 4: performance analysis of various filters to remove speckle noise at Variance=0.06 in CASIA database.



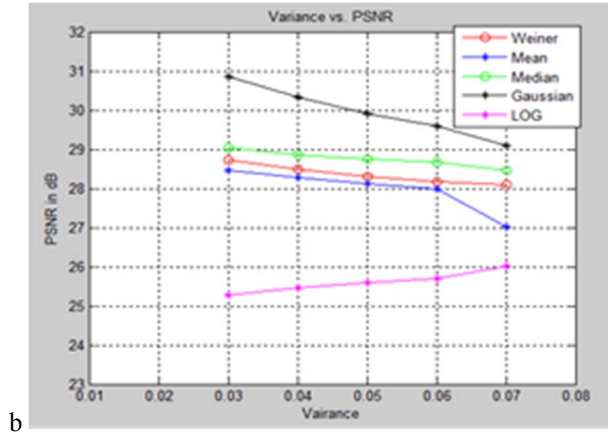


Figure 5: MSE (a) & PSNR (b) comparison curve of various filters at different variance in CASIA database.

Fig. 6 describes the effect of filtering sequentially for removing the speckle noise through different filters in MIMU database at variance 0.06. Where, fig. 6(a) represents the original images, fig. 6(b) represents speckle noise images, fig. 6(c) represents sequentially wiener filter, median filter, mean filter, log filter, gaussian filter output for removing the speckle noise from the MIMU database images.

Fig. 7 describes the comparison curves of winer filter, mean filter, median filter, gaussian filter and log filter for observing the effect in case of MSE (fig. 7(a)) and PSNR (fig. 7(b)) by varying the variance of images in MIMU database.

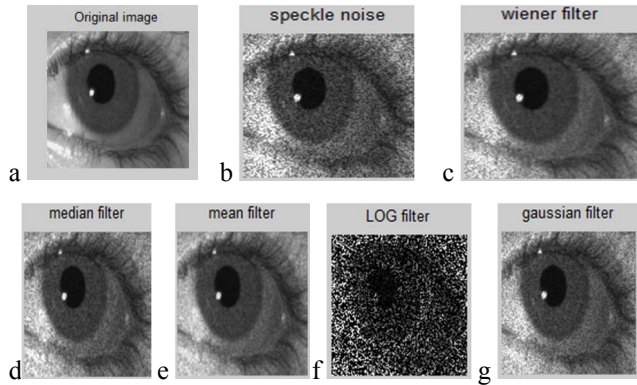
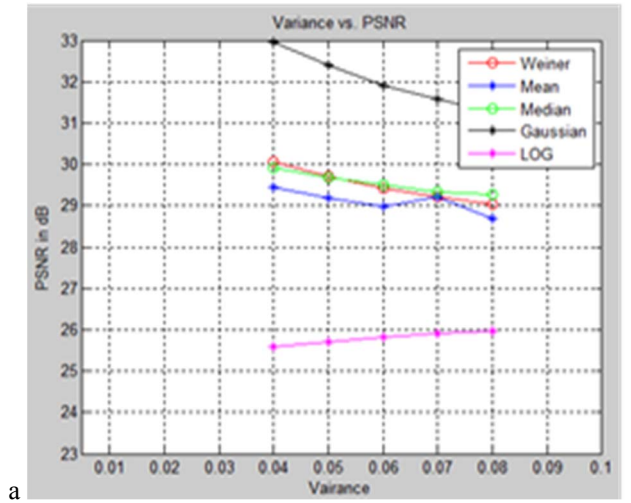
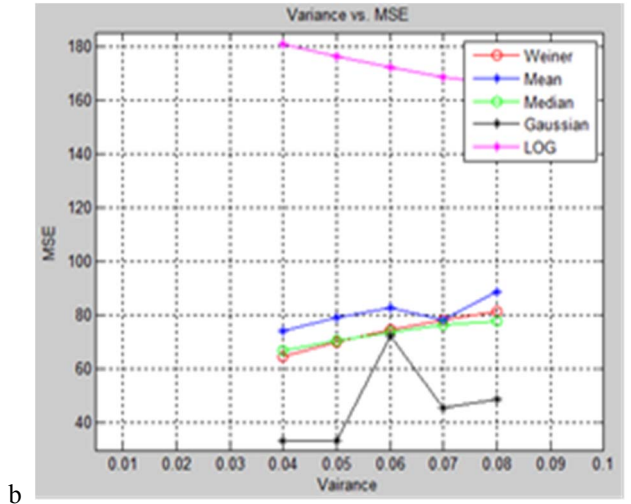


Figure 6: performance analysis of various filters to remove speckle noise at Variance=0.06 in MIMU database.



a

Table (1, 2, 3) are shows the comparisons between the MSE and PSNR for the Synthetic apertures radar (SAR) images and ultrasound images. Where if the variance is increased the value of MSE is increased but PSNR is decreased. Again if the variance is decreased the value of MSE is decreased but the value of PSNR is increased.



b

Figure 7: MSE (a) & PSNR (b) comparison curve of various filters at different variance in MIMU database.

B. Synthetic apertures radar (SAR) images

Fig. (8, 9) describes the effect of filtering sequentially for removing the speckle noise through different filters from the synthetic apertures radar(SAR) images at variance 0.06 and 0.07. Where, fig. (8(a), 9(a)) represents the original images, fig. (8(b), 9(b)) represents speckle noise images of SAR images, fig. (8(c, d, e), 9(c, d, e)) are represents sequentially wiener filter, mean filter, gaussian filter output for removing the speckle noise from the SAR images. Fig. (8(f, g)) represents the log filter and median filter output for SAR images. Fig. 9(f) shows the disk filter output for the SAR images.

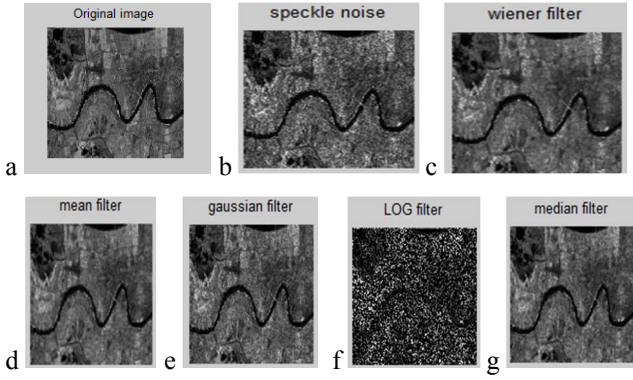


Figure 8: performance analysis of various filters to remove speckle noise at Variance=0.06.

Table I: MSE and PSNR comparisons at Variance=0.06 for SAR images

Filter types	MSE	PSNR(dB)
Weiner	68.73	29.7756871
Mean	76.55	29.3253697
Median	72.90	29.5374017
LOG	159.48	26.1376869
Gaussian	37.33	32.4436565

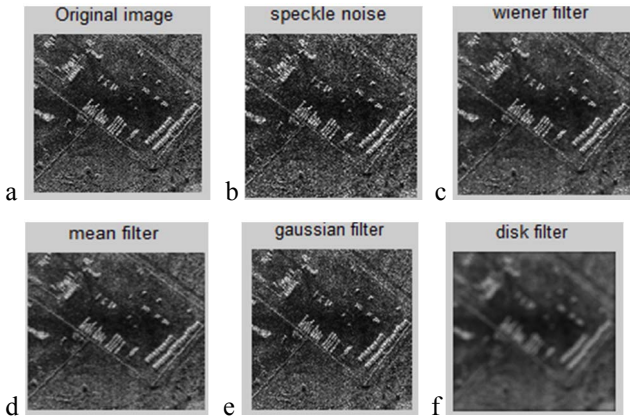


Figure 9: performance analysis of various filters to remove speckle noise at variance=0.07.

Table II: Comparisons between the MSE and PSNR at Variance=0.07 for SAR images (in case of city)

Filter types	MSE	PSNR(dB)
Weiner	66.54	29.9340106
Mean	73.02	29.5345683
Median	75.18	29.4037386
LOG	144.12	26.5774594
Gaussian	40.10	32.1335995

C. Ultrasound images

Fig. (10, 11) describes the effect of filtering sequentially for removing the speckle noise through different filters from the ultrasound images at variance 0.05 and 0.06. Where, fig. (10(a), 11(a)) represents the original images of ultrasound, fig. (10(b), 11(b)) represents speckle noise images for ultrasound images, fig. (10(c, d, e, f, g), 11(c, d, e, f, g)) are represents sequentially Wiener filter, mean filter, log filter, Gaussian filter and median filter output for eliminated the speckle noise from the ultrasound images.

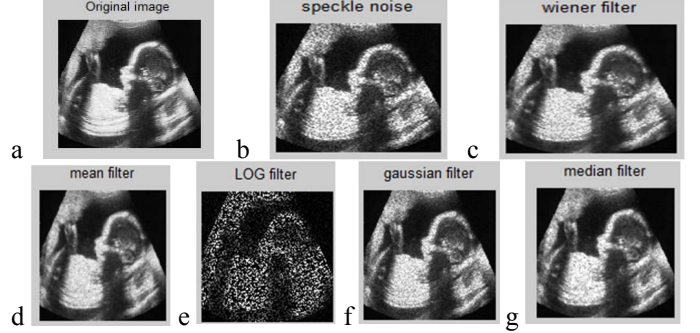


Figure 10: performance analysis of various filters to remove speckle noise at variance=0.05.

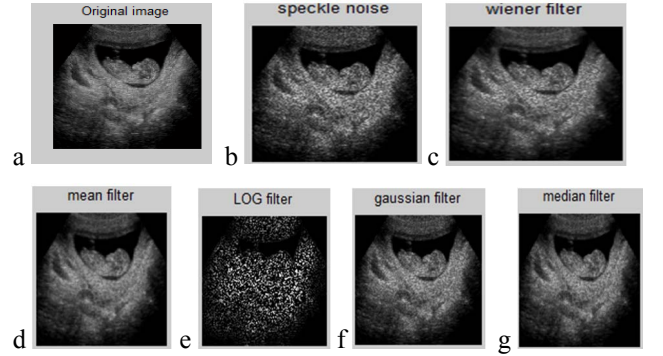


Figure 11: performance analysis of various filter to remove speckle noise at variance=0.06.

Table III: MSE and PSNR comparisons at Variance=0.06 for Ultrasound images

Filter types	MSE	PSNR(dB)
Weiner	35.60	32.6507807
Mean	48.10	31.3434541
Median	46.73	31.4697269
LOG	117.37	27.4693464
Gaussian	31.81	34.7785044

VI. CONCLUSION

Speckle noise is random in nature and deterministic in the imaging system. Texture and organ surfaces have been enhanced for the ultrasound imaging system. This paper mainly focused on the speckle reduction techniques of five different statistical based speckle suppression filters such as wiener, Gaussian, log etc. The working procedure constitutes

main parameters that have been observed on three different images (e.g. Iris, SAR image, Ultrasound image). The Experimental observation under PSNR and MSE shows the efficient filter performance at difference variance. A comparative study has been focused among those filters to effectively de-speckle the images. The outcome of this study reveals that Gaussian filter gives relatively better performance among the other filters in terms of MSE, PSNR and visual quality.

REFERENCES

- [1] R. Sivakumar, D. Nedumaran, "Comparative study of Speckle Noise Reduction of Ultrasound B-scan Images in Matrix Laboratory Environment", International Journal of Computer Applications, Volume 10– No.9, November 2010.
- [2] Askin Demirkol, "Range Density Function for Active Sensor Imaging", Defence Science Journal, Vol. 57, No. 3, May 2007, pp. 305-313.
- [3] Lee, J.-S., Jurkevich, I., Dewaele, P., Wambacq, P., and A. Oosterlinck, "Speckle Filtering of Synthetic Aperture Radar Images: A Review", Remote Sensing Review, Vol. 8, 1994, pp. 313-340.
- [4] Bruniquel, J. and A. Lopes, "Multivariate Optimal Speckle Reduction in SAR Imagery", International Journal of Remote Sensing, Vol. 8, No. 3, 1997, pp. 603-627.
- [5] Md. Mehedi Hasan, Jag Mohan Thakur, Prajoy Podder, "Design & Implementation of FHSS and DSSS for Secure Data Transmission" International Journal of Signal Processing Systems, 2014
- [6] Md. Mehedi Hasan, Jag Mohan Thakur, Prajoy Podder, "Face Detection in Color Images Using Skin Color", IJSER, Volume 5, Issue 6, June-2014.
- [7] Amandeep Kaur, Karamjeet Singh, "SPECKLE NOISE REDUCTION BY USING WAVELETS", NCCI 2010 -National Conference on Computational Instrumentation CSIO Chandigarh, INDIA, 19-20 March 2010.

Caffeine Dose Dependent Blood Perfusion: *Effects on Rhythmic Activities using Frequency Spectrum*

Md. Bashir Uddin

Department of Biomedical Engineering
Khulna University of Engineering & Technology
Khulna-9203, Bangladesh
bas.u.kuet@gmail.com

Mohiuddin Ahmad

Department of Electrical and Electronic Engineering
Khulna University of Engineering & Technology
Khulna-9203, Bangladesh
mohiuddin.ahmad@gmail.com

Abstract— The dependency of blood perfusion on caffeine doses and the effects on different rhythmic activities were investigated by spectral components of skin blood perfusion signal. The skin blood perfusion on human forearm middle finger tip was recorded by laser Doppler flowmetry technique at both before and immediately after the consumption of caffeine dose. Different caffeine doses contained different amount of caffeine but approximately same amount of sugar. Recorded and processed data was investigated using spectral component within the frequency range 0.009 – 1.6 Hz. It was found that the consumption of caffeine free dose increased the heart activity about 85.66% where other activities showed a tiny change. Consumption of low value caffeine dose resulted notable decrement in metabolic, myogenic, and heart activities. Consumption of moderate value caffeine dose resulted significant increment in sympathetic, respiratory, and heart activities. High value caffeine dose consumption resulted remarkable increment in sympathetic (230.26%) and myogenic (263.01%) activities. The results indicated significant decrement and increment for low and high value caffeine dose consumption respectively.

Keywords—*caffeine; blood perfusion; rhythmic activity; frequency spectrum*

I. INTRODUCTION

Caffeine and blood perfusion are two well known terms. A psychoactive drug [1] which is legal, unregulated and widely consumed is known as caffeine. Caffeine is the main ingredient of most of the energy drinks that are widely consumed to get a short term boost in energy. The beverages containing caffeine are commonly known as caffeinated drinks or energy drinks. Caffeinated drinks consumption has significant physical and mental effects. These effects are mainly responsible for caffeine content of consumed drinks.

There is little evidence that caffeine content have any effect. Significant improvements were reported in mental and cognitive performances [2], where excess consumption of caffeinated drinks may induce agitation, anxiety, irritability and insomnia [3]. Headache, nausea, and anxiety are recognized as short term side effects of mild caffeine consumption [4]. The effects of little amount of caffeine in form of caffeinated drinks was investigated and reported significant changes in different rhythmic activities [5].

Caffeine affects cardiovascular responses [6] as well as blood flow [7] in human body in a significant manner. A recent study reported cardiac activity with consuming different doses of caffeine using laser Doppler flowmetry technique [8] but other activities were not investigated.

Physiological and psychological effects of caffeine have been well studied but no research has been conducted to evaluate different rhythmic activities: metabolic, sympathetic, myogenic, respiratory, and heart by consuming caffeine or different doses of caffeine. Besides, no research has been done to investigate the dependency of blood perfusion with varying the amount of caffeine. That's why; the main objective of this study is to investigate the effects of different caffeine dose consumption on different rhythmic activities as well as to evaluate caffeine dose dependency of blood perfusion. It was hypothesized that different doses of caffeine would have different impacts on rhythmic activities.

II. SKIN BLOOD PERFUSION AND RHYTHMIC ACTIVITIES

Blood perfusion is the process of a body delivering blood to a capillary bed in its biological tissue. Skin blood perfusion measured on the skin indicates the blood flow in microvascular level. Laser Doppler flowmetry (LDF) is an established and reliable method for the measurement of skin blood perfusion. The principle of LDF technique is based on reflected laser light from blood. Low power laser light is used to illuminate tissue using a fiber optic; the light is scattered by the static tissue structures and moving blood cells; the moving blood cells impart a Doppler shift; an adjacent fiber detects light returned from the tissue; this light contains Doppler shifted and unshifted light; the signal is processed to extract the signal related to the moving red blood cells.

Periodic oscillations in microvascular level are detected by noninvasive technique of LDF. The spectral analysis of LDF from human forearm skin has revealed five characteristic frequencies [9]. In addition to cardiac and respiratory rhythms around 1 and 0.3 Hz, respectively [10], three frequencies have been detected in the regions around 0.1, 0.04, and 0.01 Hz in human skin [11]. It is suggested that periodic oscillations with a frequency of around 0.1 Hz reflect intrinsic smooth muscle

(myogenic) activity of blood vessels [12], whereas the frequency around 0.04 Hz represents neurogenic (sympathetic) stimulation of resistance vessels and oscillations of around 0.01 Hz (minute-rhythm) result from changes in metabolism (metabolic activity) of the perfused tissue [13]. Five different activities can be demonstrated by frequency spectrum of LDF signal as listed in Table I. This frequency domain analysis of LDF signal is an established way to evaluate different activities of human body. Fast Fourier transform (FFT) is a frequency based analysis to compute frequency component of non-stationary signal. The discrete Fourier transform of discrete time signal $x(j)$ is noted in (1) [8].

$$X_k = \sum_{j=n_s}^{n_e-1} x(j) e^{-i2\pi \frac{jk}{N}} \quad (1)$$

Here, k represents the harmonic number of frequency components of signal and X_k is the discrete frequency domain at different frequency.

III. MATERIALS AND METHODS

A. Preparation of Caffeine Doses

Different soft and energy drinks were used in this study to prepare different doses of caffeine because it was very tough to get the direct form of caffeine. Different brands of energy drinks have different amount of caffeine as the main ingredient. Different doses of caffeine were made using different brands of energy drinks as well as soft drinks. Table II shows the preparation of different doses of caffeine indicating corresponding amount of caffeine only. According to ingredient list of different brands, the listed amount of caffeine was made. Using same amount of drinks (250 ml), all doses of caffeine were prepared which may include a single drinks or combination of two drinks. Five experimental recording days was designed for corresponding five doses of caffeine as listed in Table II. Before 5 minutes of consumption, the corresponding dose of caffeine was prepared.

TABLE I. FREQUENCY RANGES OF RHYTHMIC ACTIVITIES

Different Rhythmic Activities	Frequency Ranges
Metabolic	0.009 – 0.02 Hz
Sympathetic (Neurogenic)	0.02 – 0.06 Hz
Myogenic	0.06 – 0.2 Hz
Respiratory	0.2 – 0.6 Hz
Heart (Cardiac)	0.6 – 1.6 Hz

TABLE II. CAFFEINE DOSES WITH CAFFEINE CONTENTS

Different Caffeine Doses (250 ml)	Caffeine (mg)
Sprite (250 ml) – Day 1	00
Pepsi (250 ml) – Day 2	27
RedBull (100 ml) + Pepsi (150 ml) – Day 3	48
RedBull (200 ml) + Sprite (50 ml) – Day 4	64
RedBull (250 ml) – Day 5	80

B. Recording Setup and Data Collection

In this study, healthy young male subjects free from severe physiological and psychological diseases were enrolled. Subjects were free from smoking and refrained from alcohol and caffeine containing drinks at least 20 hours prior to the test. For physical examination and medical history assessment, each participant had an initial visit to the experimental laboratory. The experiment was performed in a quite laboratory room. Throughout the whole experimental period the subjects were resting in the supine position. A familiarization day was conducted at first to remove “first-day effect” and the study design was totally unknown to the participants to minimize the effects of biases. Approximately 10 minutes were allowed for acclimatization and then LDF measurements were performed on the skin of forearm middle finger tip with consuming caffeine dose. For each caffeine dose, skin blood perfusion was measured before and immediately after the consumption. Skin blood perfusion was measured using data acquisition unit, LDF amplifier, probe, and relevant software. The recorded signals contained several noises such as 50Hz frequency component. Using filter, this high frequency component was removed. Smoothing technique was also used to find smoothed signal.

IV. EVALUATION OF RHYTHMIC ACTIVITIES

A. Caffeine Dose Dependent Blood Perfusion

In this experiment, skin blood perfusion signal using LDF was recorded with the duration of 5 minutes and 25 minutes before and immediately after the consumption of each caffeine dose respectively. Typical skin blood perfusion recordings of human forearm middle finger tip before and immediately after the consumption of caffeine free dose (0 mg caffeine) are shown in Figs. 1 & 2 respectively. It is observed that the skin blood perfusion (LDF signal) is oscillating with an average value around 1100 BPU (blood perfusion unit) and 1150 BPU at before and immediately after the consumption of caffeine free dose respectively. The mean value of skin blood perfusion has increased around 50 BPU due to the consumption of a dose of 0 mg caffeine. Besides, a tiny increment in peak to peak values of skin blood perfusion can be noticed due to the consumption of same dose of caffeine.

Typical skin blood perfusion recordings of human forearm middle finger tip before and immediately after the consumption of 27 mg caffeine dose are shown in Figs. 3 & 4 respectively. It is observed that the skin blood perfusion is oscillating with an average value around 1000 BPU and 1200 BPU at before and immediately after the consumption of 27 mg caffeine dose respectively. The mean value of skin blood perfusion has increased around 200 BPU due to the consumption of a dose of 27 mg caffeine. Besides, a tiny decrement in peak to peak values of skin blood perfusion can be noticed due to the consumption of 27 mg caffeine dose. It is very difficult to investigate the effects of caffeine doses on different activities using these time domain data, so further analysis or spectral (frequency domain) analysis is required.

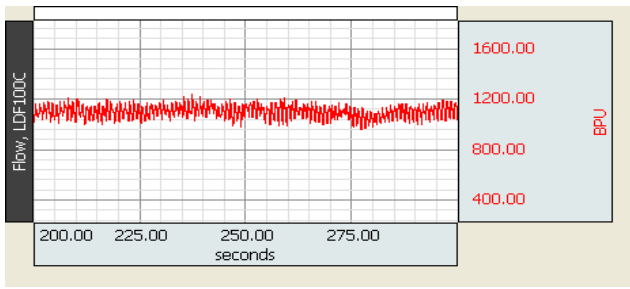


Fig. 1. Skin Blood perfusion signal before consuming caffeine free dose.

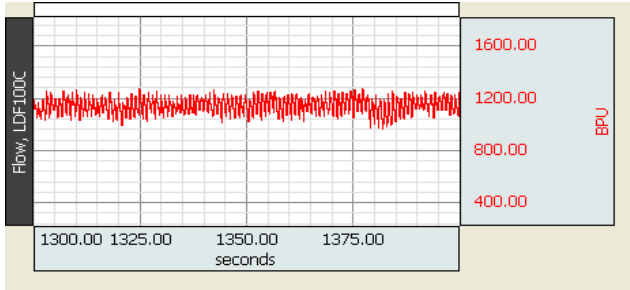


Fig. 2. Skin Blood perfusion signal after consuming caffeine free dose.

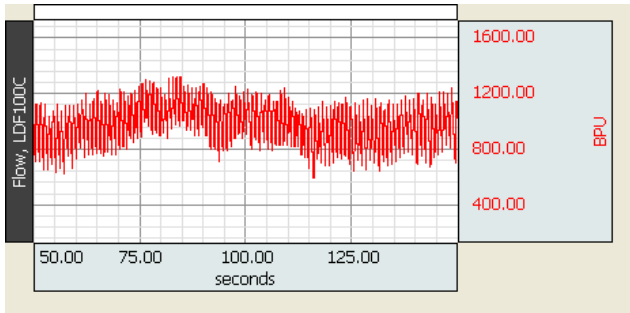


Fig. 3. Skin Blood perfusion before consuming 27 mg caffeine dose.

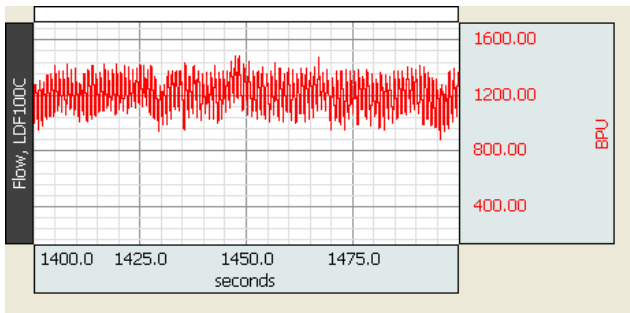


Fig. 4. Skin Blood perfusion after consuming 27 mg caffeine dose.

B. Frequency Spectrum of Blood Perfusion

Different activities can be investigated by spectral analysis (frequency domain analysis) of skin blood perfusion signal. Spectral analysis of skin blood perfusion signal from human forearm middle finger tip is effective to evaluate different activities within the frequency range 0.009 to 1.6 Hz. In this experiment, fast Fourier transform (FFT) is used as spectral analysis. Magnitudes of FFT of skin blood perfusion signal

before and immediately after the consumption of a dose of 0 mg caffeine are shown in Figs. 5 & 6 respectively. It is seen that the peak magnitudes of metabolic, sympathetic, myogenic, respiratory, and heart activities are 15.96, 2.65, 1.98, 0.81, and 2.86 BPU respectively before the consumption of 0 mg caffeine dose. After the consumption of 0 mg caffeine dose, the peak magnitudes of above activities are 14.82, 3.09, 1.85, 0.88, and 5.31 BPU respectively.

Magnitudes of FFT of skin blood perfusion signal before and immediately after the consumption of a dose of 27 mg caffeine are shown in Figs. 7 & 8 respectively. The peak magnitudes of metabolic, sympathetic, myogenic, respiratory, and heart activities are 23.9, 6.76, 6.54, 1.83, and 17.37 BPU respectively at before, 13.14, 6.96, 4.77, 2.01, and 13.52 BPU respectively at after the consumption of 27 mg caffeine dose.

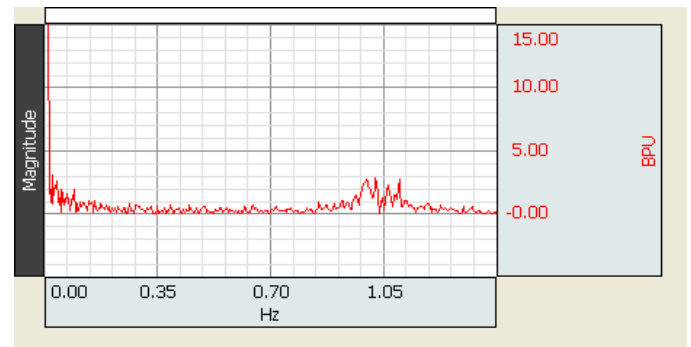


Fig. 5. Spectral analysis of LDF signal before consuming 0 mg caffeine.

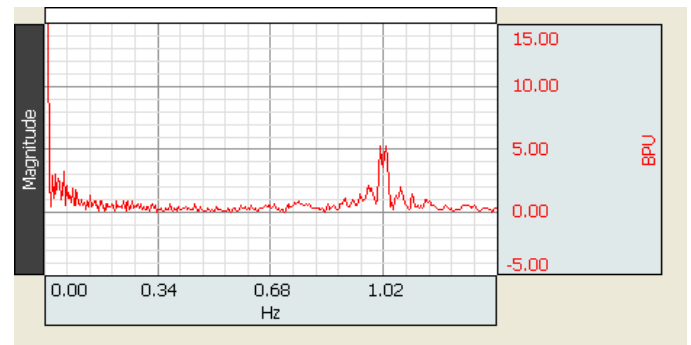


Fig. 6. Spectral analysis of LDF signal after consuming 0 mg caffeine.

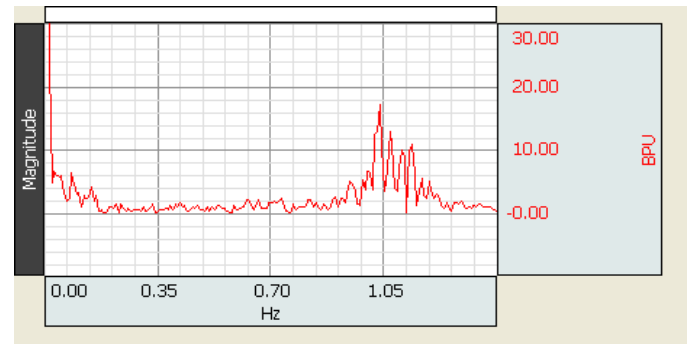


Fig. 7. Spectral analysis of LDF signal before consuming 27 mg caffeine.

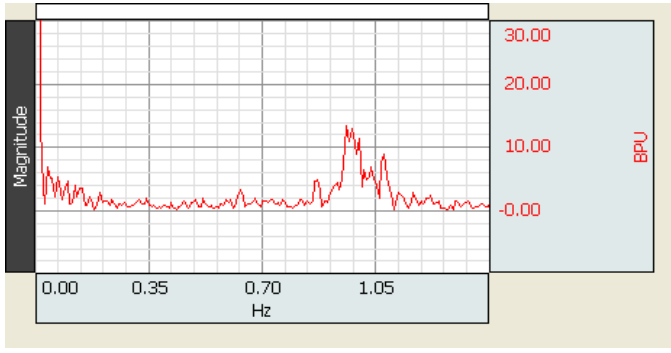


Fig. 8. Spectral analysis of LDF signal after consuming 27 mg caffeine.

C. Caffeine Doses and Rhythmic Activities

The changes in different rhythms with different amount of caffeine are listed in Table III. Peak magnitude of spectral component at before and after the consumption of different caffeine doses are listed here for different rhythmic activities. It is seen that sympathetic, respiratory, and heart activities are increased due to the consumption of caffeine free dose with a tiny decrement in metabolic and myogenic activities.

TABLE III. VARIATIONS IN RHYTHMIC ACTIVITIES WITH CAFFEINE DOSE

Amount of Caffeine (mg)	Rhythmic Activities	Peak Magnitude of Spectral Component (BPU)	
		Before consumption	After consumption
0	Metabolic	15.96	14.82
	Sympathetic	2.65	3.09
	Myogenic	1.98	1.85
	Respiratory	0.81	0.88
	Heart	2.86	5.31
27	Metabolic	23.9	13.14
	Sympathetic	6.76	6.96
	Myogenic	6.54	4.77
	Respiratory	1.83	2.01
	Heart	17.37	13.52
48	Metabolic	9.75	9.98
	Sympathetic	1.50	3.13
	Myogenic	2.19	1.92
	Respiratory	0.66	0.89
	Heart	8.93	12.85
64	Metabolic	11.72	21.52
	Sympathetic	4.70	0.76
	Myogenic	2.28	2.76
	Respiratory	1.24	0.71
	Heart	24.45	16.40
80	Metabolic	13.77	15.9
	Sympathetic	0.76	2.51
	Myogenic	0.73	2.65
	Respiratory	0.94	1.03
	Heart	12.07	15.36

Dose of 27 mg caffeine results notable decrement in metabolic, myogenic, and heart activities as well as tiny increment in sympathetic and respiratory activities. The percentage changes in different rhythmic activities can be calculated with respect to normal condition (before drink) as noted in (2).

$$\% \text{ Change} = \frac{(PM)_{AD} - (PM)_{BD}}{(PM)_{BD}} \times 100 \quad (2)$$

Here, $(PM)_{BD}$ and $(PM)_{AD}$ indicate the peak magnitude of spectral component respectively at before drink and after drink of caffeine dose. If the % change in (2) is positive, this indicates the increment as well as negative value indicates decrement in different rhythmic activities.

V. RESULTS AND DISCUSSIONS

Different rhythms have been investigated by spectral analysis of skin blood perfusion signal. This analysis has been done to determine the effects of different caffeine doses consumption on different rhythmic activities. It is well known that all the effects (physiological and psychological) of drinks (soft and energy drinks) found are due to caffeine and sugar content. It is noticed that all the caffeine doses consumed in this study contain approximately same amount of sugar and increasing order of caffeine (Table II). So the effects due to the consumption of these doses are totally responsible for caffeine variation but not for sugar content.

Effects on metabolic, sympathetic, and myogenic rhythms the percentage changes in peak magnitude of spectral component for metabolic, sympathetic, and myogenic activities are shown in Fig. 9. Metabolic activity shows a tiny decrement (7.14%) for caffeine free dose (0 mg caffeine) consumption, significant decrement (45.02%) for low value caffeine dose (27 mg caffeine) consumption and significant increment (83.62%) with the intake of high value caffeine dose (64 mg caffeine) consumption. Sympathetic activity also shows a tiny increment (16.60%) for caffeine free dose (0 mg caffeine) consumption, significant increment (108.67%) for moderate value caffeine dose (48 mg caffeine) consumption, significant decrement (83.83%) for high value caffeine dose (64 mg caffeine) consumption and remarkable increment (230.26%) with the intake of excessively high value caffeine dose (80 mg caffeine) consumption. Similarly, myogenic activity shows significant decrement (27.06%) for low value caffeine dose (27 mg caffeine) and remarkable increment (263.01%) with the intake of excessively high value caffeine dose (80 mg caffeine).

Effects on respiratory and heart rhythms the percentage changes in peak magnitude of spectral component for respiratory and heart activity are shown in Fig. 10. Respiratory activity shows a tiny increment (8.64%) for caffeine free dose (0 mg caffeine) consumption, significant increment (34.85%) for moderate value caffeine dose (48 mg caffeine) consumption and significant decrement (42.74%) with the intake of high value caffeine dose (64 mg caffeine) consumption. Heart activity shows remarkable increment (85.66%) for caffeine free dose (0 mg caffeine) consumption, significant decrement

(22.16%) for low value caffeine dose (27 mg caffeine) consumption, significant increment (43.90%) for moderate value caffeine dose (48 mg caffeine) consumption, significant decrement (32.92%) for high value caffeine dose (64 mg caffeine) consumption and significant increment (27.26%) with the intake of excessively high value caffeine dose (80 mg caffeine) consumption.

It is clear that caffeine free dose may results significant increment in heart activity but fails for other activities. Low and high value caffeine dose may result significant decrement and increment in metabolic and myogenic activities respectively. Besides, moderate and high value caffeine dose may result successive increment and decrement in sympathetic, respiratory, and heart activities.

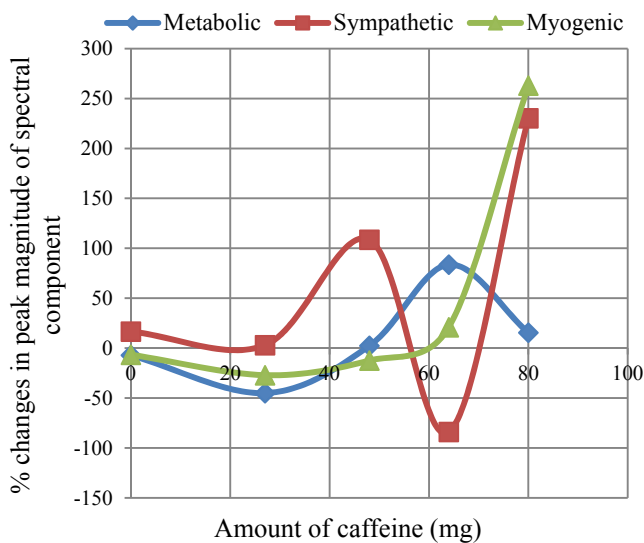


Fig. 9. Variations in metabolic, sympathetic, and myogenic activity.

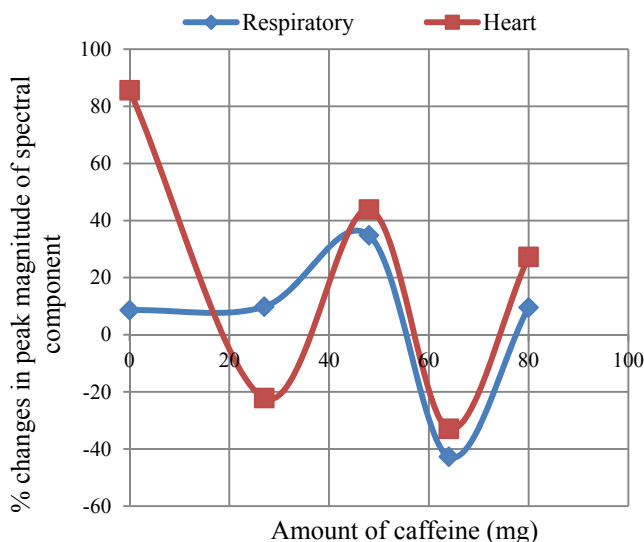


Fig. 10. Variations in respiratory and heart activity.

VI. CONCLUSIONS

The effects of different caffeine doses on different rhythmic activities are investigated in this study. It is clear that the variations in the amount of caffeine consumed have significant impacts on different rhythmic functions of human body. The consumers of different caffeinated beverages should be aware about the impacts of caffeine doses that they often consumed. The implication of the findings is that it will be helpful for consumers to choose suitable beverages to consume.

ACKNOWLEDGMENT

Authors wish to thank all participants in this study and are cordially grateful to those persons who are directly or indirectly associated to this study. For proving facilities to conduct experiment, authors are also grateful to Biomedical Engineering Department of KUET, Bangladesh. This research is partially supported by UGC, HEQEP, Sub-project "Postgraduate Research in BME" CP#3472 under the department of Biomedical Engineering.

REFERENCES

- [1] A. Nehlig, J. L. Daval, and G. Debry, "Caffeine and the central nervous system: mechanisms of action, biochemical, metabolic and psychostimulant effects," *Brain Research Reviews*, vol. 17, no. 2, pp. 139–70, 1992.
- [2] M. A. Howard and C. A. Marcinski, "Acute effects of glucose energy drink on behavioral control," *Experimental and clinical psychopharmacology*, vol. 18, no. 6, pp. 553–561, 2010.
- [3] F. Van Den Eynde, P. C. Van Baelen, M. Portzky, and K. Audenaert, "The effects of energy drinks on cognitive performance," *Tijdschrift voor psychiatrie*, vol. 50, no. 5, pp. 273–281, 2008.
- [4] R. E. Shapiro, "Caffeine and headaches," *Current Pain and Headache Reports*, vol. 12, no. 4, pp. 311–315, 2008.
- [5] M. M. Islam, M. B. Uddin, and M. Ahmad, "Determination of the Effects of Having Energy Drinks by Analyzing Blood Perfusion Signal," *Journal of Computers*, vol. 9, no. 8, pp. 1817–24, 2014.
- [6] T. L. Whitsett, C. V. Manion, and H. D. Christensen, "Cardiovascular effects of coffee and caffeine," *Am J Cardiol.*, vol. 53, no. 7, pp. 918–22, 1984.
- [7] J. W. Daniels, P. A. Mole, J. D. Shaffrath, and C. L. Stebbins, "Effects of caffeine on blood pressure, heart rate, and forearm blood flow during dynamic leg exercise," *J Appl Physiol*, vol. 85, no. 1, pp. 154–9, 1998.
- [8] M. B. Uddin, M. N. Reza, and M. Ahmad, "Effects of Caffeine Doses on Cardiac Activity using Laser Doppler Flowmetry," *Int. Conf. on Computer & Information Engineering*, pp. 118–21, Rajshahi, Bangladesh, 26–27 November, 2015.
- [9] M. Bracic and A. Stefanovska, "Wavelet based analysis of human blood flow dynamics," *Bulletin of Mathematical Biology*, vol. 60, pp. 417–433, 1998.
- [10] A. Stefanovska and P. Kroselj, "Correlation integral and frequency analysis of cardiovascular functions," *Open systems & information dynamics*, vol. 4, pp. 457–478, 1997.
- [11] M. E. MuXck-Weymann, H. P. Albrecht, D. Hager, D. Hiller, O. P. Hornstein, and R. D. Bauer, "Respiratory-dependent laser- Doppler flux motion in different skin areas and its meaning to autonomic nervous control of the vessel of the skin," *Microvascular Res.*, vol. 52, pp. 69–78, 1996.
- [12] E. G. Salerud, T. Tenland, G. E. Nilsson, and P. A. Oberg, "Rhythmical variations in human skin blood flow," *Int J Microcirc Clin Exp.*, vol. 2, no. 2, pp. 91–102, 1983.
- [13] J. Kastrop, J. Bulow, and N. A. Lassen, "Vasomotion in human skin before and after local heating recorded with laser Doppler flowmetry. A method for induction of vasomotion," *Int J Microcirc Clin Exp.*, vol. 8, no. 2, pp. 205–15, 1989.

Design and Implementation of Three-Stage Battery Charger for Lead-Acid Battery

Md Shahrier Hakim¹, Farhana Latif², Md. Imran Khan³ and Al Basir⁴

¹Dept. of Electrical Power Engineering, University of Greenwich, United Kingdom.

²Dept. of Power Engineering, BTU Cottbus, Germany.

³IICT, Bangladesh University of Engineering and Technology (BUET), Dhaka-1000, Bangladesh.

⁴School of Engineering and Materials Science, Queen Mary University of London, United Kingdom.

E-mail: ¹shahrierh@yahoo.com, ²farhanalatif@ymail.com, ³imrankhankuet@gmail.com, ⁴al.basir005@yahoo.com

Abstract—Charging method is crucial for any batteries. Over the years, many charging algorithm are developed to improve the charging method of lead acid battery. Uncontrolled charging of lead acid battery may lead to capacity loss and also reduce the life cycle of battery. To improve the charging method a simple battery charging algorithm is proposed in this paper. The IC (UC3906) is the core of the designed circuit to implement the algorithm. The result shows that the designed circuit based on the algorithm is effective during overcharging and supports the steady charging concept without consuming access charges.

Keywords—Lead-Acid, overcharging, battery safety, battery life, three-stage charging.

I. INTRODUCTION

Charging of a battery is very much important because it ensures a healthy life of the battery. Especially for applications such as Electric vehicle, it is more important where charging is often required. Sometimes unstable charging can lead to overcharging or high duration of charging [1, 2]. Even a blast of battery due to uncontrolled charging can cost life or heavy injuries. Therefore, control of battery charging is necessary. A proper charging algorithm can set control to the battery charging and can help avoiding over charging or less efficient charging. Batteries charge state should be determined accurately, otherwise, its life cycle will be affected [3]. There are some algorithm exist addressing this issue. However, most of them are either too much complex or not suitable for protecting the battery from overcharging [1, 4]. Therefore, in this paper we have proposed a simple battery charging algorithm and have shown its potential in reality.

This paper is organized as follows, section II describes battery charging systems and section III shows implementation of the algorithm. Battery charging components, design consideration and charge status indicator are presented in section IV part A, B and C respectively. Section V illustrates circuit drawing and PCB layout. Section VI explains the results and finally section VII is for conclusion.

II. BATTERY CHARGING SYSTEMS

With the advances of technology and requirements, different types of batteries are developed. Those batteries have a wide range of uses, for example, in toys, ignition, pocket

lights, PDAs, and cameras [5]. Suitable charging algorithm is chosen based on the frequency and purpose of use of a battery in a system [1]. One of the main tasks of a battery charger is to delivering appropriate charge. Besides this, a battery charger must have capability to optimize the rate of charge and to terminate the charge [1]. Different charging schemes can be used to deliver the charge to the battery of single chemistries or multiple chemistries. The purpose of a multi-chemistry charger is to charge the battery for various types of transportable appliances. The main challenge of using this kind of charging is identification of the size and type of the battery [2] which helps in insuring the charging scheme and the termination point. Additional techniques are needed to achieve this. On the other hand, single chemistry chargers have some advantages over multi-chemistry charger, e.g. it is more simple and reliable [1]. Irrespective of battery chemistry, it is important to consider the types of battery, the capacity of the battery and the application of the battery for the choice of appropriate charging algorithm [4]. Constant voltage method and constant current method are the conventional charging method [6]. Presently, other methods like constant current-constant voltage method (two-step charging) and pulse method are getting much attention for better charging. In a constant current (CC) algorithm a number of cells are connected in series [7]. Low charging currents are supplied to charge the batteries. This algorithm can be extended for the time for charging [2]. This scheme continued to charge the battery even when the battery is charged fully. Special care was needed for this charging; otherwise, overcharging can damage the battery or degrade the battery life [7]. The constant voltage (CV) charging can be implemented easily. In this algorithm, when the battery reaches to its default value, the charging voltage is hold and the charging current starts to decrease with time [7]. So, the disadvantage of CC algorithm is overcome with CV charging. But one of the drawbacks of CV algorithm is initial high charging current [2]. This high current can exceed the value for safe charging current for the battery and can damaged the plate of the battery and degraded the battery life [8]. A different approach namely constant current-constant voltage (CC-CV) composed of two charging mode helps in overcoming the drawbacks of CC and CV algorithms [2, 7, 8]. In this method, initially the charger provides a constant current until it reaches to a set value. When it reaches to that value,

the charging current starts to decrease until the battery is completely charged. High current bulk charge is responsible for transferring most of the energy [9]. Thus the damaging current is prevented both at the initial stage and at the end stage [8]. A small amount of trickle charge is considered to compensate the self-discharge of a battery [10]. Though charging process is slower than CV algorithm but it is a safe algorithm for battery. For this algorithm, voltage regulation is a key factor. Pulse charging is another algorithm which could be used instead of CC and CV charging [11]. Charging current is sent as pulses to the battery in this process. A charged pulse is followed by a discharged pulse with short duration and high amplitude. Charging time can be reduced by using large pulse current [7]. Dedicated control algorithm and converter are necessary for this method which is costly than other algorithms [8]. With a boost and a buck regulator, dual level charging can be implemented [2]. Boost converter is used in first stage and buck converter is used in second stage. IC L6561 and UC3906 are capable of providing two stage charging [12]. Among them UC3906 is dedicated for the Lead-Acid battery charging. Along with constant voltage and constant current capability this IC is also capable of charge the battery with trickle charge which was important safety feature for charging a battery. Moreover, this IC can maintain the standby or float state when the battery was fully charged and started to charge the battery automatically when the battery voltage dropped below 10 percent of float state threshold value.

III. ALGORITHM FOR THE BATTERY CHARGING SYSTEM

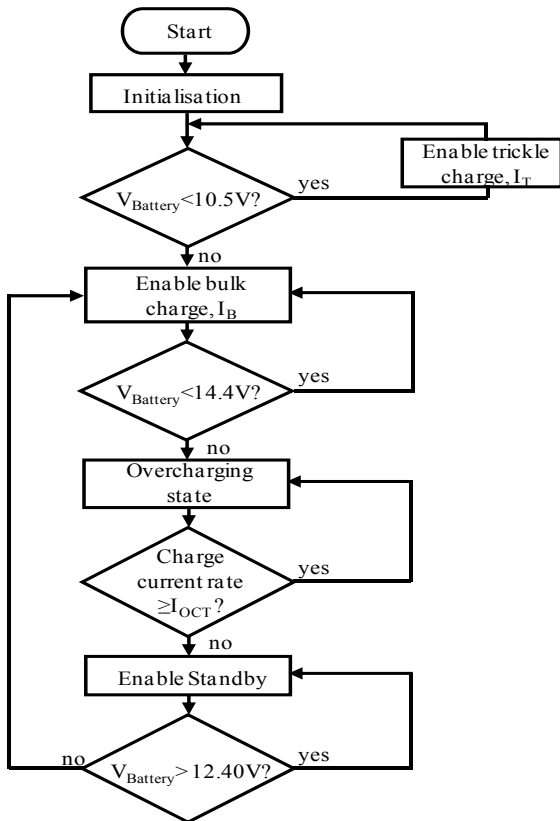


Fig. 1. Flowchart for 12V battery charging algorithm.

This charging algorithm consists of different states. Broadly, it can be explained as follows:

Step 1: Detect the voltage of the battery.

Step 2: Start charging from the appropriate state.

Step 3: If the threshold voltage for present state is reached, start the next state.

Step 4: If the battery is fully charged, keep the battery at standby.

Step 5: If the voltage drops below threshold level, start charging the battery.

Implementation of circuits was included with calculation of required components for the system, selection of components, construction of the circuits and unit testing of circuits. Based on this implementation for this system is discussed below.

IV. IMPLEMENTATION OF THE ALGORITHM

A. Battery Charging Components

A dual level float charger IC (UC3906) is selected for this system as the IC which is capable of charging the battery in three stages namely bulk charge state, over-charge state and float state. Also a trickle charge state is added to the system to turn on low current charging mode at the start. The required values to design the battery charger are calculated and shown in table 1.

TABLE I. REQUIRED BATTERY PARAMETERS FOR 12V 2.3AH LEAD ACID BATTERY

Parameter	Description	Value
V_{battery}	Nominal battery Voltage	12V
NC	No of cells	6
C_{RATE}	Battery Capacity	2.3Ah
V_T	Trickle charge voltage	$V_T = (6 \times 1.75)V = 10.5V$
V_F	Float Voltage	$V_F = (2.3 \times 6)V = 13.80V$
V_{OC}	Overcharge Voltage	$V_{OC} = (2.4 \times 6)V = 14.4V$
I_{max}	Bulk Charge rate	$I_{\text{max}} = C_{\text{RATE}}/5 = 460mA$
I_T	Start-up trickle current	$I_T = C_{\text{RATE}}/100 = 23mA \approx 25mA$
I_{OCT}	OC terminate current	$I_{OCT} = C_{\text{RATE}}/50 = 46mA$

B. Design Considerations

Design of Trickle state: The trickle charge voltage, V_T is calculated for 10.5V. When the voltage sense comparator (Pin 13) sense the battery voltage below V_T , the internal circuitry of the IC disabled the driver and enabled the trickle bias output (Pin 11). A resistor is connected with Pin 11 to get the expected trickle charge current. R_T is calculated according to [13].

$$R_T = \frac{(V_{IN} - V_T - 2.5)}{I_T} = \frac{(24 - 10 - 2.5)}{0.025} = 460\Omega$$

A 464 Ω metal film resistor with 0.125 power rating and 1% tolerance is chosen as R_T .

Design of Bulk charge state: When the battery voltage reached V_T , the internal circuit of the IC disables Pin 11. It also enables the driver and CHARGE ENABLE Pin (Pin 12). In this state a constant current, I_{max} is supplied to the battery. This current is determined with a 250mV offset in the current limit comparator, C/L (Pin 4) and an external sensing resistor, R_S . IC datasheet [13] provides the equation of R_S .

$$R_S = \frac{0.250V}{I_{MAX}} = \frac{0.250V}{0.46A} = 0.55\Omega$$

This R_S is connected in series with Pin 4. This resistor can carry high currents. So, a 0.56 Ω WW type resistor with 5W power rating and 5% tolerance is selected to ensure the stability during temperature rise.

Design of Over-charge state: The overcharge voltage, V_{OC} is calculated for 14.4 V. The voltage sense comparator (Pin 13) sense the voltage of the battery and at these state current needs to decrease which could be done by external resistor divider. IC datasheet [13] is used to calculate the resistor values.

$$R_C = \frac{2.3V}{I_d} = \frac{2.3V}{70\mu A} = 32.857K\Omega \approx 33K\Omega$$

A 33K Ω metal film resistor with 1W power rating and 1% tolerance is chosen as R_C . Now,

$$R_A + R_B = R_{SUM} = \frac{(V_F - 2.3)}{I_d} = \frac{(13.8 - 2.3)}{70 \times 10^{-6}} = 164.285K\Omega \approx 165K\Omega$$

$$R_D = \frac{(2.3 \times R_{SUM})}{(V_{OC} - V_F)} = \frac{(2.3 \times 165 \times 10^3)}{(14.4 - 13.8)} = 632.5K\Omega \approx 633K\Omega$$

A 634K Ω metal film resistor with 600mW power rating and 1% tolerance is selected as R_D .

Again,

$$R_A = (R_{SUM} + R_X) \left(1 - \frac{2.3}{V_T}\right) = [(165 \times 10^3) + (31 \times 10^3)] \times \left(1 - \frac{2.3}{10}\right) = 150.920K\Omega \approx 150K\Omega$$

$$\text{Where, } R_X = \frac{(R_C \times R_D)}{(R_C + R_D)} = \frac{(33 \times 10^3) \times (633 \times 10^3)}{(33 \times 10^3) + (633 \times 10^3)} = 31K\Omega$$

A 150K Ω carbon film resistor with 1W power rating and 5% tolerance is chosen as R_A .

$$R_B = (R_{SUM} - R_A)$$

$$= (165 \times 10^3) - (150 \times 10^3) = 15K\Omega$$

A 15K Ω wire wound resistor with 7W power rating and 5% tolerance is chosen as R_B .

Design of Standby or Float state: To terminate the over-charge state and initialize the float state internal current sense comparator, C/S (Pin 2) is used. When the charge current falls below the specified level I_{OCT} this state is terminated. This current is sensed by the 25mV offset in C/S comparator and the external resistor R_{S1} . So,

$$R_{S1} = \frac{0.025V}{I_{OCT}} = \frac{0.025V}{46mA} = 0.55\Omega$$

As $R_{S1} = R_S$, same resistor should be used with Pin 2.

The C/S output (Pin 1) generates a signal when the charger current went below I_{OCT} . This signal needs to make OVER-CHARGE TERMINATE pin (Pin 8) high. So, pin 1 needs to be connected with pin 8 of the IC. As the voltage at Pin 8 rise to high, it turns off the state level control pin (pin 10) in order to bypass the external resistor R_D (connected to Pin 10). This regulates the charger output voltage to float voltage, V_F .

Pass device selection: To charge the battery in bulk state, high current (460mA) is required. To handle this high current a pass device is necessary with sufficient power handling capability. Maximum power required for this system is $P = V \times I = 24V \times 0.46A = 11.04W$. TIP32A (PNP transistor) is chosen as a pass device which had power dissipation capability of 40W with maximum collector current 3.0A and maximum collector-emitter voltage 60V.

Diode selection: To prevent any reverse current from the battery, a diode needs to be used in series with the pass element. A 1N5007 is chosen which is able to handle current up to 1A.

C. Charge Status Indication

From the data sheet it is found that, in different charging state pin 1, pin 9 and pin 10 go to high state. Therefore, a charge status indicator can be implemented. By connecting the base of the transistor in the mentioned pins, charge status indicator is done. BC517G NPN transistor is chosen for this purpose. A logic table was shown in table 2.

TABLE II. LOGIC TABLE FOR CHARGE STATUS INDICATOR

Charging state	Pin 1	Pin 9	Pin 10
Bulk charging	0	1	0
Over-charging	0	0	1
Standby	1	0	0

V. CIRCUIT DRAWING AND PCB LAYOUT

ISIS is used to draw the circuit diagrams. All the components, except L6207 and UC3906, are available in the ISIS library. Both of IC packages are designed in ISIS by using the information from respected datasheet. After selecting and placing the components on the design sheet, they are interconnected together with the help of connecting tool based on the calculated design options as described previously.

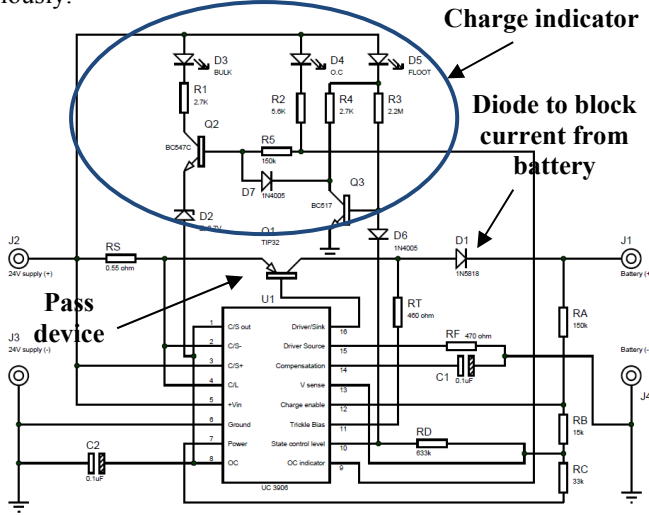


Fig. 2. Schematic capture of the battery charging circuit.

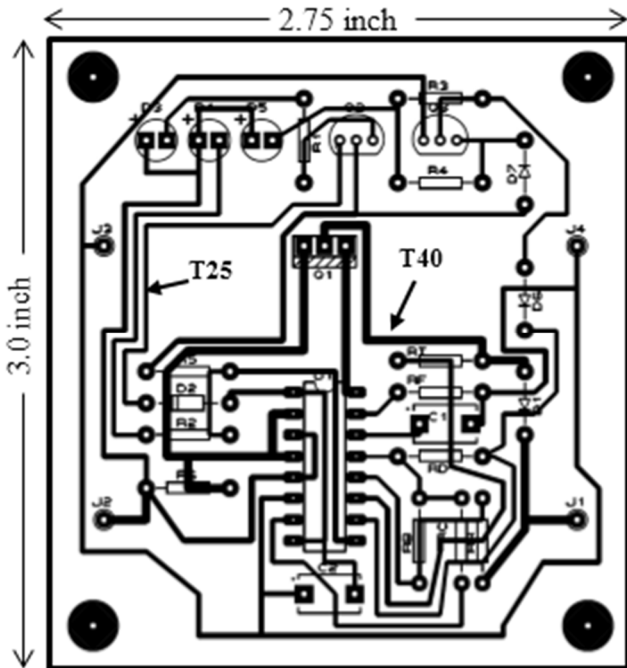


Fig. 3. PCB layout of battery charging circuit.

After drawing the circuits in ISIS separately, the netlists are transferred to ARES. At first, the “board edge” layer is chosen which defined the size of the PCB. Then all the components are placed manually in the defined board edge area. The terminals that need external access are placed at the edge of the PCB. After that, routing is done manually. When the conductor routing is finished checking routine is carried out to ensure that the design was done accordingly specified designs and standard. For battery charging circuit (Figure 2), single layer is used. During the layout design it is ensured that there should be enough space for placing heat sink properly. The size of this PCB was 2.75×3.0 inch. T40 is used for high current tracks and T25 is used for other tracks.

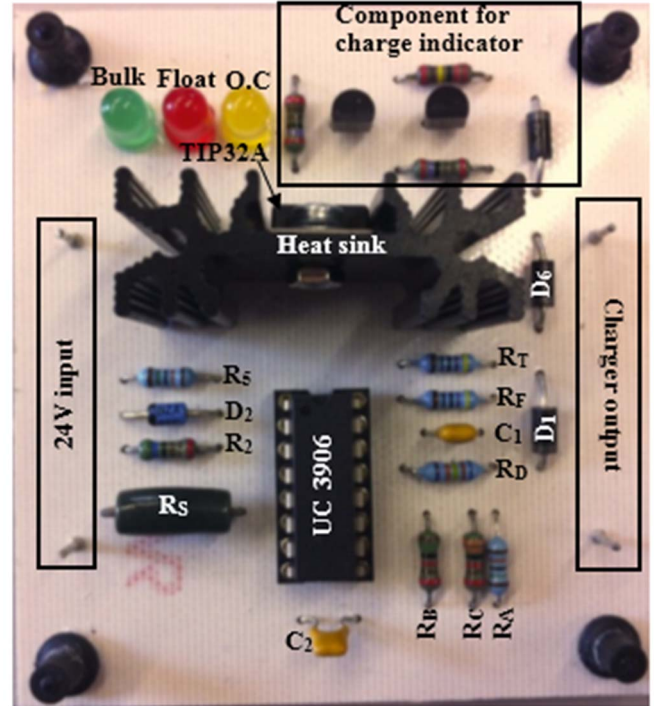


Fig. 4. Pictorial view of implemented battery charger.

VI. RESULTS

A 12V, 2.3Ah lead acid battery was used. To test the battery charger circuit, a 24V was supplied to input pin and output voltage was measured 13.8V. From the datasheet [13] it is found that without any battery connected to output, the output will be at float state and here 13.8V is the float voltage. Therefore, the circuit supports the minimum requirement for such a charging system. In the next step, the battery charger was tested with the 12V battery to check the effectiveness of the charger. Before the test, the battery was fully charged. It was discharged to 8.97V first and then it was connected to the charger. The output current from the charger was measured which is plotted against time and shown in figure 5. Initially, a small amount of current (approximately 0.25mA) was supplied to the battery from the charger to maintain the trickle state. It complied with the design as the current started to

increase and leapt to around 0.5A after certain period (T1) and then charged the battery (i.e. bulk charges), stayed almost constant for a while. However, there was a significant decrease in current and stopped to charge the battery after T3 period. Therefore, the charging does not overcharge the battery and keeps the battery standby when the battery is fully charged.

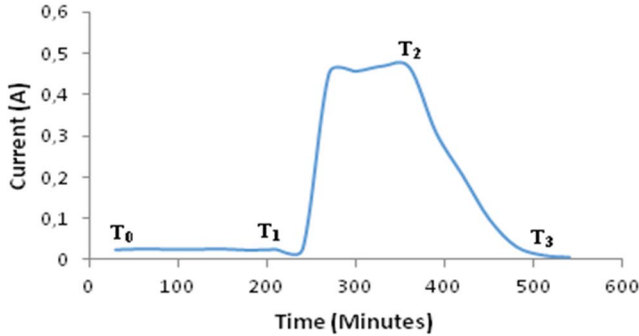


Fig. 5. Output current of the Charger.

Also, the battery voltage was recorded during the charging within every 30 minutes interval. This result is shown in figure 6. The battery was fully charged when the battery voltage reached 14.4V. It is noticeable that the battery started charging and kept increasing the amount of voltage slowly tills around 10.5V. After that there was a gradual increase in voltage and maintained a constant voltage which was represented by the steady voltage at the peak as it was fully charged. This indicates a seemingly good shape for any voltage charging characteristics.

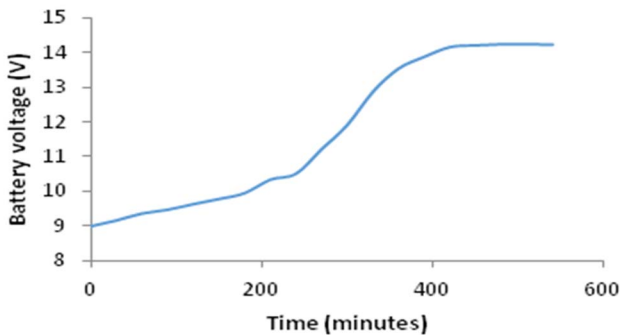


Fig 6: Time vs battery voltage.

VII. CONCLUSION

The proposed battery charging algorithm is very simple to understand. The main goal was to ensure reliable and safe charging of a vehicle battery. The result suggests that the designed circuit is capable of keeping a fixed voltage at the end and ensure efficient current consumption. The inclusion of trickle charge detection in the algorithm helps to avoid a fully charged battery. However, the circuit does not include any other protection mechanism. Especially for chemical battery, the amount of chemical and charge concentration is very much

crucial. The present circuit can be further enhance with suitable mechanism to detect chemical quantity and charge density. In addition, the present circuit is tested only for 12 volt battery. Necessary circuit design is required for any other voltages.

REFERENCES

- [1] A. A.-H. Hussein and I. Batarseh, "A review of charging algorithms for nickel and lithium battery chargers," *IEEE Transactions on Vehicular Technology*, vol. 60, pp. 830-838, 2011.
- [2] P.-H. Cheng and C.-L. Chen, "A high-efficiency fast charger for lead-acid batteries," in *IECON 02 [Industrial Electronics Society, IEEE 2002 28th Annual Conference of the]*, 2002, pp. 1410-1415.
- [3] G. Yifeng and Z. Chengning, "Study on the fast charging method of lead-acid battery with negative pulse discharge," in *Power Electronics Systems and Applications (PESA), 2011 4th International Conference on*, 2011, pp. 1-4.
- [4] J.-S. Moon, J.-H. Lee, I.-Y. Ha, T.-K. Lee, and C.-Y. Won, "An efficient battery charging algorithm based on state-of-charge estimation for electric vehicle," in *Electrical Machines and Systems (ICEMS), 2011 International Conference on*, 2011, pp. 1-6.
- [5] J. Ripley, M. R. Ansari, and J. Dehn, "Battery chargers and batteries for DC and AC back-up power systems," in *Petroleum and Chemical Industry Conference, 2001. IEEE Industry Applications Society 48th Annual*, 2001, pp. 139-144.
- [6] T. Suzuki, H. Ikeda, H. Yoshida, and S. Shinohara, "Optimum load impedance for full-bridged MOS-FET power inverter operating at 1.0 MHz to 2.5 MHz," in *Circuits and Systems, 1991., Proceedings of the 34th Midwest Symposium on*, 1991, pp. 174-178.
- [7] C.-C. Hua and M.-Y. Lin, "A study of charging control of lead-acid battery for electric vehicles," in *Industrial Electronics, 2000. ISIE 2000. Proceedings of the 2000 IEEE International Symposium on*, 2000, pp. 135-140.
- [8] C. Mi, M. A. Masrur, and D. W. Gao, "Special Hybrid Vehicles," in *Hybrid electric vehicles: principles and applications with practical perspectives*, ed: John Wiley & Sons, 2011.
- [9] B. J. Masserant and T. Stuart, "A maximum power transfer battery charger for electric vehicles," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 33, pp. 930-938, 1997.
- [10] P. Notten, H. Bergveld, and W. Kruijt, "Battery Management Systems: Design by modeling," *Kluwer Academic Publisher*, 2002.
- [11] I. A. Khan, "Battery chargers for electric and hybrid vehicles," in *Power Electronics in Transportation, 1994. [Proceedings]*, 1994, pp. 103-112.
- [12] B.-Y. Chen and Y.-S. Lai, "New digital-controlled technique for battery charger with constant current and voltage control without current feedback," *IEEE Transactions on Industrial Electronics*, vol. 59, pp. 1545-1553, 2012.
- [13] Texas Instruments. "Sealed Lead-Acid Battery Charger," [www.ti.com](https://www.ti.com/lit/ds/symlink/uc2906.pdf). [Online]. Available: <https://www.ti.com/lit/ds/symlink/uc2906.pdf> [Accessed: May 19, 2016].

High Performance Charger Design of a Dual Feed BLDC Motor Driven Electric Vehicle

MD Nasmus Sakib Khan Shabbir, Shafi Md. Kawsar Zaman, Md. Shafayet Hossain, Anik Dey Sarker, Tanvir Ahmed

Bangladesh University of Business and Technology

eng.sakib@gmail.com

Abstract—This paper presents the designing details of a soft switched fast charging PV charger and a PFC rectifier for Dual Feed BLDC Motor Driven Electric Vehicle. The vehicle is motorized by batteries which are charged through-conventional power source and PV module. In order to-facilitate fast charging and to reduce power loss substantially, a four channel interleaved boost converter has been-integrated with solar module where switching losses are minimized using zero voltage switching (ZVS) technique. In addition to this, a discontinuously switched vienna rectifier has been designed that guarantees reduced THD and improved PF. The speed control of the BLDC motor is done using hall sensor technique. The process of selecting the suitable components for this application is studied and simulated for various conditions where each of the components is practical.

Keywords—PFC, double loop control, ZVS, interleaved technique, BLDC motor, Hall sensor.

I. INTRODUCTION

Because of the hazardous particles emitted by the conventional fuel run vehicle, modern world is trying to shift into a renewable source of energy for vehicle. Electric vehicle and hybrid electric vehicle are some of the alternatives. In these multi feed vehicles, sources are generally classified into two types i.e. Main Energy System (MES) and Rechargeable Energy Source (RES) [1]. In this paper, a fast and power efficient charger has been designed for a BLDC motor driven electric vehicle where a dc battery is feed by a three phase rectifier acting as MES and PV cells as RES. As Solar irradiation and so PV cell's terminal voltage is not constant, a double feed boost converter is used which provides faster response facility during the running of the electric vehicle. To lessen switching losses ZVS technique is used here. On the other hand, large dc capacitor used in rectifier to reduce output ripple inject huge amount of harmonics in input current [2][3] that results in increased apparent power and reduced power factor(PF) [4]. PF is improved by a type of Vienna rectifier. Hall sensor technique is used to control the speed of the BLDC motor. In this paper, the modeling and the control of each portion is presented in detailed manner.

II. CHARGER DESIGN

A. Rectifier design with improved power factor

According to 'IEEE Std 519-1992 Harmonic Current Limits' for a power supply system if I_{s0}/I_L is between 50 to 100

max allowable THD for input current is 12% and for voltage it is 5%. Magnitude of the 5th and 7th current harmonics should not be greater than 10% of the fundamental component and harmonics order in any magnitude greater than 35th must not be injected in the system [10]. To satisfy the IEEE standard a pulse area modulated switching controller has been implemented which is capable of supplying regulated 72 V, 30A. Details of the circuits are described below:

1) Main circuit:

The main circuit is consist of a three phase Y-Y transformer, three ac inductor, three ac capacitor connected in Y, three IGBT switches(R,Y,B), a freewheeling diode (Dm), a dc inductor, 12 diodes and a dc capacitor. During operation, two IGBT conducts simultaneously. Because of the presence of freewheeling diode, regenerative process is absent—in this rectifier.

2) Control strategy

In the controller, output voltage is fed back to compare with the reference voltage. The differential voltage is then passed through a PID controller. Output is then modulated by a multiplier with 15° shifted input voltage and pass to the comparator as a current reference signal. Current signal is generated through the pulse area modulation (PAM) of 1V (P) 15KHz triangular wave and dc inductor current. PAM provides lesser dc inductor requirements and lessens the effect of inductor current ripple. Comparators outputs then pass to the switching generators. Switching generator is consist of an NAND gate, an OR gate and a resettable D flip flop. Switching generator use a,c; b,a and c,b signals to generate signals for R,Y and B switch respectively.

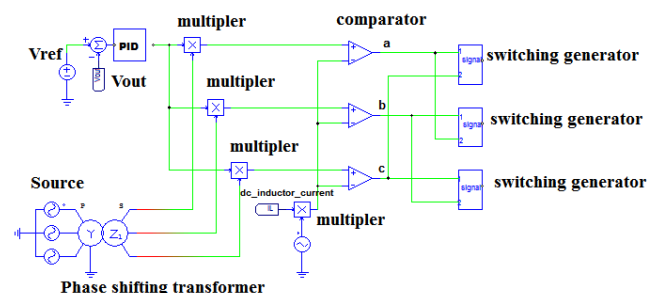


Figure 1. Block diagram of the controller

3) Switching generators

Switching generator needs two inputs. NAND and OR outputs of two inputs are connected with D and reset button respectively and first input is to the set button.

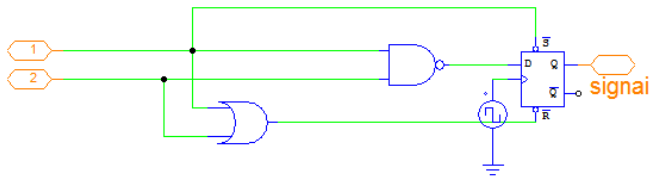


Figure 2. Switching generator

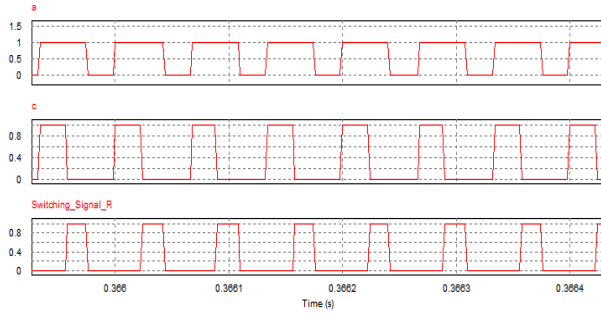


Figure 3. Switching signal generation for switch R

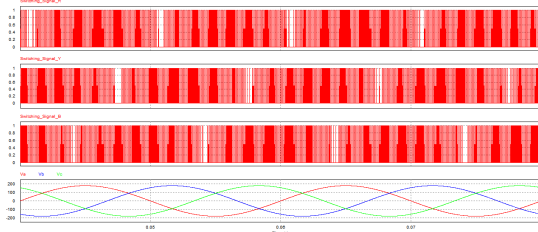


Figure 4. Discontinuous switching cycle

Above bolded regions means switches are on.

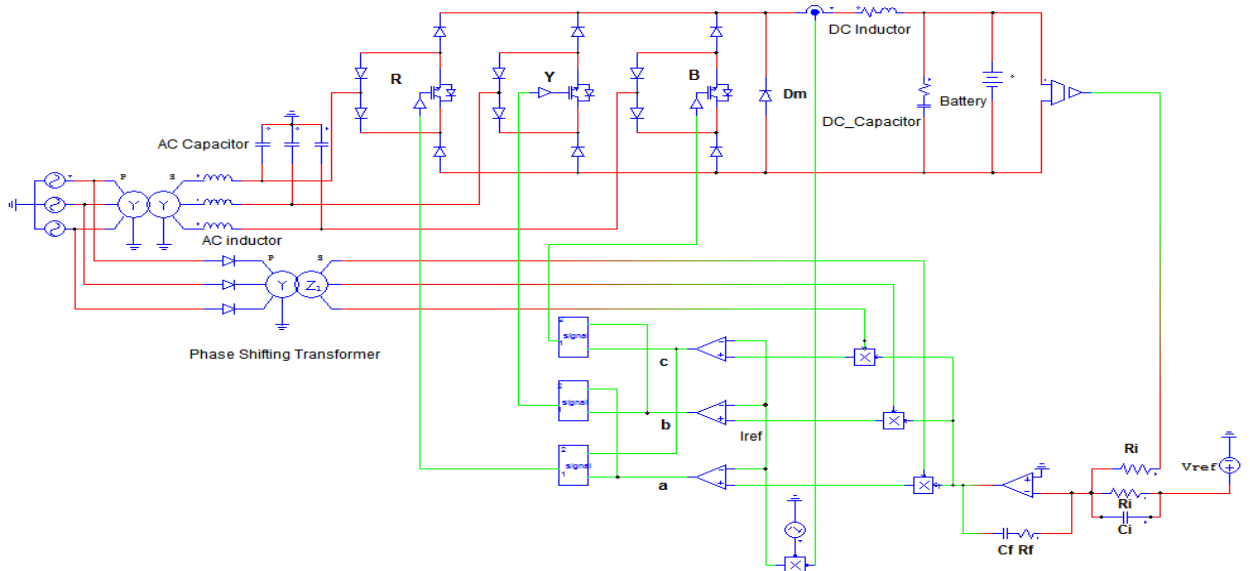


Figure 6. Rectifier circuit

4) PID Controller

An op-amp based PID controller is used here. Transfer function of the PID controller is $\frac{0.00024s^2 + s + 1000}{s}$

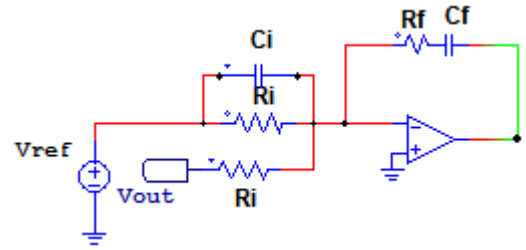


Figure 5. PID controller

5) Phase Shifting Transformer

phase shifting is done by a phase shifting transformer. The relationship between the angle and the winding turns are:

$$\frac{Ns2}{Ns1 + Ns2} = \frac{\sin(30 - \theta)}{\sin(30 + \theta)}$$

$$\frac{Np}{Ns1 + Ns2} = \frac{VAB}{Vab} \times \frac{1}{2 \times \sin(30 + \theta)}$$

where $0 \leq \theta \leq 30^\circ$

Here ,

Np (Primary winding) : 1 (Number of turns of the primary winding).

$Ns1$ (Secondary winding): 0.8966 (Number of turns of the 1st secondary winding)

$Ns2$ (Secondary winding): 0.5176 (Number of turns of the 2nd secondary winding)

V_{AB} : Primary line voltage

V_{ab} : Secondary line voltage

6) Outputs:

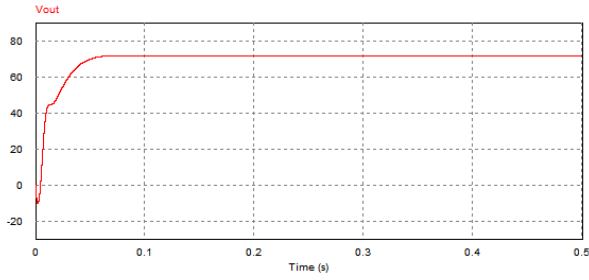


Figure 7. Rectifier output voltage

Fundamental component of the phase a input current is 7.99A.

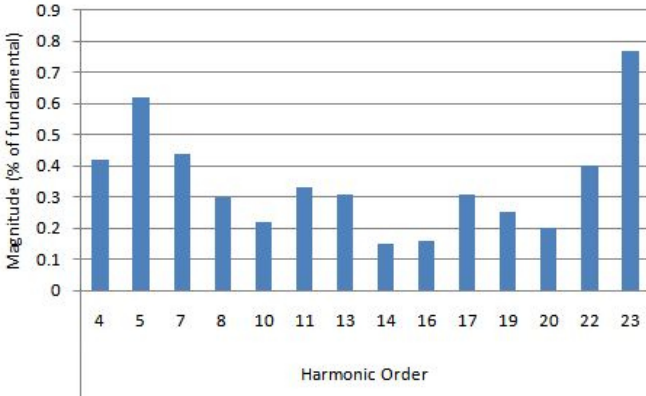


Figure 8. Harmonics distribution in phase a input current with 9.77% THD

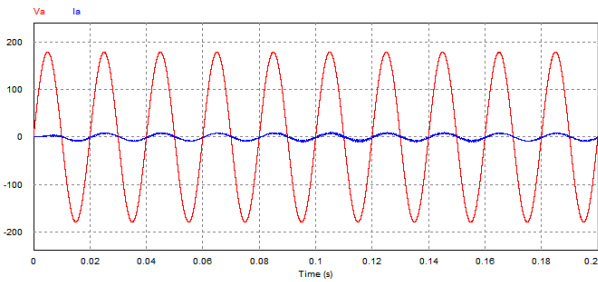


Figure 9. Input voltage and current wave shape where PF =0.967

The rectifier provides constant 72V when its PF is 0.967 and THD is 9.77%.

B. Charger feed by PV module:

To charge the batteries using solar power we have set three PV modules at the roof of the car. Module's open circuit voltage is 38.4V, but the terminal voltage of the battery is 72 V. So we have used a boost converter in between PV modules and batteries.

TABLE I. PV MODULE DATA SHEET: IRRADIATION 1000W/M²[5]

Maximum power	260 W
Open circuit voltage	38.4 V
Maximum power point voltage	31.4 V
Short circuit current	8.94 A
Maximum power point current	8.37 A
Module efficiency	15.51 %
Dimension/ Weight	(1675×1001×33)mm / 18Kg

1) Boost regulator:

PV module can supply up to 25 A current to the battery. To handle this current in a single line of boost converter, sophisticated isolation system is required and yet an enormous power loss will be associated with it. To lessen this problem, a multi channel interleaved boost converter is proposed. Fig. 11 shows a basic interleaved boost converter of 4 identical levels. To keep the output constant voltage controller (VC) compares reference and output voltage and mitigate the fluctuations which will then compared with the inductor current at current controller (CC). output of CC will then be compared with a 125K Hz triangular wave and generate switching signal. To facilitate interleaved operation the gate signals to the power switching devices are successively phase shifted by T/N where T is the switching period and N is the number of channels. Thus, the current delivered by the electric source is shared equally between each basic step-up converter level and has a ripple content of period T/N . As phase angles of the control signals are shifted by $360^\circ/N$, the frequency of the load current is N times higher than the switching frequency F . schematic diagram of the controller is shown below.

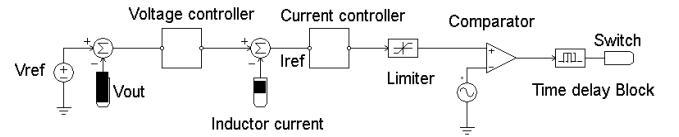


Figure 10. Block diagram of double loop Control

2) Voltage and current controller:

Here voltage controller (VC) is a type II and current controller (CC) is a type III compensator. Transfer functions of the VC and CC are $\frac{295.08S+4.1819593M}{S^2+105.467K}$ and $\frac{S^2+29.976K S+224.64M}{S^3+14.565S^2+479.44K S}$ respectively.

3) Minimum inductance and capacitance:

The minimum filter capacitor and the inductor value of each channel of the N-channel dc-dc converter are given by the following expression:

$$L = \frac{V_{out}}{4 \times F \times N \times \Delta I_{Max}} = 68 \mu H$$

$$C = \frac{I_{Max}}{4 \times F \times N \times \Delta V_{Max}} = 17.24 \mu F$$

where

N =the number of channels

ΔI_{max} = the input current ripple

F =the switching frequency

I_{max} = the maximum input current

ΔV_{max} = the maximum output voltage ripple

While vehicle is running the irradiation captured by the PV module will fluctuate and so do the module voltage. To demonstrate this, an extra 5V 5000 KHz disturbance has been added with PV module.

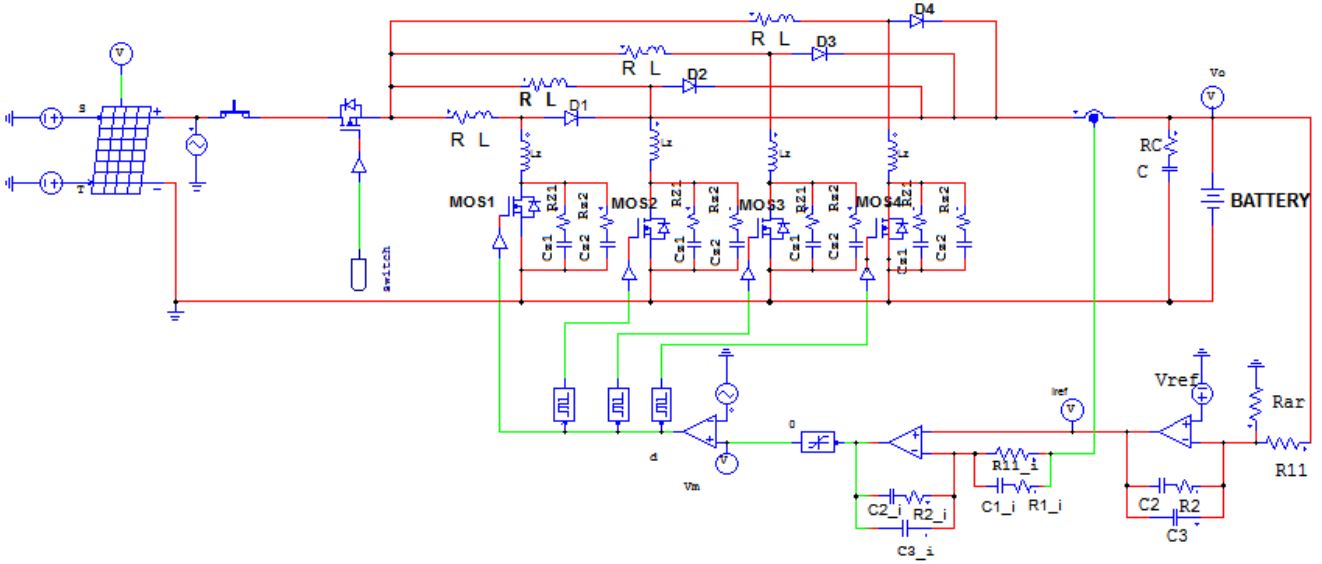


Figure 11. Charger for PV module

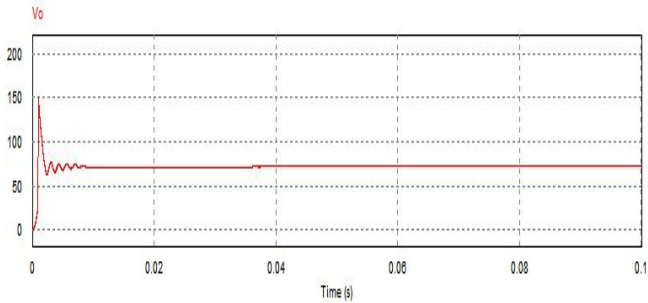


Figure 12. Output voltage with 5V 5000KHz disturbance

III. BATTERY

A. Charge-Discharge Controller:

These are two identical circuits that protect the batteries from over charging, under charging and from short circuit current. The charge controller will automatically disconnect batteries from the charger if terminal voltage is 74V or above and reconnect when it is 70V or below. In some cases, especially when solar irradiation is not adequate or at night battery voltage continues to fall, the discharge controller will disconnect the motor from the battery when terminal voltage is below 65V and connect the charger. It will reconnect the motor when voltage is 68V or above. During low voltage running especially during starting because of the flow of excess current battery terminal voltage goes below 65V but it continues for a moment. If this low voltage situation continuous more than 12 seconds only then battery will be disconnected and thus unnecessary shut down is prevented. On the contrary, maximum continuous charge or discharge current of the battery is 120A. If charging or discharging current is 120A or more and it continues for more than 3seconds only then battery will be disconnected. More ever if terminal voltage is in between 70 to 74V but current is more than 120A battery will be disconnected and automatically

reconnect when current is below 120A. Finally it disconnects only the faulty portion (Charge/ Motor) other one is kept connected

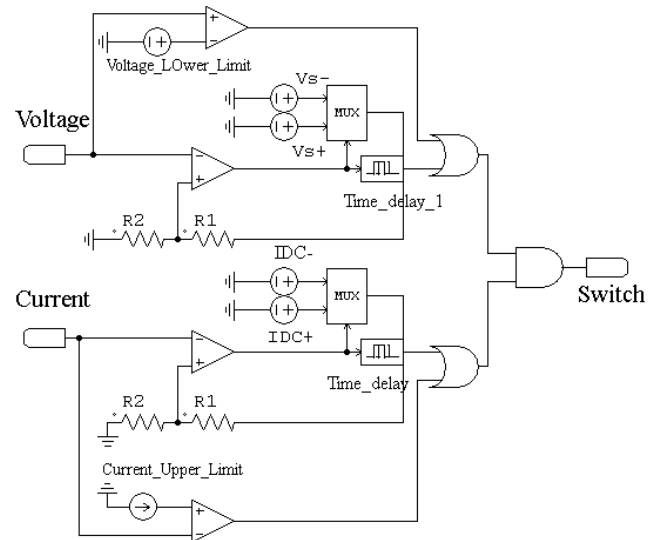


Figure 13. Discharge controller

B. Battery:

To store energy six series connected 12.8V 150AH Deep Cycle Lithium Ion (LiFePO4) Battery is used. Details of the battery are stated below [8]

TABLE II. BATTERY SPECIFICATIONS(SB150)[8]

Specifications	
Max Continuous Discharge Current	120A
Cold Cranking Amps	1500A
Efficiency	99.1%
Recharging cycles	Over 5000
Dimesion and Weight	(19.25"×10.25"×8.75") and 65 Lbs.

IV. LOSSES

The losses occurred in a power converter are usually generated by the semiconductor switches (MOSFET, Diode) and passive elements (inductor, capacitor). During practical implementation, component temperature is quiet higher than ambient temperature. During loss calculation we have considered the temperature effect and the reference temperature is taken as 40⁰ C.

TABLE III. MOSFET SPECIFICATIONS

ID Max(A)	75
Maximum operating temperature	150
RDS on(Ω)	0.0026
VGS (V)	4
tr(ns)	150
tf(ns)	93
Rth(J-C)(.C/W)	0.45
Rth(C-S) (.C/W)	.5

TABLE IV. DIODE SPECIFICATIONS

I Max(A)	50
Maximum operating temperature	150
V max(V)	600
Rth(J-C)(.C/W)	0.6
Rth(C-S) (.C/W)	.4

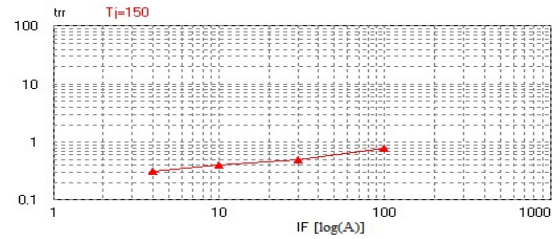


Figure 14. Diode I Vs Trr curve

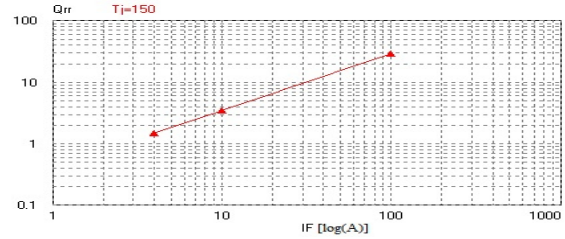


Figure 15. Diode I Vs Qrr curve

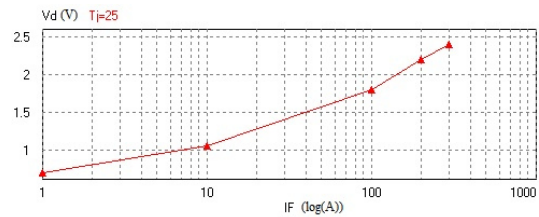


Figure 16. Diode I Vs Vd curve

A. PV feed Boost converter loss:

TABLE V. SWITCHING LOSSES:

			MOS 1	MOS 2	MOS 3	MOS 4	DIODE1	DIODE2	DIODE3	DIODE4
TRANSIENT LOSS	Without ZVS	CONDUCTION(W)	101.1	100.53	101.27	101.56	16.34	16.24	16.19	16.06
		SWITCHING(W)	38.89	39.59	39.57	39.49	11.07	11.91	11.82	11.68
	With ZVS	CONDUCTION(W)	48.526 x10-6	48.472 x10-6	48.489 x10-6	48.455 x10-6	37.439	37.439	37.438	37.438
		SWITCHING(mW)	237	237.03	236.80	236.85	2.169	5.08	2.66	1.45
STEADY STATE	Without ZVS	CONDUCTION(W)	.233	.222	.215	.205	7.22	7.01	6.87	6.67
		SWITCHING(W)	12.41	12.12	11.93	11.66	9.05	8.84	8.69	8.49
	With ZVS	CONDUCTION(W)	53.705 x10-6	53.675 x10-6	53.707 x10-6	53.695 x10-6	6.32	6.32	6.32	6.32
		SWITCHING(W)	266.79 x10-3	266.79 x10-3	266.79 x10-3	266.79 x10-3	6.54 x10-9	6.53 x10-9	6.53 x10-9	6.53 x10-9

TABLE VI. TOTAL LOSSES OF PV FEED BOOST CONVERTER

		Input Power	MOSFET loss	Diode loss	Inductor loss	Capacitor loss	Control switch loss	Output power	efficiency
TRANSIENT	Without ZVS(W)	1726.19	562	114.31	39.5	15.55	20.8	974.08	56.43
	With ZVS(W)	1797.46	.9837	161.115	39.5	8.0014	2.3026	1585.55	88.2
STEADY STATE	Without ZVS(W)	1823.54	48.99	62.84	11.89	4.82	6.74	1688.26	92.58
	With ZVS(W)	1955.13	1.067	25.28	11.89	.0786	1.9775	1914.84	97.94

From table V it can be seen that because of ZVS MOSFET switching loss falls down from around 39 W to 237mW and 12 W to 266mW and diode switching loss falls down from 11.5 W to 2.5mW and 8.5W to 6.5nW during transient and steady state period respectively. During steady state period

V. MOTOR:

A. Theory of Operation

The commutation sequence is designed in such a way that two windings are energized with positive and negative power, keeping the third one non-energized. Torque generated by the interaction of stator and rotor magnetic field, becomes maximum when their angular position is 90 apart to each other. To keep the motor running, magnetic field generated by the windings change according to the six step communication process.

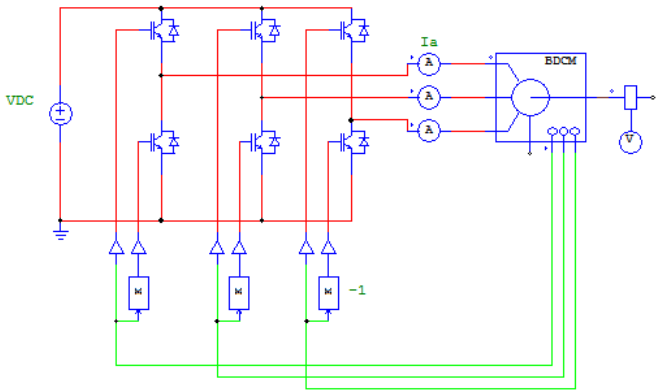


Figure 17. Hall sensor driven BLDC motor

TABLE VII. BRUSHLESS DC MOTOR DATASHEET[7]

Electrical specifications	
Resistance(Phase to Phase)	0.363
Inductance(Phase to Phase)	0.049m
No load current (A)	327m
Pole number	4
Rated current(A)	91.3
Rated voltage(V)	72
Rated Power(KW)	6
Mechanical Specifications	
Speed constant	27.78
Torque constant	.657
Rated Speed(rpm)	2000
No load speed(rpm)	2500
Rated Torque(N.m)	28.65
Max Torqu(N.m)	60
Net Weight(Kg)	48



Figure 18. BLDC motor connected with rear wheel [9]

ZVS impact on conduction loss is negligible. From Table VI it can be incurred that ZVS improves the converter’s efficiency during both transient period and steady state period.

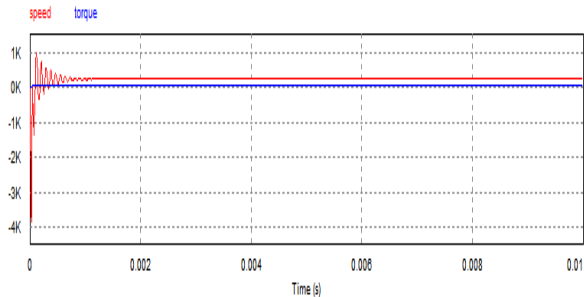


Figure 19. Speed during maximum torque

The figure above shows the torque speed relation when the torque is maximum i.e. during starting.

VI. CONCLUSION

Double loop controller of the Boost converter minimizes the effect of input disturbances to the output. ZVS reduces its switching losses and improves efficiency from 56.43% to 88.2% and 92.58% to 97.94% during transient and steady state period respectively. The proposed controller for the 2.16KW Vienna rectifier keeps its PF and THD into the IEEE standard. It takes around 5.33hours to charge the battery fully. PV modules are placed in the rooftop in such a manner that two are placed side by side and one at the bottom touching each other. Total weight of the car will be around 350 Kg including every part.

REFERENCES

- [1] A. Emadi, S. S. Williamson, and A. Kaligh "Power electronics intensive solutions for advanced electric, hybrid electric, and fuel cell vehicular power systems, " IEEE Trans. Power Electron., vol. 21, pp. 567-577, May 2006
- [2] O. Garcia, J. A. Cobos, R. Prieto, P. Alou, and J. Uceda, "Power factor correction: A survey," IEEEproceedings, PESC, 2001, pp. 8–13
- [3] "Understanding power factor", Application note; SGS-Thomson Microelectronics, 1996
- [4] C. Chavez, J. Houdek, "Dynamic Harmonic Mitigation and Power Factor Correction", 9th International Conference Electrical Power Quality and Utilisation, Barcelona, 9-11 October 2007
- [5] *SW 260 datasheet*, sunmoduleplus, Sunfields, Germany
- [6] Muhammad H. Rashid, *The Power Electronics: Circuits, Devices & Applications*
- [7] *BL6-1, Product Catalogue NO.189168* "China Electric motor Association/ GuangDongM&CElectric Power Co.Ltd
- [8] battery specifications, online: <http://www.lithiumion-batteries.com/products/12v-150ah-lithium-ion-battery/>
- [9] blde motor for electric car. Online Available: <https://www.uumotor.com/10kw-blde-motor-for-electric-car.htm>
- [10] Thomas M. Blooming, Daniel J. Carnovale 'application of ieeec std 519-1992 harmonic limits'

Android Based Health Care Service System for Aged Diabetic Patients

Farzana Alam Khan, Mohammad Ibrahim Khan

Department of Computer Science & Engineering
Chittagong University of Engineering & Technology (CUET)
Chittagong, Bangladesh.

Email: farzana.jyoti.ipsc@gmail.com, muhammad_ikhancuet@yahoo.com

Abstract— *Diabetes mellitus is an illness that requires continuing medical care and self-management system for the patients to prevent acute complications to reduce the risk of long-term complications. It is a complex disease in which the care for health is supposed to be a satisfactory daily prescribed diet and practicing regular habits, in addition to proper medication and treatment. Good diet and exercise habits, as well as continuous monitoring of blood pressure and sugar are all important for a diabetic patient in order to maintain his proper health. Although there are too many health care solutions exist in the market today, but none of them focus on providing a personalized health care management system for the patient. This proposed mobile application will take an effort to ensure a healthy balanced diet for the patients with the synchronization of balancing food energy. The system provides more facilities by giving the chance to set the medicine alarms, monitoring blood pressure and sugar levels in a chart. The system also provides more safety by making emergency calls and messages in real-time by utilizing GPS reporting technology considering the crucial moment of the patient. Two surveys are included in order to show the necessity of the implementation.*

Keywords—*Diabetes; Self-Management; Healthy diet; Balancing Food Energy; Medications; Monitoring; Emergency; GPS*

I. INTRODUCTION

Diabetes is a major, complex metabolic disease for which the key is proper diet plan and self-management system. Diabetes care is complex and requires many issues to consider. Although the internet is a burgeoning source of information and resources, the average patients often lacks the skills for finding and using the most optimal health care information and sometimes they abstain themselves from taking treatment. Patient self-management does not solely rely on the patients themselves but is also largely dependent on the availability of health care support encouraging patients to manage their health [1].

Diabetes is an arduous disease in which the body is unable to properly use and store glucose (a form of sugar). Glucose causes one's blood glucose as known as blood sugar to rise too high and it affects the patient's physical condition. Having high blood pressure is one of the several risk factors that can increase the chance of developing heart disease, kidney disease, a stroke and some other complications. Nowadays, people lack the ability to manage their health and willingness to change their lifestyles to better suit their health conditions

[2]. Numerous resources are available to help patients increase their role in the management of diabetes while simultaneously improving their metabolic parameters [3].

This proposed mobile application will take an effort to make the tasks of Diabetic patients easier. The aim of this application is to ensure a healthy balanced diet for the patients with the synchronization of balancing food energy. This application will give the user a prescribed diet plan which is made based on the information from the Diabetic Patients' Book. The user can add food item to his list every time he takes food or replace any food items. The amount of food with the calculated food energy (Calorie, Carbohydrate, Protein, Fat) is given along with each food item and if any of them crosses the calculated marginal level of energy, user will get a notification that he has taken more or less from that specific food energy. This application has a feature of medicine alarm. Through this feature, user can set the time, name and unit of medicine and gets an alarm at the given time as a reminder. This application will also provide a feature of monitoring blood Pressure and blood sugar level. Through this feature, user can make an entry of diastolic and systolic pressure. Blood sugar level can also be monitored by entering sugar level before and after the meal along with time and date. This application contains a feature of making emergency calls and messages. In any crucial moment, calls will be made and Emergency Communication Service will allow us to send messages in real-time by utilizing GPS reporting technology.

II. RELATED WORK

A large number of research guidelines are being conducted based on mobile technology to improve health care system for people. In the management of Diabetes Mellitus disease, Canagliflozin medicine is used sometimes previously. Canagliflozin is basically a drug which is used for the treatment of T2D patients [4]. Weight reduction was another significance of this research and it reduced by 4.4kg. But this reduction was successful only when Canagliflozin was added to a regimen of other anti-hyperglycemic agents. The volume of the food intake that patients take can be used to control and maintain the blood sugar. Though insulin therapy provides the most improved glycemic control and reduction of complications, but hypoglycemia and weight gain of the patients may occur [5]. But costs for the treatment of diabetic complications (40%-60% of total cost), for drug therapy for glycemic control represents 18% of the total cost. The patients

manage their vital statistics information, medical prescriptions including reminder alarms, transmission of medical information to health care providers including photographs and application settings. It was introduced basically to keep track of general fitness or provide information on medical issues [1]. But it was not specialized with a specific type of patients. The smartphone application enables the patients to tailor the application to suit their personal needs and preferences and also provides exercises information as well as the activation of the exercise to be monitored [6]. It lacks of providing a specialized treatment to the users based on their physical condition. This research study was further extended by incorporating a medical compliance system to the existing smartphone application [7].

III. SYSTEM DESCRIPTION

The system architecture of proposed health care system is illustrated in fig 1. Here, the process of profile creation, planning for proper diet plan, setting alarm for medicine, monitoring blood pressure in systolic and diastolic measure and blood sugar before and after taking meal, emergency call and message services are combined in the system architecture. Factors for evaluating the system are given below:

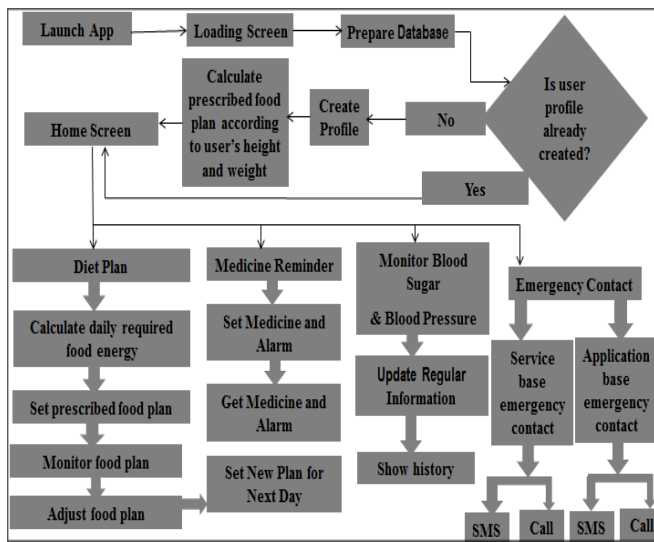


Fig. 1. System Architecture

A. Calculating Standard Height and Weight

We take the user's height, weight and date of birth at the time of profile creation. Based on these three criteria, we calculate user's standard food energy that he should take and build up a diet chart on the base of that measured calorie. At first we calculate the user's ideal weight that was given in cm.

$$\text{Height (inch)} = \text{height (cm)} * 0.39370$$

Then we will compute the person's weight from his height.

If the user has height greater than 60, Extra Height = Height (inch) – 60

If the user is male, Weight = 52 + (1.9 * Extra Height)

If the user is female, Weight = 52 + (1.7 * Extra Height)

If the user has less height than 60, Less Height = 60 – Height (inch)

If the user is male, Weight = 52 – (0.9072 * Less Height)

If the user is female, Weight = 52 – (0.9072 * Less Height)

These weights are the ideal weight that the user should have. The user can compare this calculated weight to his actual weight in order to know if he is overweight or underweight.

B. Calculating Standard BMR Value

We have to calculate the BMR (Basal Metabolic Rate) value now. BMR is the amount of energy needed while resting in a temperate environment during post-absorptive state. Basal metabolism is usually the largest component of a person's total calorie need. Daily calorie need is the BMR value multiplied by a factor with a value between 1.2 to 1.9 and BMR value is calculated basically by these two equations.

$$\text{BMR (male)} = 10 * \text{weight (kg)} + 6.25 * \text{height (cm)} - 5 * \text{age} + 5$$

$$\text{BMR (female)} = 10 * \text{weight (kg)} + 6.25 * \text{height (cm)} - 5 * \text{age} - 161$$

Now the BMR Calorie is: BMR Calorie = BMR * 1.2 [15]

Here, 1.2 is the activity level.

Now we will find out the actual calorie the person should take based on the criteria if he or she is overweight or underweight. At first we will find out integer calorie value.

$$\text{Calorie value (integer)} = \text{BMR Calorie} / 100$$

In case of Calorie value an even number, if the user has perfect weight that is if his actual and standard weight is same, Calorie = Calorie value (integer) * 100

If the user has underweight,

$$\text{Calorie} = (\text{Calorie value (integer)} + 2) * 100$$

If the user has overweight,

$$\text{Calorie} = \text{Calorie value (integer)} * 100$$

In case of Calorie value an odd number, if the user has perfect weight that is if his actual and standard weights are same, Calorie = (Calorie value (integer) + 1) * 100

If the user has underweight,

$$\text{Calorie} = (\text{Calorie value (integer)} + 1) * 100$$

If the user has overweight,

$$\text{Calorie} = (\text{Calorie value (integer)} - 1) * 100$$

The Calorie value ranges from 1000 to 2800 for all ages of people and it is proved by Estimated Energy Requirements (EER). The values fluctuate with a difference of 200 between each of them. The prescribed Calorie value is measured in such a way that it will always be from 1000 to 2800.

C. Prescribed Diet Plan

Based on user's height, weight and age, we calculate standard Calorie value that the user should take. The value should be in between 1000 to 2800. If Calorie value is greater than 2800, we will fix it in 2800. The comparison is given here,

if (Calorie > 2800)

Calorie=2800

If it is less than 1000, we will fix it in 1000.

if (Calorie < 1000)

Calorie=1000

A food energy chart will be given along with the app, from which the info about the food energy will be retrieved.

IV. SYSTEM IMPLEMENTATION

This application is developed on android platform. The application will start with a splash screen and after two seconds, the loading screen will come in front of the user. The background process will start behind the loading screen and a database will generate. The splash screen image is given in Fig. 2.

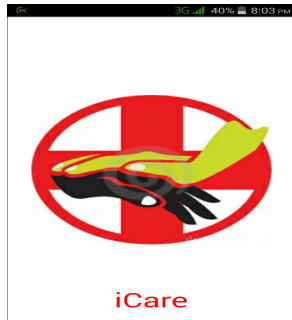


Fig. 2. Splash Screen

At first the profile needs to be created. From the info of the profile height, weight value and emergency numbers will be used. The profile creation is shown in Fig. 3.

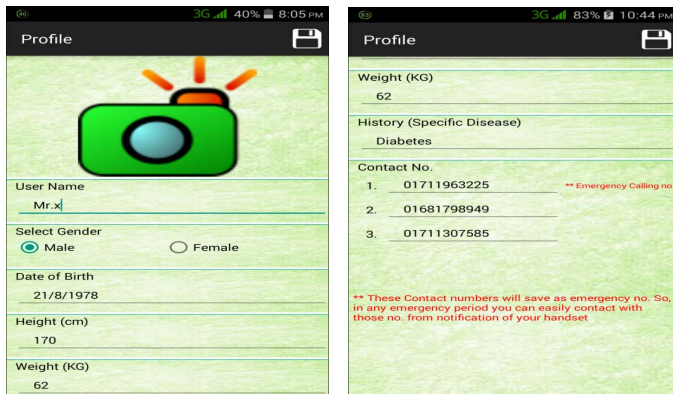


Fig. 3. Profile Creation

The user can use any of the four features which are shown in the homepage. The homepage of the application is given here in Fig. 4.

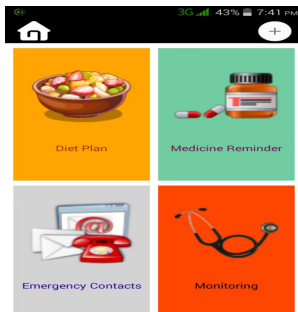


Fig. 4. Homepage of the Application

A. Balancing Food Items

When the user wishes to use diet plan feature, at first a prescribed diet plan will be viewed according to the health condition of the user. The items of prescribed diet plan are organized from a food list.

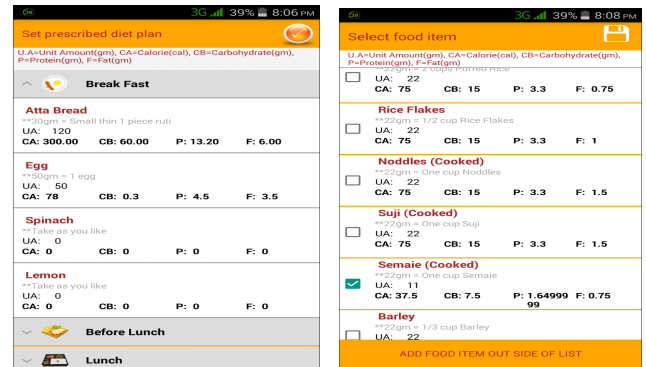


Fig. 5. (a) Prescribed Diet Plan (b) Main Food Item List

The user will be able to edit or replace the unit amount of food that was described, to add any of the items from the list.

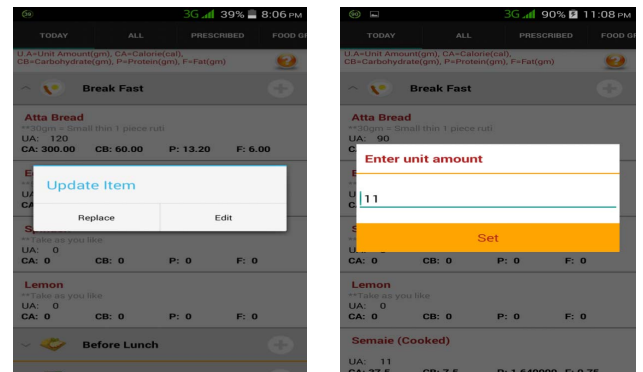


Fig. 6. (a) Edit or Replace Item (b) Entering Unit Amount

If the user wishes to add any food item out of list, he can do it by mentioning the exact food energies in specific fields. The new food item will be added to the list.

The system will give a notification to the user when he takes more or less food energy than the prescribed one.

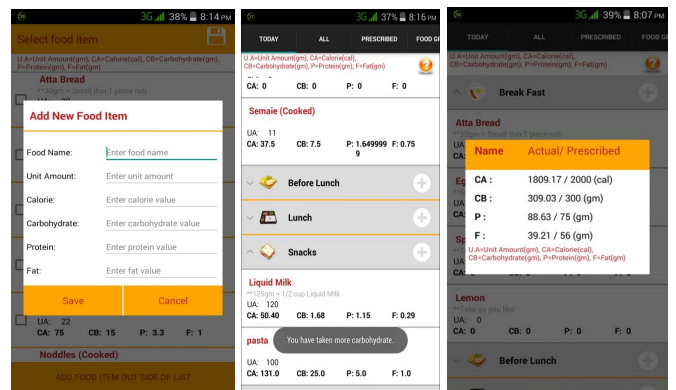


Fig. 7. (a) Adding Item out of List (Pasta) (b) The Item Added to Prescribed List (c) Actual/Prescribed Value

The previous records of diet plan are recorded in a list for the patient. All the food items are grouped together here. According to the book of Diabetes patients, these food items are organized. While replacing any item, the user can only see the food items under that list, not the whole list of foods. These groups are shown in the Fig. 8(b).

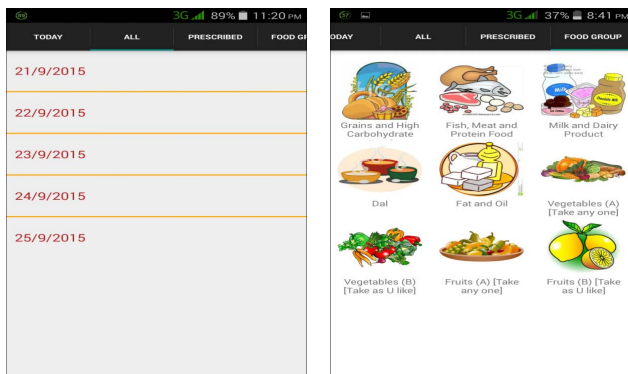


Fig. 8. (a) All Record of Food Plan (b) Basic Food Groups

A contradictory question may arise if the user has to put all the food related values as an input or not. All the food energy values according to a measured specific quantity will be given input by the system. The user only has to set his daily prescribed diet plan when a new day arrives. Then he can easily update the amount and the items those are categorized under some groups. The food items are given as input and they are grouped into here.

B. Medicine Reminder

The aged patients have tendency to forget taking their medicines timely. This app is going to help them by setting an alarm at exact time with corresponding medicine name. The patient can update the name and time of medicine anytime according to the suggestion of the physician. The medicine list will keep a track of the patient. The addition of medicine names is shown in Fig. 9(a) and 9(b).

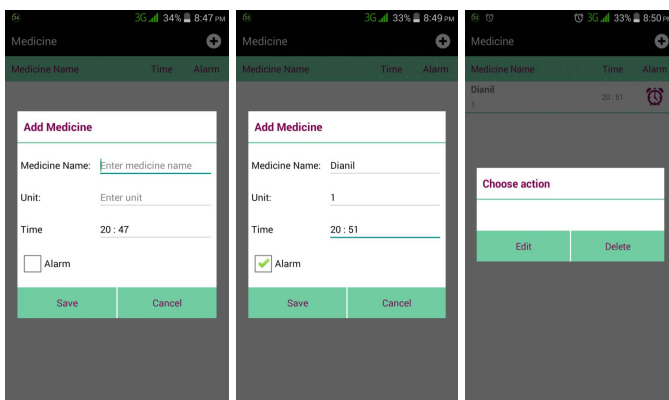


Fig. 9. (a) Indication for Adding Medicine (b) Adding Medicine with Time and Alarm (c) Updating Medicine Information

C. Monitoring Blood Pressure and Blood Sugar

This application is going to help the patients by keeping a track of their blood pressure and sugar level. It will help them to control the pressure level and sugar level. Like blood

pressure level, the patient can also keep blood sugar level in a track list. The Blood sugar level view is slightly same as Blood pressure view. But Blood sugar will be enrolled before and after taking the meal of the patient.

After taking meal the Blood sugar value becomes high. Fig. 10 and Fig. 11 have shown the addition of blood pressure level in systolic and diastolic measure and blood sugar level before and after taking the meal. The tracking list will help the patient while consulting the doctor.

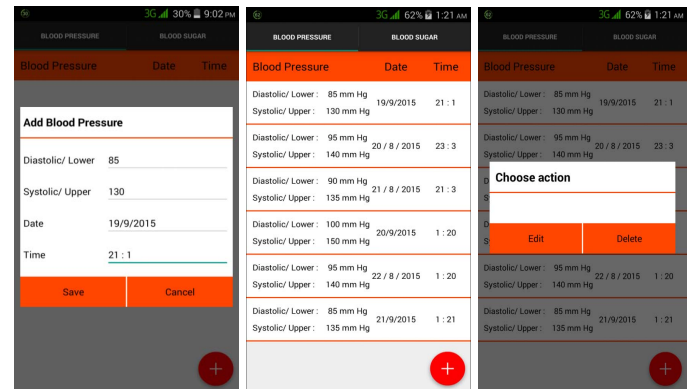


Fig. 10. (a) Adding Blood Pressure (b) Keeping Track of Blood Pressure Level (c) Updating Information

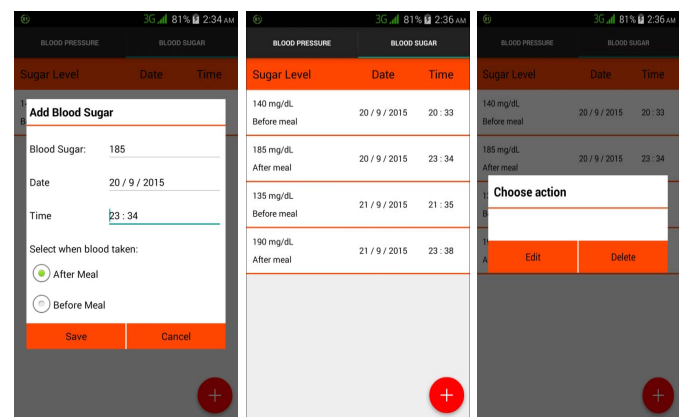


Fig. 11. (a) Adding Blood Pressure (b) Keeping Track of Blood Pressure Level (c) Updating Information

D. Emergency Call and Message Service

The patient can use emergency call feature to call any specific number and use emergency message service with his current location through a URL using GPS system. The user can choose any of the three numbers to make a call or message by pressing call or message button that is beside the number. It is shown in Fig. 12(a).

In Fig. 12(b), we can see the user can make call to the first number by pressing phone button and send messages to three numbers at a time by pressing message button. After pressing message button, the user will get a notification of message sending. So the user has to select whether he has to make a call or message and if he wants to send it to a particular one or all of them.

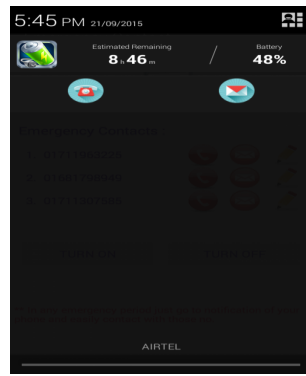
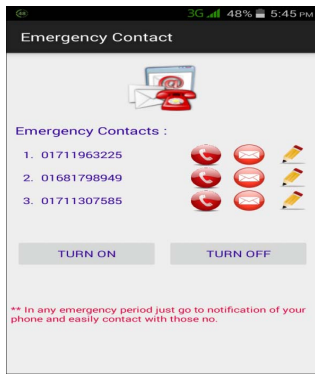


Fig. 12. (a) Emergency Contact (b) View after Pressing Turn on

The message is shown with a URL and after pressing the link the location of the patient can be identified using GPS. In Fig. 13(b) the text message is shown.

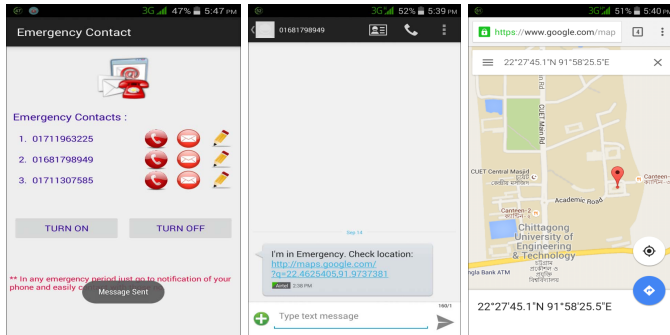


Fig. 13. (a) Delivery Report of Message Sent (b) Text Message (c) Location Identification.

V. EXPERIMENTAL RESULTS & EVALUATION

In order to evaluate the aptitude of the application we have done two real life experiments. We have talked to several Diabetic patients in Diabetic Hospitals to know their conditions and requirements before implementing the system. After implementing the system, we have taken another survey in order to find the effectiveness of our application.

A. Before Implementation

Before implementing the system, we made questionnaires to patients. We tried to find out the conditions, sufferings and requirements of patients. The survey questionnaires along with the responded numbers of participants for each question are given in Table 1. A graphical representation of questionnaires before implementing system is also shown in Fig. 14.

TABLE I. SURVEY BEFORE IMPLEMENTING THE SYSTEM

Questions	Yes	No	Not Needed
1. Can you store your prescribed diet plan with the help of any application?	7	35	8
2. Can you make a diet plan based on height, weight and age?	12	33	5
3. Is it possible to make different diet plans everyday?	10	30	10
4. Do you have any option to adjust your diet plan?	4	40	6
5. Do you know the name of alternative food item if you replace any of the items from your diet chart?	8	36	6
6. Do you know how much energy a food contains?	14	30	6
7. Can you keep the history of your diet plans?	8	28	16
8. Do you have multiple contact options in case of emergency?	8	40	2
9. Can you get your medicine reminder by any other service?	11	35	4
10. Can you track the list of medicine names, blood pressure and blood sugar values?	4	40	6

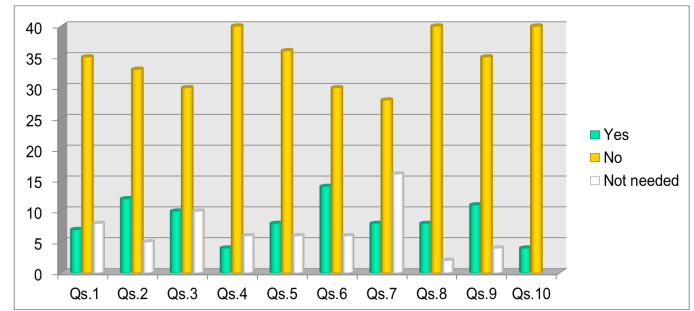


Fig. 14. Graphical Representation of Survey 1 before Implementation

If we make an analysis of the figure we can see that most of the answers are negatively given here. The patients were not getting the proper services from any other application.

B. After Implementation

After implementing the system, we made another survey with the same questions with a little modification. The list of the questions along with the responded numbers of participants for each question is accorded in Table 2. Based on the responded numbers of participants for each question, a graphical representation of the questionnaires after implementing the system is shown in the Fig. 15.

TABLE II. SURVEY AFTER IMPLEMENTING THE SYSTEM

Questions	Yes	No	Not Needed
1. Can you store your prescribed diet plan with the help of this application?	42	3	5
2. Have you got a perfect plan based on height, weight and age?	36	5	9
3. Can you make different plans everyday with this application?	38	6	6
4. Did you get any option to adjust your diet plan?	43	2	5
5. Did you get the alternative food item name if you replace any of the items from your diet chart?	38	4	8
6. Did you get idea about how much energy a food contains?	34	10	6
7. Can you keep the history of your diets with this application?	32	8	10
8. Did you use multiple contact options in case of emergency?	30	15	5
9. Did you get medicine reminder option useful to you?	42	6	2
10. Can you track the list of medicine names, blood pressure and blood sugar values with this application?	40	6	4

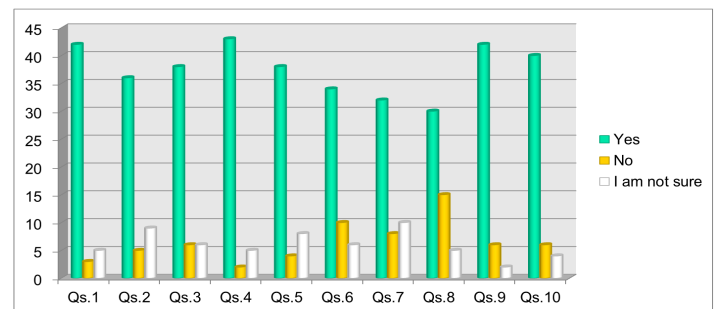


Fig. 15. Graphical Representation of Survey 2 after Implementation

In Fig. 15, we can see that most of the answers were given by patients positively. This positive feedback implies that the application is suitable and purely appropriate for the Diabetic patients as their desired requirements are fulfilling through this system.

The survey we took before and after the implementation of the system is given below here in Fig. 16(a) and Fig. 16(b). The

survey question samples along with demography are given here.

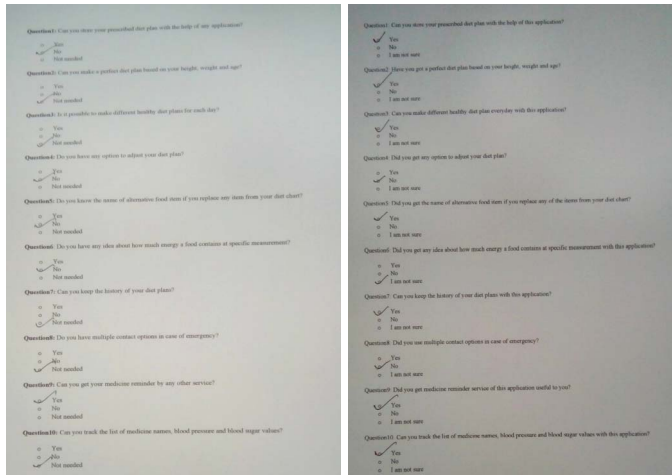


Fig. 16. (a) Survey taken before implementation (b) Survey taken after implementation

VI. CONCLUSION

In this paper, we have proposed a health care system through android application for aged Diabetic patients. It is crystal clear from our system that the chronic disease Diabetes can be reduced if the patients take some steps of self-management according to our proposed system. We have shown various features such as diet plan, medicine reminder, monitoring blood pressure and blood sugar and emergency contacts in our proposed android application. Besides, we have taken the opinions from the Diabetic patients of different Diabetic Hospitals to judge the effectiveness of our system. This system will cost very low amount of money and will be the best usable and utilizable system though the user have to use smart phone and internet connection. The limitation of our system is that it is only focused on personal user rather than multiple users. Besides Diabetes, some other arduous diseases can also be included with the system with proper information and guidance. For further improvement and for the welfare of the patients, these features can be combined together. However, the proper implementation of the system will play an important role to prevent the affliction of Diabetic patients by enabling them to self-manage their health.

REFERENCES

- [1] A. Carpio, J. Kim, R. Hoda, "Med Touch: Towards the Development of Smartphone-based Software Solutions for Mobile Health Care", SEPTA, Electrical and Computer Engineering, The University of Auckland, Auckland, New Zealand.
- [2] N. Tarara, "Designing mobile patient-centric self-help terminals for people with diabetes", In Proc. of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services, ACM, 2009.
- [3] <http://www.diabetes.org/diabetes-basics/diabetes-statistics>, American Diabetes Association: Diabetes Statistics [article online], Last day of access 22 August, 2015.
- [4] A. Nardolollo, M. P. Kane, R. S. Busch, J. Watsky, R. A. Hamilton, "A Clinical Perspective of Canagliflozin in the Management of Type 2 Diabetes Mellitus", Albany College of Pharmacy and Health Sciences, Albany, NY, The Endocrine Group, LLP, Albany, NY, Albany College of Pharmacy and Health Sciences-Vermont Campus, Colchester, VT.
- [5] A. Liebl, K. Khunti, D. O. Beltran and J. F. Yale, "Health Economic Evaluation of Type 2 Diabetes Mellitus: A Clinical Practice Focused Review", University of Leicester, UK, University of Miguel Hernandez, Spain, McGill University, Canada.
- [6] V. Gay, P. Leijdekkers, E. Barin, "A Mobile Rehabilitation Application for the Remote Monitoring of Cardiac Patients after a Heart Attack or a Coronary Bypass Surgery", Proc. of the 2nd International Pervasive Technologies Related to Assistive Environments, ACM, 2009.
- [7] I. Qudah, P. Leijdekkers, V. Gay, "Using Mobile Phones to Improve Medication Compliance and Awareness for Cardiac Patients", Proc. of the 3rd International Conference on Pervasive Technologies Related to Assistive Environments, ACM, 2012.
- [8] S. Akter, M. M. Rahman, S. K. Abe & P. Sultana, "Prevalence of diabetes and pre-diabetes and their risk factors among Bangladeshi adults: a nationwide survey", Department of Epidemiology and Prevention, Clinical Research Center, National Center for Global Health and Medicine, Toyama 1-21-1, Shinjuku-ku, Tokyo 162-8655, Japan, Department of Global Health Policy, University of Tokyo, Japan, Department of Statistics, University of Rajshahi, Rajshahi, Bangladesh. Published online:10 January 2014.
- [9] G. Bakris, J.R Sowers, "ASH Position Paper: Treatment of Hypertension in Patients with Diabetes – an Update", Department of Medicine, University of Chicago Pritzker School of Medicine, Chicago, and the Diabetes and Cardiovascular Center, Department of Medicine and Physiology, University of Missouri-Columbia School of Medicine and VA Center, Columbia.
- [10] F. Zhou, "Mobile Personal Health care System for Patients with Diabetes", Iowa State University, Ames, Iowa, United States.
- [11] IDF Diabetes Atlas, International Diabetes federation 2013. <https://www.idf.org/diabetesatlas>. Last day of access 10 September, 2015.
- [12] Qudah, P. Leijdekkers, and V. Gay, "Using mobile phones to improve medication compliance and awareness for cardiac patients", Proc. of the 3rd International Conference on Pervasive Technologies Related to Assistive Environments, ACM, 2012.
- [13] Y. WenBai, D. Yang, "Mobile Blood-glucose Monitoring of an Integrated Health Information management System", IEEE International Conference on Consumer Electronics (ICCE), pp. 1569489253.
- [14] L. Don, "Importance of cardiovascular disease risk management in patients with type 2 diabetes mellitus", Diabetes MetabSyndr Obes.2014 May 23;7:169-83.doi:10.2147/DMSO.S61438. e Collection 2014.
- [15] <http://www.calculator.net>, BMR Calculator, Last day of access 18 September, 2015.
- [16] "Diabetic Patients Book", given by Chittagong Diabetic Association.
- [17] <https://play.google.com/store/apps/details?id=com.skyhealth.glucosebuddyfree&hl=en>, Sky Health: Glucose Buddy, Accessed 24 August, 2015.
- [18] K. Rao, M. Ahmed Hanash, G.Ahmed AL-Aidaros, "Development of Mobile Phone Medical Application Software for Clinical Diagnosis", International Journal of Innovative Science and Modern Engineering (IJISME), ISSN: 2319-6386, September 2014.

Security Analysis of Approaches to Integrate Middleboxes into Software Defined Networks

Tobias Eggert

Technische Universität Darmstadt, Germany

Email: tobias.eggert@stud.tu-darmstadt.de

Rahamatullah Khondoker

Fraunhofer SIT, Darmstadt, Germany

Email: rahamatullah.khondoker@sit.fraunhofer.de

Abstract—Software-defined Networking (SDN) is a novel approach to manage enterprise and data center networks easily. Integration of middleboxes, which provides Network Functions (NF)s that are crucial for network security, performance and reliability, raises new challenges, for example, traversing middleboxes in a given order makes routing more complex. Rerouted traffic flows require that the state of middleboxes that is no longer part of the route is transferred to middleboxes which becomes part of the route. Software-defined Middlebox PoLicy Enforcement (SIMPLE) and OpenNF are two approaches to integrate middleboxes in SDNs which address these challenges. Since they are responsible to enforce middlebox policies, possible design flaws in their architecture could lead to severe vulnerabilities and put security of the network at stake. Therefore, security analysis of SIMPLE and OpenNF was conducted using Microsoft’s threat modeling approach called STRIDE, whose results show the threats on these approaches.

I. INTRODUCTION

Networks often grow over a long period of time by adding additional, at the moment of their deployment state-of-the-art network components on demand, which lead to increasing heterogeneity. Furthermore, unforeseeable events may force the implementation of ad hoc solutions that “just work”. If those ad hoc solutions are not adapted according to the networks overall topology, then these will introduce more complexity afterwards and let the networks degenerate. But, even homogeneous networks that do not degenerate over time may be hard to administer, simply due to their size. SDN is an approach to counter those and other challenges of network administration by splitting network management into two distinct planes: control and data. The *control plane* defines the routes to be taken by the traffic that flows through the network and adapts the routes dynamically to changing traffic conditions. It deploys the routes by configuring the *data plane*, which is responsible to forward the traffic.

However, beyond pure routing, SDN has to deal with middleboxes that provide a variety of crucial NFs such as Network Address Translation [1] (NAT), Intrusion Detection System [2] (IDS), (caching) proxies, firewalls [3], and load balancers (e.g. HAproxy¹). Therefore, SDN needs to enforce policies, which basically define sequences of middleboxes that have to be traversed by certain types of traffic. Unfortunately, it turns out that implementing middlebox policies in SDN is challenging. Two challenges are demonstrated on the basis

¹HAProxy project’s website: <http://haproxy.org>

Table I
THREAT MATRIX FOR ELEMENT TYPES AND THEIR CORRESPONDING DATA FLOW DIAGRAM (DFD) NOTATION. A X DENOTES THAT AN ELEMENT TYPE IS SUSCEPTIBLE TO A THREAT CLASS.

Threat Class \ Element Type	Data flow	Data store	Process	Inter-actor
Spoofing			X	X
Tampering	X	X	X	
Repudiation			X	X
Information Disclosure	X	X	X	
Denial of Service (DoS)	X	X	X	
Elevation of Privileges			X	
DFD Notation	—DF→	DS	PROC	IA

of an example network topology and corresponding policies which are described in Sec. III. The challenges which are described in Sec. IV-A and Sec. V-A pose the questions: how to route traffic flows through sequences of middleboxes, and how the NF state can be transferred when traffic flows are rerouted. The proposed approaches for those challenges are described in Sec. IV-B and Sec. V-B.

One of the goals of middleboxes (e.g. firewalls [3] or IDSs) is to increase network security. Hence, it would be contradictory to implement an approach to enforce middlebox policy enforcement which introduces new threats. Therefore, security analysis of both approaches was conducted in Sec. IV-C and Sec. V-C using the STRIDE threat modeling approach. Finally, the results are discussed in short in Sec. VII. An overview of STRIDE is given in Sec. II.

II. THE STRIDE THREAT MODELING APPROACH

STRIDE [4], a structured approach to uncover security relevant design flaws of a system, is well suited to be used in the field of SDN [5], [6]. Initially, the examined system is decomposed into its relevant elements and represented as a DFD. Elements can be data flows, data stores, processes, and interactors. Table I contains the DFD representation for each element. Beside those elements, DFDs may contain trust boundaries, denoted by dotted lines. The graphical representation of the system as a DFD helps one to make sure no relevant elements remain unconsidered. Depending on its type

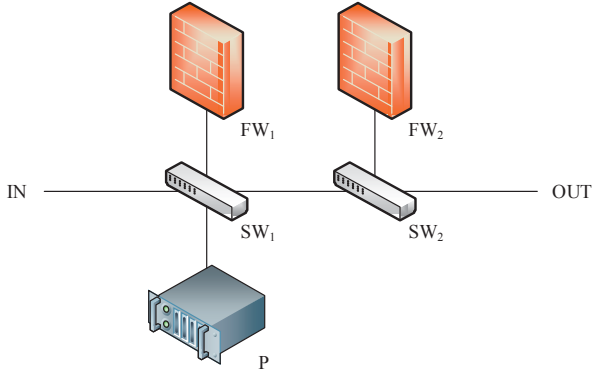


Figure 1. Exemplary network topology, where SW_1 and SW_2 are switches, FW_1 and FW_2 are firewalls and P is a proxy. IN and OUT are networks which are not part of SDN.

Table II
EXEMPLARY POLICIES WITH COMPLIANT ROUTES.

Policy	Compliant Routes
$P_1 = (\text{HTTP}, \text{IN}, \text{OUT}, (\text{FW}, \text{Proxy}))$	$R_{1,1} = (\text{IN}, SW_1, FW_1, SW_1, P, SW_1, SW_2, \text{OUT})$ $R_{1,2} = (\text{IN}, SW_1, SW_2, FW_2, SW_2, SW_1, P, SW_1, SW_2, \text{OUT})$
$P_2 = (\text{HTTPS}, \text{IN}, \text{OUT}, (\text{FW}))$	$R_{2,1} = (\text{IN}, SW_1, FW_1, SW_1, SW_2, \text{OUT})$ $R_{2,2} = (\text{IN}, SW_1, SW_2, FW_2, SW_2, \text{OUT})$

an element may be affected by threats of different classes. Six threat classes, which form the acronym STRIDE, are distinguished. Table I lists all threat classes and the element types which are endangered by them. Once the DFD was constructed, every element of the system is examined for its susceptibility to the threat classes corresponding to its type.

III. EXAMPLE SETUP

Each of the two challenges described in Sec. IV-A and Sec. V-A are demonstrated respectively by means of an example. Both examples are based on the network topology shown in Figure 1, consisting of two switches SW_1 and SW_2 , two firewalls FW_1 and FW_2 , and a proxy P . The examples use two policies which are listed in Table II together with all minimal routes that are compliant to them. A policy P_n is represented as a 4-tuple containing the traffic type², source and sink of matching flows, and a n-tuple holding the type of middleboxes the flows have to go through as the fourth element. Each flow is a sequence of multiple associated packets [7]. A route $R_{n,m}$ that complies to a policy P_n (i.e. flows following this route traverse middleboxes as required by the policy) is represented as a n-tuple, consisting of the source, the devices which pass traffic flows, and the sink.

²Switches may determine the traffic type of packets by Transport Layer (L4) properties (e.g., source and destination port).

Table III
EXCERPT OF SWITCHES' ROUTING TABLES IMPLEMENTING $R_{1,1}$.

SW_1			SW_2		
Traffic Type	Ingress Port	Egress Port	Traffic Type	Ingress Port	Egress Port
HTTP	IN	FW_1	HTTP	SW_1	OUT
HTTP	FW_1	P			
HTTP	P	SW_2			

Table IV
EXCERPT OF SWITCHES' AMBIGUOUS ROUTING TABLES IMPLEMENTING $R_{1,2}$.

SW_1			SW_2		
Traffic Type	Ingress Port	Egress Port	Traffic Type	Ingress Port	Egress Port
HTTP	IN	SW_2	HTTP	SW_1	FW_2
HTTP	SW_2	P	HTTP	FW_2	SW_1
HTTP	P	SW_2	HTTP	SW_1	OUT

IV. SIMPLE

SIMPLE [8] tags packets to keep track of their processing state. Switches use these tags once the decision, where to forward a packet, is taken depending on its processing state. SIMPLE treats middleboxes as black boxes and therefore does neither require changes to middleboxes behavior, nor in-depth knowledge of how middleboxes alter packets. Thus it is compatible with all kinds of middleboxes from different vendors.

A. Challenge

Considering $R_{1,1}$ which is compliant to P_1 , all HTTP traffic following this route has to be forwarded by SW_1 three times. When a HTTP packet arrives for the first time, the switch has to forward it to FW_1 , at the second time to P and finally to SW_2 . In this example, it is quite easy for the switch to decide where to forward the packet, since it arrives at a different port each time. Table III contains excerpts of the routing tables of SW_1 and SW_2 that implement $R_{1,1}$ based on this observation.

Now considering $R_{1,2}$ which is also compliant to P_1 , Table IV contains switches' routing table entries that were generated using the same method as for $R_{1,1}$. Note that given this routing table SW_2 is not able to decide where to forward HTTP traffic arriving at ingress port SW_1 . Actually, the routing table of SW_2 is ambiguous, because it contains two entries for HTTP traffic coming from SW_1 . Hence, it is not possible to implement $R_{1,2}$ by basing the decision where to forward a packet on its traffic type and ingress port only.

B. Solution

SIMPLE solves the challenge described in Sec. IV-A by adding *tags* to packets. A tag indicates the processing state of a packet (i.e. the middleboxes it already went through). Switches are responsible to update a packet's tags after it goes through a middlebox. Tags are implemented as Virtual Local Area Network (VLAN) tags, Multiprotocol Label Switching (MPLS) labels or by exploitation of unused Internet Protocol

Table V
EXCERPT OF SW₁ ROUTING TABLE IMPLEMENTING R_{1,2}.

Traffic Type	Tags	Ingress Port	New Tags	Egress Port
HTTP	{ }	IN	{ }	SW ₂
HTTP	{FW ₂ }	SW ₂	{FW ₂ }	P
HTTP	{FW ₂ }	P	{FW ₂ , P}	SW ₂

Table VI
EXCERPT OF SW₂ ROUTING TABLE IMPLEMENTING R_{1,2}.

Traffic Type	Tags	Ingress Port	New Tags	Egress Port
HTTP	{ }	SW ₁	{ }	FW ₂
HTTP	{ }	FW ₂	{FW ₂ }	SW ₁
HTTP	{FW ₂ , P}	SW ₁	{FW ₂ , P}	OUT

(IP) header fields. With tags, it is possible to implement R_{1,2} with the routing table entries for SW₁ and SW₂ listed in Table V and Table VI respectively.

However, some middleboxes do not just inspect packets, but alter them. Considering P doing NAT and thus rewriting packets' IP headers, tags stored inside an IP header might be destroyed by P. To solve this problem, the authors of SIMPLE added a Dynamics Handler (DH) implemented as a module of the controller. The last switch before a tag destroying middlebox will send a copy of the first N packets of every new flow to the controller. The first switch after the middlebox will do the same with the first N packets in a time window W. The DH maps the two groups of packets to each other by a similarity analysis based on rabin fingerprints. Based on this mapping, the DH installs flow rules on the switch immediately after the middlebox for tagging packets correctly.

C. Security Analysis

The proof-of-concept of SIMPLE controller was implemented by extending POX³, which in turn uses OpenFlow [7] protocol. Figure 2 shows the architecture of SIMPLE as a DFD. A trust boundary separates components which are part of the data path since the external traffic sources are beyond the control of SDN and can not be trusted (cf. [9]).

The remainder of this section discusses to which extend the particular components of SIMPLE are susceptible to threats. The results are summarized in Table VII.

1) *Data Flows*: Intracomponent data flows (4-8, 10) are considered safe as long as an attacker has no physical access to network components. Thus threats to these data flows can be mitigated by physical access control to the network hardware. However, the physical integrity of links between network components can not always be ensured due to structural conditions (e.g. consider a campus-wide network in which network components are located at different buildings).

Data flows (1-3) have to be protected against information disclosure and tampering, since an attacker would otherwise be able to get knowledge of the internal structure of SDN and

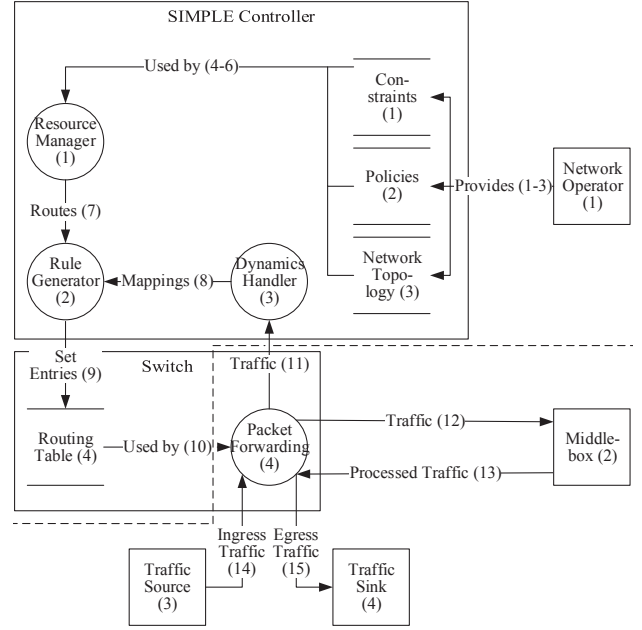


Figure 2. DFD of the SIMPLE approach.

to provide false data to the SIMPLE controller, respectively. Hence, these data flows should be protected by means of cryptography, e.g. through Transport Layer Security (TLS) [10]. An attacker could also cut the link used by data flows (1-3) (DoS). However, these data flows are only active, if the topology of the network has changed or new policies are established. Since this is expected to be rather seldom this threat can be ignored.

Data flow (9) needs to be protected, against information disclosure and tampering, too. Otherwise an attacker would be able to get information about established routes or can even establish routes that are not compliant to the policies, respectively. Because SIMPLE uses OpenFlow, this data flow is protected by TLS⁴ against the above mentioned threats. Since switches' routing tables need to be changed frequently (e.g. each time the dynamics handler generated new rules for a new flow) it has to be protected against DoS, too. This can be achieved through redundant links.

Data flows (11-13) need to be protected against tampering, since an attacker who is able to manipulate the flow tags of data packets, is able to spoof the processing state of packets and therefore interfere with the enforcement of the policies. In contrast to the former data flows the integrity can not be enforced by TLS, since tags are stored in the packets' IP headers and TLS operates on a higher Open Systems Interconnection (OSI) [11] layer. Instead, Internet Protocol Security (IPsec) which can also protect the header of IP packets against tampering [12] should be used. Some middleboxes like IDSs need traffic to be unencrypted to be able to process it. Hence, encrypting all traffic by default could even lower the overall

³The source code of POX is available under <https://github.com/noxrepo/pox>

⁴It is assumed that the network operator configured SIMPLE to use TLS.

security of SDN. If the traffic contains sensitive data, then the traffic source and sink will be responsible to protect it against eavesdropping. The implications of cutting traffic links highly depend on the network topology. These could lead to a DoS of the whole SDN in one case or could be easily mitigated by establishing new routes in another case.

The protection against tampering and information disclosure of data flows (14) and (15) has to be ensured by traffic source and traffic sink if needed. The protection against DoS of these data flows can be mitigated by redundant connections.

2) *Data Stores*: Data stores (1-3) have to be protected against tampering and information disclosure because otherwise the resource manager may generate routes that are not compliant to the policies the network operator wants to enforce and an attacker would be able to get knowledge of the internal structure of SDN. Therefore, integrity should be provided by digital signatures and confidentiality by not providing an interface to read the data from outside of the SIMPLE controller. Since the data is only read by the resource manager, which is also located inside of the SIMPLE controller, protection against DoS can be ignored, as long as physical integrity of the controller is ensured.

The routing tables of the switches (4) have to be protected against tampering and information disclosure (cf. data flow 9). These protections are expected to be implemented by the switches' manufacturers. However, the routing tables are susceptible to DoS attacks due to their limited size. As described in Sec. IV-B new routing table entries are generated for every new traffic flow. An attacker could simply establish new traffic flows until the space of the routing tables is exhausted⁵.

3) *Processes*: *Spoofing* and *tampering* processes can be mitigated through remote attestation which are provided by for example, Trusted Platform Modules (TPMs) [13]. However, not all switches' may offer such a feature for process (4).

Processes (1-3) cannot plausibly deny their actions, since there are no other processes to be accused instead, hence there is no need to mitigate *repudiation* for these processes. But a packet forwarding process (4) of a switch could set false tags and by faking its log accuse another switch. This can be detected through the logs of adjacent switches.

Processes (1) and (3) are not susceptible to *information disclosure* as they do not send data to components outside of the controller. Information sent out by processes (2) are already protected by the measures that have been taken for data flow (9). An attacker could *tamper* with a forwarding process (4) (see above) and make it forward traffic to him instead of the intended receiver. However, as already discussed above, the confidentiality of the traffic routed through SDN has to be ensured by the traffic source and sink.

If processes (1-3) fails (*DoS*), then this may have severe implications for the whole network. This could be mitigated by deploying a watchdog process which will restart a crashed process. Beyond this, an attacker could try to run a Distributed

⁵This type of attack against OpenFlow based SDN is described in detail in [9].

Table VII
SIMPLE COMPONENTS' THREAT MATRIX. M DENOTES THAT A THREAT CAN BE MITIGATED. I DENOTES THAT A THREAT CAN BE IGNORED. S DENOTES THAT A COMPONENT IS SUSCEPTIBLE TO A THREAT.

Component \ Threat Class		Spoofing	Tampering	Repudiation	Information Disclosure	Denial of Service	Elevation of Privileges
Data flow	1-3		M		M	I	
	4-6		M		M	M	
	7		M		M	M	
	8		M		M	M	
	9		M		M	M	
	10		M		M	M	
	11		M		I	M	
	12		M		I	M	
	13		M		I	M	
	14		I		I	M	
	15		I		I	M	
Data Store	1		M		M	I	
	2		M		M	I	
	3		M		M	I	
	4		M		M	S	
Process	1	M	M	I	I	M	M
	2	M	M	I	M	M	M
	3	M	M	I	I	M	M
	4	M	M	M	I	I	I
Interactor	1	M		M			
	2	M		M			
	3	I		I			
	4	I		I			

Denial of Service (DDoS) attack against processes (3), since parts of the ingress traffic get forwarded to this process which can be mitigated by carefully adjusting the values of N and W as described in Sec. IV-B. Protection against DoS of processes (4) is expected to be implemented by the switches' manufacturers.

Elevation of privileges of processes can be mitigated through isolation, e.g. by running them inside a Trusted Execution Environment (TEE) (As provided by TPM).

4) *Interactors*: Obviously it is important that the authenticity of a network operator (1) is guaranteed. Otherwise an attacker could impersonate an operator and tamper with the SIMPLE controller. The authenticity of network operators can be ensured through an access control system. Secure logging of actions of a network operator mitigates repudiation of his activities.

The spoofing of middleboxes is mitigated by their physical connection to the network. Their actions should also be logged to prevent repudiation. Traffic source (3) and sink (4) are not trusted and therefore no measures against spoofing or tampering have to be taken.

V. OPENNF

OpenNF [14] transfers and synchronizes the state between middleboxes to prevent from deducing wrong conclusions and interfering with rerouted traffic flows due to a lack of states.

A. Challenge

Considering $R_{1,1}$ and $R_{2,1}$ which implement policies P_1 and P_2 respectively⁶, both routes contain FW_1 . So if both routes are implemented concurrently, all HTTP and HTTPS traffic will go through FW_1 . This leads to a problem if HTTP and HTTPS traffic increase so much that the capacities of FW_1 will be exceeded. One solution to this problem is to reroute HTTPS traffic by replacing $R_{2,1}$ with $R_{2,2}$. HTTPS traffic will then go through FW_2 and thereby decrease the load on FW_1 . However, FW_2 may interfere with the rerouted flows. Considering the firewalls keeping track of Transmission Control Protocol (TCP) connections, if a flow containing an established FW_2 will not be aware of this connection because of not observing the handshake and therefore will drop all TCP packets related to this connection.

B. Solution

OpenNF solves the challenge by moving middleboxes' state that is related to the rerouted flows. Therefore, an Application Programming Interface (API) was implemented by each middlebox holding state which enables the controller to get and set state of middleboxes, to stop the middleboxes from updating their state, and to forward traffic flows to the controller. The manager is responsible to manage state transfer and update switches' routing tables. By doing so in an orchestrated manner, OpenNF is able to provide certain guarantees when rerouting flows. With OpenNF's *loss-free move* operation which ensures that all flow related state is moved and that all packets are processed eventually, it is possible to replace $R_{2,1}$ with $R_{2,2}$ without FW_2 interfering with rerouted flows. Figure 3 shows the operation as a Unified Modeling Language (UML) sequence diagram. The controller instructs FW_1 to stop updating state and forward every subsequent traffic to the controller⁷. Thereafter, the controller gets the state from FW_1 , sets it at FW_2 , and sends the buffered traffic to FW_2 . From this point onwards, the controller will not buffer subsequent traffic but rather directly forward it to FW_2 . Finally, the controller reroutes HTTPS traffic flows by replacing $R_{2,1}$ with $R_{2,2}$.

C. Security Analysis

The OpenNF controller was implemented based on OpenFlow. Figure 4 shows the architecture of OpenNF as a DFD. A *control application* implements the policies and enforces them by controlling OpenNF's *Manager*⁸. The control application may use additional information such as the current state of middleboxes or the network, which would imply additional data flows to the control application. However, the need for those data flows depends on the concrete implemented policies and therefore they are not considered in the OpenNF DFD. The Manager orchestrates the rerouting of flows (i.e. by configuring the switches) and the transfer of state between middleboxes. Each middlebox provides several NFs whose

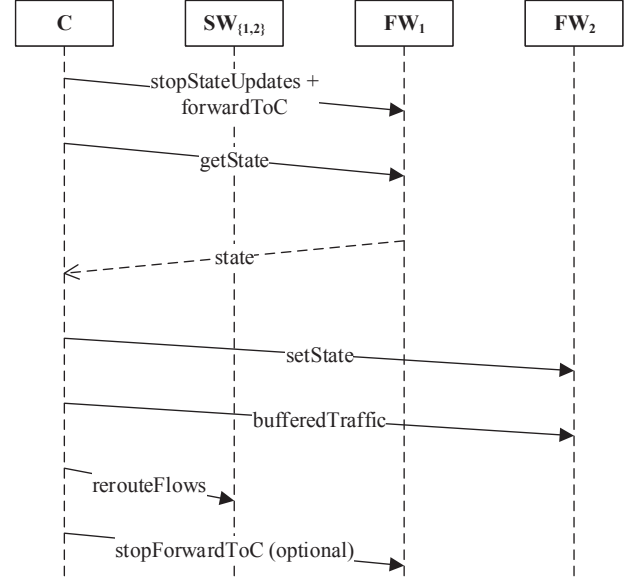


Figure 3. UML sequence diagram of the *loss-free move* operation transferring state from FW_1 to FW_2 .

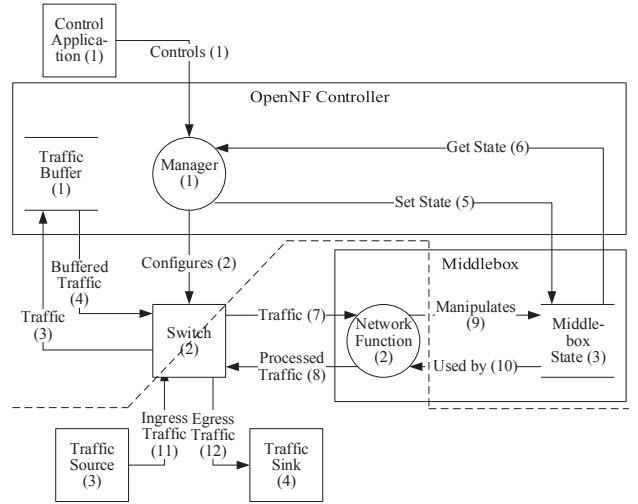


Figure 4. DFD of the OpenNF approach.

behavior are influenced by the middleboxes' state and which may manipulate the state in turn. The traffic routed to a middlebox gets processed by its NF and afterwards is sent back to a switch. Switches may also forward packets to the controller, which will buffer the traffic e.g., during rerouting of flows as described in Sec. V-B. Ingress and egress traffic of SDN are generated and consumed by external traffic sources and sinks respectively. Since the external traffic sources are beyond the control of SDN, the data created by them is not trustworthy. Hence, a trust boundary separates components which are part of the data path (cf. [9]).

The remainder of this section discusses to which extend the particular components of OpenNF are susceptible to threats.

⁶All rules except $R_{1,2}$ can be implemented without the need for tags.

⁷Only HTTPS traffic flows and related state are affected.

⁸Actually, the manager is split into two processes, the NF State Manager and the Flow Manager. Here, we consider them as one process.

The results are summarized in Table VIII.

1) *Data Flows*: Threats to intracomponent data flows (9, 10) can be mitigated through protection of network hardware against unauthorized physical access.

Data flow (1) needs to be protected against tampering, information disclosure and DoS since otherwise an attacker would be able to establish arbitrary policies, obtain knowledge about the networks state or prevent the enforcement of policies, respectively. However, since the control application can run together with the OpenNF controller on the same machine, data flow (1) can be treated like an intracomponent data flow, too.

Data flow (2) is very similar to data flow (9) of SIMPLE and since OpenNF also uses OpenFlow, it is protected by TLS as well. However, it is not active whenever a new traffic flow is routed through the network (cf. SIMPLE), but only when traffic flows are rerouted. Nonetheless its availability is crucial for SDN and therefore should be ensured by redundant links.

Data flows (3), (4), (7), (8), (11) and (12) transport traffic originated at an external traffic source and eventually routed to an external traffic sink. Therefore traffic sources and sink are responsible to protect the traffic against tampering and information disclosure, if desired. It is worthy to note that in contrast to traffic flows (11-13) of SIMPLE which also transport traffic, it is not necessary to protect the IP header from being tampered, since OpenNF does not use the IP header to store the processing state of a packet. However, the flows need to be protected against DoS, which can be achieved through redundant links.

Data flows (5) and (6) are active, whenever state needs to be transferred from one middlebox to another. Because the behavior of a middlebox is based on its state, tampering of theses data flows has to be prevented. Also information disclosure has to be prevented, since otherwise an attacker may obtain knowledge of the internal state of middleboxes like IDSs or firewalls. The attacker could exploit this information for subsequent attacks. In OpenNF the middleboxes are responsible to ensure integrity and confidentiality of state in transfer by making use of cryptography. DoS can again be prevented by redundant links.

2) *Data Stores*: Data store (1) does not need to be protected against tampering or information disclosure, since the buffered traffic will be protected by traffic's source and sink, if needed. However, the availability of the traffic buffer needs to be ensured, because otherwise OpenNF will not be able to reroute flows. Therefore, the buffer has to be large enough to store all traffic of a flow that occurs during rerouting of that flow.

It is assumed that protection of data store (3) against tampering, information disclosure and DoS was implemented by the middlebox manufacturers. Protection against these threats are needed for the same reasons as it was ascertained for data flows (5) and (6). However, as described in Sec. V-B the middleboxes need to be extended by an API which provides access to data stores (3). It is crucial that the implementation of this API does not interfere with the manufacturer's protection of the data store.

Table VIII
OPENNF COMPONENTS' THREAT MATRIX. M DENOTES THAT A THREAT CAN BE MITIGATED. I DENOTES THAT A THREAT CAN BE IGNORED. S DENOTES THAT A COMPONENT IS SUSCEPTIBLE TO A THREAT.

Threat Class \ Component		Spoofing	Tampering	Repudiation	Information Disclosure	Denial of Service	Elevation of Privileges
Data flow	1		M		M	M	
	2		M		M	M	
	3		I		I	M	
	4		I		I	M	
	5		M		M	M	
	6		M		M	M	
	7		I		I	M	
	8		I		I	M	
	9		M		M	M	
	10		M		M	M	
	11		I		I	M	
	12		I		I	M	
Data Store	1		M		M	M	
Process	1	M	M	I	M	M	M
	2	M	M	M	M	M	M
Inter-actor	1	M		M			
	2	I		I			
	3	I		I			

3) *Processes*: Spoofing of and tampering with processes can be mitigated through remote attestation (e.g. through a TPM) and elevation of privileges of processes as well as information disclosure can be mitigated by running them inside a TEE.

Protection against repudiation is not necessary for the process (1) because it is not able to plausibly deny its actions (cf. processes (1-3) of SIMPLE). For processes (2) secure logging has to be implemented to prevent repudiation.

A watchdog process, which restarts crashed processes should be implemented to protect all processes against DoS.

4) *Interactors*: Policies are implemented as a controller application (1). Thus to enforce the policies it is crucial that the application is safe against spoofing. This can be achieved through measurement of the application, e.g. by a TPM or, if the application runs on another machine than the controller, through remote attestation. Secure logging of all actions of the control application will prevent against repudiation.

The authenticity of switches (2) is ensured by their physical links and secure logging of their actions can prevent against repudiation.

Traffic source (3) and sink (4) are not trusted and therefore no measures against spoofing or tampering have to be taken.

VI. RELATED AND FUTURE WORK

Apart from the work mentioned in this paper, the STRIDE framework has also been used to accomplish security analyses of SDN SBI protocols including OpenFlow, OF-Config, and OVSDDB [15]; SDN architectures including PCE, 4D,

and SANE [16]; SDN controllers including OpenDayLight, ONOS, Rosemary, and Ryu [17]; SDN security applications including OpenFlow Random Host Mutation and Resonance [18]; SDN applications for monitoring and measurement including sFlow and BigTap [19], and SDN wireless and cellular applications such as OpenRadio and SoftRAN [20]. All of these work not only unveil potential security threats but also provide suggestions to tackle those threats using existing well-defined mechanisms. Future work will be to verify the threats using penetration testing with or without considering the suggestions. All of these work fall into the category of security-centric SDN, additionally, network security can be improved by using SDN such as OrchSec [21] and AutoSec [22] architectures.

VII. CONCLUSION

The security analysis of SIMPLE shows that the routing tables of its switches are susceptible to a DoS attack. Even an attacker with limited capabilities can exhaust the flow table by establishing new traffic flows as the Ternary Content-Addressable Memory (TCAM) of switches is very limited. This threat can not be mitigated by simple means rather requires a change in the design of SIMPLE. This vulnerability can be traced back to a similar vulnerability in OpenFlow (cf. [9]) which has not been addressed in SIMPLE.

OpenNF is not susceptible to a similar DoS attack like SIMPLE because the traffic forwarded to the controller does not trigger the creation of new rules but is just buffered. However, the buffer should be big enough otherwise packet loss will occur. Another threat of OpenNF is imposed by the necessary modification of the middleboxes. Because of these modifications, middleboxes cannot be treated as black boxes like SIMPLE. The modifications have to be implemented without tampering with the protections against all the threats that were implemented by the middleboxes' manufacturers which implies that in depth knowledge of the code is necessary. Bugs in the OpenNF application may also lead to vulnerabilities. Instead of implementing policies as an application, it would be better if OpenNF is capable to enforce policies defined by network operators.

REFERENCES

- [1] P. Srisuresh and M. Holdrege, "IP Network Address Translator (NAT) Terminology and Considerations," RFC 2663 (Informational), Internet Engineering Task Force, Aug. 1999. [Online]. Available: <http://www.ietf.org/rfc/rfc2663.txt>
- [2] K. A. Scarfone and P. M. Mell, *Guide to Intrusion Detection and Prevention Systems (IDPS)*. Gaithersburg, MD: National Institute of Standards and Technology, Feb. 2007. [Online]. Available: http://www.nist.gov/manuscript-publication-search.cfm?pub_id=50951
- [3] K. Ingham and S. Forrest, "A History and Survey of Network Firewalls," Feb. 2012. [Online]. Available: <https://www.cs.unm.edu/~treport/tr/02-12/firewall.pdf>
- [4] S. Hernan, S. Lambert, T. Ostwald, and A. Shostack, "Uncover Security Design Flaws Using The STRIDE Approach," <https://web.archive.org/web/20150208212605/https://msdn.microsoft.com/en-us/magazine/cc163519.aspx>, Nov. 2006, accessed: 2016-02-26.
- [5] M. Tasch, R. Khondoker, R. Marx, and K. Bayarou, "Security Analysis of Security Applications for Software Defined Networks," in *Proceedings of the AINTEC 2014 on Asian Internet Engineering Conference*, ser. AINTEC '14. New York, NY, USA: ACM, 2014, pp. 23:23–23:30. [Online]. Available: <http://doi.acm.org/10.1145/2684793.2684797>
- [6] D. Magin, R. Khondoker, and K. Bayarou, "Security analysis of OpenRadio and SoftRAN with STRIDE framework," ser. ICCCN '15. Piscataway, NJ, USA: IEEE, Aug. 2015. [Online]. Available: <http://publica.fraunhofer.de/dokumente/N-360563.html>
- [7] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner, "OpenFlow: Enabling Innovation in Campus Networks," *SIGCOMM Comput. Commun. Rev.*, vol. 38, no. 2, pp. 69–74, Mar. 2008. [Online]. Available: <http://doi.acm.org/10.1145/1355734.1355746>
- [8] Z. A. Qazi, C.-C. Tu, L. Chiang, R. Miao, V. Sekar, and M. Yu, "SIMPLE-flying Middlebox Policy Enforcement Using SDN," in *Proceedings of the ACM SIGCOMM 2013 Conference on SIGCOMM*, ser. SIGCOMM '13. New York, NY, USA: ACM, 2013, pp. 27–38. [Online]. Available: <http://doi.acm.org/10.1145/2486001.2486022>
- [9] R. Kloti, V. Kotronis, and P. Smith, "OpenFlow: A security analysis," in *Network Protocols (ICNP), 2013 21st IEEE International Conference on*. IEEE, 2013, pp. 1–6.
- [10] P. Dierks and E. Rescorla, "The Transport Layer Security (TLS) Protocol Version 1.2," RFC 5246 (Standards Track), Internet Engineering Task Force, Aug. 2008. [Online]. Available: <http://www.ietf.org/rfc/rfc5246.txt>
- [11] ISO/IEC JTC1, "Information Technology - Open System Interconnection - OSI Reference Model: Part 1 - Basic Reference Model," ISO/IEC 7498-1, International Organization for Standardization, Nov. 1994. [Online]. Available: [http://standards.iso.org/ittf/PubliclyAvailableStandards/s020269_ISO_IEC_7498-1_1994\(E\).zip](http://standards.iso.org/ittf/PubliclyAvailableStandards/s020269_ISO_IEC_7498-1_1994(E).zip)
- [12] P. Kent, "IP Authentication Header," RFC 4302 (Standards Track), Internet Engineering Task Force, Dec. 2005. [Online]. Available: <http://www.ietf.org/rfc/rfc4302.txt>
- [13] Trusted Computing Group, "TPM Main - Part 1 Design Principles," TCG Published, Trusted Computing Group, Inc., Mar. 2011. [Online]. Available: https://www.trustedcomputinggroup.org/files/static_page_files/72C26AB5-1A4B-B294-D002BC0B8C062FF6/TPM%20Main-Part%201%20Design%20Principles_v1.2_rev116_01032011.pdf
- [14] A. Gember-Jacobson, R. Viswanathan, C. Prakash, R. Grandl, J. Khalid, S. Das, and A. Akella, "OpenNF: Enabling Innovation in Network Function Control," in *Proceedings of the 2014 ACM Conference on SIGCOMM*, ser. SIGCOMM '14. New York, NY, USA: ACM, 2014, pp. 163–174. [Online]. Available: <http://doi.acm.org/10.1145/2619239.2626313>
- [15] M. Brandt, R. Khondoker, R. Marx, and K. Bayarou, "Security Analysis of Software Defined Networking Protocols OpenFlow, OF-Config and OVSDB," in *IEEE ICCE 2014, Special Session on SDN*, July 2014.
- [16] D. Klingel, R. Khondoker, R. Marx, and K. Bayarou, "Security Analysis of Software Defined Networking Architectures PCE, 4D and SANE," in *ACM AINTEC*, November 2014.
- [17] R. K. Arbetu, R. Khondoker, K. Bayarou, and F. Weber, "Security Analysis of OpenDayLight, ONOS, RoseMary, and RYU SDN Controllers," in *17th International Network Strategy and Planning Symposium (Networks 2016)*, September 2016.
- [18] M. Tasch, R. Khondoker, R. Marx, and K. Bayarou, "Security Analysis of Security Applications for Software Defined Networks," in *ACM AINTEC*, November 2014.
- [19] P. Dauer, R. Khondoker, R. Marx, and K. Bayarou, "Security Analysis of Software Defined Networking Applications for Monitoring and Measurement sFlow and BigTap," in *The 10th International Conference on Future Internet Technologies (CFI)*, June 2015.
- [20] D. Magin, R. Khondoker, and K. Bayarou, "Security Analysis of OpenRadio and SoftRAN Using STRIDE Framework," in *The 24th International Conference on Computer Communications and Applications (ICCCN 2015)*, August 2015.
- [21] A. Zaalouk, R. Khondoker, R. Marx, and K. Bayarou, "OrchSec: An Orchestrator-Based Architecture For Enhancing Network-Security Using Network Monitoring And SDN Control Functions," in *IEEE NOMS 2014*, May 2014.
- [22] R. Khondoker, P. Larbig, D. Senf, K. Bayarou, and N. Gruschka, "AutoSecSDNDemo: Demonstration of Automated End-to-End Security in Software-Defined Networks," in *IEEE NetSoft 2016*, June 2016.

Performance Comparison of MFCC Based Bangla ASR System in Presence and Absence of Third Differential Coefficients

Sudipto Debnath, Fatema-E-Jannat, Susmita Saha, Mohammad Tarik Aziz, Rifayet Hasan Sajol, and Md. Jakaria Rahimi

Department of Electrical and Electronic Engineering
Ahsanullah University of Science and Technology
Dhaka, Bangladesh

s.dnath91@gmail.com, f.jannat29@gmail.com, susmita.eee34@gmail.com, imran1496@gmail.com, rifayet.014@gmail.com, mjr Rahimi@gmail.com

Abstract—Present Mel Frequency Cepstral Coefficient (MFCC) based Bangla Automatic Speech Recognition (ASR) systems are mostly implemented with delta and acceleration coefficients. With delta and acceleration coefficients of MFCC and the log energy, a vector set of 39 dimensions is obtained per 10ms. In this paper, our objective is to observe the effect of third differential coefficients on the performance of Bangla ASR, which is not explored in this field yet. In doing so, we have appended 13 third differential coefficients along with previous 39 coefficients to make a vector set of 52 coefficients per 10ms frame. We have observed the performance of Bangla ASR system in the presence and absence of third differential coefficients using Hidden Markov Model (HMM) based tied-state triphone model. To make the speech corpus, 100 sentences have been uttered by a different number of speakers at different phases including both male and female of similar ages in between 22-24. Hidden-Markov-Model Toolkit (HTK) has been used here for the comparative analysis. We have considered the Sentence Correction Rate (SCR) as the performance indicator. From the experiments, it has been observed that the MFCC based system of 39 (MFCC39) and 52 (MFCC52) dimensions have average SCR of 98.89% and 98.94% respectively. Therefore, our finding is that slight improvement is possible with the inclusion of third differential coefficients when the sampling data rate is as high as 44.1 KHz.

Keywords—MFCC39; MFCC52; Bangla; third differential coefficients; HMM

I. INTRODUCTION

ASR is a remarkably swift emerging application of natural language technology that perceives the spoken speeches by computer or computerized devices. For developing and ameliorating the speech recognition system numerous experiments have been done in different languages. Despite the advancement of speech recognition technology around the world, the central of the attraction of developing ASR was always for the English language. There is an insignificant amount of research on Bangla ASR although it is the seventh

most spoken language in the world [1]. Therefore, we have taken an attempt for improving Bangla Speech Recognizer.

We have developed ASR for Bangla that employs Mel Frequency Cepstral Coefficients (MFCCs). Several techniques are used for feature extraction. Among them, we have chosen MFCC as MFCC computes the features on Mel-scale. Moreover, Mel-scale relates to human hearing as human ear responds to the varying intensity of sound in a logarithmic manner and Mel-scale is logarithmic above 1 KHz. We have observed the performance of both acceleration coefficients and third differential coefficients using HMM-based tied-state triphone model. HMM is a statistical approach that originates stochastic models from known utterances which are compared with the probability of that the unknown utterance developed by each model [2]. To achieve higher accuracy on HMM-based phonetic segmentation, an acoustic model based on tied-state triphone has been employed here as it is the most effective model for capturing the co-articulation effect. The immediate left and right phonetic contexts have been considered here. HTK toolkit has been used here with the help of MATLAB for extracting the feature, building the HMMs, and generating the results.

We have organized this paper into six sections including this introduction section. Section II presents a general review of some previous work in the field of ASR using MFCC with acceleration coefficients. In section III, the methodology is briefly described. Section IV provides the experimental setup. The experimental results are discussed in section V. Conclusion and an overview of future work are given in section VI.

II. LITERATURE REVIEW

Researchers have performed a lot of experiments in the field of ASR with different approaches to obtain acceptable accuracy of speech recognition.

Most of these approaches have used MFCC based system of 39 dimensions for extracting the feature from the formatted input signals. Therefore, we have motivated to observe the effect of MFCC based system of 52 dimensions as it has not used for extracting the feature from formatted input signals yet.

[3] worked on Bangla word recognition using the MFCC coefficients for extracting the feature vectors which were then fed to the triphone HMM-based classifiers to obtain the word recognition performance. They developed phonetic symbols for Bangla phonemes which were used to design the triphone model. They observed the Word Correct Rate (WCR) and Word Accuracy (WA) by designing MFCC38 & MFCC39 based triphone HMM with some mixture components and found MFCC39 based system performed better than the other systems.

[4] worked on Bangladeshi Dialect Recognition using MFCC based system of 39 dimensions and Gaussian Mixture Model, which was trained with the assist of iterative Expectation-Maximization (EM) algorithm. The system was tested on dialects of Barisal, Noakhali, Sylhet, Chittagong and Chapai Nawabganj regions of Bangladesh. They performed a comparison between Support Vector Machine (SVM) and Gaussian Mixture Model (GMM). Their results showed that GMM adaptation with MFCC of 39 dimensions gave an incentive accuracy rate up to 100%.

[5] applied MFCC with the addition of delta and acceleration coefficients along with one energy component and thereby made a vector set of 39 dimensions to observe the performance of isolated and continuous Bangla speech recognition for speaker dependent and independent models. For pattern classification, they used HMM-based classifier with the assist of the stochastic language model. An adaptive filter was used here for noise cancellation and zero-point detection algorithm was used here for determination of starting and ending of the speech signal. They got an average accuracy rate of 85%.

[6] presented an ASR system for Bangla alphabets. A small sized database was formed for this system. With the help of MFCC39 based system, the extracted features were saved as reference templates and then Dynamic Time Warping (DTW) algorithm was applied to compare the real time input coefficient to the saved ones. To enhance the accuracy, they employed K-Nearest Neighbors (K-NN) classifier to the output of the DTW. They got an accuracy of 90%.

III. METHODOLOGY

A. Speech Corpus

It is well known to us that to make an acoustic model for ASR, we need speech corpus. The main challenge to work with Bangla ASR is that we do not have any enrich Bangla speech corpus like British national corpus. Therefore, by uttering 100 sentences by a different number of speakers at different phases we have made our speech corpus. These 100 sentences were developed by [14]. We have used three different phases as shown in Table I. Each phase shows the number of speakers used to make the corpus for training purpose.

TABLE I. EXPERIMENTAL PHASES

Phases	No. of Speakers	Description
Phase- I	8	(8x 100) sentences have been uttered by only male speakers
Phase- II	8	(8x 100) sentences have been uttered by only female speakers
Phase- III	16	(16x 100) sentences have been uttered by an equal number of male and female speakers

On the other hand, these 100 sentences have been uttered by 10 different speakers (among them five are males and other five are females) that have been used as the test corpora. We have also used the training corpus as test corpora to evaluate the performance comparison between MFCC based system of 39 and 52 dimensions. We have recorded the speech samples on android based application named “Easy Voice Recorder” at a sampling frequency of 44.1 KHz. All the sample files have been recorded in Waveform Audio File (.wav, 16 bit PCM) format. We have recorded the speech samples in room environment in the presence of little household noises.

The architecture of an ASR system can be represented by two blocks showing in the Fig. 1 [7].

B. Feature Extraction

Computing Mel Frequency Cepstral Coefficients (MFCC) is the most prevailing and presiding method for extracting feature vector. In MFCC, from the audio signal the main contents that actually carries the information are identified and the background noises are discarded. The conventional implementation of computing MFCC is shown in the Fig. 2 [12].

1) *Pre-emphasis*: Pre-Emphasis is filtering that enhances the energy in the high frequencies. The speech signal $s(n)$ is passed through a high pass filter with a range of 0.97 to 1.0 which discards the DC components [8]. The pre-emphasized signal that we get is sharper than before and the most commonly used pre-emphasis filter can be represented by following equation [9]:

$$s_2(n) = s(n) - a * s(n-1), \quad (1)$$

where, $s_2(n)$ is the output signal and a is the pre-emphasis coefficient whose value is between 0.91 and 1.0.

2) *Framing*: For analyzing the speech signal over a short period of time, the signals are framed as the speech signal is a time varying signal. In our work, we have used the frame of 25ms as it cannot be too long or too small [4].



Fig. 1. Block representation of ASR system

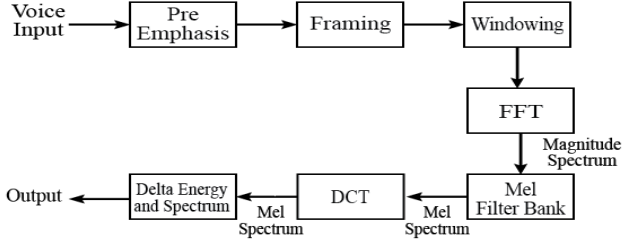


Fig. 2. Overall process of MFCC

3) *Hamming window*: Performing Fourier Transformation (FT) on framed signal arises discontinuity at the both ends which causes some undesired consequences [4]. To overcome this problem, each frame is multiplied with Hamming Window. If the signal is $s(n)$, after hamming windowing signal becomes $s(n) * w(n)$, where $w(n)$ is the hamming window and it can be represented by following formula [9]:

$$w(n, \alpha) = (1 - \alpha) - \alpha \cos(2\pi n/N - 1), \quad (2)$$

where, $0 \leq n \leq N - 1$.

4) *Fast Fourier transform*: Fast Fourier Transform (FFT) is done to transform the signal from time domain to frequency domain. The speech signal is split up into small frames and these small frames are then transformed [8].

5) *Mel-Scale filter bank*: The speech signal is not linear and human ear is less sensitive to higher frequencies. As FFT spectrum is wide ranged so Mel-Scaled Filter Bank is used to convert it to Mel-Scale. Mel-Scale is roughly linear below 1 KHz and logarithmic above 1 KHz [10]. The approximate Mel frequency can be expressed as [10]:

$$F_{\text{Mel}} = 2595 \log_{10} \left(1 + \frac{F_{\text{Hz}}}{700} \right), \quad (3)$$

where, F_{Mel} is the perceived frequency and F_{Hz} is the real frequency.

In our work, the filter bank has been constructed using 26 triangular bandpass filters. These filters are multiplied with the spectrum and then the coefficients are summed [13].

6) *Discrete cosine transform*: After converting the Mel spectrum into log Mel spectrum, Discrete Cosine Transform (DCT) is applied to convert that log Mel spectrum to the time domain. The output we get from this conversion is MFCCs. The following equation is used for computation [11],

$$c_i = \sqrt{\frac{2}{N}} \sum_{j=1}^N m_j \cos \left(\frac{\pi i}{N} (j - 0.5) \right), \quad (4)$$

where, m_j is the log filter bank amplitudes and N is the number of filter bank channels.

7) *Delta energy and delta spectrum*: Since MFCCs are static coefficients and speeches have the information in the

dynamics, to gain better performance it is needed to append time derivatives to the static parameters [4]. For computing delta coefficient, the following formula is used [11]:

$$d_t = \frac{\sum_{\theta=1}^{\Theta} \theta (c_{t+\theta} - c_{t-\theta})}{2 \sum_{\theta=1}^{\Theta} \theta^2}, \quad (5)$$

where, c_t is Mel Frequency Cepstral Feature and d_t is delta coefficient at time t .

The acceleration coefficients and third differential coefficients can be calculated in a similar way.

C. Modeling Based on HMM

HMM-based classifier works on two phases. They are (1) training phase and (2) recognition phase.

1) *Training phase*: For creating the phone based HMM models for each phoneme, global mean and variance are estimated in this training phase from the extracted feature vectors. A flat initialization of HMM parameters is required to define the HMM model for each phoneme. We have got these phonemes from [14]. Then, applying Baum-Welch algorithm, the re-estimation of HMM parameters is done for each phoneme [15]. Each of these phoneme models represents five states left to right HMM model. In our work, we have used context dependent triphone HMM model [11]. For creating the triphone HMM model, monophone transcriptions are transformed into triphone transcriptions. Then the monophones are copied to construct a set of triphone models and re-estimated again [11]. Later, for establishing strong estimation of all state distributions, similar acoustic states of these triphones are joined together.

2) *Recognition phase*: The system can recognize a word that is defined in the dictionary [5]. At first, preprocessing and feature extraction are done for recognizing an input word [5]. By feature extraction, an observance sequence is formed and then the calculation is done for the observance sequence probability and the maximum one is selected [5]. The Viterbi algorithm is used here for this probability computation [5].

IV. EXPERIMENTAL SETUP

Current ASR systems use delta, acceleration coefficients, and log energies which are appended with static MFCCs to obtain feature vector set of 39 dimensions (12-MFCC, 12- Δ MFCC, 12- $\Delta\Delta$ MFCC, E, Δ E, and $\Delta\Delta$ E, where E stands for log energy [12]) in per 10ms. To observe the performance of third differential coefficients we have appended 12 third differential coefficients (12- $\Delta\Delta\Delta$ MFCC) along with the log energy ($\Delta\Delta\Delta$ E), in total 13 coefficients with MFCC based system of 39 dimensions (MFCC39) to obtain MFCC based system of 52 dimensions (MFCC52). The configuration files for both MFCC39 and MFCC52 based systems are shown in Table II. We have built triphone HMMs with five states (among these, three are emitting states) and left to right model with the help of train corpus. Sentence Correction Rate (SCR) for both test corpora and train corpus has been evaluated using an HMM-based classifier.

TABLE II. CODING PARAMETERS OF MFCC39 AND MFCC52 BASED SYSTEM FOR HTK

Coding Parameters	Configuration	
	MFCC39	MFCC52
TARGETKIND	MFCC_0_D_A	MFCC_0_D_A_T
TARGETRATE	100000.0	100000.0
SAVECOMPRESSED	T	T
SAVEWITHCRC	T	T
WINDOWSIZE	250000.0	250000.0
USEHAMMING	T	T
PREEMCOEF	0.97	0.97
NUMCHANS	26	26
CEPLIFTER	22	22
NUMCEP	12	12
ENORMALISE	T	T

To obtain the Sentence Correction Rate (SCR) we have designed the following systems-

- MFCC39+triphone HMM
- MFCC52+triphone HMM

V. EXPERIMENTAL RESULT AND DISCUSSION

HMM-based triphone models have been generated (according to the phases mentioned in Table I) for each of the two systems which have been mentioned in the previous section. Then from the test corpora which is comprised of 10 different people's speech sample of 100 sentences, everyone's speech samples have been taken individually and given to the input of the HMM-based classifier after performing feature extraction. Then the SCR comparisons of each person's speech samples have been demonstrated in Fig. 4, Fig. 5, Fig. 7, Fig. 8, Fig. 11, Fig. 12 for each of the systems. Besides, SCR has been also evaluated for the speech samples of people which have been used to form three different phases. When triphone model for Phase-I has been generated, the SCR of the speech samples of people under this phase has also been evaluated and shown in Fig. 3. Same case for the phase-II, and its SCR comparison has been demonstrated in Fig. 6.

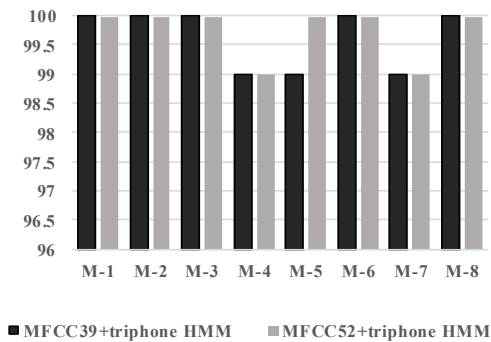


Fig. 3. Comparison of SCR between MFCC39 and MFCC52 for train corpus under Phase-I

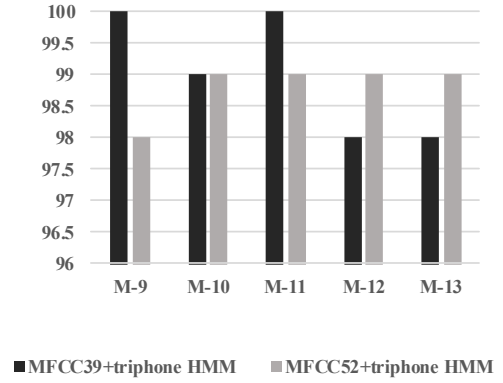


Fig. 4. Comparison of SCR between MFCC39 and MFCC52 for test corpora under Phase-I

The triphone model of phase-III has been generated accumulating 16 people's speech samples. Then SCR has been evaluated for each of these people's speech samples and their SCR variations with two different systems has been demonstrated in Fig. 9 and Fig. 10. In these figures "M" represents the speech samples of male and "F" represents the speech samples of female.

If we consider the performance of both MFCC39 and MFCC52 then it can be visualized from the Fig. 3 and Fig. 5 that MFCC52 based system has average SCR of 99.75% and 97.6% accordingly while MFCC39 holds SCR about 99.625% and 97% accordingly. But in Fig. 4 MFCC39 has an average SCR of 99% while it has 98.8% for MFCC52. From this phase, we can see an average difference between these two systems is 0.2%.

Fig. 6 to Fig. 8 represents the SCR for Phase-II. In this phase MFCC52 based system has performed better except for the test corpus (F-9 to F-13) with an average SCR of 98.77%.

Fig. 9 to Fig. 12 represents the SCR for Phase-III. In this phase, the average SCR for both the system has increased. For MFCC39 based system it becomes 99.56% while 99.36% for MFCC52 based system.

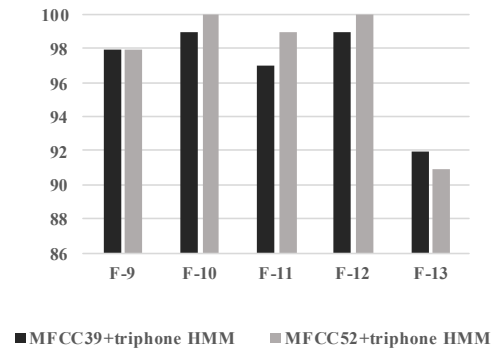


Fig. 5. Comparison of SCR between MFCC39 and MFCC52 for test corpora under Phase-I

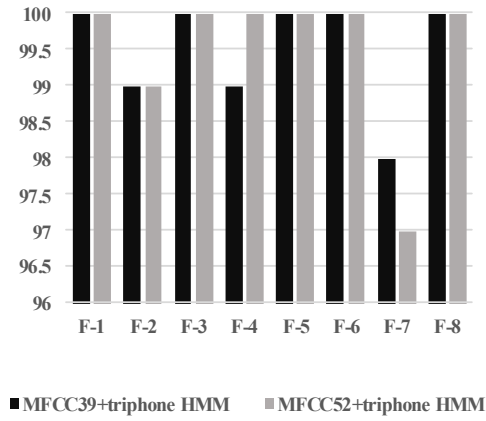


Fig. 6. Comparison of SCR between MFCC39 and MFCC52 for train corpus under Phase-II

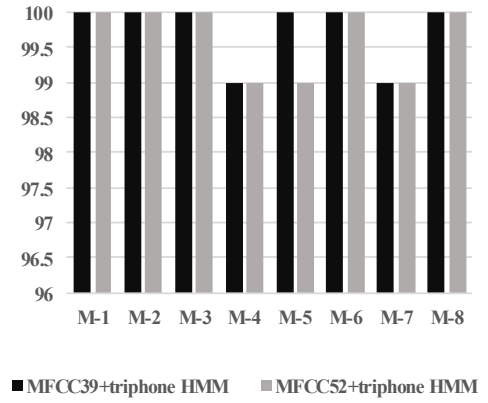


Fig. 9. Comparison of SCR between MFCC39 and MFCC52 for train corpus under Phase-III

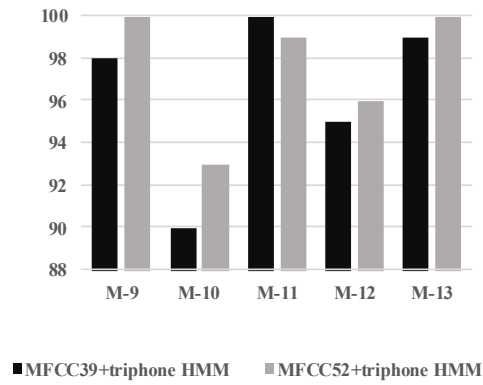


Fig. 7. Comparison of SCR between MFCC39 and MFCC52 for test corpora under Phase-II

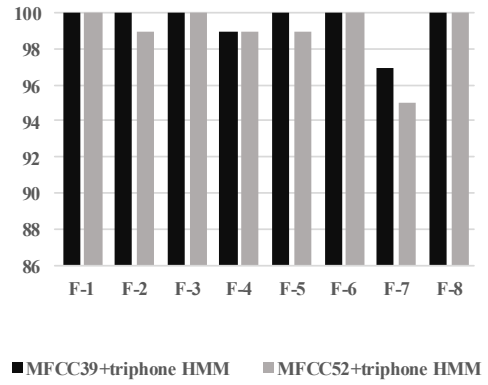


Fig. 10. Comparison of SCR between MFCC39 and MFCC52 for train corpus under Phase-III

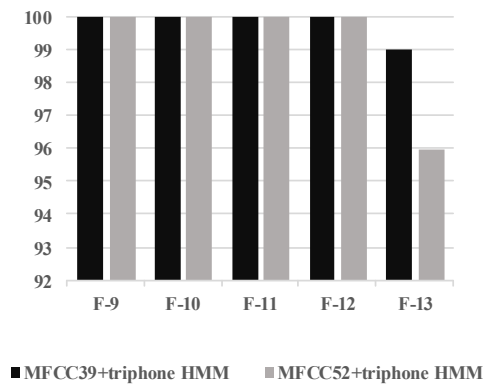


Fig. 8. Comparison of SCR between MFCC39 and MFCC52 for test corpora under Phase-II

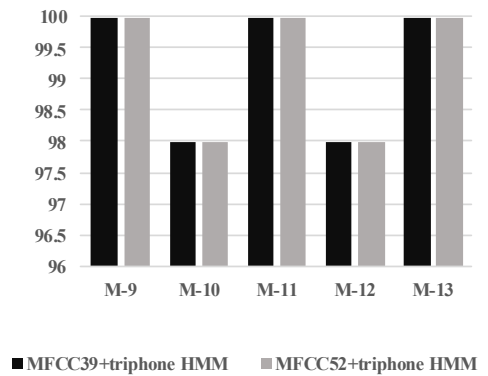


Fig. 11. Comparison of SCR between MFCC39 and MFCC52 for test corpora under Phase-III

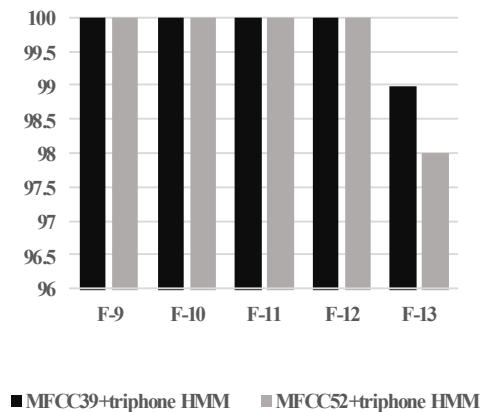


Fig. 12. Comparison of SCR between MFCC39 and MFCC52 for test corpora under Phase-III

It is also noticeable from Fig. 5 and Fig. 7 where the systems have been trained with only male and female voices respectively, MFCC52 has average SCR slightly better than MFCC39 when the systems have been tested with opposite gender. It shows average SCR of 97.6% for MFCC52 and 96.7% for MFCC39 based system.

VI. CONCLUSION AND FUTURE WORK

The performance of MFCC39 and MFCC52 based system for Bangla ASR has been compared in this paper. Experiments with MFCC52 based system show slight improvement for most of the cases except for the last phase. Therefore, we need to do a few more research on MFCC52 based system. That is why, our future works will include the study of the effect of lower sampling frequencies, noisy data, a mixture of Gaussian components and etc. on the performance of MFCC52 and MFCC39 based systems.

REFERENCES

- [1]. Wikipedia. "Bengali language". [online]. Available: http://en.wikipedia.org/wiki/Bengali_language [Accessed 27 April 2016].
- [2]. D.B. Paul, "Speech Recognition Using Hidden Markov Models," Lincoln Laboratory Journal, vol. 3, no. 1, pp. 41-62, 1990. [Online]. Available: https://www.ll.mit.edu/publications/journal/pdf/vol03_no1/3.1.3.speech_recognition.pdf [Accessed Aug. 24, 2016].
- [3]. Mohammed Rokibul Alam Kotwal et al., "Bangla Triphone HMM Based Word Recognition," in Asia Pacific Conference on Circuits and Systems (APCCAS 2010), Kuala Lumpur, 2010, pp. 883-886.
- [4]. Pronaya Prosun Das, Shaikh Muhammad Allayear, Ruhul Amin, and Zahida Rahman, "Bangladeshi Dialect Recognition Using Mel Frequency Cepstral Coefficient, Delta, Delta-delta and Gaussian Mixture Model," 8th International Conference on Advanced Computational Intelligence, Chiang Mai, Thailand, 2016, pp. 359-364.
- [5]. Md. Abul Hasnat, Jabir Mowla, Mumit Khan, (2007), "Isolated and Continuous Bangla Speech Recognition: Implementation, Performance and Application Perspective," [Online]. Available: www.researchgate.net/publication/47390524_Isolated_and_continuous_bangla_speech_recognition_implementation_performance_and_application_perspective [Accessed Aug. 24, 2016].

- [6]. Asm Sayem, "Speech analysis for alphabets in Bangla language: automatic speech recognition," Int. Journal of Engineering Research, vol. 3, no. 2, pp. 88-93, 2014.
- [7]. Youssef Zouhir and Kaïs Ouni, "A bio-inspired feature extraction for robust speech recognition," in SpringerPlus, 2014, vol. 3, doi: 10.1186/2193-1801-3-651.
- [8]. Md. Akkas Ali, Manwar Hossain and Mohammad Nuruzzaman Bhuiyan, "Automatic Speech Recognition Technique for Bangla Words," Int. Journal of Advance Science and Technology, vol. 50, Jan. 2013.
- [9]. Jyh-Shing Roger Jang, "MFCC," in Audio Signal Processing and Recognition, [Online]. Available: <http://mirlab.org/jang/books/audiosignalprocessing> [Accessed Aug. 24, 2016].
- [10]. Raghunath S.Holambe, Mangesh S. Deshpande, "Nonlinearity Framework in speech Processing," in Advances in Non-Linear Modeling for Speech Processing, 1st ed. Springer- Verlag New York, 2012, ch. 2, pp. 11-25.
- [11]. Steve Young et al., "Speech Input/Output," in The HTK Book, [Online]. Available: <http://htk.eng.cam.ac.uk/docs/docs.shtml> [Accessed Aug. 24, 2016].
- [12]. Lindasalwa Muda, Mumtaj Begam and I. Elamvazuthi, "Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques," Journal of Computing, vol. 2, issue 3, March 2010.
- [13]. J. Lyons, "Practical Cryptography," [Online]. Available: <http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfcc/> [Accessed Aug. 24, 2016].
- [14]. Nayhan Al Mahmud, "Performance Analysis of Hidden Markov Model in Bangla Speech Recognition," M.S thesis, Dept. Elect. Electron. Eng., Bangladesh Univ. Eng. Technology, Dhaka, Bangladesh, 2014.
- [15]. Biswajit Das, Sandipan Mandal, Pabitra Mitra, "Bengali speech corpus for continuous automatic speech recognition system," International Conference on Speech Database and Assessments (Oriental COCOSA), Hsinchu, 2011, pp. 51-55.

Design and Simulation of High Efficiency $Al_xGa_{1-x}As/Ga_{0.47}In_{0.53}As_yP_{1-y}$ Bandgraded 2-J Tandem Solar Cell

Mohammad Zawad Ali

Department of Electrical and Electronic Engineering
Islamic University of Technology
Gazipur-1704, Dhaka, Bangladesh
E-mail: zawad786@iut-dhaka.edu

Turja Nandy

Department of Electrical and Electronic Engineering
University of Dhaka
Dhaka-1000, Bangladesh
E-mail: turja.nandy@gmail.com

Abstract—The III–V compound semiconductor based multi-junction tandem solar cell has ultra-high efficiency and better cell stability. In this paper, a new type of AlGaAs/GaInAsP 2-J tandem solar cell has been introduced and designed by bandgrading thereby obtaining a set of descending layers having bandgap from 2.09eV – 0.84eV and grown over a Ge substrate. For matching the currents between subcells, a tunnel junction has been used as an interconnection between the two subcells. Moreover, the used materials and the characteristics of the designed cell have been discussed in the paper. After that, short circuit current, open circuit voltage, fill factor and cell efficiency have been simulated for proposed AlGaAs/GaInAsP 2-J cell under AM1.5 spectrum which have proved it as highly efficient as previously researched 3-J cell. Finally, following the newly designed AlGaAs/GaInAsP 2-J tandem solar cell, design of AlGaAs/GaInAsP/GaAsSb 3-J tandem solar cell and AlGaAs/GaInAsP/GaAsSb/GaInAs 4-J tandem solar cell design have also been introduced and proposed.

Keywords—bandgraded 2, 3, 4- junction tandem cell, optimal spectrum efficiency, fill factor and efficiency.

I. INTRODUCTION

Tandem solar cells having multiple band gaps can absorb a wide range of solar spectrum and promises to exceed the SQ limit [1]. Efficiencies of 41.3% and 32.6% have been reported for triple-junction and double-junction devices under 343 and 1026 suns respectively and such devices with a stack of single-junction cells have been proposed by many researchers [2-4]. III–V compounds have superior material characteristics, stability and bandgap energy tenability. Moreover the illumination degenerate (S-W effect) is absent in compound (III–V) tandem solar cells, thus providing higher conversion efficiency. III–V compounds are miscible therefore their energy band gap can be improved by doping of different chemical elements. This is one of the best feature for selecting a proper subcell band gap [3, 5]. The distribution of bandgap energy among the subcells in the GaInP/GaAs/InGaAs/Ge 4-J tandem solar cell are 1.8, 1.4, 1.0 and 0.7eV respectively. It has been reported that this type of band gap combination has

high spectrum efficiency and is projected to exceed 45% efficiency. The conversion efficiencies of a 3-J (InGaP/InGaAs/Ge) tandem cell are approximately 31.7% under one sun and 39.2% under 200 suns [3, 6, 7] respectively.

To design a successful tandem cell the lattice matching between subsequent subcell materials and the current matching between subcells are essential to produce a high efficiency cell. In this paper, following the high efficiency 4-J tandem cell, optimal spectrum efficiency of 2-J tandem cell is achieved by bandgrading of the subcell layers.

II. MULTI-JUNCTION SOLAR CELL

A. Tandem Cell

As a new technology, III-V multi-junction tandem solar cell offers enormously high efficiencies compared with traditional single layer solar cells [8]. The strong demand for higher efficiency photovoltaic has recently attracted considerable interest in multi-junction solar cells based on III-V compound semiconductors [9]. Depending on a particular technology, multi-junction solar cells are capable of generating approximately twice as much power under the same circumstances as traditional single layer solar cells made of silicon. The choice of materials for each subcell is determined by the requirements for lattice-matching, current-matching, and high performance optoelectronic properties [9, 10].

For optimal growth and resulting crystal quality, the crystal lattice constant of an each material must be closely matched, resulting in lattice-matched devices. A greater degree of mismatch or other growth imperfections can lead to crystal defects causing degradation in electronic properties. Since each sub-cell is connected electrically in series, the same current flows through each junction. The materials are ordered with decreasing bandgaps, E_g , allowing sub-bandgap light ($hc/\lambda < e \cdot E_g$) to transmit to the lower sub-cells where $E_{g1} > E_{g3} > E_{g2} > E_{g4}$. Therefore, for achieving current matching,

suitable bandgaps must be elected such that the design spectrum will balance the current generation in each of the subcells. Tunnel junctions are necessary since it provides a low electrical resistance connection between two subcells. It is usually highly doped and has a wide bandgap [10, 11].

B. Bandgrading

Bandgrading is a technique where by changing mole fraction the bandgap of a material can be tuned to a desired value. Several advantages can be obtained by band grading which ultimately leads to the increase in the solar cell efficiency. Different materials has a range of bandgaps that can be tuned by changing the mole fraction of the material e.g. AlGaAs has a bandgap range from 1.55 eV – 2.13eV [7] whereas GaInAsP has a much wider bandgap from 0.5eV – 2.1eV [12] both at 300K. With change of temperature the bandgap changes.

Hence by this method a wider solar spectrum can be made available to the solar cell for absorption. Theoretical analysis has shown that best results can be obtained if the energy gap gradient of the base is not too high and the energy gap gradient of the front region is high. This is due to an inverse window effect that occurs in the base when the energy gap gradient of the base is high [13]. A combination of graded material followed by a p-n junction could also be properly designed to increase the minority-carrier collection efficiency, through a built-in electric field [14, 15].

III. MATERIAL SELECTION, STRUCTURE OF 2-J TANDEM CELL AND PROPOSITION

As mentioned in the previous section that for optimal spectrum efficiency the bandgap of the subcells in a 4-J tandem solar cell were 1.8, 1.4, 1.0 and 0.7eV. To utilize this bandgap variation in a 2-J tandem solar cell bandgrading of descending order is necessary for the different layers of the two subcells and appropriate materials should be selected which not only provides a range of different bandgaps but whose lattice constants are also closer to each other.

Fig. 1 shows the overall bandgraded structural diagram of the $\text{Al}_x\text{Ga}_{1-x}\text{As}/\text{Ga}_{0.47}\text{In}_{0.53}\text{As}_y\text{P}_{1-y}$ 2-J tandem solar cell. Now, bandgap variation for $\text{Al}_x\text{Ga}_{1-x}\text{As}$ and $\text{Ga}_{0.47}\text{In}_{0.53}\text{As}_y\text{P}_{1-y}$ are shown in Fig. 2 and Fig. 3, where $\text{Al}_x\text{Ga}_{1-x}\text{As}$ and $\text{Ga}_{0.47}\text{In}_{0.53}\text{As}_y\text{P}_{1-y}$ provides a range of bandgaps with the variation of x and y [7, 16] which is suitable for the design of a 2-J tandem solar cell, utilizing the bandgap range for optimal spectrum efficiency by bandgrading. Here, Table I. represents the bandgaps of $\text{Al}_x\text{Ga}_{1-x}\text{As}$ and $\text{Ga}_x\text{In}_{1-x}\text{As}_y\text{P}_{1-y}$ for selected values of x and y.

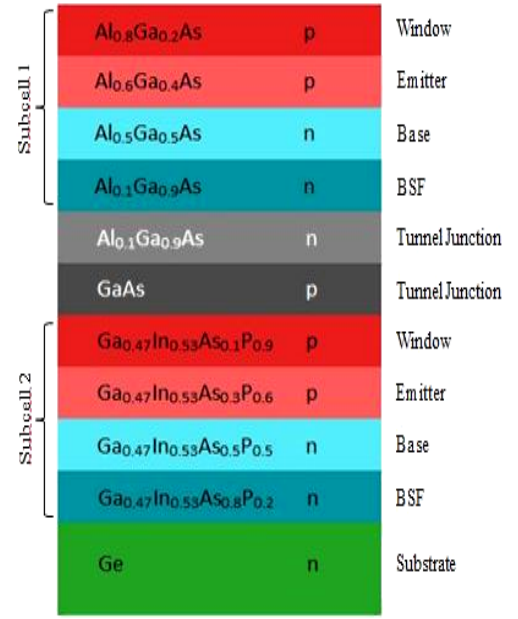


Fig 1: Schematic illustration of designed 2-Junction solar cell.

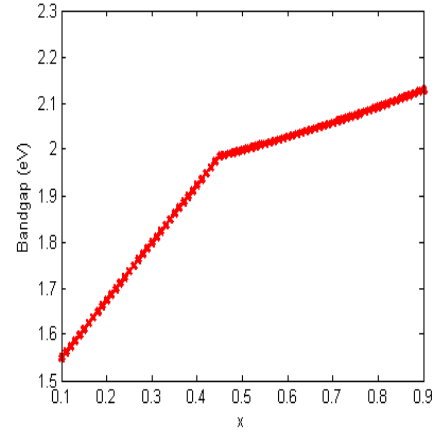


Fig 2: Bandgap vs. x for $\text{Al}_x\text{Ga}_{1-x}\text{As}$

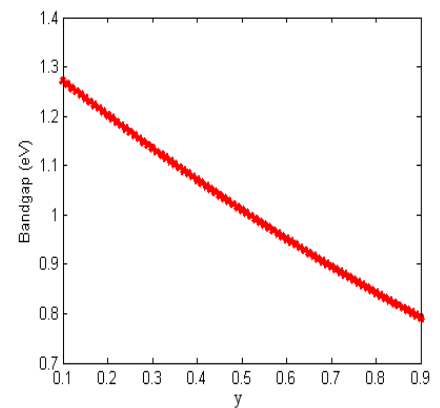


Fig 3: Bandgap vs. y for $\text{Ga}_{0.47}\text{In}_{0.53}\text{As}_y\text{P}_{1-y}$

TABLE I. Bandgaps of $\text{Al}_x\text{Ga}_{1-x}\text{As}$ and $\text{Ga}_x\text{In}_{1-x}\text{As}_y\text{P}_{1-y}$ for selected values of x and y .

Layer	Material	x	y	Bandgap (eV)
Window	$\text{Al}_x\text{Ga}_{1-x}\text{As}$	0.8	-	2.09
Emitter	$\text{Al}_x\text{Ga}_{1-x}\text{As}$	0.6	-	2.03
Base	$\text{Al}_x\text{Ga}_{1-x}\text{As}$	0.5	-	2.00
BSF	$\text{Al}_x\text{Ga}_{1-x}\text{As}$	0.1	-	1.55
Tunnel (n)	$\text{Al}_x\text{Ga}_{1-x}\text{As}$	0.1	-	1.55
Tunnel (p)	GaAs	0	-	1.42
Window	$\text{Ga}_x\text{In}_{1-x}\text{As}_y\text{P}_{1-y}$	0.47	0.1	1.27
Emitter	$\text{Ga}_x\text{In}_{1-x}\text{As}_y\text{P}_{1-y}$	0.47	0.3	1.14
Base	$\text{Ga}_x\text{In}_{1-x}\text{As}_y\text{P}_{1-y}$	0.47	0.5	1.01
BSF	$\text{Ga}_x\text{In}_{1-x}\text{As}_y\text{P}_{1-y}$	0.47	0.8	0.84

Each of the subcell consists of a window layer at the top followed by an emitter and base layer with BSF layer at the bottom. Doping level of greater or equal to 10^{18} cm^{-3} is required in window layer [17], thus we have used the doping level of $5 \times 10^{18} \text{ cm}^{-3}$. Typical width of the emitter is $\sim 100 \text{ nm}$ [18], hence we have used emitter's of width 200 nm in 1st subcell and 300 nm in 2nd subcell. For the minority carrier, diffusion length is small therefore the thickness of each subcell should be short to allow the carriers to reach the contact before they recombine thereby maximizing the subcell current density. Thickness of each subcell absorber is set as AlGaAs - $1 \mu\text{m}$ and GaInAsP - $3 \mu\text{m}$. Where, Table II, represents doping and thickness of different layers. The two subcells are connected by a tunnel junction. For sub-cell matching, improved overall cell characteristics and stability the doping concentrations of the tunnel junctions must be very high.

The stacked cells in a tandem cell are connected in series therefore for voltage addition an optically transparent and low resistance interconnects must be made between the stacked cells [3, 19]. Current through such interconnections must flow from n-type layer into p-type layer and typically tunnel junctions are used for joining the stacked solar cells, to form such an interconnection. Doping level concentrations of $10^{19} - 10^{20} \text{ cm}^{-3}$ on both sides of the junction are required for a low resistance interconnection to minimize the V_{oc} loss on it [3, 20, 21]. The solar cell is designed appropriately so that it can be grown layer by layer over the Ge substrate. The thickness of the tunnel junction has been set to be 20 nm . Tunnel junction matches the lattice constant between the top and bottom subcell and creates good light permeability [3, 22-24].

TABLE II. Doping and thickness of different layers.

Layer	Doping type	Doping level, 10^{17} cm^{-3}	Thickness (μm)
Window	P	50	0.02
Emitter	P	1	0.2
Base	N	1	1
BSF	N	10	0.1
Tunnel (n)	N	100	0.01
Tunnel (p)	P	100	0.01
Window	P	50	0.03
Emitter	P	1	0.3
Base	N	1	3
BSF	N	10	0.2

Thus a 2-J tandem solar cell was designed whose optimal spectral efficiency is attained by bandgrading.

IV. SIMULATION AND RESULTS

To analyze the whole cell, wxAMPS has been used to simulate the I-V characteristic curves of the cell. Table I and Table II show the bandgaps and doping concentrations with layer thicknesses respectively for the different layers. Due to the heavy doping of tunnel junction, defects occurred by the heterojunction between tunnel junction and subcell are negligible. The parameters and curve are observed under AM1.5G and temperature of 300 K .

A. Efficiency Measurement

With the range of bandgaps offered by $\text{Al}_x\text{Ga}_{1-x}\text{As}$ and $\text{Ga}_{0.47}\text{In}_{0.53}\text{As}_y\text{P}_{1-y}$ a set of descending bandgap layers were obtained by bandgrading. The I-V characteristic curve of the 2-J tandem solar cell is shown in Fig. 4, and the designed solar cell produces an open circuit voltage (V_{oc}) of 1.262 V , short circuit current density (J_{sc}) of 32.1187 mA/cm^2 , fill factor (FF) of 0.8169 and an overall cell efficiency (η) of 29.55% .

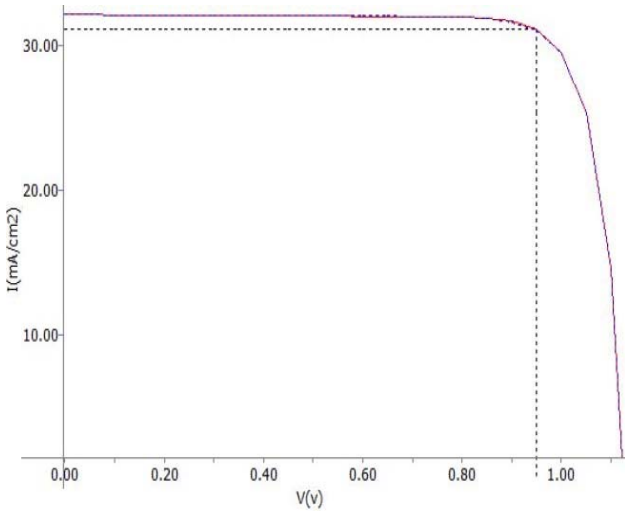


Fig 4: I-V characteristic curve of the designed 2-Junction Solar cell.

B. Tunnel Junction Optimization

The tunnel junction has been made from $\text{Al}_x\text{Ga}_{1-x}\text{As}$ of n & p type. By varying the mole fraction x , different bandgap combinations of the n type & the p type materials, of the tunnel junction, were applied and the respective results were obtained. The n type material of the tunnel junction has been given the bandgap of 1.55 eV while the bandgap of p type material has been changed from 1.424 eV [25] (for $x = 0$, i.e. GaAs) to 1.55 eV [7] (for $x = 0.1$) and the corresponding solar cell efficiency has been calculated. The variation of the solar cell efficiency with the change in bandgap is shown in Fig. 5. From the above figure it is found that a larger bandgap difference between the n & p type material of the tunnel junction provides lower resistance connection between subcells hence produces a higher efficiency of the solar cell.

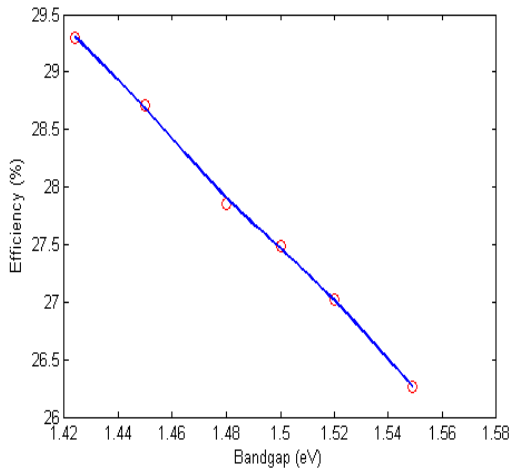


Fig. 5: Variation of efficiency with the change of bandgap of p-type material of the tunnel junction.

Hence for the p type material of the tunnel junction, the mole fraction $x=0$ have been used, which gives the bandgap that of GaAs (1.424 eV), thus obtaining a higher bandgap difference of (1.55 eV - 1.42 eV) 0.13 eV between the n & p type material of the tunnel junction and correspondingly achieving a higher efficiency.

IV. PROPOSED DESIGN FOR 3-J AND 4-J TANDEM SOLAR CELLS

An AlGaAs/GaInAsP 2-J tandem cell has been designed to simulate the cell's overall characteristics by wxAMPS under AM1.5 spectrum. In consideration of the 2-J tandem cell, designs for the AlGaAs/GaInAsP/GaAsSb 3-J tandem cell and AlGaAs/GaInAsP/GaAsSb/GaInAs 4-J tandem cell have been proposed and depicted in Fig. 6(a), and Fig. 6(b), respectively.

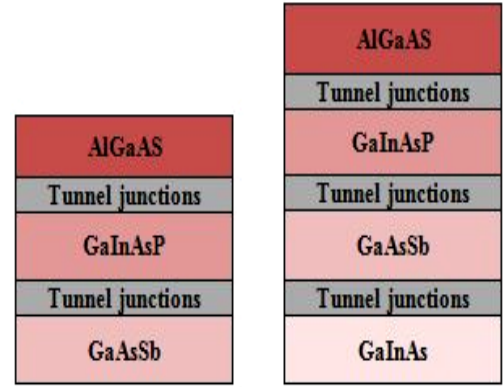


Fig. 6(a). Design proposal for 3-J tandem solar cell.

Fig. 6(b). Design proposal for 4-J tandem solar cell.

V. CONCLUSION

In this research, by bandgrading a range of bandgaps from 2.09 eV – 0.84 eV has been used in designing the $\text{Al}_x\text{Ga}_{1-x}\text{As}/\text{Ga}_{0.47}\text{In}_{0.53}\text{As}_y\text{P}_{1-y}$ 2-J solar cell structure of about 4.5 μm thick. From the simulation of the cell with wxAMPS, a cell efficiency, $\eta = 29.55\%$ has been observed. In case of InGaP/InGaAs/Ge 3-J tandem solar cell, it has a conversion efficiency of around 31.7% under one sun (AM1.5G). Therefore it can be concluded that with one less junction and one less compound type being required the performance in terms of conversion efficiency of the AlGaAs/GaInAsP 2-J tandem solar cell grown on Ge substrate and InGaP/InGaAs/Ge 3-J tandem solar cell are quite similar. Following the design of the 2-J tandem solar cell an AlGaAs/GaInAsP/GaAsSb 3-J tandem solar cell and an AlGaAs/GaInAsP/GaAsSb/GaInAs 4-J tandem solar cell have also been proposed. This overall proposition and investigation can be used in further future researches and developments of micro and nano-structured multijunction tandem solar cells.

ACKNOWLEDGMENT

We would like to acknowledge the use of AMPS-1D program developed by Dr. Fonash's group at Pennsylvania State University (PSU) and wxAMPS program developed by UIUC in collaboration with Nankai University of China.

REFERENCES

- [1] Green, Martin A. "Third generation photovoltaics: solar cells for 2020 and beyond." *Physica E: Low-dimensional Systems and Nanostructures* 14, vol. 1, pp. 65-70, 2002.
- [2] Luque, Antonio, and Steven Hegedus, eds. *Handbook of photovoltaic science and engineering*. John Wiley & Sons, 2011.
- [3] Siyu, Lü, and Qu Xiaosheng, "AlGaAs/GaAs tunnel junctions in a 4-J tandem solar cell." *Journal of Semiconductors* 32, no. 11, p. 112003, 2011.
- [4] Green, Martin A., Keith Emery, Yoshihiro Hishikawa, and Wilhelm Warta. "Solar cell efficiency tables (version 36)." *Progress in Photovoltaics: Research and Applications* 18, vol. 5, pp. 346-352, 2010.
- [5] Sa, B. E., N. Szabo, H. Döschner, U. Seidel, C. Höhn, K. Schwarzburg, and T. Hannappel. "Lifetime and performance of ingaasp and ingaas absorbers for low bandgap tandem solar cells." In *2009 IEEE 34th Photovoltaic Specialists Conference (PVSC)*, pp. 001090-001093, 2009.
- [6] Cornfeld, Arthur B., Daniel Aiken, Benjamin Cho, A. Vance Ley, Paul Sharps, Mark Stan, and Tansen Varghese. "Development of a four sub-cell inverted metamorphic multi-junction (IMM) highly efficient AM0 solar cell." In *2010 IEEE 35th Photovoltaic Specialists Conference (PVSC)*, pp. 000105-000109, 2010.
- [7] Adachi, Sadao. "GaAs, AlAs, and Al_xGa_{1-x}As: Material parameters for use in research and device applications." *Journal of Applied Physics* 58, vol. 3, R1-R29, 1985.
- [8] Luque, A., A. Martí, C. Stanley, N. López, L. Cuadra, D. Zhou, J. L. Pearson, and A. McKee. "General equivalent circuit for intermediate band devices: Potentials, currents and electroluminescence." *Journal of Applied Physics* 96, vol. 1, pp. 903-909, 2004.
- [9] King, Richard R. "Multijunction cells: Record breakers." *Nature Photonics* 2, vol. 5, pp. 284-286, 2008.
- [10] Yamaguchi, Masafumi, Tatsuya Takamoto, and Kenji Araki. "Super high-efficiency multi-junction and concentrator solar cells." *Solar Energy Materials and Solar Cells* 90, vol. 18, pp. 3068-3077, 2006.
- [11] Walukiewicz, Wladyslaw, Joel W. Ager III, and Kin Man Yu. "Low resistance tunnel junctions for high efficiency tandem solar cells." U.S. Patent Application 12/528,394, filed April 9, 2008.
- [12] A.T. Gorelenok, A.G. Dzigasov, P.P. Moskvina, V.S. Sorokin, I.S. Tarasov, *Sov. Phys.Semicond.*, 15, no.12, pp.1400-1402, 1981.
- [13] Sassi, G. "Theoretical analysis of solar cells based on graded band-gap structures." *Journal of applied physics* 54, no. 9, p. 5421-5427, 1983.
- [14] M. Wolf, *Proc. IRE* 48, 1246 (1960).
- [15] Ellis, Brian, and Trevor Simpson Moss. "Calculated efficiencies of practical GaAs and Si solar cells including the effect of built-in electric fields." *Solid-State Electronics* 13, vol. 1, p 1-24, 1970.
- [16] Goldberg Yu.A. and N.M. Schmidt Handbook Series on Semiconductor Parameters, vol.2, M. Levinstein, S. Rumyantsev and M. Shur, ed., World Scientific, London, pp. 153-179, 1999.
- [17] Luque, Antonio, and Steven Hegedus, eds. *Handbook of photovoltaic science and engineering*. John Wiley & Sons, 2011.
- [18] Hamaker, H. C. "Computer modeling study of the effects of inhomogeneous doping and/or composition in GaAs solar-cell devices." *Journal of applied physics* 58, vol. 6, pp. 2344-2351, 1985.
- [19] Hauser, John R., Zach Carlin, and S. M. Bedair. "Modeling of tunnel junctions for high efficiency solar cells." *Applied Physics Letters* 97, vol. 4, pp. 042111, 2010.
- [20] Nishioka, Kensuke, Tatsuya Takamoto, Takaaki Agui, Minoru Kaneiwa, Yukiharu Uraoka, and Takashi Fuyuki. "Evaluation of InGaP/InGaAs/Ge triple-junction solar cell and optimization of solar cell's structure focusing on series resistance for high-efficiency concentrator photovoltaic systems." *Solar Energy Materials and Solar Cells* 90, vol. 9, pp. 1308-1321, 2006.
- [21] Braun, Avi, Baruch Hirsch, Eugene A. Katz, Jeffrey M. Gordon, Wolfgang Guter, and Andreas W. Bett. "Localized irradiation effects on tunnel diode transitions in multi-junction concentrator solar cells." *Solar Energy Materials and Solar Cells* 93, vol. 9, pp. 1692-1695, 2009.
- [22] Wheeldon, Jeffrey F., Christopher E. Valdivia, Alex Walker, Gitanjali Kolhatkar, Trevor J. Hall, Karin Hinzer, Denis Masson et al. "AlGaAs tunnel junction for high efficiency multi-junction solar cells: Simulation and measurement of temperature-dependent operation." In *2009 34th IEEE Photovoltaic Specialists Conference (PVSC)*, pp. 000106-000111, 2009.
- [23] Wanlass, Mark, Phil Ahrenkiel, David Albin, Jeff Carapella, Anna Duda, Keith Emery, Daniel Friedman et al. "Monolithic, ultra-thin GaInP/GaAs/GaInAs tandem solar cells." In *2006 IEEE 4th World Conference on Photovoltaic Energy Conference*, vol. 1, pp. 729-732, 2006.
- [24] Seidel, U., B. E. Sağol, N. Szabó, K. Schwarzburg, and T. Hannappel. "InGaAs/GaAsSb-interface studies in a tunnel junction of a low band gap tandem solar cell." *Thin Solid Films* 516, vol. 20, pp. 6723-6728, 2008.
- [25] Streetman, Ben G., and Sanjay Banerjee. *Solid state electronic devices*. Vol. 4. New Jersey: Prentice Hall, 2000.

A Compact Multiband Antenna for 4G/LTE and WLAN Mobile Phone Applications

Foez Ahmed*, Md. Halim Miah Chowdhury[†] and Nisab Hasan[‡]

*Department of Information and Communication Engineering
Faculty of Engineering, University of Rajshahi, Rajshahi-6205
Email: *foez28@ru.ac.bd, [†]halimice08@gmail.com, [‡]nisabhasan16@gmail.com

Abstract—A very simple multiband monopole antenna system is presented. The proposed antenna system consists of two alphabetic letter shaped elements: one is C-shaped and another is inverted L-shaped – all together can combinedly operate the wideband frequency range of LTE-1, 2, 3, 7, 13, 14, 22, 40, 43 and WLAN bands. The C-shaped is responsible for leading the lower bands and L-shaped for higher bands. The reflection coefficient is achieved 6 dB in case of LTE-13 and LTE-14 bands, whereas for all other bands is greater than 10 dB. This multiband antenna element has a compact volume of $20 \times 24 \times 0.8 \text{ mm}^3$ with total system size of $20 \times 70 \times 0.8 \text{ mm}^3$, which is suitable for 4G-LTE ultra slim mobile phones and wireless devices.

Keywords—Handset Antenna, Multiband, LTE, WLAN, 4G, Monopole.

I. INTRODUCTION

Due to growing applications and communication needs, multiband antenna system is essential for multifunctional handsets [1]–[3]. On the other hand, compact and light weight mobile handsets are also demanded [4]. Several structures are available for designing multiband antenna such as planar inverted-F antenna (PIFA), monopole antenna and slot antenna [5]. By using monopole element, multiple resonances can be excited to cover a large bands with reasonable system size.

Over the last few years, a large number of multiband antenna systems have been investigated for mobile handsets. A wideband antenna system was proposed [6] to cover a frequency range of 1541–2103MHz at 10dB and 2288–2534MHz at 8 dB. These frequency bands were also covered by [7] and [8] at reflection coefficient greater than 10dB. The antenna system for LTE/GSM/UMTS mobile phone application with total system size of $40 \times 115 \times 10 \text{ mm}^3$ was proposed in [9]. However the system size is not suitable for a light weight and slim handsets. Another multiband antenna [10] was also developed for GSM and UMTS bands but failed to cover the most essential WLAN band and even the system size ($40 \times 100 \times 5 \text{ mm}^3$) was quite large. Recently a multiband antenna was reported [11] for GSM 850/900, DCS1800, PCS1900, UMTS2100, and LTE2300/2500 bands, but its total system size ($40 \times 115 \times 10 \text{ mm}^3$) makes handsets bulky and uncomfortable. However, all of these antenna system work either for UMTS/WLAN or GSM/LTE/WLAN bands. Even they are unable to cover large number of LTE bands including LTE-700 band with reasonable system size for 4G mobile handsets. In this paper, a multiband monopole antenna has been developed for 4G wireless standards that can operate

for LTE-1(1920-2170MHz), LTE-2(1850-1990MHz), LTE-3(1710-1880MHz), LTE-7(2500-2690MHz), LTE-13(746-787MHz), LTE-14(758-798MHz), LTE-40(2300-2400MHz), LTE-22(3410-3590MHz), LTE-43(3600-3800MHz), and WLAN(2400-2484MHz) bands. The total system size of the proposed antenna is $20 \times 70 \times 0.8 \text{ mm}^3$, which could be a good candidate for future generation handsets.

II. ANTENNA CONFIGURATION AND PARAMETER STUDY

A. Antenna Structure

The detailed dimensions of the proposed monopole antenna are illustrated in Fig. 1. The antenna consists of two alphabetic letter shaped elements: C-shaped and an inverted L-shaped. Both C- and L- shaped elements are printed on the front side of a low cost thin FR-4 substrate (dielectric constant $\epsilon_r = 4.4$, loss tangent = 0.02 and thickness = 0.8mm). A ground plane of size $20 \times 46 \text{ mm}^2$ is printed on the back side of the substrate. A $1.2 \times 46 \text{ mm}^2$ 50 Ω microstrip line is used to feed the monopole. However, the proposed antenna was optimized and analyzed by using commercial software Ansoft HFSS [12]. The optimized geometric parameters and values are listed in Table I.

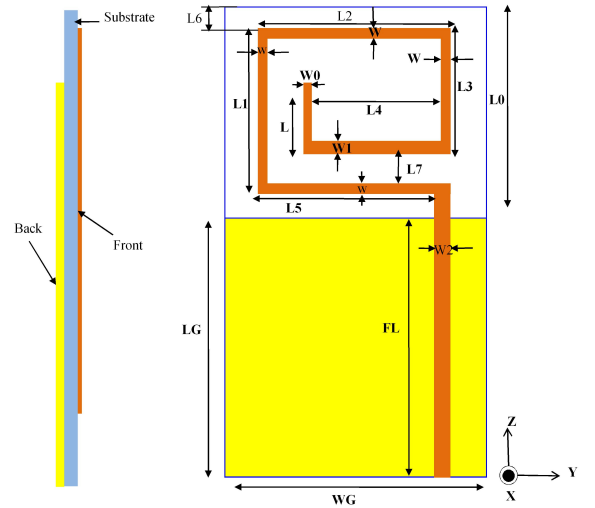


Fig. 1. Optimized structure of the proposed multiband Antenna.

The proposed antenna system can achieve wide bandwidths which are around 55MHz (at lower band, for reflection coefficient (S_{11}) > 6dB, and Voltage Standing Wave Ratio (VSWR) < 3), 1300MHz (at middle band, for S_{11} > 10dB, and

TABLE I. OPTIMIZED PARAMETERS & VALUES FOR THE PROPOSED ANTENNA (UNIT:MILLIMETERS)

Parameters	Values	Parameters	Values
LG	46	WG	20
LO	24	L1	22.75
L2	16	L3	20.6
L4	12.3	L	4.9
L5	14.8	L6	0.25
L7	1	W	1
W0	0.7	W1	1.1
W2	1.2	FL	46

VSWR<2) and 440MHz (at higher band for S11>10dB, and VSWR< 2) shown in Fig. 2 and Fig. 3 which is better than those of reported in [6]–[8].

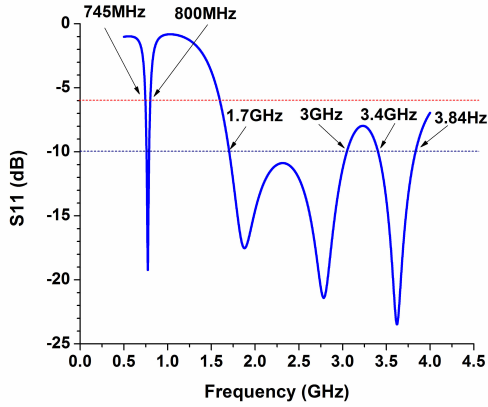


Fig. 2. Simulated S-parameter of the proposed multiband Antenna.

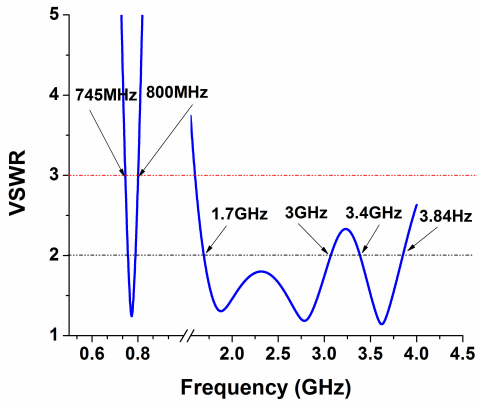


Fig. 3. VSWR of the proposed multiband Antenna.

B. Operating Principle and Parameter Study

The basic operating principles are illustrated in Fig. 4 and simulated reflection coefficient is plotted in Fig. 5. For the C-shaped element, it is observed that three wide range of resonant modes are excited at around 1.14GHz, 2.4GHz and 2.9GHz. However, it is required to improve the impedance matching to cover the higher order frequency bands. The length of this element from its feed point is about 117 mm that can excite a half-wavelength ($\lambda/2$) resonant mode at 1.14 GHz and

other two modes may be excited due to the combined effects of ground plane and C-shaped elements. Next by adding an inverted L-shaped element, the 4th resonant mode is excited at around 3.6 GHz. In the mean time, the first, second and third resonant modes are shifted towards 770MHz, 1.87GHz and 2.8GHz respectively with impedance matching.

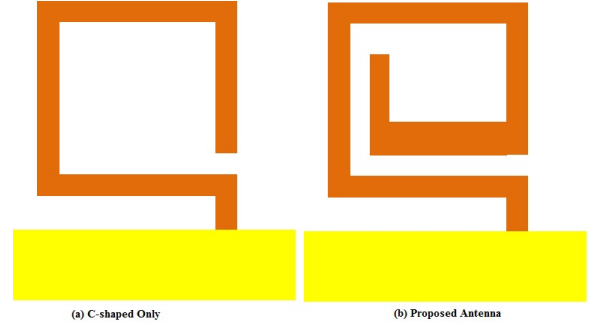


Fig. 4. Antenna structure in case of (a) C-shpaed Only, and (b) Proposed model.

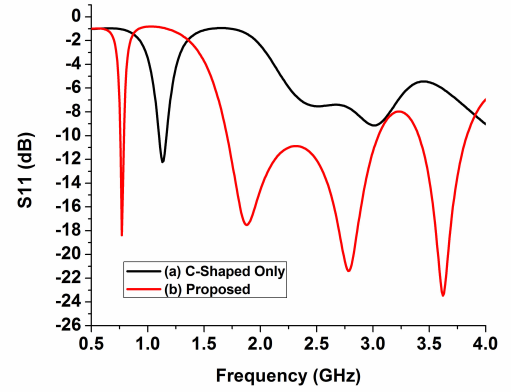


Fig. 5. Simulated S-parameter in case of (a) C-shpaed only, and (b) Proposed model.

Furthermore, the resonant characteristics over the entire bands can be controlled by tuning the length, L of L-shaped element (Fig. 1) that is shown in Fig. 6. In Fig. 6, it is observed that the reflection coefficient is changed at varying lengths from L = 4mm to 6mm. While L is decreased, all the resonant frequencies are shifted towards the higher frequency band or vice-versa.

III. RESULT AND DISCUSSION

A. Radiation Characteristics

The 3D radiation pattern of the proposed antenna is plotted in Fig. 7. It is observed that the radiation pattern is the same as like as that of a dipole antenna. It shows an omnidirectional radiation pattern which is desirable in case of mobile phone applications.

B. Peak Gain (dBi) and Efficiency(%)

The simulated peak gain and antenna efficiency of the proposed antenna system are plotted in Fig. 8 and Fig. 9

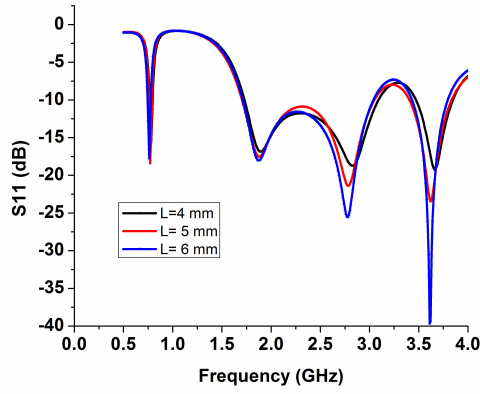


Fig. 6. Effect of the length (L) on S_{11} of the proposed multiband Antenna.

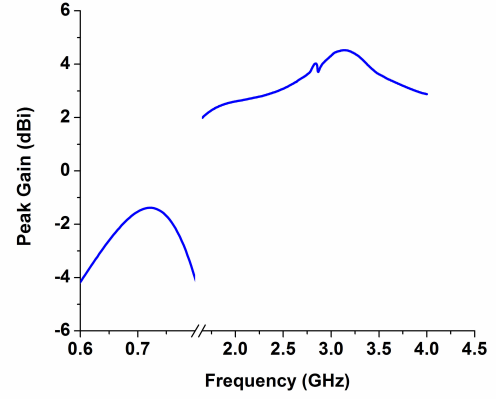


Fig. 8. Peak Gain (dBi) of the proposed multiband Antenna.

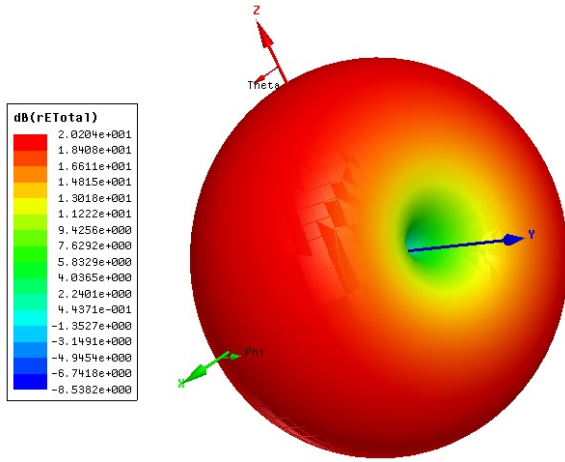


Fig. 7. 3D Radiation pattern of the proposed multiband Antenna.

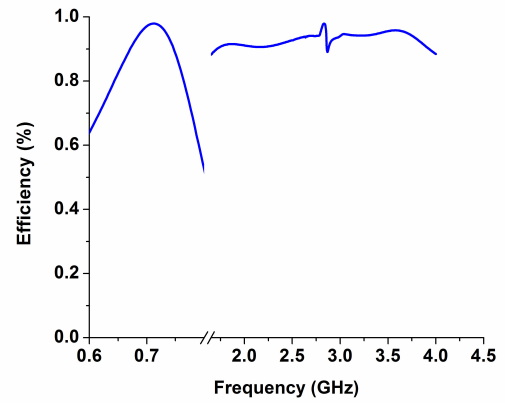


Fig. 9. Efficiency (%) of the proposed multiband Antenna.

respectively. At the lower band, the peak gain is varied from -1.58 dBi to -4.28 dBi, whereas at the upper band peak gain is fluctuated from 2 dBi to 4 dBi. The graph shows lower gain at LTE 700 band, perhaps due to the compactness of antenna structure. The radiation efficiency is changed from 90% to 50% at the lower bands. However, for the higher bands, the antenna efficiency is almost constant and reached at 95%. In terms of peak gain and efficiency, our proposed antenna outperforms as compared of [8]–[11].

IV. CONCLUSION

A simple and very low-profile compact multiband antenna is presented. The total system size is $20 \times 70 \times 0.8 \text{ mm}^3$ and the antenna occupies a small area of $20 \times 24 \text{ mm}^2$. The proposed antenna is successfully implemented and it covers multiple bandwidths which are 745–800 MHz, 1.7–3 GHz and 3.4–3.8 GHz for LTE and WLAN bands. The optimized antenna system satisfies the reflection coefficient condition, $|S_{11}| > 6 \text{ dB}$ for lower band and $|S_{11}| > 10 \text{ dB}$ for the higher band. The simulated results show an omnidirectional radiation pattern, high gain and high efficiency. Hence, the proposed antenna could be applied in today's smart phone applications for wireless communication.

ACKNOWLEDGMENT

The authors would like to thank the Key Laboratory of **Antenna for Wireless Communications**, Department of Information and Communication Engineering, University of Rajshahi, Rajshahi-6205, for supporting all types of technical as well as simulation facilities.

REFERENCES

- [1] Y. X. Guo, M. Y. W. Chia and Z. N. Chen, *Miniature built-in Quadband Antennas for Mobile Handsets*, IEEE Antenna Wireless Propag. Lett., vol. 2, pp. 3032, 2004.
- [2] I. Ang, Y. X. Guo, and Y. W. Chia, *Compact internal quad-band antenna for mobile phones*, Micro. Opt. Technol. Lett., vol. 38, no. 3, pp. 217223, Aug. 2003.
- [3] B. K. Yu, B. Jung, H. J. Lee, F. J. Harackiewicz and B. Lee, *A folded and bent internal loop antenna for GSM/ DCS/ PCS operate of mobile handset applications*, Microw. Opt. Technol. Lett., vol. 48, no. 3, pp.463467, 2006.
- [4] Y. X. Guo, I. Ang and M. Y. W. Chia, *Compact internal multiband antennas for mobile handsets*, IEEE Antennas Wireless Propag. Lett., vol. 2, pp. 143146, 2003.
- [5] K. L. Wong, *Planar Antennas for Wireless Communications*, New York: Wiley, 2003.
- [6] Z. LI, and J. WANG, *A Novel Broad Band Antenna of Planar Monopole*, Microwave Conference Proceedings, 2005. APMC 2005. Asia-Pacific Conference Proceedings. Vol. 4. IEEE, 2005.

- [7] T. Cho, S. NAM. and H. Lee , *Penta- Band Internal Antenna for Mobile Handset*, Microwave Conference, 2009. APMC 2009. Asia Pacific. IEEE, 2009. .
- [8] Y. J. Cho, Y. S. Shin, S. H. Hwang, K. H. Kim and S. O. Park, *A Miniature Wideband Internal Antenna with Dual Resonant Structures* , Asia-Pacific Microwave Conference Proceedings. Vol. 2. IEEE, 2005.
- [9] Kin- L. W. and W. Y. Chen, *Small- size printed loop- type antenna integrated with two stacked coupled- fed shorted strip monopoles for eight- band lte/ gsm/ umts operation in the mobile phone* , Microwave and Optical Technology Letters, Vol. 52, No. 7, July 2010
- [10] S. H. Lee , K. J.Kim , B. N. Kim , J. K. Oh , Y. J. Yoon, *Multi-band Coupled Feed Loop Antenna for Mobile Handset* , In Microwave Conference, 2009. APMC 2009. Asia Pacific, pp. 2703-2706. IEEE, 2009.
- [11] C. Q. Lei, Y. L. Ban, J. H. Chen, and Joshua Le- Wei Li , *Small-size slot antenna for seven- band LTE/ WWAN mobile phone operation* , New York: Wiley, 2003
- [12] HFSS v.13, *Ansoft Corporation*, Pittsburgh, PA, USA.

Different Control Schemes of Entire Microgrid: A Brief Overview

Md. Ashib Rahman* and Md. Rabiul Islam[†]

*Department of Electrical and Electronic Engineering, North Bengal International University, Rajshahi, Bangladesh

[†]Department of Electrical and Electronic Engineering, Rajshahi University of Engineering & Technology, Rajshahi, Bangladesh

Email: m.rahman.bd@ieee.org and rabiulbd@hotmail.com

Abstract—The main objective of this very literature to present a brief overview of the overall control schemes used so far to control an entire microgrid system. The control schemes are classified according to their applications and operations. Brief descriptions of them are also provided. The overall control architecture is divided into four basic control levels, such as: tertiary, secondary, primary, and the last is internal control scheme. Trends and different strategies used so far, in above control levels are clearly showed in block diagrams to take a quick review.

Keywords—Microgrid, smartgrid, hierarchical control.

I. INTRODUCTION

Energy sector is approaching more and more towards decentralized, distributed system to pass the problems faced by the conventional power network: increased carbon di-oxide emissions, voltage and service disturbances, lack of reliability and stability, rise of certain power losses etc. [1], [2]. To cope with this changing genre, new innovative grid architectures are emerging. Microgrid (MG), nanogrid (NG) [3], virtual power plant (VPP) [4] are some examples of the modern distributed power network. Microgrid, as the name implies, is a small low voltage electrical network, situated at the distribution level for efficient power distribution to its end users [5]. All of the networks mentioned above form the revolutionary idea of bidirectional Smart Grid (SG).

MG is the basic key element to implement the SG network. Most interestingly, in SG environment, *consumer* will be termed as *prosumer* [6]. From the last few years, extensive research and small scale to large scale project works are going on to conceptualize the true meaning of MG. This results in to produce 4GW of power worldwide based on MG concept, where North America single-handedly contribute 67% of it. [7], [8].

The notable components of basic MG is distributed generator (DG, 10-15 MW), energy storage system (ESS), local loads, power electronic (PE) converters, point of common coupling (PCC) and, intelligent breaker for making MG to operate in grid-tied or islanded mode [9]–[11]. MG presents diverse benefits including, superior steady-state and transient characteristics with integrated DER units and abrupt load alteration, positive environmental impact, greater reliability and efficiency, improved power quality and control and, better energy scheduling and market strategy [1]. But these benefits

come with challenges like: reliable, economic and schedule dispatch of load units, planning apposite demand side management (DSM) scheme, examining the viability of protection tactics to guarantee bi-directional power drift at delivery level, output voltage and frequency handling technique of integrated PE devices to extract optimum power flow from distributed energy resources (DERs) and to facilitate plug-n-play capability to MG and, devising user friendly market policy [11]. So, various researchers use numerous control techniques in different control levels to keep the MG stable under any faults and disturbances.

To win the challenges mentioned above, three types of basic controller can be used: 1) Distributed generation (DG) governor linked with every DG unit and storage device, 2) MG central controller (MGCC) and, 3) Load regulator, as stated in [4]. Authors in [4] and [12], investigated zero to three levels of MG control architecture to adapt to ISA-95 std. model for finer energy management control with greater flexibility and smartness. Droop control schemes, adopting the hierarchical control architecture (tertiary control, secondary control, primary control, inner control) from the UCTE (Union for the Co-ordination at Transmission of Electricity, Continental Europe) for appropriate power sharing in decentralized MG system are discussed in [13]. An excellent and informative review on distributed control techniques is highlighted in [14]. Here, the authors intelligently combine the centralized and decentralized control schemes and new distributed control schemes emerged. Hierarchical droop control review is analyzed in [10], where the authors divided the whole MG control operation in three particular control levels, grid level, management level and field level and, their control functions are determined. Another good literature on hierarchical control scheme is [15] where theory, mathematical formulation and modification of droop based methods are discussed comprehensively. In [8], detail comparison of centralized and decentralized control (secondary) of EMS is reviewed. The application of IEC/ISO 62264 standards to MG is proposed in [16] from hierarchical control perspectives. Here the standard can be explained in five levels. The extra level is level 4-market structure and business model. Operation and control strategies are explained here [1] but not in a categorized way. Power management strategies for hybrid MG are presented in [17]. But categorization is not done based on control levels rather based on coupled MG system. In

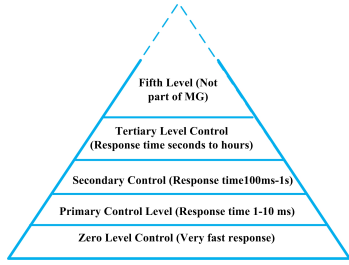


Fig. 1. IEC/ISO 62264 MG control hierarchy

[18], state-of-the-art primary control and inner loop control methods, applied to parallel inverters are reviewed, neglecting the other control levels. Review of hierarchical control of DC MG is presented in [19], where a good comparison of control methods are added in the end of the literature. A very good review on MG primary control strategies happens so far in [20] where controllers are divided on the basis of using communication and no-communication in the local DG units. An excellent, detailed and comprehensive review on power sharing strategies of islanded AC-MG, specially on modified droop control schemes are discussed in [22]. Detailed explanation of MG control trends is presented in [11]. A quick review of overall control schemes are not provided though. In [23], classification of control strategies are categorized clearly in a block diagram table, specially for hybrid MG. But the table contents are not up-to-date.

On the basis of the reviews mentioned above, not any of them covers the entire MG control system. This paper is unique in this way that, it covers the most up-to-date control techniques of entire MG in a brief yet comprehensive way. Anybody who wants to start research on MG control field, he/she will definitely find clear a cut idea, from where to start to where to end. The next section of this paper states about MG hierarchical control. Section III, IV, V and, VI detect and discuss about the four hierarchical MG control levels respectively.

II. HIERARCHICAL CONTROL

Grid-tied or grid connected (GC) and islanded or standalone modes are two basic modes of operation of MG [24]. Regulation and operation of MG in those two modes can be achieved by using local source controller and central controller (CC) [1]. CC control strategies, as reported in [1] are:

- Decentralized and hierarchical control;
- Expert control;
- Optimization based control;

This article presents a brief overview on hierarchical control architecture, based on IEC/ISO 62264 int. std. model, reported in [16]. The five levels of control hierarchy is given in Fig. 1. The top hierarchy, i.e., level 4-market structure and business model, is not discussed here because the control nature of it, is more like main grid rather than MG.

III. INTERNAL LOOP CONTROL

This is also known as inner loop control or low level voltage (V) and current (I) controller and, certain controlling issues are guaranteed here [4], [10], [11], [13], [14], [25]–[28]:

- Responsible for local voltage and current control;
- Import/Export power from/to main power grid;
- Transfer of energy from DERs (controllable or uncontrollable) to MG;
- Determine the operating states of DG units and ESS;
- Maintaining stability with high bandwidth performance;

To acquire those objectives, three types of converter topologies are introduced in [25], [26]: grid feeding, grid forming and, grid supporting converter. Next subsections discuss their operation and control strategies in brief.

A. Grid Feeding Converter

Grid feeding converter works in GC mode and collects voltage and frequency (f) set points from the grid. Also known as current source inverter (CSI) which injects current to the host grid and should always be synchronized with the grid [26]. CSI uses two regulating loops: first one is to regulate current injected to the grid and the second one is for setting current reference to deliver proper power to the grid [29]. Different current regulating techniques are invented by many researchers and those can be categorized into mainly two types: linear and non-linear controllers [30], [31], shown in Fig. 2.

B. Grid Forming Converter

Grid forming converters are made to work in standalone condition to adjust the frequency and amplitude of local grid voltage internally, by utilizing two regulating loops: inner current regulating loop and outer voltage regulating loop [32]. Reference and calculated voltages are compared, regulated and then, reference current signal is generated in voltage regulating loop. The current regulating loop just tracks and regulates this current signal. On the basis of this characteristics, voltage source inverters (VSI) are made, which not only have to supply fundamental power but also have to neutralize harmonics and abnormal conditions [33], because, autonomous MG is more prone to disturbances due to lower inertia, more line impedance and, unbalanced dispatch of local loads. Voltage control techniques [32] are shown in Fig. 2. Current control techniques shown in Fig. 2, can also be applicable for grid forming converter.

C. Grid Supporting Converter

Grid supporting power converter can be characterized by the combination of grid feeding and grid forming power converter. It delivers MG loads and adjusts voltage, frequency parameters of host grid with the main grid simultaneously [25], [26].

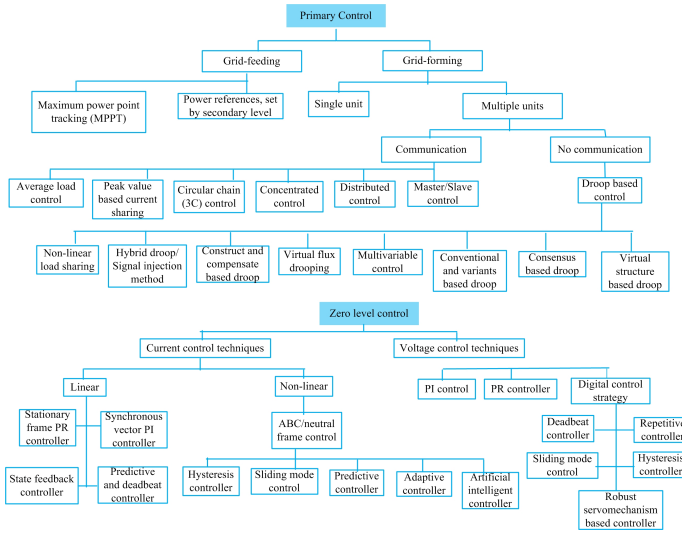


Fig. 2. Primary and zero level controllers

IV. PRIMARY CONTROL LEVEL

Also known as power generation control has following desirable characteristics [11], [15], [20], [25], [34], [35]:

- Reliability ensuring and performing stable operation for zero level internal control;
- Ensuring fast response, islanding detection, output voltage, current and, power sharing control;
- Correcting magnitude and frequency of output voltage to attain references for local voltage and current regulators by ensuring power allocating control;
- Offering plug-n-play capability of DER units;
- Avoiding circulating currents among the parallel connected power electronic converter modules;

A. Primary Control in Autonomous Mode

In autonomous or islanded mode, grid forming or supporting converters can operate as discussed in second section where VSIs are the main controllable interfaces, consist of two main controlling stages [11], [35]: 1) DG power sharing controller (for correcting real (P) and reactive (Q) power mismatches); and, 2) Inverter output controller (output voltage, current control).

1) *Control without Communication*: Here, for proper controlling of power sharing among the parallel connected PE converters, there is no need of any communication link (like, inductor) between them. Droop method is extensively investigated to work in this level of control. This technique imitates the drooping features of synchronous generators [36]. Three droop controlling stages for paralleled inverters are [11], [12], [25], [37]:

- Internal voltage/current regulation loop (section III);
- Intermediate virtual impedance loop to minimize line impedance effect on droop based control;
- External power sharing loop for proper sharing of P and Q among the parallel connected inverters, for preventing

circulating currents without communication and for regulating the V, I references to be supplied in the internal regulating loop using P/f and Q/V droop regulator;

Although, droop method possesses some inbuilt advantages like no-communication, useful for dispersed system, offering flexibility, it suffers from many adverse effects discussed in [10], [11], [14], [25] which limit its application. So, to overcome those shortcomings, several modified and advanced droop techniques emerged which are shown clearly in Fig. 2. These methods [22] are just stated below.

- *Conventional and variants based droop*, including, conventional droop, voltage active power droop/frequency reactive power boost (VPD/FQB) based droop, complex line impedance based droop, angle droop and, voltage based droop control;
- *Multivariable based droop control*;
- *Virtual structure based methods*, including, virtual output impedance loop based, virtual frame transformation based and, enhanced virtual impedance loop based droop;
- *Consensus based droop control*;
- *Construct and compensate based method*, including, adaptive voltage droop, synchronous reactive power compensation based, droop control based synchronous operation, Q-V dot droop, common variable based droop;
- *Hybrid droop/signal injection based droop method*;
- *Virtual flux drooping*;
- *Non-linear load sharing based droop*;

Various hybrid MG systems use low bandwidth communications [38] (CAN, PLC, wireless RF network) for appropriate power allocating goals, overcoming the drawbacks of typical droop methods.

2) *Control with Communication*: Communication based controllers use high bandwidth communication link (LAN/Ethernet, Optical fiber transmission [38]) between the control modules which achieve good voltage regulation and proper power sharing [20], [22]. But for a highly dispersed system, long distance communication lines are expensive, can easily get affected which limits the system reliability and plug-n-play capability. However, some communication based methods are stated here:

- *Concentrated control* [20];
- *Master/Slave (M/S) control method* including auto M/S control strategy [39], control with or without central controller [40];
- *Distributed control method* including distributed networked control [37], distributed two layer control [41], robust control with partial feedback linearization [42];
- *Circular chain (3C) control* [43];
- *Average load control* [44], [45];
- *Peak value based current sharing* [46], [47];

B. Primary Control in Grid Tied Mode

Grid feeding converters mostly operate in grid tied operation. The host grid sets up voltage (V) and frequency (f) of MG system. So, the sole purpose of local controllers in current

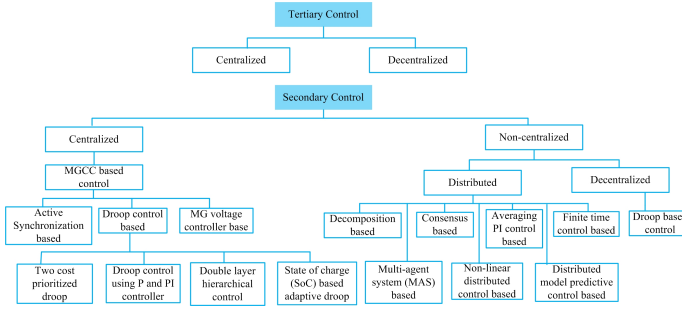


Fig. 3. Tertiary and secondary level controllers

control mode is to extract maximized power from the RESs, constructing P and Q set points by exploiting maximum power point tracking (MPPT) algorithm [23], [26]. Also, P and Q set points can also be delivered from higher (secondary) control level. Then the regulators can operate in non-optimal zone and, can exert all the effort to maximize the power allocating tactic of the network [23], [26].

V. SECONDARY LEVEL CONTROL

This control level, also known as energy management system (EMS), involves PI controller, known as load frequency controller (LFC) or automatic gain controller (AGC), ensures below mentioned control objectives [11], [14], [25]:

- Restoring deviations in voltage amplitude and frequency, from the primary control level;
- Producing output variables which serves as references for the primary control;
- Minimizing total operating cost in stand alone operation and maximization of MG revenue in grid tied mode;
- Finding the optimal unit commitment (UC) and dispatch of the DER units;
- Synchronizing MG with the main grid before converting MG operating mode from autonomous to grid-tied mode;

EMS can be accomplished by using mainly two controlling concepts: centralized and decentralized. Where centralized technique uses high speed communication link whereas decentralized technique does not. A new control technique, called distributed control technique works as the combination of the above two techniques, reported in [14]. A brief explanation is given on those techniques in below subsections.

A. Centralized Secondary Control

Centralized control is apprehended by utilizing MG central controller, involving three fundamental control levels [21]:

- Distribution management system (DMS);
- Local micro-source (MC) and load controller;
- MG central controller (MGCC);

Primary control level incorporates local controllers. MGCC interfaces the DMS and MG, using a limited bandwidth interaction link. MGCC activates online optimization framework, with the generation of relevant information at a singular point instantaneously [11], regulates load in MG by accurately

handling energy stability in the network [21], examines power produced by renewable energy generators and co-ordinates with the load demand and, serves or sheds controllable loads [10]. The major shortcomings of this type of control is that, it restrains the system from plug-n-play ability, particularly suitable for small MG where system arrangement is almost stable [8] and, single point breakdown in MGCC or any communication system causes system wide collapsing. However, centralized control is the most applied scheme in power system. Some centralized scheme based controllers [21] like: active synchronize based controller [48], microgrid voltage controller [49], droop based control techniques [50]–[53] are detected and shown in Fig. 3.

B. Distributed Secondary Control

An excellent work is reported in [14] on distributed control. Latest advancement in networked control scheme, communication systems (Zigbee, WiFi), ICT etc. makes distributed and decentralized control happen for large, flexible smart MG system [8]. In distributed and decentralized control, decision making is performed based on local information, rejecting the need of MGCC or any other high bandwidth communication facilities [54]. Distributed control technique monitors interaction among the units, allocating control duty to various units depending on operation in various time frame [14]. Some recent works based on distributed control scheme, reported in [11], [14], [21], [54] are presented (Fig. 3) here:

- *Multi agent system (MAS) based* techniques, including, combination of MAS with cooperative control [55], distributed cooperative control [56], hierarchical MAS based energy management system [57] and, agent based control of islanded microgrid [58];
- *Consensus based control* [59]–[61];
- *Distributed model predictive control* [62];
- *Decomposition based technique* [63];
- *Averaging PI control* [64];
- *Finite time control* [65];
- *Non-linear distributed control* [66];

C. Decentralized Secondary Control

There is no necessity of having any communication facility among the subsystems in decentralized controlling. Decentralized technique does not consider any contact or communication framework among DG units [54], which is the primary dissimilarity between distributed and decentralized controlling. This feature of decentralization presents poor system wide performance [14]. The only controller falls in this category is droop controller, discussed previously.

VI. TERTIARY LEVEL CONTROL

The final controlling stage of MG is tertiary or global control level. It mainly acts when MG operates in grid-tied mode; remains inactive in autonomous operation. Some features and control objectives [4], [8], [10]–[13], [15], [16], [25], [26], [38] of tertiary control are presented below:

- Facilitating energy generation and market policy by utilizing market operator (MO) and distribution network operator (DNO) - where MO acts on market and DNO connects with different MGs;
- Regulating undisturbed, high quality power flow between main grid and MG;
- Utilizing transmission system operator (TSO) and MG operating manager (MOM), set points of f and V for secondary control level are produced;
- Constituting a path between host grid and MG or additional MG sub-systems and, predicting storage capabilities and instantaneous load changes;
- Non-planned islanding detection;
- Overall supervision of reactive power;
- Offering best economical performance by utilizing game theory and gossiping based methods;

Generally, secondary and tertiary controllers are applied interchangeably in grid supporting/feeding converter to regulate MG reference positions [26]. At the instant of unexpected faults, tertiary controller possesses active power of host grid as nonexistent, which results in frequency deviation. While passing a preset value, the MG will lose its contact. Islanding finding technique works as a vital factor in this context.

VII. CONCLUSION

During the last two decades, control of inner loop converters to use in industry level applications makes significant progress. In primary control level, communication based controlling shows speedy dynamics control, superior power sharing and V/f quality but the system flexibility is lost and communication lines get easily interrupted. Wireless communication based droop control is promising and trending in this situation. Though it suffers from some drawbacks, continuous efforts are being made, which produce diverse droop based controlling techniques mentioned earlier. A compromise between centralized and decentralized based scheme is introduced in secondary control hierarchy and distributed controlling based research field is emerged. It offers promising solutions in optimal power flow and economic load dispatch problem ranging from single MG to MG with multiple units. This paper discusses all the control techniques of a whole MG system, emerged so far in a concise way. Hierarchical control level and the sub levels: tertiary, secondary, primary, local control are presented. Controllers operating in those sub-levels, ranging from smart to conventional are also detected and summarized in block diagrams.

REFERENCES

- [1] H. A. Rahman, M. S. Majid, A. R. Jordehi, G. C. Kim, M. Y. Hassan, and S. O. Fadhl, "Operation and control strategies of integrated distributed energy resources: A review," *Renew. Sust. Energy Rev.*, vol. 51, pp. 1412 – 1420, 2015.
- [2] P. Basak, S. Chowdhury, S. H. nee Dey, and S. Chowdhury, "A literature review on integration of distributed energy resources in the perspective of control, protection and stability of microgrid," *Renew. Sust. Energy Rev.*, vol. 16, no. 8, pp. 5545 – 5556, 2012.
- [3] A. Werth, N. Kitamura, and K. Tanaka, "Conceptual study for open energy systems: Distributed energy network using interconnected dc nanogrids," *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 1621–1630, July 2015.
- [4] J. J. Justo, F. Mwasilu, J. Lee, and J.-W. Jung, "Ac-microgrids versus dc-microgrids with distributed energy resources: A review," *Renew. Sust. Energy Rev.*, vol. 24, pp. 387 – 405, 2013.
- [5] N. Hatziaargyriou, H. Asano, R. Iravani, and C. Marnay, "Microgrids," *IEEE Power Energy Mag.*, vol. 5, no. 4, pp. 78–94, July 2007.
- [6] C. H. Lo and N. Ansari, "Decentralized controls and communications for autonomous distribution networks in smart grid," *IEEE Trans. Smart Grid*, vol. 4, no. 1, pp. 66–77, March 2013.
- [7] N. Lidula and A. Rajapakse, "Microgrids research: A review of experimental microgrids and test systems," *Renew. Sust. Energy Rev.*, vol. 15, no. 1, pp. 186 – 202, 2011.
- [8] L. Meng, E. R. Sanseverino, A. Luna, T. Dragicevic, J. C. Vasquez, and J. M. Guerrero, "Microgrid supervisory controllers and energy management systems: A literature review," *Renew. Sust. Energy Rev.*, vol. 60, pp. 1263 – 1273, 2016.
- [9] I. Patrao, E. Figueres, G. Garcer, and R. Gonzalez-Medina, "Microgrid architectures for low voltage distributed generation," *Renew. Sust. Energy Rev.*, vol. 43, pp. 415 – 424, 2015.
- [10] E. Planas, A. G. de Muro, J. Andreu, I. Kortabarria, and I. M. de Alegria, "General aspects, hierarchical controls and droop methods in microgrids: A review," *Renew. Sust. Energy Rev.*, vol. 17, pp. 147 – 159, 2013.
- [11] D. E. Olivares, A. Mehrizi-Sani, A. H. Etemadi, C. A. Caizares, R. Iravani, M. Kazerani, A. H. Hajimiragha, O. Gomis-Bellmunt, M. Saeedifard, R. Palma-Behnke, G. A. Jimenez-Estvez, and N. D. Hatziaargyriou, "Trends in microgrid control," *IEEE Trans. Smart Grid*, vol. 5, no. 4, pp. 1905–1919, July 2014.
- [12] J. M. Guerrero, J. C. Vsquez, and R. Teodorescu, "Control of droop-controlled dc and ac microgrids-a general approach towards standardization," in *Ind. Electron., 2009. IECON '09. 35th Annual Conf. of IEEE*, Nov 2009, pp. 4305–4310.
- [13] J. M. Guerrero, M. Chandorkar, T. L. Lee, and P. C. Loh, "Advanced control architectures for intelligent microgrids-part i: Decentralized and hierarchical control," *IEEE Trans. Ind. Electron.*, vol. 60, no. 4, pp. 1254–1262, April 2013.
- [14] M. Yazdani and A. Mehrizi-Sani, "Distributed control techniques in microgrids," *IEEE Trans. Smart Grid*, vol. 5, no. 6, pp. 2901–2909, Nov 2014.
- [15] A. Bidram and A. Davoudi, "Hierarchical structure of microgrids control system," *IEEE Trans. Smart Grid*, vol. 3, no. 4, pp. 1963–1976, Dec 2012.
- [16] O. Palizban, K. Kauhaniemi, and J. M. Guerrero, "Microgrids in active network management—part I: Hierarchical control, energy storage, virtual power plants, and market participation," *Renew. Sust. Energy Rev.*, vol. 36, pp. 428 – 439, 2014.
- [17] F. Nejabatkah and Y. W. Li, "Overview of power management strategies of hybrid ac and dc microgrid," *IEEE Trans. Power Electron.*, vol. 30, no. 12, pp. 7072–7089, Dec 2015.
- [18] A. Mohd, E. Ortjohann, D. Morton, and O. Omari, "Review of control techniques for inverters parallel operation," *Electric Power Systems Research*, vol. 80, no. 12, pp. 1477 – 1487, 2010.
- [19] C. Papadimitriou, E. Zountouridou, and N. Hatziaargyriou, "Review of hierarchical control in dc microgrids," *Electric Power Systems Research*, vol. 122, pp. 159 – 167, 2015.
- [20] T. Vandoorn, J. D. Kooning, B. Meersman, and L. Vandevelde, "Review of primary control strategies for islanded microgrids with power-electronic interfaces," *Renew. Sust. Energy Rev.*, vol. 19, pp. 613 – 628, 2013.
- [21] A. Kaur, J. Kaushal, and P. Basak, "A review on microgrid central controller," *Renew. Sust. Energy Rev.*, vol. 55, pp. 338 – 345, 2016.
- [22] H. Han, X. Hou, J. Yang, J. Wu, M. Su, and J. M. Guerrero, "Review of power sharing control strategies for islanding operation of ac microgrids," *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 200–215, Jan 2016.
- [23] E. Unamuno and J. A. Barrena, "Hybrid ac/dc microgrids—part II: Review and classification of control strategies," *Renew. Sust. Energy Rev.*, vol. 52, pp. 1123 – 1134, 2015.
- [24] F. Katiraei, M. R. Iravani, and P. W. Lehn, "Micro-grid autonomous operation during and subsequent to islanding process," *IEEE Trans. Power Del.*, vol. 20, no. 1, pp. 248–257, Jan 2005.
- [25] A. M. Bouzid, J. M. Guerrero, A. Cheriti, M. Bouhamida, P. Sicard, and M. Benghanem, "A survey on control of electric power distributed

- generation systems for microgrid applications," *Renew. Sust. Energy Rev.*, vol. 44, pp. 751–766, 2015.
- [26] J. Rocabert, A. Luna, F. Blaabjerg, and P. Rodriguez, "Control of power converters in ac microgrids," *IEEE Trans. Power Electron.*, vol. 27, no. 11, pp. 4734–4749, Nov 2012.
 - [27] M. R. Miveh, M. F. Rahmat, A. A. Ghadimi, and M. W. Mustafa, "Control techniques for three-phase four-leg voltage source inverters in autonomous microgrids: A review," *Renew. Sust. Energy Rev.*, vol. 54, pp. 1592–1610, 2016.
 - [28] J. M. Guerrero, L. Hang, and J. Uceda, "Control of distributed uninterruptible power supply systems," *IEEE Trans. Ind. Electron.*, vol. 55, no. 8, pp. 2845–2859, Aug 2008.
 - [29] A. Timbus, M. Liserre, R. Teodorescu, P. Rodriguez, and F. Blaabjerg, "Evaluation of current controllers for distributed power generation systems," *IEEE Trans. Power Electron.*, vol. 24, no. 3, pp. 654–664, March 2009.
 - [30] M. P. Kazmierkowski and L. Malesani, "Current control techniques for three-phase voltage-source pwm converters: a survey," *IEEE Trans. Ind. Electron.*, vol. 45, no. 5, pp. 691–703, Oct 1998.
 - [31] G. Shahgholian, J. Faiz, and P. Shafaghi, "Nonlinear control techniques in uninterruptible power supply inverter: A review," in *Computer and Electrical Engineering, 2009. ICCEE '09. Second International Conference on*, vol. 1, Dec 2009, pp. 51–55.
 - [32] M. R. Miveh, M. F. Rahmat, A. A. Ghadimi, and M. W. Mustafa, "Power quality improvement in autonomous microgrids using multi-functional voltage source inverters: A comprehensive review," *Journal of Power Electronics*, vol. 15, no. 4, pp. 1054–1065, 2015.
 - [33] J. M. Guerrero, P. C. Loh, T. L. Lee, and M. Chandorkar, "Advanced control architectures for intelligent microgrids—part II: Power quality, energy storage, and ac/dc microgrids," *IEEE Trans. Ind. Electron.*, vol. 60, no. 4, pp. 1263–1270, April 2013.
 - [34] A. Mehrizi-Sani and R. Iravani, "Potential-function based control of a microgrid in islanded and grid-connected modes," *IEEE Trans. Power Syst.*, vol. 25, no. 4, pp. 1883–1891, Nov 2010.
 - [35] J. A. P. Lopes, C. L. Moreira, and A. G. Madureira, "Defining control strategies for microgrids islanded operation," *IEEE Trans. Power Syst.*, vol. 21, no. 2, pp. 916–924, May 2006.
 - [36] Q. C. Zhong and G. Weiss, "Synchronverters: Inverters that mimic synchronous generators," *IEEE Trans. Ind. Electron.*, vol. 58, no. 4, pp. 1259–1267, April 2011.
 - [37] Q. Shafiee, J. M. Guerrero, and J. C. Vasquez, "Distributed secondary control for islanded microgrids — a novel approach," *IEEE Trans. Power Electron.*, vol. 29, no. 2, pp. 1018–1031, Feb 2014.
 - [38] X. Lu, J. M. Guerrero, K. Sun, J. C. Vasquez, R. Teodorescu, and L. Huang, "Hierarchical control of parallel ac-dc converter interfaces for hybrid microgrids," *IEEE Trans. Smart Grid*, vol. 5, no. 2, pp. 683–692, March 2014.
 - [39] Y. Pei, G. Jiang, X. Yang, and Z. Wang, "Auto-master-slave control technique of parallel inverters in distributed ac power systems and ups," in *Power Electronics Specialists Conference, 2004. PESC 04. 2004 IEEE 35th Annual*, vol. 3, June 2004, pp. 2050–2053 Vol.3.
 - [40] R. Ramos, D. Biel, F. Guinjoan, and E. Fossas, "Master-slave sliding-mode control design in parallel-connected inverters," *AUTOMATIKA*, vol. 42, no. 1-2, pp. 37–44, 2001.
 - [41] A. Bidram, A. Davoudi, and F. L. Lewis, "A multiobjective distributed control framework for islanded ac microgrids," *IEEE Trans. Ind. Informatics*, vol. 10, no. 3, pp. 1785–1798, Aug 2014.
 - [42] M. A. Mahmud, M. J. Hossain, H. R. Pota, and A. M. T. Oo, "Robust nonlinear distributed controller design for active and reactive power sharing in islanded microgrids," *IEEE Trans. Energy Convers.*, vol. 29, no. 4, pp. 893–903, Dec 2014.
 - [43] T.-F. Wu, Y.-K. Chen, and Y.-H. Huang, "3c strategy for inverters in parallel operation achieving an equal current distribution," *IEEE Trans. Ind. Electron.*, vol. 47, no. 2, pp. 273–281, Apr 2000.
 - [44] A. M. Roslan, K. H. Ahmed, S. J. Finney, and B. W. Williams, "Improved instantaneous average current-sharing control scheme for parallel-connected inverter considering line impedance impact in microgrid networks," *IEEE Trans. Power Electron.*, vol. 26, no. 3, pp. 702–716, March 2011.
 - [45] X. Sun, Y.-S. Lee, and D. Xu, "Modeling, analysis, and implementation of parallel multi-inverter systems with instantaneous average-current-sharing scheme," *IEEE Trans. Power Electron.*, vol. 18, no. 3, pp. 844–856, May 2003.
 - [46] Q. Chen, P. Ju, K. Shi, Y. Tang, Z. Shao, and W. Yang, "Parameter estimation and comparison of the load models with considering distribution network directly or indirectly," *Int. Journal of Electrical Power & Energy Syst.*, vol. 32, no. 9, pp. 965–968, 2010.
 - [47] C. L. Chen, Y. Wang, J. S. Lai, Y. S. Lee, and D. Martin, "Design of parallel inverters for smooth mode transfer microgrid applications," *IEEE Trans. Power Electron.*, vol. 25, no. 1, pp. 6–15, Jan 2010.
 - [48] C. Cho, J. H. Jeon, J. Y. Kim, S. Kwon, K. Park, and S. Kim, "Active synchronizing control of a microgrid," *IEEE Trans. Power Electron.*, vol. 26, no. 12, pp. 3707–3719, Dec 2011.
 - [49] M. A. Tabrizi, G. Radman, and A. Tamers, "Micro grid voltage profile improvement using micro grid voltage controller," in *Southeastcon, 2012 Proc. of IEEE*, March 2012, pp. 1–6.
 - [50] I. U. Nutkani, P. C. Loh, P. Wang, and F. Blaabjerg, "Cost-prioritized droop schemes for autonomous ac microgrids," *IEEE Trans. Power Electron.*, vol. 30, no. 2, pp. 1109–1119, Feb 2015.
 - [51] L. Xu and D. Chen, "Control and operation of a dc microgrid with variable generation and energy storage," *IEEE Trans. Power Del.*, vol. 26, no. 4, pp. 2513–2522, Oct 2011.
 - [52] T. Dragievi, J. M. Guerrero, J. C. Vasquez, and D. Krlec, "Supervisory control of an adaptive-droop regulated dc microgrid with battery management capability," *IEEE Trans. Power Electron.*, vol. 29, no. 2, pp. 695–706, Feb 2014.
 - [53] X. Lu, K. Sun, J. M. Guerrero, J. C. Vasquez, and L. Huang, "State-of-charge balance using adaptive droop control for distributed energy storage systems in dc microgrid applications," *IEEE Trans. Ind. Electron.*, vol. 61, no. 6, pp. 2804–2815, June 2014.
 - [54] F. Chen, M. Chen, Q. Li, K. Meng, J. Guerrero, and D. Abbott, "Multiagent based reactive power sharing and control model for islanded microgrids," *IEEE Trans. Sust. Energy*, vol. PP, no. 99, pp. 1–1, 2016.
 - [55] A. Bidram, A. Davoudi, F. L. Lewis, and J. M. Guerrero, "Distributed cooperative secondary control of microgrids using feedback linearization," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 3462–3470, Aug 2013.
 - [56] T. Morstyn, B. Hredzak, and V. G. Agelidis, "Distributed cooperative control of microgrid storage," *IEEE Trans. Power Syst.*, vol. 30, no. 5, pp. 2780–2789, Sept 2015.
 - [57] M. Mao, P. Jin, N. D. Hatziaargyriou, and L. Chang, "Multiagent-based hybrid energy management system for microgrids," *IEEE Trans. Sust. Energy*, vol. 5, no. 3, pp. 938–946, July 2014.
 - [58] Q. Li, F. Chen, M. Chen, J. M. Guerrero, and D. Abbott, "Agent-based decentralized control method for islanded microgrids," *IEEE Trans. Smart Grid*, vol. 7, no. 2, pp. 637–649, March 2016.
 - [59] J. Schiffer, T. Seel, J. Raisch, and T. Sezi, "Voltage stability and reactive power sharing in inverter-based microgrids with consensus-based distributed voltage control," *IEEE Trans. Control Syst. Technol.*, vol. 24, no. 1, pp. 96–109, Jan 2016.
 - [60] L. Y. Lu and C. C. Chu, "Consensus-based droop control synthesis for multiple dics in isolated micro-grids," *IEEE Trans. Power Syst.*, vol. 30, no. 5, pp. 2243–2256, Sept 2015.
 - [61] H. Zhang, S. Kim, Q. Sun, and J. Zhou, "Distributed adaptive virtual impedance control for accurate reactive power sharing based on consensus control in microgrids," *IEEE Trans. Smart Grid*, vol. PP, no. 99, pp. 1–13, 2016.
 - [62] M. Moradzadeh, R. Boel, and L. Vandevelde, "Voltage coordination in multi-area power systems via distributed model predictive control," *IEEE Trans. Power Syst.*, vol. 28, no. 1, pp. 513–521, Feb 2013.
 - [63] A. G. Beccuti, T. H. Demiray, G. Andersson, and M. Morari, "A lagrangian decomposition algorithm for optimal emergency voltage control," *IEEE Trans. Power Syst.*, vol. 25, no. 4, pp. 1769–1779, Nov 2010.
 - [64] J. W. Simpson-Porco, Q. Shafiee, F. Drfler, J. C. Vasquez, J. M. Guerrero, and F. Bullo, "Secondary frequency and voltage control of islanded microgrids via distributed averaging," *IEEE Trans. Ind. Electron.*, vol. 62, no. 11, pp. 7025–7038, Nov 2015.
 - [65] F. Guo, C. Wen, J. Mao, and Y. D. Song, "Distributed secondary voltage and frequency restoration control of droop-controlled inverter-based microgrids," *IEEE Trans. Ind. Electron.*, vol. 62, no. 7, pp. 4355–4364, July 2015.
 - [66] M. A. Mahmud, M. J. Hossain, H. R. Pota, and A. M. T. Oo, "Robust nonlinear distributed controller design for active and reactive power sharing in islanded microgrids," *IEEE Trans. Energy Convers.*, vol. 29, no. 4, pp. 893–903, Dec 2014.

Performance Limitations of Optical Transmission System Due to the Combined Effects of Chirping, GVD and SPM

Kanis Fatima¹ and Md. Saiful Islam²

^{1,2}Institute of Information and Communication Technology
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
Email: kanis.eee@diu.edu.bd

Abstract—Due to the technological enhancement of fiber-optic communication, for large data rate, intense optical power, increasing link length and diverse channel in the wavelength division multiplexing (WDM), the simultaneous effect of chirping, group velocity dispersion (GVD) and self-phase modulation (SPM) should take into account in the optical transmission system. SPM introduces nonlinear refractive index that depends on intensity and GVD results for the refractive index that relies on wavelength. Both SPM and GVD impose limitations on the propagating optical pulse when chirped Gaussian pulse transmits into the fiber. To investigate the simultaneous effect of SPM and GVD with chirping, numerical analysis and simulation is carried out by split step Fourier method based on nonlinear Schrodinger equation in 3-D environment using different fibers. Results show that large effective area fiber (LEAF) gives better performance than standard single mode fiber (SSMF) or nonzero dispersion shifted fiber (NZDSF) for same bit rate, power, link length and chirping factor.

Keywords—SPM, GVD, Chirping, NLSE, SSMF, LEAF, NZDSF.

I. INTRODUCTION

For long distance, high bandwidth, large capacity, intense optical power communication system operating at speed about gigabit per second or higher, optical transmission is an advance medium. This high speed optical fiber transmission system face an interruption in the form of GVD that limits the rates in optical transmission systems [1-2]. On the other hand, for long distance data transmission in optical fiber, the fiber must carry high input power. And at high input power, nonlinear effects are commenced in the optical fiber that in turn limits the system performance [3]. SPM is one of the relevant nonlinear effects that influence the system performance greatly [4-5]. When the Kerr non-linearity are the result of the pulses own intensity then the SPM effects observed. The SPM produces spectral changes in the pulse shape with the simultaneous effect of chirping and GVD in high speed optical transmission systems [6-9]. It is possible to compensate the dispersion using different technologies in optical channel [10-11]. Hamza *et al.* (2014) evaluated performance behavior of chirped and un-chirped pulses for

single mode optical fiber [12]. Pre-chirping can be used to inhibit waveform distortion caused by SPM–GVD effect [13]. Soliton propagation due to interaction between SPM and GVD exhibits in the anomalous regime with LEAF [14]. Most of the analysis is usually performed considering SPM and GVD alone. So, it is essential to analyze the combined effect of chirping, SPM and GVD for different fibers to interpret the performance limitations in high speed transmission systems. In this paper, we have investigated the effect of SPM in presence of chirping and GVD in WDM optical transmission system using 3-different types of fiber such as standard single mode fiber, large effective area fiber, and nonzero dispersion shifted fiber for two different dispersion regimes. The main goal of this paper is to analyze graphically the simultaneous effects of chirping, SPM and GVD on the propagating pulses and power for different fibers. The initial pulse distribution is taken as a chirped Gaussian pulse shape. Due to computational complexity, we could not solve the nonlinear Schrodinger equation (NLSE) analytically in a close form. In this work, NLSE has been solved numerically by split step Fourier technique [15] and the combined effect of chirping, SPM and GVD has been visualized for different fibers of optical transmission system through varying various parameters.

II. THEORITICAL BACKGROUND

In the slowly varying envelope theory, the pulse propagation along the transmission system is governed by the NLSE. The NLSE which takes into account both dispersion and nonlinearity can be written as [1].

$$\frac{\partial A}{\partial z} + \frac{i}{2}\beta_2(z) \frac{\partial^2 A}{\partial T^2} - \frac{1}{6}\beta_3 \frac{\partial^3 A}{\partial T^3} + \frac{\alpha(z)}{2}A = i\gamma|A|^2A \quad (1)$$

Where $A(z,t)$ is the time retarded slowly varying complex amplitude of the field, $i = \sqrt{-1}$, α is the attenuation constant, $\gamma = \frac{n_2\omega_0}{cA_{eff}}$ is a nonlinear parameter, ω_0 represents the carrier frequency, c is the velocity of light, n_2 is the nonlinear refractive index and A_{eff} states the effective cross sectional area of the fiber. β_2 and β_3 illustrate the first and second order GVD effects respectively. T is measured in the frame of reference moving with the pulse at the group velocity v_g

($T = t - z/v_g$) and z is the propagation distance. The effect of GVD is included through dispersion parameter that's involved with β_2 as $\beta_2 = \frac{-D\lambda^2}{2\pi C}$ in equation (1), where λ is the operating wavelength and D is the dispersion parameter. If chirped Gaussian input pulse is transmitted into the fiber then for the different chirp parameters, they show interesting behavior. A chirped Gaussian pulse can be mathematically represented by [1].

$$A(0, T) = \exp - (1 + iC) \frac{T^2}{2T_0^2} \quad (2)$$

Where, T_0 is the half width at 1/e intensity point, C illustrates the chirp parameter which may be negative or positive. If we put $C=0$, this equation will convert to a un-chirped Gaussian pulse. The numerical value of C can be accounted from the spectral width of the Gaussian pulse. Unfortunately, NLSE can't be solved analytically. Split-step Fourier method is an efficient technique to solve Eq.1 numerically [15] to find the combined effect of chirping, SPM and GVD. By Split Step Fourier method Eq.1 can be represented in the following form,

$$\frac{\partial A}{\partial z} = (\hat{D} + \hat{N}) A \quad (3)$$

$$\text{Where } \hat{D} = -\frac{i}{2}\beta_2 \frac{\partial^2}{\partial T^2} - \frac{\alpha}{2} \text{ and } \hat{N} = i\gamma|A|^2$$

Where, the operator $\hat{D}$ is used for loss and GVD, whereas the operator $\hat{N}$ includes the nonlinear effect. As a method, divide the optical fiber into small pieces, each of length h meters. The optical pulse is propagated through each segment from z to $z + h$, where z is a running variable for distance along the fiber. In first step, the fiber loss and GVD effects is included using $\hat{D}$ operator over distance h . In the next step, the output obtained in the first step is propagated using $\hat{N}$ operator by the same segment of fiber of length h meters. The process of these two steps can be mathematically represented as [1],

$$A(z+h, T) \approx e^{\hat{D}h} e^{\hat{N}h} A(z, T) \quad (4)$$

Where the factor $e^{\hat{D}h}$ is observed in the Fourier domain using the following mathematical formula,

$$e^{\hat{D}h} B(z, T) = \{F^{-1} e^{\hat{D}h i\omega} F\} B(z, T) \quad (5)$$

Where $e^{\hat{D}h}$ denotes the Fourier-transformation, and Eq.1 is replaced with $(i\omega)$ in accordance of the basic rules of Fourier transformation. Following this procedure, the optical pulse is propagated from one end to the other end. The output obtained at the receiving end can be used to investigate various interesting phenomena in optical fiber transmission. In this paper, we have used the following parameters to find our desired result.

TABLE I. SIMULATION PARAMETERS FOR DIFFERENT FIBER

Parameter(Unit)	Fiber types		
	SSMF	LEAF	NZDSF
Nonlinear Refractive Index, $n_2(\text{m}^2/\text{W})$	2.35×10^{-20}	2.35×10^{-20}	2.35×10^{-20}
Wavelength, λ (nm)	1550	1550	1550
Input Power, $P(\text{mW})$	5	5	5
Bit Rate, $B(\text{Gb/s})$	10	10	10
Dispersion parameter, $D(\text{ps/nm-km})$	17	3.5	5
First order GVD, $\beta_2(\text{ps}^2/\text{km})$	-21.7	-4.46	-6.38
Effective area, $A_{eff}(\mu\text{m}^2)$	80	72	50
Chirp Parameter	Varied	Varied	Varied

III. RESULTS AND DISCUSSIONS

In this section, numerical simulations have been carried out by varying different fiber parameters. Here, we have used 3-types of fibers, such as SSMF, LEAF and NZDSF.

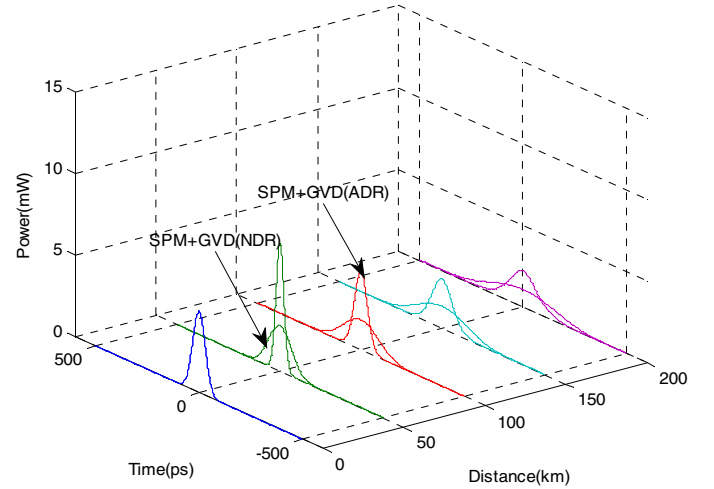


Fig. 1. Combined effect of Chirping, SPM and GVD in SSMF in normal and anomalous dispersion regime

To compare the effects of SPM and GVD among different fibers we have plotted the temporal evolution of the output pulses in Fig.1, Fig.2 and Fig.3 as a 3-D plot showing the pulse power versus time and distance considering chirped Gaussian pulse. The initial pulse width of the Gaussian pulse is 42 ps.

From Fig. 1 it can be observed that for SSMF in the anomalous dispersion regime, the pulse broadening is less compared to the normal dispersion regime. As in the anomalous regime, SPM induces positive chirp while

dispersion induces negative chirp. As a consequence, these two chirps result in a less pulse broadening. There is no pulse compression in SSMF in both regimes. As the pulse propagates through the transmission medium, the output pulse power degrades in both regimes. At propagation distance $z = 200$ km, the power of the output pulse in the anomalous and normal regime is 1.82 mW and 1.306 mW respectively. So, in the normal dispersion regime less power is obtained than anomalous dispersion regime for the same distance.

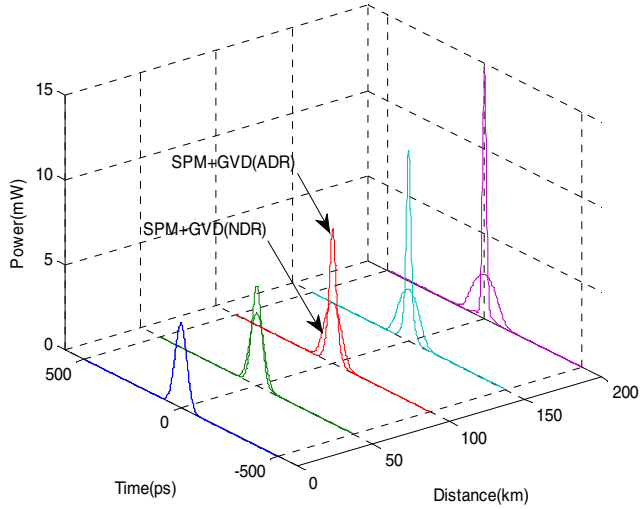


Fig. 2. Combined effect of Chirping, SPM and GVD in LEAF in normal and anomalous dispersion regime

From Fig. 2 the combined effect of SPM and GVD with chirping can be observed for LEAF. In the anomalous dispersion regime with the increase of propagation distance the pulse power increases but with pulse compression.

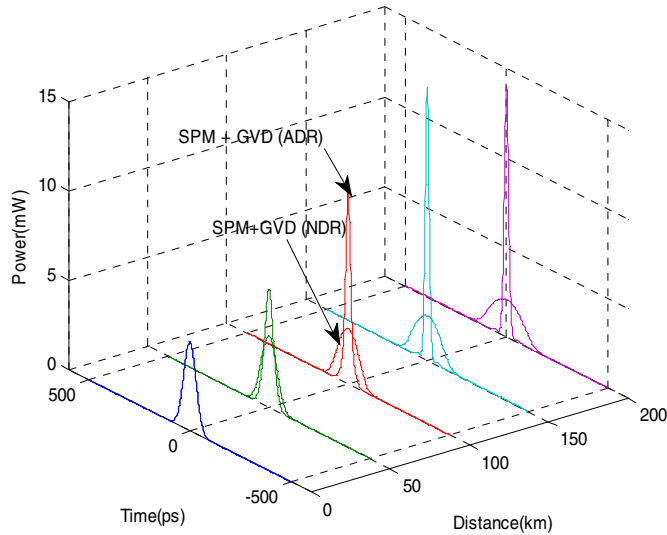


Fig. 3. Combined effect of chirping, SPM and GVD in NZDSF in the normal and anomalous dispersion regime.

At $z = 200$ km, the output pulse power is high approximately 15.3 mW. In the normal dispersion regime with the increase of

propagation distance the output pulse broadens by lowering its peak power. At 200 km, the power is very low approximately 2.8 mW. So, it is observed that LEAF shows better performance in the anomalous dispersion regime than the normal dispersion regime.

It is analyzed from Fig. 3 that the output pulse power is more in NZDSF than LEAF, almost 10.8 mW in anomalous regime at $z = 100$ km. On the other hand, in LEAF the power is approximately 7.92 mW at 100 km in the anomalous regime. So, it is observed that NZDSF can be used to transmit signal in short distances. The output pulse propagation characteristics is almost same in both fibers except at $z = 200$ km at the anomalous dispersion regime, the power is approximately 14.9 mW and in the normal dispersion regime it is 2.5 mW which is more than LEAF fiber.

The chirping effect in 3-D environment for different fibers is shown in Fig.4, Fig.5 and Fig.6. In Fig.4 for the increasing positive chirp parameter the output pulse compresses at 50 km but after that distance the output pulse broadens rapidly. So, the output pulse power degrades drastically for increasing positive chirping effect with SPM and GVD in SSMF. On the other hand, it can be observed that the broadening rate is more severe for negative chirping than positive chirping and there is no pulse compression. At $z = 200$ km the output pulse power is very low compared to the input power for negative chirping.

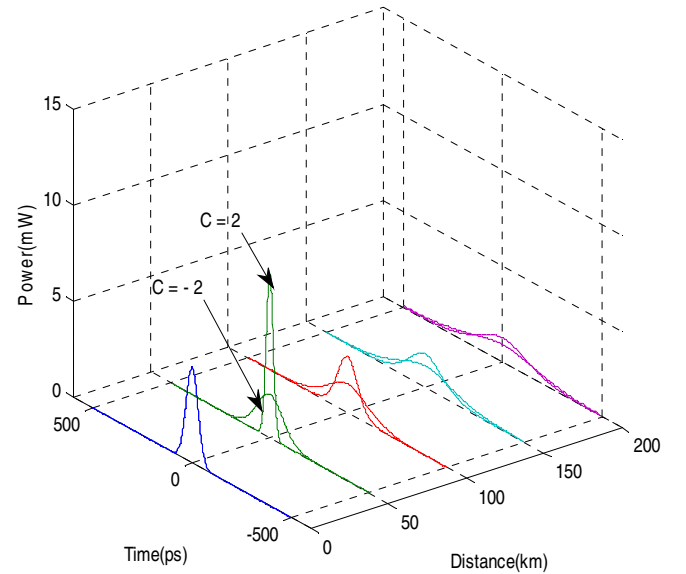


Fig. 4. Positive ($C = +2$) and negative ($C = -2$) chirping effect on SSMF in anomalous dispersion regime

From Fig. 5, it is found that for LEAF fiber for positive chirping, the output pulse gets compressed up to a distance of 200 km. The peak output power increases up to a distance of 150 km and then decreases in case of positive chirping. For

negative chirping the pulse broadens lowering its peak power. Though the pulse broadening for negative chirping is greater in LEAF fiber but still shows better performance than SSMF fiber.

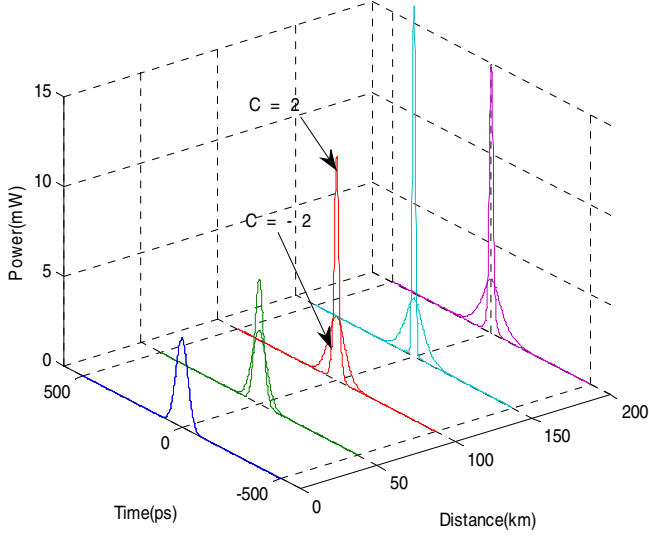


Fig. 5. Positive ($C = +2$) and negative ($C = -2$) chirping effect on LEAF in anomalous dispersion regime

The effect of increasing chirping parameter for NZDSF fiber is shown in Fig.6. Thus for large positive chirp value the output pulse initially compresses and increases peak power up to 100 km, after that the pulse starts broadening with lowering peak power.

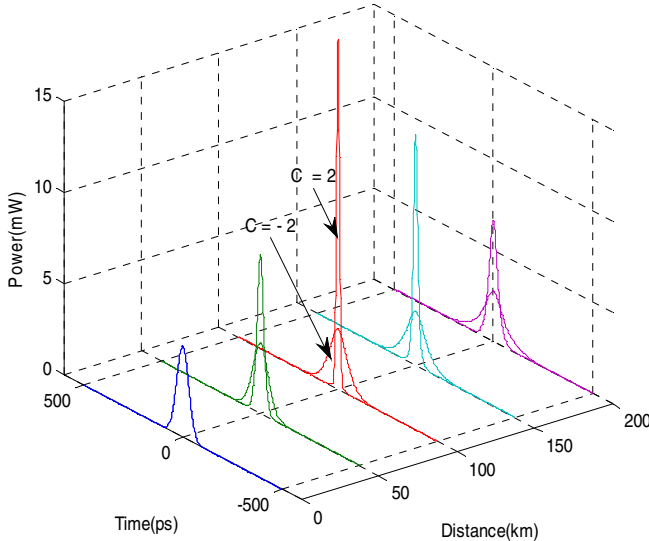


Fig. 6. Positive ($C = +2$) and negative ($C = -2$) chirping effect on NZDSF in anomalous dispersion regime

Again, for negative chirping the pulse broadening rate is high with lowering its peak power. At 200 km, the peak power is approximately 3.2 mW. Thus from this comparison, we can conclude that in LEAF, the combined effect of chirping, SPM and GVD is less compared to SSMF or NZDSF and it can be considered as a good media to transmit pulses at longer distances.

A comparison of pulse degradation behavior in normal and anomalous dispersion regime for the combined effect of SPM and GVD for different fibers is shown in Fig. 7 in 2D environment. It can be observed from Fig. 7 that the pulse degradation is more severe in the normal regime than the anomalous regime for all fibers.

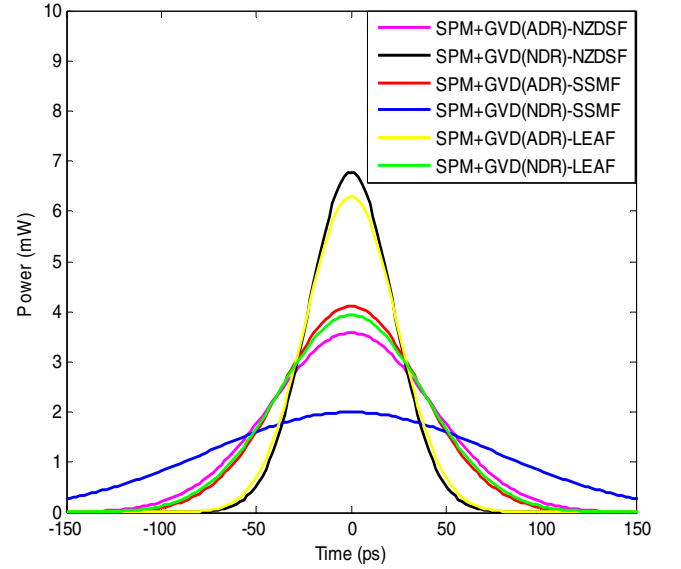


Fig. 7. Pulse degradation behavior in normal versus anomalous regime in different fibers

In LEAF fiber the pulse degradation effect is less than NZDSF and SSMF. In the anomalous regime, the power is approximately 6 mW and in the normal regime the power is approximately 3 mW for LEAF. In NZDSF, the pulse degradation effect is greater than LEAF. As in the anomalous regime, the power is approximately 6.5 mW and in the normal regime the power is approximately 3.0 mW.

IV. CONCLUSION

The combined effect of chirping, GVD and SPM on the propagating pulses in optical transmission system has been investigated for various kinds of fiber. A numerical simulation based on NLSE considering the said effect with the chirped Gaussian pulse has been done by split step Fourier method. The output pulse behavior and power at a certain transmission distance have been calculated and visualized graphically. The results showed that the combined effect of chirping, SPM and GVD are more severe in SSMF than LEAF or NZDSF. It is also found that LEAF can be an effective transmission media

to transmit pulses with lower performance degradation than other fibers. In addition, the negative chirping effect on pulses is more detrimental than positive chirping in all the types of fibers for both normal and anomalous dispersion regimes. These findings will help to design long haul fiber-optic transmission link considering different various impairments. Using the above methods transmission characteristics can also be understood for other fibers as well.

REFERENCES

- [1] G. P. Agrawal, "Nonlinear fiber optics", Academic Press, New York, 2002.
- [2] M. Nishimura, "Optical fibers and fiber dispersion compensators for high speed optical communication", Journal of optical fiber communication report 2, vol. 2, no.2, pp.115-139, 2005.
- [3] Y. K. Karfaa, M. Ismail, F. M. Abbou, S. Shaari, and S. P. Majumder, "Effects of cross-phase modulation crosstalk in WDM networks on received power and number of channels for various fiber types", Proceedings of Asia Pacific conference on applied electromagnetic (APCAE 2007), pp 22-27, December 2007.
- [4] S.C. Pinault and M. J. Potasek, "Frequency broadening by self -phase modulation in optical fibers", Journal of Opt.Soc.Am.B2, vol.2, no.8, pp.1318-1319, 1985.
- [5] S. M. Salakhov, "Evaluation of the effect of self-phase modulation in transmission system with CPFSK", Proceedings of 6th international Siberian workshop and tutorials EDM, pp.177-179, July 2005.
- [6] M. Y. Hamza, S. Farah, S. Yang, and N. Sarwar, "Investigations of un-chirped and chirped Gaussian pulse propagation through single mode fiber in normal versus anomalous dispersion Regimes", Proceedings of 33rd International conference on Telecommunication and Signal Processing (ISP2010), pp. 468-473, August 2010.
- [7] M.J. Islam and M. R. Islam, "Impact of SPM and GVD on the performance of DS-OCDMA", Journal of networks, vol.6, no.1, pp.12-17, 2011.
- [8] J. Maeda and S. Ebisawa, "SPM-GVD effect on radio-over-fiber transmissions using OSSB+C signals", J.Opt.Soc.Am. B.31, vol.31, no.7, pp.14433-50, 2014.
- [9] N. Sultana and M. S. Islam, "Comparative study on the effects of XPM with SPM and GVD and SPM with CD in WDM fiber optic transmission system", Proceedings of 6th International conference on Electrical and Computer Engineering (ICECE), pp.405-409, December 2012.
- [10] O. Rozen, D. Sadat, G. Katz, A. Levey and Mahlab, "Dispersion compensation of self - phase modulation impairment in optical channel using MLSE", Proceedings of ICTOM, pp.178-181, June 2008.
- [11] M. Y. Hamza, S. Tariq and L. Chen, "Dispersion compensation in the presence of nonlinearity in optical fiber communication", Proceedings of 10th IEEE Singapore International Conference on Communication System (ICCS), pp. 1-5, October 2006.
- [12] M. Y. Hamza, S. Saeed, N. Sarwar and S. Yang, "Investigations for the evaluation behavior of Cos-Gauss pulse in dispersion dominant regime of single mode optical fiber", Engineering Letters, vol. 22, no.2, pp.83-86, 2014.
- [13] T. Ito, Y. Yano, T. Ono and, K. Emura, "Pre-chirp assisted normal dispersion region transmission for highly marginal dense WDM transoceanic system", Optical fiber communication conference, USA, pp.367-368, February 1998.
- [14] B. Mani, K. Chitraand, A. Sivasubramanian, "Study of fundamental soliton propagation and their interaction in enhanced-large effective area fiber (E-LEAF)", International conference on Computer communication and systems, pp.4-9, February 2014.
- [15] J. R. Costa, C. R. Paiva, and A. M. Barbosa, "Modified split-step Fourier method for the numerical simulation of soliton simplification in erbium- doped fibers with forward propagation noise", IEEE journal of Quantum Electronics, vol.37, no.1, pp.145-152, 2001.

Developing Demand Side Management Program for Commercial Customers: A Case Study

Md. Aminur Rahman, Riazul Islam, Kazi Fatima Sharif, Tareq Aziz

Dept. of Electrical and Electronic Engineering,
Ahsanullah University of Science and Technology,
Dhaka, Bangladesh.

Email: aminur.rahman.eee@gmail.com, iriazul74@gmail.com, kazi.fatima.sharif@gmail.com, taziz@ieee.org

Abstract—In the new era of smart grid in power system, demand side management (DSM) plays a vital role which helps the customers to make the right decision regarding their power consumption thus minimize the peak load demand and provide the best load profile. Due to DSM carbon emission level and financial cost has been deduced. This paper represents a strategy on load shifting technique for demand side management for commercial customers of future smart grids. A complete methodology for implementing a DSM program has also been outlined. Case study has been carried out taking Ahsanullah University of Science and Technology (AUST), Dhaka as a commercial customer. Results prove that implementation of proposed methodology would provide substantial economic benefit to customer and distributors.

Keywords—Demand Side management, Smart Grid, Load Shifting.

I. INTRODUCTION

Bangladesh is a developing country where electricity demand is on the rise. Rapid development in recent years and scarcity in primary fuel sources make it difficult to keep balance between electricity supply and demand. Currently only 47% of the total population have access to electricity, which is heavily concentrated in urban areas [1]. At present total installed capacity in the power sector, private and public sector combined was around 12.339 MW in April, 2016 [2]. Energy infrastructure of Bangladesh consists of aging generation, transmission, distribution, protection and monitoring devices, which is not sufficient to handle the expected demand for reliable power. In contrary, developed countries are adopting new facilities under the umbrella of “Smart grid” and upgrading their existing network to ensure reliable power supply with maximum efficiency [3]. Communication and control between different nodes in power grid go through wide-scale improvement in the smart grid network. Some primary characteristics of smart grid according to the U.S. Department of Energy include: consumer oriented system, self-healing technology, able to accommodate different types of generation and storage equipment, improved power quality and efficiency [4]. In smart grid scheme, demand side management (DSM) is a component which consists of planning, implementing and monitoring the levels and patterns of electricity consumption [5]. DSM ensures great benefit for both utilities and consumers. For example, in Florida, USA customers are willingly participating in DSM programs and

have saved more than \$1.25 billion since 1981 and their energy usage got reduced by more than 119 million kWh in 2011 alone [6]. Katoomba Demand Management Project in New South Wales, Australia was launched back in late 1990s focusing on energy efficiency in the private sector. This resulted in deferral of pricey investment and ensured efficient network operation [7].

This paper proposes a DSM framework for commercial customers and estimates its impacts on electricity load profile as well as assesses its economic feasibility. The paper is organized as follows: the generalized concept of DSM program is presented in section II. Section III describes the proposed methodology. Ahsanullah University of Science and Technology (AUST) has been taken as a sample case to evaluate the impact of proposed methodology. Results have been presented in section IV along with analysis of the outcome of DSM implementation. At the end, section V summarizes the major findings and produces scope for future work.

II. GENERAL CONCEPT OF DSM

The planning and development of DSM strategy is done primarily by the utility provider. However, other implementers like government organizations and regulatory board, profit and nonprofit groups or a collaboration of different stakeholders may also be involved in the development of DSM program-planning and implementation strategy. Analysis from various data from the different stake holder reflects the current status, performance and operational costs of the system. It then acts as a standard reference point. These statistics help to study and compare the advantages and costs of alternative solutions in commercial DSM program [8].

Different types of demand forecasting such as mid-term or short-term are significant in planning and development of a DSM program. Experiment of load involves collection and analysis of the load data such as system location, customer class, time, magnitude, etc. which are collected from different locations of the distribution system [9]. After forecasting of demand and acquiring enough knowledge about customer load, DSM planner specifically selects and describes a load shape objective.

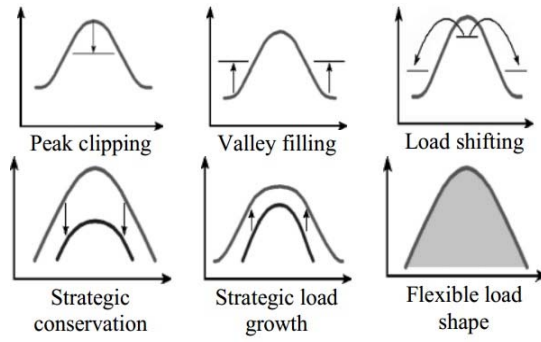


Fig. 1. General Load Shape Objectives [10]

Even though there can be infinite combinations of load shape changing possibilities [8], [11] the most commonly used patterns in globally have been illustrated in Fig. 1. Such patterns include peak clipping, valley filling, load shifting, strategic conservation, strategic load growth and flexible load shape [12]. These are not mutually special and may frequently be engaged in combinations. Peak clipping, valley filling and load shifting techniques are usual forms of load management. DSM program requires large investments and it requires large financing to cover administrative costs and cost sharing investments [13].

The significant benefits of a DSM program are the savings achieved through the number of kWhs per year and reductions achieved in peak demand. This reductions and savings are dependent on the consumer tariff plan (electricity price and demand charges). Feasible benefits include maintenance and labor saving aids related to well-organized technology. DSM program supports in reducing greenhouse gas emissions and provides indirect economic and environmental benefits. It also creates the employment opportunities and consumer savings that are reinvested in other services [14], [15], [16].

The most significant part in the description of demand side alternatives is the selection of perfect market implementation methods. Influencing the customer adoption and approval of demand side management program can be selected from various methods by planners and policy makers [17], [18]. In order to encourage greater customer acceptance of DSM program, customer education is the most important step in the implementation methods available and should be used in combination with one or more other market implementation methods for maximum effectiveness [19]. Out of many different types of customers such as residential, commercial and industrial one, present paper deals only with commercial customers. Proposed methodology of DSM program is illustrated in next section.

III. PROPOSED METHODOLOGY

The proposed DSM methodology is shown in Fig. 2. Obtaining electricity consumption and load data of a customer is the first step for designing a DSM program. The load data can be obtained from various sources such as: meter surveys, electricity bills, data loggers etc. After obtaining the load data,

overall total load curve is plotted for the building, the load curve is analyzed and various demand parameters such as peak load, base load, average load, load factor etc. are determined. Based on the data the end user based load curves are plotted and analyzed. After analyzing the individual end user, the system should read load shifting options.

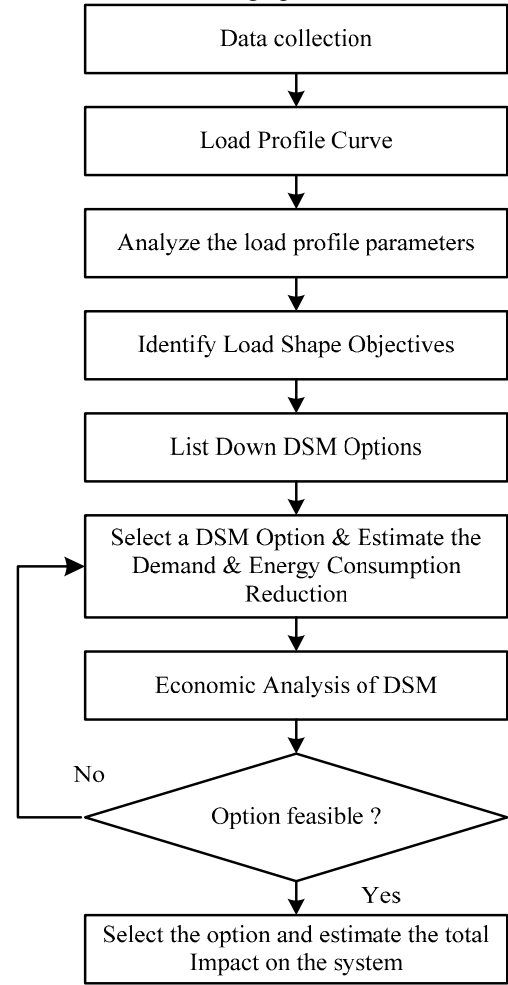


Fig. 2. Proposed DSM algorithm

After analyzing the DSM options, the energy savings and demand reductions arising out of these options are calculated. Further, the tests mandated by the regulators are also performed to assess the economic feasibility of the DSM options. These tests essentially determine the difference of net present value of costs and benefits. This value needs to be positive for the option to be economically viable. Finally, the viable economic options are accumulated to estimate the impacts of DSM on the load profile of the customer.

There can be significant price reduction by using load shifting and load reduction methods. Formulas used to calculate reduction in cost due to load shifting has been given below [20]:

$$\text{Cost Reduction from Load Shifting} = SC * \Delta P \dots \dots \dots (1)$$

$$\text{Cost Reduction from Load reduction} = RC * \Delta P_{avg} \dots \dots (2)$$

Here SC is the Shifted consumption during peak hour and difference between the peak and off-peak electricity price is (ΔP). RC is the reduced cost.

DSM is a wide term which considers efficiency of the load, demand response and dynamic demand. To find out the performance of any electrical equipment efficiency of energy conversion, power factor of the equipment, harmonics generated by the equipment, standby power consumption, peak current has to be taken into account. Also to investigate valley filling, load shifting, peak clipping option for a system, load variation data/trend and its analysis are vital. In case of valley filling the energy storage capacity, efficiency, life cycle need to be considered. Section IV applies the proposed methodology on the test customer i.e. AUST and explains the findings.

IV. CASE STUDY

For analysis purpose Ahsanullah University of Science and Technology (AUST) has been chosen as the subject of study. AUST consumed 1,359,000 kWh and 282,000 kVARh in 2015 which was worth 258,026.9USD. From analysis it can be seen majority of the power consumption is due to various cooling loads. Computers and illumination loads also contributed significantly to the overall consumption.

A. Relevant data for a DSM study

According to regulation created by Bangladesh Power Division, High Voltage users (11KV and up) are charged based on KW and KVAR power consumption. Consumers are required to keep their system power factor at 0.95 or higher [21]. Failing to maintain the power factor close to the value is penalized by additional charge. Power Factor is used by the electricity distribution company to calculate the real and reactive power consumption and the overall electricity bill.

In recent years, a large share of load has been identified as non-linear load [22]. Traditionally majority of the electrical equipment used by the general population were linear loads. Incandescent lights, florescent light with magnetic ballasts, fan with autotransformer for speed control, air conditions, motor are mostly linear load, which means it does not change the current waveform much. But the incandescent lights are replaced by LED/CFL lights with dimming option, air coolers use inverter, TRIAC based choppers are used to control the fan speed, Motors use Variable Frequency Diver (VFD) which optimize the motor efficiency by changing the speed, reducing the inrush current and so on. All the equipment's mentioned later use different power electronic devices and they are nonlinear loads. In order to get an idea of how various non-linear load utilize power along with their reactive power consumption an experimental set up has been established and described in the next few sub sections.

B. Data collection for varieties of loads

The load consumption trend of the university was extracted from energy meter provided by Dhaka Power Distribution Company (DPDC). AUST uses 11KV HV line from DPDC and then the voltage is dropped down to 400V (L-L) for

general consumption. Traditionally by definition power factor can be defined as

$$Pf = \cos(\theta_v - \theta_i) \dots \dots \dots (3)$$

But this definition of power factor is only restricted to linear loads. For nonlinear load the power factor or true power factor can be calculated as:

$$Distortion\ power\ factor(dPF) = \frac{1}{\sqrt{1+(THD_i)^2}} \dots \dots \dots (4)$$

$$Displacement\ Power\ Factor(DPF) = \frac{W}{VA} = \cos(\theta_v - \theta_i) \dots (5)$$

$$True\ Power\ Factor = DPF * dPF \dots \dots \dots (6)$$

Where THD_i is the Total Harmonic Distortion of the current waveform. An important factor to be considered while extracting data is, some old electromechanical power meters are not designed to measure high frequency current harmonic content but they give the average power consumption based on line frequency. A true R.M.S. wattmeter/voltmeter/ammeter need to be used to get the correct result.

For this particular experiment a hall effect sensor with 1A/5A current shunt was used to sense current and differential amplifier was used to measure the line voltage. Both voltage and current data was sampled at a 500MS/s by a data acquisition system and later analyzed in MATLAB to calculate the THD real and reactive power of all the equipment. The experimental procedure can be seen in Fig. 3.

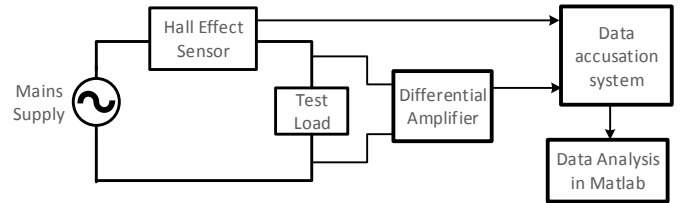


Fig. 3. Experimental setup for data acquisition

The experimental setup can be seen in Fig. 4. Fig. 5 shows the findings for fluorescent tube set. Real power and Reactive power was also calculated from the instantaneous power curve.

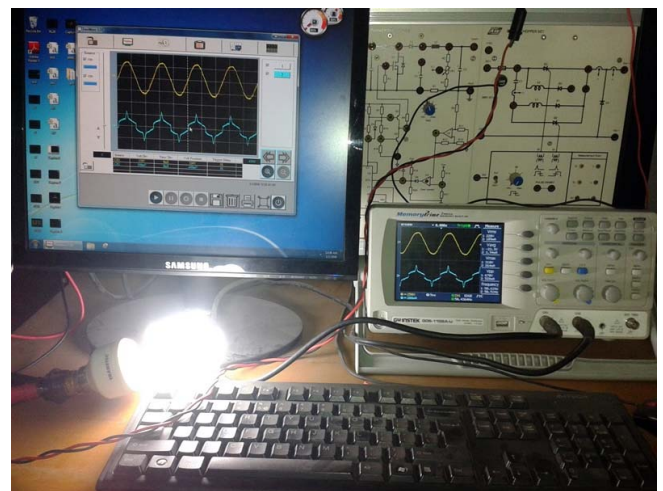


Fig. 4. Photo of a CFL bulb under test

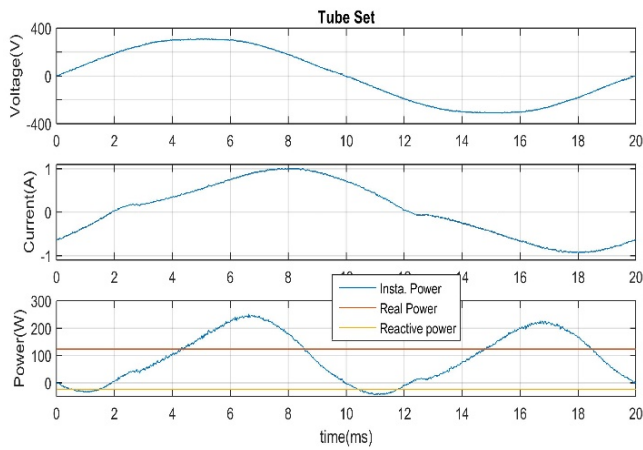


Fig. 5. Voltage, current and power of florescent tube set

Fig. 6 show the different harmonics present in a typical fluorescent tube set. The total harmonic distortion as well as true power factor can be calculated from the graph.

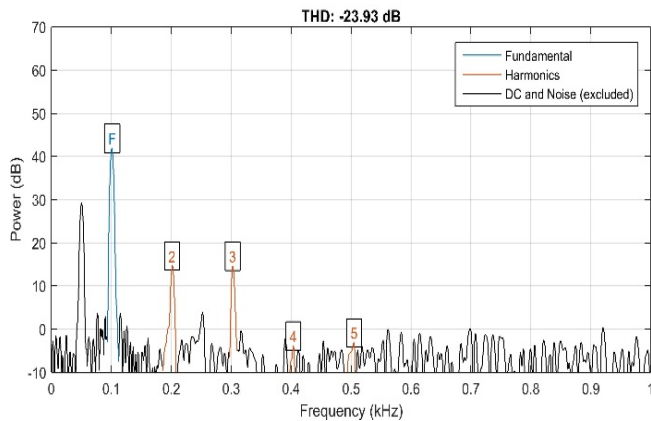


Fig. 6. Harmonic content of a florescent tube

Using the method described earlier, a comparison of power factor as well as power consumption of various types of lights can be seen in Fig. 7 and 8 respectively. For the experimental setup a 100W incandescent bulb, a 14W LED, 23W CFL, and normal 4ft florescent tube was used as they are commonly used in the university.

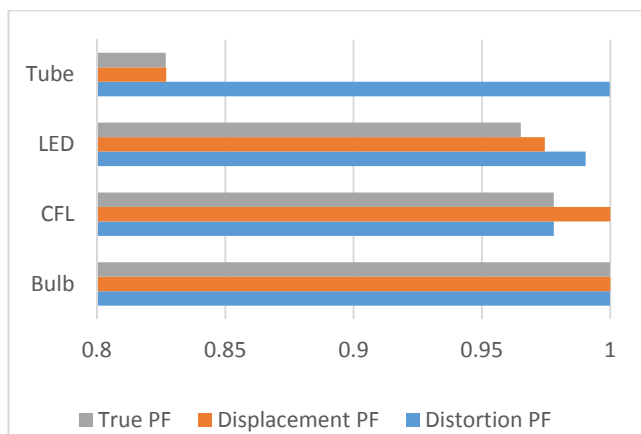


Fig. 7. Power factor comparison of different lights

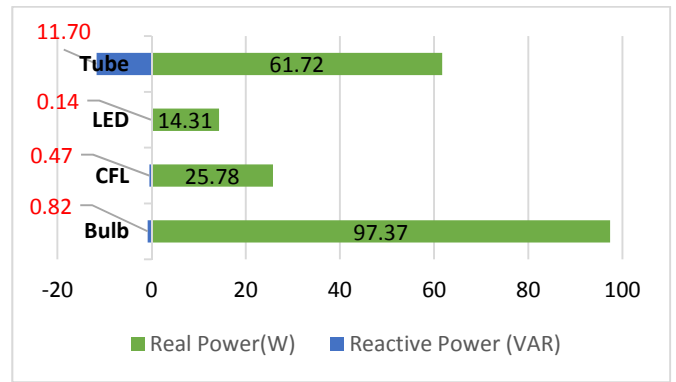


Fig. 8. Power consumption comparison of different lights

Significant amount of electricity wastage in any organization is due to the standby power consumption of various electrical devices such as computer, television, microwave oven and so on. Fig. 9 shows standby power consumption of an Intel Pentium 4 computer with CRT monitor. It was found that the computer consumes 17.55 W of real and 0.7966 VAR of reactive power during the standby condition.

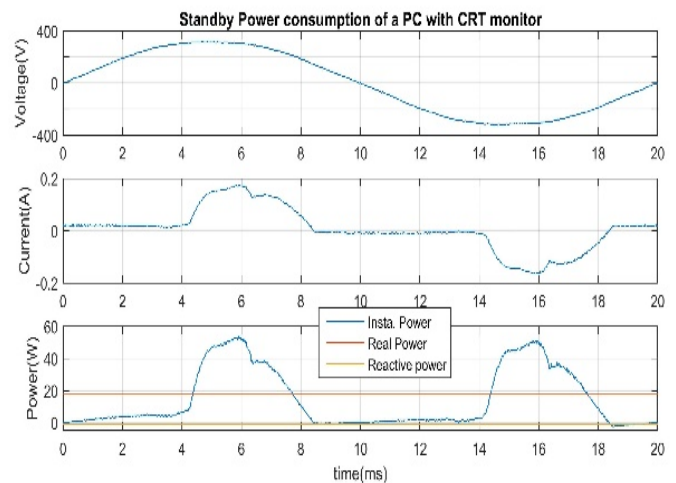


Fig. 9. Standby energy consumption of a PC

C. Collected Bill analysis

DPDC calculates tariff differently during off peak and peak hours [23]. The peak hours are from 5PM up to 11PM and off peak hours are from 11PM to 5PM. In Fig. 10 it can be seen that most of the energy consumed by AUST is during the off peak hours. It is also visible in Fig. 11 that AUST consumes large amount of reactive power. The main sources of reactive powers are air cooler, water pump, fan and the fluorescent tube lights with magnetic ballasts.

From electricity bill/consumption trend it is apparent that there is a direct relationship between the class schedule and ambient temperature. Academic Calendar of spring semester ranged from June, 2015 to September, 2015 which is within the summer months. Hence, the peak consumption and highest bill was found in the month of August, 2015. Classes for fall semester were held from January, 2015 to March, 2015. Because the ambient temperature was relatively low, the

consumption in Fall was less than August, 2015. Though the university activity was on full swing during the month of December, 2015 the energy consumption was low due to the lower ambient temperature in the winter months.

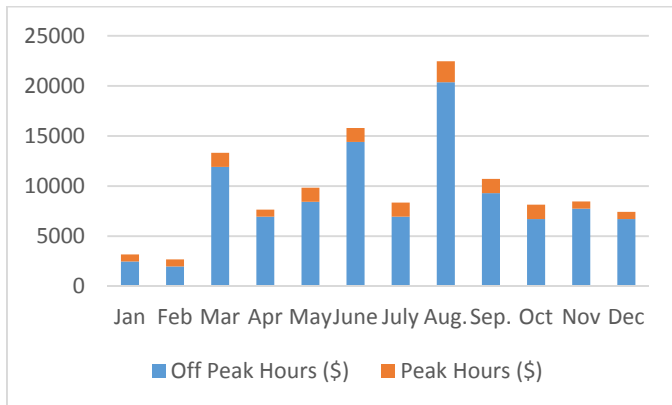


Fig. 10. Electricity bill for the year 2015

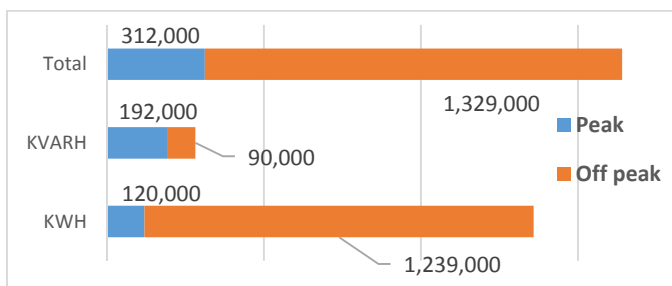


Fig. 11. Real and reactive power consumption by AUST in the year 2015

D. Candidates for DSM and Economic Analysis

In this work, we considered load shifting DSM method on water pump. AUST have five three phase 11KW induction motor coupled with water pump which is used to pump water from reservoir to the tank on the top of the building. While running, each of the pump draws around 11A current from feeder. The water pumps are manually operated. When necessary the pumps are turned on, both during the peak and off peak hours. This make the pumps good candidate for load shifting.

Power consumption analysis of lighting equipment such as fluorescent tubes, incandescent bulb, LED and CFL are candidates for study of efficient electrical equipment. Also standby power consumption reduction is also included in the proposed methodology.

For the water pump/motor it is clear that it should be operated during the off peak hours to achieve the lowest operating cost. The motor array consumes 25kWh energy and the price reduction for shifting load from peak to off-peak hour in DPDC area is 3.3 cent (USD)/KWh [23]. So the cost saving from motor load shifting from peak hour to off peak is =1,227 USD/year.

From the data it can be seen LED lighting have the lowest energy consumption and the true power factor is better than fluorescent tube. Although incandescent bulb has the lowest THD, it has the worst efficiency of all the lighting solutions.

Considering 8 hours of consumption, the cost savings from changing 123-watt tube set to 28 watt LED has been found to be 23,362USD/year. Also the Standby power consumption data states that PC with CRT monitor waste large amount of energy in the standby mood. This can be improved by introducing LCD monitor or by complete turning off the equipment when it is not being used. Cost savings by avoiding stand-by mode operation would save significant amount of energy due to the large number of PC with CRT monitor used in current scenario.

The biggest energy consumer in AUST is the air cooler. From the previous analysis it can be seen during the summer time AUST has the highest energy demand. To minimize the cost, the class schedule can be shifted as such that most of the class are held during relatively cooler time of the year. Also proper maintenance like cleaning the air filters regularly and checking the cooling gas for optimum pressure will improve the performance of the air cooler. Also some passive measures like turning off the air cooler when the rooms are empty, keeping the cooler at 25°C, improving the heat insulation of the building will also increase the efficiency of the coolers and lower the energy consumption.

V. CONCLUSION:

Demand side management has the potential to provide many benefits to the entire smart grid community, particularly at distribution network level. The proposed strategy is a generalized technique based on load shifting and energy efficiency improvement. Results show that along with multiple benefits for generation as well as utility company, this framework would offer economic benefits to various commercial customers. Scaling-up this framework will enable intelligent load management for a city, which should be the objective of any smart or efficient grid system.

REFERENCES

- [1] Electricity scenario in Bangladesh – unnoyononneson. Available at: www.bdresearch.org.bd/home/.../Electricity%20Scenario%20in%20Bangladesh.pdf
- [2] Present Installed Generation Capacity, Available: http://www.bpdb.gov.bd/bpdb/index.php?option=com_content&view=article&id=5&Itemid=6
- [3] Arnab Kumar Paul, Md. Rajib Hossen, Bishnu Sarker, and Monalisa C. Urmi, "An approach to demand side load curtailment for the future intelligent and smart power grid of Bangladesh" In International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT), 2014, pp. 1-7. IEEE, 2014, MIST, Dhaka.
- [4] P. Agrawal, "Overview of DOE microgrid activities," in Proc. SympMicrogrid, Montreal, QC, Canada, 2006 [Online]. Available: http://der.lbl.gov/2006microgrids_files/USA/Presentation_7_Part1_Poonumgrawal.Pdf
- [5] EPRI, "Principles and Practice of Demand Side Management," [Online]. Available: <http://www.epri.com/Pages/Default.aspx>
- [6] Energy efficiency and DSM Programs, Available: <https://www.progress-energy.com/commitment/corporate-responsibility-report/customers/energyefficiency.page>

- [7] Demand management activities applicable to electricity networks, Available at <https://www.ausgrid.com.au/>
- [8] V. S. K. V. Harish and A. Kumar, "Planning and implementation strategy of Demand Side Management in India," Automation, Control, Energy and Systems (ACES), 2014 First International Conference on, Hooghy, 2014, pp. 1-6
- [9] Darby, Sarah J., and Eoghan McKenna. "Social implications of residential demand response in cool temperate climates." Energy Policy Volume 49, October 2012, Page: 759-769.
- [10] C. W. Gellings, J.H. Chamberlin, "Demand-Side Management: Concepts and Methods", 2nd Ed. The Fairmont Press, Lilburn, GA, 1993.
- [11] T. Logenthiran, D. Srinivasan and T. Z. Shun, "Demand Side Management in Smart Grid Using Heuristic Optimization," in IEEE Transactions on Smart Grid, vol. 3, no. 3, pp. 1244-1252, Sept. 2012
- [12] P. K. Srivastava, "Financial analysis of investment under DSM", CORE International, Inc, 2012.
- [13] Palensky, P., & Dietrich, D. (2011). Demand side management: Demand response, intelligent energy systems, and smart loads. Industrial Informatics, IEEE Transactions on, 7(3), 381-388.
- [14] Matallanas, E., Castillo-Cagigal, M., Gutiérrez, A., Monasterio-Huelin, F., Caamaño-Martín, E., Masa, D., & Jiménez-Leube, J. (2012). Neural network controller for active demand-side management with PV energy in the residential sector. Applied Energy, 91(1), 90-97.
- [15] Arteconi, A., N. J. Hewitt, and F. Polonara. "State of the art of thermal storage for demand-side management." Applied Energy 93 (2012): 371-389.
- [16] Gelazanskas, Linas, and Kelum AA Gamage. "Demand side management in smart grid: a review and proposals for future direction." Sustainable Cities and Society 11 (2014): 22-30.
- [17] Aghaei, Jamshid, and Mohammad-Iman Alizadeh. "Demand response in smart electricity grids equipped with renewable energy sources: A review." Renewable and Sustainable Energy Reviews 18 (2013): 64-72.
- [18] Heussen K, You S, Biegel B, Hansen LH, Andersen KB. Indirect control for demand side management-A conceptual introduction. In Innovative Smart Grid Technologies (ISGT Europe), 2012 3rd IEEE PES International Conference and Exhibition on 2012 Oct 14 (pp. 1-8). IEEE, Berlin, Euroope.
- [19] Alvarez, Carlos, Antonio Gabaldón, and Angel Molina. "Assessment and simulation of the responsive demand potential in end-user facilities: application to a university customer." Power Systems, IEEE Transactions on 19.2 (2004): 1223-1231
- [20] Available at https://www.unido.org/fileadmin/media/documents/pdf/training_manual_on_sustainable_energy_regulation_and_policymaking_for_Africa.pdf
- [21] Power Factor Correction Charge, Bangladesh Energy Regulatory Commission, BERC Order no- 2012/01, Available : http://www.bpdb.gov.bd/bpdb/index.php?option=com_content&view=article&id=35&Itemid=33
- [22] Tareq Aziz, Sanjib Kumar Nandi, Md Siddikur Rahman, Ridown Rashid Riadh, "Study of Power Quality with changing customer loads in an urban distribution network," International Conference on Green Energy and Technology (ICGET), Dhaka, 11~12 September, 2015.
- [23] Electricity tariff rate in DPDC Areas, Available at <https://www.dpdc.org.bd/index.php/customer-service/tariff-rates>

FPGA Implementation of LBlock Lightweight Block Cipher

Md. Nazmul Hasan¹, Md. Tariq Hasan², Rafia Nishat Toma³, Md. Maniruzzaman⁴

Electronics and Communication Engineering Discipline, Khulna University

¹md.nazmul.hasan@outlook.com, ²mdthasan@gmail.com, ³rafiatoma.eceku@gmail.com, ⁴m_m_zaman@hotmail.com

Abstract—The ubiquitous computing era is characterized by wide deployment of resource constraint devices. Secure communication between such devices encounters a big challenge due to their resource limitation. The lightweight cryptographic algorithms are developed to address this challenge. In this work, implementation of LBlock, a lightweight encryption algorithm is investigated in Altera DE1 FPGA board. The encryption algorithm involves a Feistel structure and requires 32 clock cycles to encrypt 64-bit data. The implementation results show that it requires 326 Logic Elements (LEs) in Cyclone II EP2C20F484C7 platform. The maximum clock frequency was reported as 156.32 MHz and a corresponding power consumption of 107.62 mW. Also, the throughput is found as 312 Mbps. These results are better as compared to some other lightweight encryption such as XTEA and Hummingbird for the same FPGA platform.

Keywords—MCD, LBlock, Quartus II, FPGA.

I. INTRODUCTION

The trend of computing has gone through many significant changes over the last 50 years. In present times the computing environment is characterized by many small computing devices which had been termed as Ubiquitous Computing Devices [1]. These small computing devices surrounding us are being used in many forms of applications as electronic gadgets, security purposes, healthcare, environment and pollution monitoring, WSNs and so many mentioned in [2]. All these devices are indeed connected into the internet and exchange data that is processed, analyzed and used in many purposes. These devices are characterized by smaller size, having low power consumption and cost, mobility and inter connection [3]. They also have small computational capabilities. These limitations lead to a problem to ensure the security of the data transmitted by them as conventional ciphers, which are a means of protecting information, are not suitable for such resource limited devices. To address this problem a new class of cipher termed as the Lightweight cipher is developed. Majority of the hardware implementation of lightweight ciphers are carried out targeting ASICs due to their lower power consumptions and low costs when produced in large volumes. But the availability of low power and low cost FPGAs can set it as a potential medium for implementing lightweight ciphers [4]. They can also be easily reconfigured and take less time to market.

Some FPGA implementations are investigated for lightweight block ciphers targeting Xilinx and Altera FPGA

boards. Donggeon Lee *et al.* performed hardware implementation of three version of the lightweight block encryption algorithm LEA in both Xilinx Vertex 5 series and Altera Cyclone III series in [5]. They investigated the area (in terms of total LEs), maximum frequency, and throughput for LEA-128, 192 and 256 versions. Their results show that the number of LEs increased as the key size is increased. Two lightweight algorithms, Present and HIGHT, are implemented in Xilinx Spartan-3 FPGAXC3S50-5 device by Panasayya Yalla and Jens-Peter Kaps in [4]. It is found that Present requires 117 slices whereas HIGHT needs 91 slices. HIGHT provides a higher throughput of 65.48Mbps in contrast to 28.64 Mbps by Present. Another lightweight algorithm Hummingbird is also been implemented in Xilinx Virtex-5 xc5v1x20t-2-ff-323 FPGA board by Nikita Arora and Yogita Gigras in [6] where they have found that the Hummingbird algorithm implementation requires 4242 slices. Implementation of the same in Altera Cyclone II EP2C20F484C7 FPGA device by Miss. Pranali Umap and Dr. A.S.Joshi in [7] has required 420 LEs and 240 dedicated registers. Issam Damaj, Samer Hamade and Hassan Diab analyzed the performance of extended tiny encryption algorithm (XTEA) in [8] for Xilinx Vertex II Pro 2VP2fg256 FPGA board where 634 LUTs are required. They also implemented the same algorithm in two different devices EP2S15F484C and EP2S60F672C of Altera's Stratix II FPGA board. For both devices, 573 LUTs were required.

In this paper we have developed the RTL code of the LBlock lightweight cipher using Verilog HDL and implemented in an FPGA board. The design is synthesized and analyzed with Quartus II EDA tool.

This paper has been organized as follows. Section II illustrates the LBlock cryptographic algorithm. An overview of implementation methodology is discussed in section III. In section IV the results have been presented. Finally section V concludes the paper.

II. LBLOCK LIGHTWEIGHT ALGORITHM

LBlock is a lightweight block cipher introduced by Wenling Wu and Lei Zhang in [9]. It has a block size of 64-bit and key size of 80-bit. The cipher requires 32-round operation to encrypt 64-bit data following a variant of Feistel structure. In each round of encryption and decryption a round key of 32-bit is involved which is produced from the 80-bit key through a key scheduling process. LBlock also involves Substitution box (S-box) layer and permutation layer (P). The substitution

layer is performed using 10 S-boxes of size 4×4 as given in **TABLE I**.

TABLE I. S-BOXES USED IN LBLOCK.

S ₀	E	9	F	0	D	4	A	B	1	2	8	3	7	6	C	5
S ₁	4	B	E	9	F	D	0	A	7	C	5	6	2	8	1	3
S ₂	1	E	7	C	F	D	0	6	B	5	9	3	2	4	8	A
S ₃	7	6	8	B	0	F	3	E	9	A	C	D	5	2	4	1
S ₄	E	5	F	0	7	2	C	D	1	8	4	9	B	A	6	3
S ₅	2	D	B	C	F	E	0	9	7	A	6	3	1	8	4	5
S ₆	B	9	4	E	0	F	A	D	6	C	5	7	3	8	1	2
S ₇	D	A	F	0	E	4	9	B	2	1	8	3	7	5	C	6
S ₈	8	7	E	5	F	D	0	6	B	C	9	A	2	4	1	3
S ₉	B	5	F	0	7	2	9	D	4	8	1	2	E	A	3	6

A. Encryption using LBlock

At the beginning the plaintext is divided into two equal segments, $M = X_0 \parallel X_1$ then the encryption process is as following:

1. $X_i = F(X_{i-1}, K_{i-1}) \oplus X_{i-2} \lll 8$ for $i = 2 \dots 33$
2. The encrypted message is $C = X_{32} \parallel X_{33}$.
3. Here the round function is defined as $F(x, k_i) = P(S(X \oplus k_i))$. Where, S is the confusion function which denotes the non-linear layer of the round function, and it consists of eight 4-bit S-boxes ($S_0 \dots S_7$). Finally, P is defined as the permutation of eight 4-bit words.

B. Decryption using LBlock

Decryption using LBlock cipher is similar but inverse in process that of the encryption. It also requires 32-rounds and follows a Feistel structure. In this case, $C = X_{32} \parallel X_{33}$ is the 64-bit encrypted message and the decryption procedure is as follows:

1. $X_j = F(X_{j+1}, k_{j+1}) \oplus X_{j+2} \ggg 8$ for $j = 31 \dots 0$
2. The restored 64-bit plaintext is $M = X_1 \parallel X_0$.

C. Key scheduling in LBlock

From the 80-bit key $K = (k_{79}, k_{78}, \dots, k_0)$, the leftmost 32-bit is taken as the first round subkey (k_1). Then the 80-bit key is updated in each round as follows:

1. K is circularly left shifted by 29 bits, $K \lll 29$.
2. S-boxes S_8 and S_9 are used to substitute $k_{79}, \dots, k_{72}$ as:
 $[k_{79} \ k_{78} \ k_{77} \ k_{76}] = S_9 [k_{79} \ k_{78} \ k_{77} \ k_{76}]$
 $[k_{75} \ k_{74} \ k_{73} \ k_{72}] = S_8 [k_{75} \ k_{74} \ k_{73} \ k_{72}]$
3. Then, the value of the round counter is XORed with $k_{50}, \dots, k_{46}$
 $[k_{50} \ k_{49} \ k_{48} \ k_{47} \ k_{46}] = [k_{50} \ k_{49} \ k_{48} \ k_{47} \ k_{46}] \oplus [i]_{bin}$

Finally, the left most 32-bit of the updated key $[k_{79}, \dots, k_{48}]$ is taken as the round sub key.

III. IMPLEMENTATION METHODOLOGY

The main focus of this work is to observe the performance of LBlock algorithm in FPGA platform. For this purpose, an Altera Cyclone II DE1 board is chosen. Quartus II 13.0 Web Edition is used as the EDA tool to perform synthesis and configure the FPGA board.

First, the RTL design for the algorithm is developed with Verilog as the hardware description language. Initially ModelSim PE 10.4 is used to develop the RTL and check the functionality of the algorithm.

After developing a full functional RTL of the algorithm, the design is then synthesized in Quartus II software. After synthesis the functionality is verified by observing the encrypted and decrypted message with the help of waveform editor in Quartus II.

Next, static timing analysis (STA) is performed for the synthesized design in order to find out the maximum clock frequency at which the design can work. The TimeQuest Timing Analyzer tool is used for this. A synchronous design constraint (.sdc) file is created and used for timing analysis.

Later, the power analysis is performed using the PowerPlay Power Analyzer tool. This tool calculates the total thermal power consumed by the design when operated at the maximum clock frequency as determined by STA during timing analysis. Total thermal power dissipation comprises of core dynamic thermal power, core static thermal power, and I/O thermal power dissipation.

Finally, the encrypted bits are observed in the DE1 board using the output LEDs (LEDR) of the board that resembles with the ModelSim simulations.

IV. RESULT AND DISCUSSION

A. Functional simulation

To check the functional correctness of the RTL design of the encryption algorithm it is necessary to look on the encrypted form of the plaintext. A sample of the plaintext vector and their corresponding encrypted message is tabulated in **TABLE II**.

TABLE II. INPUT PLAINTEXT AND RESULTING CIPHER TEXT

64-bit Plaintext	80-bit key (K)	64-bit cipher text
0000000000000000	FFFFFFFFFFFFFFFFFFFF	9B8F81DAA8A99D95
FFFF11110000AAAA		B47A3C33E1BBB55A
FFFE222200015554		424DD286FC90CF9E
FFFD33330001FFFE		4DD4CF7E330BD42D

The resulting cipher texts are decrypted using the decryption algorithm which restored the similar plaintexts.

B. Synthesis report

The synthesis reports the resource utilization by the design in terms of the number of logic elements (LEs) required for implementing it in FPGA board. The Verilog RTL code

synthesis report in Altera DE1 board is provided in **TABLE III**.

TABLE III. SYNTHESIS RESULTS OF LBLOCK.

Family	Cyclone II
Total logic elements	326 (available 18,752)
Total combinational functions	280 (available 18,752)
Dedicated logic registers	183 (available 18,752)
Total registers	183

The RTL design of the algorithm requires only 326 logic elements to implement in DE1 FPGA board. This indicates that the area required by the synthesized circuit is very small which can be observed from the chip planner view as given in Fig. 1.

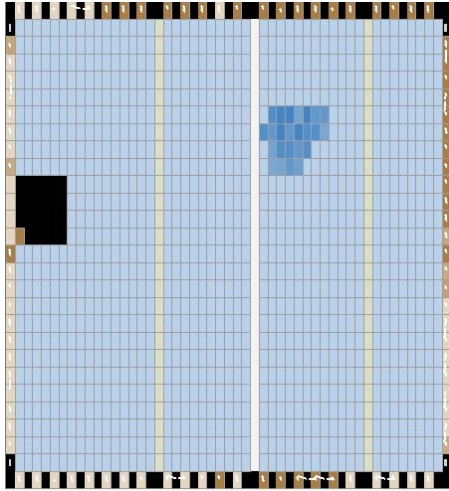


Fig. 1. Chip planner view of synthesized RTL design.

The chip planner view shows the placement and amount of the CLBs to implement the design inside the FPGA chip. Fig. 1 shows that the area occupied by the synthesized circuit (area occupied by rectangles bluer than the others) is quite small.

C. Static timing analysis report

The STA report generated by the TimeQuest Timing Analyzer tool provides the maximum operating clock frequency of the circuit. The summary reports timing analysis is provided in **TABLE IV**.

TABLE IV. STA SUMMARY REPORT.

Delay model	Slow
Setup slack	+1.903
Hold slack	+0.445
F_{max}	156.32 MHz

STA summary shows the setup slack and hold slack values for the longest or critical path inside the circuit. Both the slack

values are positive for the critical path which means there will be no timing violation at a clock frequency of 156.32 MHz.

D. Power analysis report

The power analysis report generated from PowerPlay Power Analyzer tool is given in **TABLE V**.

TABLE V. POWER ANALYSIS REPORT

Power models	Final
Core dynamic thermal power dissipation	11.3 mW
Core static thermal power dissipation	47.45 mW
I/O thermal power dissipation	65.28 mW
Total thermal power dissipation	124.06 mW

Power analysis report shows that, I/O thermal power constitutes the most of the power dissipation by the design. It is because; the number of I/O ports in this design is 130. The static thermal is the power consumed by the FPGA when no signal is toggling. The static leakage current through the logic elements and routing constitutes this power dissipation. Static thermal power depends upon the number of logic elements present there in the circuit. The dynamic thermal power is caused by signal toggling and charging-discharging of capacitive loads. It is related with the operating frequency. These values of power consumption are also device specific. The target FPGA device is based on 90 nm technology, so for other FPGA device these power consumptions will be different.

The findings are comparable with the lightweight cipher XTEA [10] implemented in the same Cyclone II DE1 FPGA board. The comparison is presented in **TABLE VI**.

TABLE VI. COMPARISON BETWEEN XTEA AND LBLOCK.

Lightweight Algorithm	Number of LEs	Number of Registers	Throughput	Frequency
XTEA [10]	850	625	200 Mbps	50 MHz
LBLOCK	326	183	312 Mbps	156.32 MHz

It can be seen that LBlock requires 65% less LEs than XTEA and can operate at a frequency three times faster than XTEA and the throughput is also higher.

The LBlock also consumes less LEs and less core dynamic power than KATAN₆₄, analyzed in [11], which requires 1560 LEs and a core dynamic power consumption of 45.69 mW.

LBlock also exhibits compact design than Hummingbird cipher investigated in [7]. The comparison with Hummingbird is given in **TABLE VII**.

The comparison shows that our implementation of LBlock is better than Hummingbird in terms of LEs and dedicated registers though it requires more combinational functions.

TABLE VII. COMPARISON BETWEEN HUMMINGBIRD AND LBLOCK.

Algorithms	Total Logic Elements	Total Combinational Functions	Total Dedicated Registers
Hummingbird [7]	420	260	240
LBlock	326	280	183

The simplified block diagram of the data path of the synthesized design is presented in Fig. 2. The block diagram depicts simplified representation of flow of data between different registers and the associated mathematical operations.

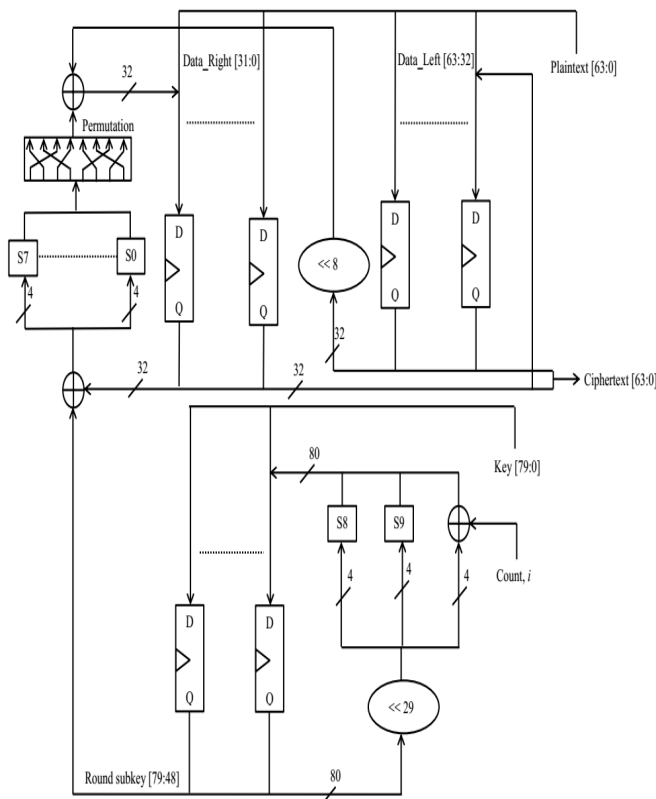


Fig. 2. Simplified data path of the synthesized design

V. CONCLUSION

In this paper a full functional RTL is designed for LBlock lightweight cipher which requires 32-clock cycles to encrypt a block of 64-bit message. The design is then implemented in an Altera DE1 FPGA board. It is found that, the design requires 326 LEs to be implemented in the FPGA board and can operate at a clock frequency of 156.32 MHz. These are better than XTEA, Hummingbird, and KATAN_64 implementations. The thermal power dissipation by the design is found as 124.06 mW of which the core dynamic thermal power dissipation contributes 11.3mW. The core dynamic thermal power dissipation by our design is much less than that of KATAN_64. This design may be optimized further by using structured memory. Also if hard key is utilized then the size of the design may reduce.

REFERENCES

- [1] Mark Weiser. (1996, March 17) Ubiquitous Computing. [Online]. <http://www.ubiq.com/hypertext/weiser/UbiHome.html>
- [2] N. Jeon, C. Leem, M. Kim and H. Shin, "A taxonomy of ubiquitous computing applications", *Wireless Pers Commun*, vol. 43, no. 4, pp. 1229-1239, 2007.
- [3] J. Kim and H. Ryou, "Safe RFID System Modeling Using Shared Key Pool in Ubiquitous Environments", in *International Conference on Intelligent Computing, ICIC 2005*, Hefei, 2005, pp. 831-840.
- [4] P. Yalla and J. Kaps, "Lightweight Cryptography for FPGAs", in *International Conference on Reconfigurable Computing and FPGAs, Quintana Roo*, 2009, pp. 225 - 230.
- [5] D. Lee, D. Kim, D. Kwon and H. Kim, "Efficient Hardware Implementation of the Lightweight Block Encryption Algorithm LEA", *Sensors*, vol. 14, no. 1, pp. 975-994, 2014.
- [6] N. Arora and Y. Gigras, "FPGA Implementation of Low Power and High Speed Hummingbird Cryptographic Algorithm", *International Journal of Computer Applications*, vol. 92, no. 16, pp. 42-47, 2014.
- [7] M. Umap and A. Joshi, "Performance Analysis of Hummingbird Cryptographic Algorithm using FPGA", *Journal of Multidisciplinary Engineering Science and Technology (JMEST)*, vol. 2, no. 11, pp. 3274-3279, 2015.
- [8] I. Damaj, S. Hamade and H. Diab, "Efficient tiny hardware cipher under verilog", *HIGH PERFORMANCE COMPUTING & SIMULATION*, 2008.
- [9] W. Wu and L. Zhang, "LBlock: A Lightweight Block Cipher", in *9th International Conference on Applied Cryptography and Network Security, Nerja*, 2011, pp. 327-344.
- [10] C. Baskar, C. Balasubramanian and D. Manivannan, "Establishment of Light Weight Cryptography for Resource Constraint Environment using FPGA", in *International Conference on Information Security & Privacy (ICISP2015)*, Nagpur, 2016, pp. 165-171.
- [11] B. Mohd, T. Hayajneh and Z. Khalaf, "Optimization and modeling of FPGA implementation of the Katan Cipher", in *6th International Conference on Information and Communication Systems (ICICS)*, Amman, 2015, pp. 68 - 72.

Design and Optimization of Optical Add-Drop Filter and Sensor using 2D Photonic Crystal

Shakila Naznin*, Rajib Mandal, Shamlendu Sarker, Md. Tanveer Alam, Md. Shamim Ahsan and Md. Younus Ali

Electronics and Communication Engineering Discipline

Khulna University, Khulna 9208, Bangladesh

*ratna_kuece@yahoo.com

Abstract— In this paper, we have investigated the performance of a 2D photonic crystal add-drop filter (ADF) and sensor based on a quasi-square ring resonator cavity. High quality factor (Q) is a crucial requirement for both filter and sensor. We have focused on improving the Q of the structure by applying different optimization techniques. The designed $20\ \mu\text{m} \times 19\ \mu\text{m}$ photonic crystal has a square lattice of GaAs rods in air with refractive index of 3.40 at operating wavelength of $1.55\ \mu\text{m}$. The effect of scattering rods at four corners and inner rods of the ring resonator cavity are analyzed at first. After observing slight improvement in Q, it is found that comparatively high Q is achieved by the application of a hetero ring resonator cavity to the structure. Besides, the optimized ADF is tuned to two different wavelengths simultaneously available from two output ports, $1537.40\ \text{nm}$ $1548.5\ \text{nm}$. Finally, the performance of the optimized structure is investigated as a sensor and it is found that the sensitivity is $11.6\ \text{nm/RIU}$ (Refractive Index Unit). Simulation results are obtained using finite difference time domain (FDTD) method.

Keywords—*photonic crystal; add-drop filter; quasi-square ring resonator; hetero-ring; sensor.*

I. INTRODUCTION

Future world is going to be highly dependent on all optical devices to meet the huge demand of high performance systems. Photonic crystals (PhCs) have attracted great interest for designing all optical integrated circuits due to their exceptional photonic band gap (PBG) and strong photon confinement effect [1-3]. PhCs exhibit PBG property because of the inherent periodicity and electromagnetic propagation is almost zero in the PBG region. This periodicity can be broken by the introduction of defects (point or line defect) and the propagation of light can be localized in the PBG region around the defect. All the PhC devices are based on this attractive band gap engineering technique. These devices are also gaining popularity because of their compact size (10-100 times smaller than the conventional waveguides), long life time and suitability for photonic integrated circuits (PICs) [4-5]. In recent years, many optical devices such as add-drop filters [6-7], power splitters [8-9], multiplexers and de-multiplexers [5], [10], [11], switches [12], logic gates [13-15], sensors [16-21] etc. have been proposed and fabricated experimentally. Optical add drop filter (ADF) is one of the fundamental building blocks of optical add drop multiplexers (OADMs), optical modulators, optical switches etc. needed for PICs and wavelength division multiplexed (WDM) optical

communication systems [6], [7]. High Q increases the selectivity of a filter. On the other hand, PhC based biosensors have recently attracted the researchers due to the high quality factor [19-21]. Bio sensing includes sensing of bio- molecules, bio-chemicals, proteins etc. Ultra high sensitivity is required in bio-sensing to detect very low amount of analyte and small bio-molecules because of small change in refractive index. High quality factor ensures high sensitivity [21].

So, high Q is a crucial requirement for filters as well as for sensors. Photonic crystal ring resonator (PhCRR) based ADFs and sensors are very promising because of their high quality factor, scalability in size and flexibility in tuning. Different optimization techniques are examined to improve the performance of the PhCRR based structures. Several parameters of the PCRR such as refractive index, lattice constant and radii of rods (coupling rods, inner/core rods, and adjacent rods) are varied to improve the PhC performance [22-26]. Scattering rods at each corner of the PCRR cavity have a great effect on the performance of the device. These rods act like reflectors which reduce back-reflections and radiation losses from their corresponding corners and as a result, quality factor increases [25], [26]. Properties of inner rods such as size, refractive index, and position also play an important role in the light confinement and tuning. PCRR actually controls the total behavior of the PhCRR based PhCs. So there is a great scope to optimize the performance of these devices by changing the PCRR parameters.

In this paper, we have focused on the performance of a 2D photonic crystal add-drop filter and sensor based on a quasi-square PhCRR cavity. All the evaluations are based on the improvement of Q. At first, the effect of the scattering rods at four corners and inner rods of the cavity has been analyzed. Subsequently, the performance of the PhC structure is observed as an ADF and a sensor with a whole hetero PCRR acting as a photonic well.

II. STRUCTURE DESIGN AND ANALYSIS

A. Design Considerations

The structure used in this paper is a two dimensional $20\ \mu\text{m} \times 19\ \mu\text{m}$ square lattice of GaAs ($n = 3.40$) rods in the air host. The lattice constant (a) and radius (R) of GaAs rods are $0.540\ \mu\text{m}$ and $0.185a$ respectively. The structure consists of four ports labeled as input port, port B, port C and port D as shown in Fig. 1. Operating wavelength is chosen as of

1550 nm. A Gaussian pulse which is wide enough in frequency domain to cover the range of the frequencies we are interested in is launched into the input port.

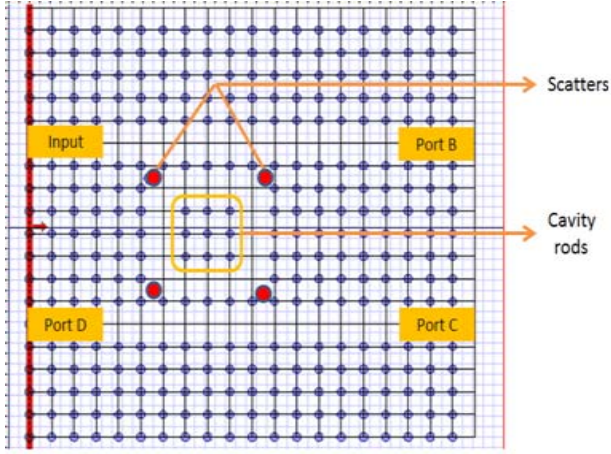


Fig. 1. Schematic structure of the add-drop filter with four ports.

By putting a ring resonator between the two waveguides, energy can be coupled from the input waveguide to the other output waveguides at the resonant frequency of the PCRR cavity. The four rods placed at the corner of the ring resonator are called scatters/ scattering rods as shown in Fig. 1. These scattering rods have a great impact on the performance of the device. They can improve the spectral selectivity, drop efficiency and suppress the counter propagating modes [25], [26]. Other considered design parameters are listed in Table I.

TABLE I. DESIGN PARAMETERS OF THE UNOPTIMIZED STRUCTURE

Design Parameters	Values
Radii of the rods	0.1 μm
Lattice constant	0.54 μm
Refractive indices of rods	3.40
Background index	1
Size	20 $\mu\text{m} \times 19 \mu\text{m}$
PBG range	1300 nm – 1700 nm

B. Bandgap Calculation

The band structure diagram of this lattice is calculated by the 2D plane wave expansion (PWE) method as shown in Fig. 2.

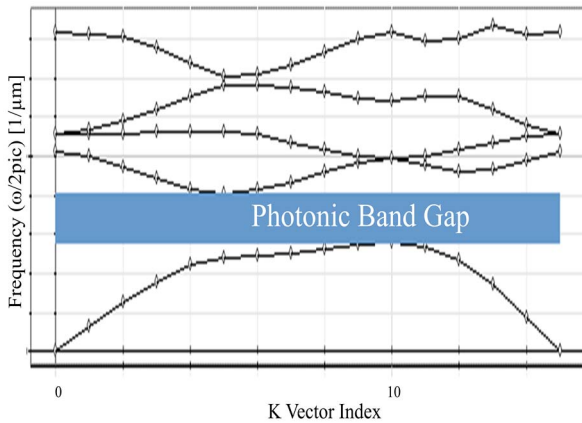


Fig. 2. Band diagram of the unoptimized (fundamental) structure.

III. RESULTS AND DISCUSSION

Initially, the quality factor of the unoptimized structure is measured. Then different optimization techniques are applied on the structure to increase the quality factor and finally the performance of the device is observed as a filter and sensor.

A. Performance of the Unoptimized PhC Structure

Transmission spectra of port B and port D are obtained for the unoptimized structure by conducting fast fourier transform (FFT) of the fields and calculating by 2D FDTD method. At normal condition, input signal entirely reaches port B while at resonance condition, the input signal reaches one of the dropping waveguides depending upon the symmetry of the ring resonator coupler. Thus the PhC structure can work as an ADF. All rods are of same radius (0.1 μm) and same refractive index (3.40) in the unoptimized structure. It is easy to understand from Fig. 3 that poor spectral selectivity is found for port B and port D.

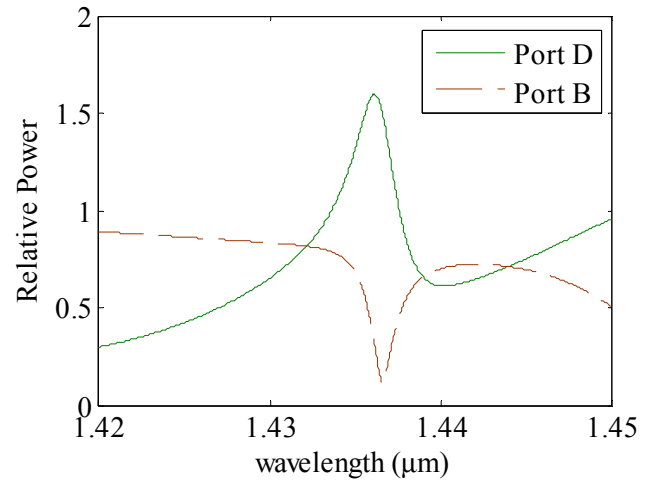


Fig. 3. Transmission spectra of the unoptimized add-drop filter.

B. Optimization of the PhC Structure

1) Different radii of cavity inner rods and scattering rods:

The radius of the scattering rods and inner rods has a great effect on the coupling, output power and spectral selectivity.

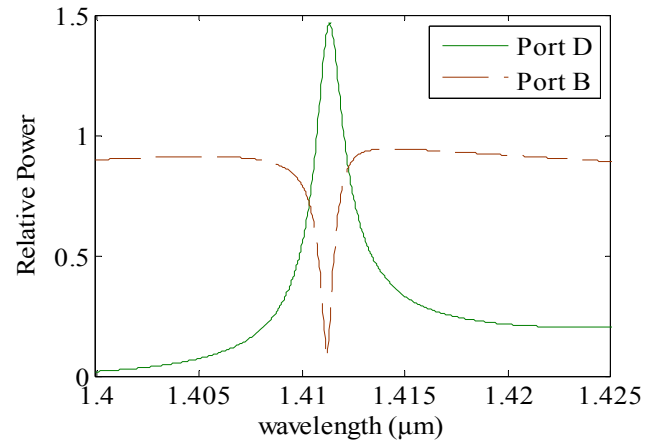


Fig. 4. Transmission spectra of the add-drop filter for optimized structure-1.

2) *Different refractive index of the scattering rods*: The refractive index of the scattering rods also play an important role in the performance of the PhC structure. The scattering rods are assigned different refractive index in the ‘optimized structure-2’. As a result, the position and width of the photonic gap changes and new opportunities of photon management are created. The new refractive index is chosen as of 3.47. It has been observed in Fig. 5 that the center wavelength shifts to the longer wavelength and the quality factor improves because of the refractive index difference.

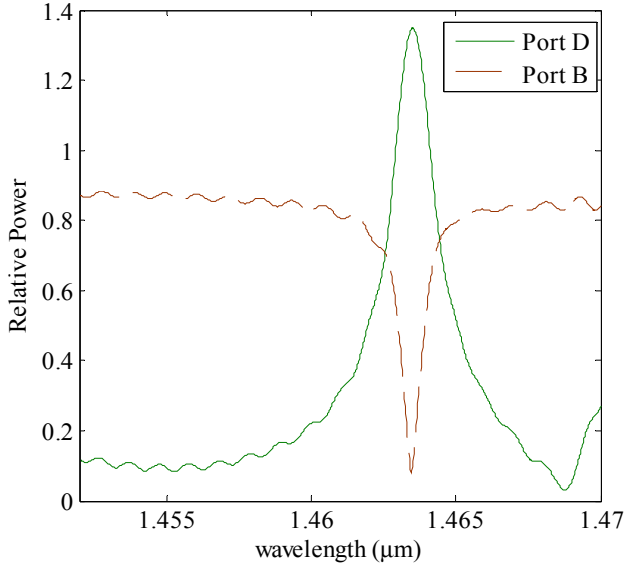


Fig. 5. Transmission spectra of the add-drop filter for optimized structure-2.

3) *Hetero-ring resonator cavity*: Upon confirming that the quality factor improves because of the different refractive indices of the scattering rods, the PhC performance is then observed for a whole hetero ring cavity. The hetero ring is created in the ‘optimized structure-3’ by assigning the refractive index $n = 3.47$ to the whole ring resonator as shown in Fig. 6.

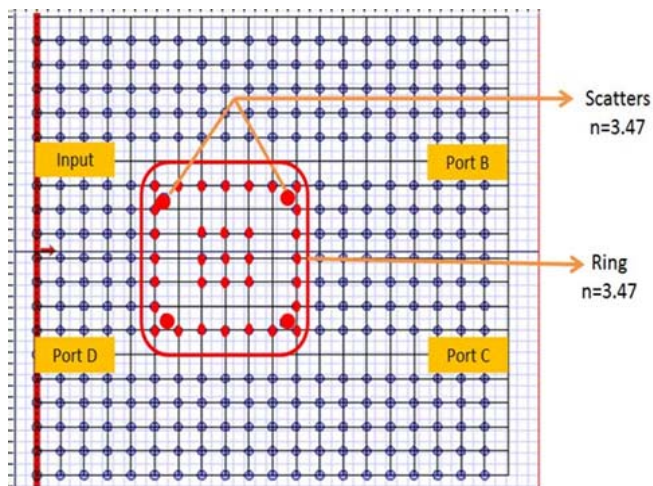


Fig. 6. Schematic of the optimized structure-3 with a hetero-ring.

As a result, the bandgap and modal behavior of the structure changes dramatically. The scenario is that the center part has a different refractive index than the surroundings and the ring resonator cavity works like a photonic well. So, theoretically it is expected that the performance of the structure will be improved.

From the transmission spectra as shown in Fig. 7, it is noticed that the ports quality factor have been improved. It is also clear that pass band and stop band have been reduced.

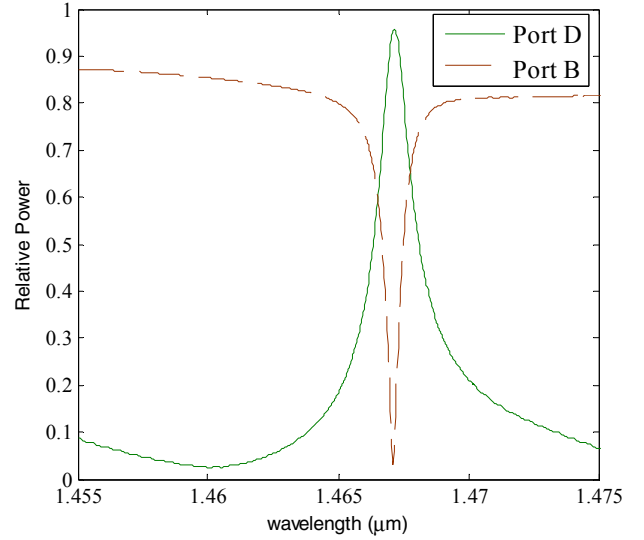


Fig. 7. Transmission spectra of the add-drop filter for optimized structure-3.

C. Quantitative Measurements

Different optimization techniques are applied to the PhC structure to improve the quality factor. The optimized structures are denoted as ‘optimized structure-1’, ‘optimized structure-2’ and ‘optimized structure-3’. The gradual improvements in the Q have been summarized in Table II and Table III. Finally the structure is tuned to be operated in the optical third communication window.

TABLE II. QUALITY FACTOR OF THE INVESTIGATED ADF FOR PORT D

Structure	λ (nm)	$\Delta\lambda$ (nm)	$Q(\lambda/\Delta\lambda)$
Unoptimized Structure	1436.00	4.6330	310.81
Optimized Structure 1	1411.40	2.2950	614.98
Optimized Structure 2	1462.20	1.6540	884.83
Optimized Structure 3	1467.20	1.4560	1007.54

TABLE III. QUALITY FACTOR OF THE INVESTIGATED ADF FOR PORT B

Structure	λ (nm)	$\Delta\lambda$ (nm)	$Q(\lambda/\Delta\lambda)$
Unoptimized Structure	1436.60	1.5930	901.82
Optimized Structure 1	1411.12	0.6830	2066.18
Optimized Structure 2	1462.47	0.6830	2141.24
Optimized Structure 3	1467.10	0.5720	2564.86

D. Separation of Wavelengths at Output Ports

Point defects can be used also for tuning purposes. To separate output wavelengths, different point defects can be placed in the output channels [24]. To accomplish this, port B and port C are used where port D is closed. Tuning can be done

by changing the radius or the refractive index of the point defects. Wavelengths are separated here by the change of point defects radii as shown in Fig. 8.

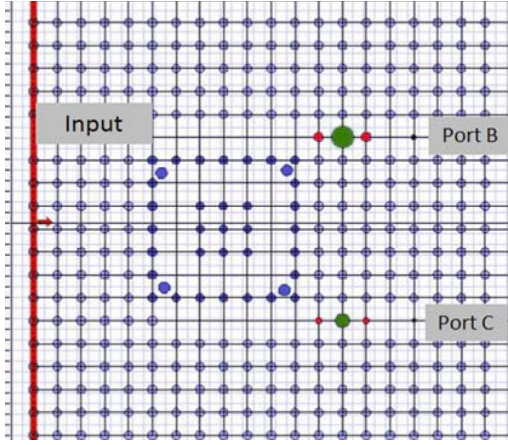


Fig. 8. Tuning of the output signals by placing point defects in output ports.

The values of point defect radii have been provided in Table IV. From the transmission spectra as depicted in Fig. 9, it is observed that two separate wavelengths, 1548.4 nm and 1537.40 nm can be dropped simultaneously from port B and port C respectively. This technique leaves the option to tune the output from respective ports keeping other properties constant.

TABLE IV. RADII OF DEFECT RODS AT OUTPUT PORTS

Defect Rods	Port B	Port C
Red Rods	0.20a	0.13a
Green Rods	0.46a	0.30a
Black Rods	0.09a	0.06a

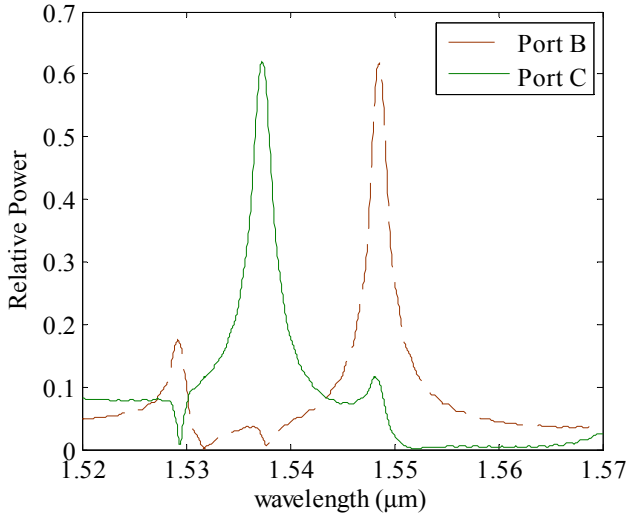


Fig. 9. Two separate wavelengths 1548.4 nm and 1537.40 nm available from port B and port C respectively.

E. Optimized PhC Structure as a Sensor

The optimized PhC structure can act as a sensor if a virtual sensing environment can be created. Sensor can sense the physical or biological change such as temperature, humidity, displacement, pressure, bio-chemicals, bio-molecules etc.

Because of this change, optical properties of the device can be monitored in the transmission spectra by the shift of resonant wavelengths or intensity variations. To create a virtual sensing environment, it is assumed that the refractive indices of the inner rods have been changed due to analyte binding or any other interaction. This change is highlighted by red color in the structure as shown in Fig. 10. As sensing can be accomplished by one output port, port B and port C are closed.

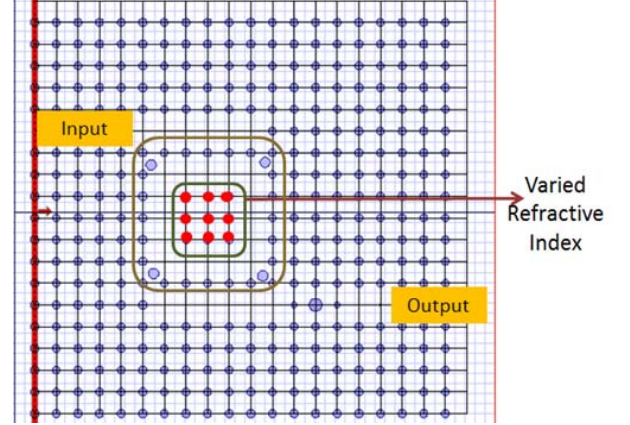


Fig. 10. Schematic of the optimized structure as a sensor.

Refractive index of the inner rods is changed in an interval of 0.04 and shift of resonant wavelengths are mentioned in Table V. From the transmission spectra as shown in Fig. 11, it can be seen that the device can detect the change of 0.04 in the refractive index effectively.

TABLE V. SHIFT OF WAVELENGTHS FOR SENSING

Refractive Index	Center Wavelength, nm	Quality Factor
3.47	1510.3	1015.67
3.51	1510.7	1037.4
3.55	1511.1	1059.75
3.59	1511.5	1081.81

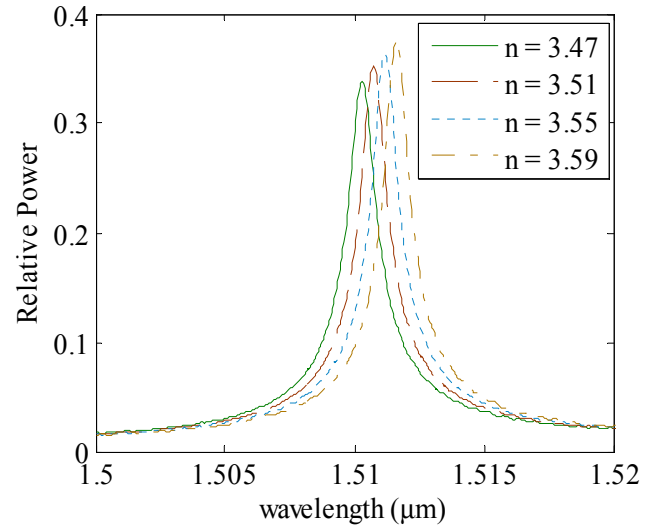


Fig. 11. Wavelength shifts visible in the transmission spectra for sensing.

To determine the sensitivity of the device, finally the refractive indices of inner rods are varied from 3.47 to 3.87 with an

interval of 0.04 and a linear response curve is found as shown in Fig. 12. The sensor is based on the resonant wavelength shift scheme and it is noticed that the sensor has the sensitivity of 11.6 nm/RIU. This device can be used in different types of sensing applications including bio-sensing.

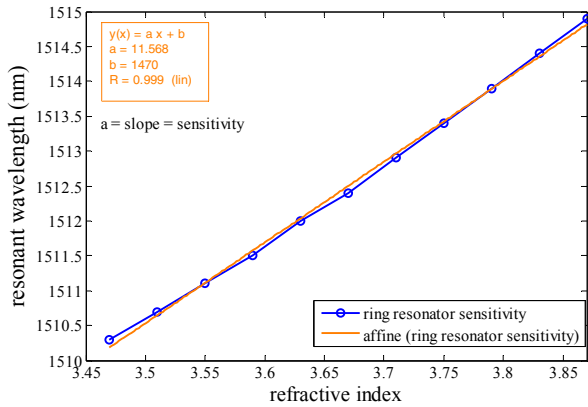


Fig. 12. Response curve of the optimized structure as a sensor.

IV. CONCLUSION

The paper is focused on 2D Photonic Crystal (PhC) add-drop filter and sensor. Because of the photonic band gap property, PhCs can confine and route the light effectively with compactness and have recently become popular as add-drop filters and sensors. Add-drop filters are playing an important role in WDM applications whereas bio-sensors based on PhCs are gaining popularity day by day. High Q is a crucial requirement for both filter and sensors. So, we have tried to improve the Q by applying and comparing different optimization techniques. It is found that the radius and refractive index change in the scattering rods contribute a lot in the improvement of Q. It is also found that comparatively high Q can be achieved if the whole ring resonator works like a hetero-ring, a photonic well. Besides, the optimized add-drop filter is tuned to two different wavelengths 1537.40 nm and 1548.5 nm, simultaneously available from two ports. Finally, the performance of the optimized structure is investigated as a sensor and it is found that the sensitivity is 11.6 nm/RIU (Refractive Index Unit).

REFERENCES

- [1] K. Sakoda, Optical Properties of Photonic Crystals, Springer-Verlag, Berlin, 2001.
- [2] F. Mehdizadeh, and H. Alipour-Banaei, "Band gap management in two dimensional photonic crystal thue-morse structures", *J. Opt. Commun.* vol. 34, pp. 61-65, 2013.
- [3] Z. Wu, K. Xie, and H. Yang, Band gap properties of two dimensional photonic crystals with rhombic lattice, *Optik*, vol. 123, no. 06, pp. 534-536, March 2012.
- [4] S. Robinson, "Photonic crystal ring resonator based optical filters for photonic integrated circuits", AIP Conf. Proc. vol. 1620, pp. 131-138, March 19-21, 2014, Kerala, India.
- [5] T. Saraniya, S. Robinson, and K. Vijaya Shanthi, "Design and simulation of two dimensional photonic crystal ring resonator based four port wavelength demultiplexer", *IJISSET*, vol. 1, no. 2, pp. 255-261, April 2014.

- [6] Z. Qiang, W. Zhou, and Richard A. Soref, "Optical add-drop filters based on photonic crystal ring resonators", *Opti. Express*, vol. 15, no. 4, pp. 1823-1831, 2007.
- [7] S. Robinson, and R. Nakkeeran, "Characteristics of add-drop filter using single and dual PCRR in square lattice", *Optik*, vol. 124, no. 23, pp. 5918-5922, December 2013.
- [8] M. David, A. Ghaffari, F. Monifi, and M.S. Abrishamian, "Analysis of photonic crystal power splitters with different configurations", *Journal of Applied Sciences*, vol. 8, no. 8, pp. 1416-1425, 2008.
- [9] N. Nozhat and N. Granpayeh, "Analysis and simulation of a photonic crystal power divider", *Journal of Applied Sciences*, vol. 7, no. 22, pp. 3576-3579, 2007.
- [10] G. Manzacca, D. Pacioti, A. Marchese, M.S. Moreolo, and G. Cincotti, "2D photonic crystal cavity based WDM multiplexer", *Photonics and Nanostructures-Fundamentals and Applications*, vol. 5, no. 4, pp. 164-170, 2007.
- [11] A. Gaffari, F. Monifi, M. Djavid, and M. S. Abrishamian, "Photonic crystal bends and power splitters based on ring resonators", *Optics Communications*, vol. 281, no. 23, pp. 4028-4032, 2008.
- [12] Qiong Wang, Yiping Cui, Haiyu Zhang, and Changchun Yan, "The position independence of heterostructure coupled waveguides in photonic crystal switch", *Optik*, vol. 121, no. 8, pp. 684-688, 2010.
- [13] Pallavi Singh, Devendra Kr. Tripathi, Shikha Jaiswal and H. K. Dixit, "All-optical logic gates: designs, classification, and comparison," *Advances in Optical Technology*, vol. 2014, pp. 1-13, March 19, 2014.
- [14] Weijia Liu et al., "Design of ultra-compact all-optical XOR, XNOR, NAND and OR gates using photonic crystal multi-mode interference waveguides," *Optics and Laser Technology*, vol. 50, pp. 55-64, September 2013.
- [15] Aryan Salamanpour, Shahrar Mohammadnejad, and Pedram Taghnejad Omran, "All-optical photonic crystal NOT and OR logic gates using nonlinear Kerr effect and ring resonators", *Optical and Quantum Electronics*, vol. 47, no. 12, pp. 3689-3703, December 2015.
- [16] Y. Liu, and H. W. M. Salemink, "Photonic crystal-based all optical on-chip sensor", *Optics Express*, vol. 20, no. 18, pp. 19912-19921, 2012.
- [17] S. Olayee, and A. A. Dehghani, "High resolution and wide dynamic range pressure sensor based on two-dimensional photonic crystals", *Photonic Sensors*, vol. 2, no. 1, pp. 92-96, 2012.
- [18] S. Kita, K. Nozaki and T. Baba, "Refractive index sensing utilizing a CW photonic crystal nanolaser and its array configuration", *Optics Express*, vol. 16, no. 11, pp. 8174-8180, 2008.
- [19] F. I. Hsiao and C. Lee, "Computational study of photonic crystals nanoring resonator for biochemical sensing", *IEEE Sensors Journal*, vol. 10, no. 7, pp. 1185-1191, 2010.
- [20] S. Olayee and S. Najafgholizadeh, "Computational study of a label-free biosensor based on photonic crystal nanocavity resonator", *Applied Optics*, vol. 52, no. 29, pp. 7206-7213, 2013.
- [21] Zhixuan Xia and Zhiping Zhou, "Sensitivity analysis of microring resonator based biosensor: the quality factor perspective", International Symposium on Optoelectronics, pp. 44-46, November 1-4, 2006, China.
- [22] Mayur Kumar Chippa, "The simulation & improved design of tunable channel drop filter using hexagonal photonic crystal ring resonator", AIP Conf. Proc. vol. 1620, pp. 282-288, March 19-21, 2014, India.
- [23] Farhad Mehdizadeh, Hamed Alipour-Banaei and Somaye Serajmohammadi, "Channel-drop filter based on a photonic crystal ring resonator", *J. Opt.*, vol. 15, no. 7, pp. 075401/1-075401/7, 2014.
- [24] L. Li, and G. Q. Liu, "Photonic crystal ring resonator channel drop filter", *Optik*, vol. 124, no. 17, pp. 2966-2968, 2013.
- [25] Hamed Alipour-Banaei, Mahsa Jahanara, and Farhad Mehdizadeh, "T shaped channel drop filter based on photonic crystal ring resonator", *Optik*, vol. 125, no. 18, 2014.
- [26] Alireza Tavousi, and Mohammad Ali Mansouri-Birjandi, "Performance evaluation of photonic crystal ring resonators based optical channel add-drop filters with the aid of whispering gallery modes and their Q-factor", *Optical and Quantum Electronics*, vol. 47, no. 7, pp. 1613-1625, 2015.

Throughput Analysis for TCP NewReno

Lutfun Nahar

Department of Computer Science and Engineering
University of Chittagong
Chittagong, Bangladesh
lutfacecu@gmail.com

Tarana Nasrin

Department of Computer Science and Engineering
University of Chittagong
Chittagong, Bangladesh
nasrintarana@yahoo.com

Md. Mohibur Rahaman

Department of Computer Science and Engineering
University of Chittagong
Chittagong, Bangladesh
mohib.csecu@yahoo.com

Kazi Ashrafuzzaman

Department of Computer Science and Engineering
University of Calgary
Canada
ashraf@cu.ac.bd

Abstract— Today's Internet suffers from a very complex problem-congestion. It is one of the top-ten listed fundamentally important problems in networking. It causes many important data to be lost. Wastage of network resources results from this. It makes the network easily become gridlocked, with little or no data being transported end-to-end. To protect data transfer from this problem various models of TCP are proposed. These variants differ from each other based on the congestion control algorithm and segment loss recovery techniques they use. One of the techniques is TCP NewReno. In this paper throughput analysis of a simple stochastic model for TCP NewReno is described as a mathematical relation of time that it takes for a signal to be sent plus the length of time it takes for an acknowledgment of that signal to be received and also loss behavior which is based on some previous work done over TCP Reno. It has dissimilarity in three characteristics: model for fast recovery, formulation of timeout behavior, and use of loss event. Here a two-parameter loss model is used that can better represent the diverse packet loss framework encountered by TCP on the Internet. The significant performance advantages of TCP NewReno over TCP Reno are shown using the ns-2 simulator. This performance evaluation is considered in case of two queue management mechanisms- DropTail and RED. The main outcomes from the experiments are: 1) In a wide range of network conditions for TCP NewReno there is an analysis of proposed model which can exactly forecast the stable state throughput 2) At the time of congestion TCP NewReno perform better than the Reno and 3) The packet loss against time during data transmission suffers from less loss rate than the existing Reno model.

Keywords—TCP; Reno;RED; Drop Tail; NewReno

I. INTRODUCTION

In Internet two distinct transport-layer protocols are used- User Datagram Protocol (UDP) and Transmission Control Protocol (TCP). TCP provides reliable data transfer by using flow control, sequence numbers, acknowledgements, and timers. It ensures data is delivered from the sender to the receiver, correctly and in-order. It is a window-based protocol data transferred between sender and receiver is governed by the window size that is advertised by the receiver. It provides full-duplex, unicast data transfer in the Internet. It offers several services to prevent data from lost or corruption. One of

these services is congestion control that prevents a router from overflowing. In Internet a major problem is congestion in the network. This causes many important data to be lost and wastage of network resources. There are two broad approaches to detect network congestion: end-to-end congestion control and network assisted congestion control. The first one performs congestion control based on the network behavior, whereas the later approach uses the feedback provided by the network layer components. The congestion control mechanism of TCP follows the end to end congestion control approach. It dynamically increases or decreases the size of congestion window of packet in transit by observing the network behavior. In TCP for the “best effort” internet the stability is provided by the congestion control schemes. Starvation of flows is prevented by this procedure, increase network utilization, and it also assures the fairness of inter protocol. The congestion control algorithm uses two incidents to detect congestion: triple ACKs and timeout.

TABLE I

Model	Comparison of TCP throughput models(segments per Round Trip Time R	
	TCP Reno[8]	TCP NewReno[1]
Simple (NoTO)	$\frac{\sqrt{\frac{3}{2p}}}{R}$	$\frac{\frac{1}{p} + \frac{W^2 q}{1 + Wq}}{\left(\frac{W}{2} + Wq + 5/2\right) R}$

TABLE II

Model	Comparison of TCP throughput models(segments per Round Trip Time R	
	Full Model	
TCP Reno[8]	$\frac{1}{R\sqrt{\frac{2p}{3}} + RT0min(1,3(\sqrt{(3P/8))P(1 + 32P^2)})}$	
TCP NewReno[1]	$\frac{\frac{1}{p} + \frac{W^2 q}{1 + Wq}}{NR + P_{To}(1 + 2P + 4P^2)R + (1 + \log(W/4) R)}$	

The major algorithms for congestion control is slow-start, congestion avoidance, fast retransmit, and fast recovery.

Depending on the use of these algorithms TCP has several variants: Tahoe [2], Reno [3], Vegas [4], SACK [5], NewReno [1]. In this paper an analysis of throughput for TCP NewReno [1] is described that builds upon the well known Reno model [8]. But model [1] differs from the Reno model [8] in three important ways: 1) fast-recovery algorithm- in the fast-recovery phase NewReno [1] may finish remarkable time depending on the segment loss characteristics 2) timeout behavior- timeout during fast-recovery is also considered, and 3) two-parameter loss model considers both segment losses burstiness and loss event rate.

Table-I&II summarize the Reno [8] and NewReno [1] models. By using of the simple version of Reno [8] overestimates throughput. Again parameterizing the Reno [8] model with packet loss rate alternative to loss event rate tends to underestimate throughput. In some cases, these two opposing errors lead to good predictions. The Full TCP Reno [8] model has been applied extensively in diverse areas TCP friendly rate control, active queue management, and overly bandwidth management. The results from the model [1] show that the Full Reno [8] model highly estimates throughput for Reno and NewReno [1] bulk transfers. Here we proposed an improved fast-recovery algorithm for TCP based on model [1]. The rest of this paper is organized as follows: In segment 2 some relevant materials, segment 3 the improved fast-recovery algorithm for TCP NewReno [1] and some assumption used in this model [1] are described. In segment 4 analysis of the model [1] studied here. It has two parts model without timeout and full model (including both types of loss event). Segment 5 describes the simulation process that is used to compare the performance of the Reno [1] and the NewReno [1] flows.

II. PRELIMINARIES

In this preliminaries many terms are clarified, which is used in the analytic model [1] and simulation result. Huge amount of data are seized and store in a segment has the limitation called Maximum Segment Size (MSS). It is naturally start by first consider the size of the largest link layer frame, Maximum Transmission Unit (MTU). A signal passed between communicating processes, is consider as Acknowledgement (ACK). In TCP an ACK show that a packet is received in order and perfectly. The response (ACK) indicates that one of the packets in transit has left the network and it is safe to input a new packet into the network. But the entry of 4 back-to-back ACK (also called triple duplicate ACK) for the same segment clarify that there is network congestion.

As the Transmission Control Protocol TCP is a connection-oriented, window based transport-layer protocol, it controls a state variable for each and every connection- Congestion Window (cwnd) the maximal amount of data that is used by sender during transmission before receiving an ACK. The receiver also provides a limit, rwnd, on the amount of outstanding data. Data transmission is always monitored by the window size, $W = \min(cwnd, rwnd)$. And cwnd is not

permitted to fall below the 1MSS. One of the variables is cwnd that are used in congestion control algorithms. The 50% of the current cwnd is assumed as a threshold value ssthresh, that are used in the slow-start phase after timeout happens. The Round Trip Time (RTT) is the time that acquire for a packet to travel from client to server and then back to the client, that is the time between the sending of a packet and the reception of ACK for that packet. It covers packet queuing delays, packet propagation delays in instant routers, switches and packet processing delays. The RTT can be estimated as $RTT = (1-\alpha) * old_RTT + \alpha * SampleRTT$. Where the SampleRTT for a segment is the total time for sending the segment and receiving acknowledgement for the segment and old_RTT express the previous RTT. An individual timer is associated in TCP with every transmitted segment. If an ACK is not received in this limited amount of time, that segment is retransmitted by TCP. TCP sets a function for timeout using both the average of RTT and the standard deviation in that average. While triple duplicate ACKs receive or a timeout occurs at a TCP sender, it assumes this event as a packet drop because of network congestion. We use two-parameter to analysis the model [1] loss model- loss event rate that triggers either a fast-recovery or a timeout and burstiness of segment losses that affects the interval of the fast recovery time. Losses of the packet play a major role at the throughput- average rate of successful packet delivery over a link. When packet losing occurs throughput of the connection is decreased. Each router must implement some queuing discipline in a network that monitors how packets are buffered when waiting to be transmitted. There are several technique to control the queue in a router- either drop the receiving packet when the queue is full or drop a packet before the queue is full. The technique to drop packets that arrived at the tail of the queue is known as Drop-tail. Drop-tail is used when there is not adequate memory to buffer an incoming packet. By using this process, when the queue is fulfilled to its maximum size, the newly arriving packets are dropped until the queue has adequate room to accept incoming traffic. Thus this process has several limitations. It shares buffer space with biased among traffic flows. In this methodology when a queue has been filled, the router begins discarding all extra datagrams. The loss of datagrams affects the TCP sender to enter into slow-start, which decrease throughput in that TCP connection until the sender start to receive acknowledgements again and grows its congestion window. A major problem occurs when datagrams from many TCP connections are dropped; causing all of the connected TCP senders enter into slow-start. This happens because, instead of scraping multiple segments at individual's connection, one segment from each connection is scraped by router. Another process of queue management is Random Early Detection (RED) that permits to loss a packet before the buffer becomes full. Two queue length are used by Random Early Detection(RED). One is minimum threshold (minthresh) and the another is maximum threshold (maxthresh). If the average queue length is not greater or equal than the minimum threshold, when a packet is arrived, then the packet is insert to the queue. If the average queue length is greater than the

maximum threshold, or full then the arriving packet is lost. At last if the packet comes, (after finding an average queue length between the minimum threshold and maximum threshold), then the packet is marked or dropped with a probability that is naturally some function of the average queue length, maximum threshold and minimum threshold. Another application-layer protocol-FTP is used by TCP in transport-layer. Using FTP a user can send or receive files to or from a remote host.

By using a clear-text sign-in protocol FTP users may authenticate themselves, generally in the form of a user name and password, but can connect anonymously when the server is ready to permit it. Constant Bit Rate (CBR) is an application-layer protocol which is used for sessions that transport traffic at a constant bit rate, where there is a hierarchy reliance on time synchronization. It generally uses UDP in transport-layer. It is used for constant bit rate audio, video traffic that carrying real-time. By using CBR, a flow is carried across the network that a cell's end-end delay and the fraction of cells that are drop or delivered late are all ensure to be less than corresponding values. The performances of Reno [8] and NewReno [1] are evaluated using ns-2 simulator. The ns-2 simulator is written in C++, with an OTcl (an object oriented extension of Tcl) interpreter shell to execute the input model files. Most network elements in ns2 simulator are developed as classes, oriented fashion. The simulator carries a class hierarchy in C++, and a similar class hierarchy in OTcl. Through the OTcl interpreter users create new simulator objects, and then these objects are mirrored by corresponding objects in the class hierarchy in C++. A traffic generator named `pareto` embodied in the OTcl class `Application/Traffic/Pareto`. It creates traffic according to a Pareto On and Off service. Packets are sent at a constant rate during this time, and no packets are sent during off time. At our paper we analyzed the modified TCP Reno [3] which is called TCP NewReno [1]. In the event of multiple packet losses, it is able to detect multiple packet losses that makes TCP NewReno[1] more efficient than Reno[3]. It uses slow-start, congestion-avoidance algorithm, and fast-retransmit as like as Reno [3], but to recover from multiple losses in a single window, it offers an advanced fast-recovery algorithm. The modified fast-recovery algorithm for TCP NewReno [1] is described in next section.

III. FAST-RECOVERY ALGORITHM FOR TCP-NEWRENO

In this section we described the fast-recovery algorithm that illustrated how TCP NewReno performs better than the Reno [8]. At the time of congestion avoidance if triple duplicate ACKs is received by the sender, then fast-retransmission is performed and enters into fast-recovery. In fast-retransmit the sender retransmits the lost segment, sets the threshold value for slow-start as $ssthresh = cwnd/2$, and sets congestion window, $cwnd = ssthresh + 3$ segments. In fast-recovery, sender's congestion window is growing by one segment when every subsequent duplicate ACK is received, that mark that a segment is reached at the destination. If the

congestion window gives the permission to transmit then the transmission of new segments will happen. In Reno, when a non-duplicate ACK is received by sender it sets congestion window ($cwnd$) to slow-start threshold ($ssthresh$), terminates fast-recovery, and resumes normal congestion avoidance phase. For many segments loss, from the same window of data, Reno may enter and goes to a fast-recovery many times, causing multiple reductions of the congestion window. TCP NewReno [1] alters this fast recovery characteristics of Reno. It differs the non-duplicate ACKs as- full ACK (FA) from partial ACK (PA). In Reno only incomplete ACK is used to stop the fast recovery. But NewReno [1] again transmits the segment one after another in orderly based on the limited ACK, and decrease the congestion window by one less than the number of segments acknowledged by the partial ACK. It recuperates from several parts of segment losses at the same window by again transmitting one dropped segment per RTT, present in fast-recovery until a full ACK is received. After finding a full ACK, NewReno [1] sets congestion window to slow-start threshold, stop fast-recovery, and restart congestion avoidance.

A. Assumptions

In this section we focus on some assumptions, by analyzing the model [1], about the application layer, the sender/receiver, latency Model and the network. Except the loss model of segment, all these characteristic are as previous work based on which this model [1] is established.

Application Layer: The model [1] works on the steady-state throughput for TCP bulk transfers.

TCP Sender and Receiver: This model [1] concludes that the TCP NewReno [1] congestion control algorithm is used by sender. At all-time sender transmits full-sized (MSS) segments whenever the congestion window permits it to do so. Also, the receiver gives one ACK for each received segment, and ACKs are never lost. The model[1] neglect TCP's three-way connection creation phase and also beginning slow start phase due to the congestion avoidance algorithm control through a long-lived TCP bulk data transfer.

Latency Model: The latency of the TCP transfers focus on by measuring "rounds". When the congestion avoidance begins, first round starts and its interval is one round trip time. All other rounds start instantly after the prior round has completed. The interval of special round could be less than one round trip time which stops fast recovery and shift to congestion avoidance. It is also consider that in a round the round duration is much greater than the time needed to transmit segments, and that the round duration is not dependent of the congestion window size. Transmission of Segment may be bursty or arbitrarily spaced within the round.

Loss Model: The model [1] focus on introducing a novel two-parameter segment loss model that captures both the frequency of loss events and the burstiness of segment losses within a loss event.

IV. ANALYTICAL MODEL

The analytical model is developed in two steps – 1) Model without timeout (NoTO) and 2) Model with both timeout and duplicate ACKs (full model). It uses following notations: RTO= Average duration of first timeout in a series of timeouts P= Loss event rate, q= loss rate of segment within a loss event. R= Average RTT, W=Average of the peak congestion window size.

A. Model without timeout(NoTo)

It is considered that by 4 duplicate ACKs, all loss events are detected, no timeout occurs. Each cycle in this model consists of a congestion avoidance period, followed by two factors. One is fast recovery period and the other is detection of segment loss. These cycles is called a (CAFR) named Congestion Avoidance/Fast Recovery Period. Average throughput for NoTo can be written as $T_{NoTo} = S_{CAFR} / D_{CAFR}$. Here, S_{CAFR} =expected number of successfully transmitted segment. D_{CAFR} = expected time duration of this period. In a CAFR period a round starts with congestion avoidance phase. During this time the congestion window increases the number of transmitting segment by one per round. When a loss is detected by three duplicate ACKs, congestion avoidance terminated, fast retransmission occurs, then TCP starts fast-recovery. Here a term is used named drop window that contains segments starting from the first loss segment in around to the segment transmitted before the receipt of the first duplicate ACK. It is supposed that the loss event occurs at the round $W/2+1$ and m segments are lost in the drop window. Then the fast-recovery continues for m RTTs where $W/2$ new segments are sent per RTT. The congestion avoidance phase will be resumed after receiving a full acknowledgement. It is considered that, If two adjacent and identical CAFR periods are found then the expected number of segments S_{CAFR} can be defined as the summation of expected number of segments between two LE, α , and the expected number of segments between first and last loss. Here it is assumed that the value of α is $1/p$. Then S_{CAFR} can be written as, $S_{CAFR} = 1/p + \delta$. Again, in the linear increase phase to find out the value of W by using p and q the S_{CAFR} can be defined as a function of summation of the expected number of segments, S_{LI} , triple ACKs terminate congestion avoidance S_{β} , and during fast recovery more segments transmitted, S_{FR} . Therefore $S_{CAFR} = S_{LI} + S_{\beta} + S_{FR}$.

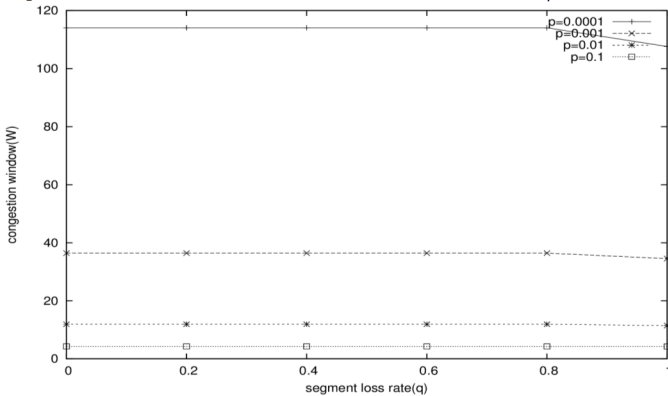


Fig.1. Effect of p and q on Window value W

The equation for congestion window size

$$W = (10pq - 5p + \sqrt{(p(24 + 32q + 49))}) / (p(3 + 4q)) \quad (1)$$

This graph (Fig.1.) indicates congestion window is inversely proportional to the segment loss rate. When p is small the impact of q over W is noticeable but for large it is negligible.

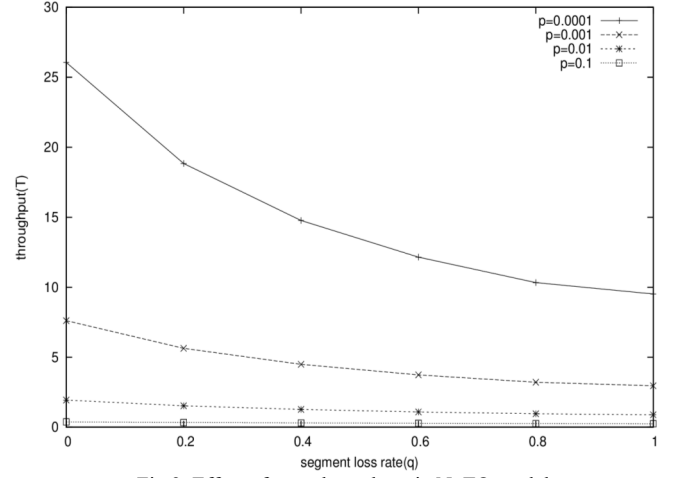


Fig.2. Effect of q on throughput in NoTO model

Fig.2. indicate effect of q on throughput in NoTO model. The equation for NoTO model is given at Table I.

B. Model with both timeout and duplicate ACK(Full model):

In this model besides fast recovery and congestion avoidance phase timeout is considered. Here different CAFR periods are used in each cycle followed by a CATOSS period. CATOSS period contains timeout periods along with congestion avoidance and slow-start. Therefore the throughput can be written as,

$$T_{Full} = (1 - p_{TO}) S_{CAFR} + p_{TO} (S_{CA} + S_{TO} + S_{SS}) / (1 - P_{TO}) D_{CAFR} + P_{TO} (D_{CA} + D_{TO} + D_{SS})$$

Here, P_{TO} = probability of loss event leads to a timeout.

S_X = number of successful S segment transmitted in a period of type X and D_X =duration of a period of type X . From Fig.3 it can be concluded that the segment loss rate has an inverse effect on the throughput. An increasing segment loss rate decreases the throughput. Using equation from Table II segment loss vs. throughput change graph is given below.

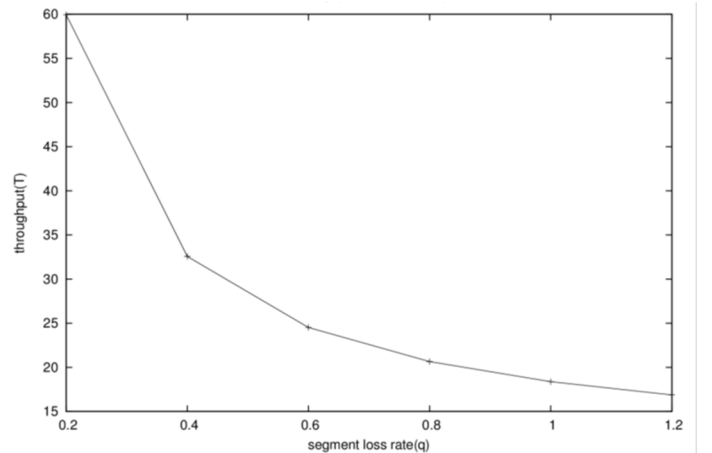


Fig.3. Throughput change against segment loss rate according to the full model

V. SIMULATION RESULTS

In this segment we illustrate TCP NewReno performs how much better than TCP Reno [8], using ns-2 simulator. The simulation is done for both single and multiple bottleneck link. In case of single bottleneck-link two type of queue management mechanisms are used- DropTail and RED. All experiments have two long duration flows: one NewReno flow and the other Reno flow. Here three TCP sources and one UDP source are used. The bottleneck bandwidths vary from 0.7Mbps to 2Mbps. Here two types of traffic flows are used- FTP and constant bit rate UDP flows. The propagation delays vary from 1ms to 20ms. The minthresh and the maxthresh for RED queue management are set to 1/10 and 3/10 of the corresponding queue size limit. Here identical client/server pair is used in each session. The waiting time of a client between the reception of a file and issuing the next request is distributed exponentially and has an average of 0.3ms. Pareto distribution is used to draw the file sizes with a mean of 10KB and shape is 1.5. The maximum congestion window size is 8KB and packet size is 0.5KB. The UDP flow uses a rate of 0.01Mbps, the packet size is 1KB, and the propagation delay is 7ms. The FTP flows begins at times between 0 and 7 seconds.

A. DropTail Queue

In case of Drop-Tail queue router a single bottleneck link is used that has 0.7Mbps bandwidth, buffer limit is of 20 packets. The congestion window is of 8KB, the round trip propagation delay for each TCP flow is 20ms.

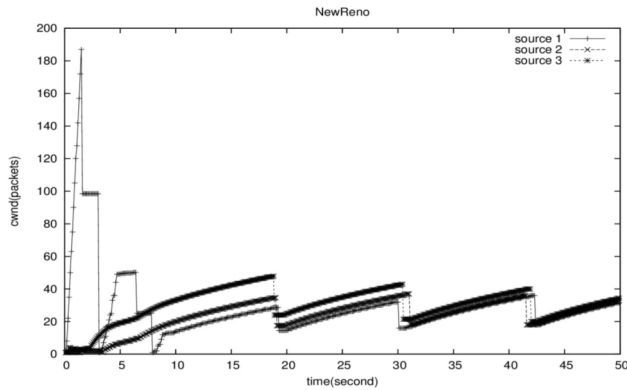


Fig.4. Congestion window change in NewReno

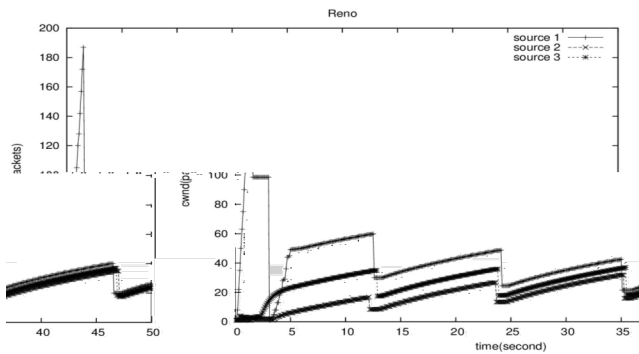


Fig.5. Congestion window change in Reno

In this simulation process the change of congestion window size over time are graphically shown in Fig.4 and Fig.5. From the Fig.5 it is shown that the Reno needs more multiple reductions of the congestion window than the NewReno because of its multiple entries to fast recovery. Comparison of throughput changes over time between these two models has shown in Fig.6. The loss rate scenarios are shown in Fig.7. The simulation result is shown in table III.

TABLE III

Model	Comparison of TCP NewReno over Reno
	<i>DropTail Queue</i>
Throughput	NewReno 15%-25% higher than Reno
Loss rate	Reno gives 2%-3% more loss than NewReno.

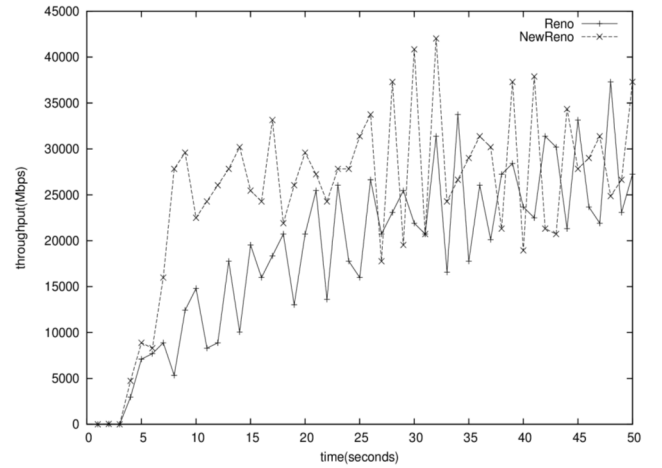


Fig.6. Throughput change over time

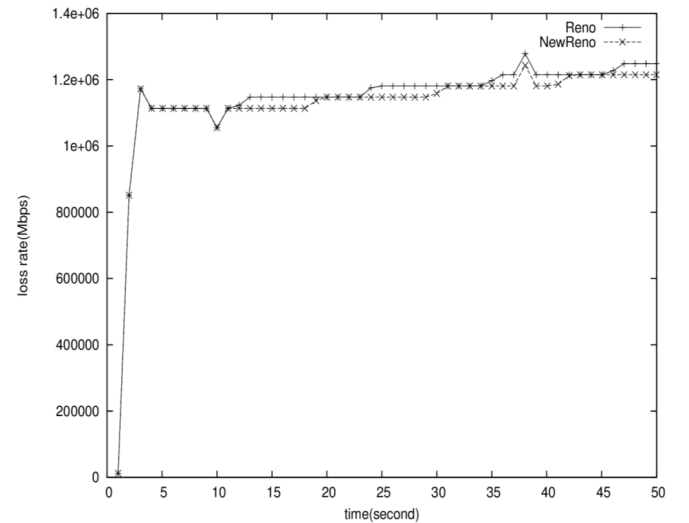


Fig.7. Packet loss rate with time

B. RED Queue

With RED queue the bottleneck link has bandwidth of 2Mbps, propagation delay is 1ms, buffer limit is of 50 packets, the minthresh is set to 5, maxthresh is 15, and congestion window size is 2KB. Here 25 flows are used with per TCP source.

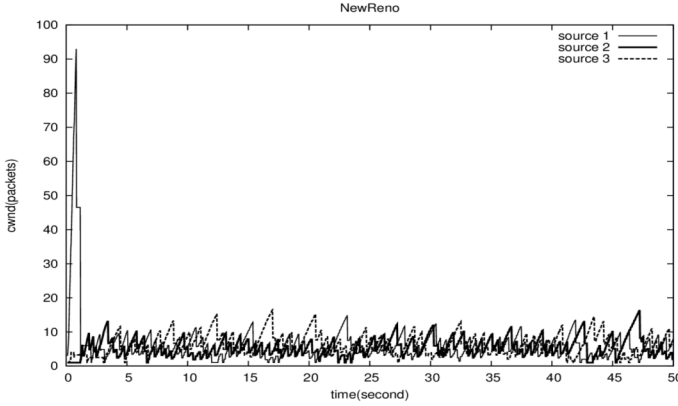


Fig.8. Congestion window change in NewReno

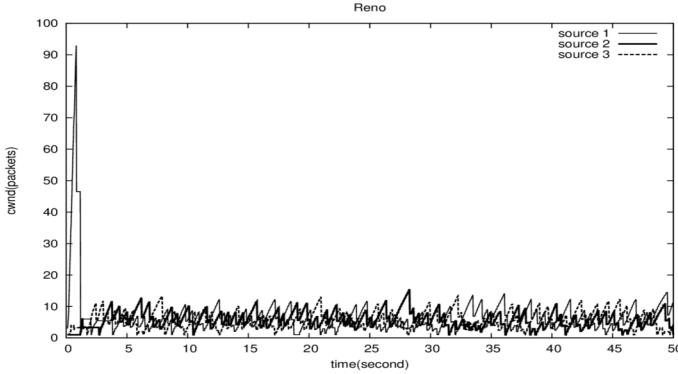


Fig.9. Congestion window change in Reno

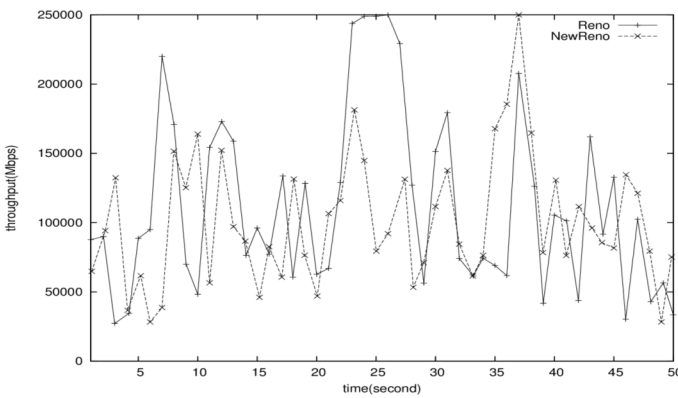


Fig.10. Throughput change over time

The congestion window changes over time during this session are shown in Fig.8 and Fig.9. Comparison of throughput changes over time between these two models has shown in Fig.10. The loss rate scenarios are shown in Fig.11. The simulation results are illustrated at Table IV.

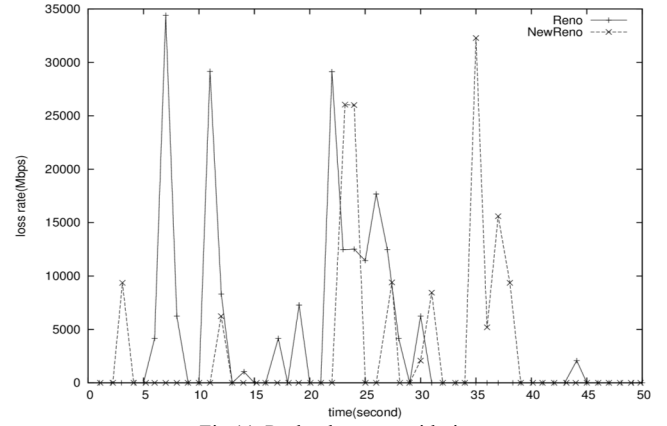


Fig.11. Packet loss rate with time

TABLE IV

Model	Comparison of TCP NewReno over Reno
	RED Queue
Throughput	NewReno 10%-15% higher than Reno
Loss rate	Reno gives 20%-25% more loss than NewReno.

VI. CONCLUSION

In this paper analysis of throughput for TCP NewReno bulk data transfer is discussed based on model [1]. This analysis of model provides a simple and accurate steady state throughput for NewReno in terms of RTT and loss rate. The congestion avoidance mechanism of TCP NewReno is more efficient and utilizes network resources efficiently. Because of its modified fast-recovery algorithm data loss is less than the Reno. In the simulation result the advantages of TCP NewReno over TCP Reno are shown using ns-2 simulator. The performance is evaluated in case of two queue management mechanisms- DropTail and RED. The results from the simulation illustrate the significant performance advantages of NewReno over Reno.

REFERENCE

- [1] Nadim Parvez, Anirban Mahanti, and Carey Williamson, An Analytic Throughput Model for TCP NewReno, IEEE/ACM Transactions on Networking, Vol. 18, No.2, April 2010.
- [2] V. Jacobson. Congestion Avoidance and Control. In Proc. Of ACM SIGCOMM, pages 314-329, Stanford, CA, USA, August 1988.
- [3] V. Jacobson. Berkeley TCP evolution from 4.3-Tahoe to 4.3 Reno. In Proc. Of the 8th IETF, Vancouver, Canada, August-1990.
- [4] L. Brakmo, S. O'Malley, and L. Peterson. TCP Vegas: New Techniques for Congestion Detection and Avoidance. In Proc. of ACM SIGCOMM, pages 24-35, New York, USA, August 1994.
- [5] K. Fall and S. Floyd. Simulation-Based Comparisons of Tahoe, Reno, and SackTCP. ACM Computer Communication Review, 26(3):5-21, July 1996.
- [6] C. Samios and M. Vernon. Modeling the Throughput of TCP Vegas. In Proc. of ACM SIGMETRICS, San Diego, USA, June 2003.
- [7] A. Misra and T. Ott. The Window Distribution for Idealized TCP Congestion Avoidance with Variable Packet Loss. In Proc. of IEEE INFOCOM, pages 1564-1572, New York, USA, March 1999.
- [8] J. Padhye, V. Firoiu, D. Towsley, and J. Kurose. Modeling TCP Throughput: A Simple Model and its Empirical Validation. In Proc. of ACM SIGCOMM, Vancouver, Canada, September 1998.

Effects of Higher Concentration of Hole Transport Material on Organic Light Emitting Diode Performance

Riku Chowdhury^{1*}, Md. Rashedul Haq¹, Md. Sarwar Uddin Chowdhury², Sharmin Afrose¹ and Sukanta Paul³

¹Department of Electrical & Electronic Engineering, FSET, USTC, Chittagong, Bangladesh

²Department of Electrical & Electronic Engineering, CUET, Chittagong, Bangladesh

³Department of Computer Science & Engineering, FSET, USTC, Chittagong, Bangladesh

*Corresponding email ID: eeriku@gmail.com

Abstract— The effects of higher hole transport material (HTM) composition on single emissive layer (EML) based organic light emitting diode (OLED) performance are investigated quantitatively. Due to higher ratio of HTM in EML, the turn-on voltage is reduced from 3.5 V to 2.7 V. In parallel, OLED luminescence and current efficiency also increased with higher HTM concentration. This improvement of OLED performance because of higher HTM concentration in mixture of transport materials has been explained in terms of charge transport, carrier recombination and charge balancing in EML.

Keywords— mixed-host composition; charge density; electric field; carrier recombination; charge balancing

I. INTRODUCTION

First emission of light from ‘anthracence crystal’ (organic material) was observed by applying more than 100 V_{dc} in 1963 [1]. But the research did not generate interest to produce light from organic material due to require very high input voltage which quickly reduced the device lifetime and performance. In 1987, C. W. Tang and S. A. Vanslyke [2] group reported on the first bi-layer organic light emitting diode (OLED) using a relatively low DC voltage (under 10 V. In that device, tris (8-hydroxyquinoline) aluminum (Alq3) was deposited on an aromatic diamine layer by the vacuum evaporation technique. This triggered an intensive research based on the bi-layer OLED, which consisted of a hole-transport layer (HTL) and an electron-transport layer (ETL) between a low work-function metal cathode and a high work-function anode. A sharp barrier at the HTL/ETL interface used to confinement charge carriers. This increased device efficiency. However, chemically unstable cations were formed in the light emitting layer (EML) due to accumulated holes, which increased the formation of non-radiative trapping centers leading to decay in luminescence. This increased the operating voltage and reduced lifetime [3]. To date, the reliability of OLED remains as one of the challenging issues for the researchers. To improve the OLED performance, efficiency and life time of OLED, as well as a high external quantum efficiency (η_{EQE}) and high power efficiency (η_p), multilayers OLED was fabricated by the researchers. This device structure typically consisted of an electron-transport, hole-transport and light transport layers for improving the charge injection, transport,

balance and recombination [4]. Lower operating voltage was achieved in multilayer OLED by introducing conductive materials for doping and injection layers [4-5].

However, the fabrication process of multilayers OLED is time consuming and also it is a difficult process. Therefore, researcher reduced the number of organic layers in the OLED and modified the EML structure by using a mixture of HTM and EML in a single layer, which is known as a mixed-host (MH) device. The ratios of ETM and HTM into the MH-EML can be controlled in the fabrication to result in difference in the carrier concentration profile. It was reported that uniform mixture (UM) of hole and electron transport materials in a single layer could increase the operation lifetime of OLED [6-7]. In this UM-OLED, the HTL/EML interface was eliminated and it broadened the light emission zone. The operation lifetime was increased by 6 times as compared to the conventional, bi-layer OLED [8].

It may be clear from the above that further improvement in OLED efficiency and lifetime over conventional heterojunction (HJ) structure can be obtained with uniform MH structure. However, there remain free parameters in these MH-OLEDs for proper optimization of MH concentration through the experimental investigations of the fundamental physics [9-10]. And numerical simulations of these MH-OLEDs may relieve some costly and time-consuming experiments with better understanding of the underlying mechanisms.

Motivating from this point of view, a numerical simulation based work has been presented to optimize the mixed-host based UM-OLEDs performance from the different composition of HTM and ETM in MH-EML. At first, current density-luminescence-voltage characteristic and efficiency curves of three different MH based OLEDs are simulated and analyzed their device performances. After that, the numerical simulation of charge density profile, distribution of electric field and carrier recombination rate in the EML those are difficult to measure directly through experiments, able to investigate the underlying MH device mechanisms those are leading to improve device performance and increase the operation lifetime.

II. MH-OLED MODELLING

Drift-diffusion electrical model has been implemented through a numerical simulator called ‘SETFOS’ (Semiconducting Thin Film Optics Simulation Software) for analyzing the electrical characteristics of single mixed-host emissive layer based OLED. The Poisson equation and continuity equation as presented in “(1),” and “(2),” correspondingly considered in this numerical simulation as electrical model to solve the electric field distribution and time evaluation of the MH device.

$$\frac{dE}{dx} = \frac{q}{\epsilon\epsilon_0}(p-n) \quad (1)$$

$$\frac{dn}{dt} = \frac{dJ_n}{qdx} - R - T \quad (2)$$

where, E is the electric field inside the OLED, q the elementary charge, ϵ the dielectric permittivity of the organic materials, ϵ_0 the dielectric constant, p and n are the hole density and the electron density respectively, J_n the electron current density, R the recombination rate and T the trapping rate for free carriers. The carrier density distribution can be obtained by solving the drift and diffusion equation based on the Langevin recombination (R) process those are calculated using “(3),” and “(4),” respectively. The conventional electric field dependent ‘Poole-Frenkel mobility (PFM)’ model as expressed in “(5),” applied for carrier mobility calculation in the device. The current balance factor (CBF) (b) in “(6),” used to measure the device quantum efficiency [11] which is defined as the recombination current with the total current density.

$$J_n = qp\mu_n E + qD_n \frac{dn}{dx} \quad (3)$$

$$R = \frac{q}{\epsilon\epsilon_0}(\mu_p + \mu_n) \quad (4)$$

$$\mu(E) = \mu_0 \exp(\gamma \sqrt{E}) \quad (5)$$

$$b = \frac{\int_0^L qR(x)dx}{J_T} \quad (6)$$

Here, μ_p and μ_n are the hole and electron mobility respectively, D_n the diffusion constant, $\mu(E)$ is the charge carrier mobility under electric field (E), μ_0 is the mobility under zero electric field, γ is the Poole-Frenkel field dependent factor, J_T is the total current density and L is the total device thickness.

III. STRUCTURES OF DIFFERENT MH-OLEDs AND INPUT PRAMETERS FOR DEVICE SIMULATION

For numerical simulation, N,N'-diphenyl-N,N'-bis(1,1'-biphenyl)-4,4'-diamine (NPB) and Tris-(8-hydroxyquinoline) aluminum (3) (Alq₃) are used as HTM and ETM materials correspondingly, LiF/Al as cathode and indium-doped tin oxide (ITO) as anode. Three uniform MH-OLEDs, namely S₁,

S₂ and S₃ where HTM and ETM are mixed in the single EML with three uniform compositions: 2:3 (HTM:ETM), 3:3 (HTM:ETM) and 3:2 (HTM:ETM) correspondingly. The basic structure of three MH-OLEDs is as follow: ITO / NPB : Alq₃ (100 nm) /LiF/Al. From device S₁ to device S₃, the concentration of HTM (NPB) is increased in the EML and reverse is true for ETM (Alq₃) composition. The detail of device structure is shown in Figure 1.

The parameters setting in UM-EML is the main problem due to the mixture of bi-polar transport materials. It is reported [12] that HOMO and LUMO in mixed-host EML may not be changed for hole-transport (HTM) and electron-transport materials (ETM), respectively. Following the experimental observations [13], the mobilities in the EML are taken to be power-dependent of the concentration of mixture, therefore Equations (6) and (7) are used to calculate these mobilities:

$$\mu_{mix,p} = \mu_{HTM,p}^{C_{HTM}} * \mu_{ETM,p}^{(1-C_{HTM})} \quad (6)$$

$$\mu_{mix,n} = \mu_{HTM,n}^{C_{HTM}} * \mu_{ETM,n}^{(1-C_{HTM})} \quad (7)$$

where $\mu_{mix,p(n)}$ is the hole (or electron) mobility in mixed-host layer, $\mu_{HTM,p(n)}$ the hole (or electron) mobility of hole-transport material (HTM), and $\mu_{ETM,p(n)}$ is the hole (or electron) mobility of electron-transport material (ETM). C_{HTM} is the concentration of hole transport material in the EML. The performances of UM-OLEDs are simulated and analyzed from the numerical results after defining the energy level and mobilities of the mixed host layer. The energy level and calculated mobilities of the mixed-host materials in the emissive layer using the mobility of pure electron transport material (ETM) and hole transport material (HTM) are listed in Table 1.

Parameters	NPB [13]	Alq ₃ [13]	MH Device		
			S ₁	S ₂	S ₃
HOMO (eV)	-5.4	-5.8	-5.4	-5.4	-5.4
LUMO (eV)	-2.4	-3.1	-3.1	-3.1	-3.1
μ_{p_0} (cm ² /Vs)	2.6×10 ⁻⁴	1.4×10 ⁻⁶	1.1×10 ⁻⁵	1.9×10 ⁻⁵	3.2×10 ⁻⁵
γ_{p_0} (cm ^{1/2} /V ^{1/2})	1.3×10 ⁻³	2×10 ⁻⁴	1.4×10 ⁻³	1.4×10 ⁻³	1.4×10 ⁻³
μ_{n_0} (cm ² /Vs)	1.4×10 ⁻⁶	2×10 ⁻⁸	1.1×10 ⁻⁷	1.6×10 ⁻⁷	2.5×10 ⁻⁷
γ_{n_0} (cm ^{1/2} /V ^{1/2})	2.8×10 ⁻⁴	9.9×10 ⁻³	1.4×10 ⁻³	1.4×10 ⁻³	1.4×10 ⁻³
ITO (eV)	- 4.8 [13]				
LiF-Al (eV)	- 3.2 [13]				

TABLE I. INPUT PARAMETERS FOR MH-OLEDs SIMULATION.

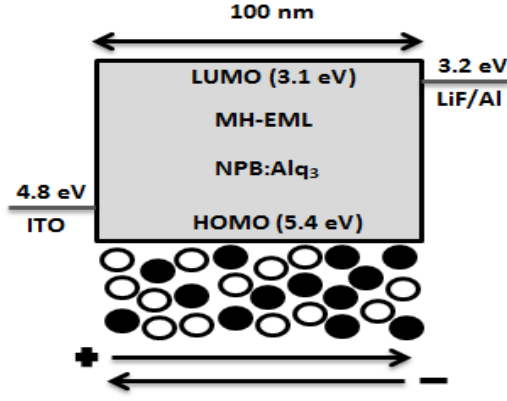


Fig. 1. Schematic Diagram of the compositions of HTM (filled circle) and ETM (open circle) in the uniformly MH-OLED.

IV. RESULTS AND DISCUSSIONS

A. J-V-L Characteristic and Efficiency Curves

Figure 2 (a) shows the current density-voltage-luminance (J-V-L) characteristics curve of three uniformly MH-OLEDs. Under the same operating voltage, the current density increased significantly from device S_1 to S_3 where the concentration of ETM decreased from 60% to 40%. The turn-on voltage (the voltage as measurable luminescence of 1 cd/m^2) decreased from 3.5 V for device S_1 to 2.7 V for device S_3 . The luminescence-voltage characteristic is related to the current density. At 8V, the luminescence (current density) of device S_1 , S_2 and S_3 are 1820 cd/m^2 (190 mA/cm^2), 4000 cd/m^2 (293 mA/cm^2) and 6806 cd/m^2 (490 mA/cm^2) respectively. Due to optimum MH ratio, device S_3 shows the maximum current and power efficiencies (Figure 2 (b)) as like as maximum luminescence and current density over device S_1 and S_2 . The optimum performance of device S_3 can be explained from the simulation results of charge concentration, carriers current density, distribution of electric field, carrier recombination rate and current balance factor in the following section.

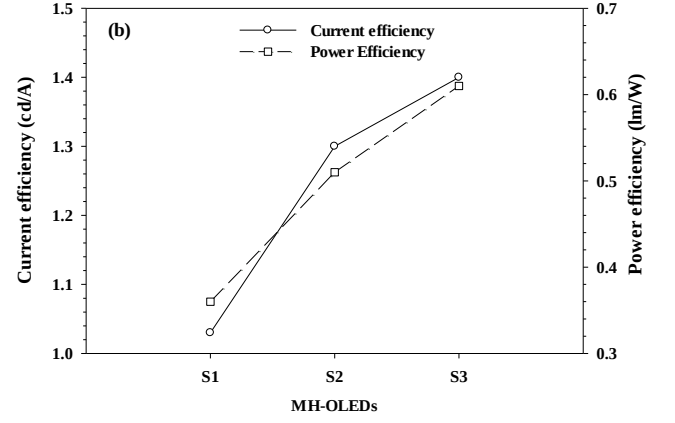
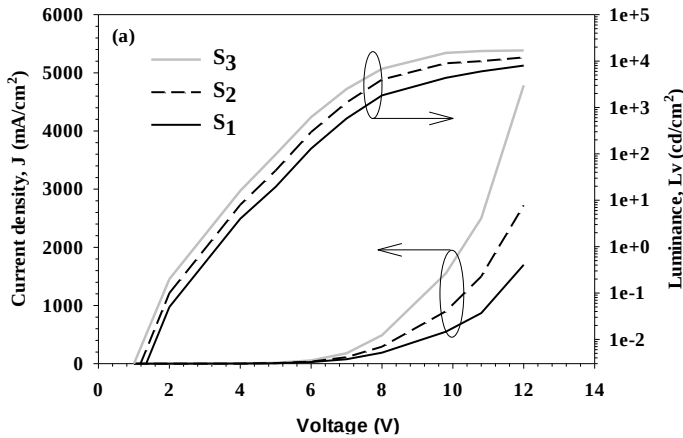


Fig. 2. Characteristics curves of MH-OLEDs in terms of (a) Current density-Voltage- Luminance and (b) Maximum Current (Open circle) and Power Efficiency (Open Square)

B. Effects of different MH compositions on charge transport, recombination and balancing

Figure 3 (a) shows the density of hole and electron in the bulk organic EML at 8V. In device S_1 , a higher concentration of ETM increased the magnitude of electron mobility than the hole mobility through the EML. Therefore, the average hopping distance for electrons are decreased and increased the hopping distance for holes. Hence, electrons are accumulated easily at the ITO/EML interface and increased the electron density with respect to the other MH-OLEDs (S_2 and S_3). Due to the balanced mixture of HTM and ETM (3:3), the magnitudes of hole and electron mobility are similar in device S_2 maintained the charge density in between the charge density of device S_1 and device S_3 as they have unbalanced mixture of transport materials. The scenario of carriers mobility in device S_3 (HTM:ETM=3:2) is reversed with respect to device S_1 due to shorter hopping distance for holes which results of higher hole density at EML/LiF-Al interface.

The distribution of electric field and carrier recombination rate through the EML are shown in Fig. 3(b). In device S_1 , the number of carriers accumulation at the ITO/EML interface to middle of EML is higher with respect to device S_3 . Therefore, along this EML length, the generation of electric field is higher in case of device S_1 which is result of higher carriers recombination in this corresponding region. But the scenario of charge accumulation becomes reverse at the EML/cathode interface for device S_1 . As result, the magnitude of electric field generation and recombination rate inverted for device S_1 with respect to device S_3 . As balanced mixture of transport materials in device S_2 , the electric field and carrier recombination rate are kept in the middle of device S_1 and S_3 .

Moreover, an area-integration for assessing the total recombination rate under respective curves is conducted which give S_1 : $1.6 \times 10^{29} \text{ cm}^{-3} \text{ s}^{-1}$; S_2 : $1.8 \times 10^{29} \text{ cm}^{-3} \text{ s}^{-1}$ and S_3 : $2.1 \times 10^{29} \text{ cm}^{-3} \text{ s}^{-1}$. The computed characteristic of CBFs are shown in Fig. 3(c) also indicates that the performance of different MH-OLEDs followed as $\text{CBF}_{S3} > \text{CBF}_{S2} > \text{CBF}_{S1}$. These two

parameters values: total recombination rate and CBFs agreed with the earlier ranking of MH-OLEDs performance by the current efficiency (Fig 3): $\gamma (S_3) > \gamma (S_2) > \gamma (S_1)$.

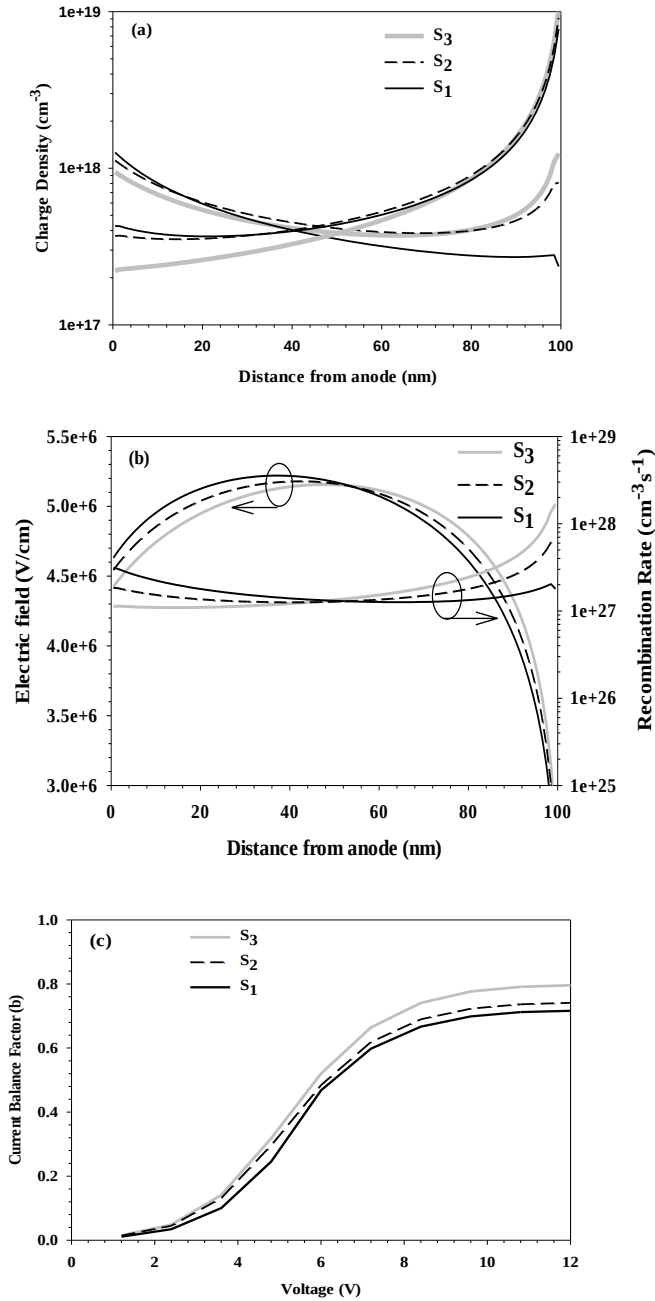


Fig. 3. (a) Carriers concentration profile at 8V, (b) Distribution of Electric field and Recombination zone at 8V and (c) Current Balance Factor for three MH-OLEDs

C. Discussions

Through this numerical simulation, it is clearly visualized that different MH composition in EML influenced the OLED performance by changing the carrier transport and recombination. Especially, higher composition of HTM in MH

ratio increased the overall carrier recombination rate by balancing the hole and electron current density in EML as evidence of higher CBF of around 0.8 in Fig. 3(c). Hence, the device S₃ which EML consists of around 60% of HTM, presented the best performance in terms of luminescence and both current and power efficiencies over other MH-OLEDs.

V. CONCLUSIONS

The effects of different MH compositions on carrier transport, balancing and recombination in single EML-OLEDs has been presented by the numerical simulation. A quantitatively electrical model, based on PFM model, is used to analyze the performances of three MH-OLEDs. Among them, higher concentration of HTM based MH-OLED shows the superior performance over rest of devices. Higher hole density in this MH device maintain higher hole current density which leading towards to generation of higher electric field, good carrier recombination and higher charge balancing factor which is proportional to the experimental quantum efficiency.

REFERENCES

- [1] Pope, M., Kallmann, H. P., & Magnante, P., "Electroluminescence in Organic Crystals," *The Journal of Chemical Physics*, vol. 38, no. 8, p 2042, 1963.
- [2] C. W. Tang and S. a. VanSlyke, "Organic electroluminescent diodes," *Appl. Phys. Lett.*, vol. 51, no. 12, p. 913, 1987.
- [3] Kondakov, D. Y., Sandifer, J. R., Tang, C. W., & Young, R. H., "Nonradiative recombination centers and electrical aging of organic light-emitting diodes: Direct connection between accumulation of trapped charge and luminance loss," *Journal of Appl. Phys.*, vol. 93, no.2, p.1108, 2003.
- [4] Chopra, N., Swensen, J. S., Polikarpov, E., Cosimbescu, L., So, F., & Padmaperuma, A. B., "High efficiency and low roll-off blue phosphorescent organic light-emitting devices using mixed host architecture," *Applied Physics Letters*, vol. 97, no. 3, p. 033304, 2010.
- [5] Quirino, W. G., Teixeira, K. C., Legnani, C., Calil, V. L., Messer, B., Neto, O. P. V., Cremona, M., "Improved multilayer OLED architecture using evolutionary genetic algorithm," *Thin Solid Films*, vol. 518, no. 5, p. 1382–1385, 2009.
- [6] Chwang, A. B., Kwong, R. C., & Brown, J. J., "Graded mixed-layer organic light-emitting devices," *Applied Phys. Lett.*, vol. 80, no. 5, p. 725, 2002.
- [7] Chwang, A. B., Kwong, R. C., & Brown, J. J., "Graded mixed-layer OLEDs," *International Symposium on Optical Science and Technology*, pp. 55–61, 2003.
- [8] Kalinowski, J., Palilis, L. C., Kim, W. H., & Kafafi, Z. H., "Determination of the width of the carrier recombination zone in organic light-emitting diodes," *Journal of Appl. Phys.*, vol. 94, no. 12, p. 7764, 2003.
- [9] Kim, K., & Won, T., "Numerical Simulation on the Charge Transport in Tris(8-hydroxyquinolino)aluminum-Based Organic Light-Emitting Diode Structure," *Japanese Journal of Appl. Phys.*, vol. 52, no. 10, 2013.

- [10] E. Knapp and B. Ruhstaller, "Numerical analysis of steady-state and transient charge transport in organic semiconductor devices," *Opt. Quantum Electron.*, vol. 42, no. 11–13, pp. 667–677, 2011.
- [11] J. C. Scott, S. Karg, and S. A. Carter, "Bipolar charge and current distributions in organic light-emitting diodes," *J. Appl. Phys.*, vol. 82, no. 3, p. 1454, 1997
- [12] H. Wang, K. P. Klubek, and C. W. Tang, "Current efficiency in organic light-emitting diodes with a hole-injection layer," *Appl. Phys. Lett.*, vol. 93, no. 9, p. 093306, 2008
- [13] S.-W. Liu, J.-H. Lee, C.-C. Lee, C.-T. Chen, and J.-K. Wang, "Charge carrier mobility of mixed-layer organic light-emitting diodes," *Appl. Phys. Lett.*, vol. 91, no. 14, p. 142106, 2007.

Computational Overhead Reduction and Side Information Free Transmission with Low PAPR for Riemann Sequence Based OFDM System

Md. Moshir Rahman*, Md. Nur Al Safa Bhuiyan[†], Muhammad Sajjadur Rahim[‡] and Sabbir Ahmed[§]

Dept. of Information and Communication Engineering, Faculty of Engineering

University of Rajshahi, Rajshahi-6205, Bangladesh

Emails: *mintu_apee@yahoo.com, [†]safa@ru.ac.bd, [‡]sajid_ice@ru.ac.bd, [§]sabbir@ru.ac.bd

Abstract—In Orthogonal Frequency Division Multiplexing (OFDM) system, high value of Peak-to-Average Power Ratio (PAPR) is an operational problem that may cause non-linear distortion resulting in high Bit Error Rate (BER). Selected Mapping (SLM) is a well known PAPR reduction technique that shows good performance but inflicts added computational overhead. In this paper, using Riemann sequence based SLM method, we apply reverse searching technique to find out low PAPR yielding phase sequences with significant reduction in computation complexity. In addition, we show side-information free transmission is possible that supports higher throughput with very little sacrifice in PAPR reduction. simulation results show that the proposed method achieves reduced computational complexity as well as side information free transmission with low PAPR.

Keywords—OFDM, PAPR, SLM, Riemann Matrix, Computational Overhead and Side Information.

I. INTRODUCTION

Orthogonal Frequency Division Multiplexing (OFDM) is an attractive technique for future wireless transmission systems which supports high spectral efficiency, immunity from inter-symbol interference (ISI) and robustness against multi-path channel fading [1]. A receiver can recover the transmitted data without interference. Specially, high data rate applications are more suitable with this OFDM technique. Because of these advantages, OFDM is used in many existing wireless and cable based standards such as IEEE 802.11a standard for Wireless Local Area Networks (WLAN), IEEE 802.16 standard for Wireless Metropolitan Area Network, Terrestrial Digital Video Broadcasting (DVB-T), Very-high-bit-rate digital subscriber line (VDSL) [2] and so forth.

However OFDM is not free from limitations. One critical limitations associated with it is high PAPR. High PAPR signal requires a large range of dynamic linearity for analog circuits and eventually lowers the efficiency of the power amplifiers. Thus, it is very important to reduce PAPR of the signal before transmission. Regarding this, researchers have proposed various PAPR reduction techniques [3]. These techniques can be broadly classified into signal scrambling techniques and signal distortion techniques. Block coding, Selected Mapping (SLM) and Partial Transmit Sequences (PTS) fall into signal

scrambling techniques [4, 5], whereas amplitude clipping, companding etc. represent signal distortion techniques [6].

Among the above mentioned techniques, due to the achievement of better PAPR reduction without causing signal distortion, SLM scheme is relatively attractive. In this scheme, the choice of proper phase sequence to achieve good PAPR reduction is very important; hence researchers have proposed several ways to generate such sequences. For example, random values from the set $(\pm 1, \pm j)$ was chosen in the very first work on SLM [4]. Later a significant number of phase sequences were proposed. These includes Riemann Matrix, Hadamard Matrix, Hilbert Matrix based phase sequence and so on [7–9]. In our previous work, Conde and GCD matrix based phase sequence were introduced which showed good PAPR reduction with low computational overhead [10, 11].

Previous research results show that the SLM technique using 64-length Riemann phase sequence gives improved results regarding PAPR reduction capability [7]. But the results also show that it takes many iteration to find out the proper phase sequence for a required level of PAPR reduction. Higher number of iterations implies higher computational overhead, which in turn implies higher power consumption. Motivated by Riemann sequence's remarkable capability of PAPR reduction, we were interested in exploring the possible ways of reducing the associated computational overhead. Looking into existing Riemann sequence based PAPR reduction works, we observed that to find out the proper phase sequence, searching is performed in forward direction [12]. Here, forward direction implies that index of Riemann sequence starts at 1 and increases at each iteration. For both minimum value oriented PAPR reduction and pre-defined threshold value oriented PAPR reduction, we observed that preferred phase sequences were almost always found in the last quadrants, i.e., indices 40 to 61. Hence, we came up with an idea that if searching is started from the ending phase sequence of 64 and finished at starting phase sequence of 1, there is a higher probability of finding out proper phase sequences with less computational complexity. This is called reverse searching technique in our current research work. In this reverse searching, for 64-length Riemann sequence, the

starting index is 64 which is the maximum and ending index is 1. Our simulation outcome revealed significant reduction of computational complexity for the proposed searching technique.

We are also interested in finding out the default index of Reimann sequence. By default index, we indicated the Reimann sequence index that satisfies minimum PAPR threshold for maximum number of OFDM data blocks. The underlying motivation is, if such sequence exists, the receiver can be informed about it a-priori. And as a result, there would be no need of sending the information about selected phase sequence as side information with every data block. This will save valuable data throughput, otherwise lost due to side information. Now, several researches can be found to determine the default index [13]. However, in this research, the default index can be found using our proposed reverse searching technique with less computational complexity.

In the proposed method, through simulations, we obtained a reduction in PAPR of 8.6 dB and 7.8 dB compared to the original OFDM signal considering minimum PAPR and with default phase sequence, respectively.

II. SYSTEM MODEL AND PAPR

Let us consider an OFDM system consisting of N sub-carriers in the frequency domain, where, vector $X = [X_0, X_1, \dots, X_{N-1}]$ denotes the input data in blocks. Each symbol in X is used to modulate a sub-carrier and the input symbol vector is also called the input symbol sequence. For N -th sub-carrier frequency, let $f_k, k = 0, 1, \dots, N-1$. In the OFDM system as the adjacent sub-carriers are orthogonal to each other, let, $f_k = k \cdot \Delta f$, where, $\Delta f = 1/(NT)$ and T is the symbol duration. Therefore, the complex baseband representation of the transmitted OFDM signal is given by:

$$x(t) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_k e^{j2\pi k t \Delta f}, 0 \leq t \leq NT \quad (1)$$

where, $j = \sqrt{-1}$, Δf is the sub-carrier spacing, and NT denotes the useful data block period. In OFDM the sub-carriers are chosen to be orthogonal (i.e., $\Delta f = 1/NT$). If $x = [x_0, x_1, \dots, x_{N-1}]$ denotes a discrete time OFDM signal vector, then x corresponds to the Inverse Fast Fourier Transform (IFFT) of X , that is, $x = QX$, where, Q is the IFFT matrix.

PAPR is defined as the ratio between the maximum power and the average power for the envelope of a baseband complex signal during an OFDM symbol period. The PAPR of the OFDM signal $x(t)$ can be written as follows:

$$\text{PAPR} = \frac{\max_{0 \leq t \leq T} |x(t)|^2}{E(|x(t)|^2)} \quad (2)$$

where, $E[\cdot]$ is the expectation operator which denotes the average value over the time duration of OFDM symbol. The

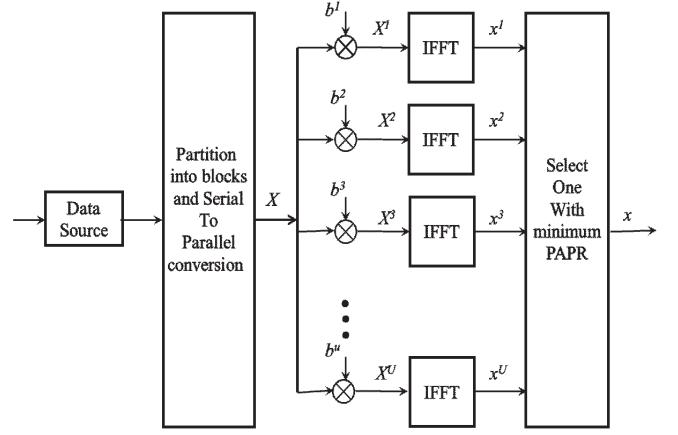


Fig. 1. Block diagram of OFDM transmitter with SLM.

PAPR reduction techniques are concerned with reducing $\max |x(t)|$.

Some of the important performance factors that should be considered while reducing the PAPR are Complementary Cumulative Distribution Function (CCDF), spectral spreading, bit error rate, computational complexity, data rate loss etc.

III. SELECTED MAPPING TECHNIQUE

Selected Mapping (SLM) is one of the most attractive methods of PAPR reduction. In this technique, a set of sufficiently different candidate data blocks are generated at the transmitter, all representing the same information as the original data blocks, and then the data block with least PAPR is selected for transmission [4]. In Fig. 1, the block diagram of SLM is shown. Here, each data block is multiplied by U number of different phase sequences, each of length N , resulting in U number of modified data blocks. Among the modified data blocks, the one with the lowest PAPR is selected for transmission.

Let us consider the input data stream as a vector $X = [X_0, X_1, \dots, X_{N-1}]$. U -modified data blocks can be generated by multiplying each input data block with U different phase sequences, each of length N . Mathematically modified data blocks can be represented by $B^{(u)} = [b_{u,0}, b_{u,1}, \dots, b_{u,N-1}]$, $u = 1, 2, \dots, U$. The modified data block for the u th phase sequence can be written as $X^{(u)} = [X_0 b_{u,0}, X_1 b_{u,1}, \dots, X_{N-1} b_{u,N-1}]$, $u = 1, 2, \dots, U$, and after applying Inverse Fast Fourier Transform (IFFT) the multicarrier signal becomes $x^{(u)}$.

Finally, the one with the lowest PAPR is selected for transmission. Receiver needs to know this selected phase sequence to recover the original data. Hence, along with the signal information, selected phase sequence is transmitted as side information. Different types of phase sequence techniques can be used in SLM approach [4].

IV. RIEMANN PHASE SEQUENCE GENERATION

The columns of normalized Riemann matrix can be selected as phase sequence set in SLM [7]. The Riemann Matrix [14] is obtained by removing the first row and first column of the matrix A , where

$$A(i, j) = \begin{cases} i - 1 & \text{if } i \text{ divides } j \\ -1 & \text{otherwise} \end{cases}$$

The Riemann matrix has the following properties:

- Each eigenvalue $e(i)$ satisfies $\text{abs}(e(i)) \leq m - 1/m$, where $m = n + 1$.
- $i \leq e(i) \leq i + 1$ with at most $m - \text{sqrt}(m)$ exceptions.
- All integers in the interval $(m/3, m/2)$ are eigenvalues.

Using above equation, Riemann Matrix (A) of order 4 can be written as:

$$A = \begin{bmatrix} 1 & -1 & 1 & -1 \\ -1 & 2 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ -1 & -1 & -1 & 4 \end{bmatrix} \quad (3)$$

From the above equation $A^{(u)} = [A_0^{(u)}, A_1^{(u)}, \dots, A_{N-1}^{(u)}]$ can be defined with each row of the normalized Riemann matrix A . Each input data block $X = [X_0, X_1, \dots, X_{N-1}]$ is multiplied with U different phase vectors $A^{(u)}$, $u = 1, 2, \dots, U$. Then, all U subcarrier vectors are transformed into time domain to get the following:

$$x^{(u)} = \text{IDFT}\{X^{(u)}\} = \text{IDFT}\{X_{(n)} \cdot A_n^{(u)}\} \quad (4)$$

Finally, the one with the minimum PAPR is selected for transmission.

V. PROPOSED METHOD

This section is divided into two sections. These are: Proposed method and Determination of default index.

A. Proposed Method

In the proposed method, following process are repeated for reverse direction.

- The input data streams are mapped as a vector symbols $X_0, X_1, X_2, \dots, X_{N-1}$ into N sub-blocks.
- Each block is multiplied (point wise multiplication) by B Riemann phase sequence vector, each of length u . Mathematically, $B^{(u)} = [b_{u,0}, b_{u,1}, \dots, b_{u,N-1}]$ and the modified data block for the u th phase sequence can be written as $X^{(u)} = [X_0 b_{u,0}, X_1 b_{u,1}, \dots, X_{N-1} b_{u,N-1}]$, $u = 64, 63, \dots, 1$. This generates a modified data block and its PAPR is calculated.
- A particular data block is selected for transmission, and the phase index value is used as side information, if $\text{PAPR} \leq \text{Threshold}$.
- The process is repeated for the same data block with the different phase vector at the decremented index, if $\text{PAPR} > \text{Threshold}$.

B. Determination of Default Index

Investigation is carried out in our research to find out the possibility of side-information free PAPR reduction mechanism using the Riemann Sequence. For this, we tried to single out one particular Riemann sequence that would yield minimum PAPR for maximum number of OFDM symbols. We name the corresponding Riemann sequence index as the Default index. The underlying motivation is, if such sequence is found, the receiver can be informed about the default sequence only once at the start of the transmission. Thus, it is not required to send the sequence number used as side-information with every OFDM symbol and thereby improve the system efficiency. However, on the downside, we may have to sacrifice PAPR reduction capability.

VI. SIMULATION RESULTS

We performed simulations to compare the performance of SLM technique using Bauml and Riemann phase sequence with original OFDM system (OFDM with no PAPR reduction applied). The simulation parameters are listed in Table I. In Fig. 2, based on SLM technique, the phase sequence index and the corresponding number of OFDM symbols for minimum PAPR with Bauml phase sequence is shown. It can be seen from the figure that out of total 100,000 symbols, only 10,000 and 3,000 number of symbols satisfied the minimum PAPR at phase sequence with index 1 and 64, respectively. The remaining 87,000 symbols are distributed among the phase sequence from 2 to 63. It can be observed that, with the exception of phase sequence 1 and 64, almost all the other phase sequences are equiprobable of becoming the least PAPR generating phase sequence. Hence, it takes many iterations to find out the proper phase sequence in SLM technique with Bauml phase sequence. We can readily find the difference in Fig. 3 which showed the phase sequence index and the corresponding number of OFDM symbols with minimum PAPR when the Riemann phase sequence is used. In this figure, most of the least PAPR generating symbols are concentrated between phase sequence no. 40 to 61. Hence, searching can be focused in this zone requiring 22 iterations. In this way, computational complexity is reduced up to 65% compared to the Bauml Phase sequence.

Fig. 4 showed the comparative Complementary Cumulative Distribution Function (CCDF) of PAPR for Riemann and Bauml phase sequence based SLM techniques with that of original OFDM. It can be seen from this figure that, for a

TABLE I
SIMULATION PARAMETERS

Number of Subcarriers (N)	64
Modulation Technique	BPSK
Phase Sequences	Bauml and Riemann Matrix
Length of Phase Sequences	64
Number of OFDM Symbols	10^5
Simulation Tool and Version	MATLAB [7.8.0.347 (R2009a)]

probability of 10^{-3} , the corresponding PAPR of Riemann, Bauml and original OFDM are 1.6 dB, 5.3 dB and 10.2 dB, respectively. Thus, in Riemann and Bauml phase sequence with SLM technique, PAPR reduction of 8.6 dB and 4.9 dB is achieved compared to the PAPR of original OFDM signal. Therefore, Riemann performs better than Bauml phase sequence.

Coming back to Fig. 3 it can also be seen that at phase sequences 60 and 61, most of the symbols with least PAPR appeared. Moreover, there is a clear tendency of low PAPR yielding sequences being present in the region of sequence no. 40 to 61. Hence, in our proposed method, to find out the proper phase sequences with low computational complexity, we adopt a reverse searching technique which starts from index 64 and end at 1. Additionally, to minimize the computational complexity even further, some predefined values of PAPR is selected. The selected PAPR threshold values are 1.8 dB and 2 dB which are slightly higher than the minimum PAPR for Riemann Phase sequence with SLM. Using each case of selected threshold values of PAPR for Riemann sequence with SLM, searching for proper phase sequence is analyzed for both forward and reverse direction.

Fig. 5(a) and Fig. 5(b), showed the number of symbols that achieve PAPR threshold of 1.8 dB versus the number of iterations required in forward and reverse searching, respectively. Here, searching in forward direction implies that the index of Riemann sequence starts at 1 and increases at each iteration, whereas in reverse searching the starting index is 64 and the ending index is 1. From Fig. 5(a), it is observed that for forward searching, about 98,000 out of total OFDM symbols yields the PAPR threshold in 57 iterations (from 1 to 57). As there are 64 total iterations, only 11% complexity reduction is found. In contrast, for reverse searching, almost 98,200 symbols satisfying the PAPR threshold requirement with 7 iterations (from 64 to 58) is shown in Fig. 5(b). Hence, a big complexity reduction of 89% is achieved.

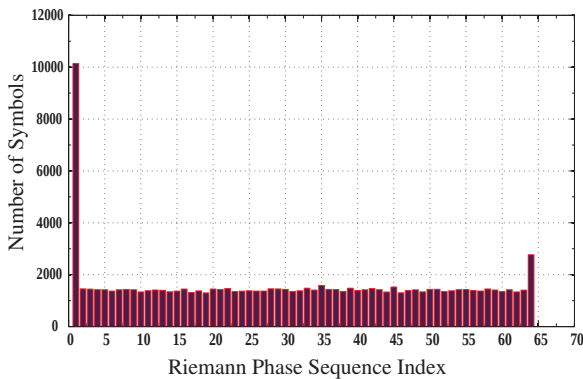


Fig. 2. Complexity Diagram for SLM using Bauml Phase Sequence.

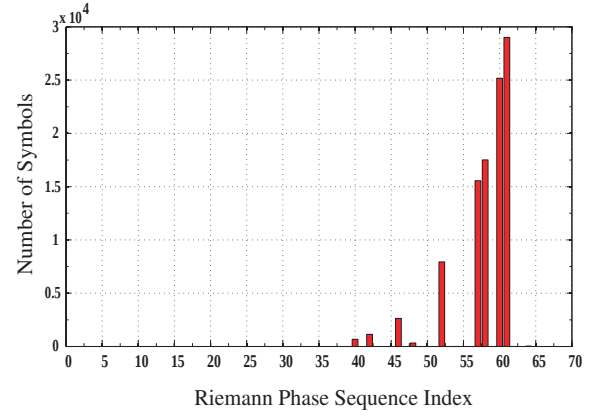


Fig. 3. Complexity Diagram for SLM using Riemann sequence.

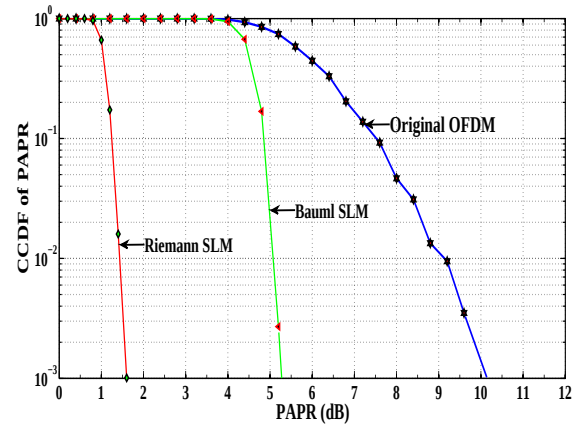
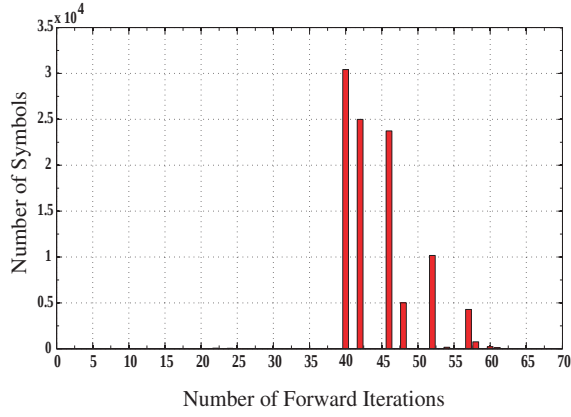


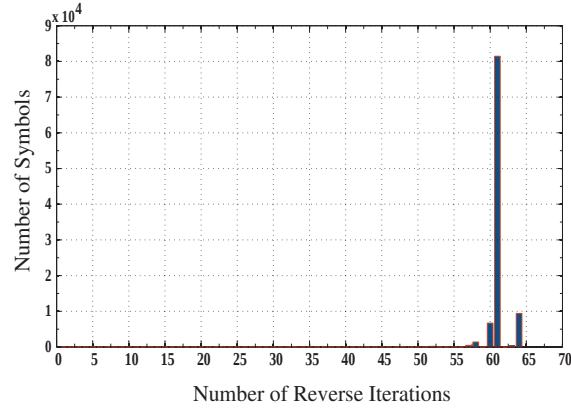
Fig. 4. CCDF vs. PAPR for SLM with Riemann phase sequence, SLM with Bauml phase sequence and Original OFDM system with $N=64$, BPSK modulation.

Similar simulation is repeated while increasing the PAPR threshold value slightly. In Fig. 6(a) and (b), the simulation results for forward and reverse searching is given, respectively when the PAPR threshold is set at 2 dB. From Fig. 6(a), it can be seen that 52 iterations are required (from 1 to 52) to generate 98,000 number of symbols that met the 2 dB threshold requirement in forward searching. It implies that complexity reduction is 19%. Whereas, in case of reverse direction, much improved performance is achieved with only 5 iterations (from 64 to 60). In fact, 98,300 number of the symbols met the requirement in 5 iterations only which is shown in Fig. 6(b). This time, a bigger complexity reduction of 92% is achieved.

From the above discussion it is evident that the reverse searching technique outperforms the forward searching technique by a large margin and significantly reduces the complexity of the system. It is also observed that with the expense of increase in PAPR threshold values for



(a)

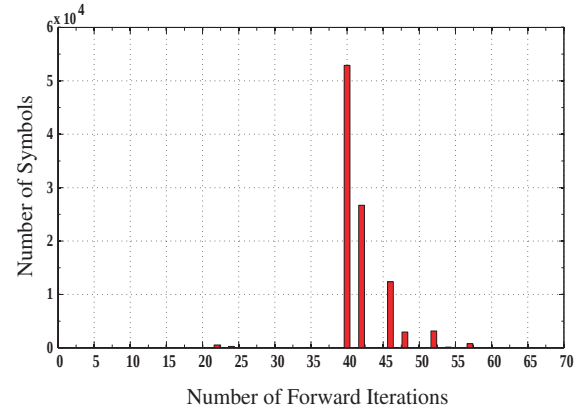


(b)

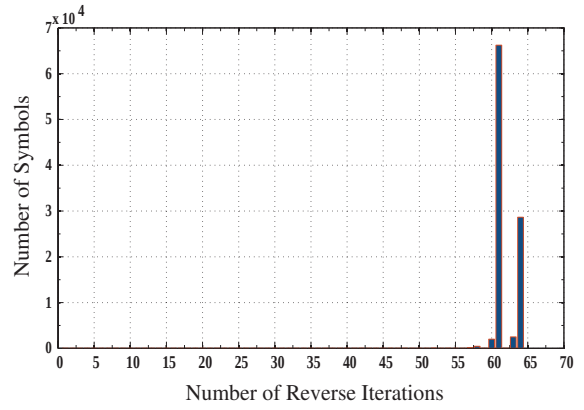
Fig. 5. Searching in Riemann matrix: (a) forward direction, and (b) reverse direction (Threshold level 1.8 dB).

the proposed reverse searching technique, computational complexity decreases. Hence, regardless of the threshold, the reverse searching mechanism always performs better than forward searching.

Another important point is to find out the default Riemann phase sequence which will permit side information free transmission with increase in system throughput. It can be observed from Fig. 3, most of the symbols satisfying minimum PAPR at phase sequence index 61st. It can also be observed from Fig. 5(b) and 6(b) that in the proposed reverse searching, more than 70% symbols satisfying threshold level always appeared at 61st sequence. Hence, index with 61st phase sequence can be chosen as the default Riemann sequence. Additionally, it is seen from Fig 5(b) and 6(b) that it took only 4 iterations to find out this special phase sequence. Whereas, for forward searching in Fig. 5(a) and 6(a), to find out the default index (40th phase sequence), it would take 40 number of iterations. Similarly, in Fig. 3 which is also forward searching with minimum PAPR, required 61 number of iterations to find out the default index. Hence, compared to forward searching with both threshold and minimum PAPR, proposed reverse searching technique



(a)



(b)

Fig. 6. Searching in Riemann matrix: (a) forward direction, and (b) reverse direction (Threshold level 2 dB).

finds this default index with less computational complexity. It is always advantageous to have high number of symbols on that default phase sequence to minimize the PAPR value. In the proposed reverse searching, number of symbols appeared on the default index are 81,000 and 66,000 for Fig. 5(b) and 6(b), respectively. In contrast, in forward searching in Fig. 5(a) and 6(a) number of symbols appeared on the default index are 31,000 and 53,000, respectively. Therefore, considering the number of symbols and number of iterations to find out the default index, our proposed reverse searching technique outperforms the forward one.

From Fig. 7, it is seen that CCDF of PAPR for Riemann Phase sequence with default index 40th is 3.7 dB for forward searching which showed the PAPR reduction of 6.5 dB compared to the PAPR of original OFDM signal. Since, considering all the phase sequences with minimum PAPR of Riemann phase sequence is 1.6 dB, PAPR sacrifice of 2.1 dB is observed for 40th default index. On the contrary, with the proposed reverse searching, the CCDF of PAPR with default index 61st is 2.5 dB which showed the PAPR reduction of 7.8 dB compared to the PAPR of original OFDM signal. In that case, PAPR sacrifice of only 0.9 dB is observed which

is much better than forward searching. Hence, in proposed method, side information free transmission is possible using Riemann default index with little sacrifice in PAPR value.

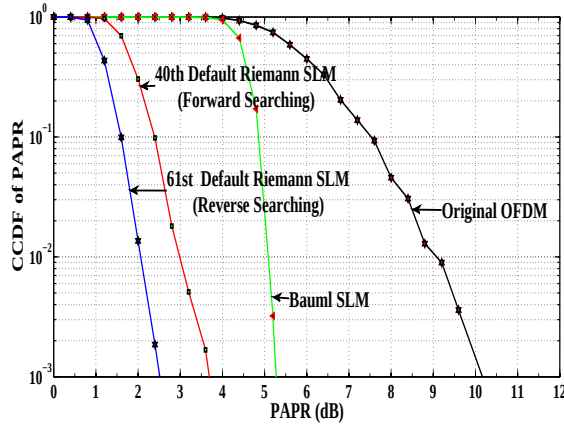


Fig. 7. CCDF vs. PAPR for SLM with Default Riemann phase sequence using Forward and Reverse searching, SLM with Bauml phase sequence and Original OFDM with $N=64$, BPSK modulation.

VII. CONCLUSIONS

Mitigating the adverse effect of high PAPR is one of the most challenging issues in OFDM system. While there are quite a significant number of existing PAPR reduction methods, none of them can achieve all the desirable outcomes, namely, large reduction in PAPR, no spectral leakage, low computational complexity, no or little loss of data rate etc. In this paper, we focused on the complexity reduction issue along with PAPR reduction. Our proposed idea showed a new approach of reverse searching using Riemann Sequence in SLM technique. Using this technique, we showed that it is possible to simultaneously achieve significant reduction in computational overhead along with low PAPR value. The results showed that the computational complexity can be reduced up to 92%. It is also found, using proposed reverser searching, selection of the default Riemann phase sequence is possible with less computational complexity and lower PAPR value. Hence, the issues of reducing the computational complexity and side information free transmission with low PAPR value is achieved in the proposed method.

REFERENCES

- [1] S. H. Wang, C. P. Li, K. C. Lee, and H. J. Su, "A novel low-complexity precoded OFDM system with reduced PAPR," *IEEE Trans. on Sig. Proc.*, vol. 63, no. 6, pp. 1366-1376, Mar. 2015.
- [2] J. G. Andrews, A. Ghosh, and R. Muhamed, *Fundamentals of WiMAX*, Prentice Hall, 2007.
- [3] Y. Jiang, "New companding transform for PAPR reduction in OFDM," *IEEE Commun. Lett.*, vol. 14, no. 4, pp. 282-284, Apr. 2010.
- [4] R. W. Bauml, R. F. H. Fischer, and J. B. Huber, "Reducing the peak-to-average power ratio of multicarrier modulation

- by selected mapping," *IEEE Elect. Lett.*, vol. 32, no. 22, pp. 2056-2057, Oct. 1996.
- [5] S. H. Muller and J. B. Huber, "OFDM with reduced peak-to-average power ratio by optimum combination of partial transmit sequences," *IEEE Elect. Lett.*, vol. 33, no. 5, pp. 368-369, Feb. 1997.
- [6] Y. Rahmatallah and S. Mohan, "Peak-to-average power ratio reduction in OFDM systems: A survey and taxonomy," *IEEE Commun. Surveys and Tutorials*, vol. 15, no. 4, pp. 1567-1592, 2013.
- [7] N. V. Irukulapati, V. K. Chakka, and A. Jain, "SLM based PAPR reduction of an OFDM signal using new phase sequence," *IEEE Elect. Lett.*, vol. 45, no. 24, pp. 1231-1232, Nov. 2009.
- [8] D. W. Lim, J. S. No, C. W. Lim, and H. Chung, "A new SLM OFDM scheme with low complexity for PAPR reduction," *IEEE Sig. Proc. Lett.*, vol. 12, no. 2, pp. 93-96, Feb. 2005.
- [9] S. L. Marple Jr., "Computing the discrete-time analytic signal via FFT," *IEEE Trans. on Sig. Proc.*, vol. 47, no. 9, pp. 2600-2603, Sept. 1999.
- [10] M. M. Rahman, M. S. Rahim, M. N. A. S. Bhuiyan, and S. Ahmed, "PAPR reduction of OFDM system using condex matrix based SLM method with low computational overhead," in *Proc. Int. Conf. on Elect. Inf. and Commun. Techn. (EICT)*, KUET, Khulna, Bangladesh, Nov. 2015, pp. 294-297.
- [11] M. M. Rahman, M. S. Rahim, M. N. A. S. Bhuiyan, and S. Ahmed, "Greatest common divisor matrix based phase sequence for PAPR reduction in OFDM system with low computational overhead," in *Proc. 1st Int. Conf. on Elect. and Electronic Eng. (ICEEE)*, RUET, Rajshahi, Bangladesh, Nov. 2015, pp. 97-100.
- [12] P. M. Z. Goyoro, I. J. Moumouni, and S. Abouty, "SLM using riemann sequence combined with DCT transform for PAPR reduction in OFDM communication system," *World Acad. Sci. Eng. Techn.*, vol. 6, no. 4, pp. 313-318, Apr. 2012.
- [13] M. Chandwani, A. Singhal, N. V. Irukulapati, & V. K. Chakka, "A low complexity SLM technique for PAPR reduction in OFDM using riemann sequence and thresholding of power amplifier," in *Proc. Annual IEEE India Conf.(INDICON)*, Gujara, India, Dec. 2009, pp. 1-4.
- [14] Y. C. Cho, S. H. Han, and J. H. Lee, "Selected mapping technique with novel phase sequences for PAPR reduction of an OFDM signal," in *Proc. IEEE 60th Vehicular Techn. Conf.*, LA, USA, vol. 7, Sep. 2004, pp. 4781-4785.

Highly Birefringent Octagonal Shaped Photonic Crystal Fiber with Two Zero Dispersion Wavelengths in Ti:sapphire Oscillator Range

Md. Mamunur Rashid^{1,*}, Md. Shamim Anower¹, Md. Rabiul Hasan¹

¹Department of Electrical & Electronic Engineering, Rajshahi University of Engineering & Technology, Rajshahi-6204, Bangladesh

*Email: mamun722_ruet@yahoo.com

Abstract— A simple octagonal photonic crystal fiber (OPCF) is proposed to have simultaneously two zero dispersion wavelengths (ZDWs), high birefringence, large nonlinear coefficient and low confinement loss. The finite element method (FEM) with perfectly matched layer boundary condition is used for the simulation of the presented OPCF. It demonstrates to a fact that for optimum design parameters this OPCF exhibits two ZDWs at 0.84 μm and 1.31 μm for fast axis and at 0.98 μm and 1.36 μm for slow axis, respectively. Moreover, high birefringence of 2.04×10^{-2} , large nonlinear coefficient of $33 \text{ W}^{-1}\text{km}^{-1}$ and low confinement loss about 0.0003965 dB/km are obtained at 1.55 μm . The proposed fiber may have important applications in supercontinuum generation, optical parametric amplification and four-wave mixing.

Keywords—Photonic crystal fiber, finite element method, birefringence, nonlinear coefficient, confinement loss

I. INTRODUCTION

Photonic crystal fibers consist of silica glass having a particular arrangement of microscopic air-holes that gives design flexibility of tuning dispersion properties, nonlinear coefficient, birefringence and so on. Over the last decades, design of PCF with high value of nonlinear coefficient has become an exceptional competitor to generate supercontinuum (SC) [1]. In a nonlinear photonic crystal fiber, the combination of air-holes in the cladding region can be made to have anomalous dispersion at wavelengths smaller than that of zero material dispersion. Due to this type of few wavelengths, light becomes bounded into the core portion of PCF, which minimizes the effective area [2]. Whereas it is usual thing that small effective area gives high nonlinearity, which can easily create broadband supercontinuum. Because of much more important advantages, for example high brightness, wide bandwidth and SC has been employed heavily in various sectors starting from optical coherent tomography to three-photon microscopy [3]. In addition, photonic crystal fibers with ZDW and large nonlinearity may also be applicable in optical parametric oscillator, optical parametric amplifier, multi-wavelength lasers, signal switcher and optical code division multiplexer [4].

Birefringence is another important parameter is PCF. It is deliberately created in PCF with changing distortion in the

core such as using missing air holes, using elliptical air holes etc. Due to birefringence, light splits into two orthogonal polarization components having different refractive indices. Large birefringence is necessary to eliminate the effect of PMD [1]. Besides, photonic crystal fibers with birefringence have significant uses in optical sensors, optical filters and coherent optical communications. For the production of broadband supercontinuum, it is necessary for maintaining the state of polarization of the guided electromagnetic waves. Therefore, birefringent PCF can also be employed for SC generation [1].

In order to obtain high birefringence, high nonlinearity and ZDWs, various different structural geometries have been shown over some years. Agrawal *et al.* [5] proposed a spiral PCF having two ZDWs to generate broadband supercontinuum. Besides, modal birefringence of 0.022 was reported, but there is no information about nonlinear coefficient in Ref. 5. In Ref. 6, Ademgil had proposed a PCF structure based on modified hexagonal lattice, which shows birefringence of 2.65×10^{-2} and two ZDWs. Fabrication of this PCF is difficult due to complex structure. Very recently, Hasan *et al.* [1] proposed defected-core hexagonal PCF having two ZDWs in both fundamental slow and fast axis modes. The authors also reported high birefringence of 0.0315 at 1550 nm. In our paper, an easy octagonal structure has been reported to obtain two ZDWs and high birefringence. The FEM strategy has been employed to characterize the guiding properties. The proposed OPCF contains the air-holes of only circular shape that certainly allows the fabrication easier. The key characteristics of this structure are the two ZDWs in both fast and slow axis modes with high level of birefringence. Using the optimum value of design specifications, the suggested structure gives large valued birefringence of 2.04×10^{-2} , high nonlinear coefficient of $33 \text{ W}^{-1} \text{ km}^{-1}$ with effective area about $2.82 \mu\text{m}^2$ at 1550 nm.

II. DESIGN METHODOLOGY AND EQUATIONS

The proposed OPCF is made of pure silica which has arrangement of air-holes running along its length. Fig. 1 shows the cross-sectional view of the proposed OPCF. The PCF contains circular air-holes in all five rings and 1st, 2nd, 3rd, 4th and 5th ring hold 6, 16, 24, 32 and 40 air-holes,

respectively. Two missing air-holes in first ring along y-axis are to have high birefringence. Three different sized air-holes have been used in the design with diameter of d for 1st, 2nd and 5th ring and diameter of d_3 , d_4 for 3rd and 4th ring, respectively. There is a common pitch Λ between two rings and optimum value is set to $\Lambda = 1 \mu\text{m}$, $d/\Lambda = 0.76$, $d_3/\Lambda = 0.5$, $d_4/\Lambda = 0.6$. Here, OPCF is a good one because octagonal symmetry gives hard mode field confinement capability and poor confinement losses [7].

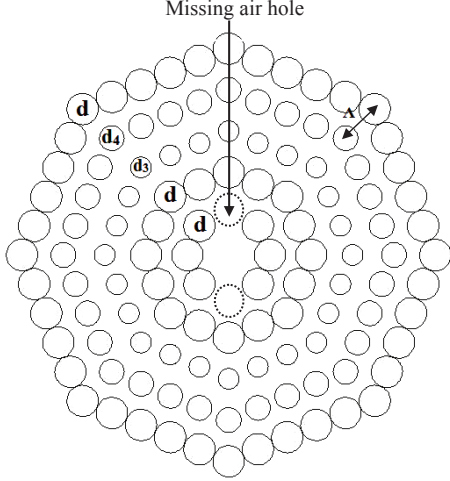


Fig. 1. Cross-section view of the proposed OPCF for optimum design parameters. The dotted circle in the 1st ring represents missing air holes.

In order to calculate chromatic dispersion, confinement loss nonlinear coefficient, and effective area the following equations have been used [8]

$$D = -\frac{\lambda}{c} \frac{d^2 \text{Re}[n_{\text{eff}}]}{d\lambda^2} \quad (1)$$

$$\alpha_{\text{CL}} = 8.686 \times \frac{2\pi f}{c} \text{Im}(n_{\text{eff}}) \quad (2)$$

$$\gamma = \left(\frac{2\pi}{\lambda} \right) \left(\frac{n_2}{A_{\text{eff}}} \right) \quad (3)$$

$$A_{\text{eff}} = \frac{\left(\iint |E|^2 dx dy \right)^2}{\iint |E|^4 dx dy} \quad (4)$$

where c is the velocity of light in vacuum, λ is the wavelength, $\text{Re}[n_{\text{eff}}]$ is the real part, $\text{Im}[n_{\text{eff}}]$ is the imaginary part of effective refractive index, f the operating frequency, A_{eff} is effective area, n_2 is Kerr constant of pure silica and E is the electric field amplitude of the guided wave.

III. SIMULATION RESULTS

We have used commercially available full-vector finite element method (FV-FEM) with COMSOL Multi-physics software of version 4.2 to simulate the structure of our proposed OPCF. The simulation results are reported for the design with optimum settings as $\Lambda = 1.00 \mu\text{m}$, $d/\Lambda = 0.76$, $d_3/\Lambda = 0.50$, $d_4/\Lambda = 0.60$. Fig. 2 shows the fundamental mode

field distribution due to horizontal- polarization and vertical- polarization at the excitation wavelength of 1550 nm. It can be evident that both modes are tightly bounded in the central core.

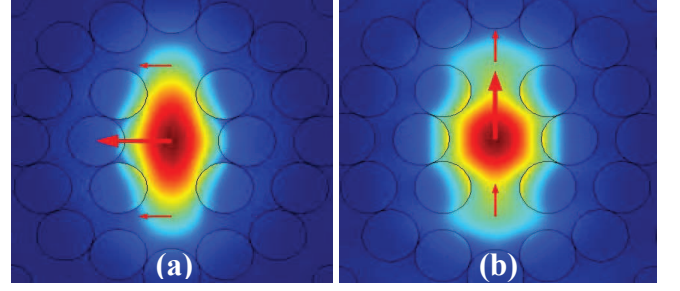


Fig. 2. Fundamental optical field mode distribution at 1550 nm for (a) horizontal-polarization (b) vertical-polarization.

The effective refractive index of our offered photonic crystal fiber according to wavelength is illustrated in Figure 3. From the figure, we can observe that there exists a reverse relation between effective refractive index and wavelength i.e. the index decreases according to the increment of wavelength.

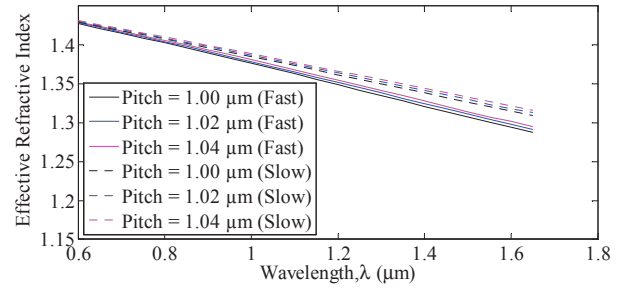


Fig. 3. Effective refractive indices as a function of wavelength.

Birefringence of the proposed fiber with varying wavelength is presented in Figure 4. This can be demonstrated from the figure so that the birefringence rises smoothly according to wavelength. In the first ring of the OPCF, there are two missing air holes. These air holes are responsible to induce high level of birefringence. At optimum design parameters, high birefringence about 2.04×10^{-2} has observed at the excitation wavelength 1.55 μm . In Ref. [9], we have found birefringence of 1.77×10^{-2} , which is lower value than that of the proposed OPCF. Moreover, Ref. [9] uses rotational hybrid cladding which has design difficulty.

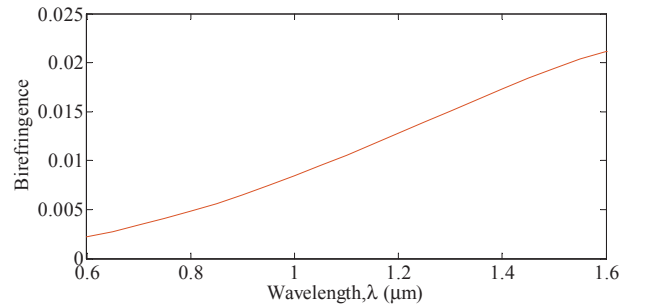


Fig. 4. Birefringence versus wavelength curve due to $\Lambda = 1.00 \mu\text{m}$.

In this paper, now our investigation is to reveal the findings of the chromatic dispersion versus wavelength curve, which is shown in Fig. 5. We can modify the chromatic dispersion by tailoring some structural parameters like the value of the ratio of air-hole diameter to pitch [10]. For this reason, we fixed up the optimum value for the guiding parameters. The investigation about Fig. 4 is that there are two ZDWs in the curve denoted by ZDW1 and ZDW2. The ZDWs are lying at the 0.84 μm and 1.31 μm for fast axis and at 0.98 μm and 1.36 μm for slow axis over 0.70 to 1.60 μm , which covers Ti:sapphire oscillator range. In the design, we can see that there are air holes of comparatively smaller diameter in third and fourth ring to create coupling between cladding and core portion, which helps to get zero dispersion wavelengths into the curve [1]. A photonic crystal fiber with high nonlinearity and nearly staying two zero dispersion wavelengths can be employed fruitfully to generate supercontinuum [11]. This is feasible to create large-spectral density and supercontinuum with poor noise through selecting a pump wavelength point between two zero dispersion wavelengths [12].

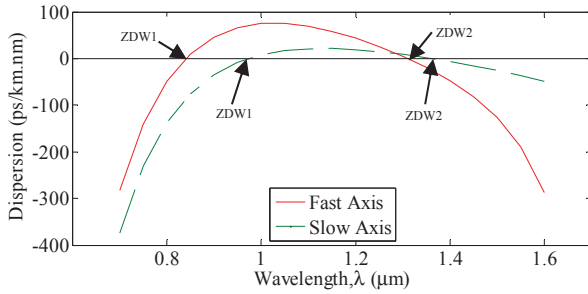


Fig. 5. Dispersion curve with zero dispersion wavelengths for $\Lambda = 1.00 \mu\text{m}$.

Fig. 6 shows the effect of changing Λ on the zero dispersion wavelengths by investigating the dispersion curves with three different values of pitch. For $\Lambda = 1 \mu\text{m}$ keeping d/Λ , d_3/Λ and d_4/Λ constant, the curve cuts the zero level line two times at about wavelength 0.84 μm and 1.31 μm for fast axis and at about 0.98 μm and 1.36 μm for slow axis. For choosing $\Lambda = 1.02 \mu\text{m}$, we get two ZDWs over 0.70 μm -1.60 μm wavelength bands. In this case, the positions of ZDWs are at about wavelength 0.84 μm and 1.34 μm for fast axis and at about 0.97 μm and 1.41 μm for slow axis. When $\Lambda = 1.04 \mu\text{m}$ and d/Λ , d_3/Λ and d_4/Λ are kept constant, we observe that there are also two zero dispersion wavelengths at wavelength of 0.85 μm and 1.37 μm for fast axis and at wavelength of 0.97 μm and 1.45 μm for slow axis respectively. Here, since the air-holes of first ring and second ring should be large and in this case $d/\Lambda = 0.76$ is the maximum possible value. Because if we choose the value of pitch less than 1 μm then the air-holes of 1st, 2nd and 5th rings will be overlapped which creates design complexity. So setting d/Λ at maximum value of 0.76, the pitch value should be greater than or equal to 1 μm . Therefore, we have studied the cases of pitch value 1 μm , 1.02 μm and 1.04 μm . When $\Lambda = 1 \mu\text{m}$, the distance between two ZDWs is about 0.47 μm and 0.38 μm due to horizontal and vertical axis respectively. Again, if $\Lambda = 1.02 \mu\text{m}$, the distance between ZDW1 and ZDW2 is about 0.50 μm and

0.44 μm because of x- and y-axis respectively. Finally when $\Lambda = 1.04 \mu\text{m}$, the distance between two ZDWs is about 0.52 μm and 0.48 μm due to horizontal and vertical axis respectively. Since the distance between ZDWs must be small as possible, so we can select $\Lambda = 1.00 \mu\text{m}$ as the best one among three studied cases. Here we can choose a pump wavelength between two zero dispersion wavelengths which is useful to create SC. Since the offered OPCF provides two zero dispersion wavelengths for x- & y- polarizations and ZDW1 of both polarizations are lower than zero material dispersion wavelength, so our photonic crystal fiber is eligible to generate supercontinuum. In addition, the ZDW1's of both horizontal and vertical-polarization are positioned in the working wave band of Ti:sapphire oscillator range (700 nm to 980 nm), which may be suitable to be utilized for frequency conversion of Ti:sapphire femtosecond laser [12].

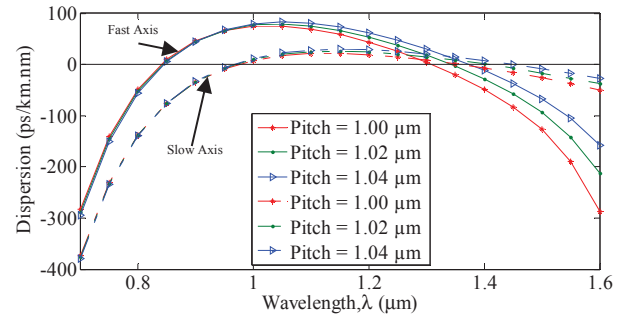


Fig. 6. Effect of changing Λ (pitch) on dispersion property as a function of wavelength.

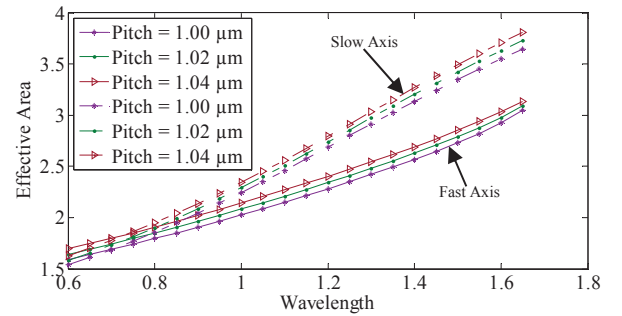


Fig. 7. Effect of change of pitch on effective area as a function of wavelength.

Now, we observe the effective area of the proposed PCF for varying wavelength. Fig. 7 shows the curve of effective area for optimum design parameters. From the figure, it is evident that the effective area increases with the increasing of wavelength. The value of effective area is 2.82 μm^2 and 3.45 μm^2 for y- & x-mode respectively at the operating wavelength of 1550 nm. When the value of Λ is set to 1.02 μm , we obtain the effective area value 2.88 μm^2 and 3.53 μm^2 for y- & x-mode at 1550 nm. Again when $\Lambda = 1.04 \mu\text{m}$, then the effective area is 2.94 μm^2 and 3.60 μm^2 for y- and x-polarization, respectively at 1.55 μm . As small value of effective area is expected and observing three above cases, effective area is smallest for pitch of 1 μm . Besides, Because of small effective area, a photonic crystal fiber may be insensitive for high bending loss [13-14].

Now we investigate the nonlinear coefficients for different pitch values through Fig. 8. The high value of nonlinear coefficient is specifically beneficial for optical parametric amplification and for generation of stable supercontinuum [15]. Due to the pitch value of 1.00 μm , 1.02 μm and 1.04 μm , we have found the nonlinear coefficient of 33, 32.39 and 31.73 $\text{W}^{-1}\text{km}^{-1}$ respectively for fast axis (y - polarization) and 27.02, 26.42 and 25.86 $\text{W}^{-1}\text{km}^{-1}$ respectively for slow axis (x -polarization). It is known from Ref. [16] that photonic crystal fibers of large nonlinear coefficient are perfect for dispersion compensation while four-wave mixing is prone for appearing.

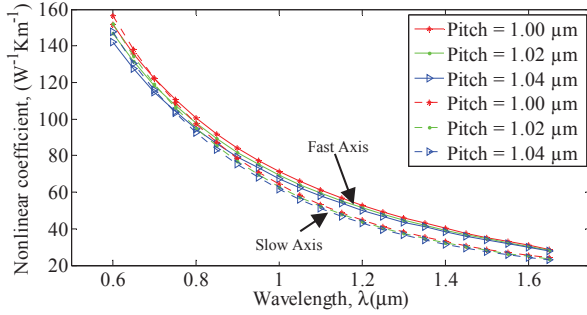


Fig. 8. Effect of change of pitch on nonlinear coefficient versus wavelength curve.

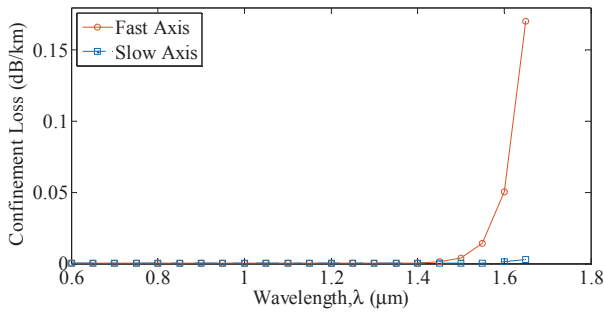


Fig. 9. Confinement losses for fast and slow axis as a function of wavelength.

Confinement loss is a challenge in fiber optic communication and it is very important to minimize the confinement loss. Fig. 9 is to expose confinement losses for only five rings ($N_r=5$). The loss can be decreased with using larger sized air-holes in the outermost rings of PCF. Again, the loss can be decreased through inserting a number of air-hole rings into the cladding section [17] while inserting more air-hole rings in cladding portion does not influence the dispersion curve [18]. The confinement loss of the proposed fiber is 0.01441 and 0.0003965 dB/km for fast axis and slow axis respectively at the operating wavelength of 1.55 μm . Whereas, there is confinement loss of about 42.4 dB/km in Ref. [19], which is very high value compared to the proposed OPCF. Finally, we would like to discuss about the fabrication possibility of the proposed OPCF. There are several viable ways that can be used to practical realization of the designed structure. Standard stack and draw method is applicable for circular shaped air holes having triangular lattice. Therefore, this technique can be used for fabrication. On the other hand,

sol-gel casting technique can fabricate almost any structure; therefore, it can effectively used to fabricate this OPCF.

IV. CONCLUSION

In summary, we propose a simple octagonal PCF for obtaining two ZDWs and ultra-high birefringence in the telecommunication bands. The designed structure shows two ZDWs for both fundamental slow and fast axis modes in Ti:sapphire oscillator range, which is eligible to be utilized for frequency conversion of Ti:sapphire femtosecond laser. The proposed OPCF can be realized for practical fabrication since it consists of circular air holes with octagonal symmetry. Due to excellent index guiding properties, this structure would be a prominent candidate in the field of optical fiber communications. Investigation on bending loss for small bending radius will be the future work.

REFERENCES

- [1] M. R. Hasan, M. S. Anower and M. I. Hasan, "Polarization maintaining highly nonlinear photonic crystal fiber with closely lying two zero dispersion wavelengths," *Opt. Eng.*, vol. 55, no 5, pp. 056107, May 2016.
- [2] J. K. Ranka, R. S. Windeler and A.J. Stentz, "Visible continuum generation in air-silica microstructure optical fibers with anomalous dispersion at 800 nm," *Opt. Lett.*, vol. 25, no 1, pp. 25-27, 2000.
- [3] W. Bi, X. Li, J. Gao, L. Zhang, and M. Liao, "Numerical simulations of the ultrabroadband supercontinuum generation by dual-wavelength pumping in photonic crystal fiber with two zero dispersion wavelengths," *Appl. Opt.*, vol. 54 no. 11, pp. 4542-4548, May 2015.
- [4] Z. Ya-Ni, "Optimization of highly nonlinear dispersion-flattened photonic crystal fiber for supercontinuum generation," *chin. Phys. B*, vol. 22, no. 1, pp. 014214, 2013.
- [5] A. Agrawal, N. Kejalakshmy, B. M. A. Rahman, and K. T. V. Grattan, "Polarization and dispersion properties of elliptical hole golden spiral photonic crystal fiber," *Appl. Phys. B*, vol. 99, no. 4, pp. 717-726, 2010.
- [6] H. Ademgil and S. Haxha, "Highly nonlinear birefringent photonic crystal fiber," *Opt. Commun.*, vol. 282, no. 14, pp. 2831-2835, 2009.
- [7] S. M. A. Razzak and Y. NamiHira, "Tailoring dispersion and confinement losses of photonic crystal fibers using hybrid cladding," *J. Lightw. Technol.*, vol. 26, no. 13, pp. 1909-1914, Jul. 1, 2008.
- [8] M. I. Hasan, S. M. A. Razzak and M. S. Habib, "Design and characterization of highly birefringent residual dispersion compensating photonic crystal fiber," *J. Lightw. Technol.*, vol. 32, No. 23, December 1, 2014.
- [9] M. S. Habib, M. Selim Habib, M.I. Hasan, S. M. A. Razzak, M. A. Hossain and Y. NamiHira, "Polarization maintaining large nonlinear coefficient photonic crystal fibers using rotational hybrid cladding," *Optik*, vol. 125, no. 3, pp. 1011-1015, 2014.
- [10] C. Xia, G. Zhou, Y. Han, and L. Hou, "Highly birefringent octagonal photonic crystal fibers with two zero-dispersion wavelengths," *chin. Opt. Lett.*, vol. 9, no. 10, pp. 100609, 2011.
- [11] K. M. Hilligsoc et al., "Supercontinuum generation in a photonic crystal fiber with two zero dispersion wavelengths," *Opt. Express*, vol. 12, no. 6, pp. 1045-1054, 2004.
- [12] S. Murugkar et al., "Coherent anti-stokes Raman scattering microscopy using photonic crystal fiber with two closely lying zero dispersion wavelengths," *Opt. Express*, vol. 15, no. 21, pp. 14028-14037, 2007.
- [13] H. Ademgil, S. Haxha, and F.A. Malek, "Highly nonlinear bending insensitive birefringent photonic crystal fibers," *Engineering*, vol. 2, no. 8, pp. 608-616, 2010.

- [14] S.-G. Li, X.-D. Liu and L.-T. Hou, "Numerical study on dispersion compensating property in photonic crystal fibers," *Acta Phys. Sin.*, vol. 53, no. 6, pp. 1880-1886, 2004.
- [15] M. H. Frosz, P. Falk, and O. Bang, "The role of the second zero-dispersion wavelength in generation of supercontinua and bright-bright soliton-pairs across the zero-dispersion wavelength," *Opt. Express*, vol. 13, no. 6, pp. 6181-6192, 2005.
- [16] Q. Li, C. Liu, H. Li, and T. Sun, "Research on dispersion compensating property and nonlinear coefficient of photonic crystal fiber," in *Proc. Symp. Photon. Optoelectron.*, pp. 14-16. Aug. 2009.
- [17] M. R. Hasan, M. I. Hasan, and M. S. Anower, "Tellurite glass defect-core spiral photonic crystal fiber with low loss and large negative flattened dispersion over S + C + L + U wavelength bands," *Appl. Opt.*, vol. 54, no. 32, pp. 9456-9461, 2015.
- [18] M. R. Hasan, M. S. Anower, and M. I. Hasan, "A polarization maintaining single mode photonic crystal fiber for residual dispersion compensation," *IEEE Photonics Technol. Lett.*, vol. 28, no. 16, pp. 1782-1785, 2016.
- [19] M. I. Hasan, M. S. Habib, and S. M. A. Razzak, "An elliptical-shaped core residual dispersion compensating octagonal photonic crystal fiber," *IEEE Photonics Technol. Lett.*, vol. 26, no. 20, October 15, 2014.

A Smart Solution for Solar DC Nano Grid integrated with Pre-paid billing and Mobile Application

Mahmud Ibrahim (mahmud.ibrahim021@gmail.com)
United International University, Dhaka, Bangladesh

Shahriar Ahmed Chowdhury (shahriar.ac@gmail.com)
United International University, Dhaka, Bangladesh

Mohammad Nurul Huda (mnh@cse.uiu.ac.bd)
United International University, Dhaka, Bangladesh

M. Rezwan Khan (rezwanm@uii.ac.bd)
United International University, Dhaka, Bangladesh

Abstract— Nowadays, renewable energy specially solar energy has become immensely popular as well as a sustainable and alternative source of energy harvesting and utilization. In Bangladesh, there are more than 4.5 million Solar Home System (SHSs) which have been installed till January, 2016 [1]. It is a natural tendency in rural areas of Bangladesh that people live in clusters and these clusters are arranged by some tens of households to few hundreds. Formation of a small scale Solar DC Nano grid would be a practical affordable solution for generating electricity and also providing basic electricity demand to these clusters. In a research work [2], it has been seen that in SHS system, more than 30% of its produced energy cannot be stored or utilized at any given time. Therefore, a smart solution for Solar DC Nano grid is needed. In this paper, a smart Solar DC Nano grid is proposed for better energy utilization. Proposed Solar DC Nano grid formed a small energy sharing (both selling and buying) cluster between SHS systems using Bi-directional DC-DC converter without large change in existing SHS systems. Pre-paid billing using mobile application is also introduced in this paper for smooth cash in and cash out after buying or selling excess energy to the neighboring SHS system. This paper illustrates the design and development of a Bi-directional DC-DC converter with Solar Charge Controller mechanism and also Pair to Pair (P2P) communication between mobile app and this device.

Keywords— *SHS, Solar DC Nano Grid, Bi-directional DC-DC converter, IDCOL, Moible app, P2P Communication.*

I. INTRODUCTION

In Bangladesh, approximately 38% of its total population does not have any access to the grid electricity [3]. In addition, there are more than 20% areas of total area of Bangladesh where providing grid electricity is costly [4]. Solar power can be an alternative clean energy source for the electrification of these off-grid areas. The majority of these rural households' electricity demands are basic lighting load, fan load, charging mobile phones, watching TV, irrigation and rice husking. A standard 50Wp SHS system can only fulfill basic lighting load, fan load and charging mobile phones. From the investigation report of "Technical Audit, 2014" by Centre for Energy Research (CER), United International University (UIU) and Infrastructure Development Company Limited (IDCOL), it has been seen that off-grid areas people are now interested about advance systems such as color TV, rice husking, preserving foods and also irrigation. In addition, it is also reported that people are also price sensitive about adding new items with their existing SHS system. Therefore, it would be a smart solution if Solar DC Nano grid system is introduce in these off-

grid areas to ensure an affordable and cost effective solution to them.

Development of Solar DC Nano grid system reduces the unused energy as the system is shared by many people otherwise this excess energy can be used in productive use for the community [5][6]. A Solar Nano-Grid is a small grid network with a power output from few hundreds watts to few kilowatts, whereas a solar home system (SHS) is a standalone system supplying a fixed amount of energy to the consumer household [7].

From the concept of Solar Nano grid system, it's a source of 500W to 5kW power generation system which fulfills electrical energy demand of up to 50 household within short range of area. Where solar PV is installed in a central position and transmitted power in the form of DC current. In Solar DC Nano grid there is no need of DC-AC conversion.

DC system has the following advantages over AC system

- All the modern appliances such as LED lights, mobile phone, laptops etc. that run on AC are also fully functional on DC in the same voltage range.
- Solar PV generates DC power, and also the generated energy is stored in DC form. Therefore no need of inverter for the conversion of DC to AC.
- Battery DC voltage (12V) can't be transferred for long distance due to voltage drop. But on the other hand, the high voltage DC (110V to 220V) can be easily transferred in few hundreds meters without any significant losses.
- Adding up of DC system (energy sources) in parallel is much easier than by AC. Paralleling of AC sources need synchronizing of the systems.
- Power factor is not a concern in DC systems.
- DC system has no inductive or capacitive losses.
- Management of harmonics distortion is easier in DC than AC.
- DC cooling fans are more efficient than conventional AC cooling fans (BLDC motors are efficient than induction motors).
- DC power transmission is more efficient and less expensive than AC power transmission. [8]

There is another concept of Solar DC Nano grid, where power generation is decentralized and by sharing (both selling and buying) excess energy from the neighbors can be possible [9]. The proposed smart Solar DC Nano's power generation is decentralized. For example, suppose 5-50 households are connected with each other via a smart device. In figure 1, there are a DC grid line of 125V and existing SHS systems. It's assumed that all the SHS systems are 20Wp system. These SHS systems are connected with the 125V DC grid line via a bi-directional DC-DC converter and formed an energy cluster. This bi-directional DC-DC converter can charge the SHS system's battery while presence of excess energy in the grid line and also feed energy to the grid line while presence of excess energy in a SHS system. The buying or selling of energy depends on customers of this cluster. If someone need energy he can buy it from other customers of that cluster and vice versa. At extreme condition, if all the cluster customers need energy and no one has to sell excess energy on that time the smart device will turn off and also when all the cluster customers have excess energy but no one wants to buy on that time this device will be turned off.

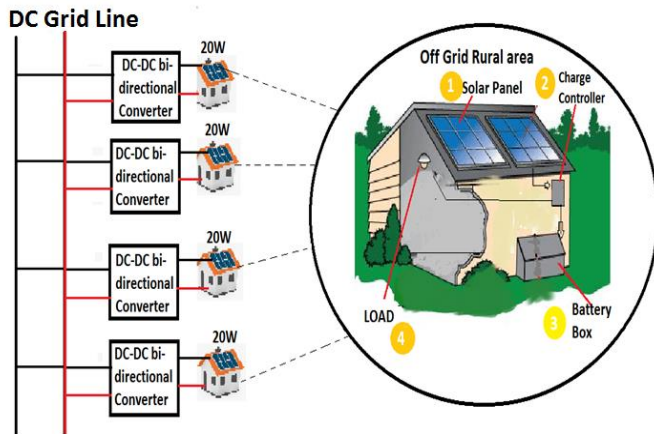


Figure 1: Decentralized energy sharing theme using bi-directional DC-DC converter.

II. RESEARCH OBJECTIVE

The objective of the work is to develop a bi-directional DC-DC converter with integration of solar charge controller mechanism to form a 125V DC grid line. Design and development of a mobile app for virtual money transaction between this device and mobile app for smooth cash in and cash out. Real cash transaction will remain as a manual process which ensures safe hand to hand money transfer, better collection efficiency, an authority to overview these clusters as well as a better system management.

III. METHODOLOGY

The proposed system would be something like that suppose there are 10 households in a cluster. All of them are connected to a 125V DC grid line via bi-directional DC-DC converter. Grid line will cover 5-50m radial distance. Each household has a separate SHS. In each household, panel, battery and load are connected with bi-directional DC-DC converter instead of solar charge controller. One of the I/O of this converter is connected

to the 125V DC grid line. Energy selling or buying will be performed by this I/O line. Each device has a small wifi module integrate inside the device to communicate with mobile app. Furthermore, via this module, one can transfer virtual recharge money into the device (cash in) and also collect virtual money (cash out) by contacting local agent of the device manufacturer. The local agent will help individual customer by recharging his device after collecting real cash in hand and also giving him real cash after collecting user balance information from the device. In this part, the whole device level communication will be done by creating a P2P communication link between an android mobile phone and device's wifi module. Each device has a unique ID number by which one can get every information about that device user such as user name, location, balance update, mobile number, no. of transactions etc. In figure 2, a 32-bit micro-controller (MCU) has been used to control and communicate with bi-directional DC-DC converter, bi-directional energy flow meter and wifi module (combined name: "Smart box"). Logical communication for instance voltage-current measurement, control pulse, energy flow direction and measurement is done by MCU. P2P communication is also done via MCU. Unique user ID number, balance update and other information is stored in internal EEPROM of the MCU device.

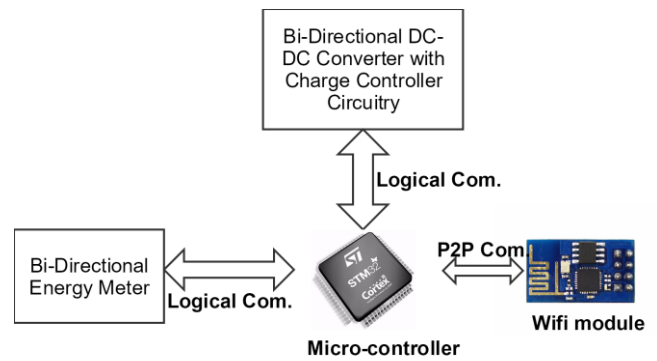


Figure 2: Logical internal communication between micro-controller and other sections block diagram of the device.

A bi-directional DC-DC converter is used in applications where bi-directional power flow is required [10]. A typical system consists of a full-bridge power stage on the high-voltage (HV) side, which is isolated from a full-bridge or a current-fed push-pull stage on the low-voltage (LV) side. In step down mode, a phase-shifted full-bridge (PSFB) with synchronous rectification controls power flow from 125V bus to the 12V battery. On the other hand in step-up mode, another PSFB with synchronous rectification used to control power flow from 12V battery to 125V DC grid bus line. Both voltage control and average current control mode is used here to ensure optimal operation. Various PWM switching schemes have been implemented with seamless transitions between switching modes and also between the two operation modes.

In figure 3, the logical circuit diagram of the bi-directional DC-DC converter has been shown. Here, two full-bridge power stages are used. Q1, Q2, Q3 and Q4 MOSFETs are used for

step-down (buck) mode full-bridge power stage. On the other hand, Q5, Q6, Q7 and Q8 MOSFETs are used for step-up (boost) mode full-bridge power stage.

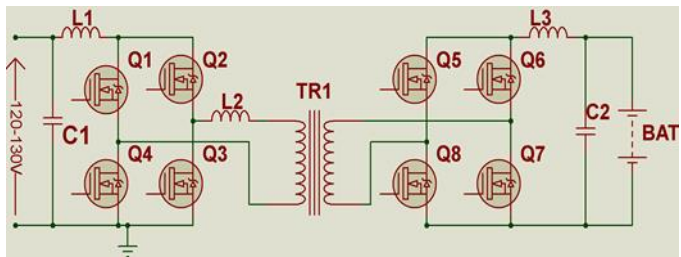


Figure 3: Bi-directional DC-DC converter circuit diagram.

Inductor L1 and capacitor C1 used for filtering on boost stage and inductor L3 and capacitor C2 used for filtering on buck stage. Inductor L2 is used for zero-voltage switching (ZVS) to reduce switching losses. A high frequency transformer TR1 is used here for voltage level conversion. Battery (BAT) is used to store energy.

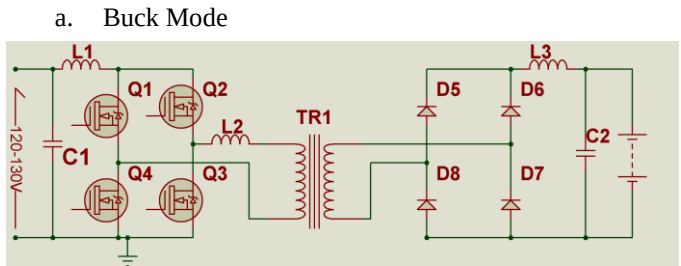


Figure 4: Buck mode power stage.

In figure 4, buck mode power stage has been represented. In this stage, four power MOSFETs (Q1,Q2,Q3 and Q4) form a phase shifted full bridge on the primary side of the isolation transformer and secondary side MOSFETs (Q5,Q6,Q7 and Q8) act like a synchronous rectifier (diode D5,D6,D7,D8). Q1 and Q4 are switched at 50% duty and 180 degrees out of phase with each other. Same thing done also for Q2 and Q3. The PWM switching signals for leg Q2-Q3 of the full bridge are phase-shifted with respect to those for leg Q1-Q4. The amount of energy transferred is depends on the amount of phase shifted between the amount of overlapping diagonal switches. Inductor L2 provide support to the transformer leakage inductance for resonance operation with MOSFET capacitance and facilitates zero voltage switching (ZVS). Inductor L2 and capacitor C2 form the output filter. The output of this stage is 15V (open-circuit) to charge the 12V battery. Buck mode PWM waveforms are displayed in figure 5.

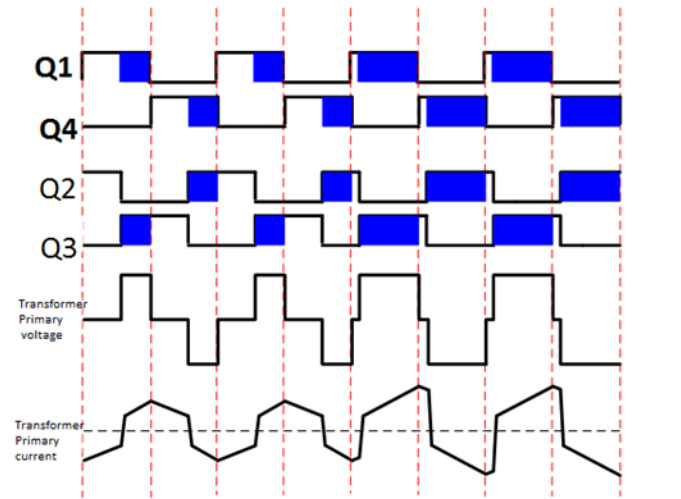


Figure 5: Buck mode PWM waveforms.

b. Boost mode

In figure 6, boost mode power stage has been displayed. Here, MOSFETs (Q5,Q6,Q7 and Q8) formed a phase shifted full bridge on the secondary side of the isolation phase transformer and primary side MOSFETs (Q1,Q2,Q3 and Q4) act like a synchronous rectifier (diode D1,D2,D3,D4). Q5 and Q8 are switched at 50% duty and 180 degrees out of phase with each other. Same thing done also for Q6 and Q7. The PWM switching signals for leg Q6-Q7 of the full bridge are phase-shifted with respect to those for leg Q5-Q8.

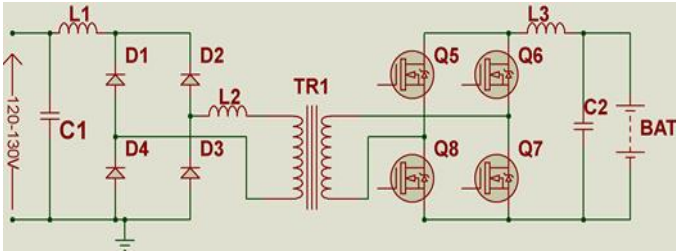


Figure 6: Boost mode power stage.

The amount of energy transferred is depending on the amount of phase shifted between the overlapping amounts of diagonal switches. Inductor L1 and capacitor C1 formed the output filter. The output voltage in this stage is 125V nominal which is feed from the 12V battery. Boost mode PWM waveforms are displayed in figure 7.

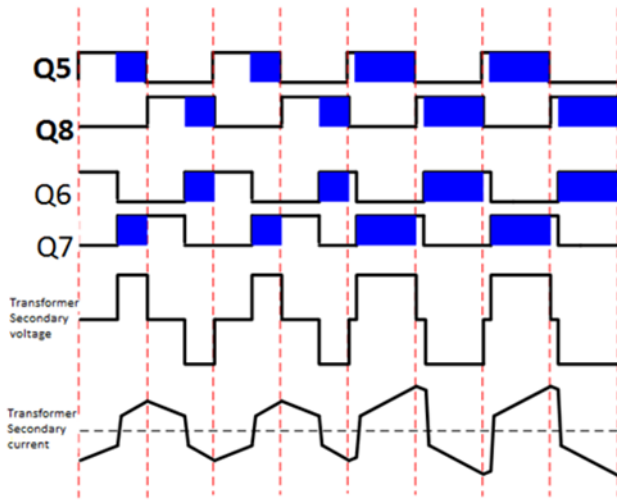


Figure 7: Boost mode PWM waveforms.

In this bi-directional DC-DC converter both HV and LV side output voltage and current is measured via micro-controller. In the figure 8, a micro-controller (MCU) is displayed which measure buck stage voltage and current from low side (Secondary side of transformer) mentioned as LV_V and LV_I simultaneously.

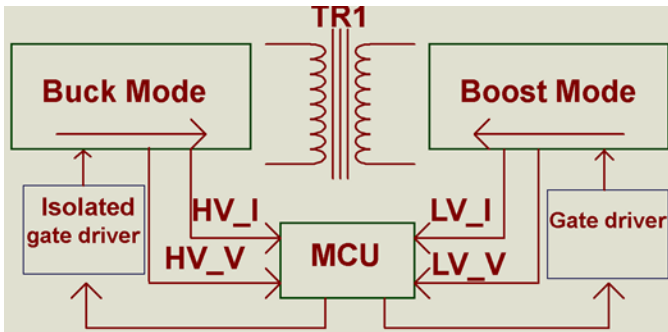


Figure 8: Voltage-Current feedback and gate drive diagram.

Also measure boost stage voltage and current from high side (primary side of transformer) mentioned as HV_V and HV_I using isolated measurement scheme. There are two MOSFET gate driver circuit used here to control MOSFETs.

Here is table below which explains when these two MOSFET drivers are turned on and turned off.

Table 1: Gate driver turn on and turn off schedule.

Grid line voltage (Volt)	Battery voltage (Volt)	Isolated gate driver (HV side)	Gate driver (LV side)	Power Stage
$V_{grid} > 130$	$V_{bat} > 14.5$	OFF	OFF	No feed or charge
$V_{grid} > 125$	$V_{bat} < 13.8$	ON	OFF	Charge
$V_{grid} < 125$	$V_{bat} > 14.5$	OFF	ON	Feed
$V_{grid} < 120$	$V_{bat} < 11.6$	OFF	OFF	No feed or charge

There are four possible conditions which can be happened for any individual/all customer in a cluster which is shown in table First, if grid line voltage increases and battery is fully charged (measured by MCU via HV_V and LV_V connection) then both MOSFET driver will be turned off. For example, if grid line voltage (V_{grid}) is greater than 130V and battery voltage (V_{bat}) is 14.5V then both MOSFET driver is OFF. Another MOSFET drivers turned off stage is that when V_{grid} is less than 120V and V_{bat} is also less than 11.6V. MOSFET driver will alternatively turn ON when other two conditions arrived. Firstly, if V_{grid} is greater than 125V and V_{bat} is less than 13.8V, then isolated MOSFET gate driver will turn ON and charge the battery. Otherwise, if V_{grid} is less than 125V and V_{bat} is higher than 14.5V (battery fully charged) MCU will turn ON the low side gate driver to feed power in to the grid line. In addition, there are two user modes which is pre-programmed in the MCU. One is automatic mode, in which selling and buying are done automatically by MCU. Manual mode, where user input helps MCU to decide when to sell or buy for his own system. In manual mode, if battery voltage V_{bat} is higher than 12.8V then only a user can capable to sell energy (leveled as "Forced selling") before that threshold voltage there is no option for selling. And another option is no sell option under manual user mode. Buying is always automatic and controls by MCU. Apart from the voltage measurement line, in addition there are two current measurement connection. The high side current measurement connection (HV_I) used here to measure power feeding current and low side current measurement connection (LV_I) used here to measure battery charging current. Apart from that, these two lines are also used to prevent faults on both sides such as ground fault, short circuit, overload, and line to neutral fault by measuring excess amount of current (assume 100W maximum load) flow on both sides. The overall device charging and power feeding capacity is maximum 100W at a single time.

c. Solar charge controller operation

Solar charge controller is basically a charge and discharge limiter of a battery. It prevents a battery to get over charged as well as over discharged.

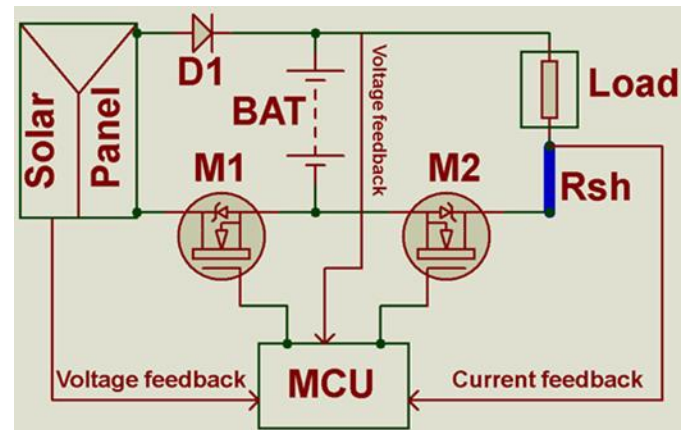


Figure 9: Logical block diagram of solar charge controller.

In figure 9, a logical solar charge controller diagram is presented. Here, two power MOSFET M1 and M2 are used to control charge and discharged consequently. Diode D1 prevents reverse current leakage protection during night. A shunt resistor Rsh is used to measure load current. This low value resistor helps MCU to detect overload, short circuit conditions via current feedback path. Two voltage feedback, one from battery (BAT) and another from solar panel (PV) is used to control discharge and charge stages of battery. MCU gives control pulses to MOSFETs M1 and M2 when battery reaches different threshold levels. For example, if battery voltage (Vbat) is higher than 14.5V then MCU will turn OFF MOSFET M1 and again turn ON when Vbat drops below 13.8V. It ensure battery will never get over charged. On the other hand, if Vbat is lower than 11.6V then MCU will turn OFF MOSFET M2 and prevent over discharge. MOSFET M2 will turn ON if Vbat is higher than 12.6V.

The overall self-consumption measured for the device is 25mA at 12V. The peak current consumption is 115mA at 12V while communicating with mobile app via wifi module. The overall device efficiency is recorded in various stages. Figure 10 represents, buck mode efficiency vs load curve at three different voltage levels where LV side output voltage is 12.8V. When Vin=128V then the efficiencies are around 85% at half load and 87% at full load. When Vin=122V then the efficiencies are around 82% at half load and 86% at full load. Finally, when Vin=125V then the efficiencies are around 84% at half load and 88% at full load. Therefore, it is clear from the experimental data of the efficiency vs load curve that the overall device efficiency at buck mode is greater than 82% in half rated load and 87% rated load conditions. However, device efficiency is recorded below 75% when it exits rated load. Implementation of overload protection in LV side is the reason behind this behavior of the device.

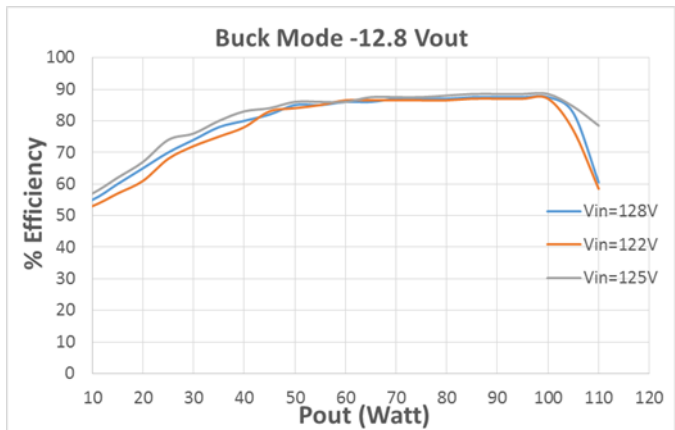


Figure 10: Efficiency vs Load (Buck mode with different Vin)

Figure 11 represents, boost mode efficiency vs load curve at three different voltage levels where HV side output voltage is 125V. When Vin=12.5V then the efficiencies are around 69% at half load and 83% at full load. When Vin=13.8V then the efficiencies are around 71% at half load and 85% at full load.

Finally, when Vin=14.5V then the efficiencies are around 80% at half load and 86% at full load.

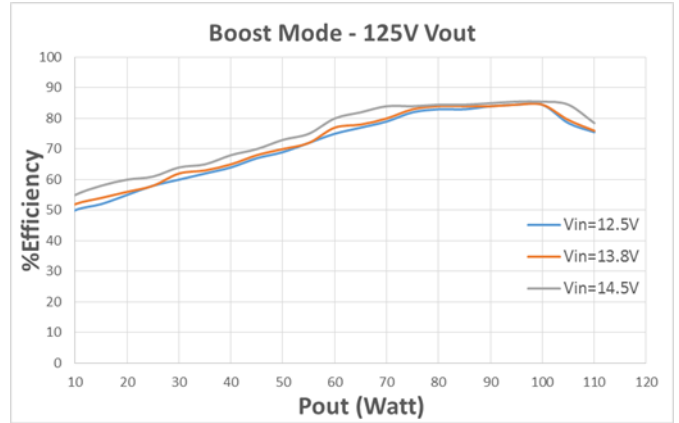


Figure 11: Efficiency vs Load (Boost mode with different Vin)

Therefore, it is clear from the experimental data of the efficiency vs load curve that the overall device efficiency at boost mode is greater than 70% in half rated load and 86% rated load conditions. However, device efficiency is recorded below 80% when it exits rated load. Implementation of overload protection in HV side is the reason behind this behavior of the device. In addition, boost stage efficiency is little bit lower than the buck stage. Because in boost stage, one additional inductor L2 heat loss (I^2R loss) is incorporated. The device efficiency is also related to the output filters power consumption, heat loss in high frequency transformer and switching losses. If all these losses get reduces, than the overall device efficiency can be greater than 90%.

IV. MOBILE APP DEVELOPMENT

Mobile app is an application software that developed only for use on small, wireless computing devices like as smartphones and tables. In this research work a mobile app has been designed for communicating wirelessly with device's (bi-directional converter) MCU. By this wireless communication, one can check device status, balance update of specific device ID (customer ID), cash in and cash out. Here, pair to pair (P2P) communication is used by the help of wifi module and mobile app. Here is the outlook on the mobile app.

Figure 12 describes about the password protected user login option in first scene (Account login). It ensures mobile app user's authentication and then give him the option of choosing further facilities in second scene (user account verified). Then after two second time delay mobile will let in to further activities options for this user.

From figure 13, it has been seen that after entering customer ID number ("Smart box" number) and pressing enter button, it will give three different options cash in, cash out and status. By pressing status button it will display customer present status. For example, customer name, his location, mobile number, current balance, total energy produced till date, today's energy

production rate, amount of total cash in and cash out with number of recharge as well as cash out times.

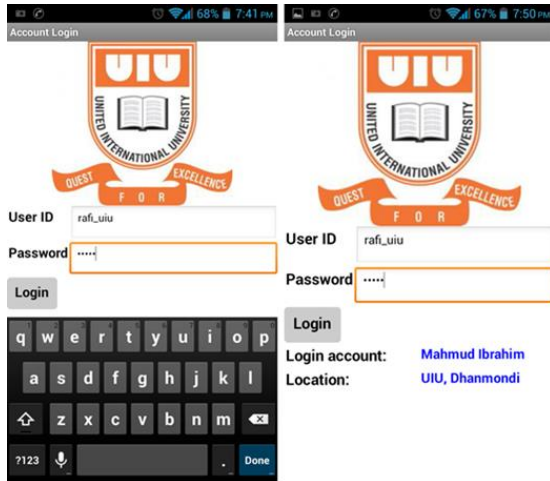


Figure 12: Password protected user login option.

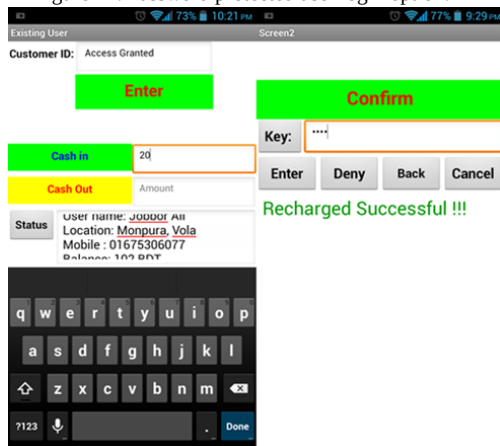


Figure 13: Mobile app user activities.

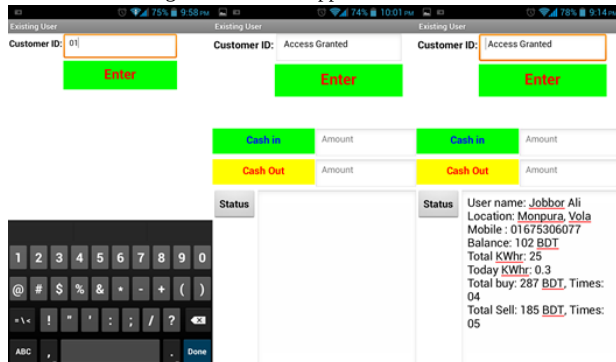


Figure 14: "Cash In" input and confirmation message by mobile app.

In figure 14, after putting the cash in or recharge amount (first screen) mobile app will automatically ask confirmation (second screen). Then user need to put the confirmation key number in the key box and press enter. Then mobile app will communicate with device's MCU via wifi module and update the present balance.

This mobile app has been develop by using MIT app inventor software. It's an open source graphical user interface (GUI) programming platform.

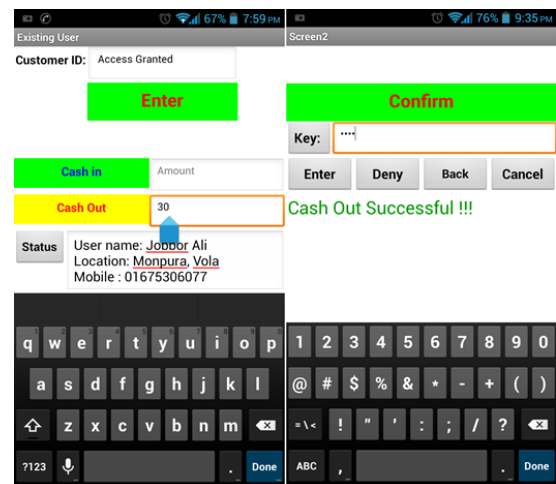


Figure 15: "Cash Out" input and confirmation message by mobile app.

V. DISCUSSION AND CONCLUSION

SHS program has been a great success in Bangladesh. Utilization of unutilized energy of individual SHS will be a step further if Solar DC Nano grid topology is applied in existing SHS systems. It will enhance living standards of the people of the off-grid areas. It will also enhance socio-economic development of off-grid areas of country like Bangladesh. Therefore, bi-directional DC-DC converter with pre-paid billing integrated device ("Smart box") would be better solution for implementing Solar DC Nano grid without large change in existing SHS system. This research work also open the scope of further research where multiple energy clusters can share energy between each other and then finally feed into country's national grid.

REFERENCES

- [1] [online] solar home system program,2014, <http://idcol.org/home/solar>
- [2] Hannes Kirchhoff , Identifying hidden resources in solar home systems as the basis for bottom-up grids, Book- Decentralized solutions for developing economics,chnp-2, published by Springer
- [3] BPDB, B. P. (2014). *Annual Report 2012 -13*. Dhaka: BPDB.
- [4] MoPEMR, Ministry of Power Energy and Mineral Resouyrces, Annual Report 2014-2015
- [5] M.Rezwan Khan, A Stand Alone DC Microgrid for Electricity Supply in Rural Bangladesh, ICDRET-2012
- [6] M. Rezwan Khan, DC Nanogrids: A Low Cost PV Based Solution for Livelihood Enhancement for Rural Bangladesh, ICDRET-2014
- [7] [online] Song or SHS : A Comparative Overview, 2014, <http://solarnanogrids.org/prospect-of-solar-nano-grid-in-kenya-bangladesh-a-comparative-overview/>
- [8] Microgrids: Benefits and Barriers, Paul Savage, R. R. (2011), pp-51-66, Book- From Silos to Systems: Issues in Clean Energy and Climate Change
- [9] Sebastian Groh, Swarm Electrification: Investigating a paradigm shift throug the building of microgrids bottom-up, Book- Decentralized solutions for developing economics,chnp-2, published by Springer
- [10] Texas Instrument design note: Bi-directional DC-DC converter-2015, Texas Instrument,US

Efficiency Improvement of Organic Light Emitting Diodes (OLEDs) Using Micro/Nano-Structures

Md. Shamim Ahsan, Md. Rezvi Hossain,
Sabiha Mehbuba, K. M. Sarif Islam, Md. Ziaul Amin
Electronics and Communication Engineering Discipline
Khulna University
Khulna 9208, Bangladesh
shamim@kaist.ac.kr, rezvi.amity@gmail.com,
sabiha.bithi@yahoo.com, khan.anik.sarif@gmail.com,
ziaulece@gmail.com

Ik-Bu Sohn
Advanced Photonics Research Institute (APRI)
Gwangju Institute of Science and Technology (GIST)
Gwangju 500-712, Republic of Korea
ibson@gist.ac.kr

Abstract— This paper demonstrates the design of high efficiency OLEDs by incorporating micro/nano-metric structures on the surface of and inside the substrate layer of the OLEDs. The OLEDs have been designed and simulated using Lumerical FDTD simulator. We also analyze the performance of the OLEDs from various perspectives such as quantum efficiency, normalized transmitted power, intensity, and angular distribution of light. To increase light extraction from the substrate layer, various micro/nano-metric structures, made of Si_3N_4 , have been incorporated as light scattering structures. Single layer structures of various size, shape (plano-convex & plano-concave lenses and pillars), and period have been investigated during our simulation. Simulation results confirm higher light extraction, angular distribution, and efficiency in micro/nano-metric structures induced OLEDs, compared to traditional OLEDs. Plano-concave MLA engraved OLEDs show high out-coupling efficiency, whereas, nano-pillars engraved OLED shows consistent light distribution in a wide angular region. We strongly believe that, the proposed OLEDs will have wide range of commercial applications.

Keywords—organic light emitting diode; micro/nano-structures; quantum efficiency; finite difference time domain (FDTD); Si_3N_4

I. INTRODUCTION

The discovery of the electroluminescence in organic crystals by Martin Pope and his research group in 1963 has opened a new era, especially in the field of OLEDs [1]. However, the development of organic electroluminescence based devices is very slow due to the requirement of high voltage as well as for low efficiency. About 29 years ago, in 1987, Ching W. Tang and Steve Van Slyke have discovered the first efficient light-emitting device based on small molecule by using tris (8-hydroxyquinoline) aluminum Alq_3 in its active layer [2]. The external quantum efficiency of the first OLED was only 1%. In 1990, Richard Friend's group invented a PPV based OLED at Cambridge University, shortly called PLED or polymer-LED, having quantum efficiency of only 0.05% [3]. Afterwards, a good number of research works have been demonstrated to improve the efficiency, color gamut, and device reliability of the OLEDs [4,5].

The fundamental limitations of conventional OLEDs are their low out-coupling or light extraction efficiency and viewing angle. We know that, a conventional OLED has a

maximum out-coupling efficiency of 30% because some of the generated light is confined by the interface between the glass substrate and transparent conducting oxide. To resolve the low out-coupling or light extraction efficiency, many attempts have been taken by introducing random texturing on the substrate layer on glass substrate [4,5] and photonic crystals [6]. Micro-lens arrays (MLAs) or mesh pattern have also been tested by the researchers [7–9]. Due to the addition of cylindrical MLAs on substrate surface, the external quantum efficiency of the OLEDs has increased to $\sim 45\%$ [9]. Nano-scale Si_3N_4 layer has also been investigated on the substrate layer of OLEDs to increase the external quantum efficiency up to 50% [10,11]. Although the external quantum efficiency has increased significantly because of the addition of micro/nano-structures in OLEDs, still there is scope the external quantum efficiency of the OLEDs. In addition, it is a challenge to increase the distribution angle of light in the forward direction.

In this paper, we demonstrates the design of highly efficient OLEDs by incorporating several unique light scattering structures of Si_3N_4 on the surface of and inside SiO_2 substrate to improve the light out-coupling efficiency and the viewing angle of OLEDs. Starting with plano-convex micro-lenses, we investigate plano-concave & conical micro-lenses, and rounded nano-pillars in single layer. We examine various optical properties of the OLEDs such as normalized transmitted power, out-coupling efficiency of light, light intensity, and angular distribution of light using Lumerical FDTD simulator. All of the micro/nano-structures engraved OLEDs show better efficiency compared to traditional OLEDs where plano-concave MLA shows highest efficiency. However, nano-pillars engraved OLED expresses consistent light distribution in different angles. These OLEDs show promises for commercial applications in display devices and as a lighting source.

II. FUNDAMENTAL CONCEPTS

A typical OLED has four fundamental components: substrate, anode, organic layers, and cathode. One or more thin films made of semiconducting organic molecules are sandwiched between two electrodes among which one must be transparent [4]. The fabrication of an OLED is done by sequentially depositing thin films on the substrate, which is made of glass or polymers. The anode is made of transparent materials that can pass the light produced inside the OLED. The anode is responsible for removing electrons when current

flows through the device. Organic layers are the most important components of the OLED, which are made of polymers or small organic molecules. Conducting layer is made of organic plastic molecules, which is responsible for sending electrons out from the anode. Emissive layer is also made of plastic molecules but different from that is used in conducting layer. The emissive layer transport electrons from the cathode and produces light. The cathode is the component which injects electrons into the emissive layer when a forward current flows through the device. The cathode may be made of transparent (for transparent OLED) or non-transparent (for conventional OLED) metal. Fig. 1 shows a simplified schematic diagram of a traditional OLED.

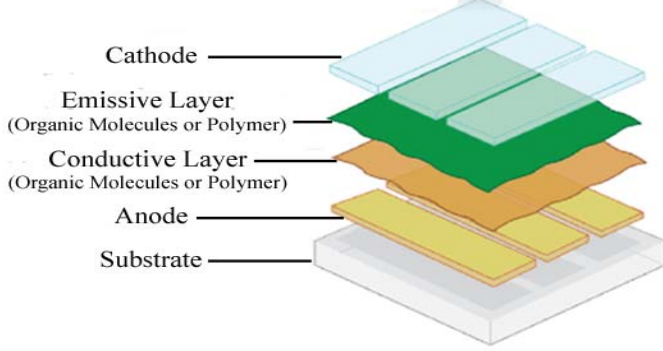


Fig. 1. Schematic diagram of a typical OLED.

Internal quantum efficiency is an important parameter for characterizing OLED. When the forward bias voltage or current is applied to the terminal of an OLED, the electrons generated by the cathode and the holes generated by the anode make pairs for their attraction to each other. Then, the electron-hole pairs are recombined. This recombination can be of two types: radiative recombination—responsible for producing light and non-radiative recombination—responsible for generating heat as loss. Internal quantum efficiency (η_{in}) means to the percentage of electron-hole pairs which are recombined in radiative manner, which can be expressed as the ratio of the radiative recombination rate (R_r) to the total recombination rate, as in (1).

$$\eta_{in} = \frac{R_r}{R_r + R_{nr}} \quad (1)$$

where R_{nr} represents the non-radiative recombination rate. The most important parameter to characterize the performance of an OLED is the external quantum efficiency (η_{ext}), also termed as light out-coupling efficiency. External quantum efficiency can be defined as the ratio of power of the transmitted light (P_t) to the power of the light source (P_d) (in our case, dipole power), as in (2).

$$\eta_{ext} = \frac{P_t}{P_d} \quad (2)$$

The out-coupling efficiency of OLEDs in air for a smooth substrate can be estimated as $\eta_{ext} = 0.5 \times n^{-2} \times 100\%$ where n represents the refractive index of the substrate [12]. However, addition of micro/nano-metric structures can improve the external quantum efficiency significantly, as shown in Fig. 2.

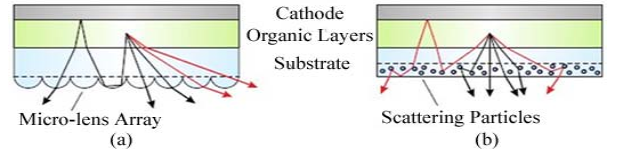


Fig. 2. Improvement of light out-coupling by micro/nano-structures. (a) Micro-lens array; (b) scattering particles.

The efficiency is directly related to the size and shape of the structures. Light paths indicated in red correspond to radiation that would normally not be able to leave the device due to total internal reflection at the glass/air interface. Multiple reflections inside the stack reduce the efficiency due to absorption in the organic layers as well as due to reflection losses at the cathode.

III. MATERIALS AND METHODS

A. Materials

Our proposed OLEDs consist of basic five layers without light scattering structures: anode, cathode, hole-transport molecular layer, electron-transport molecular layer, and substrate. Fig. 3 represents the basic structure of the proposed OLEDs without additional light scattering structures.

Cathode (Ag)
Electron Transport Molecular Layer (Alq₃)
Hole Transport Molecular Layer (α-NPD)
Anode (ITO)
Glass Substrate (SiO₂)

Fig. 3. Basic structure of the proposed OLEDs without having any light scattering structures.

Anode: We used indium tin oxide (ITO) as anode because of its wide acceptance. It has a high work function and high transparency (90%) to visible light, good electrical conductivity, and excellent adhesion to the substrates.

Cathode: We considered silver (Ag) as cathode due to its high electrical conductivity.

Hole-transport molecular layer: We placed α -NPD (N,N'-di(1-naphthyl)-N,N'-diphenyl-(1,1'-biphenyl)-4,4'-diamine) as hole-transport molecular layer.

Electron-transport molecular layer: We utilized Tris (8-hydroxyquinolino) aluminum (Alq₃) as electron-transporting material due to its wide use in OLEDs as a good electron transporter. In addition, Alq₃ is thermally and morphologically stable to be evaporated into thin films, robust, and easily synthesized & purified.

Substrate: In all of our designs, we considered silicon-dioxide (SiO₂) glass as the substrate of the OLEDs. To increase the out-coupling efficiency of our proposed OLEDs, a large variety of micro/nano-structures of Si₃N₄ have been incorporated on the surface as well as inside the SiO₂ substrate. Table I summarizes the materials' and physical properties of various layers of the OLEDs.

TABLE I. MATERIALS AND CORRESPONDING PHYSICAL PARAMETERS OF VARIOUS LAYERS OF THE OLEDs

Layer name	Length (μm)	Thickness (μm)	Material	Refractive index
Cathode	50	0.6	Ag	6.67 at 688 nm
Anode	50	0.15	ITO	1.8
Hole-transport molecular layer	50	0.05	α -NPD	1.763 at 683 nm
Electron-transport molecular layer	50	0.6	Alq ₃	1.68
Substrate	50	2.0	SiO ₂	1.158 at 675 nm

B. Physical Parameters of Various Layers of the OLEDs

To design our OLEDs, we varied the length and thickness of different layers. We considered fixed length (50 μm) for all the layers although the thickness of various layers was different, as summarized in Table I. The thickness of any particular layer was constant for all kinds of design.

C. Physical Parameters of the Micro/nano-metric Structures

As mentioned before, we designed a large variety of OLEDs by adding various micro/nano-structures including micro/nano-lenses (plano-convex, plano-concave, and conical) and nano-pillars in the SiO₂ glass substrate. Initially, we added an array of plano-convex circular lenses having radius of 0.3 μm on the substrate surface without any spacing in between the lenses. Later on, the radius was increased to 0.7 μm . Fig. 3 shows the design of plano-convex lens engraved OLEDs.

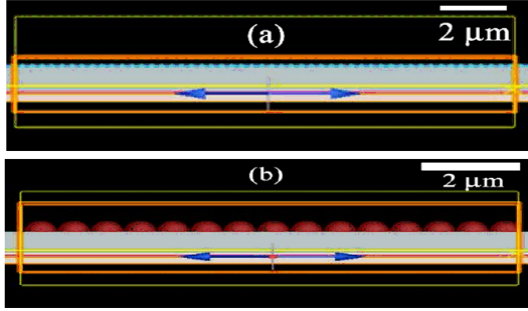


Fig. 3. Design of OLEDs consisting plano-convex Si₃N₄ micro/nano-lenses of different radius on SiO₂ glass surface. (a) Radius: 0.3 μm ; (b) radius: 0.7 μm .

Afterwards, we replaced the plano-convex Si₃N₄ lenses by plano-concave circular Si₃N₄ micro-lens arrays (MLAs) of different radius (0.3 μm and 1 μm) without having any space in between (as depicted in Figs. 4(a-b)). On the contrary, we incorporated Si₃N₄ concave lenses having radius of 1 μm and spacing of 0.5 μm among the lenses in both horizontal and vertical directions, as shown in in Fig. 4(c). Furthermore, we placed Si₃N₄ conical lenses of various sizes inside SiO₂ glass substrate. The height, width, and angle of the conical structures were varied to investigate the impact of physical shape of the lenses on the intensity and efficiency of the OLEDs. The height (H) of the conical structures was varied from 0.3 μm to 1 μm , whereas, the solid angle (A) of the conical structures was varied from 50–90 degrees. The spacing (S) among the conical structures was either 0 μm or 0.4 μm . Fig. 5 shows various conical lenses incorporated OLEDs.

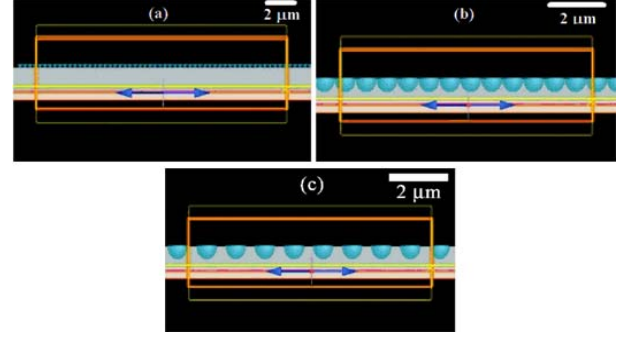


Fig. 4. Design of OLEDs consisting plano-concave Si₃N₄ micro/nano-lenses of different radius in SiO₂ glass substrate. (a) Surface, Radius: 0.3 μm , spacing: 0 μm ; (b) inside, radius: 1 μm ; spacing: 0 μm ; (c) inside, radius: 1 μm , spacing: 0.5 μm .

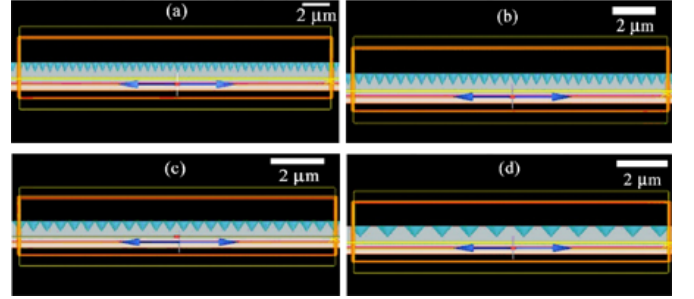


Fig. 5. Design of OLEDs consisting conical lenses of different physical parameters inside SiO₂ glass substrate. (a) H : 0.8 μm , A : 50°, S : 0 μm ; (b) H : 1 μm , A : 60°, S : 0 μm ; (c) H : 1 μm , A : 80°, S : 0 μm ; (d) H : 1 μm , A : 90°, S : 0.4 μm .

We also designed OLEDs by placing round-shaped nano-pillars array on top of the SiO₂ glass substrate. The height, width, and spacing of the nano-pillars were 300 nm, 100 nm, and 50 nm, respectively. Fig. 6 represents the design of nano-pillars engraved OLED.

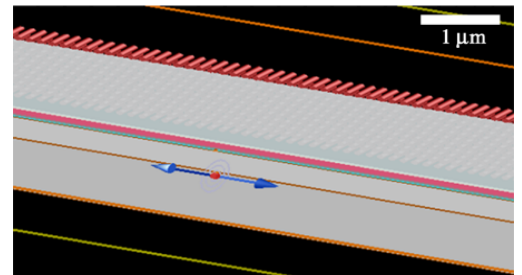


Fig. 6. Top view of the OLED having nano-pillars array on top of the SiO₂ glass substrate.

D. Simulation Parameters

Our designs were simulated in the 2D version of Lumerical FDTD simulator. Firstly, we selected the simulation region from one micrometer below the cathode ($-1\ \mu\text{m}$) to few micrometers above the substrate ($2.5\ \mu\text{m}$) through the y-axis and from $-15\ \mu\text{m}$ to $+15\ \mu\text{m}$ in the x-axis in order to cover the whole area of light spreading. To investigate the optical properties of the OLEDs and analyze the performance of light coming out from the OLEDs at different locations, we selected three different observation points: Far field-1, Far field-2, and Near field. The Far field-1, Far field-2, and Near field were $0\ \mu\text{m}$, $2.5\ \mu\text{m}$, and $> 2.5\ \mu\text{m}$ apart from the top surface of the substrate. We selected the mesh order as 6. The simulation was conducted at 300 K temperature and the total simulation time was 500 fs. During our simulations, we considered visible spectrum of light (wavelength range: 400–700 nm) as the wavelength of the built-in dipole. We applied 301 simulation points for each wavelength to simulate the transmitted power, dipole power, and the angular distribution of light. The normalized dipole power is plotted in Fig. 7.

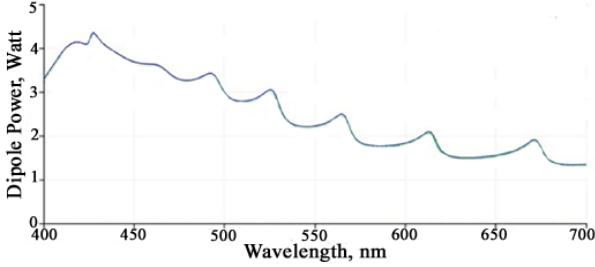


Fig. 7. Normalized dipole power.

IV. RESULTS AND DISCUSSION

We simulated the basic OLED structure along with various light scattering structures engraved OLEDs. We investigated angular distribution of light, intensity profile, and power distribution for all kinds of the designed OLEDs.

A. Basic OLED Device Without Light Scattering Structures

The angular distribution of light in Far field-1 and Far field-2 regions of the basic OLED (without having any light scattering structures) is illustrated in Fig. 8. In the Far field-1 region (Fig. 8(a)), there is a blur portion between -40° to $+40^\circ$ indicating poor light intensity in this region. The maximum intensity at Far field-1 region is approximately 2.8×10^{-5} cd. In the Far field-2 region (Fig. 8(b)), the angular distribution has increased from -70° to $+70^\circ$. However, the maximum value of intensity at Far field-2 region has decreased significantly (0.8×10^{-5} cd) indicating poor performance of the basic OLED device. The normalized transmitted power of the basic OLED for different field regions is plotted in Fig. 8(c). This result indicates that, a great portion of the dipole power is trapped inside the OLED and results in poor performance of the basic OLED device.

B. Plano-convex MLA Engraved OLEDs

The angular distribution of light in Far field-1 and Far field-2 regions of various plano-convex MLA engraved OLEDs are illustrated in Fig. 9. Addition of plano-convex

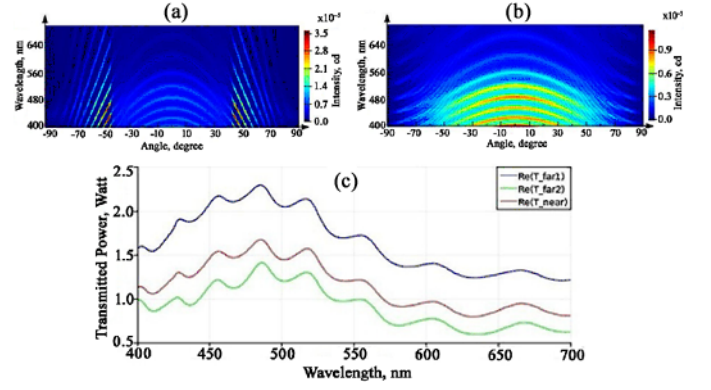


Fig. 8. Intensity distribution and power distribution of light for basic OLED. (a–b) Intensity distribution: (a) Far field-1, (b) Far field-2; (c) power distribution.

MLA of different radius on the surface of SiO_2 substrate has improved the light extraction performance of the OLEDs. The angular distribution of light and the intensity were increased as well. The maximum intensity at Far field-2 for the $0.3\ \mu\text{m}$ and $0.7\ \mu\text{m}$ MLA engraved OLEDs were 2.5×10^{-5} cd and 3×10^{-5} cd. The angular distribution of light for $0.7\ \mu\text{m}$ MLA engraved OLED was increased up to $\pm 74^\circ$, which shows continuous intensity in a wide angular range. Figs. 3(e–f) represents the normalized transmitted power at different locations for the convex MLA engraved OLEDs. The power has increased compared to basic OLED structure.

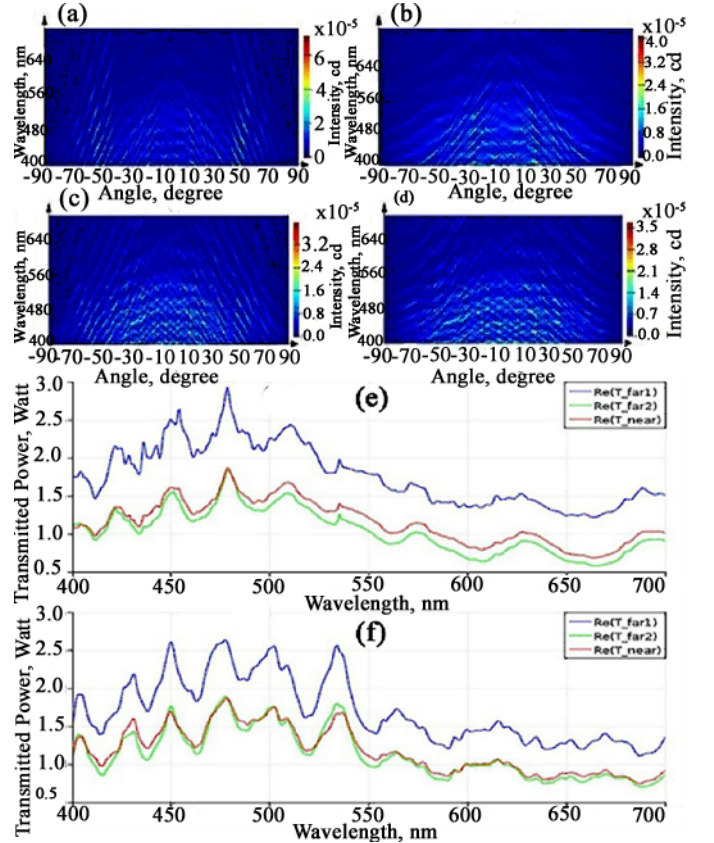


Fig. 9. Intensity distribution of light for plano-convex MLA engraved OLEDs: (a–b) Far field-1 and Far field-2 for lenses having radius (R) of $0.3\ \mu\text{m}$ (Fig. 3(a)), (c–d) Far field-1 and Far field-2 for lenses having $R = 0.7\ \mu\text{m}$ (Fig. 3(b)); (e–f) power distribution for lenses of Figs. 3(a) and 3(b).

C. Plano-concave MLA Engraved OLEDs

Replacing the convex MLA with concave MLA having radius of $0.3\ \mu\text{m}$ provided higher intensity level at the Far field-2 (maximum of about $3 \times 10^{-5}\ \text{cd}$) but showed poor angular distribution ($\pm 55^\circ$), as depicted in Fig. 10(b). Increase in the radius ($1\ \mu\text{m}$) of concave MLA causes the increase in the angular distribution ($\pm 88^\circ$) of light where the maximum intensity reached to only $2.4 \times 10^{-5}\ \text{cd}$ at Far field-2 region, as shown in Fig. 10(d). Addition of $0.5\ \mu\text{m}$ between the lenses decreased the angular distribution ($\pm 85^\circ$) but increased the intensity up to $4 \times 10^{-5}\ \text{cd}$, as represented in Fig. 10(f). Figs. 10(g–i) showed significant improvement of normalized transmitted power of the plano-concave MLA engraved OLEDs compared to plano-convex MLA engraved OLEDs and basic OLED.

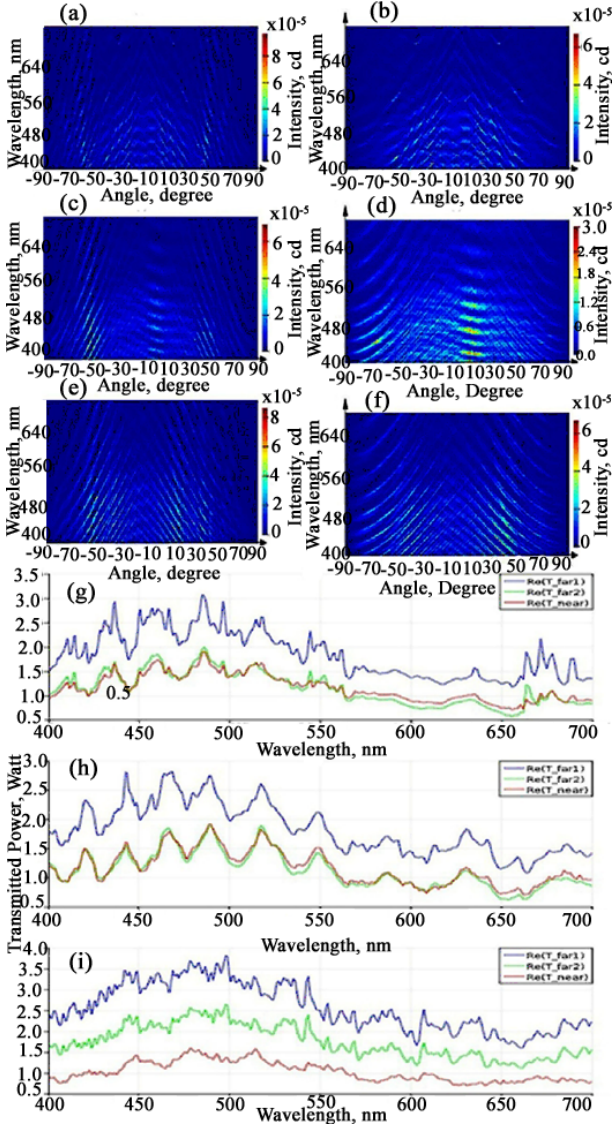


Fig. 10. Intensity distribution of light for plano-concave MLA engraved OLEDs: (a–b) Far field-1 and Far field-2 for lenses having $R = 0.3\ \mu\text{m}$ and spacing (S) of $0\ \mu\text{m}$ (Fig. 4(a)), (c–d) Far field-1 and Far field-2 for lenses having $R = 1\ \mu\text{m}$ and $S = 0\ \mu\text{m}$ (Fig. 4(b)), (e–f) Far field-1 and Far field-2 for lenses having $R = 1\ \mu\text{m}$ and $S = 0.5\ \mu\text{m}$ (Fig. 4(c)); (g–i) power distribution for lenses of Figs. 4(a), 4(b), and 4(c), respectively.

D. Conical MLA Engraved OLEDs

Conical MLA engraved OLEDs showed better result in terms of intensity compared to other micro/nano-structured OLEDs. However, in most of the cases, the angular distribution of light showed poor result ($\pm 60^\circ$ to $\pm 70^\circ$), as shown in Figs. 11(a–f).

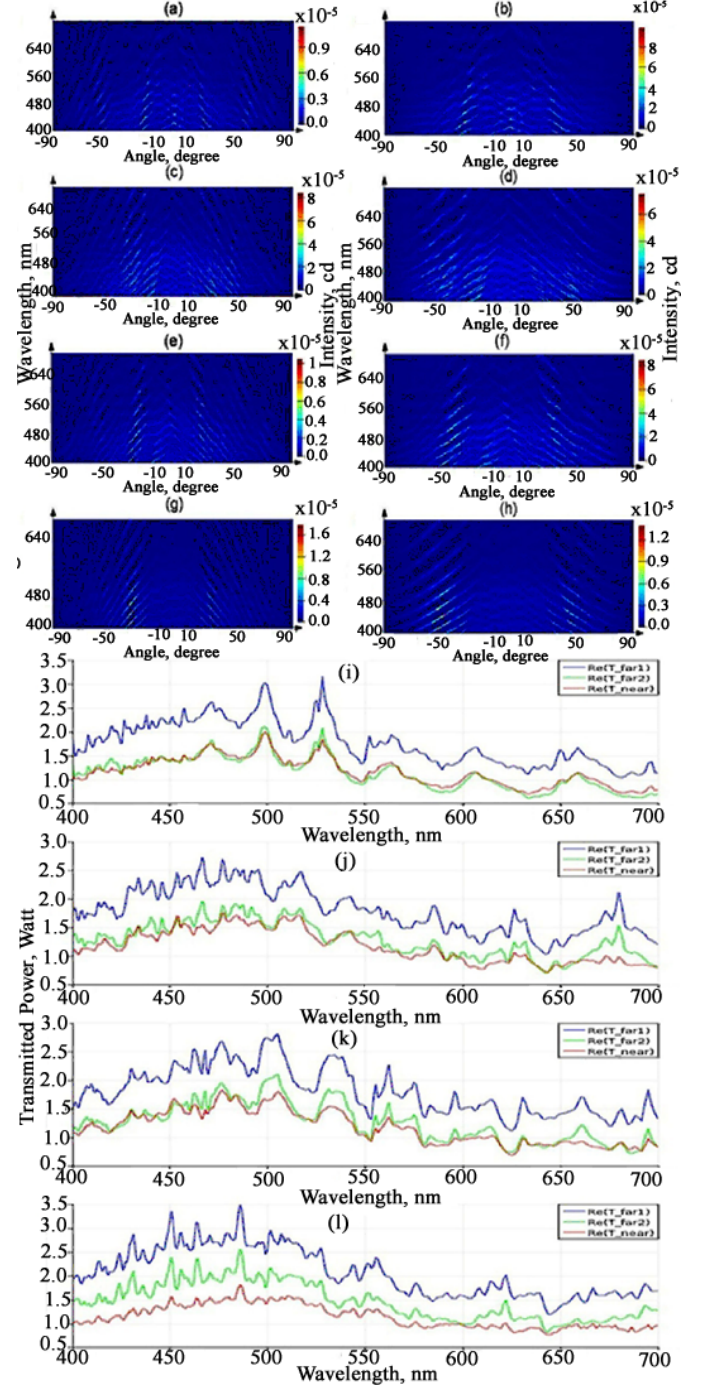


Fig. 11. Intensity distribution of light for conical MLA engraved OLEDs: (a–b) Far field-1 and Far field-2 for lenses having solid angle (A) of 50° and $S = 0\ \mu\text{m}$ (Fig. 5(a)), (c–d) Far field-1 and Far field-2 for lenses having $A = 60^\circ$ and $S = 0\ \mu\text{m}$ (Fig. 5(b)), (e–f) Far field-1 and Far field-2 for lenses having $A = 80^\circ$ and $S = 0\ \mu\text{m}$ (Fig. 5(c)), (g–h) Far field-1 and Far field-2 for lenses having $A = 90^\circ$ and $S = 0.4\ \mu\text{m}$ (Fig. 5(d)); (i–l) power distribution for lenses of Figs. 5(a), 5(b), 5(c), and 5(d), respectively.

However, when the angle was 90° and spacing was $0.4 \mu\text{m}$, the angular distribution of light was increased to $\pm 80^\circ$. Highest intensity ($3.7 \times 10^{-5} \text{ cd}$) at the Far field-2 was observed when the angle of the cones were 80° without any spacing in between the cones (Fig. 11(e–f)). The average power coming out from the conical structures incorporated OLEDs were higher than the basic OLED device.

E. Nano-pillars Array Engraved OLEDs

The nano-pillars array engraved OLEDs showed good angular distribution of light ($\pm 80^\circ$) and consistency in intensity, as depicted in Fig. 12(b). The maximum intensity of this OLED was $\sim 1 \times 10^{-5} \text{ cd}$, which is higher than the basic OLED but lower than other micro/nano-structured OLEDs. The transmitted power from the nano-pillars array engraved OLED was almost similar to the basic OLED, as shown in Fig. 12(c).

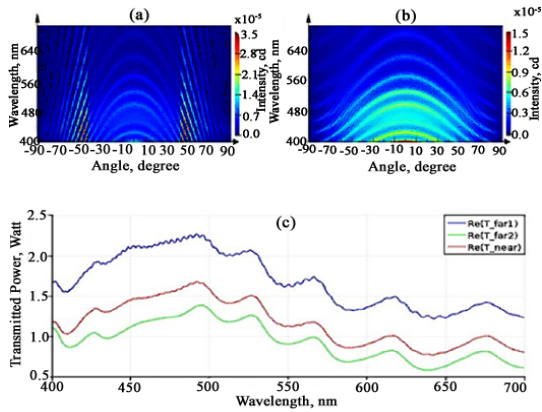


Fig. 12. Intensity distribution and power distribution of light for nano-pillars (height: 300 nm , width: 100 nm , spacing: 50 nm) array engraved OLED (Fig. 6): (a) Far field-1; (b) Far field-2; (c) power distribution.

Furthermore, we calculated the out-coupling efficiency of various OLEDs under different wavelengths, which are summarized in Table II. High and consistent out-coupling efficiency in all the wavelengths were observed for the plano-concave Si_3N_4 lenses (radius: $1 \mu\text{m}$, spacing: $0.5 \mu\text{m}$) engraved OLED of Fig. 4(c).

TABLE II. OUT-COUPLING EFFICIENCY OF VARIOUS OLEDs

Figure No.	Out-coupling Efficiency		
	474 nm (Blue)	550 nm (Green)	610 nm (Red)
Fig. 8(c)	36.36%	41.30%	25.00%
Fig. 9(e)	53.03%	47.83%	35.00%
Fig. 9(f)	54.55%	43.48%	52.50%
Fig. 10(g)	42.42%	58.70%	37.50%
Fig. 10(h)	43.94%	60.87%	39.00%
Fig. 10(i)	74.24%	76.09%	77.50%
Fig. 11(i)	46.97%	43.48%	55.00%
Fig. 11(j)	55.15%	63.04%	57.50%
Fig. 11(k)	59.09%	50.00%	85.00%
Fig. 11(l)	54.54%	71.74%	65.00%
Fig. 12(c)	37.88%	36.96%	40.00%

From the simulation results we find that, plano-concave MLA engraved OLEDs, especially MLAs having spacing in between the lenses (Fig. 10(f)) shows best transmitted power among all the designed OLEDs, as shown in Fig. 13. In addition, this OLED shows good angular distribution of light with consistent intensity.

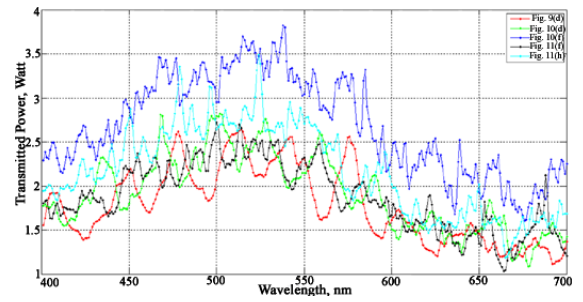


Fig. 13. Comparison of normalized transmitted power of the OLEDs in the visible spectrum.

V. CONCLUSION

In summary, we designed high efficiency OLEDs by incorporating micro/nano-metric structures of Si_3N_4 in SiO_2 substrate of OLEDs. We investigated the out-coupling efficiency, power, intensity, and angular distribution of light. OLEDs were designed using plano-convex & plano-concave lenses, conical structures, and nano-pillars of different size, shape, and period. Plano-concave MLA engraved OLEDs showed highest average out-coupling efficiency with good intensity in a wide angular range. On the contrary, nano-pillars engraved OLED showed highest consistency in intensity distribution among all the designed OLEDs. The proposed OLEDs are expected to be valuable for various applications.

REFERENCES

- [1] M. Sano, M. Pope, and H. Kallman, "Electroluminescence in organic crystals," *J. Chem. Phys.*, vol. 38, pp. 2042–2043, 1963.
- [2] C. W. Tang and S. A. Vanslyke, "Organic electroluminescence diodes," *Appl. Phys. Lett.*, vol. 51, pp. 913, 1987.
- [3] R. H. Friend, R. W. Gymer, A. B. Holmes, J. H. Burroughes, R. N. Marks, C. Taliani, D. D. C. Bradley, D. A. D. Santos, J. L. Brédas, M. Lögdlund, and W. R. Salaneck, "Electroluminescence in conjugated polymers," *Nature*, vol. 397, pp. 121–128, 1999.
- [4] J. C. Scott, P. J. Brock, J. R. Salem, S. Ramos, G. G. Malliaras, S. A. Carter, and A. Bozano, "Charge transport processes in organic light-emitting devices," *Synth. Met.*, vol. 111–112, pp. 289–293, June 2000.
- [5] O. Vyavahare, "Fabrication and characterization of organic light emitting diodes for display applications," Thesis, Rochester Institute of Technology, New York, USA, August 2009.
- [6] L. S. Hung and C. H. Chen, "Recent progress of molecular organic electroluminescent materials and devices," *Mat. Sci. Engg.*, vol. 39, pp. 143–222, 2002.
- [7] J. Yeon, T. W. Koh, H. Cho, J. Chung, S. Yoo, and J. B. Yoon, "Actively transparent display with enhanced legibility based on an organic light-emitting diode and a cholesteric liquid crystal blind panel," *Opt. express*, vol. 21, pp. 10358–10366, 2013.
- [8] Markus Schwoerer and Hans Christoph Wolf, *Organic Molecular Solids*, 1st ed., WILEY_VCH, 2007.
- [9] M. K. Wei, J. H. Lee, H. Y. Lin, Y. H. Ho, K. Y. Chen, C. C. Lin, C. F. Wu, H. Y. Lin, J. H. Tsai, and T. C. Wu, "Efficiency improvement and spectral shift of an organic light-emitting device by attaching a hexagon-based microlens array," *J. Opt. A: Pure Appl. Opt.*, vol. 10, pp. 055302, 2008.
- [10] S. Nowy, N. A. Reinke, J. Frischeisen, and W. Brütting, "Light extraction and optical loss mechanisms in organic light-emitting diodes," *Proc. SPIE*, vol. 6999, 2008.
- [11] S. J. Park, Y. D. Kim, H. W. Lee, H. J. Yang, J. Y. Cho, Y. K. Kim, and H. Lee, "Enhancement of light extraction efficiency of OLEDs using Si_3N_4 -based optical scattering layer," *Opt. express*, vol. 22, pp. 12392–12397, 2014.
- [12] Jörg Frischeisen, "Light extraction in organic light-emitting diodes," Thesis, University of Augsburg, Augsburg, Germany, June 2011.

Power Factor Correction for Single Phase Loads using Modified Vienna Rectifier

Mir Humainul Islam¹ and M. Abdur Razzak²

Department of Electrical and Electronic Engineering

Independent University, Bangladesh

Plot-16, Block-B, Aftabuddin Ahmed Road, Bashundhara, Dhaka-1229, Bangladesh

iub.tutul@gmail.com¹, razzak@iub.edu.bd²

Abstract—This paper presents the design of a modified Vienna rectifier for single phase power factor correction (PFC) under RL load. This research proposes a newly developed scheme consist of a modified Vienna rectifier for converting AC to DC, a simple control circuit for controlling single phase modified Vienna rectifier switching, a single phase inverter (H-bridge) for converting power from DC to AC, a step down transformer and a T-LCL filter circuit. This article presents a simple control system for single phase modified Vienna rectifier which is supposed to control the DC output voltage and almost pure sinusoidal single phase input current to gain almost unity power factor. For switching pulse of the H-bridge inverter, a combination of square wave signal and SPWM (sinusoidal Pulse Width Modulation) under grid synchronization is employed. The simulation of proposed system has been done using PSIM simulation software. The propose topology provides almost ripple free input current, lower input current THD which is found to be about 5.6% and improved power factor up to 0.992.

Keywords—Single phase modified Vienna rectifier, Control circuit for modified Vienna rectifier, H-bridge Inverter, Step down transformer, T-LCL filter circuit, Power factor, THD.

I. INTRODUCTION

Power electronics are the main application of solid state engineering and control systems which are using for conversing and controlling of electric power. At present days, very high switching frequency devices are broadly used, due to power electronic technologies [1-2]. A device for two reasons have a poor power factor-1. It marks source current out of phase (phase displacement) with the source voltage. 2. It marks source current in a non-sinusoidal waveform [3]. Out of phase current components cannot deliver usable power [3] and non-sinusoidal waveforms have lot of harmonics that are creating huge problems like overheating, circulating currents, high voltage, component failures, cable damaging etc. To overcome these limitations we need power factor correction.

There are many topology to correct power factor. Vienna rectifier, boost converter, buck boost converter are used to correct power factor. Boost converter generates a high voltage stress in the wide input voltage range of adjustable electronic equipment [4]. Boost and buck boost both are used very complicated controller [4-5]. They provide huge path losses thereby decrease the efficiency. High voltage stress, low power density, huge path losses and switching

losses are the main disadvantages of boost and buck boost converter. Vienna rectifier normally uses for three phase power factor correction but it is also used for single phase power factor correction. Vienna rectifier used approximately 100KHz [6-7] as switching frequency. For using high switching frequency, the switching losses are very high and the sustainability of switches become low thereby decrease the efficiency. There are four conduction path in Vienna rectifier [6-8] which provides huge path losses. To overcome these limitation we proposed modified Vienna rectifier. Lower number of component are used in single phase modified Vienna rectifier and 21kHz are used as switching frequency which decreases the switching losses thereby increase the efficiency. The designed control topology of proposed rectifier is very light and reliable to compare with Vienna rectifier as well as buck-boost converter. The proposed system is used a highly efficient inverter aiming to provide high efficiency, low cost, low Distortions (less than 0.1%) and light controller [9-11].

The proposed system configuration consists of main six parts. There are: 1. single phase input voltage source, 2. A single phase modified Vienna rectifier to achieve boost type output DC, 3. An single phase H-bridge inverter to achieve AC, 4. A T-LCL filter circuit to deliver filtered as well as constant current & voltage, 5. A simple control circuit for single phase modified Vienna rectifier switching, 6. A step down transformer which used for switching of inverter [9]. The Block diagram of our proposed topology is shown in Fig.1.

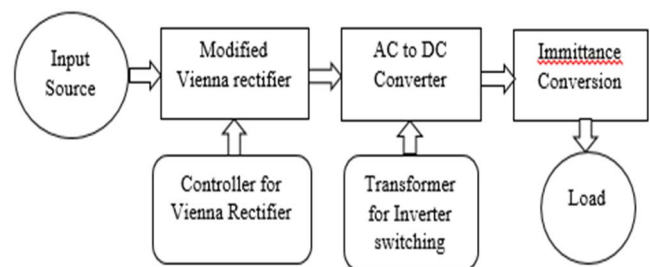


Fig.1. Block Diagram of Proposed Topology

II. DESIGN OF PROPOSED SINGLE PHASE MODIFIED VIENNA RECTIFIER

This section presents the design of modified Vienna rectifier aiming almost unity power factor and low input

current distortion. The proposed single phase modified Vienna rectifier is shown in Fig.2. V_{in} is the input voltage of grid side and SW is switch. L and C are the input side inductor and capacitor. In input side of proposed modified Vienna rectifier, the inductor is connected in series and capacitor is connected in parallel. Vienna rectifier consist of two output capacitor [6-9] but here we use only one capacitor. The switching strategy is very simple and reliable compare to conventional Vienna rectifier. In the proposed modified Vienna rectifier, there are two conduction paths that discussed in section IIC. The major concern is output capacitor selection and very large value of output capacitor is chosen so that the output voltage of rectifier can be observed as a constant value. This capacitor also acts as a filter that reduced DC ripples. The design parameters are listed in Table-I.

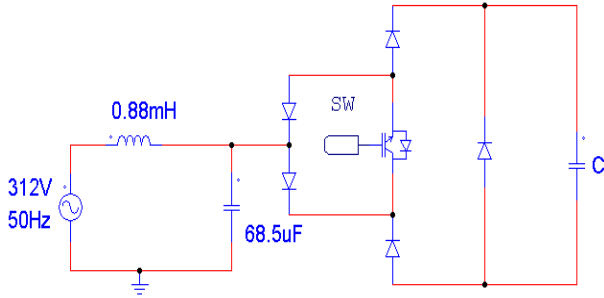


Fig.2. Proposed single-phase Modified Vienna Rectifier

TABLE-I DESIGN PARAMETERS OF MODIFIED VIENNA RECTIFIER

Symbol	Actual Meaning	Value
V_{in}	Single phase input voltage	312V
V_{out}	Desired output DC Voltage of rectifier	345V
I_{out}	Output current	16A
ΔI	Desired output current ripple (8.92% output current)	1.606A
I_L	Induction current (Max)	5.6A
f_s	Switching frequency of controller	21KHz
D	Duty cycle	0.50

A. Inductor Selection

Inductor is used in series with single phase input voltage to reduce the input ripple. In this conversion, inductor value can't be chosen from the data sheets because here the voltage conversion is happened at a huge scale. Therefore, the inductor value is selected using the following equation [12][13]. Input inductor L ,

$$L = \frac{V_{in} \times (V_{out} - V_{in})}{\Delta I \times f_s \times V_{out}} = \frac{312(345 - 312)}{1.6 \times 21000 \times 345} \approx 0.88mH$$

B. Capacitor Selection

In proposed modified Vienna rectifier, the capacitor is used in parallel with the input voltage and the capacitor is also used to reduce the input current ripple. So the capacitor value is selected using the equation below [12][13]. Input capacitor C ,

$$C = \frac{I_{out} \times D}{f_s \times I_L} = \frac{16 \times 0.5}{21000 \times 5.6} \approx 68.5uF$$

C. Operating Principle

Boost and Buck-Boost converter both are provided huge path losses thereby decrease the efficiency. There are 4

conduction paths [6-8] in Vienna rectifier which also provided huge path losses. But there are two conduction paths only in proposed modified Vienna rectifier as shown in fig.3, which reduce the path loss significantly.

Here we get output voltage across the capacitor only when switch is ON. If control switch SW is ON then conduction path is illustrated in Fig.3(A) for positive half circle. Similarly, if control switch SW is ON then conduction path is illustrated in Fig.3(B) for negative half circle. We get no output voltage across the capacitor when control switch is OFF.

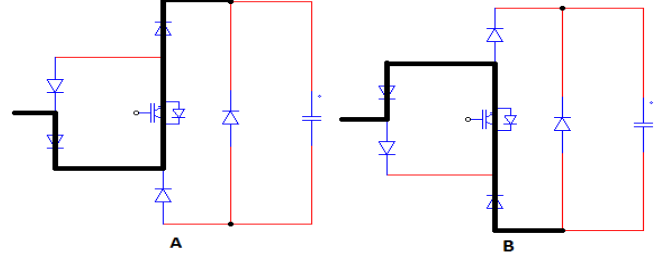


Fig.3. Conduction paths for proposed modified Vienna rectifier switching cell. Fig.3(A). For positive half circle. Fig.3(B). For negative half circle.

III. SYSTEM MODELING

The performance of the proposed system were performed using PSIM Simulation software. Fig.6. illustrates the model of modified Vienna rectifier for single phase power factor correction using R-L load and an highly efficient single phase H-bridge inverter. A p-MOSFET has been used for switching purpose in the proposed 1-ph modified Vienna rectifier and four p-MOSFETs have been used for switching purpose of inverter. The complete topology covered with a single phase modified Vienna rectifier, a single phase highly efficient inverter, a step down transformer, a simple control circuit for modified Vienna rectifier and a controller for inverter switching. In modified Vienna rectifier, the single phase input AC voltage is converted to boost type DC voltage. Here the rectifier rectified from 312V peak AC voltage to 345V DC Voltage. The sinusoidal voltage source is sampled from power grid by using step down transformer that step down from 220Volt RMS to 5Volt RMS. The H-bridge Inverter consists of a pair of two parallel MOSFETs. Only one type of switching technique is used in conventional inverter but here a combination of SPWM and square wave are used as switching technique to minimize switching losses by minimized the switching frequency thereby increases the efficiency. A T-LCL immittance conversion [14-15] circuit is used as filter circuit to reduce THDs (Total Harmonics Distortions) of load current as well as input current.

A. Design of Control Circuit for Modified Vienna Rectifier

This section provides the design of the controller of single phase modified Vienna rectifier. The control circuit has a summer which adds the reference voltage, V_r and the output voltage of rectifier, V_{out} and then the output of summer passes through the PI (Proportional-Integral) controller. The proportional integral controller (PI) is used to regulate the output voltage of summer. The output of PI controller multiplied with sinusoidal voltage source by a multiplier. The multiplier have a sinusoidal input of 50 Hz.

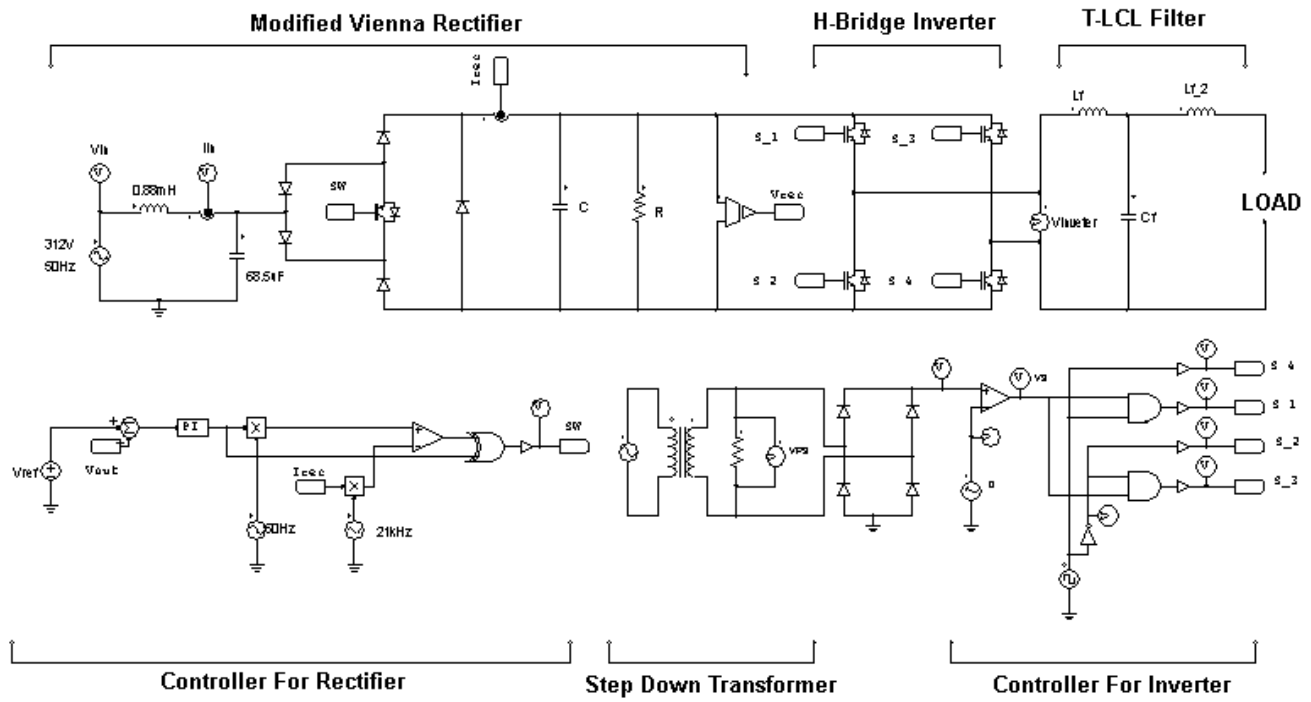


Fig.4. Schematic diagram of proposed single-phase modified Vienna rectifier connected to H-bridge inverter for simulation in PSIM software

The outputs of the multiplier is compared with the output of another multiplier whose inputs are the output current of rectifier and a high frequency triangular wave source by using a comparator. The frequency and the duty cycle of triangular wave are 21KHz and 0.5 respectively. This comparator is used to build SPWM (Sinusoidal Pulse width Modulation). The controller has a XOR Logic Gate whose inputs are the output of comparator and the output of PI controller. The modified Vienna rectifier requires one set of gate signal as it uses one p-MOSFET in single-phase modified Vienna rectifier circuit. The output of XOR is the set of gate signal, SW. The gate switching pulse, SW is shown in Fig.5.

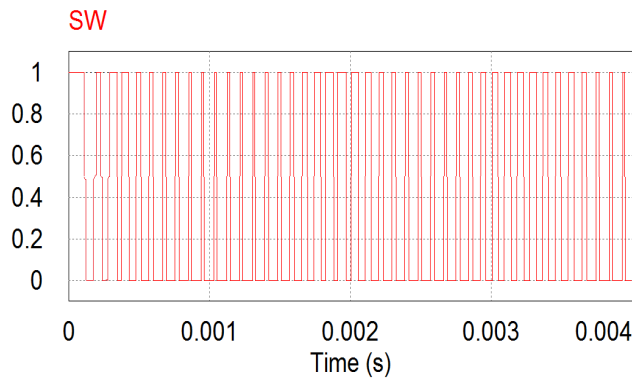


Fig.5. Switching pulse from controller for single-phase modified Vienna rectifier, SW.

B. Single Phase H-bridge Inverter

This section presents an highly efficient H-bridge inverter. The fundamental principle of this inverter [10][16] is converted from a DC voltage to an AC voltage by controlling the Switching components. The inverter has two parallel p-MOSFETs as shown in Fig.6. An abbreviation of analog and digital circuit is used to generate the gating signals of the MOSFETs [11].

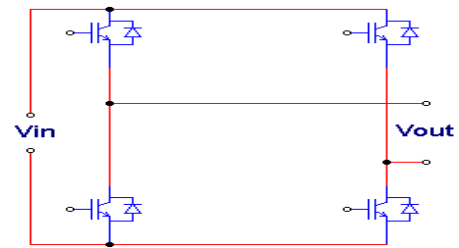


Fig.6. H-bridge Inverter with four MOSFET switches

C. Switching Circuit for Inverter

This section describes the switching controller of the inverter. This inverter circuit consists of two pair of two parallel p-MOSFET gates. Only one type of switching pulse technique is used in conventional inverter but here an abbreviation of SPWM and Square wave [10][16] is used as switching technique to minimize switching losses [16] by minimizing the switching frequency. The sinusoidal voltage source is sampled from power grid by using step down transformer that step down [9] from 220Volt RMS to 5Volt RMS. Then the output of step down transformer after rectifying and triangular wave of high frequency (25KHz) are passed through a comparator to build the SPWM signal. The square wave of 50Hz is passed through a NOT logic gate to generate 180 degree out of phase of the original square wave signal [9][13]. The inverter has four MOSFETs so it requires four gating signals. ANDs operation are performed between square wave signal or 180 degree out of phase of square wave signal and the SPWM to create four gating signal for inverter. The four sets of gating signals of inverter can be classified in two groups. The first group consists of gating signals S_1 & S_4 while the second group consists of gating signals S_2 & S_3. When S_4 is ON by the square wave pulse, SPWM takes place at S_1 and at that time S_2 and S_3 are OFF. Again when S_2 is ON by the 180° out of phase of square wave pulse, SPWM takes place

at S₃, at this time S₁ and S₄ remain OFF. The gate signals of inverter S₁, S₂, S₃ and S₄ are shown in Fig.7.

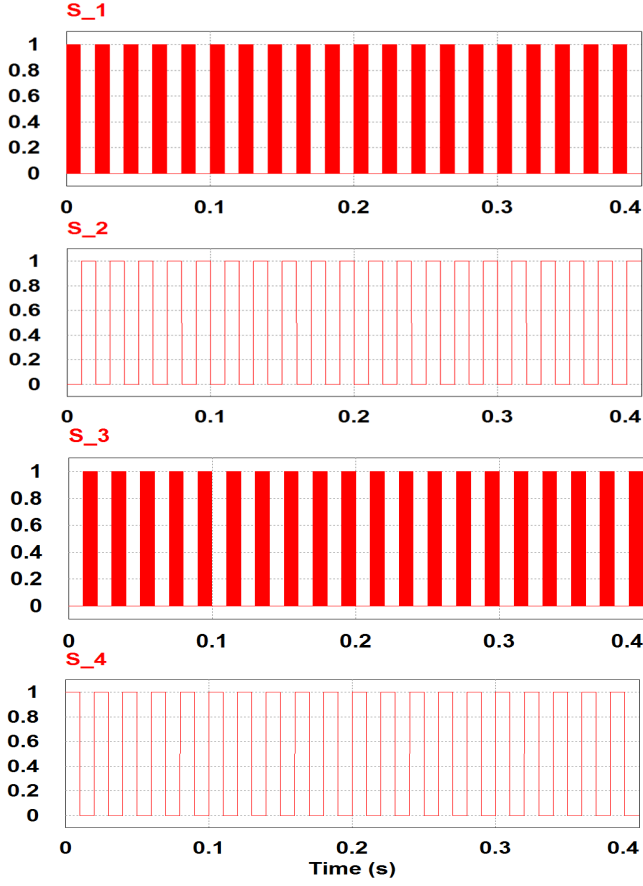


Fig.7. Switching Pulses from controller of H-bridge inverter to p-MOSFETs, S₁, S₂, S₃ and S₄ respectively.

D. Design of Filter Circuit

LC filter is used in conventional inverter as filter circuit but here we used a T-LCL immittance converter circuit as filter circuit as shown Fig.8. The immittance circuit converted impedance to admittance. The load does not depend on admittance. So the T-LCL immittance circuit helps to keep up the desired constant load current and voltage as well as acts as a filter circuit to reduce the harmonics. This filter circuit consists of two inductors (L_f & L_{f2}) and one capacitor (C_f) in T shape. The equations of the output current, I_2 & efficiency are found as [14][15][11].

$$I_2 = \frac{V_1}{Z_0} \left[1 - \frac{1}{Q} \frac{Z_1}{Z_0} \right] \quad (1)$$

$$\eta = \frac{1}{1 + \frac{1}{Q_1} \frac{Z_1}{Z_0} + \frac{1}{Q_2} \frac{Z_1}{Z_0}} \quad (2)$$

Here, the input voltage is V_1 , the load impedance is Z_2 , the quality factor is Q . So the characteristic impedance, Z_0

$$Z_0 = \sqrt{\frac{L}{C}} \quad (3)$$

Quality factor,

$$Q = \frac{\omega L}{r} \quad (4)$$

where ω ($\omega = 2\pi f$) is the angular frequency and r is the inductor's internal resistance. If $Z_0 = Z_2$ & $Q \equiv Q_1 \equiv Q_2$, then the efficiency,

$$\eta = 1 - \frac{2}{Q} \quad (5)$$

where value of r is very low so r is negligible then the Q becomes infinity, under this condition,

$$I_2 = \frac{V_1}{Z_0} \quad (6)$$

From the equation. (6), it is observed that the output current of T-LCL immittance conversion circuit (T-LCL filter) is not dependent of load. So the T-LCL immittance circuit helps to keep up the desired constant load current and voltage as well as acts as a filter circuit to reduce the harmonics. The characteristic impedance,

$$Z_0 = \frac{1}{2\pi f C} \quad (7)$$

Assuming Z_0 as 30 ohm and cutoff frequency, f_c as 50Hz, we get the values of inductor and capacitor by using equation (7) and (3).

$$\text{Capacitor, } C = \frac{1}{2\pi f_c Z_0} = \frac{1}{2 \times \pi \times 50 \times 30} \approx 0.105 \text{ mF}$$

$$L_f = L_{f2} = L = C Z_0^2 = 0.105 \times 10^{-3} \times 30^2 \approx 95.6 \text{ mH}$$

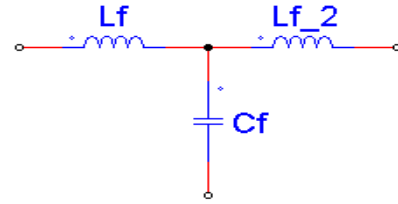


Fig.8. A T-LCL immittance conversion circuit (Filter Circuit).

IV. SIMULATION RESULTS AND DISCUSSION

Simulation have been carried out in the PSIM simulation software to verify the proposed system with different single phase PFC topology. The single phase input voltage of single-phase modified Vienna rectifier is 312V (peak) as depicted in Fig.9. Our proposed modified Vienna rectifier used 21KHz as switching frequency. The output of proposed single-phase rectifier is 345V (constant) DC which is illustrated in Fig.10. In this research, we proposed single phase grid side power factor correction under RL load. The single-phase input voltage and current and power factor (0.992) are shown in Fig.11 for RL load. A highly efficient inverter (H-bridge) is used for using load.

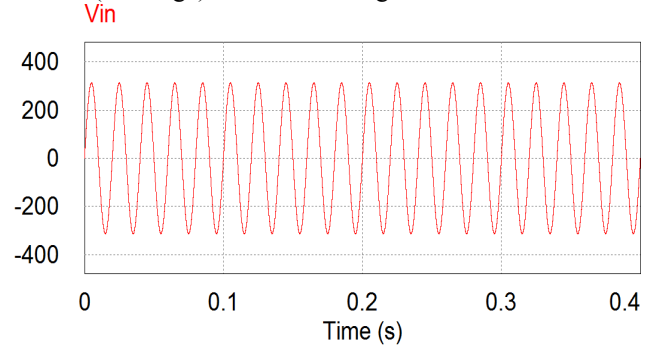


Fig.9. Single phase input voltage (220RMS) of proposed single-phase modified Vienna rectifier

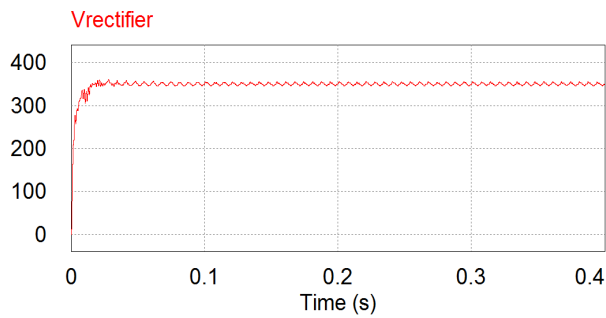


Fig.10. DC output voltage (constant 345V) of proposed single-phase modified Vienna rectifier

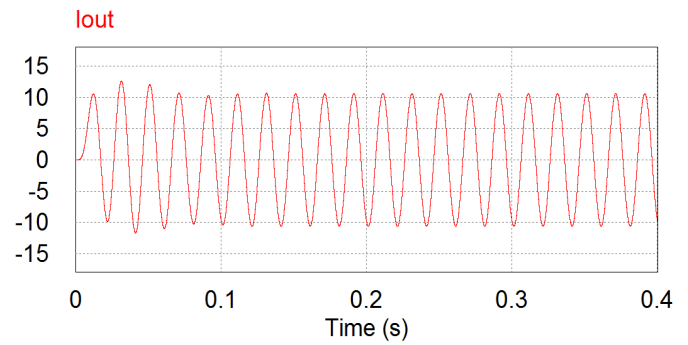


Fig.14. Output current waveform of after filtering

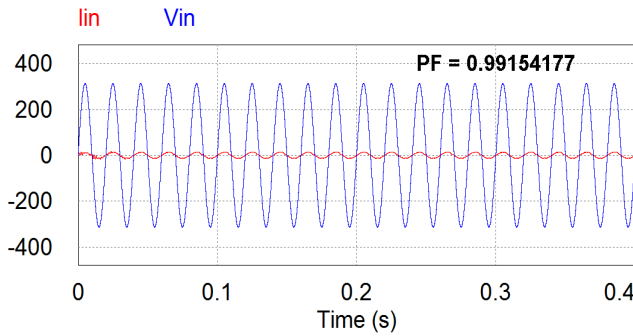


Fig.11. Input voltage and current waveform and almost unity power factor using RL load

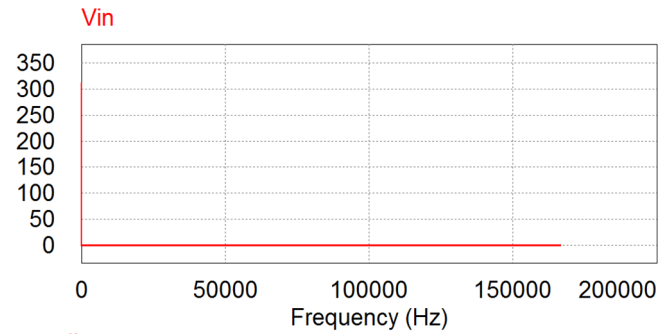


Fig.15. FFT analysis of input voltage and current in PSIM

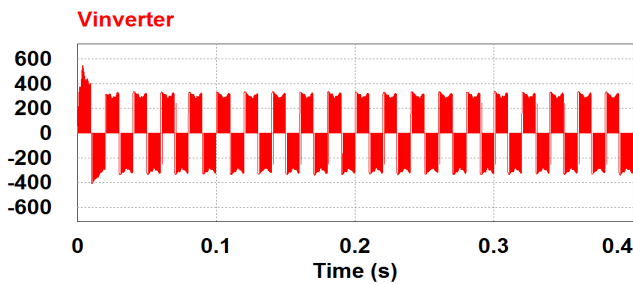


Fig.12. Output waveform of inverter without filtering

Fig.12 shows the output of the inverter which is distorted and contains lot of ripples. Here we use a T-LCL immittance filter circuit to minimize the distortions thereby achieving pure sinusoidal output voltage and current. After filtering almost ripple free voltage and current waveform of the inverter are achieved as shown Fig.13 and Fig.14 accordingly. Fig.15 represents the input voltage and current's FFT analysis. The FFT analysis confirms that the input voltage and current has only fundamental component with a frequency of 50Hz.

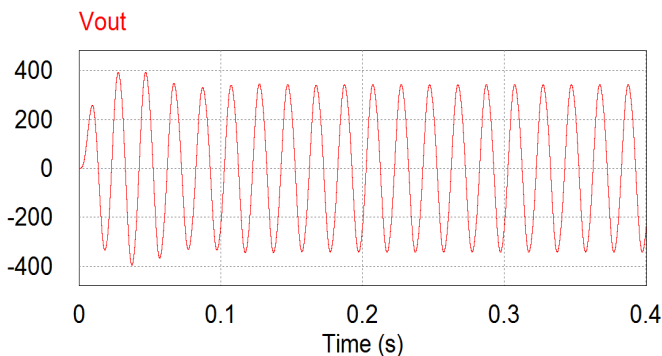


Fig.13. Output voltage waveform of inverter after filtering

V. PERFORMANCE ANALYSIS

The proposed system is very effective for the single phase generation side power factor improvement. Almost unity power factor ($0.99154177 \approx 0.992$) is achieved (when RL load is connected) and the designed rectifier is tested that the single phase input current THD is found to be about 5.6%. The efficiency of the proposed system is found to be about 96.8%. The controller of single phase modified Vienna rectifier is very simple. The output current and voltage of the H-bridge inverter have been found pure sinusoidal waveform after T-LCL filtering. Before this there was huge harmonics. Total harmonics distortions of output current and voltage of inverter are found to be less than 0.01%.

Table-II shows the performance comparison between Modified Vienna, conventional Vienna and Buck-Boost converter under RL load.

It is observed that, power factor and efficiency of this proposed single-phase modified Vienna rectifier are much better than any conventional single phase PFC topology. Table-IV The performance comparison between modified Vienna, conventional Vienna and buck-boost converter according to power factor, distortions and efficiency under RL load.

TABLE-II COMPARISON OF PERFORMANCE AMONG DIFFERENT POWER FACTOR CORRECTION TECHNIQUES

Parameters	Modified Vienna	Vienna	Buck-Boost
Power Factor	0.992	0.94	0.97
Distortion	5.6%	8.9%	8.3%
Efficiency	96.8%	93%	95.2%

VI. CONCLUSION

In this research, a single phase modified Vienna rectifier and single phase two stage inverter (H-bridge) has been proposed. The proposed topology has been designed for acquiring corrected power factor, lower Total Harmonics Distortions (THD), low cost, higher efficiency and simple control system. By analyzing the simulation result in PSIM simulation software of proposed single-phase PFC topology, we got nearly unity power factor. Here, it is tested that the THD of single phase input current is found to be about 5.6% and the efficiency of proposed single-phase PFC system is about 96.8%. The simulation result of inverter views that the Inverter (H-bridge) generates a very high quality ripple free voltage and current with less than 0.01% THD that is much lower than IEEE 519 standard. It is observed that the proposed PFC topology has higher efficiency (96.8%) lower cost, lower THD, simple control, almost unity power factor (0.992) compared with any conventional single phase PFC scheme.

VII. FUTURE WORK

Based on this simulation, hardware prototype of this proposed system will be implemented in near future.

REFERENCES

- [1] Bor-Ren Lin, "High Power Factor AC/DC/AC Converter with Random PWM". IEEE Transactions on Aerospace and Electronic Systems, Vol. 35, no. 3, pp.935-944 July1999.
- [2] Chaoyong Guo, Fei Lin, Xiaojie You and Trillion Q. Zheng "Modeling and power quality enhancement to AC drives based on synchronous PI current control technology of PWM rectifier"Industrial Electronics and Applications, 2008. ICIEA 2008. 3rd IEEE Conference on, Pages: 438 - 442.
- [3] Mir Humainul Islam and M. A. Razzak, "Design of a Modified Vienna Rectifier for Power Factor Correction under Different Three Phase Loads", The 5th International Conference on Informatics, Electronics and Vision (ICIEV-2016), 13-14 May, 2016, Dhaka, Bangladesh.
- [4] Laslo Huber, Yungteck Jang, Milan M. Jovanovic, "Performane Evaluation of Bridgeless PFC Boost Rectifiers" IEEE Transactions on power electronics, vol.23,No.3, may 2008.
- [5] Bindu KV, B. Justus Rabi, "A novel Power Factor Correction Rectifier for Enhancing Power Quality ", International journal of power electronics and drive system (IJPEDS), Vol.6, No.4, December 2015, PP. 772-778.
- [6] Chongming Qiao, K. M. Smedley "Three-phase unity-power-factor VIENNA rectifier with unified constant-frequency integration control" Power Electronics Congress, 2000. CIEP 2000. VII IEEE International, Pages: 125 – 130
- [7] Kolar, J.W, Drofenik.; Zach, F.C. "Current handlingcapability of the neutral point of a three-phase/switch/level boost-type PWM(VIENNA)rectifier". PESC 96 Record. p.1329- 36vol.2.
- [8] 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 anovel integratedpower semiconductor module".APEC'96,SanJose, CA. IEEE, 1996, p.514-23,vol.2.
- [9] Srabanty Ahmed Shaon, K. M. A.Salam, "Study of Vienna Rectifier and a HighlyEfficient Single Phase Two Stage Inverter with lowTHD", 8th International Conference on Electrical and Computer Engineering 20-22 December, 2014, Dhaka,Bangladesh
- [10] Abdalrahman, A. Zekry, and A. Alshazly, "Simulation and implementation of grid-connected inverters", International Journal of Computer Applications, vol. 60, no. 4, December2012
- [11] Sajib Chakraborty, Md. Abdur Razzak "Design of a Transformer-less Grid-Tie Inverter Using Dual-Stage Buck and Boost Converters" International Journal of Renewable Energy Research-IJRER, Chakraborty et al., Vol.4, No.1, 2014
- [12] M. H. Rashid, Power Electronics, New Delhi, India: Prentice-Hall of India Private Limited, 2007, pp. 147- 265.
- [13] N. Mohan, T.M. Undeland and W. P. Robbins, Power Electronics, Kundli, India: Replica Press Pvt. Ltd., 2003 pp.-204-211.
- [14] S. B. Afzal, M. M. Shabab and M. A. Razzak, "A combined π - and T-type immittance converter for constant current applications", Proc. 2nd IEEE International Conference on Informatics, Electronics & Vision (ICIEV), May 17-18, 2013, DhakaBangladesh.
- [15] Irie H. et al. Utility interactive inverter using immittance converter for photovoltaic power system. Trans IEE Japan, Vol.120-D,pp.410-416.
- [16] 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 May2013.

Analysis of Power Demand Response for Smart Home in Germany from End-User's Perspective Based on High Resolution Power Demand Model

Nayeem Hasan Mallick
Department of EEE
World University of Bangladesh
Dhaka, Bangladesh
nayeem.mallick@alumni.uni-ulm.de

Md Hasan Maruf
Department of EEE
Green University of Bangladesh
Dhaka, Bangladesh
maruf.eee@green.edu.bd

Abstract—This study forecasts the load profile of Germany for household scenarios till 2050 based on Markov chain process. Taking into account the penetration level of photovoltaic (PV) solar technology, technological improvement of efficient home appliances, battery support, load management and demand response scheduling, a high resolution smart home demand profile is created. At the end, combining of all, a highly annual aggregated demand profile is produced which shows daily variations on demand load profile for fixed tariff scenarios up to 44% with respect to the present demand profile. On the other hand it is found that, Time of Use (TOU) tariff shows a variation of 64% with respect to the present demand profile.

Keywords—smart home; demand response; load management; activity pattern; markov process, high resolution model

I. INTRODUCTION

Day by day, the demand of renewable energy creates revolutionary growth over the world. With the development of technology, the demand of power has been increased to meet the requirement of modern activities. In these purposes, developed country like Germany is running first by using renewable energy sources as well as photovoltaic (PV). On May 8, 2016 Germany produced about 87% renewable energy out of total consumption [1]. On that day they had produced so much energy from PV and wind that they had to pay to the customers to use electricity. Normally in summer surplus amount of PV is produced in a household. As the excess power can't be wasted so the way is to store the electricity or to sell it to the grid. And in winter the situation is absolutely opposite. So the way is to meet the demand is to provide power from grid which cost lots of money.

On the other hand the utilities are trying to promote demand response programs that offer customers financial incentive to shift load from peak hour to off peak hour. Combining all these factors from the end users aspect it is not that much clear how efficient would be the total scenario now and later. To provide details, a high resolution power demand model was initially developed [2]. Taking that as an input, in this paper an aggregated demand profile is produced with the projections of till 2050. In this regard, the simulation is done with the

different levels of photovoltaic integration, storage analysis, technological improvement of home appliances, demand load management and two level tariffs. The model presented here is the improvement of demand side analysis proposed in [3-9].

A detailed model description is presented in the Section II. In section III how power demand setup is done using demand management model is presented. Simulation results are discussed in the next section IV. Finally the conclusion is drawn in section V.

II. SMART HOME MODEL

A. High Resolution Model

The work flow of the model is presented in Fig. 1. The bottom-up modeling technique [3, 4] is used here in this work. This technique first generates synthetic activity patterns for each household member. Later this activity pattern is converted into power demand model for the household. To approach these, effects of technological improvements, smart home requirements and automation algorithms are accounted from end users perspective. On the left side of the flow chart representing the demand load profile is established from a report on energy use of households in Germany [10] and another annual report on the demand load profile [11] by using top down technique. In this part the demand load model is calibrated to produce a German household scenario which is taken as reference scenario. Then the model generator construct the smart home's load profile which is outlined on the right side of the chart. In this part different considerations and trends has been taken to evaluate the future scenarios. Combining all the results has been analyzed.

B. Generation of Synthetic Activity Patterns

Modeling individual's behavior is a complicated task due to the stochastic nature of the activities that performed in a day. This model uses non-homogeneous Markov chain to model occupant daily activity and predict the associated energy consumption. In order to estimate the transition probability matrices used in the behavioral Markov chain model, detailed time usage data of different person is desired. In this case as Germany's household load profile is needed, a German time

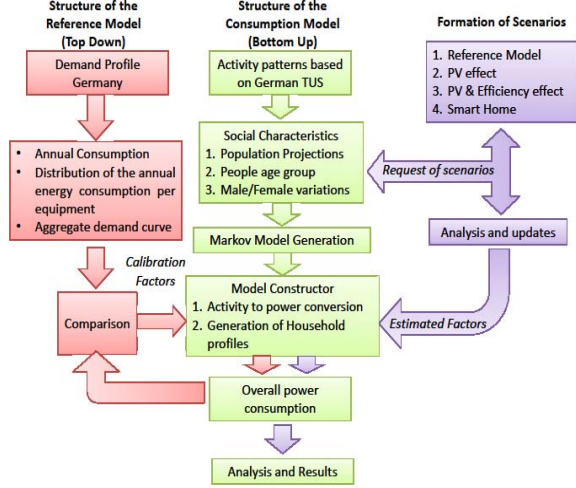


Fig. 1. Flow chart of the methodology for the high-resolution power demand model.

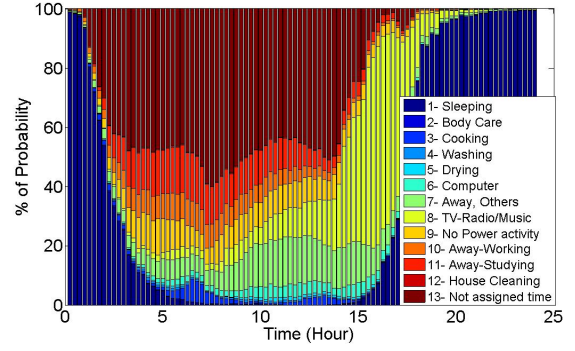
use survey data is used to resolve the vectors of probabilities for the transition probability matrices for the Markov chain as:

$$X(k)=(x_1(k), \dots, x_n(k))^T \quad (1)$$

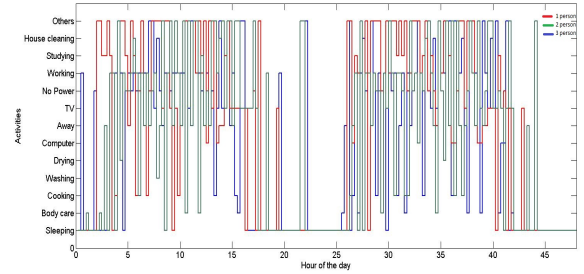
Where $x_i(k)$ is the probability that activity number i^{th} will be performed in time interval number k with $i=1, \dots, 13$. An example of this individual's behavior probability for the different activities of a German household is shown on fig. 2a. It has been marked with 13 different colors for 13 different activities. The probability of each activity group is shown as length relative to the ordinate for each hour of the day. Fig. 2b. represents the activities performed by a 3 person household for 2 days generated by the Markov chain model constructor. Any of the 13 activities took place in these 24 hours by 3 persons. Sometimes same activity like "sleeping" appeared for 2 persons at the same time. Another example can be illustrated here that, the activity named "washing" activated only one time by only one person in this total 48 hours time span.

C. Error Analyzing

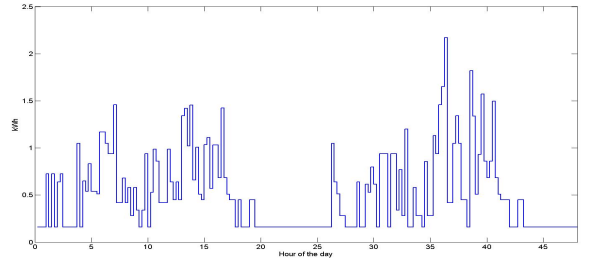
For validation purpose normalized root mean square error (NRMSE) has been included into the model. Several types of interactions between different number of persons and days have been performed and error has been checked. Some error curves have been shown in Fig. 3 with the activity pattern. First one has been done with 100 persons and 10 days with total 1000 interactions and is clearly observable that the error is so high at the beginning when the number of interactions is less and going down in later stages nearly to zero but still higher enough. From the activity pattern diagram it is also seen that the curves showing some scatter values. From here it can be extractable that the simulated data fit with the actual data with an R^2 of 0.4201 when the minimum interaction is 400. This interaction can be equals for 100 persons and 4 days



a.



b.



c.

Fig. 2. Activity to power-conversion. a. Example of different attractors for the Markov chain model. b. Resultant activity patterns for 3 persons in a household for two consecutive days. c. Generated power demand model for the corresponding activity pattern

which can be said as minimal sample size of the simulations to get more accurate data. In the next fig. 3b, 500 persons for 500 days simulation result is plotted. The activity diagram is now more and more smooth and the error is closer to zero. It can also be said that the NRMSE 0.01651 is the median value; most of the errors are not more than that value.

D. Activity-to-power Conversion

The activities found in the Markov Chain Model correspond to different electricity end uses. The 13 activities converted into 12 different end usages. For one activity it may corresponds to different appliance usage. For example, Body care is an activity which uses different appliances like Toothbrush, Hairdryer & Electric razor etc. To model this in briefly would require a detailed data of use or TUS data for each activity with each individual appliance which is not available and is also very complicated. To resolve this

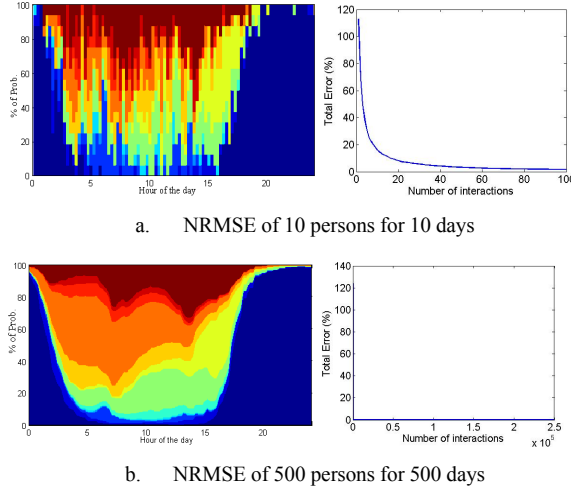


Fig. 3. Error analyzing of activity pattern

problem, Widén [8] found a simple but more accurate technique to assume a constant power demand during the period of activity. Each appliance model is associated with one of the three activity-to-power conversion arrangement such as: sharing scheme, constant demand during the period of activity and a variable duty-cycle demand that begins with the activity. When an appliance is shared by many, the power demand is determined from the combined pattern. The technique is much alike as in [5, 6, 12, and 13] but load patterns are combined with the representation of standby power as in [8].

Activity to power demand conversion curve is shown in fig. 3. Same activity pattern is converted into power demand curve. When there is an activity started shown on fig. 3a. that corresponds any power activity is shown on the fig. 3b. But when the activity doesn't represent any power consuming activity then the demand curve also doesn't shows any power activity. Though a standby power is assigned throughout the activity period.

Fig. 3c shows the amount of energy required for 2 days for each activity as the activity pattern presented in fig. 2c for the same 3 persons household. As an example it can be seen that in the time between 20 to 24 hours and between 44 to 48 hours of the day only the activities 'sleeping' and 'others' took place by the 3 persons which doesn't consume any power. So power demand curve also reflects no power activity with only standby constant power in the same time frame.

III. POWER DEMAND MODEL

In this section, the previously built load profile of household coordinated with photovoltaic and battery support to comply with the household energy management (HEM) system to evaluate the electrical impacts for three demand response scenarios, as follows.

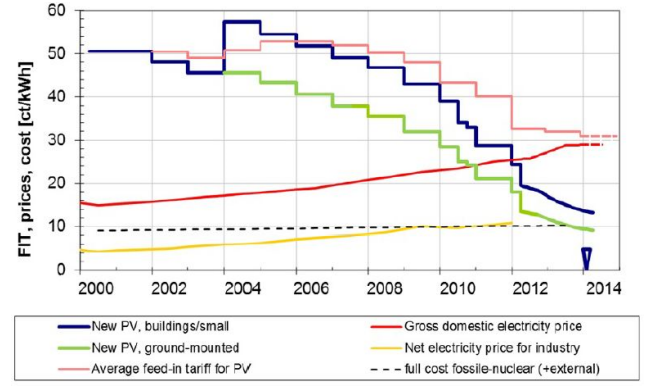


Fig. 4. Comparison of different tariff, prices, cost (ct/kWh) in recent years

A. Feed in Tariff

The feed in tariff (FIT) define as the price to sell the energy to the grid [9]. In Germany, The German Renewable Energy Act (EEG) fixed the level of the FIT for all the energy sources and in case of solar power it's getting the priority. This tariff is valid for the time period of 20 years from the date of installation of the plants. Just 7/8 years before the tariff was much higher. But the situation has been changed. From the beginning of year 2012, the FIT of household gets grid parity which is a notable stage for energy policy in Germany. From June 2014 the tariff is fixed in between 9.01 and 13.01 €-cents/kWh for different size of plants which is valid for next 20 years who will install plant now [14]. As the present domestic electricity price is almost 0.295 €/kWh [15], so FIT is much lower comparing that. From this statistics it can be said that self-consumption is beneficial than selling energy to the grid. Fig. 4 showing the comparison of net prices of different PV setup, gross domestic electricity price and the average feed in tariff for PV [11]. Average FIT for PV represents the tariff which activated years before and running for the period of 20 years. The average feed in tariff is now going down significantly though still higher than domestic electricity price. But depending on the current net FIT the effect will be observed after 10/12 years later.

B. Time of Use Rate

Instead of using a single flat rate for energy, time of use (TOU) rate is introduced in various countries in several states [16]. The reason behind the use of TOU tariff is to shave the demand of energy use during peak hour. Customers can also avail the benefit from this. Normally during off peak time the electricity price is set to much lower than that of normal price. Using electricity on that time customer can reduce their electricity bill significantly. Charging battery on off peak hour is also a great advantage for the customer which has been checked in later stages. Here in these model two level prices is introduced as normal Fixed rate 29 cents/kWh from 6am to 23pm and TOU rate 20 cents/kWh from 23pm to 6am. The current flat rate of electricity is used only for simulating the present years but for future electricity price is used the forecasted data collected from IBESA [4].

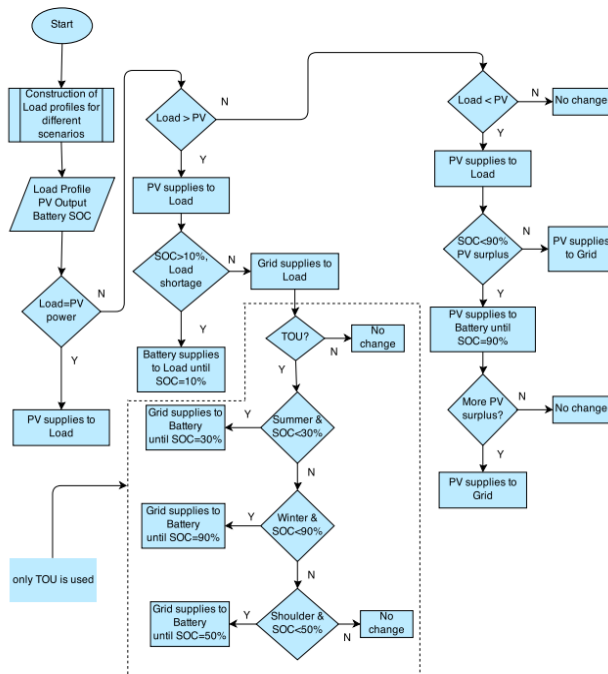


Fig. 5. Flow chart of the proposed demand load management algorithm

C. Demand Management Model

The model architecture is presented in fig. 5. The load profile built in [2] is used as a data input in this model. Power generation from different rated photovoltaic also included with real time and forecasted data over the whole year with every 15 minutes time steps. Energy storage of different capacity is also included in the model with State of Charge (SOC) as an input which initial value is fixed to 80%. The model can produce load management scenarios of three different seasons like before [2] which is named as summer, shoulder & winter. After taking the inputs load profile, PV output & battery SOC, at first the model constructor compares the load with PV output. If PV power is sufficient to support the load then nothing happen. When PV cannot support the load, then at first it will take power from battery, if it's not enough then grid will supply the rest. In case when PV output is higher than the demand load, then surplus energy will store in battery and after full filling both of them the rest will go to grid. Battery SOC has given the limit in between 10% to 90% for the longer life of battery. PV is the first option of battery charging, and only period of TOU battery will charge from grid to support the high price energy period later. Depending on the seasonal variations charging of battery from grid has some bar. In summer season photovoltaic can support a longer period. So battery charging bar is set to 30% from the grid on summer and for winter there is no limit. For shoulder season the limit is set to 50%. In case of non TOU tariff time there is no option of charging battery from grid as there is no need to do so.

IV. SIMULATION RESULTS

The model constructor can re-produce the energy projections of different household scenarios till 2050 which

also reflect the change in activity pattern of German's people based on their age & sex, and technological improvement as well as variation on seasons. In next section the results has been discussed in perspective of different scenarios.

A. Smart Home with Demand Load Management

The model constructor can re-produce the activity patterns of individual household members with power demand associated with each energy consuming home appliance. So it can also generate the energy consumption curves throughout the year of single household consisting of multiple individuals. In this part, the model has been constructed with a demand load management algorithm with technological improvement taken in mind as follows:

- 1) *Integration of photovoltaic:* Germany's target is to reach the goal of 80% renewable energy of their total energy production by 2050, and Fraunhofer ISE forecast that the share of PV will be 30% in Germany by 2050 [5]. Different study shows that it is possible to generate the whole power renewably by the date. Keeping all that in mind the high penetration of PV in household is taken into the model using calibration factor to forecast the next years.

- 2) *Seasonal variations:* Fig. 6 represents the load management of different scenarios. Subfigure (a) and (b) represent the load management curve of shoulder and winter season respectively with the capacity of 5kWh battery and using TOU tariff. In shoulder season it can be observe that the generation of PV is moderate in a sunny day whereas in winter the portion from PV is much less. So surplus of PV power supplies to grid in shoulder season occur for a longer time and also it supplies power to battery. But in case of winter season there is no sign of supply to grid but have little portion supply to battery in day time when the demand load is not that much.

- 3) *Battery analysis*: In winter season the battery effect is more significant than any other seasons. In this time solar power generation is very lower or sometime close to zero. In this time battery have a big role on daily power demand curve. To discuss about the battery effect into the model the fig. 6b. can be seen. Battery is taking charge little bit in day time when there is solar power generation. And supporting it in evening time when the electricity cost is flat rate. Again battery taking charges from the grid in the period of TOU rate, and supporting the load when the TOU period is over from the morning. Thus battery saving money by utilizing the TOU tariff period and PV power. As discussed previously that now FIT is lower, so self-consumption is the best way of utilizing the PV power where only battery can support to use the PV power in later time by not selling the power to grid.

- 4) *Effect of TOU rate*: Fig. 6c represents the load curve of a winter day with fixed tariff. Here no sign of battery charging from grid are seen. Most of the time grid directly supplies to the load. As it is a winter day load curve, so PV generation is also lower. Comparing this fixed tariff with the TOU tariff curve of subfigure (b) will show the changes made by the TOU rate boldly. As discussed before that in TOU period

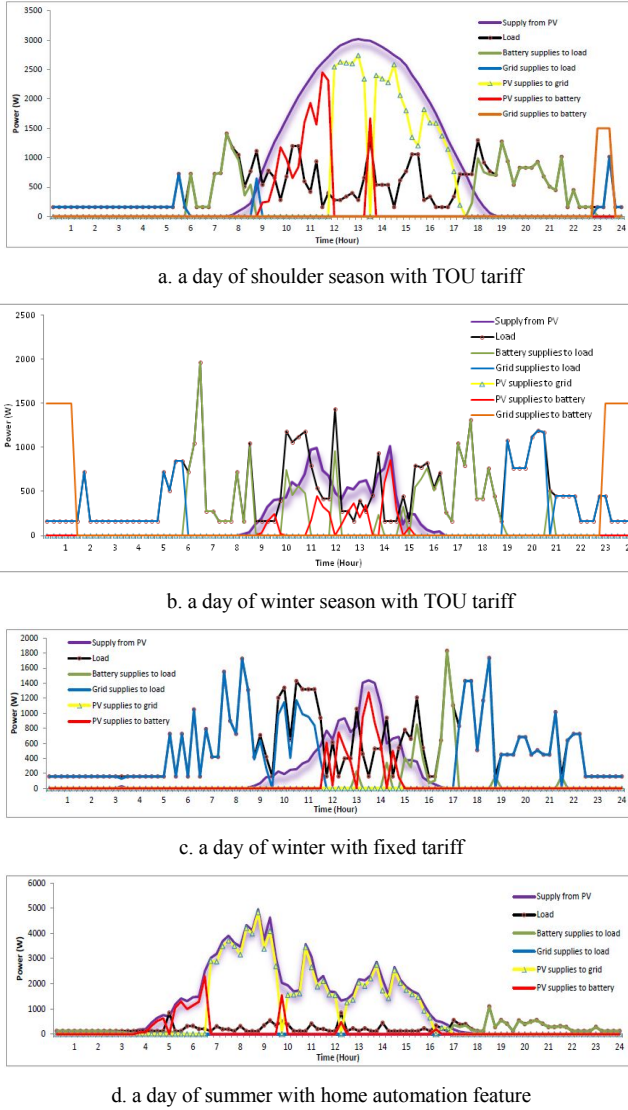


Fig. 6. Comparison of different types of load management scenarios

battery taking charges and supporting the load during morning time when the TOU period is over. Thus saving money a lot for the customers. In combination with TOU tariff and battery could be a great money saver for customers.

5) *Automation analysis*: Fig. 6d represents the results from doing some sort of automations that includes load shifting and using technology. This power demand curve is for a summer day with battery included. Automation does a power cut in day time and also shifts the load when demand is high which leads battery to support more in the evening. As it is fixed tariff curve, so no charging from grid is enabled for battery.

B. Aggregated Households Demand

Previously built all the individual household is merged together to get the aggregated average demand load profile

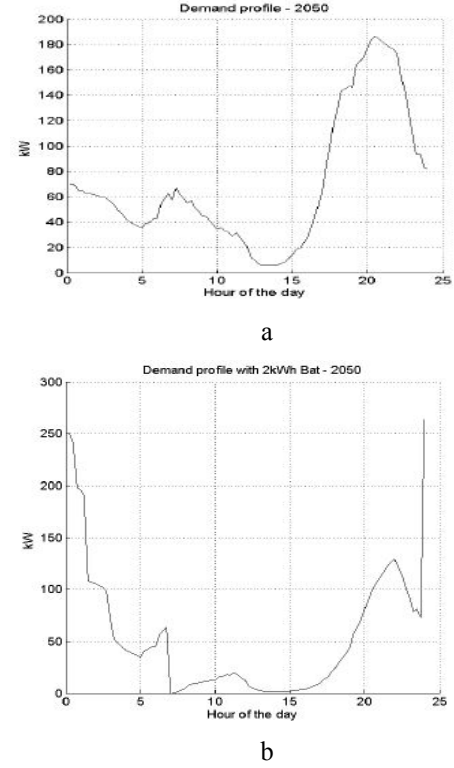


Fig. 7. Demand profile of the household in 2050. a. reference scenario with fixed tariff. b. combination of different scenario with TOU tariff

from the grid. Fig. 7 is presented with the demand profile of the household in the year of 2050 with two different combinations. First (a) shows the reference scenario with fixed tariff, (b) shows the demand profile with TOU tariff integrating 2kWh battery capacity. In reference scenario there is no automations or storage back up, so demand from the grid is much higher compare to the other one. Using TOU tariff figure (b) shows a sharp change in the beginning of TOU period when the battery begins its charging. But because of using battery in the early morning demand from grid decreases significantly which allow to save lot money for customer.

In the fig. 8 projected variations of demand profile is shown taking into account the demand profile of reference scenario of 2050. The model reference scenario is the projected load

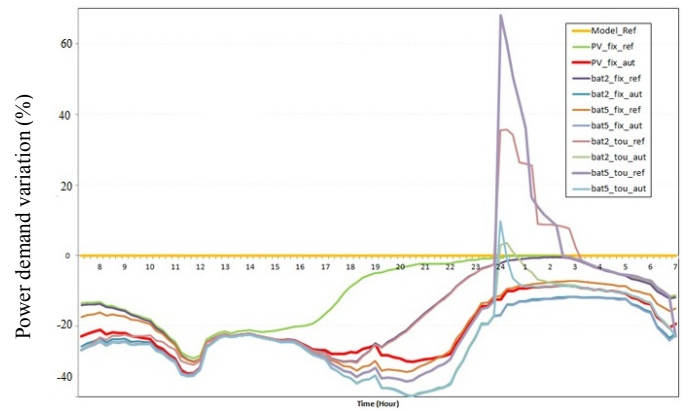


Fig. 8. Projected variation in demand profile for different scenarios with respect to the demand profile of reference scenario for 2050

profile of 2050 with fixed tariff system. There is no other effect counted for this one. From the reference scenario other 10 different scenarios has been deducted and shown as percentage of power demand variations. In this case two different type battery capacities of 2 kWh and 5 kWh are taken. From the fig. it can be understand that the curve which have positive response are the scenarios related with TOU tariff. In the beginning of TOU period in night time the demand from the grid increases lot to charge the battery as the electricity price is less on that time. After taking charge battery support the demand of household in the early morning when the TOU period is over. On that time demand from the grid also decreases.

For fixed tariff scenarios the variation on the load profile is up to 44% with respect to the current demand profile. That's mean up to 44% power can be less consumed from the grid at a certain time which will shave the peak load. And for TOU tariff that goes up to 64% of positive side of the curve with respect to the current demand profile which represents the demand from the grid on TOU period. But on other time it shave the peak load as well. For demand side management TOU could be a good option to encourage the customer to shift their peak hour load which will also ease the pressure from utility side. In this model two level price is used which could be more level.

V. CONCLUSION

An analysis of power demand response of smart home in Germany was presented based on high resolution power demand model from end-user's perspective. Interactions between smart home load profile and supply from grid was simulated and analyzed. In this regard different levels of PV penetration, two types of tariff (fixed & TOU), different types of storage capacity were undertaken to simulate highly realistic demand response profile. All these are based on the time of use survey (TUS) of Germany's people using Markov chain model process. Finally the study reveals that the power demand response of a smart home can create high variation in the daily demand profile up to 44% with respect to the present demand profile with fixed tariff and goes up to 64% for TOU tariff.

REFERENCES

[1] IFL Science, [Online] Available: <http://www.iflscience.com/environment/germany-just-produced-so-much-renewable-energy-it-had-pay-people-use-it> (Accessed- 20 May, 2016)

[2] E. B. Alzate, N. H. Mallick and J. Xie, "A high-resolution smart home power demand model and future impact on load profile in Germany," *Power and Energy (PECon), 2014 IEEE International Conference on*, Kuching, 2014, pp. 53-58. doi:10.1109/PECON.2014.7062413 URL: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7062413&isnumber=7062393>

[3] A. Chrysopoulos, C. Diou, A.L. Symeonidis, P.A. Mitkas, Bottom-up modeling of small-scale energy consumers for effective Demand Response Applications, *Engineering Applications of Artificial Intelligence*, Volume 35, October 2014, Pages 299-315, ISSN 0952-1976, <http://dx.doi.org/10.1016/j.engappai.2014.06.015>.

[4] Ammon, Martin. Electricity Storage PV: Current Status Market Overview Germany. June 2013

[5] Ian Richardson, Murray Thomson, David Infield, A high-resolution domestic building occupancy model for energy demand simulations, *Energy and Buildings*, Volume 40, Issue 8, 2008, Pages 1560-1566, ISSN 0378-7788, <http://dx.doi.org/10.1016/j.enbuild.2008.02.006>.

[6] Ji Hoon Yoon, Ross Bladick, Atila Novoselac, Demand response for residential buildings based on dynamic price of electricity, *Energy and Buildings*, Volume 80, September 2014, Pages 531-541, ISSN 0378-7788, <http://dx.doi.org/10.1016/j.enbuild.2014.05.002>.

[7] Matteo Muratori, Matthew C. Roberts, Ramteen Sioshansi, Vincenzo Marano, Giorgio Rizzoni, A highly resolved modeling technique to simulate residential power demand, *Applied Energy*, Volume 107, July 2013, Pages 465-473, ISSN 0306-2619, <http://dx.doi.org/10.1016/j.apenergy.2013.02.057>.

[8] Joakim Widén, Ewa Wäckelgård, A high-resolution stochastic model of domestic activity patterns and electricity demand, *Applied Energy*, Volume 87, Issue 6, June 2010, Pages 1880-1892, ISSN 0306-2619, <http://dx.doi.org/10.1016/j.apenergy.2009.11.006>.

[9] Cajsa Bartusch, Fredrik Wallin, Monica Odlaere, Iana Vassileva, Lars Wester, Introducing a demand-based electricity distribution tariff in the residential sector: Demand response and customer perception, *Energy Policy*, Volume 39, Issue 9, September 2011, Pages 5008-5025, ISSN 0301-4215, <http://dx.doi.org/10.1016/j.enpol.2011.06.013>.

[10] Federal Statistical Office of Germany, "German Statistical Yearbook 2012", Chapter 2. March 2013

[11] ISE, Fraunhofer institute for solar energy. "Recent facts about photovoltaics in Germany" 2014

[12] V. S. K. Murthy Balijepalli, V. Pradhan, S. A. Khaparde and R. M. Shereef, "Review of demand response under smart grid paradigm," *Innovative Smart Grid Technologies - India (ISGT India), 2011 IEEE PES*, Kollam, Kerala, 2011, pp. 236-243.

[13] P. Du and N. Lu, "Appliance Commitment for Household Load Scheduling," in *IEEE Transactions on Smart Grid*, vol. 2, no. 2, pp. 411-419, June 2011.

[14] ISE, Fraunhofer institut for solar energy. "Levelized cost of electricity renewable energy technologies". November 2013

[15] Eurostat, Electricity prices for domestic consumers- URL <http://appsso.eurostat.ec.europa.eu/nui/submitViewTableAction.do>

[16] Muratori M., Schuelke-Leech B.-A., Rizzoni G. " Role of residential demand response in modern electricity markets". From: *Renewable and Sustainable Energy Reviews* 33 (2014) 546-553.

Transforming a Multi-value Database System into a Relational Database System for Faster Querying

Md. Zahid Hasan¹

Department of Computer Science and Engineering
East West University
Dhaka, Bangladesh
mail2zahid_kuet@yahoo.com¹

Mohammad Rezwanaul Huq²

Department of Computer Science and Engineering
East West University
Dhaka, Bangladesh
mrhuq@ewubd.edu²

Abstract: When the data is fluid and relationships among attributes are getting complex and are changing rapidly over time, a Multi-value Database Management System (MVDBMS) can provide a great deal of support to handle that. MVDBMSs can handle attributes taking a list of values unlike a Relational Database Management System (RDBMS) that can only accept single-valued attributes to conform first normal form (1NF). It is suitable to use multi-valued database in case where a list of values can be assigned to a single attribute in a particular record. In this case, MVDBMS can save a lot of storage space and it is quite efficient to retrieve a specific record. However, when it comes into retrieving data based on Select-Project-Join (SPJ) queries, a multi-valued database cannot handle these queries in a straightforward manner. In this paper, we propose techniques to transform a MVDBMS into a RDBMS for faster querying and retrieval. During the transformation, we carefully considered the trade-off in between execution time of queries and storage space consumed by data. Therefore, we introduce two novel techniques: one of them will generate database schema conforming to 3NF and the other will be in 1NF. We also evaluate the performance of these two techniques for different types of queries.

Keywords: Multi-value database management systems (MVDBMS); Multi-value (MV); Sub-value (SV); Associated value (AV); Multi dimensional database; NoSQL.

I. INTRODUCTION

There are some systems which use multi-value database as their data storage; a type of NoSQL and multidimensional database. In this system a single field can store more than one value in a single record. These values are called multi-value of that field. It can be extend up-to sub-value for a specific field. Storing and updating a particular record in this method works fine in applications where the relationships are complex and can change rapidly and drastically over time. But when we consider the case of querying information by joining table or extracting information of sub-values/multi-values within same table; performance will degrade as number of records increase.

In multi-value database management systems there is no relationship between data sets or among multi-value/sub-value fields or between one table with other tables. Database native functionality like indexing, constraints can't be imposed in multi-value database systems. Each column data are independent with each other within a record as well as tables are also independent. In this scenario it is difficult to extract data.

Some business cases where multi-value database suits more than RDBMS for storing, retrieving (only a specific record) or updating record. If we efficiently query/extract data from more than one table of multi-value database systems then its searching issue would be resolved and it can be more usable. We are

proposing a technique to convert multi-value database systems to RDBMS, so that we can define relationship between tables, convert multi-value/sub-value into a number of single values; so that we can search information by join table or retrieve information of multi-value, sub-value in efficient way. In our paper we use MVDBMS and NoSQL as interchangeable.

II. BACKGROUND

Our area of interest include multi value database. In [1] [5] it states for today's interactive web and mobile applications the importance of flexibility and scalability in data model can't be over-stated and NoSQL databases are currently being used by Google, Amazon, Facebook and many other major organizations operating in the era of Web 2.0. In addition to SQL databases, cloud applications usually store their data in NoSQL databases as in [2]. However, in [2] discussed most NoSQL databases do not support join operations. It [4] shows every day people and companies generate enormous amounts of data and this data may be structured, unstructured, semi-structured or a combination of all. This has called [5] for the need to design NoSQL databases which can store this type and volume of data.

However, in [4] shows the NoSQL solution to this Big Data problem has also lead to several other problems. The challenges of NoSQL databases specially in [2] [4] [6] the field of 'Business Intelligence and Analytics' can't be overlooked. It shows [6] how different NoSQL databases uses third party API for ad-HOC query generation ; even some don't have any option to do that.

III. WHAT IS MVDBMS ?

In case of relational database, it is not allowed to keep more than one instance of value in any column, so we have to keep each instance separately against same record id. It will make data redundant which can be solved using multi-valued database. The biggest benefit is that it means you don't need to maintain more tables like RDBMS to fulfill the full operational process of a business model. In case of MVDBMS, in one table you can store/maintain large number of fields; sometimes all field values of a record in a single table. As MVDBMS store data in flat file, it doesn't establish relationship between them. So it easy to store, retrieve, update that particular record. For this, sometimes refer to them as record based database systems. jBASE, Revelation, Ladybridge, InterSystems are some examples of multi-valued database.

Data is stored using two separate files: a file to store raw data and a dictionary to store the format for displaying the raw data. Note that in case of MVDBMS term 'file' and 'table' are interchangeable.

MVDBMS just simply use different methods to store values. It uses a set of delimiters to separate the values, multi-values and sub-values of the record. Suppose we have meta data of a table 'Account' under multi value database system is:

FIELD_NAME	LEN	MULTI_MARK	ASSOC
id	20		
account title	35		
email id	35	M	
mobile no	20	M	
txn date	8	M	Y
no of txn	2	M	Y
txn_type	2	S	Y
txn_amt	20	S	Y

Suppose there is a record with ID 10001 of 'Account' file:
10001

ZahidHasan,a@yahoo.com::b@yahoo.com,8801921033322::88
01721033322,20141220::20141221,1::2,CR::CR~DR,10::20~30

Data will be mapped against dictionary file as follows:

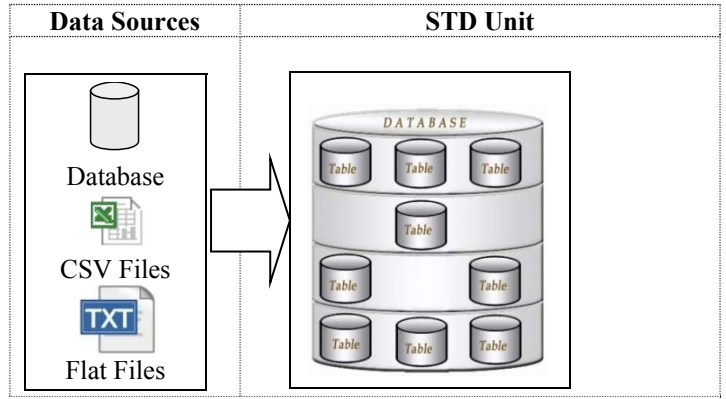
FIELD_NAME	FIELD VALUES
id	10001
account title	Zahid Hasan
email_id	a@yahoo.com::b@yahoo.com
mobile_no	8801921033322::8801721033322
txn_date	20141220::20141221
no of txn	1::2
txn_type	CR::CR~DR
txn_amt	10::20~30

Note that email_id, mobile_no, txn_date, no_of_txn are multi-value fields and separated by '::' delimiter. txn_type, txn_amt are sub-value fields and separated by '~' delimiter.

Another important property of MVDBMS is field association. In dictionary file example note that txn_date, no_of_txn, txn_type, txn_amt fields are associated (ASSOC column is Y). We can observe the impact of this property on data part example. For Account id 10001 there are two transaction date; one on 20/12/2014 with transaction type 'Credit' with transaction amount 10. For second multi-value of txn_date all other association can be understood. For associated field set, values are grouped together. So for an associated field set, no of multi-value for each member field under association must be same. Same way no of sub-value for each sub-valued member field under association must be same. In example txn_date, no_of_txn, txn_type, txn_amt all have 2 multi-values and under each multi-value of a sub-valued field, no of sub-value is also equal. In example there are two sub-valued field txn_type, txn_amt under association; for each field, for first multi-value set, no of sub-value under this is 1 and for second multi-value set, no of sub-value under this are 2. We will discuss the mechanism to transform MVDBMS to RDBMS for efficient query by joining tables as well as extract multi-value, sub-value data to overcome the limitation of MVDBMS on data query.

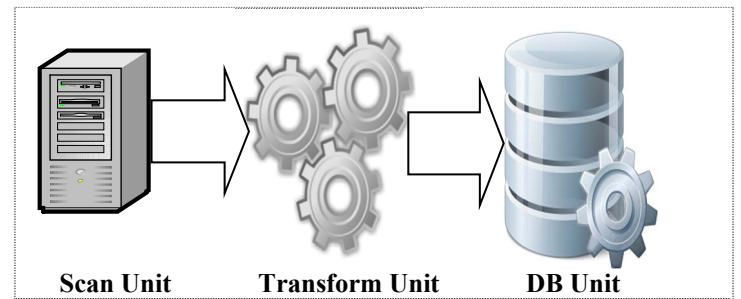
IV. ARCHITECTURE

Data sources can be database/ csv files/ flat file. Source data will be handed over to STL unit, STL unit's output is required relational schema with data.

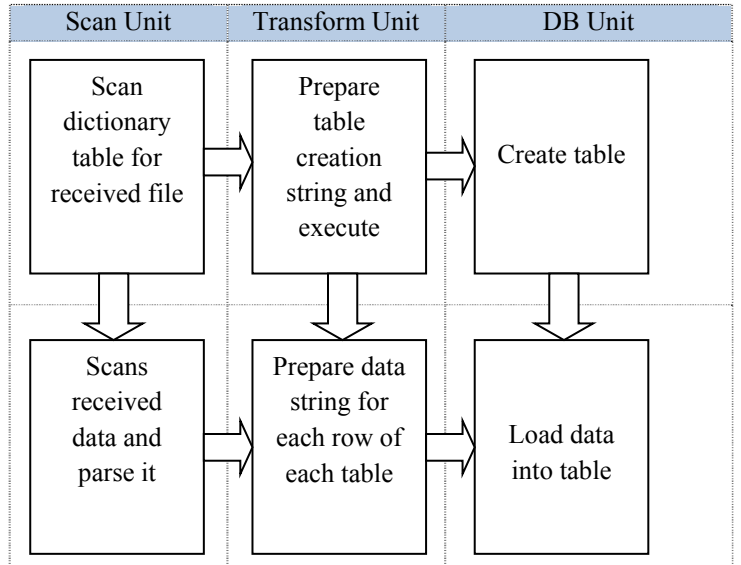


STD unit consists of three units:

1. Scan unit
2. Transform unit
3. Database unit.



Building block of STL unit:



V. WORKING PRINCIPAL

Our objective is to convert data from multi value database systems to RDBMS so that data can be searched more efficiently. We propose two evaluation criteria to transform the system:

- A. Space efficient transformation
- B. Time efficient transformation

A. Space efficient transformation

This technique transform source data into normal form as higher as possible. In this technique source file generally split into many tables into RDBMS. In file, it will have data consisting field values, multi values and sub values of records; values can be found in association or non association mode. If is association found then a set of values are related/grouped to each other within that associated group. In association it may have only multi value set or mixed of multi value and sub value set. If multi value columns are not associated then to normalize each multi value column of a records it will require a separate table. Also in case of associated multi value set of column if it has no sub value then for that whole associated set it will require one table. So it will require n number of separate columns to normalize n number of associated multi value set of columns. Additionally columns which are single valued; means which are not multi valued or sub valued, for all single valued columns it will require one table. Now consider case when an associated value set consists of multi value and sub value set. In this scenario for multi value data set it require one table as well as for sub value data set it will require one more table to normalize the associated data set. So we can list out the cases how data values can form a record:

Case-1: Data values will have some single-valued columns.

Case-2: Data values may have one or more non associated number of multi- valued columns.

Case-3: Data values may have one or more associated number of columns where association exists only within multi-valued columns.

Case-4: Data values may have one or more associated number of columns where association exists with mixed of multi-valued and sub-valued columns.

Note that if a sub value column is exists in any data position it must be under association of a multi value column. A sub value column will never exist separately out of association. In this technique we have taken first multi value column as its identity. In the same way multi value column will never exists without a single value column but unlike sub value it may come without association. Actually this multi value column is associated with the first single value column (first column of a record, normally id of the record). So in case of sub value column it will require record id along with first multi value column of its association to fully identify it.

So we can set an expression for number of tables requires normalizing multi value data systems. Suppose,

A =Number of non associated multi-valued column.

B = Number of associated value set which contain only multi-valued column set.

C = Number of associated value set which contain both multi-valued and sub-valued column set.

Total number of table require in this technique,

$$X = 1 + A + B + 2C$$

Suppose we have Meta data of a table 'account' under MVBDBMS is:

FIELD NAME	LEN	MULTI MARK	ASSOC
account_id	20		
category	5		

FIELD NAME	LEN	MULTI MARK	ASSOC
first_name	35		
email_id	35	M	
txn_date	8	M	Y
txn_type	2	S	Y
txn_amt	20	S	Y
int_rate	6		
doc_name	20	M	Y
doc_id	20	M	Y
branch_id	20		

Fig: TABLE-A

In this method, number of table required as follows:

Number of non associated multi value column, $A = 1$ (Column email_id)

Number of associated value set which contain only multi value column set, $B = 1$ (Columns doc_name, doc_id)

Number of associated value set which contain both multi value and sub value column set, $C = 1$ (txn_date, txn_type, txn_amt)

Single value columns are account_id, category, first_name, int_rate, branch_id.

So total number of tables require to normalize according to this method, $X = 1 + A + B + 2C = 5$

The table structure will be as follows:

For all single value columns, it create table account as:

```
CREATE TABLE account
(
    account_id    varchar2(35),
    category      number(3),
    first_name     number(4),
    int_rate      varchar2(1),
    branch_id     varchar2(1)
)
```

Fig: SCHEMA-A1

For non associated multi value column, it create table named account_email_id as follows:

```
CREATE TABLE account_email_id
(
    account_id    varchar2(35),
    email_id      varchar2(35)
)
```

Fig: SCHEMA-A2

Note that all table (except single value columns) table name convention will be name of the columns one after another prefixed by main table but maximum length will not exceed 30.

For associated value set which contain only multi value column set, it create table account_doc_name_doc_id as:

```
CREATE TABLE account_doc_name_doc_id
(
    account_id    varchar2(35),
    doc_name      varchar2(20),
    doc_id        varchar2(20)
)
```

Fig: SCHEMA-A3

For associated value set which contain both multi value and sub value column set, it will generate two tables. One named as account_txn_date as follows:

```
CREATE TABLE account_txn_date
(
  account_id      varchar2(35),
  txn_date        varchar2(8)
)
```

Fig: SCHEMA-A4

Another one named as account_txn_date_txn_type_txn_amt as follows:

```
CREATE TABLE
account_txn_date_txn_type_txn_amt
(
  account_id      varchar2(35),
  txn_date        varchar2(8),
  txn_type        varchar2(2),
  txn_amt         varchar2(20)
)
```

Fig: SCHEMA-A5

Now concentrate on how data is going to be parsed and get populate into table in this technique. Consider the case that already discussed above(TABLE-A)

Suppose value of the following column account_id, category, first_name, email_id, txn_date, txn_type, txn_amt, int_rate, doc_name, doc_id, branch_id is stored like this:

0004121000001,6001,A,a@yahoo.com::b@yahoo.com,20141220::20141221::20141223,CR::CR~DR::DR,10::20~30::40,5.00,PASSPORT_NO::NID::BIRTH_REG,A1230::123450::98760,BD0010004

Here ',' is field value separator, '::' is multi value separator, '~' is sub value separator.

Columns which have only single value will get populated into single value column table (as per SCHEMA-A1); figure TABLE-A1.

So it populate data columns of single value as follows:

account_id	category	first_name
0004121000001	6001	A
int_rate	branch_id	
5.00	BD0010004	

Fig: TABLE-A1

Columns which are non associative multi value will get populated into separate table TABLE-A2 as SCHEMA-A2. In this case number of inserted rows for that record will be number of multi value of that column. So it populate data columns of non associative multi value set as follows:

account_id	email_id
0004121000001	a@yahoo.com
0004121000001	b@yahoo.com

Fig: TABLE-A2

Columns which are associated with only multi value set will get populated into a separate table TABLE-A3 as SCHEMA-A3. So if there is more than one associated multi value set than it will get populated on separate table. On both cases number of

inserted rows for that record will be number of multi value of any of the associated column. Note that no of multi value will be equal for each column that are associated with each other.

So it populate data columns of associative multi value set as follows:

account_id	doc_name	doc_id
0004121000001	PASSPORT_NO	A1230
0004121000001	NID	123450
0004121000001	BIRTH_REG	98760

Fig: TABLE-A3

Columns which are associated with both multi value and sub value set will get populated into a two different table TABLE-A4 and TABLE-A5 as per SCHEMA-A4 and SCHEMA-A5. One table to relate multi value data with record id say main_multi_set and another to relate sub value data with record id as well as multi value say main_first_multi_sub_set.

In main_multi_set table, columns which are multi-valued among the association get populated. If there is more than one associated value set like this then it will populated on same way on separate main_multi_set table. In all cases number of inserted rows for that record will be number of multi value of any of the associated column. Number of multi value will be equal for each multi value column that are associated with each other.

In main_first_multi_sub_set table, columns which are sub value among the association get populated. In addition the first multi value of that association will be also taken to relate this sub value. So if there is more than one associated value set like this then it will get populated on same way on separate main_first_multi_sub_set table. In all cases no of inserted rows for that record will be sum of sub values of each multi value of the associated column. Number of sub value will be equal for each sub value column that are associated with each other.

It populate data columns of associative multi value set as:

account_id	txn_date
0004121000001	20141220
0004121000001	20141221
0004121000001	20141223

Fig: TABLE-A4

account_id	txn_date	txn_type	txn_amt
0004121000001	20141220	CR	10
0004121000001	20141221	CR	20
0004121000001	20141221	DR	30
0004121000001	20141223	DR	40

Fig: TABLE-A5

B. Time efficient transformation

In this technique we concentrate on how much time it will take to process a query after loading into data warehouse. In this technique simply we will create only two tables.

One for all the single value columns. Another one for all the multi-value and sub-value columns whether it is associated or not. So number of table require on this technique is 2.

Considering scenario of figures TABLE-A, it will create one table for single value columns named as account_main as follows:

```
CREATE TABLE account_main
(
  account_id    varchar2(20),
  category      varchar2(5),
  first_name    varchar2(35),
  int_rate      varchar2(6),
  branch_id     varchar2(20)
)
```

Fig: SCHEMA-B1

Another one for all the multi value and sub value columns whether it is associated or not as account_multi_sub as follows:

```
CREATE TABLE account_multi_sub
(
  account_id    varchar2(20),
  mv_seq        number(3),
  sv_seq        number(3),
  email_id      varchar2(35),
  txn_date      varchar2(8),
  txn_type      varchar2(20),
  txn_amt       varchar2(20),
  doc_name      varchar2(20),
  doc_id        varchar2(20)
)
```

Fig: SCHEMA-B2

Now concentrate on how data is going to be parsed and get populate into table in this technique. We will use same data set as used in previous technique.

Columns which have only single value will get populated into account_main table as per SCHEMA-B1.

account id	category	first name
0004121000001	6001	A
int rate	branch id	
5.00	BD0010004	

Fig: TABLE-B1

Columns which have multi value and sub value set will get populated into account_multi_sub as per SCHEMA-B2.

account id	mv_seq	sv_seq
0004121000001	1	1
0004121000001	2	1
0004121000001	2	2
0004121000001	3	3
email id	txn_date	txn_type
a@yahoo.com	20141220	CR
b@yahoo.com	20141221	CR
b@yahoo.com	20141221	DR
	20141223	DR
txn amt	doc name	doc id
10	PASSPORT NO	A1230
20	NID	123450
30	NID	123450
40	BIRTH REG	98760

Fig: TABLE-B2

In account_multi_sub table, there will be a row for each multi-value position as well as for each sub-value position under that multi-value position of a particular record.

Consider multi-value set of a record, for each multi-value element there will be a separate row. So in this case actually it will insert maximum number of multi-value for any of the fields on that record. Say if there are 3 multi-valued fields on a value set; first one has 1 multi-value, second has 3 and third has 2. When considering whole value set, it create 3 rows to separately identify multi values and identify relationship between them.

Now consider case where there is sub-valued field among the whole value set (on each multi-value, it may have more than one sub-value). Suppose there are two sub-valued fields on a value set, where first one has 1 multi-value and under it 2 sub-values, it will require 2(maximum number of sub-value for its only multi-value position) rows to be inserted. Second has 2 multi-value where it has 2 and 3 number of sub-values under its first and second multi-value respectively. For first multi-value position maximum number of sub-value is 2 and for second multi-value position is 3. So for first multi-value position of the value-set it will require 2(maximum number of sub-value for first multi-value position) rows to be inserted and for second multi-value position of the value-set it will require 3(maximum number of sub-value for second multi-value position) rows to be inserted.

Say n is the number of multi-value element of a field among a value set and S1,S2,.....,Sn are the maximum number of sub-value for 1st,2nd,.....,nth multi-value element respectively. So total no of rows inserted for each record will be:

$$X = S1+S2+.....+Sn = \sum_{i=1}^n S_n$$

VI. IMPLEMENTATION

Working principal has been implemented using following:

- Java as programming language
- jdk-7 as development kit
- netbeans-7.4 as java editor
- oracle as RDBMS
- Toad as oracle editor

VII. PERFORMANCE EVALUATION

We measure space and time on both space and time efficient case and evaluate space versus time to evaluate performance.

First we take a look back on how tables created and inserted data in space efficient technique. By applying this method we will get schemas in third normal form; so except foreign key we will not get data duplication. On the other hand in case of time efficient method only two tables will be created. By applying this method we will get schemas in first normal form; so in table there will be huge data duplication can be found.

Suppose there is a subvalue field which have 9 multivalues m1 ,m2 ,m3,.....,m9 and multivalues have 2, 4, 6, 8, 10, 12, 14, 16, 18 subvalues respectively. And another subvalue field (not associated) have 9 multivalues n1 ,n2 ,n3,.....,n9 and multivalues have 18, 16, 14, 12, 10, 8, 6, 4, 2 subvalues respectively. Now we will calculate number of rows inserted by each technique.

In space efficient technique, number of inserted rows will be=2+4+6+8+10+12+14+16+18=50*2=100

In time efficient technique, number of inserted rows will be=18*9=162. Also other multi values and sub values data will get duplicated for same multi value and sub value position. So it is clear that first method will take less space than second one.

Now take a look on time efficiency. Space efficient method generates schema in 3NF; number of tables on this method is good in number (depends on dictionary file). So it requires frequent joining operation to get expected query result. On the contrary time efficient method generates schema in 1NF; number of tables on this method is only 2. So it requires very less joining operation to get expected query result. Number of tables created in space efficient method against a table follows:

No of non associated MV column(A)	No of associated MV column set(B)	No of associated set which contain both MV and SV column set(C)	$X=1+A+B+2C$
1	1	1	5
2	2	2	9
3	3	3	13
4	4	4	17
5	5	5	21

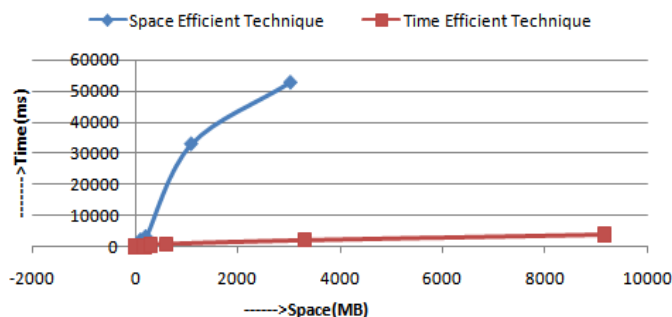
As the number of joining operation increase, time require executing query also increase. It summarize that in space efficient technique it save space over time efficient technique; first one is space efficient. But time efficient technique save time over space efficient technique; second one is time efficient.

Sample data of MVDBMS taken and convert it into both techniques, calculate occupied space. Then we perform various queries to get expected result set and calculate elapsed time. Space and time measurement for both techniques are as follows:

Sample No	Space(MB) X		Time(ms) Y	
	Space Efficient Technique	Time Efficient Technique	Space Efficient Technique	Time Efficient Technique
1	0.43	1.28	162	9
2	2.59	7.74	345	23
3	59.91	180.33	540	227
4	99.45	300.34	2005	655
5	199.59	604.75	3107	904
6	1087.06	3304.66	33060	2137
7	3025.01	9150.51	52903	4007

Space versus Time table

Space vs Time Graph



VIII. CONCLUSION AND FUTURE WORK

In this paper, we have identified that a multi-valued database cannot handle complex queries well. Therefore, we have proposed and then elaborated two approaches to transform multi-value database systems into relational database systems for faster querying and supporting complex SPJ queries. One of the proposed approach works well when only limited amount of storage space is available while the other approach executes queries faster than the former which makes it more suitable to online or streaming data platform.

In future, we would like to extend this work and want to build an intelligent system that will be capable of analyzing the current system and data characteristics to determine the appropriate platform, i.e. either MVDBMS or RDBMS, as well as to transform MVDBMS into RDBMS (if required) using which of these two proposed approaches. We would like to also investigate the possibility of building a hybrid approach which can offer us the goods of both of these proposed approaches.

REFERENCES

- [1] Karamjit Kaur, Rinkle Rani; "Modeling and querying data in NoSQL databases " in IEEE International Conference on Big Data, 2013
- [2] Xiang Li, Zhiyi Ma , Hongjie Chen; "QODM: A query-oriented data modeling approach for NoSQL databases" in IEEE Workshop on Advanced Research and Technology in Industry Applications (WARTIA), 2014
- [3] Anderson Chaves Carniel, Aried de Aguiar Sá, Vinícius Henrique Porto Brisighello, Marcela Xavier Ribeiro, Renato Bueno, Ricardo Rodrigues Ciferri , Cristina Dutra de Aguiar Ciferri; "Query processing over data warehouse using relational databases and NoSQL" in XXXVIII Conferencia Latinoamericana En Informatica (CLEI), 2012
- [4] Kudakwashe Zvarevashe, Tatenda Trust Gotora; "A Random Walk through the Dark Side of NoSQL Databases in Big Data Analytics" in International Journal of Science and Research (IJSR); ISSN (Online): 2319-7064
- [5] Mohamed A. Mohamed , Obay G. Altrafi, Mohammed O. Ismail; "Relational vs. NoSQL Databases: A Survey " in International Journal of Computer and Information Technology (ISSN: 2279 – 0764) Volume 03 – Issue 03, May 2014
- [6] Biswajeet Sethi, Samaresh Mishra, Prasant ku. Patnaik; " A Study of NoSQL Database " in International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Vol. 3 Issue 4, April - 2014
- [7] <http://www.jbase.com/products/jbase-multivalue-database/>
- [8] Zhao, Y.; Ramasamy, K., Tufte; K., Naughton, J.F., 'Array-based evaluation of multi-dimensional queries in object-relational database systems', 14th International Conference on Data Engineering 1998 pp-241-249.
- [9] R. Kimball, 'Data Warehousing Toolkit', 1995 :John Wiley & Sons
- [10] M. Gyssens and L. V.S. Lakshmanan 'A Foundation for Multi-Dimensional Databases', VLDB'97, pp.106 -115 1997
- [11] R. Agrawal , A. Gupta and S. Sarawagi, 'Modeling Multidimensional Databases', ICDE'97, vol. 97, pp.232 -243
- [12] Suchitra Reyya; T. Summallika; G.V.M. Vasuki, 'An approach to store spatial big-data using multi-valued database'. International Journal of Research in Engineering & Technology (IMPACT: IJRET) ISSN(E): 2321-8843; ISSN(P): 2347-4599 Vol. 1, Issue 6, Nov 2013, 15-22

Improvement of PF and Reduction of THD using PID Controlled Current Sensored Three Phase Rectifier

MD Nasmus Sakib Khan Shabbir, Mahbubur Rahman , Farhana Afrin , Shamsun Naher Shila , Md. Shafayet Hossain
Bangladesh University of Business and Technology
eng.sakib@gmail.com

Abstract— a new 3-phase current controlled PFC rectifier has been proposed in this paper. Phase difference in between input voltage and current is minimized using an a-periodic control signal generated using input current, phase angle of input voltage, and output voltage. A 1V (P-P) 5KHz triangular wave is used here to generate PWM signal. Simple PID controllers have been added with the controller of the rectifier which reduce fluctuations and thereby ensure better performances. The major advantages of the proposed rectification technique are reduced harmonic distortion along with improved power factor. Further the PSIM simulation supporting the merits of proposed rectifier is recorded and demonstrated in this work.

Keywords—clarke transformation, parktransformation, PID controller, PFC, THD

I. INTRODUCTION

Reactive power compensation and power factor correction is much desirable in AC power system. Improving the power factor allows financial and technical advantages because national power supply companies host high tariffs on reactive energy [1]. The existence of harmonics increases the apparent power and as a result power factor declines [2],[3]. In large industry reactive Power compensation is traditionally lessen by large capacitors but it doesn't remove harmonics whereas it boost harmonics if not applied properly. A resonance frequency introduces at which Thevenin impedance matches the capacitor bank reactance [4]. If nonlinear load is applied and resonance frequency is near to harmonics caused by the load, there will be harmonic intensification creating over voltages and over currents of this frequency and causes the destruction of the capacitor bank [4]. There are so many flaws in traditional system for using bridge rectifiers and contains current harmonics which causes the rise of Total Harmonic Distortion (THD). Modern rectifiers are required to bind current harmonics according to international Standards IEC 1000-3-2, EN61000-3-2, and IEEE519. According to 'IEEE Std 519-1992 Harmonic Current Limits' for a power supply system if I_{sc}/I_L is between 20 to 50 maximum allowable THD for input current is 8% and for voltage it is 5%. Magnitude of the 5th and 7th harmonics should not be greater than 7% of the fundamental component and harmonics order in any magnitude greater than 35th must not be injected in the system. [10] In writings, various algorithms of three phase PWM rectifiers have been suggested to achieve such desires. The one-cycle current control method was proposed in [5]. Though the method is modest to implement, the calculation of duty cycle of switches are directly based on measured currents in

the open loop method. Hence, Comparing with other current closed loop methods, the sensitivity of control plan is highly reliant on the quality of current measurement .In [6], without using current control method, the power factor control system is based on instantaneous active power and reactive power control. The Virtual assessed flux and line currents are used to compute power. In [7], an improved hysteresis current control method was proposed to minimize ripple currents and number of switching per cycle but in digital execution this proposed method requires high frequency digital controller. In this paper, a PID controller based current oriented control with lower frequency switching scheme is proposed.

II. CONTROLLER DETAILS

A. Working Principle

In conventional PWM rectifier converter voltage and phase angle is generally controlled which indirectly control amplitude and phase of input current and thus phase difference is created in between voltage and current. [8] problem arises as this change is not consider during PWM switching signal generation. To mitigate this we propose a controller where output voltage, input current and input voltage phase has been considered. using a phase detector phase of the input voltage is sensed, using this as a reference angle input current is converted into dqo co-ordinate. As it is a balance three phase system so I_o will be zero. I_q controls the phase of the control signal so that input PF remain unity or near of unity where as I_d adjusts control signal amplitude to keep the output voltage constant. Controller will keep the converter power invariant. PWM signal is generated using a reference signal of 1V(P-P) 5Khz saw-tooth wave. In spite of making continuous signal PID controller helps to generate adjustable signal which indeed helps to get better performance at lower frequency. In this paper, a PID controller based current oriented control with lower frequency switching scheme is proposed.

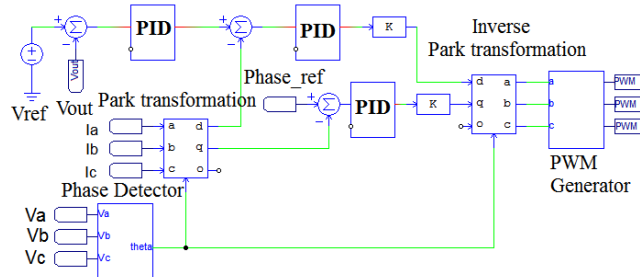


Fig. 1. Block diagram of the controller

B. Phase Detector:

To detect time to time phase of the input voltage a clarke or abc to alpha-beta-gama transformation block is used here. As input is balanced, so gama component will be zero [9]. Using an inverse tangent block to the ratio of beta and alpha theta (θ) is detected which is in radian scale. Here θ is the phase angle of the voltage of phase a. equations can be written as

$$\theta = \tan^{-1}\left(\frac{I_\beta}{I_\alpha}\right) \text{ where}$$

$$I_\alpha = \sqrt{\frac{2}{3}}\left(I_a + \frac{I_b}{2} - \frac{I_c}{2}\right)$$

$$I_\beta = \sqrt{\frac{2}{3}}\left(\frac{\sqrt{3}I_b}{2} + \frac{\sqrt{3}I_c}{2}\right)$$

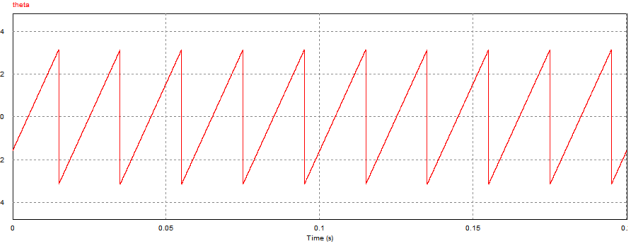


Fig. 2. Output of phase detector

C. Mathematical Modeling of the switching signal:

Using theta (θ) obtain from phase detector, three phase line current is converted into dqo co-ordinate. Because of symmetry component of O is zero. Equations are stated below

$$I_d = \frac{2}{3} \left\{ (\cos\theta)I_a - \cos\left(\theta - \frac{2\pi}{3}\right)I_b + \cos\left(\theta + \frac{2\pi}{3}\right)I_c \right\}$$

$$I_q = \frac{2}{3} \left\{ -(\sin\theta)I_a - \sin\left(\theta - \frac{2\pi}{3}\right)I_b + \sin\left(\theta + \frac{2\pi}{3}\right)I_c \right\}$$

Reference of amplitude controller I_d is generated from V_{ref} and V_{dc} , transfer function of the PID controller used here is $1K \times \frac{s+5K}{s^2 30n + s 200\mu}$.

$$I_{dref} = -\frac{R_{F2}}{R_3}(V_1) \text{ Where}$$

$$V_1 = -\frac{R_{F1}}{e} \frac{t}{R_{F1}C_{F1}} \left(\frac{V_{ref}}{R_1} + C_1 \frac{d}{dt}(V_{ref}) - \frac{V_{dc}}{R_2} - C_2 \frac{d}{dt}(V_{dc}) \right)$$

Which is then pass through a PID controller with I_d and become d_{in} of inverse park transformation block.

$$d_{in} = \left(1 + \frac{R_{F4}}{R_6}\right)(V_2)$$

$$V_2 = -\frac{R_{F3}}{e} \frac{t}{R_{F3}C_{F3}} \left(\frac{I_{dref}}{R_4} + C_4 \frac{d}{dt}(I_{dref}) - \frac{I_d}{R_5} - C_5 \frac{d}{dt}(I_d) \right)$$

On the contrary phase controller I_q passes through an PID controller with P_{ref} and become q_{in} of inverse park transformation block.

$$q_{in} = \left(1 + \frac{R_{F4}}{R_6}\right)(V_3)$$

$$V_3 = -\frac{R_{F3}}{e} \frac{t}{R_{F3}C_{F3}} \left(\frac{P_{ref}}{R_4} + C_4 \frac{d}{dt}(P_{ref}) - \frac{I_q}{R_5} - C_5 \frac{d}{dt}(I_q) \right)$$

Based on sensed theta (θ), d_{in} and q_{in} are then converted into abc co-ordinates, Which act as reference for switching signal generation. These control signals contain input current, phase of voltage, desired out voltage, so signal generated through these will tend to keep the output constant, improve input PF and reduce THD of input current.

$$\begin{bmatrix} I_a \\ I_b \\ I_c \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta & 1 \\ \cos\left(\theta - \frac{2\pi}{3}\right) & \sin\left(\theta - \frac{2\pi}{3}\right) & 1 \\ \cos\left(\theta + \frac{2\pi}{3}\right) & \sin\left(\theta + \frac{2\pi}{3}\right) & 1 \end{bmatrix} \begin{bmatrix} d_{in} \\ q_{in} \\ o_{in} \end{bmatrix}$$

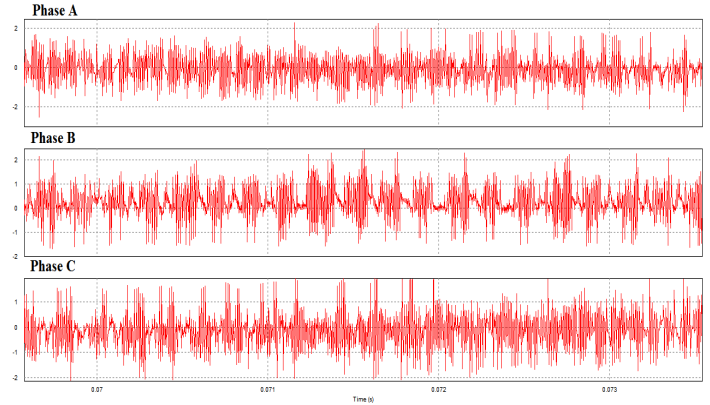


Fig. 3. Control signals

D. Switching sequence:

To avoid dead time losses switches connected to the same phase (upper and lower) are complementary. To have an initial idea about our switching pattern we have divided one cycle into six broad regions spanned from one to next crossover point (fig.4) During crossover upper switch of the highest voltage and lower switch of the lowest voltage are on, complementary switches are off and rest are PWM. Here on means duration of on time of those switches will be maximum in spite of continuously on during that period. To illustrate this, phase switching signals of phase C is given below in (fig.5).

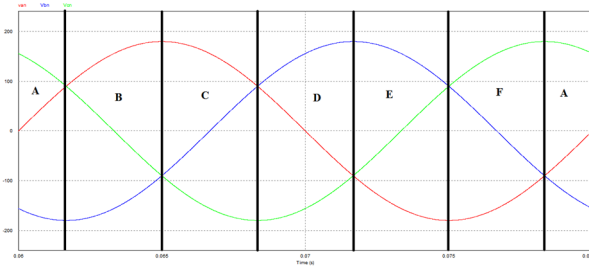


Fig. 4. Switching regions

TABLE I. SWITCHING PATTERN IN DIFFERENT PHASES

Phase	a low	a up	b low	b up	c low	c up
A	PWM	PWM	on	off	PWM	on
B	off	on	on	PWM	PWM	PWM
C	PWM	on	PWM	PWM	on	off
D	PWM	PWM	off	on	on	PWM
E	on	off	PWM	on	PWM	PWM
F	on	PWM	PWM	PWM	off	on

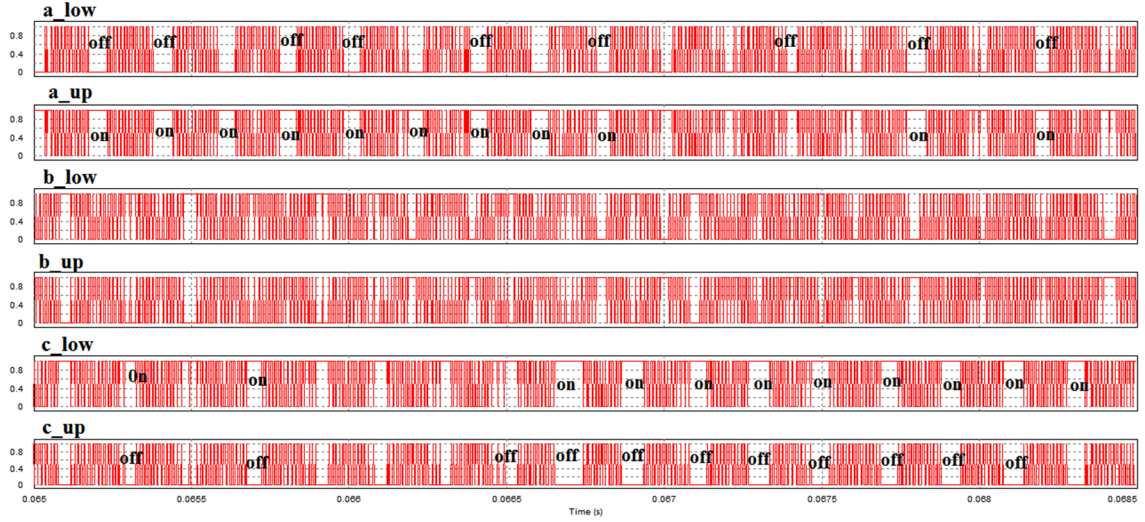


Fig. 5. Switching sequence during phase C

III. PARAMETERS SPECIFICATIONS

A. Maximum Value of line inductance:

Rectifier performance vastly depends on line inductance. It makes input behaves like a current source and provides boost facility. Higher inductance reduces current ripple but also reduces rectifier operation range. For proper operation

$$V_{dc} > \sqrt{3(V_m^2 + \omega L_{max} I^2)}$$

Which gives

$$L_{max} < \frac{\sqrt{\frac{V_{dc}^2}{3} - V_m^2}}{\omega I}$$

And it is 16.5mH for our system.

B. PID controller:

General transfer function of the PID controller used here is

$$H(s) = R_{F1} \times \frac{(S + \frac{1}{C_{F1} R_{F1}})}{S^2(C_1 + C_2) + S(\frac{1}{R_1} + \frac{1}{R_2})} \text{ Which is } 1K \times \frac{S + 5K}{S^2 30n + S 200\mu}$$

for the PID controller used to generate I_{dref} and $23.055 \times \frac{S + 455.75}{S^2 24n + S 2m}$ for the rest. According to the Routh-Hurwitz criteria all are limitedly stable. Input capacitance C2 of the

controllers used to generate d_{in} and q_{in} is higher than C1 as inputs of this node are I_d and I_q which are generated from park transformation of input current has higher fluctuation.

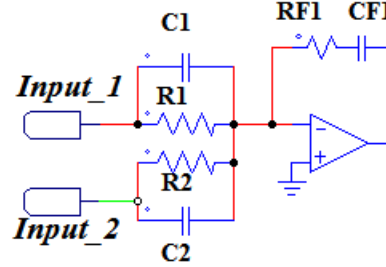


Fig. 6. PID controller

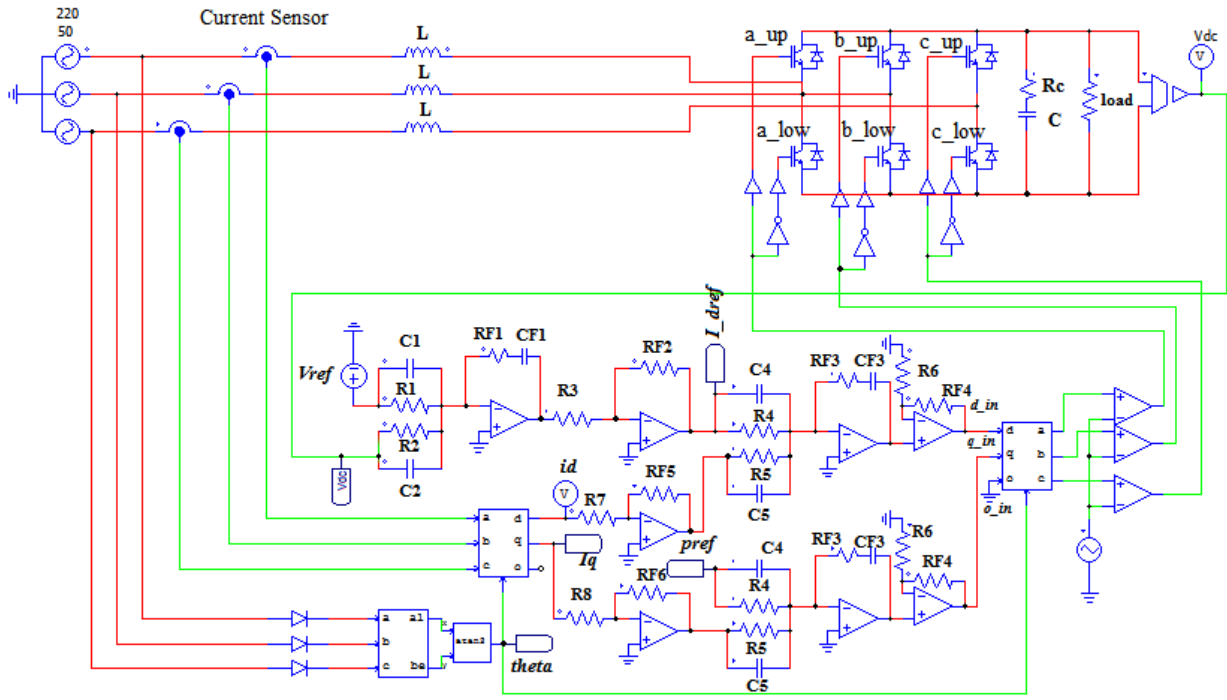


Fig. 7. Complete circuit

IV. RESULTS AND DISCUSSION

The controller has kept the power factor near of unity. It bounds input currents THD within 4.22% that satisfies the 'IEEE Std 519-1992 Harmonic Current Limits' which says the maximum allowable THD for input current is 8%. Fundamental components of I_a , I_b and I_c contains 106.08A, 107.49A and 107.73A. Total harmonic distortion for V_a , V_b and V_c are 0.0188%, 0.0167% and 0.0167% respectively which is much less than the maximum allowable 5% THD for voltage according to 'IEEE Std 519-1992'. Using PSIM simulator, the PF of the phase A, B, C are found to be 0.997, 0.981 and 0.988. Efficiency of the converter is 94.2%.

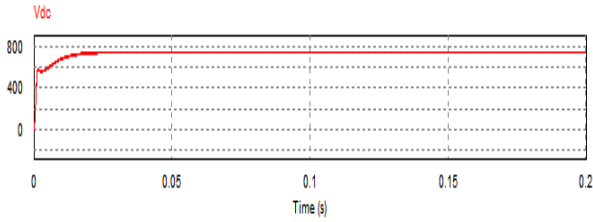


Fig. 8. Output voltage

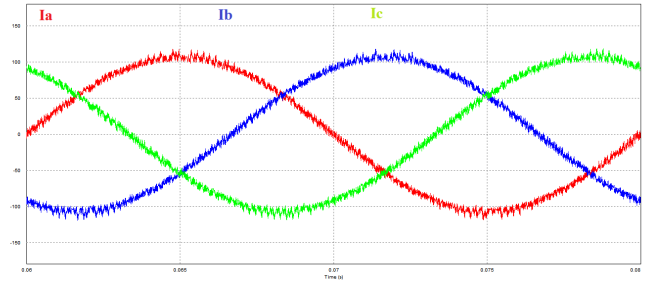


Fig. 10. Input currents wave shapes

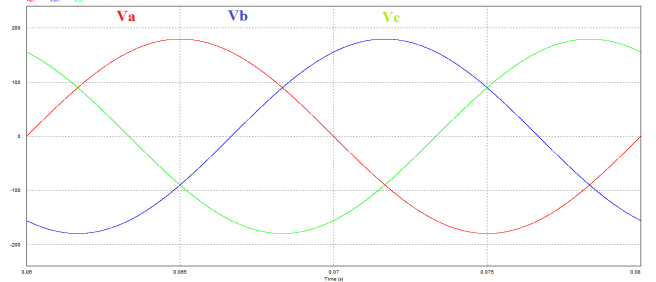


Fig. 11. Input voltage

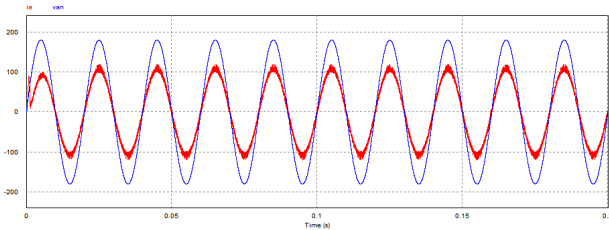


Fig. 9. Input voltage and current for phase a with 0.997 PF

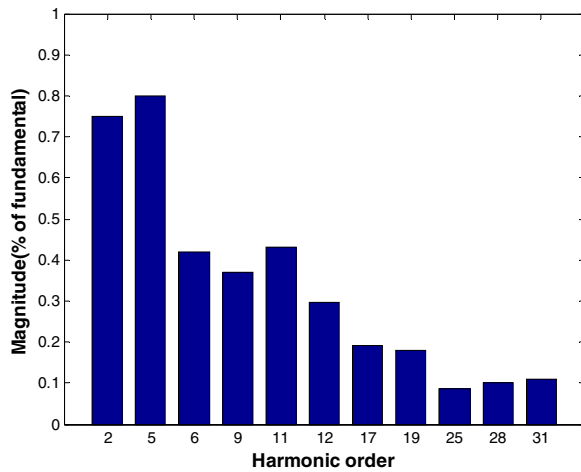


Fig. 12. Harmonics profile of phase A current

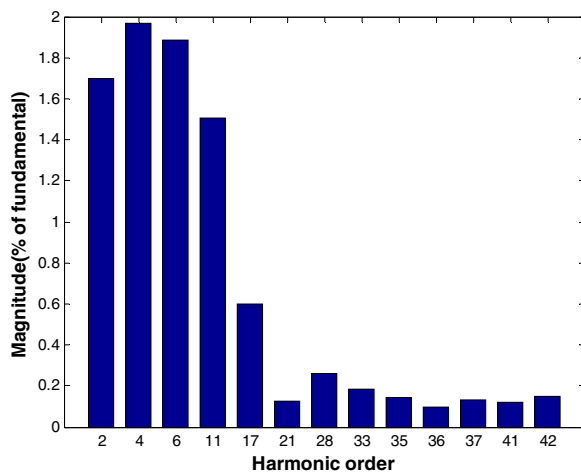


Fig. 13. Harmonics profile of phase B current

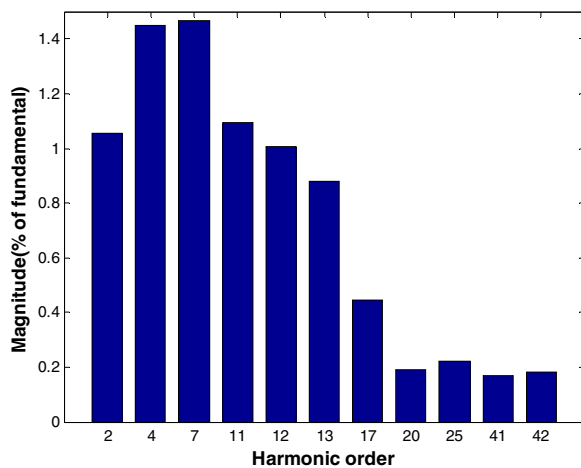


Fig. 14. Harmonics profile of phase C current

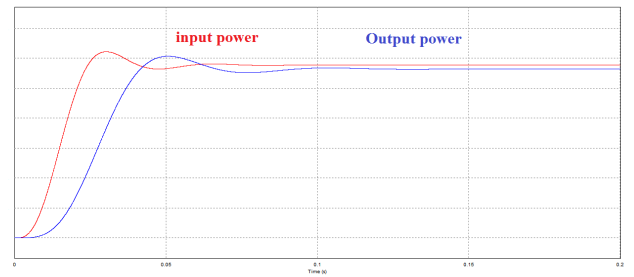


Fig. 15. Input and output power

V. CONCLUSION

Performance of the converter improves significantly. PF of the phase A, B, C are 0.997, 0.981 and 0.988 whereas PF using the PI controller was 0.96. Total harmonic distortion for V_a , V_b and V_c are 0.0188%, 0.0167% and 0.0167% respectively and of input currents are 4.15%, 4.21% and 4.22% respectively whereas using PI controller THD was 10.06%. Improved PF comes with improved efficiency which is 94.2%. Because of PID controller better performance is possible at lower triangular frequency.

- [1] H. K. Tyll, Dr. F. Schettler, "Historical overview on dynamic reactive power compensation solutions from the begin of AC power transmission towards present applications", Power Systems Conference and Exposition, Seattle, WA, pp. 1-7, March 15-18, 2009.
- [2] V.M. Quang, H. Wei, W. Dazhi, W. Xuming, "A New Type of PWM Rectifier with Function of Harmonic Suppression and Reactive Power Compensation", 25th Chinese Control and Decision Conference, pp. 3013 - 3017 May 25-27, 2013
- [3] C. Chavez, J. Houdek, "Dynamic Harmonic Mitigation and Power Factor Correction", 9th International Conference Electrical Power Quality and Utilization, Barcelona, 9-11 October 2007.
- [4] S. Acevedo, A. Llamas, J. Baez, J. De los Reyes, "Filters for Power Factor Correction in the presence of Non-Linear Loads", Proceedings of the Power Quality 99 Conference, Chicago, Illinois, USA, November 9-11, 1999, pp. 32-39.
- [5] G. Chen and K.M. Smedley, "Steady-State and Dynamic Study of One-Cycle-Controlled Three-Phase Power-Factor Correction", IEEE Trans. On Industrial Electronics, Vol. 52, No. 2, pp. 355-362, April 2005.
- [6] M. Malinowski, M. Jasinski, and M.P. Kazmierkowski, "Simple Direct Power Control of Three-Phase PWM Rectifier Using Space-Vector Modulation (DPC-SVM)", IEEE Trans. On Industrial Electronics, Vol. 51, No. 2, pp. 447-454, April 2004.
- [7] A.N. Tiwari, P. Agarwal, and S.P. Srivastava, "Modified hysteresis controlled PWM rectifier", IEE Proc-Electr. Power Appl., Vol. 150, No. 4, pp. 389-396, July 2003
- [8] KASMIERKOWSKI MARIAN P., BLAABJERG F., KRISHNAN R.- "Control in power electronics, selected problem". Elsevier Science, USA. 2002. 518p. ISBN : 0-12-402772-5
- [9] W. C. Duesterhoeft; Max W. Schulz; Edith Clarke (July 1951). "Determination of Instantaneous Currents and Voltages by Means of Alpha, Beta, and Zero Components". Transactions of the American Institute of Electrical Engineers 70 (2): 1248–1255. doi:10.1109/T-AIEE.1951.5060554. ISSN 0096-3860
- [10] Thomas M. Blooming, Daniel J. Carnovale 'application of iecce std 519-1992 harmonic limits'

Automated Soil Clustering for Crops Cultivation

Afrida Islam
Akash Bhuiyan
Junaeid Ahmed
Computer Science and Engineering
United International University
Dhaka, Bangladesh
e-mail: aislam121041@cse.uiu.ac.bd
e-mail: abhuiyan121019@bscse.uiu.ac.bd
e-mail: mislam123013@bscse.uiu.ac.bd

Md. Aminul Islam
Mohammad Nurul Huda
Computer Science and Engineering
United International University
Dhaka, Bangladesh
e-mail: mislam11012@cse.uiu.ac.bd
e-mail: mnh@cse.uiu.ac.bd

Abstract— This paper focuses on the idea of cultivating a certain crop to a certain District-Sub District-Union of Bangladesh. The cluster and classification are done first on the basis of soil features and then comparing the soil features with the common features of the crops to be cultivated, is being searched for. Here k-mean Clustering, Fuzzy c-Mean clustering and SOM (Self-Organizing Map) based clustering algorithms are used. These algorithms are analyzed according to their results to find the better classification, by which one can find the most perfectly matched cultivable crops in a certain area. The big challenge of this paper is data collection, which was gathered from Sub-districts level books containing soil and crops features, both organic and physical property of Bangladeshi soil. These books were borrowed from Soil Research and Development Institute (SRDI), Khamarbari, Dhaka, Bangladesh. In this study, soil data of Louhgonj Sub-district of Munsigonj District in Bangladesh are analyzed.

Keywords— *k-Mean clustering; Fuzzy c-Mean clustering, Self Organizing Map; Soil clustering;*

I. INTRODUCTION

Bangladesh is a land of agriculture where a lot of people depend on farming. Regarding this if a better way can be introduced for crops cultivation, that will help this country. Soil clustering for better cultivation would be a great help in this sector. Clustering algorithms [1-5] have various applications such as soil clustering, food clustering, student group clustering, etc. Here it is used for soil clustering.

Previously some investigations were done on soil classification [6-7]. Food and Agriculture Organization (FAO) of United Nations [8-9] has recommended a Universal soil classification based on soil series. Some authors in Australia worked on soil classification [10]. But, previous works of soil classification and clustering were certainly limited with k-means clustering or Fuzzy clustering, where works are being done on basis of chemical elements or only soil series or

taxonomy. Moreover, all the other works are concentrated only in soil classification based on various features and properties of soil, which are totally not related with crops features or idea of crops cultivation. However, in Bangladesh this type of research work are not available. In some books [11-12] few analysis on soil features was given.

This paper highlights on soil clustering. Soil is called the “Skin of Earth”, a complicated and complex natural element which is combination of various particles organic and inorganic. In this study we have found that soil provides a wide range of properties like physical, chemical and biological. Grouping of soil can be done on basis of Taxonomy, Texture, Chemical elements, Soil Series variation or even the Physical location of the soil series in a certain area. Here, soil clustering is done based on k-mean Clustering, Fuzzy c-Mean clustering and SOM (Self-Organizing Map). Besides, the best cultivable crops are identified by matching the cluster properties with the soil properties of those crops.

The challenges of our research work was data collection. In this connection we have collected the “Land and Soil Resources Utilization Guide” for different upazilas (sub-districts) of Bangladesh. This guide [13] was the initiative of Soil Resource Development Institute (SRDI) based on the program of semi-detailed soil. The “Land and Soil Resources Utilization Guide” – is popularly known as Upazila Nirdeshika. This guide discusses about soil characteristics like-soil texture, soil reaction, consistence, horizon nature.

The paper is organized as follows: Section II discusses the data collection, while Section III elaborates the soil feature extraction process. Again, Section IV illustrates the soil clustering mechanism and crops selection based on clustered soil. Finally, Section V gives the experiments and their set ups, Section VI analyses the experimental results and Section VII concludes the paper with future works.

II. DATA COLLECTION

A. Corpus [13]

Since this paper is focusing on crops cultivation, some features that must be considered are Land type, Water recession status, Soil reaction, Soil juice of that very certain area. Upazila Nirdeshika, the guide book is a Sub-District level book where information about Mapping Unit, Physiographic Landform, Land type, Soil group, Water recession during dry season, Analytical data, Soil sampling points, Crop suitability, Fertilizer recommendation etc for that sub-district are given. This guide has been used for land use planning, crop specific fertilizer recommendation and post disaster agricultural management too.

B. Features Specification

- Land Type : High Land, Medium High Land, Medium Low Land, Low Land and Very Low Land
- Water Recession : Very Early, Early, Normal, Late and Very Late
- Drainage Status: Fine, Moderately Fine, Little Bad, Bad and Very bad
- Soil components (a part of soil texture) : Sandy, Sandy loamy, Loamy, Clay loamy and Clay
- Soil Rigidity : Loose, Crumbly, Firm and Sticky
- Soil Juice : Low, Medium, High and Very High
- Soil Reaction : Very Highly Acidic, Highly Acidic, Mild Acidic, Neutral, Mild Base, Highly Base and Very Highly Base

C. Data Processing

Upazila Nirdeshika is huge guide book with so many data, only for a single Sub-District. These guide books also provide a Map with it for easy understanding of data. But these data are not in the form to be clustered so easily. For that reason the data were being processed and nicely configured in such a way so that the clustering algorithm works finely. First of all it is being find out, how many unions is there on that certain Sub-District. A single Union can have different types of soil. According to soil series the areas of the map is shown in different color. They are numbered sequentially mention in the guide book by numbering. Then all these are being mapped according to the features which are being considered for clustering. As a single Union can have more than one type of soil so for area code “3 characters of District - 3 character of Sub District-the whole union name-serial number” is used.

For example, the Bejgao union of Louhojong sub district in Munsigonj Distrcit has Soil components- Clay, Clay Loamy and Loamy, Soil Reaction- Neutral, Mild Acidic, Land Type- Medium High Land, Medium Low Land and other variety too. So for this single BEJGAO union all possible combination of data is considered and area codes are given as: MunLouBejgao1, MunLouBejgao2, MunLouBejgao3, etc.

For considering crops data, the features which are present for a certain crop is being considered and according to that the features are being numbered with 0/1 or numeric. So that crops data can easily match with the soil data.

For example, Local Aus is cultivated in Soil Component of- Loamy or Clay or Clay Loamy, where the land type may be- High Land or Medium High Land with Drainage- Little Bad or Very Bad. So the every feature which are present are considered.

According to this process a good number of data are made for a single union to be matched with the crops. That huge numbers of distinct dataset are made for a single Sub-District which is further being clustered through the clustering algorithms.

III. FEATURE EXTRACTION

Before implementing the algorithms some task were performed to make a specific data set which could be used for clustering. Among them the first task was feature extraction. Three types of feature vectors were found during feature extraction. Table 1.1 shows them clearly.

In Table 1.1 Vector-I has 25 Dimensions because at first in a certain Sub-Districts only those features were considered which were present, and according to that in a certain Area code the present sub-features were given 1(one) and non-present were denoted with 0(zero). For example MunLouBejgao4 have "Medium High Land" Sub-feature which was given 1 but MunLouBejgao5 don't have "Medium High Land" Sub-feature so in that column was given 0.

In Table 1.1 Vector-II has 36 Dimensions because here each feature all the Sub-features were considered and each Feature's Sub-features were given numeric value sequentially, instead of binary 0-1. For example for Feature Land Type: High Land, Medium High Land, Medium Low Land, Low Land and Very Low Land were given numeric value 1,2,3,4 and 5 following. If a sub-feature is not present then that sub-feature is given the value 0.

In Table 1.1 Vector-III have only 7 Dimensions because here only the considered Features are denoted by their numeric value that were assigned in Vector-II. Table 1.2 will give a bird eye view of Vector-III.

Table 1.1: Soil Feature Vector Types

Feature Vector Types	Features
Vector-I (24 dimensions)	Medium Juice, Low Juice, Bad Drainage, Medium High Land, Normal Water Recession, Neutral Reaction, Mild Acidic Reaction, Very Late Water Recession, Low Land, High Juice, Very Bad Drainage, Loose Rigidity, Medium Low Land, Clay, Firm Rigidity, Late Water Recession, Loamy, Crumbly Rigidity, Mild Base Reaction, N/A Juice, Sandy, Early Water Recession, Sandy Loamy, Clay and Loamy.
Vector-II (36 dimensions)	High Land, Medium High Land, Medium Low Land, Low Land, Very Low Land, Very Early Water Recession, Early Water Recession, Normal Water Recession, Late Water Recession, Very Late Water Recession, Fine Drainage, Moderately Fine Drainage, Little Bad Drainage, Very Bad Drainage, Sandy, Sandy Loamy, Loamy, Clay Loamy, Clay, Loose Rigidity, Crumbly Rigidity, Firm Rigidity, Sticky Rigidity, Low Soil Juice, Medium Soil Juice, High Soil Juice, Very High Soil Juice, Very Highly Acidic Reaction, Highly Acidic Reaction, Mild Acidic Reaction, Neutral Reaction, Mild Base Reaction, Highly Base Reaction and Very Highly Base Reaction.
Vector-III (7 dimensions)	Land Type, Water Recession, Drainage Status, Soil components, Soil Rigidity, Soil Juice, Soil Reaction

Table 1.2: Feature Set III

Area Code	Land Type	Water Recession	Drainage	Soil	Rigidity	Soil juice	Reaction
Mun-Lou-Beigao1	2	2	4	4	3	2	4
Mun-Lou-Beigao2	2	2	4	4	3	2	3

In Table 1.2 only the present Feature's Sub-features numbers are assigned to an Area code which makes it more compatible for further use.

IV. PROPOSED SYSTEM

Figure 1.1, the proposed system diagram is depicted. The system comprises two parts: i) Soil clustering for determining the soil of common properties and ii) Crops selection for clustered soil. Soil Clustering and Crops selection are discussed in Subsections IV.A and IV.B respectively.

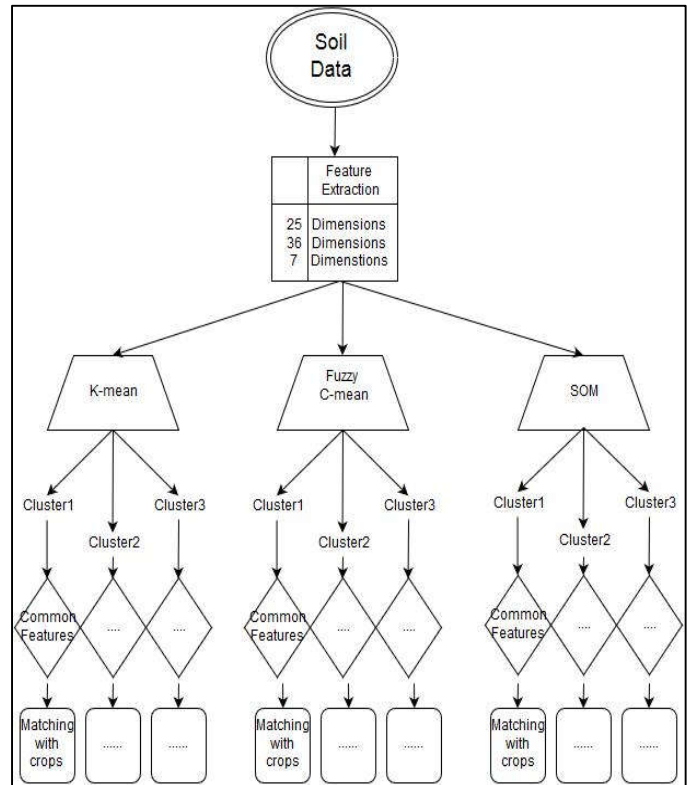


Figure 1.1: Proposed System

IV.A SOIL CLUSTERING

Soil Clustering is done by K-mean Clustering, Fuzzy C-mean Clustering and Self Organizing Map (SOM) Clustering. The descriptions are given in the following Subsections IV.A.1, IV.A.2 and IV.A.3 subsequently.

IV.A.1 K-mean Clustering [3]

For k-mean clustering, three classes were first decided to be classified and hence three centroids were randomly chosen from the output data set which is also addressed as clustered dataset. For each and every feature, a mean were determined. According to the algorithm a new centroid is updated which were then again processed for mean calculation. When the centroid become fixed the mean update process stops with $\epsilon = 10^{-3}$.

IV.A.2 Fuzzy C-mean Clustering [4]

In Fuzzy C-mean clustering the output data set is the clustered data from where also three number of cluster is being determined. A random number need to given to all the clustered data for all the three clusters. But for this the summation of the random numbers of those three clusters must not be greater than 1. For example, let a single data MunLouBejgao1, 3 clusters random numbers were given following 0.6, 0.3, and 0.1. These random numbers were generated by a "random number generation algorithm" which was addressed as initial weight vector. After implementing the random numbers it were being multiplied with all the features and then divided by the summation of random numbers. By this a new weight vector is generated for each data. Subtracting this new weight vector from the previous initial weight vector the absolute weight vector is determined for each cluster. Among these absolute weights the maximum one is being compared with an ϵ (epsilon value) until the maximum value is greater than ϵ . By this the weight is updated. The final weight is denoted with percentage which is the desirable result of this clustering.

IV.A.3 SOM (Self Organizing Map) Clustering [5]

For SOM clustering initially a learning rate (0.6, 0.5, 0.1) randomly generated initial weight and the output data addressed as clustered data set are needed. In this method the initial weight vector is denoted in percentage. Subtracting initial data from clustered data and squaring the result, then summation of that square result are being numbered as the cluster. Among these three clusters the minimum one is selected as the winner cluster. Next the initial weight is subtract from the winner cluster value and multiplied with the learning rate and finally update in the initial weight. This process is done for every cluster until the learning rate ends.

IV.B CROPS SELECTION FOR CLUSTERED SOIL

The output set had a lot of noise. These noises were tried to being reduced so that a common feature of a certain cluster can be determined. For this, one cluster is picked and checked how many features have zeros and how many of them have values. If more than 55% values are being noticed in a sub-feature type then that feature is not being eliminated, otherwise if more than 45% are found zeros then that feature is being omitted. All this were done by using brute force in per cluster. And finally some common features are determined in all the three clusters.

After that a data sets were build for crops data using the same feature extraction which have been discussed in the *section-III Feature Extraction*. The final data sets of Feature Vector-III with 7 dimensions, were furthered matched with crops data to get our desired output of cultivable crops. A linear search is performed with the crops output data set. A threshold 70% value is decided to keep after trying other values.. In the search if at least this threshold value is being observed then that certain crop is announced to be perfectly cultivated in that certain area. This matching were also performed based on clusters too, where a search is done to determine crops for each cluster. This process for each

cluster gives a result about, which certain crops would be good to cultivate on that certain cluster's feature types.

V. EXPERIMENTS

A. Experimental Database

The database is described in the *II.A Corpus* section, about how the database is found and further designed.

B. Experiment Setup

The following methodologies for soil clustering are implemented.

K-mean based:

- I. K-mean + Binary Value
- II. K-mean + Numeric Value
- III. K-mean + Numeric reduced Features

Fuzzy C-mean based:

- IV. Fuzzy C-mean + Binary Value
- V. Fuzzy C-mean + Numeric Value
- VI. Fuzzy C-mean + Numeric reduced Features

Self Organizing Map based:

- VII. SOM + Binary Value
- VIII. SOM + Numeric Value
- IX. SOM + Numeric reduced Features

In K-mean algorithm, $\epsilon = 10^{-3}$ is used for all the three feature sets. In Fuzzy C-mean algorithm, ϵ value 0.011967724983176997, 0.011967724983176997 and 0.01848768188909855 are used for the three feature set respectively. In Self Organizing Map algorithm learning, rate = 0.6, 0.5, 0.1 are used for the three feature sets.

VI. EXPERIMENTAL RESULTS AND ANALYSIS

Figure 1.2 shows that in which algorithm what percentage of data fits by using Feature Vector Type-I. By using K-means algorithm cluster1, cluster2 and cluster3 have the following 35.60%, 29.77% and 34.63%. By using Fuzzy C-mean algorithm cluster1, cluster2 and cluster3 have 26.54%, 44.66% and 28.80% accordingly. And by using SOM cluster1, cluster2 and cluster3 have 1.29%, 69.90% and 28.80% respectively.

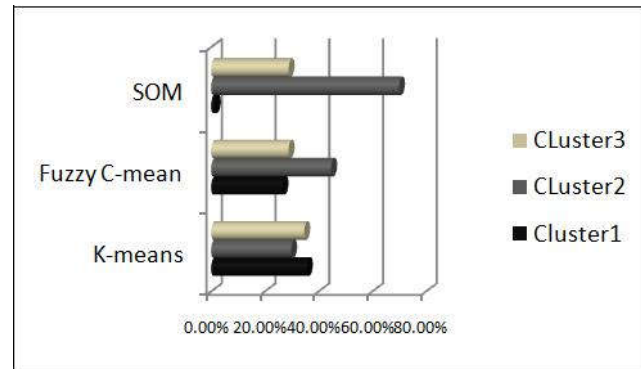


Figure 1.2: Division of 309 feature vectors into three clustering using k-mean, Fuzzy C-Mean and SOM for Feature Set-I binary value

Figure 1.3 shows that in which algorithm what percentage of data fits by using Feature Vector Type-II. By using K-means algorithm cluster1, cluster2 and cluster3 have the following 23.62%, 45.95% and 30.42%. By using Fuzzy C-mean algorithm cluster1, cluster2 and cluster3 have 21.36%, 42.07% and 36.57% accordingly. And by using SOM cluster1, cluster2 and cluster3 have 1.29%, 33.66% and 65.05% respectively.

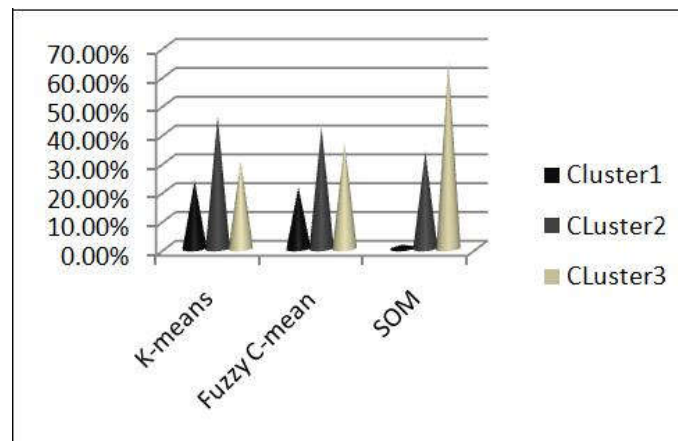


Figure 1.3: Division of 309 feature vectors into three clustering using k-mean, Fuzzy C-Mean and SOM for Feature Set-II numeric value

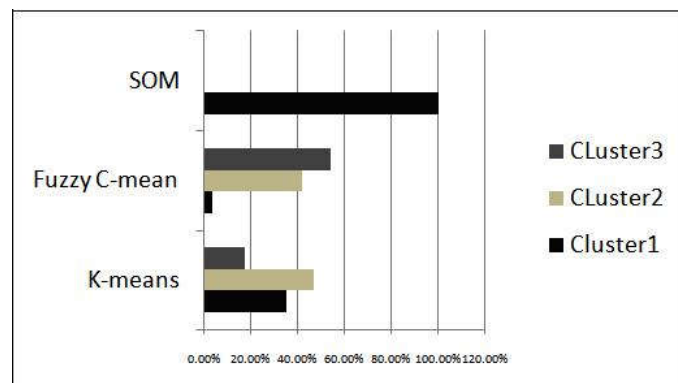


Figure 1.4: Division of 309 feature vectors into three clustering using k-mean, Fuzzy C-Mean and SOM for Feature Set-III numeric reduced value

Figure 1.4 shows that in which algorithm what percentage of data fits by using Feature Vector Type-II. By using K-means algorithm cluster1, cluster2 and cluster3 have the following 35.28%, 47.25%, and 17.48%. By using Fuzzy C-mean algorithm cluster1, cluster2 and cluster3 have 3.88%, 42.07% and 54.05% accordingly. And by using SOM cluster1, cluster2 and cluster3 have 100%, 0% and 0% respectively.

Table 1.3 reflects the common features of soil. Although all the algorithm are used with all the feature sets, but here only the result with K-mean algorithm using Numeric features set (Feature Vector-II) is shown. The other methods also give nearly similar results.

Table 1.3: Common features found by K-mean by using Numeric

Cluster Number	Features
Cluster1	Low Land, Firm Rigidity, Bad Drainage, Normal Water Recession
Cluster2	Low Land, Bad Drainage, Early Water Recession
Cluster3	Medium Juice , Early Water Recession, Clay loamy, Medium Low Land, Bad Drainage

Table 1.4 defines the features cluster wise which is found with matching the crops. Although all the algorithm are used with all the feature sets but the best result is found by using K-mean numeric value feature set (Feature Vector-II).

Table 1.4: Common features after matching with crops that found by K-mean of Numeric value

Cluster1	Cluster2	Cluster3
Low Land, Firm Rigidity, Bad Drainage, Normal Water Recession	Low Land, Bad Drainage, Early Water Recession	Medium Juice , Clay loamy, Medium Low Land, , Early Water Recession

Table 1.5 summarizes the final output of the system using K-mean algorithm. The soil common properties are found while considering the crops which are good for cultivate according to those properties. In the table the crops name are also given. In regarding crops name, with irrigation, without irrigation all are considered. Some name of crops that may be found are not changed and they remain in Bangladesh's Local crops name like: Upasi Aus (a type of paddy), Upasi R.Aman (a type of paddy)

Table 1.5: Final output using K-mean

Cluster	Soil Common Properties after matching with crops	Crops
I	Low Land, Firm Rigidity, Bad Drainage, Normal Water Recession	transplanted with irrigation upasi Aus, transplanted with irrigation upasi R.Aman, Boro with irrigation upasi, transplanted without irrigation upasi Aus, transplanted with irrigation upasi R.Aman.

II	Low Land, Bad Drainage, Early Water Recession	with irrigation chickpea, Local Aus, Boro with irrigation upasi, transplanted without irrigation upasi R.Aman, transplanted with irrigation upasi Aus, without irrigation wheat, without irrigation barley, transplanted without irrigation upasi Aus
III	Medium Juice , Clay loamy, Medium Low Land, Early Water Recession, Bad Drainage	with irrigation chickpea, Local Aus, Boro with irrigation upasi, transplanted without irrigation upasi R.Aman, transplanted with irrigation upasi Aus, with irrigation wheat, with irrigation corn (robi), transplanted with irrigation upasi R.Aman, without irrigation chickpea, Local Aman, without irrigation wheat, without irrigation barley.

Similar type of result like Table 1.5 can be obtained using Fuzzy C-mean algorithm and Self Organizing Map (SOM) algorithm.

VII. CONCLUSION

In this paper the whole process of automated soil clustering for crops cultivation is discussed. Regarding this, the methodologies that are used are K-mean clustering, Fuzzy C-mean clustering and Self Organizing Map based clustering, where K-mean clustering gives the better result. However, Self Organizing Map algorithm gives nearly good result but the

worst case that was being faced was with Fuzzy C-mean algorithm. As per the discussion on Section-I Introduction, there are many more features of soil based on which clustering may be performed. Nevertheless, only these three clustering algorithms can be denoted as the applicable one. So, considering many more features a better result can be performed in future. On the contrary, this paper work is done only for Louhgonj Sub District, so further the research can be done on broader perspective. The Authors will also like to do some research work by merging some clustering algorithms to obtain a new approach in near future.

REFERENCES

- [1] Anderberg, M.R., "Cluster Analysis for Applications," Academic Press, New York, 1973, pp. 162-163.
- [2] Charu C. Aggarwal and Chandan K. Reddy, "Data Clustering, Algorithms and Applications," CRC Press, (2014)
- [3] Hartigan, J. A.; Wong, M. A. (1979). "Algorithm AS 136: A K-Means Clustering Algorithm". *Journal of the Royal Statistical Society, Series C* **28** (1): 100–108
- [4] Ahmed, Mohamed N.; Yamany, Sameh M.; Mohamed, Nevin; Farag, Aly A.; Moriarty, Thomas (2002). "A Modified Fuzzy C-Means Algorithm for Bias Field Estimation and Segmentation of MRI Data," *IEEE Transactions on Medical Imaging* **21** (3): 193–199
- [5] Kohonen, Teuvo (2005). "Intro to SOM," *SOM Toolbox*. Retrieved 2006-06-18
- [6] Elma Hot ,Vesna Popović-Bugarin, "Soil Data Clustering By Using K-means and Fuzzy K-means Algorithm," *Telecommunications Forum Telfor (TELFOR)*, 2015 23rd pages- 890 - 893
- [7] Ping Han, Jihua Wang, Zhihong Ma, Anxiang Lu, Miao Gao, and Ligang Pan, " Application of Fuzzy Clustering Analysis in Classification of Soil in Qinghai and Heilongjiang of China," Beijing Research Center for Agrifood Testing and Farmland Monitoring, Beijing 100097, P.R. China .
- [8] Food and Agriculture Organization of United Nations (<http://www.fao.org/soils-portal/soil-survey/soil-classification/en/>)
- [9] Natural Resources Conservation Service Soils (<http://www.nrcs.usda.gov/wps/portal/nrcs/main/soils/survey/class/>)
- [10] R Isbell, " The Australian Soil Classification," National Committee on Soil and Terrain (NCST), 1996
- [11] Land and Soil Resources Database for Grass-Root Agriculture Development in Bangladesh by S.M Imamul Huq and A.F.M Manzurul Hoque (Senior Scientific Officer, Soil Resource Development Institute, SRDI, Dhaka).
- [12] Hari Eswaran, Thomas Rice, Robert Ahrens, Bobby A. Stewart, "Soil Classification- A global desk reference," December 26, 2002 by CRC Press, 280 Pages
- [13] Resource Development Institute, SRDI, Dhaka, "Land and Soil Resources Utilization Guide," (Upazila Nirdeshika) District: Munshigonj, Sub-District: Louhgonj.

Online Multi-thread Polling Algorithm with Predicted Window Size for DBA in Long Reach PON

Rehnuma Bushra, Monir Hossen, and Mohammad Mahbubur Rahman

Department of Electronics and Communication Engineering
Khulna University of Engineering & Technology
Khulna, Bangladesh

Abstract— Passive optical network (PON) has become the vital technology for the next generation access networks. The advancement of optical technology extends the span of the PON from 20 to 100 km. Such an extended PON is known as Long-Reach PON (LR-PON). But the data transmission in the existing LR-PONs suffer from longer propagation delays between the optical line terminal (OLT) and optical network units (ONUs) that degrade the performance of the traditional dynamic bandwidth allocation (DBA) algorithms in the upstream bandwidth allocation of the LR-PONs. This challenge is the main reason to propose a new algorithm which can reduce the effect of longer propagation delay, i.e., due to larger round trip time (RTT), between the transmission of Report messages from the ONUs and reception of Gate messages from the OLT in the LR-PONs. This proposed algorithm allocates some additional bandwidth based on the prediction during the RTT between the OLT and ONUs if the requested bandwidth is less than the maximum bandwidth of the ONUs. The proposed algorithm is a modified version of the existing online multi-thread polling algorithm. The proposed prediction based online multi-thread polling algorithm reduces the end-to-end packet delay and increases the bandwidth utilization and throughput significantly than any other centralized DBA algorithms which are proposed for the LR-PONs.

Keywords—LR-PONs; online polling algorithm; Delay; Bandwidth utilization; Throughput.

I. INTRODUCTION

A remarkable escalation in internet traffic has been perceived from the last two decades. As a result, bandwidth demand has been turned into a significant issue. To mitigate this challenge, the changes that had been applied in the access network system were not satisfactory enough. The backbone network in the communication system is still an issue to be settled, basically used for access network which makes a link between the central offices (COs) of any service provider and the end terminals. Usually, all of these modern access networks are based on optical fiber due to its bandwidth efficiency [1].

To meet the demand of higher bandwidth in access network system, digital subscriber line (DSL) technologies were set up by different companies [2]. They had deployed fiber on a limited length, i.e., about few kms, within their own premises. Since, all these existing network structures were not completely fiber based; they were not that much effective to fulfill the required demand of huge end terminals.

To ensure the demand of high data traffic, the operators started with the ideas like fiber to the building (FTTB) and fiber to the home (FTTH) solutions. These new solutions set up fibers from their Cos to the end subscriber's premises and can provide a data rate about Gigabits per second (Gbps).

All these passive optical access systems do not consume any power between the CO and end users because passive optical components, e.g., splitters, combiners and couplers, are used. That is why, these access networks are called passive optical networks (PONs).

With the extensive deployment of the short range PONs [2], current research attentions have shifted to increase the fiber distance that is longer reaches and higher split ratios which arises the idea of Long-Reach PON (LR-PON). The LR-PON is very effective because it reduces the requirement of the number of COs or optical line terminals (OLTs). Because the LR-PON combines multiple COs of the short range PONs that is also more effective in terms of cost for the broadband optical access networks [2]. The LR-PON can extend the coverage of a PON from the traditional 20 km to 100 km. This LR-PON can maintain the power budget while capable of increasing the split ratio from 64 to 1024 or more.

The practical deployment of the LR-PON becomes possible at the cost of many research challenges. One of the main challenges is dynamic bandwidth allocation (DBA) algorithm having longer round trip time (RTT) delays between the CO and optical network units (ONUs). The traditional DBA algorithms designed for short range PONs cannot provide the satisfactory performance with the LR-PON. In centralized DBA algorithms of traditional PONs, the OLT distribute the upstream channel into the ONUs by time division multiple access principle. For accurate timeslot assignments, OLT needs to know the exact buffer state of a given ONU. Usually, polling schemes have been used for this purpose, which are based on two control messages, i.e., Report and Gate. The centralized DBA algorithm includes two coexisting processes, i.e., determination of the timeslot sizes (grant sizing) and determination of slot order (grant scheduling). These two processes can be carried out in either offline or online polling. In the offline polling, the OLT first collect all the Report messages from the entire ONUs in the network and then allocate the granted window sizes to the entire ONUs based on their demands. However, using a simple offline centralized polling algorithm will result in an idle time

between collecting all the Report messages and sending the first Gate message [1].

Since the RTTs are larger in the LR-PON it will increase the ideal time and reduces the bandwidth utilization. However, online multi-thread polling (MTP) algorithm reduces the effects of large RTT moderately. In this research, utilization of this longer waiting time in bandwidth allocation and minimizing the effects of longer RTT in the LR-PON are our main target. Therefore, in this paper, a new DBA algorithm for the LR-PON is proposed which is a modification of the existing online MTP algorithm. The proposed scheme also allocates some extra bandwidth to the lightly loaded ONUs in addition to the requested bandwidth of each lightly loaded ONU. This extra allocated bandwidth is calculated by the prediction of data traffic arrival in the queue of an ONU during the RTT. The performances of the proposed scheme have been evaluated by using computer simulations. We also have compared the performances of the proposed scheme with the existing online and offline MTP schemes. Our proposed online MTP scheme effectively reduces the end-to-end packet delay, increases the bandwidth utilization (BWU) and throughput than those of the existing DBA schemes.

The next section contains the details of our proposed online MTP scheme with the analysis of important performance parameters. Section III presents the simulation environment and results. Finally, the paper is concluded in section IV.

II. PRINCIPLE OF THE PROPOSED ONLINE MULTI THREAD POLLING ALGORITHM

In this section, we provide detail explanation of the existing and proposed online MTP schemes and the parameter analysis of the proposed scheme.

A. Existing Online MTP Algorithm

Similar to the offline MTP algorithm [3], in the online MTP the OLT also launches multiple threads in a time cycle. Fig. 1 represents data transmission principle of the existing online MTP algorithm. Here, the solid arrows in the upward direction denote the Report messages from the ONUs to the OLT and dashed arrows in the downward direction denote the Gate messages from the OLT to ONUs in a time cycle z . For simplicity, here only two threads in the upstream direction for N ONUs and a single OLT are considered.

In the online MTP algorithm, at the beginning of thread 1 the ONU 1 sends a Report message to the OLT. Unlike the offline MTP algorithm, as soon as the OLT receives the Report message, it sends the Gate message to the ONU 1 according to the requested window size without waiting for the Report messages of the ONUs 2 to N . Similarly, when OLT receives the Report message of ONU 2, it sends Gate message instantaneously. But the granted window size is pre-defined and that must not be greater than the maximum window size. Here, the length of the maximum window size W^{max} depends on the length of a time cycle, i.e., cycle time T_{cycle} , and number of ONUs in the network, i.e., $W^{max} = T_{cycle}/N$. After completing the transmission of Gate message of thread 1 to the ONU N in the time cycle z , the OLT receives new Report message and sends Gate message of thread 2 of

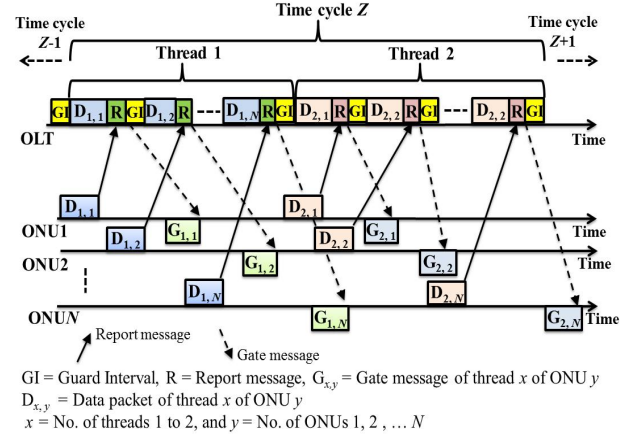


Fig. 1. Data transmission diagram of the existing online MTP algorithm.

the ONU 1, simultaneously, at the same time cycle z . Similarly, the Reports and Gates messages between the OLT and remaining ONUs of thread 2 are exchanged simultaneously. Then in the next time cycle $z+1$, the data transmission processes between the OLT and ONUs are repeated. Due to these simultaneous data transmissions between the OLT and ONUs the online MTP algorithm reduces the overall packet delay significantly than the offline MTP algorithm [3].

B. Proposed Online MTP Algorithm

To reduce the effects of longer propagation delay as well as end-to-end packet delay in the LR-PON and to improve the BWU, in this paper, a modification in the existing online MTP algorithm is proposed. In the proposed online MTP algorithm, some predicted window sizes are allocated in addition to the requested window size if the requested window size of any ONU in any time cycle is smaller than the W^{max} , i.e., lightly loaded ONUs. The calculation of the predicted window size of an ONU in a time cycle depends on the RTT of that ONU and the aggregation time requested data packets after the last upstream transmission. Fig. 2 represents scratch idea of the proposed online MTP algorithm with the predicted window sizes of ONUs.

In the figure, solid arrows in the upstream direction and dashed arrows in the downstream direction denote the Report and Gate messages, respectively, between the OLT and ONUs of an LR-PON. In the MTP scheme, the total number of threads is not limited to 2. However, for simplicity, in the proposed online MTP scheme only 2 threads are shown in the figure. As 2 threads are considered, each thread gets the 50% of a cycle time, i.e., $T_{cycle}/2$ and the maximum window size of each thread is $W^{max}/2$, for transmission of the upstream data in a time cycle. Here, same priority is given for both the threads of all the ONUs. In the network, total N ONUs are considered to be connected to an OLT in a random manner.

The proposed online MTP algorithm will bring change in the granted window size of each thread for all the lightly loaded ONUs in the upstream direction of the LR-PON.

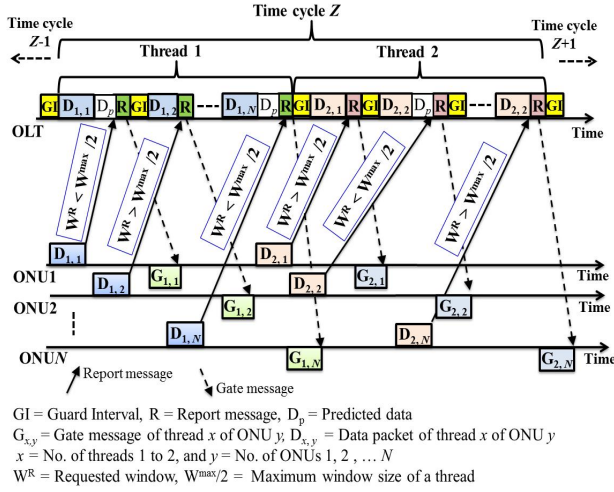


Fig. 2. Data transmission diagram of the proposed online MTP scheme.

Because, in all the existing online MTP schemes, the granted window sizes for all the lightly ONUs in a time cycle are same as the requested window sizes. These entire existing schemes do not consider about the data packets arrival to the queue of an ONU during the longer RTT that is after sending the Report message and before receiving the Gate message of the LR-PON. However, in the proposed scheme, before sending the Gate message to a lightly loaded ONU, the OLT predicts the amount of data packets that come to the queue of an ONU and allocates requested bandwidth plus additional bandwidth that is equal to or less than the predicted bandwidth depending on the remaining bandwidth to the $W^{\max}$ of that ONU. Here, the prediction is done depending on the RTT of the ONU, requested window size, and the aggregation time for the requested window size. If the requested window size of an ONU is greater than the $W^{\max}$ then no predicted window will be allocated to that ONU. This method will reduce the packet delay, as some of the data packets on the next time cycle are transmitted with the data packets of the present time cycle by the means of the predicted window size.

Fig. 3 shows the upstream time frame format based on the proposed online MTP scheme of this paper. Here, the predicted window W^{pred} is added to the requested window size $W^R_{1,1}$ of thread 1 of ONU 1 as the W^R is less than the $W^{\max}/2$ while no W^{pred} is allocated to the W^R of thread 1 of ONU N. Based on the same algorithm of thread 1 the predicted windows are allocated to all the ONUs of thread 2. These predicted window size allocation scheme does not affect the maximum length of a time cycle $T_{\text{cycle_max}}$.

C. Parameter Analysis and Bandwidth Allocation Principle of the Proposed Scheme

Performance of the LR-PON system depends on the efficient bandwidth allocation and longer RTT management for the ONUs in a time cycle. Here, the bandwidth allocation means granted window size W^{grant} to an ONU by the OLT in a time cycle. In the proposed scheme, the granted window size calculation of an ONU i in a time cycle z is shown in (1).

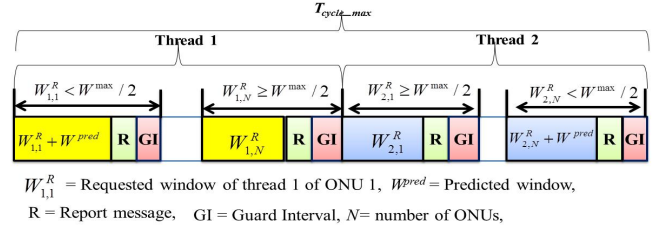


Fig. 3. Upstream time frame format of the proposed online MTP scheme.

$$W^{\text{grant}}_i = \begin{cases} W^R_i + W^{\text{pred}}_i & \text{if } W^R_i < W^{\max}/2 \\ W^{\max}/2 & \text{if } W^R_i \geq W^{\max}/2 \end{cases} \quad (1)$$

here, W^R_i and W^{pred}_i are the requested and predicted window sizes of ONU i , respectively.

The calculation principle of predicted window size W^{pred}_i of the ONU i is shown in (2).

$$W^{\text{pred}}_i = \frac{RTT \times W^R_i}{T_{\text{ag}}} \quad (2)$$

where, T_{ag} is the data aggregation time for W^R_i in a time cycle.

Packet delay calculation: In the proposed online MTP scheme, the average packet delay of the LR-PON can be calculated as follow:

$$D_T = \frac{1}{N} \left(N(T_{\text{EO}} + T_{\text{GI}}) + \sum_{i=1}^N \left(\frac{W^{\text{grant}}_i}{R_u} + T_C \right) \right) \quad (3)$$

here, T_{EO} , T_C and T_{GI} are the time delay for Ethernet overhead, congestion, and guard interval, respectively, R_u is the data transfer rate of the LR-PON.

Calculation of bandwidth utilization (BWU): The BWU is the ratio of the total grants of N ONUs without overhead to the total grants of N ONUs with overhead in a time cycle. Following equation presents the calculation of BWU in the proposed scheme:

$$BWU = \frac{W^{\text{grant}}_T}{W^{\text{grant}}_T + OH_T} \quad (4)$$

here, W^{grant}_T is the total granted windows of N ONUs in a time cycle of the LR-PON, OH_T is the total overhead in a time cycle.

Calculation of throughput: Throughput is the rate of successful message delivery in a time cycle over a communication channel. (5) shows the calculation of throughput.

$$Th = \frac{W^{\text{grant}}_T}{W^R_T} \quad (5)$$

here, Th is the throughput, and W_T^R is the total requested windows of N ONUs in a time cycle.

III. PERFORMANCE ANALYSIS AND SIMULATION RESULTS

This section consists of the simulation scenarios and the performance of the proposed online MTP algorithm and its comparison with the existing online and offline MTP algorithms.

To analyze the performances of the proposed scheme laboratory made computer programs were used. We also considered a tree topology based LR-PON with 1 OLT and 16 ONUs. A highly bursty self-similar network traffic model [1, 8] was used to generate the data packets for the simulation purposes. As the LR-PON is considered here, the physical distances between the ONUs and the OLT were considered in the range of 80 – 100 km. In our analysis, we have considered non-uniform offered loads in the range 0 to 1.0 with different time cycles of 1.0–3.5 ms. To perform the simulation an ideal environment was considered, i.e., no packet loss, and infinity buffer size [9]. The simulation parameters are listed in the following Table I.

TABLE I. SIMULATION PARAMETERS

Quantity	Values
No. of ONUs	16
Channel capacity for both upstream and downstream	1 Gbps
Ethernet overhead	576 bits
Report message	304 bits
Packet length	1500 Bytes
Distance between OLT and ONUs	80 – 100 km
Guard interval	5 μ s
Maximum cycle times	1.0 – 3.5 ms
Maximum number of packets	10

The end-to-end packet delay is an important parameter in any network. One of the main goals of our proposed scheme was to reduce this end-to-end packet delay by allocating predicted bandwidth in addition to the requested bandwidth of the lightly loaded ONUs. Fig. 4 shows the comparison of the end-to-end packet delay vs offered load between the proposed online MTP and existing online and offline MTP algorithms for 2 ms cycle time. Though the packet delay increases with offered loads for all the three cases, the comparison graph shows that the delay for higher offered load in the proposed online MTP scheme is lower than both the online and offline MTP schemes. This is because in the upstream direction of the proposed scheme, the predicted window size is allocated for the lightly loaded ONUs so that some of the next cycle data is transmitted in the present cycle.

In Figs. 5(a) and 5(b), the contour plots show the end-to-end packet delays for different offered loads and cycle times for the proposed and existing online MTP schemes,

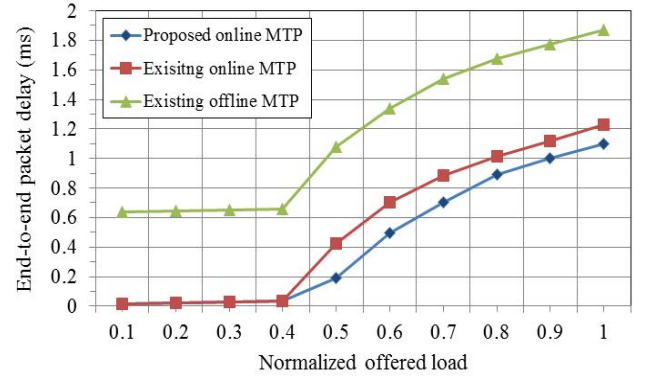


Fig. 4. Comparison of end-to-end packet delay for 2ms cycle time.

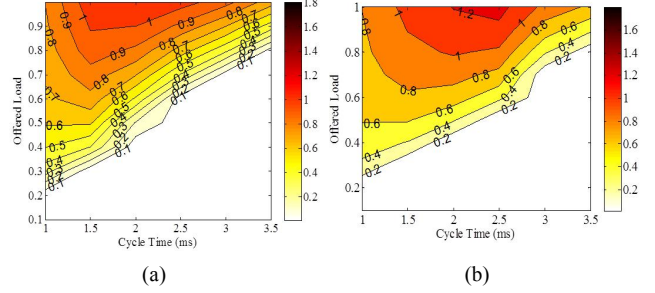


Fig. 5. End-to-end packet delay (a) Proposed online MTP algorithm and (b) Existing online MTP algorithm.

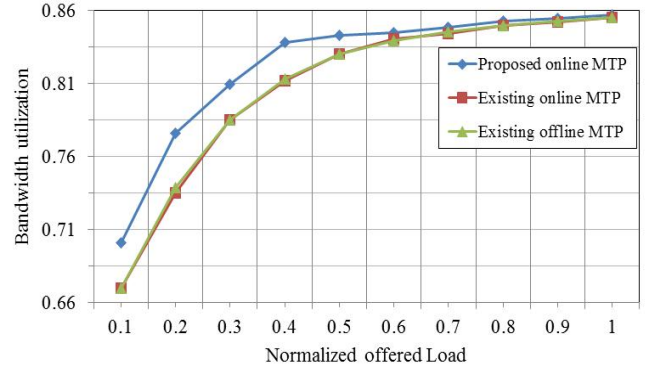


Fig. 6. Comparison of bandwidth utilization for 2 ms cycle time.

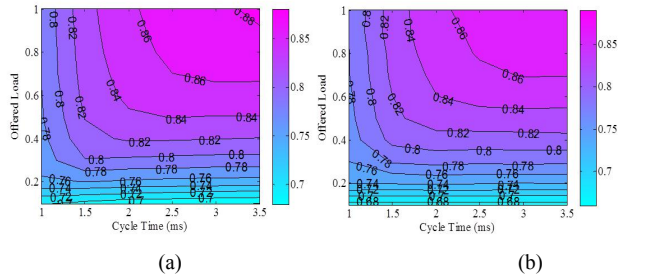


Fig. 7. Bandwidth utilization (a) Proposed online MTP algorithm and (b) Existing online MTP algorithm.

respectively. From the comparison of contour plots, it is clear that the proposed scheme provides less end-to-end packet delay than the existing scheme for any value of cycle time and offered loads.

BWU is the most important parameter to analyze the performance of a DBA algorithm. Fig. 6 demonstrates the comparative results of the BWU among the proposed and existing DBA algorithms. From the figure, it is clear that the BWU for higher offered loads remain approximately same for the entire DBA algorithms discussed in this paper. However, the proposed online MTP scheme provides higher BWU at the lower offered loads.

Figs. 7(a), and 7(b) signify the contour plots of the BWU in terms of different offered loads and cycle times for the proposed and existing schemes, respectively. Due to the usage of predicted window size, the proposed scheme provides better BWU than the existing centralized DBA algorithm. From these two contour plots it is found that the proposed scheme provides higher BWU for entire ranges of cycle times and offered loads.

From the comparison of throughput among the proposed and existing algorithms, shown in Fig. 8, we can say that the proposed scheme provides higher throughput from very lower offered load to the maximum offered load of 1.0. The differences of throughputs among the proposed online MTP and existing schemes are increased with the offered loads. From the comparisons of contour plots in Figs. 9(a), and 9(b) it is also confirmed that the proposed scheme provides better throughput than the existing schemes for any value of offered load and cycle time. At the offered load of 0.8 to 1.0 and cycle time of 3.0 to 3.5 ms the proposed scheme provides 100%

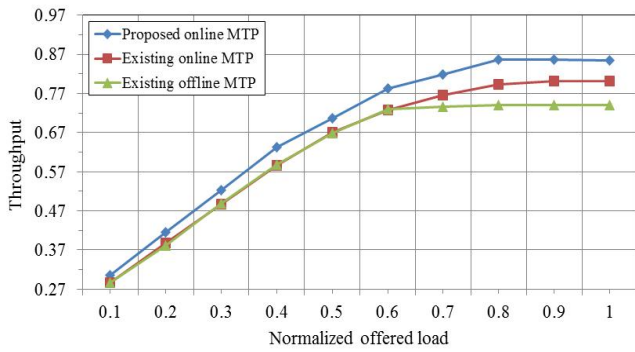


Fig. 8. Comparison of throughput for 2 ms cycle time.

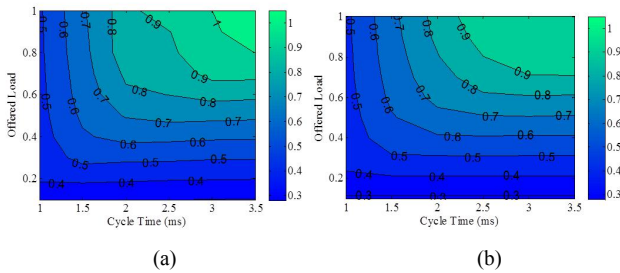


Fig. 9. Throughput (a) Proposed online MTP algorithm and (b) Existing online MTP algorithm.

throughput while the existing online MTP scheme provides 90% throughput.

IV. CONCLUSION

In this paper, we have proposed a new online MTP algorithm with predicted window size for improving the QoSs of the LR-PON system. This paper evaluates and compares the performances of the proposed scheme with the existing online and offline MTP schemes in terms of the end-to-end packet delay, bandwidth utilization and throughput. From the simulation results we can assure that the proposed scheme provides less delay than the existing schemes. The proposed online MTP scheme also provides better bandwidth utilization in the upstream direction for lower offered loads than both the existing online and offline MTP schemes for any length of the time cycle. Moreover, the simulation result proves that the proposed scheme provides higher throughput from very lower offered load to the maximum offered load of 1.0.

REFERENCES

- [1] M. Kamran, "A framework for dynamic bandwidth allocation algorithms in TDM Ethernet passive optical networks" in Proc. International Symposium on High Capacity Optical Networks and Enabling Technologies, Dubai, Nov., 2007.
- [2] A. Helmy, H. Fathallah, and H. Mouftah, "Interleaved polling versus multi-thread polling for bandwidth allocation in Long-Reach PONs", J. Optical Commun. Network, vol. 4, no. 3, Mar. 2012.
- [3] H. Song, B. W. Kim, and B. Mukherjee, "Multi-thread polling: a dynamic bandwidth distribution scheme in long-reach PON," IEEE J. Select. Areas Commun., vol. 27, no. 2, pp. 134-142, Feb. 2009.
- [4] A. Mercian, M. P. McGarry, and M. Reisslein, "Offline and online multi-thread polling in long-reach PONs: A critical evaluation", J. Lightwave Technology, vol. 31, no. 12, Jun. 2013.
- [5] M. Hossen and M. Hanawa, "Network architecture and performance analysis of multi-OLT PON for FTTH and wireless sensor networks", International J. Wireless & Mobile Networks (IJWMN) Vol. 3, No. 6, Dec. 2011.
- [6] G. Kramer, B. Mukherjee, "Interleaved polling with adaptive cycle time (IPACT) a dynamic bandwidth distribution scheme in an optical access network" J. Photonic Network Commun., vol. 4, no. 1, pp. 89-107, Jan. 2002.
- [7] H. Song, B. W. Kim, and B. Mukherjee, "Long-reach optical access networks: A survey of research challenges, demonstrations, and bandwidth assignment mechanisms," IEEE Commun. Surv. Tutorials, vol. 12, no. 1, pp. 112-123, 2010.
- [8] S. Basu, M. Hossen, M. Arefeen, and B. Kundu, "An Efficient Multi-OLT and Multi Wavelengths Passive Optical Network for Differentiated Classes of Services," in Proc. International Conference on Electrical Engineering and Information & Commun. Technology (iCEEICT), Dhaka, Bangladesh, May, 2015.
- [9] M. Hossen, and M. Hanawa, "Adaptive limited dynamic bandwidth allocation scheme to improve bandwidth sharing efficiency in hybrid PON combining FTTH and wireless sensor networks" IEICE Transaction on Commun., Vol. E96-B, no.1, pp.127-134, Jan. 2013.

Electrical Characteristics of Two Dimensionally Modeled InGaN/InN/InGaN-based Double Channel HEMTs (DHEMTs)

Md. Rashedul Haq
Electrical & Electronic
Engineering,
FSET, USTC
Chittagong, Bangladesh
rashedul.ustc.kuet.eee@gmail.com

Riku Chowdhury
Electrical & Electronic
Engineering,
FSET, USTC
Chittagong, Bangladesh
eekeriku@gmail.com

Md. Sarwar Uddin
Chowdhury
Electrical & Electronic
Engineering,
CUET,
Chittagong, Bangladesh
sarwarcueta@gmail.com

Sharmin Afrose
Electrical & Electronic
Engineering,
FSET, USTC
Chittagong, Bangladesh
sharminafrose99@gmail.com

Abstract: In this paper, InGaN/InN/InGaN-based double channel HEMTs are investigated by numerical simulation from the self-consistent solution of Schrodinger-Poisson systems. The influences of polarization charges and quantum effects are incorporated into this model. The simulation model predicts the two-dimensional electrical characteristics (internal electric fields and distribution of charges in the channel) that are associated with specified physical structures and bias conditions through the proper material parameters, the polarization field and surface states. Corresponding analysis and discussion based on the simulation results are subsequently given. This analysis provides a platform to investigate the InN-based double channel HEMTs and to optimize their design.

Keywords: DHEMT; 2DModeling; n-channel; InGaN/InN/InGaN

I. INTRODUCTION :

In recent years, III-Nitride based HEMTs have been gaining more interest among the researchers due to unprecedented power performance capability [1]. Among III-Nitride materials, InN offers the lower effective electron mass, $m_e = 0.04\text{--}0.11m_0$ (in comparison to $0.2m_0$ for GaN) and the higher maximal steady-state drift velocity $v_{max} \sim 5 \times 10^5$ m/s (in comparison to $\sim 3 \times 10^5$ m/s for GaN) [2]. Consequently, assuming $v \sim v_{max}$, InN-based HEMTs may be considered as fast among III-Nitride transistors. Theoretical analysis has shown that, for ideal transistors with gate length (L) $0.1 \mu\text{m}$, by replacing the GaN channel with InN, it is expected that the frequency may increase from 480 to 680 GHz. Moreover, low m_e in the InN channel also gives a prospect for electron non stationary dynamics, and transit frequency f_T over 1 THz has been predicted for $0.1\text{-}\mu\text{m}$ gates in that case. Though InN-based HEMTs possess superior performance in many aspects, they often suffer from low current densities due to the relatively small number of carriers in the channel and spill over into the buffer or barrier layers. This spillover increases low frequency noise and decreases transconductance [3]. To overcome these limitations the sheet carrier density should be increased in the material system. The sheet carrier density can be improved to some extent by increasing the doping in the donor layer at the expense of lower breakdown voltages. Furthermore, a better approach to achieve high current driving capability, distributing the doping into double donor regions by employing double heterojunctions is promising. In this way double 2DEGs are formed and a high current density can be expected [4]. Thus, with combined merits of high power and

high speed, InN-based DHEMTs have high potential in high frequency and high power applications particularly for fabrication of high performance microwave transmitter for high speed devices.

The principal objective of this work is to establish a viable concept for the physical two-dimensional numerical modeling of InGaN/InN/InGaN DHEMTs, to understand the operation mechanism with a focus on main issues improving device performance. In this study, a comprehensive physical 2D model based on the solution of 2D Schrödinger-Poisson-Drift diffusion equations is also developed. We have theoretically analyzed the mechanism of channel formation and the distribution of charges in the channel in InGaN/InN/InGaN-based double heterostructure system. Finally, this 2D model is extended to calculate the dc output characteristics i.e. current-voltage characteristics and transconductance so as to predict the microwave performance of the InN-based DHEMT devices.

II. DEVICE STRUCTURE AND MODEL FORMULATION :

A. DEVICE STRUCTURE

The device model considered for this thesis is shown in Fig. 1. InN has been considered as a channel material. Both sides of InN were sandwiched by InGaN alloy. Due to the lattice mismatch between InGaN/InN and InN/InGaN heterostructures, there will be intentional stress at the interfaces. Electron will be defused to the lower energy InN layer where they are confined due to the energy barrier at the heterointerfaces. The technique of modulation doping is a perfect means of introducing electrons into the InN channel layer without the adverse effects of donors. Even though the electrons and donors are separated spatially, their close proximity allows an electrostatic interaction called Columbic scattering. By setting the donors away from the interface, Columbic scattering by donors can be reduced. It has been shown that increasing the spacer layer thickness enhances the electron mobility by reducing Columbic scattering [5]. However it reduces the 2DEG carrier density as well, which is not desired due to reduction in electron transfer. Therefore, a compromise should be made between the donor density in InGaN, conduction band edge discontinuity (ΔE_c), which is controlled by the In content (x) in the n-InGaN and the thickness of the undoped InGaN spacer layer to maximize both saturation velocity and the electron density of 2DEGs in the InN channel of DHEMTs.

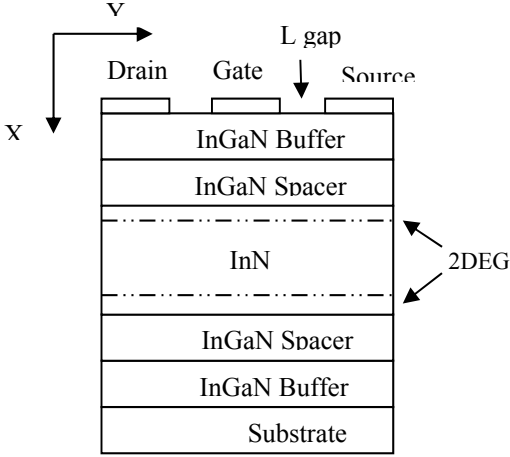


Fig. 1. Schematic of the simulated device

B. TWO DIMENSIONAL MODELING

Schrödinger's equation coupled with Poisson and Charge conservation equations must be solved self-consistently in two dimensional DHEMTs simulation to get an insight into the physical operation and carrier control mechanism of InN based DHEMT.

At the beginning of analysis, it is assumed that all the carriers are attached to their associated dopant atoms, which in fact means that, the net charge density is zero in the whole body of device. As a result, the corresponding electric potential can be obtained from the Poisson equation-

$$\nabla^2 \Phi = -\frac{\rho}{\epsilon} \quad (1)$$

Where, Φ is potential and ρ denotes electrical charge density. The charge density ρ can be written in terms of the electron density n and the total density of ionized dopants, $N_D^+ - N_A^-$.

$$\rho = e [-n + N_D^+ - N_A^-] \quad (2)$$

Schrödinger's equation is solved numerically to determine energy level and electron wave functions, and thereby the electron density.

$$-\frac{\hbar^2}{2m} \frac{d^2 \psi_i}{dx^2} + U_t \psi_i = E_i \psi_i \quad (3)$$

Equation (3) denotes the Schrödinger's equation, where U_t is the potential found from the Poisson solution. Energy level E_i and wave function ψ_i describe electrons in the quantum well. One dimensional (1D) drift-diffusion simulation using MATLAB is used for proposed device.

TABLE I. MATERIAL PARAMETER USED FOR SIMULATION OF DHEMTS

Doping concentration, $N_d=10^{25} \text{ m}^{-3}$	Charge concentration (m^{-3})			
	-0.4 V	2.32 V	2.62 V	2.80 V
Upper channel	-	2.216×10^{24}	8.14×10^{24}	3.233×10^{25}
Lower channel	7.5×10^{23}	*	*	*

III. RESULT AND DISCUSSION:

The conduction band energy and carrier concentration for the quantum well, formed by the double hetero-junction at the barrier and channel interface, for various gate biases based on numerical simulations have been investigated here. Here the doping density is considered to be $N_d=10^{25} \text{ m}^{-3}$ in the InGaN barrier layers. The analysis shows the electron density distribution as a function of gate voltage. In one dimensional simulation at a gate voltage $V_g=-0.4\text{V}$ (near threshold) only the bottom channel is conducting, but also the carriers are confined mainly to the lower interface of the well. The peak charge concentration is found to be $7.5 \times 10^{23} \text{ m}^{-3}$ at this condition.

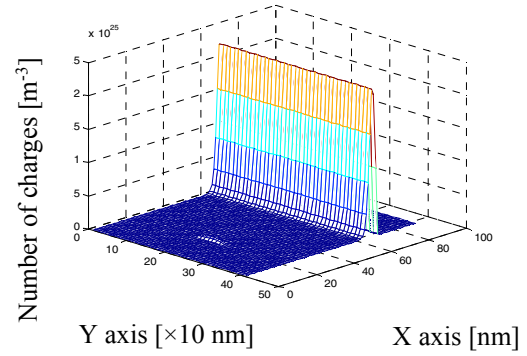


Fig. 2. Total electron distribution for proposed InN-based DHEMT at $V_g=-0.4\text{V}$

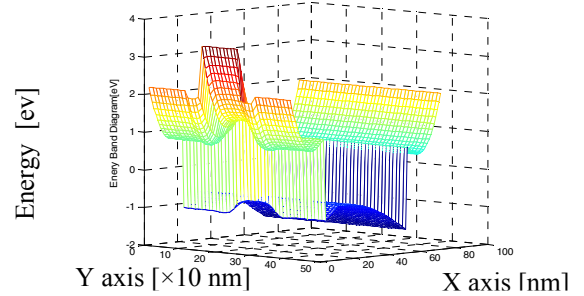


Fig. 3. Energy Band Diagram corresponding to Fig. 2

But in two dimensional simulation at a gate voltage $V_g = -0.4\text{V}$, only bottom channel is conducting and the peak charge concentration is found to $2.216 \times 10^{25} \text{ m}^{-3}$. Fig. 2 and Fig. 3 show the charge distribution. The energy band diagram refers that the lowest point at the upper channel is -0.7508 eV while for lower channel it is -1.531 eV . Drain to source voltage has been kept zero.

The sheet charge density at the bottom channel rises proportionally with V_g until the channel becomes saturated and the upper channel starts to conduct at the same time. The peak charge at second channel is $2.216 \times 10^{25} \text{ m}^{-3}$ at $V_g=2.32\text{V}$ as shown in Fig. 4 and corresponding energy band diagram is shown in Fig. 5. With the increasing gate voltage the upper channel charge develops. The lower channel is saturated at that instant. Fig. 6 and Fig. 7 indicates that, at $V_g=2.62 \text{ V}$ the upper channel peak electron density is $8.148 \times 10^{24} \text{ m}^{-3}$ & The

minimum energy at the upper channel is -1.608 eV respectively.

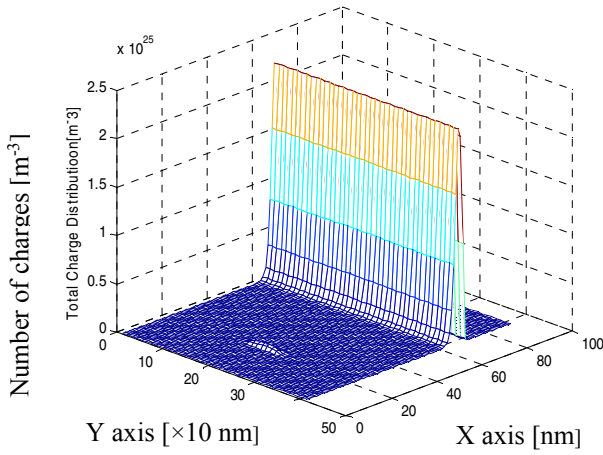


Fig. 4. Total electron distribution for proposed InN-based DHEMT at $V_g=2.32$ V

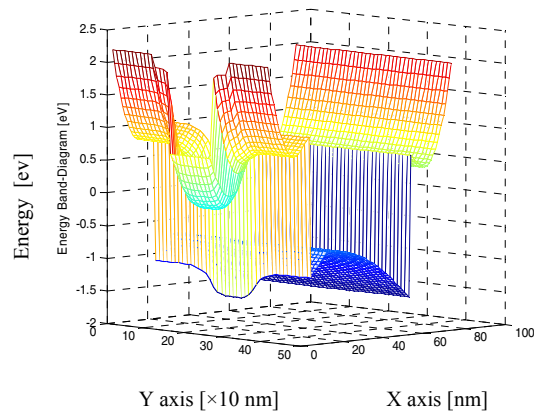


Fig. 5. Energy Band Diagram corresponding to Fig. 4

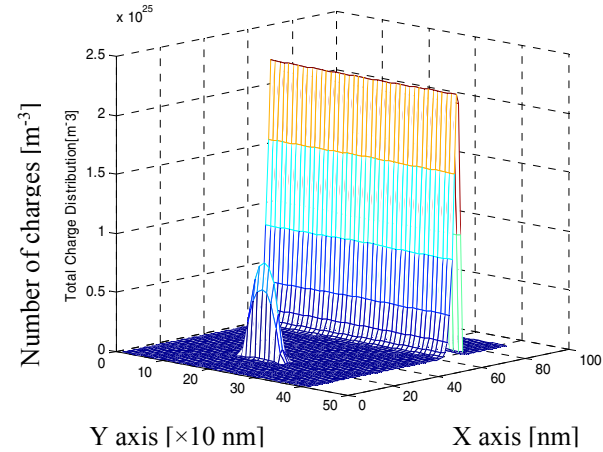


Fig. 6. Total electron distribution for proposed InN-based DHEMT at $V_g=2.62$ V

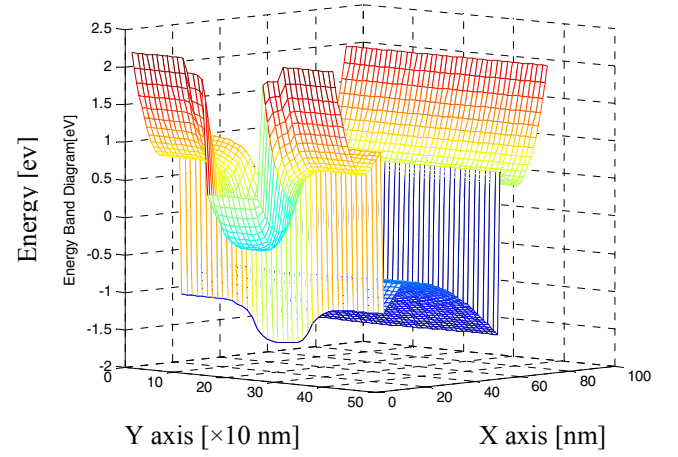


Fig. 7. Energy Band Diagram corresponding to Fig. 6

With further increase in gate voltage, the upper channel also gets saturated. First and Second channel charge peaks at this instant at $V_g=2.80$ V. Peak charge at the first channel is $3.233 \times 10^{25} \text{ m}^{-3}$ which is obvious from Fig. 8. At specified gate voltage the lowest conduction band energy at the upper channel is -1.658 eV as shown in Fig. 9.

Any more increase of gate voltage beyond saturation voltage causes a negative result. More gate voltage will decrease the channel charge density which is not desired region of operation for the device. Fig. 10 and Fig. 11 are the outcome of numerical calculation for total charge distribution and energy level at increased gate voltage, $V_g=3.7$ V. At this condition both the channel get depleted.

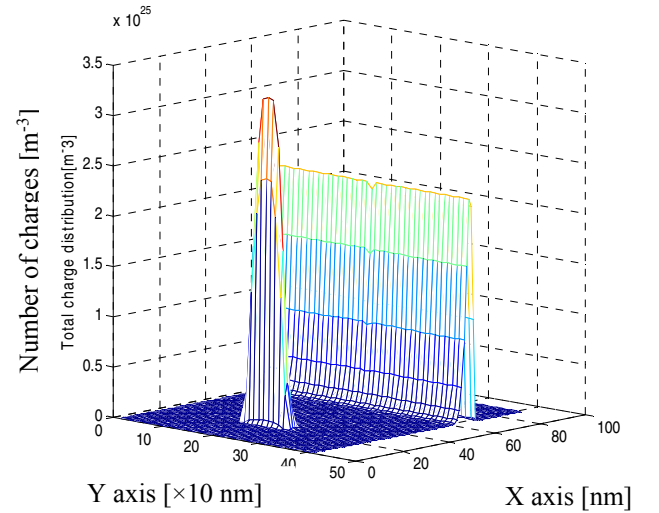


Fig. 8. The total charge distribution at $V_g=2.80$ V

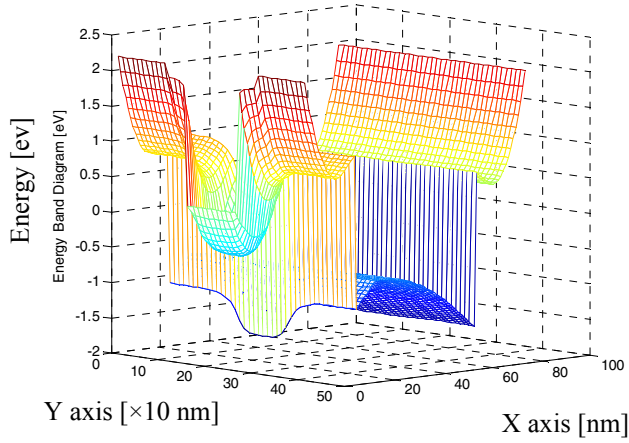


Fig. 9. Energy Band Diagram corresponding to Fig. 8

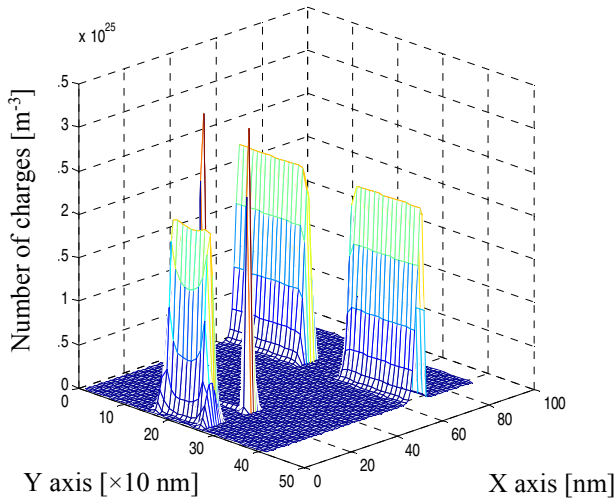


Fig. 10. The total charge distribution $V_g=3.7$ V

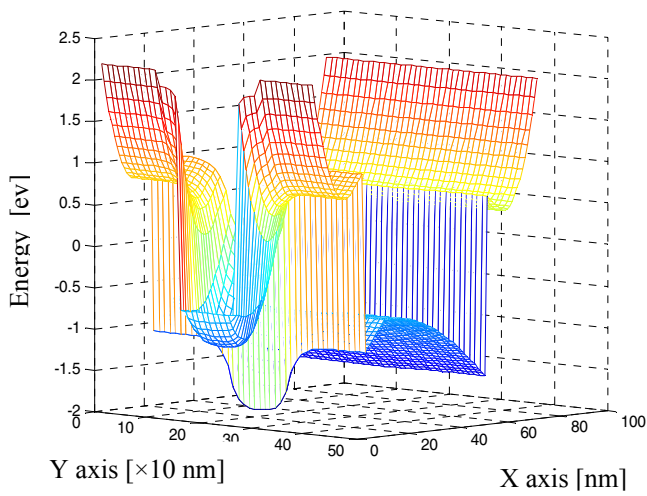


Fig. 11. Energy Band Diagram corresponding to Fig. 10

Source	Gate	Drain
Highly Doped Region	Spacer	Highly Doped Region
	Buffer	
2DEG		
Channel		
Spacer		
Buffer		
Substrate		

Fig. 12. Simulated device with extra doping at first channel

Source	Gate	Drain
Highly Doped Region	Spacer	Highly Doped Region
	Buffer	
2DEG Channel		
2DEG		
Highly Doped Region	Spacer	Highly Doped Region
	Buffer	
Substrate		

Fig. 13. Simulated device with extra doping at both channel

It has been intentionally extra doped with $N_d=10^{25} \text{ m}^{-3}$ up to the 1st channel. The device has been shown in Fig. 12. The drain to source voltage has been kept constant. It is clear that the aim is to get charge reservoir at both sides of the first channel. At this condition, the peak reservoir charge is $3.354 \times 10^{25} \text{ m}^{-3}$. Fig. 14 shows that, the peak 1st channel charge at this instant is $3.228 \times 10^{25} \text{ m}^{-3}$ and peak 2nd channel charge is $2.21 \times 10^{25} \text{ m}^{-3}$.

Now the intentional extra doping with $N_d=10^{25} \text{ m}^{-3}$ in the lower portion likewise the upper portion of the device shown in Fig. 13. Extra doping continues to the lower doped region beyond the second channel (lower). It can be easily predict that, by this time the reservoir will be formed in the two side of the second channel. Fig. 15 indicates that, the second channel reservoir peak charge is $2.815 \times 10^{25} \text{ m}^{-3}$. While the second channel peak charge is $2.21 \times 10^{25} \text{ m}^{-3}$ and the first channel peak charge is $2.874 \times 10^{25} \text{ m}^{-3}$.

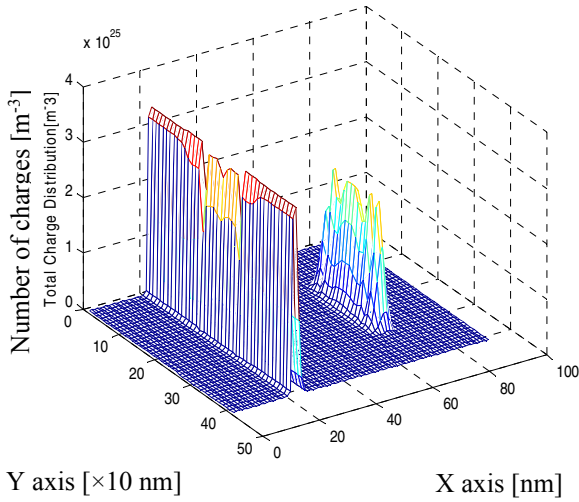


Fig. 14. Total Electron Distribution at $V_g=2.69$ V and $V_{ds}=0$ V

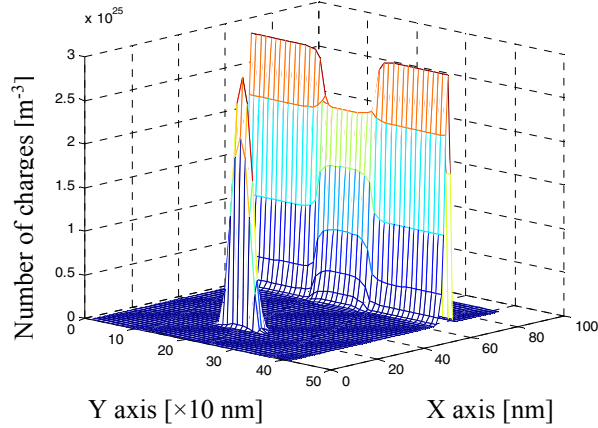


Fig. 15. Total Electron Distribution at $V_g=2.85$ V after extra doping at both channel.

IV. CONCLUSION

In this paper, InN-based DHEMTs have been investigated through theoretical studies, physical modeling and numerical simulations of device operation. A quantum mechanical charge control model of DHEMTs based on InGaN/InN/InGaN double hetero-structure using self-consistent solution of 2D Schrödinger-Poisson equations is developed for the characterization and optimization of the device performances. In order to realize the device performance, a comprehensive quasi-2D model based on the solution of 1D Schrödinger-Poisson drift-diffusion equations is developed.

The transport properties provide an ideal platform for profound investigation of InN-based DHEMTs. Among the transport properties, electron density and mobility of 2DEGs are two most important concerns. As the operation principle of DHEMTs is based on the carrier control of 2DEGs by applying an external gate voltage, an accurate carrier control

characteristics with gate voltage is very much essential. At a gate voltage of -0.4 V, the bottom channel starts to conduct. The sheet charge density at the bottom channel increases with the increase of gate voltage until the bottom channel becomes saturated and simultaneously the top channel starts to conduct. The peak electron concentration at the bottom channel is $2.211 \times 10^{25} \text{ m}^{-3}$. The threshold voltage for upper channel i.e. while the upper channel starts to conduct $V_g=2.32$ V. The sheet charge density at the upper channel also increases with the increase of gate voltage until the upper channel becomes saturated. At $V_g=2.80$ V both the channel charges peak. The peak upper channel electron concentration is $3.233 \times 10^{25} \text{ m}^{-3}$. The proposed device remains conducting between 2.32 V and 2.80 V. Any cumulative increase in gate voltage results decrease in channel charge and the charge starts to accumulate near the gate. At $V_g=3.7$ V, both the channel vanished. The channel performance was improved by forming charge reservoir by extra doping into the device. Intentional doping with $N_d=10^{25} \text{ m}^{-3}$ is applied to the upper channel to get continuous charge distribution. The peak reservoir electron density is found to be $3.354 \times 10^{25} \text{ m}^{-3}$. For same intentional doping from the lower channel to the highly doped lower region the continuous electron distribution will be shifted to the lower channel. The peak electron distribution for the lower channel reservoir is $2.815 \times 10^{25} \text{ m}^{-3}$. It is almost analogous to a water tank. The sheet carrier densities obtained are higher than the conventional InN-based devices [6] and GaN-based double channel HEMTs [8].

REFERENCES

- [1] C. Ostermaier, G. Pozzovivo, J. F. Carlin, B. Basnar and W. Schrenk, "Ultrathin InAlN/AlN Barrier HEMT with High Performance in Normally off Operation," *IEEE electron device letter*, vol. 30, no. 10, October 2009.
- [2] Kuzmik, J. and Georgakilas, "A Proposal of High-Electron Mobility Transistors with Strained InN Channel," Volume 58 Issue 3, *IEEE Transactions on Electron Devices*, 2011.
- [3] Md. Rejvi Kaysir, Md. Tanvir Hasan, Md. Sherajul Islam, Ashraf G. Bhuiyan, Md. Rafiqul Islam, A. Hashimoto and A. Yamamoto, "2DEG properties in InGaN/InN/InGaN-based double channel HEMTs," *Physica Status Solidi (C)* June 10, 2010.
- [4] S. Vitanov, V. Palankovski, and S. Selberherr, "Hydrodynamic Models for GaN-Based HEMTs," *Advanced Material and Device Analysis*, Group Institute for Microelectronics, Technische Universität Wien.
- [5] Md. Rejvi Kaysir and Md. Rafiqul Islam, "Theoretical Charge Control Investigations in InN-Based Quantum Well Double Heterostructure High Electron Mobility Transistors (QW-DHEMTs)," *Advanced Materials Research (AMR) Journal*, November 11, 2011.
- [6] Supriya Datta, "Quantum transport Atom to transistor," *Cambridge University Press*, Jun 16, 2005.
- [7] Hasan, Md. Tanvir, Md. Arafat Hossain, M. Majharul Haque, Ashraf G. Bhuiyan, and Akio Yamamoto, "Effects of cap layer on 2DEGs in InN-based heterostructures," *International*

- [8] Chaturvedi and Nidhi, "Development and study of AlGa_N/Ga_N microwave transistors for high power operation," *Technische Universitat Berlin*, 2008.
- [9] T wang, C H Hsieh and Hsin-chu "Numerical analysis of non-equilibrium electron transport in AlGaAs/InGaAs/GaAs pseudomorphic MODFETs," *IEEE Trans Electron Devices*, vol 9 pp 1930-1038,(2002).
- [10] M.S.Islam, S.M.Muhtadi, Md.T.Hasan, A.G.Bhuiyan,Md.R.Islam, A.Hashimoto and A.Yamamoto, "AlInN/InN metal oxide semiconductor hetero-structure field effect transistor," *Phys. Stat. Sol.(C)*,vol.7,no.7/8,pp.1983-1987,Jul. (2010).
- [11] C.Q.Chen,J.P.Zhang, V. Adivarahan, A.Koudymov, H. Fatima, G. Simin, J. Yang and M. Asif Khan, "AlGa_N/Ga_N/AlGa_N double hetero-structure for high power field effect transistors," *Applied Physics Lett*,vol,82,no 25, 23 June, (2003).
- [12] H. Benedickter, E. Fe;tin and J. F. Carlin, "205-GHz (Al,In)N/Ga_N HEMTs," *IEEE electron letters*, vol. 31, no. 9, September 2010.
- [13] B.E Foutz, S.K.OLeary, M.S. Shur and L. F. Eastman, "Transient electron transport in wurtzite Ga_N,InN and AlN," *J.Appl.Phys*, vol. 85, no. 11, pp. 7727- 7734, Jun (1999).

Performance Analysis of a Memristor-based Hybrid Memory Cell with Rapid Bidirectional Storage Capability

Mohammad Nazmus Sakib, Rakibul Hassan and Satyen Biswas

Department of Electrical and Electronic Engineering

Ahsanullah University of Science and Technology

Dhaka, Bangladesh

shuvo.sakib@gmail.com, rakib_h@y7mail.com, sbiswas@linuxmail.org

Abstract— Technology scaling has been aggressively developed during last several years and almost close to the final states. In order to cope with high density new technology, silicon based memory cell also needs to be replaced by alternate devices. Memristor is one of the promising novel elements for memory cell. This paper proposes a new memristor based hybrid memory cell, which is capable of bidirectional data storage. The present model requires less number of transistors than other recently proposed model. Simulation results of the proposed circuit demonstrate superior competency in terms of read and write time and power consumption as compared with other models.

Keywords—*Memristance; Precharge; Mobility; Non-volatility; Memory cell.*

I. INTRODUCTION

Conventional memory technologies have reached its zenith as the scaling of CMOS technology has been constricting in nano scale regime [1-3] to achieve higher density. The end of the scaling paradigm is imminent. It will be pragmatic to say that CMOS technology will fail to meet Moore scaling in near future. To overcome the obsession of shrinking of transistors, significant emphasis is given by the researchers in search for a new infinitesimal device despite having highly capable. While searching for the supplant of CMOS technology in future non-volatile arrays researchers have integrated different technologies such as carbon nanotube field effect transistors (CNTFET) [4], metamaterials based optical nanocircuits [5], spintronics [6], and one of the most auspicious “the memristor” [7] in recent times. Some fascinating features of memristor like nonlinearity at the boundaries, nonvolatility and high packing density in crossbar arrays have attracted the researchers to gain extensive attention towards it. By keeping in mind about the limited endurance of the flash memory, the focus of this research is given to develop a hybrid memristor based rapid bidirectional memory cell.

The memory cell proposed in this research is composed of only five NMOS transistors, two PMOS transistors and a passive two terminal memristor which can hold data in terms of resistance. Less number of transistors than [8-10] and memristor than [11] are required to perform rapid bidirectional storing operation in the proposed hybrid memory

architecture. As well as proposing the design following contributions are also done in this research.

- Several memristor models and their limitations have been introduced and a comparison of data storing time is also provided for various memristor models.
- Effect of mobility and R_{OFF}/R_{ON} ratio of the memristor model on storing operation is also observed in this research.
- We have varied the feature size of the transistors and W/L ratio of the access transistors to observe the performance of the proposed cell while storing binary data.

The following section provides a review of memristor. Operation of the proposed circuit is described in section III. Simulation results are shown in section IV while in section V comparison between the proposed memory cell and other memristor based memory cells as well as flash memories is provided. Section VI provides the concluding remarks.

II. BACKGROUND

Based on the circuit theoretical reasonings Chua [12] postulated the existence of memristor which is considered as the so called fourth fundamental circuit element. The missing static relationship between flux and electric charge is shown by the memristor which makes the two terminal passive device canonical in nature [13]. This canonical nature makes memristor to be the fourth fundamental circuit element. The relationship between flux and electric charge of the memristor having memristance M is given by [7]

$$d\phi = M dq \quad (1)$$

where ϕ and q are the flux and electric charge respectively.

As described in the physical model of memristor by Strukov et al. [7] there is a proportional relationship between the memristance and electric charge which is passed through the memristor. The memristor is mainly a resistance with memory which does not change even if the supply voltage is removed. The physical model of memristor shown in Fig. 1 is composed of a two layer of TiO_2 thin film which is sandwiched

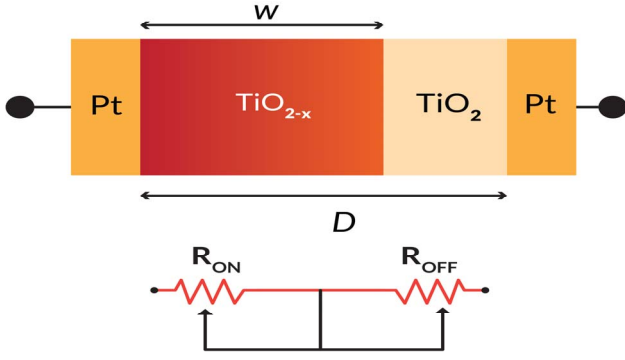


Fig. 1. Physical memristor model

between two platinum nanowires. State of the memristor can be changed according to the direction of voltage across or current through the device. R_{ON} and R_{OFF} state can be achieved by forward and reverse biasing the memristor respectively. If a positive voltage is applied to the doped region (TiO_{2-x}) of the memristor, the boundary of the doped region shifts right to the memristor (see Fig. 1) resulting in lower resistance (R_{ON} state) as oxygen vacancies drift from the doped (TiO_{2-x}) to the undoped (TiO_2) region. Higher resistance state (R_{OFF} state) can be achieved by applying positive voltage to the undoped region (TiO_2).

The total resistance of the memristor is given by

$$M(t) = \left\{ R_{ON} \frac{w(t)}{D} + R_{OFF} \left(1 - \frac{w(t)}{D} \right) \right\} \quad (2)$$

Where R_{ON} and $w(t)$ are the resistance and width of the completely doped region while R_{OFF} is the resistance of the completely undoped region. The parameter D denotes the thickness of TiO_2 . The width of the doped region with respect to time is given by

$$w(t) = \mu_v \frac{R_{ON}}{D} q(t) \quad (3)$$

where μ_v is the average dopant mobility. From (3) the changing rate of the width of the doped region can be known.

III. PROPOSED MEMORY CELL

The hybrid memory architecture is designed by employing the memristor as a storage element and exploiting the non-volatility of memristor. CMOS-based logic circuitry is employed for data selecting and sensing purpose. The proposed memory cell shown in Fig. 2 is composed of two and five PMOS and NMOS transistors respectively having one memristor to store binary data. Raw data is fed to the memristor through the bit line (BL) and the inverse bit line (BL') of the circuit. Line selection of the proposed cell is served by the word line (W/L). To send the data through bit line (BL) and inverse bit line (BL'), the access transistors T_1 and T_2 are turned ON by setting the word line (W/L) to high (VDD). The biasing of transistor T_3 determines whether WRITE or READ operation is performed in the proposed circuit. The

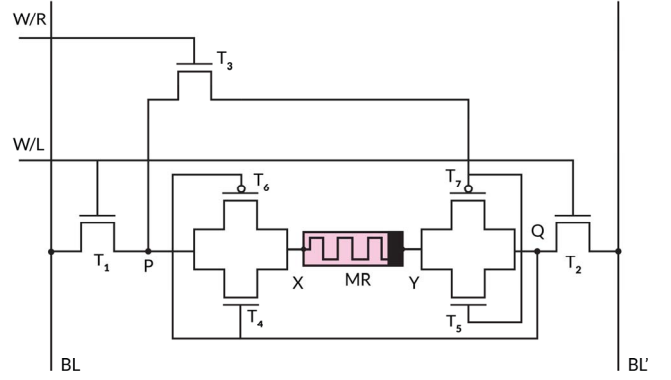


Fig. 2. Proposed memory cell

biasing of the transistor T_3 is done by the W/R line. W/R is set to high (VDD) and low (GND) respectively for data storing and retrieving purpose. When BL is "1" current flows from node P to Q and "0" is stored in the cell as memristor is biased to R_{OFF} state. Conversely, when BL is "0" current flows from node Q to P and "1" is stored in the cell by biasing the memristor to R_{ON} state. This is an evident of bidirectional storage operation performed by the proposed circuit. Detailed explanation of the READ and WRITE operations of the proposed hybrid architecture are carried out next.

A. Write Operation

The necessity of different biasing in the proposed memory circuit during write "0" and "1" operation is explained here. The word line (W/L) and the W/R line is always set to high for both writing "0" and "1". So, the access transistors T_1 , T_2 and T_3 are always ON while performing data storing operation. By taking the time to change the memristance of the memristor to the minimum and maximum value, the write time for "1" and "0" is determined respectively. Fig. 3 shows the variation of memristance of the memristor at different biasing conditions for consecutive write "1" and "0" operations.

1) Write a "0"

The memristance of the memristor has to be maximum for writing a "0" in the proposed design. To get the maximum memristance value BL and BL' are set to high and low respectively. So the transistors T_5 and T_6 are on during this biasing ensuring the flowing of current from node P to node Q. Thus, the memristor will be reverse biased enhancing the boundary of the undoped region over the doped region completely. As a result, the memristor will be in R_{OFF} state having maximum memristance and ensuring the completion of write "0" operation. The time taken to perform a write "0" operation in the proposed circuit is 10.06 ns.

2) Write a "1"

The memristance of the memristor has to be minimum for writing a "1" in the proposed design. To get the minimum memristance value BL and BL' are set to low and high respectively. So the transistors T_4 and T_7 are on during this biasing ensuring the flowing of current from node Q to node P. Thus, the memristor will be forward biased enhancing the

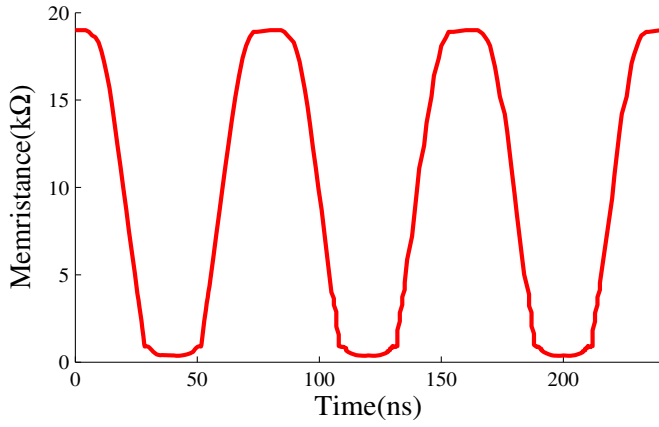


Fig. 3. Memristance versus time for consecutive write “1” and “0”

boundary of the doped region over the undoped region completely. As a result, the memristor will be in R_{ON} state having minimum memristance and ensuring the completion of write “1” operation. The time taken to perform a write “1” operation in the proposed circuit is also 10.06 ns.

B. Read Operation

Likewise the biasing of write “0” operation, the read operation gets started by having the bit line (BL) and inverse bit line (BL') to be precharged to VDD and GND respectively. The word line (W/L) is high likewise the data storing operation. To retrieve the binary data that has already been stored in the memory cell, the circuit proposed here gets isolated from the bitlines after 0.6 ns of the read operation being started. This isolation is done by setting the W/R line to low at 0.6 ns after the starting of read operation. The isolation time of the circuit from the bitlines is 0.6 ns because it is clearly depicting from Fig. 4 that the current through the memristor tends to zero for both read “0” and “1” operation at 0.6 ns while the W/R line is still high. The reason behind setting W/R to low after 0.6 ns of the read operation is the change of memristance during repetitive read operation. If W/R is set to low, the circuit will be completely isolated from the bitlines for every read operation as transistors T_5 and T_7 will not be able to turn ON. As a result, the change of memristance will be futile to change the memristor state (i.e. R_{OFF} to R_{ON}).

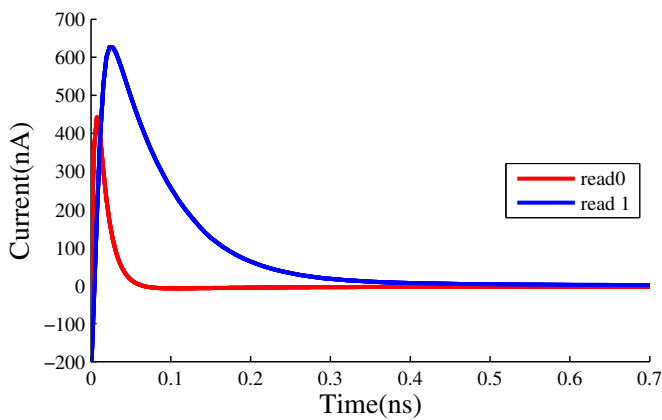


Fig. 4. Current through the memristor versus time during read operation

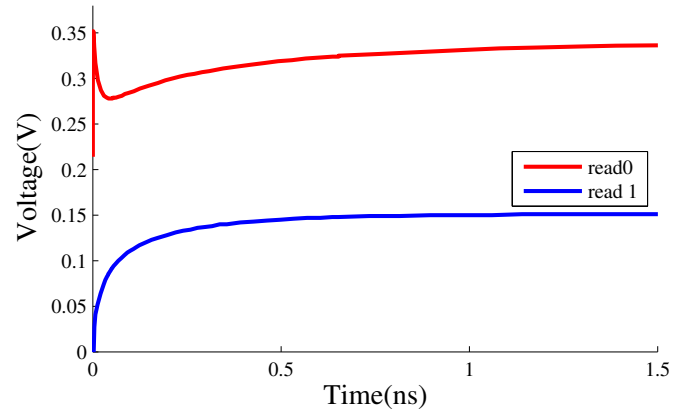


Fig. 5. Bitline voltage difference versus time during read operation

The detection of the binary data that has already been stored is done by the disparate bitline voltage differences. Fig. 5 delineates that the voltage difference between the bitlines for retrieving “0” is superior to the one of retrieving “1”. The reason behind this is

- If a “1” is stored in the memory cell, the memristor is already biased to the R_{ON} state having lower memristance influences the rapid transferring of voltage from the bit line (BL) to the inverse bitline (BL'). The rapid transformation of voltage from BL to BL' results in reducing the voltage difference across the bitlines.
- If a “0” is stored in the memory cell, the memristor is already biased to the R_{OFF} state having higher memristance. This higher memristance impedes the transferring of voltage from BL to BL' causing sluggish abatement of voltage from BL. This sluggish transformation of voltage from BL to BL' is the reason of high voltage difference across the bitlines for retrieving “0”.

The read time of the proposed memory cell is 0.6 ns as the balancing of voltage from the bitlines tends to very low at that period of time. The bitline voltage difference for retrieving “0” and “1” for a 0.6 ns read operations are 323 mV and 147 mV respectively.

IV. SIMULATION RESULTS AND ANALYSIS

To evaluate the feasibility of the proposed memory architecture comprehensive simulation experiments have been carried out by employing LTspice as a simulation tool. The simulation results proposed in the previous section have been found by utilizing the memristor model of [19] having the following parameters: $R_{ON} = 380\Omega$, $R_{OFF} = 19K$, $\mu_v = 450 \times 10^{-9} \text{ m}^2\text{s}^{-1}\text{V}^{-1}$, $x_0 = 0.5$, $a = 50$, $p = 10$. The ratio of the resistance of the undoped and the doped region has been taken 50 to ensure hard switching as indicated in [7]. 32 nm feature size [14] of the transistors have been taken to implement the proposed circuit.

Performance of the proposed memory cell has been analyzed to further evaluate the competency of the proposed rapid bidirectional memory architecture. The change in write time is estimated in this section while changing the W/L ratio of the access transistors and the feature size [13] of the transistors. The variation of time to store binary data is also observed by varying the R_{OFF}/R_{ON} ratio and mobility of the memristor. Furthermore, Different memristor models [15-19] are also employed to examine the variation of time to store data.

If the W/L ratio of the access transistors is increased, the voltage across the memristor increases due to the rapid transfer of voltage to the memristor from the bitline and inverse bitline. The resistance of the NMOS transistors gets lower while increasing the W/L ratio which eventually ended up with the increase in voltage across the memristor (see Fig. 6). It is unequivocal from the Fig. 7 that, write time decreases considerably with the increment of W/L ratio of the access transistors.

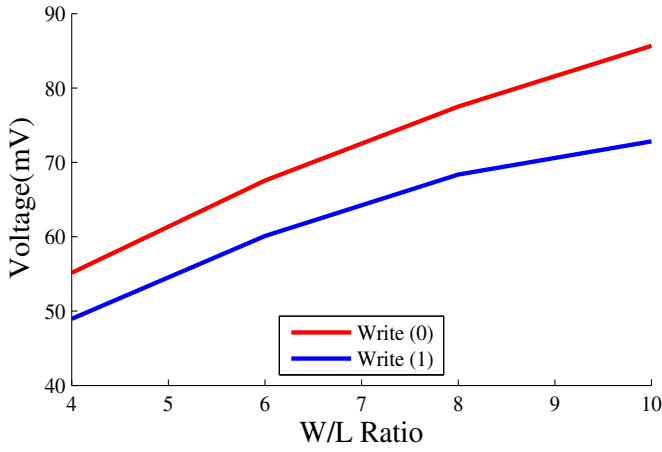


Fig. 6. Voltage across the memristor in different W/L ratio of access transistors.

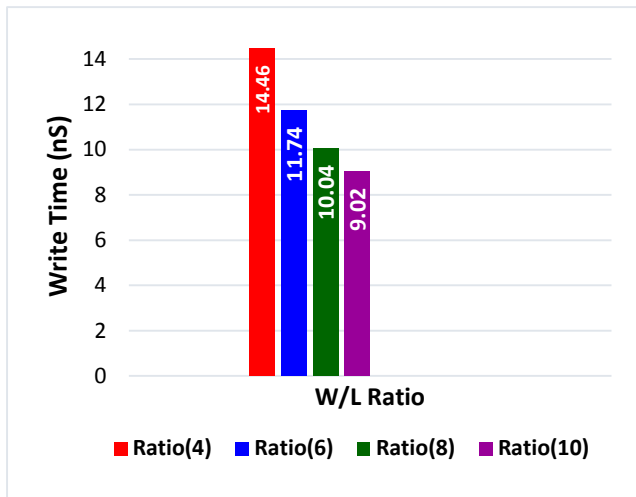


Fig. 7. Write time in different W/L ratio of the access transistors

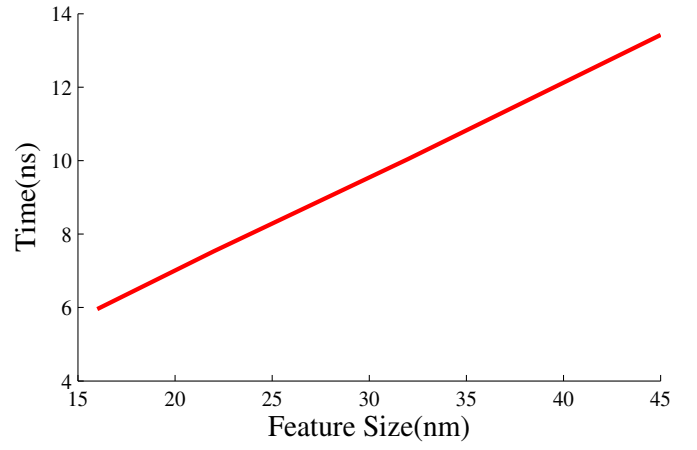


Fig. 8. Write time versus transistor feature size

A. Feature Size

The proposed circuit is also analyzed by varying the CMOS feature size [14]. The circuit has been simulated by taking 16nm, 22nm, 32nm, 45nm fabrication technologies. With the reducing feature size, the voltage across the memristor increases noticeably as voltage drop across the transistor gets lesser. The shifting of boundaries (see Fig. 1) to the left (i.e. R_{OFF}) or right (i.e. R_{ON}) of the memristor becomes significantly faster with the increased voltage across the memristor. It is evident from Fig. 8 that there is a proportional relationship between transistor feature size and write time.

B. Ratio

The variation of write time is considered next while varying the R_{OFF}/R_{ON} ratio of the memristor model of [19]. If the ratio is increased write time is increased noticeably. Emphasis should be given on the ratio of the resistance of the undoped to the doped region of memristor to design a faster memory cell. Fig. 9 demonstrates the change in storing data for different R_{OFF}/R_{ON} ratio of memristor.

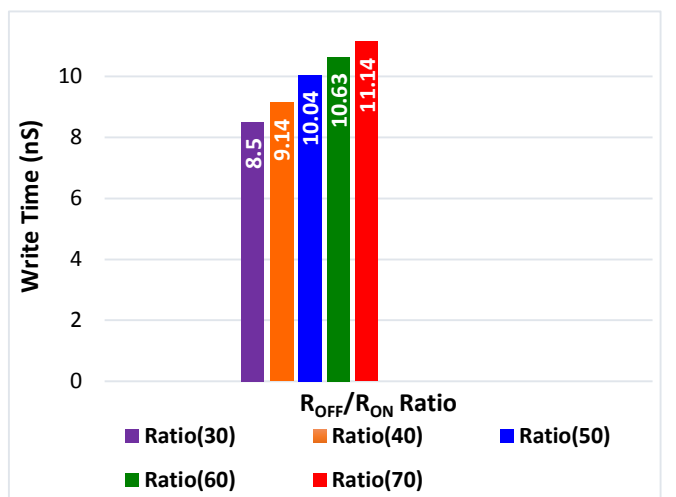


Fig. 9. Write time versus different R_{OFF}/R_{ON} ratio.

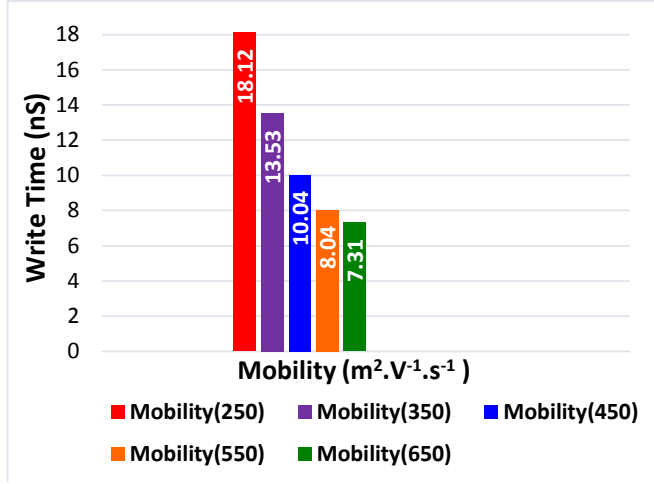


Fig. 10. Write time for different mobility of the memristor.

C. Mobility

Fig. 10 clearly depicts that write time maintains an inverse relationship with mobility. Oxygen vacancies can easily transverse to oxide in higher mobility [20]. Variable width of conductive region can be formed by many parallel conductive paths in higher mobility region. Thus mobility affects the write time.

D. Memristor Model

To investigate the performance of the proposed memory cell several memristor models [15-19] having the same memristor parameters provided previously are also employed and simulation results are shown in Fig. 11. From Fig. 11, it is unequivocal that the memristor model proposed by Jogelkar [17] outperforms other memristor models by taking 7.77 ns to store data. Due to very low scalability and behavior like getting stuck at the terminal states, this model is not taken to propose write timing in section III. Memristor model of [16] suffers same problem. Time required to store data is also less in nonlinear dopant drift model [15]. But discontinuity at the boundary region having limited scalability makes the model vulnerable. Conversely, Prodromakis memristor model [18] solves the terminal state problem as a feedback implementation has been adopted to solve the problem. But noticeable delay in storing data from other window function based memristor models makes the storing operation inefficient. The generic memristor model of [19] does not use window function as nonlinear transforming function is used between the physical and native state variables of the memristor. The performance of the proposed circuit is satisfactory by employing the memristor model [19] which also eradicates the above mentioned problem. This is why extensive simulation is done by employing the memristor model of [19] to evaluate the performance of the proposed cell in this paper.

E. Power Dissipation

Power dissipation during data storing and retrieving operation is estimated by using LTspice. The following formula is used to measure the power dissipation during

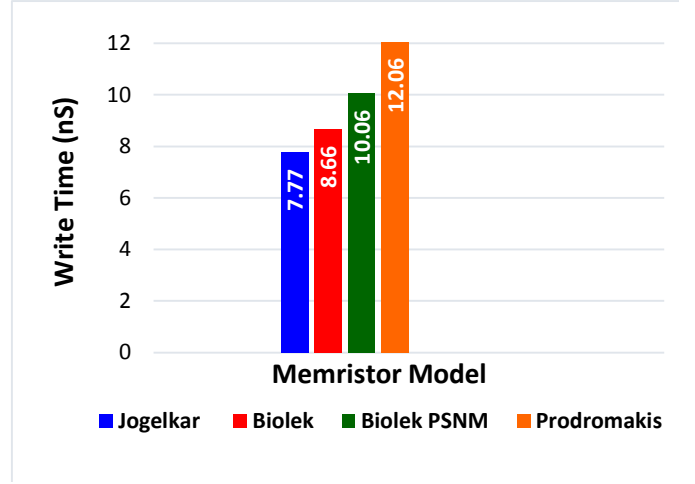


Fig. 11. Write time for different memristor model

TABLE I. POWER DISSIPATIONS DURING DIFFERENT OPERATIONS

Operation	AVERAGE POWER	ENERGY (fJ/cycle)
Write 0	272.18nW	2.7218
Write 1	69.061nW	0.69061
Read 0	2.8808pW	1.7285e-006
Read 1	1.2376pW	7.4259e-007

different operations:

$$P_{av} = \left[\left(\frac{1}{T} \right) \int_0^T Idt \right] \times V \quad (4)$$

Five NMOS transistors and two PMOS transistors along with one memristor are used to design the proposed memory architecture. Power source is not short circuited with the ground in the proposed cell. Furthermore, state change of the memristor is not occurred while there is no supply voltage after storing a “0” or a “1”. This is the reason of significant reduction of power during different memory operations in comparison with the existing memory technology despite the presence of dynamic power dissipation. The proposed memory cell dissipates lower power than the ambipolar based memory cell [10] and the MCAM [11] because of the requirement of less number of transistors and lower supply voltage respectively.(See Table I)

V. COMPARATIVE DISCUSSION AND EVALUATION

In order to evaluate the performance of the proposed memory cell, a comparison between the proposed memory cell and recently developed memristor based memory designs such as ambipolar transistor based memory cell [10] and MCAM [11] is conducted in this section. The comparison is shown in Table II by using the same operating voltage and transistor feature size. Due to less number of transistor requirement of the

TABLE II. COMPARISON OF THE PROPOSED MEMORY CELL WITH OTHER MEMRISTOR BASED CELL [10-11] AT SAME OPERATING VOLTAGE

	Proposed Memory	Ambipolar Based Memory [10]	MCAM [11]
<i>VDD (V)</i>	0.9V	0.9V	0.9V
<i>Write Time</i>	10.04ns	219ns	51.2ns
<i>Read Time</i>	0.6ns	1.095ns	12ns

TABLE III. COMPARISON OF THE PROPOSED MEMORY CELL WITH NAND/NOR FLASH MEMORIES [21-25]

	Proposed cell	NAND Flash [21,22,24]	NOR Flash [21,23,25]
<i>Write Time</i>	10.04ns	1/.1ms	1μm/10ms
<i>Read Time</i>	0.6 ns	.1ms	15ns
<i>Write Voltage</i>	0.9V	15V	10V
<i>Read Voltage</i>	0.9V	1.8V	1.8V

proposed cell, the voltage drop across the transistor and memristor is low and high respectively than other memristor based memory architecture [10]. The memristance of the memristor changes rapidly in the presence of higher voltage across the memristor in the proposed design. This high changing rate of the memristance facilitates the faster storage operation of the proposed circuit.

Data retrieving is done by checking the bitline voltage difference of the stored data in the proposed memory cell. The stored data in MCAM needs to be matched with the search data to discharge the match line voltage. So, state checking is done by search operation in MCAM. No search operation is required in the proposed memory cell which makes the read operation faster than MCAM [11].

Furthermore, performance evaluation is done by comparing the proposed cell with NAND and NOR CMOS-based flash memories [21-25] predicted and manufactured recently. The performance of the proposed cell outstrips the flash memories in every aspect despite having very low supply voltage. (See Table III)

VI. CONCLUSION:

In this proposed memristor based memory model, only 5-NMOS and 2-PMOS transistors are employed, while other models used large number of transistors. This memory cell is non-volatile and also bidirectional in nature for storing data. Due to less number of transistors it is faster and requires less amount of power to perform rapid bidirectional operation. This model can be widely used because of its linearity and scalability.

REFERENCES

- [1] S. Borkar, T. Karnik, and V. De, "Design and reliability challenges in nanometer technologies," in Proc. 41st Annu. Design Autom. Conf., 2004, p. 75.
- [2] C. T. Chuang, S. Mukhopadhyay, J. J. Kim, K. Kim, and R. Rao, "High-performance SRAM in nanoscale CMOS: Design challenges and techniques," in Proc. IEEE Int. Workshop MTD, Dec. 2007, pp. 4–12.
- [3] K. Kuhn, "Considerations for ultimate CMOS scaling," IEEE Trans. Electron Devices, vol. 59, no. 7, pp. 1813–1828, Jul. 2012.
- [4] D. Akinwande, S. Yasuda, B. Paul, S. Fujita, G. Close, and H. S. P. Wong, "Monolithic integration of CMOS VLSI and CNT for hybrid nanotechnology applications," in Proc. 38th Eur. Solid-State Device Research Conf., 2008, pp. 91–94.
- [5] N. Engheta, "Circuit with light at nanoscales: Optical nanocircuits inspired by metamaterials," Science, vol. 317, no. 5845, pp. 1698–1702, Sep. 2007.
- [6] G. I. Bourainoff, P. A. Gargini, and D. E. Nikonov, "Research directions beyond CMOS computing," Solid-State Electron., vol. 51, no. 11–12, pp. 1426–1431, Nov./Dec. 2007.
- [7] D. B. Strukov, G. S. Snider, D. R. Stewart, and R. S. Williams, "The missing memristor found," Nature, vol. 453, pp. 80–83, May 2008.
- [8] P. Junsangsri, F. Lombardi, "A Memristor-Based Memory Cell using Ambipolar Operation" Proc IEEE ICCD 2011, pp. 148-153, Oct 10 2011, MA USA.
- [9] P. Junsangsri, F. Lombardi, "A memristor-based memory cell with no refresh" Proceedings of the 14th IEEE International Conference on Nanotechnology Toronto, pp.947-950, 2014.
- [10] P. Junsangsri, F. Lombardi, "Design of a Hybrid Memory Cell Using Memristance and Ambipolarity" IEEE Trans. Nanotechnology, Vol.12, Issue 1, pp.71-80, Jan 2013.
- [11] K. Eshraghina, K.R. Cho, O. Kavehei, S.K Kang, D. Abbott, S.M.Steve Kang, "Memristor MOS Content Addressable Memory (MCAM): Hybrid Architecture for Future High Performance Search Engines" IEEE Transactions on VLSI Systems, vol. 19, no. 8, pp.1407-1417, Aug 2011.
- [12] L. O. Chua, "Memristor—the missing circuit element," IEEE Trans. Circuit Theory, vol. CT-18, no. 5, pp. 507–519, Sep. 1971.
- [13] L. O. Chua, C. A. Desoer, and E. S. Kuh, Linear and Nonlinear Circuits, McGraw-Hill, 1987.
- [14] <http://ptm.asu.edu/>
- [15] Z. Biolek, D. Biolek, and V. Biolková, "SPICE model of memristor with nonlinear dopant drift," Radioeng. J., vol. 18, no. 2, pp. 211–211, 2009.
- [16] S. Benderli and T. A. Wey, "On SPICE macromodelling of TiO2 memristors," Electron. Lett., vol. 45, no. 7, pp. 377–379, 2009.
- [17] N. Joglekar and S. J. Wolf, "The elusive memristor: Properties of basic electrical circuits," Eur. J. Phys., vol. 30, no. 4, pp. 661–661, 2009.
- [18] PRODROMAKIS, T., PEH, B. P., PAPA VASSILIOU, C., TOUMAZOU, C. A versatile memristor model with nonlinear dopant kinetics. IEEE Transactions on Electron Devices, 2011, vol. 58, no. 9, p. 3099–3105. DOI: 10.1109/TED.2011.2158004
- [19] Z. Biolek, D. Biolek, and V. Biolkov, "Reliable Modeling of Ideal Generic Memristors via State-Space Transformation," Radio eng. J., vol 24, no. 2, p. 393, 2015.
- [20] Chang, T., Jo, S.-H., Kim, K.-H., Sheridan, P., Gaba, S., Lu, W.: Synaptic behaviors and modeling of a metal oxide memristive device. Appl. Phys. A 102(4), 857–863 (2011)
- [21] International Technology Roadmap for Semiconductors (ITRS).(2011).[Online].p.6.Available:www.itrs.net/Links/2011ITRS/2011Chapters/2011ERD.pdf
- [22] International Technology Roadmap for Semiconductors (ITRS). (2011). [Online]. Available: <http://public.itrs.net>
- [23] International Technology Roadmap for Semiconductors (ITRS). (2011).[Online].p.27.Available:<http://www.itrs.net/Links/2011ITRS/2011Chapters/2011SysDrivers.pdf>
- [24] S. Kim and C. H. Lam, "Transition of memory technologies," in Proc VLSI Technol. Syst. Appl., Apr. 2012, pp. 1–3.
- [25] T. Ogura, M. Mihara, Y. Kawajiri, K. Kobayashi, T. Sakaniwa, K. Nishikawa, S. Shimizu, S. Shukuri, N. Ajika, and M. Nakashima, "A fast rewritable 90 nm 512 Mb NOR "B4-Flash" memory with 8F2 cellsize," in Proc. Symp. VLSI Circuits, Jun. 2011, pp. 198–199.

Extracting and discriminating selective brain signals in non-invasive manner and using them for controlling a device: a cost-efficient approach to Brain Computer Interface (BCI)

Md. Arif Abdulla Samy

Department of EEE
Primeasia University
Dhaka, Bangladesh
arifabdulla.samy@primeasia.edu.bd

Md. Mirazur Rahman

Department of EEE
Primeasia University
Dhaka, Bangladesh

Tafsir Ahmed Khan

Department of ETE
Dhaka International University
Dhaka, Bangladesh

Abstract—The interface through which a human brain establishes links with external devices is generally called Brain Computer Interface. Although there are some significant amounts of ongoing researches on how an overall efficient BCI can be developed are going on, making a cost-efficient approach while dealing with limitless brain patterns is found to be more challenging. In this work, feasibility of a cheaper but appropriate way of extracting and discriminating of several non-invasive EEG signals and using those for controlling devices such as a wheel chair has been proved. To assist the argument of this project, numerous experimental data has been processed to produce several signals, such as, right turn, moving forward, stop etc. for the wheel chair. In the experiment the above mentioned three signals were well distinguished from each other. A microcontroller has been used for processing the signals collected from the brain and hence sending to the wheel chair controlling motors. Despite the challenges of dealing with very low but noise sensitive brain signals, their limitless patterns, and limited scope of necessary circuitries, this work has opened up the scope of feasibility of BCI technology in practical life with a simpler and easier approach.

Keywords—BCI, EEG, Wheel Chair, Microcontroller.

I. INTRODUCTION

In this work an approach of controlling devices by using Brain Computer Interface (BCI) technology has been demonstrated. This project was proved as successful by extracting some brain signals as responses of body movement of a human subject and controlling a device by using those signals as some selective commands. The signals were extracted by performing non-invasive EEG and hence they were processed through a microcontroller to send as control signals to the motors attached with a wheel chair. In the whole project, extensive analysis has been performed on several EEG responses such as clapping, eye blinking, movement of hands etc. Additionally, the responses from a conscious and a relaxed mind were also investigated. Extraction of some

selective and well distinguished patterns from hand movement was a success to control a wheel chair in an accurate manner. Although only few brain patterns were considered despite having limitless patterns, this project exhibited a very simple and cost efficient approach. Implementing an automated system was not the primary goal of this work. The primary concern was to prove the feasibility of applicability of BCI in practical arena.

This kind of research is needed to make devices smarter and more compatible with us to make rapid use. This method can be applied to drive wheelchairs for paralyzed person, or can be used to communicate with deaf and mute persons.

II. LITERATURE REVIEW

BCI is one of the newest technologies under research that creates the communication pathway between a brain and a computer. In another sense it is also called Man Machine Interface (MMI) as brain signals are extracted to control machines. The natural process is that a brain following a person's intent sends signal via peripheral nervous system to the analogous muscles. BCI is the process where the brain activity is measured according to a person's intent and those activities are converted to control signals for the external devices. The above concept was demonstrated by Wolpaw et al. where BCI was referred as communicating and controlling device for brain which is accomplished by non-muscular networks. Invasive and non-invasive: there are two ways of how BCI can be performed. The earlier one requires surgery and the process is called Electrocorticogram (ECoG). In comparison, for the later, only some electrodes needed to be placed on the scalp to extract signals by performing the process called Electroencephalography (EEG). The electrical and magnetic activities produced by brain are sensed by those electrodes in the EEG process. A brain can exhibit different modes of activities with difference frequencies such as- delta, theta, beta and gamma wave with the ranges of 0.5-4 Hz, 4-8 Hz, 14-35 Hz and above 35 Hz respectively. EEG signals are also

classified based on those modes [2]. All but Gamma waves are not taken into account for BCI research to control external devices.

In BCI approach producing certain useful brain patterns which are convertible and controllable for certain operations is called mental or experimental strategy. The most common two ways of strategies used in current BCI research are-selective attention and motor imagery. In the case of selective attention, external stimuli, either auditory [3] or somatosensory [4] are required in BCI approach. For the later case, BCI research focuses on detecting a pattern of Event Related Synchronization (ERS). Any movement or imagination of movement of muscles creates or changes activities in brain's cortex region which is called Sensorimotor Rhythm (SMR). This rhythm generates different brain waves. Any increase in brain activity in a specific frequency band is termed as ERS. In BCI research, a control signal for particular movement of a device comes from a particular pattern of ERS after detection.

III. METHODOLOY AND RESULTS

A. Extracting the Brain Wave

For extraction, two wet type passive electrodes have been used which were placed on the fore head of the subject as shown in figure 1 as only two channels were considered in the experiment.

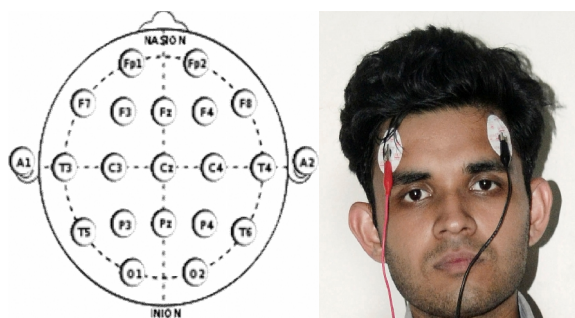


Fig. 1. Acquiring signals: position of electrodes

The necessary steps for extraction of the signals are shown in the figure 2. Before sending it to microcontroller it was needed to amplify and filter the signals as signals are of very low voltage and highly noise sensitive.

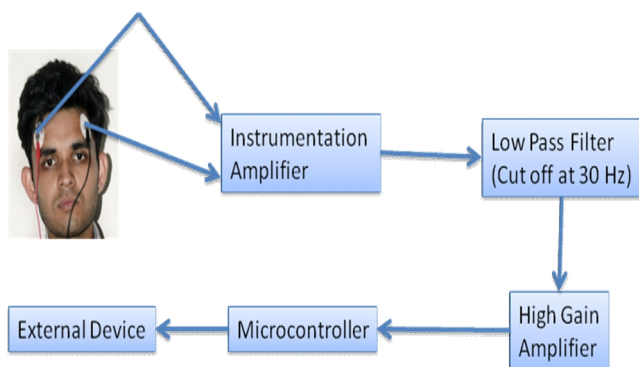


Fig. 2. Brainwave Signal Extraction

After performing some exhaustive trial and error in filtering process a fourth order Chebyshev Low Pass Filter with a cutoff at 30Hz was used. Again, a sharp roll-off -15 dB/Decade was crucial as signals were extremely noise sensitive. The circuit's lay out is shown in figure 3, whereas, gain and phase as responses are shown in figure 4. As the extracted signals were considered to be Alpha (8-13 Hz) and Beta (13-30 Hz) waves, a cutoff of 30 Hz was chosen in the filter.

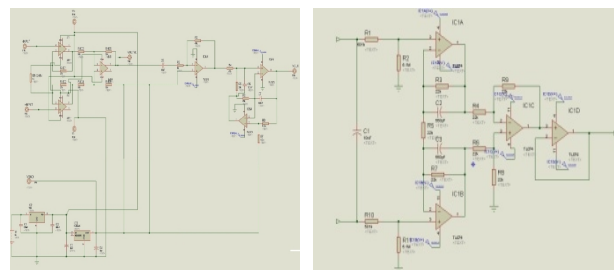


Fig. 3. Amplification and Filtering

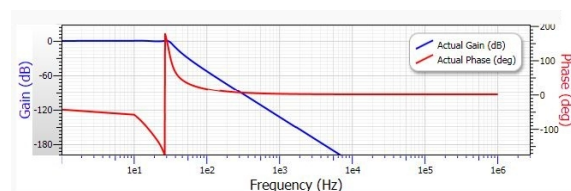
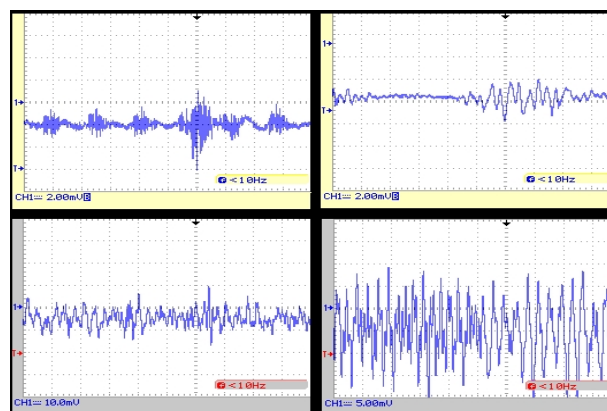


Fig. 4. Phase and Gain vs. Frequency of LPF with cut off at 30 Hz

It is important to note that even the signals are necessarily amplified and filtered, a BK Precision 2530B Digital Oscilloscope was used where all the signals were fed to that before delivering to the microcontroller. The purpose of this was to check and compare those display waveshapes with theoritcal waveshapes and whether the signals have expected harmonis. Figure 5 shows the comparison where upper half shows the experimental waveshapes and lower half shows the theoritcal waveshapes.



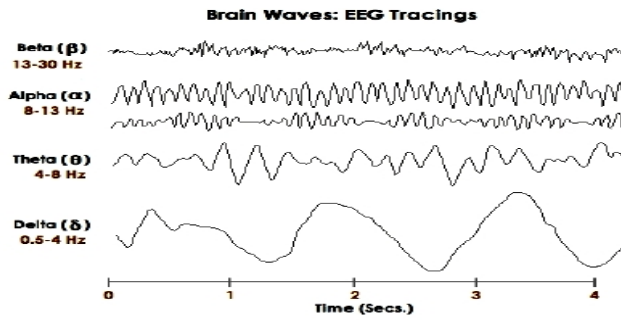


Fig. 5. Comparison between experimental (top) and theoretical results (bottom) [9]

B. LED control with brain wave

In this part of experiment the signals were justified by setting a LED on. It was performed by waking up the subject from relaxed state which exhibited the discrimination between conscious and relaxed states. This is shown in figure 6.

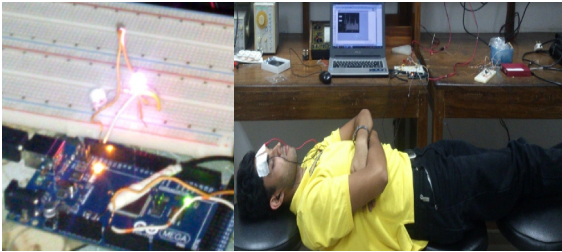


Fig. 6. LED responding to brain wave signal

The wave shapes below (figure 7) show the two states of mind mentioned above. It is noteworthy that alpha wave and beta wave represent relaxed state and conscious state respectively. In the figure below top part shows relaxed state whereas bottom one shows conscious state.

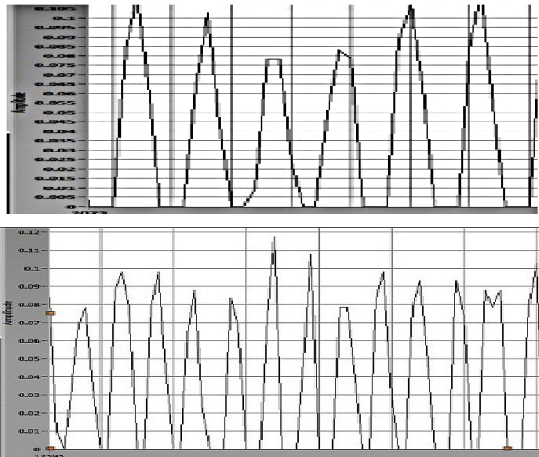


Fig. 7. Two states of mind: Relaxed (top) and Conscious (bottom)

The software named LabVIEW was used for a real-time plot of the subject. A firmware installed for the

microcontroller (Arduino) enabled the microcontroller to be controlled in LabVIEW platform. This microcontroller was then programmed by “G” language for continuous sampling purpose of the signals and for varying the sampling rates (figure 8).

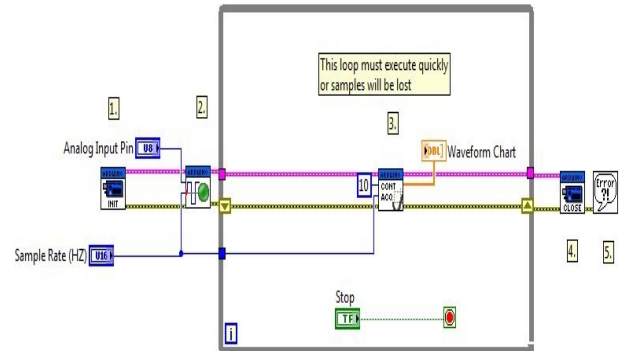


Fig. 8. Continuous Sampling Diagram (Arduino-LabVIEW interface)

LabVIEW enables Arduino to use the plotted signals from into plotting application after reading from its analog port as shown earlier (figure 8). The microcontroller continuously reads 10 samples from the input pin and then the microcontroller is disconnected. Then the loop is started unless it the application stops it. Figure 9 shows how the analog brain signals are read in analog form, converted to digital and then encoded finally.

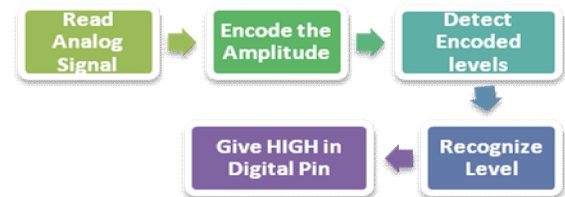


Fig. 9. Controller Block diagram

As the amplitude of the signals are in millivolt ranges, encoding was accomplished by following an algorithm. In this algorithm +5 V was quantized or divided in 1024 levels, which means 0-5 V can be represented in wider range of 0-1024. Thus the output signals were programmed where different states have shown voltage in wide range of levels.

C. Wheelchair control

In this stage the final patterns received after previous stages were used to control the wheel chair which is the ultimate purpose of this project. The challenge here is to drive a large load that draws high current. While looking for the appropriate motors, two DC gear motors have been chosen to drive the wheel chair. Before choosing the motors, high torque and minimum gear requirements were some major factors. The design finally was accomplished where the motors could drive a 100 kg load at 0.5m/s and which required only 12 V and 120 W to rotate at 180rpm. This generic design is shown in figure

10 where arrangements of gears are shown. Efficiency enhancement was the purpose of including the chain.

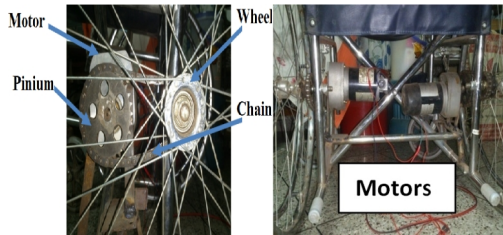


Fig. 10. BCI application: a Wheelchair modified with gears and motors

A relay was added to turn on the motors after processing the signals. Four physical inputs such as no-touch, one finger touch, one hand grip and full hand grip from the users were considered and four different directions were achieved by setting up an algorithm which are shown in figure 11, 12, 13 and 14.

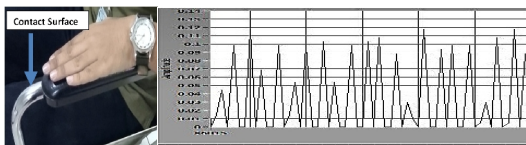


Fig. 11. Hand position: no touch mode and the corresponding wave shape

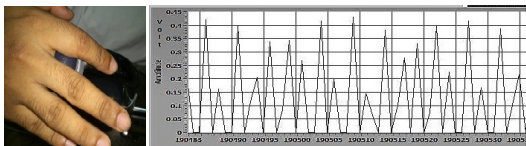


Fig. 1. Hand position: one finger mode and the corresponding wave shape

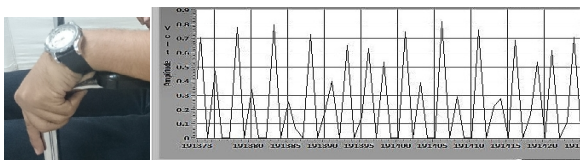


Fig. 2. Hand position: one hand grip mod and corresponding wave shape

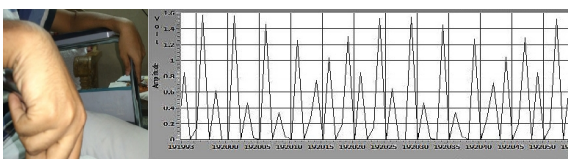


Fig. 3. Hand position: two hands grip mode and corresponding wave shape

It is noteworthy that in the whole design the wheel chair's all metallic parts was considered as the sensitive part, where, the foam pad for resting of the hands was separated from the handle bar.

It should be noted that the resting foam pad of the handle bar is isolated from the wheel chair. All the metallic parts of the wheel chair are considered as sensitive areas in this design.

In the first case (figure 11), the user was not touching the handle bar and a maximum record of 0.12 V was found from the wave shape. This did not provide any motion in the wheel chair.

In the second case (figure 12), user touched the handle bar and a maximum level of 0.45 V was recorded. In this case the level of quantization was one and the signal sent to the motors and made the wheel chair to move forward. The wheel chair stopped as soon as touch was removed.

The third mode (figure 13) exhibited a higher voltage level of 0.8 V with a quantization level of two. Here, the user had a grip on the handle bar. In this case the command received by the motor was to turn right by turning only the left motor on and again it stopped as soon as the grip removed.

In the final mode (figure 14), with a full two hands grip, the voltage level found 1.6 V and it was quantized as level 3. In this case the right motor was set on and hence, the wheel chair turned left.

CONCLUSION

Exploring the feasibility of how BCI can be applied in practical field was the main objective of this study which was performed with the aid of controlling external devices such as wheelchair. The approach was proven as very simple and cost-efficient one with the support of sufficient experimental data and the whole work has created more scopes to investigate BCI applications further. Although, only two channels EEG extraction were used, improvement on circuitries and including more channels may enhance the efficiency of the whole system. Additionally, extensive training on subjects will give a significantly improved outcome. The scopes of this work may be extended to develop a system to control prosthetic limbs, a system for military applications, a thought controlled device, bionic limbs etc. Even the wheelchair designed in this project can be further improved after extracting and discriminating more channels, and it was already suggested by some Japanese scientists that using an EEG sensor cap and a computer program that can process the EEG signals into directional signals a user's thought read wheel chair can be built [5]. Emotiv epoc headset can also be used to read EEG signals, as it claims that it can read both muscle movements and brain waves.

REFERENCES

- [1] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, and T.M. Vaughan, "Braincomputer interfaces for communication and control." Clin Neurophysiol, 113, Jun., 767–791, (2002)
- [2] Bernhard Graimann, Brendan Allison, and Gert Pfurtscheller "Brain-Computer Interfaces: A Gentle-Introduction " The Frontiers Collection, DOI 10.1007/978-3-642-02091-9_1, C _ Springer-Verlag Berlin Heidelberg 2010
- [3] D.S. Klobassa, T.M. Vaughan, P. Brunner, N.E. Schwartz, J.R. Wolpaw, C. Neuper, and E.W. SELLERS, "Toward a high-throughput auditory P300-based brain-computer interface." Clin Neurophysiol, 120, Jul., 1252–1261, (2009).
- [4] G.R. Müller-Putz, R. Scherer, C. Neuper, and G. Pfurtscheller, "Steady-state somatosensory evoked potentials: suitable brain

- signals for brain-computer interfaces" IEEE Trans Neural Syst Rehabil Eng, 14, Mar., 30–37, (2006).
- [5] "RIKEN Brain Science Institute (RIKEN BSI)." RIKEN Brain Science Institute (RIKEN BSI). 07 Apr. 2014 <http://www.brain.riken.jp/en/>
 - [6] J. G. Makin, M.R. Fellows, P.N. Sabes, "Learning Multisensory Integration and Coordinate Transformation via Density Estimation." PLoS Computational Biology, vol. 9, NO. 4, e1003035
 - [7] M.G. Bleichner, J.M. Jansma, Z.V. Freudenburg, E.J. Aarnoutse, M.J. Vansteensel, N.F. Ramsey, "Give me a sign: The possibilities of using hand gestures as control signal for implanted brain computer interfaces", Rudolf Magnus Institute of Neuroscience, Dept. of Neurology and Neurosurgery, University Medical Center Utrecht, NL
 - [8] Y. Hashimoto, T. Ota, M. Mukaino, J. Ushiba "Motor recovery of chronic writer's cramp by brain-computer interface rehabilitation: A pilot study" Department of Biosciences and Informatics, Faculty of Science and Technology, Keio University, Yokohama, Kanagawa, JP
 - [9] M. Shانهچی, R. Hu, M. Powers, G. Wornell, E. Brown, Z. Williams "A concurrent brain-machine interface for sequential motor function", "Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA Massachusetts General Hospital, Boston, MA, USA Harvard Medical School, Boston, MA, USA Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA
 - [10] (2014) National Geographic Website. [Online]. [Cited January 5, 2014]. Available: <http://news.nationalgeographic.com/news/2013/08/130829-mind-brain-control-robot-brainwave-eeeg-3d-printing-music/>
 - [11] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, and T.M. Vaughan, "Braincomputer interfaces for communication and control." Clin Neurophysiol, 113, Jun., 767–791, (2002).
 - [12] J.P. Donoghue, "Connecting cortex to machines: recent advances in brain interfaces." Neurosci. 5 (Suppl), Nov., 1085–1088, (2002).
 - [13] Bernhard Graimann, Brendan Allison, and Gert Pfurtscheller "Brain-Computer Interfaces: A Gentle-Introduction" The Frontiers Collection, DOI 10.1007/978-3-642-02091-9_1, C Springer-Verlag Berlin Heidelberg 2010
 - [14] Anderson, A., III., Wegener, C., Thome, R., Shikha, & Lorenz, M. "Portable Electroencephalogram Biofeedback Device", UW-Madison BME Design, Fall 2005.
 - [15] D.S. Klobassa, T.M. Vaughan, P. Brunner, N.E. Schwartz, J.R. Wolpaw, C. Neuper, and E.W. SELLERS, "Toward a high-throughput auditory P300-based brain-computer interface", Clin Neurophysiol, 120, Jul., 1252–1261, (2009).
 - [16] G.R. Müller-Putz, R. Scherer, C. Neuper, and G. Pfurtscheller, "Steady-state somatosensory evoked potentials: suitable brain signals for brain-computer interfaces", IEEE Trans Neural Syst Rehabil Eng, 14, Mar., 30–37, (2006).
 - [17] A. Schlögl, J. Kronegg, J. Huggins, and S. Mason, "Evaluation criteria for BCI research: Toward brain-computer interfacing", MIT Press, Cambridge, MA, pp. 342, 327, (2007).
 - [18] D.J. McFarland, W.A. Sarnacki, and J.R. Wolpaw, "Brain-computer interface (BCI) operation: optimizing information transfer rates", Biol Psychol, 63, Jul., 237–251, (2003).
 - [19] B. Graimann, B. Allison, C. Mandel, T. Lütth, D. Valbuena, and A. Gräser, "Non-invasive brain computer interfaces for semi-autonomous assistive devices", Robust Intell Syst, 113–138, (2009).
 - [20] R. Leeb, D. Friedman, G.R. Müller-Putz, R. Scherer, M. Slater, and G. Pfurtscheller, "Self-Paced (Asynchronous) BCI control of a wheelchair in virtual environments: A case study with a Tetraplegic", Comput Intell Neurosci, 79642, (2007).
 - [21] C. Cruz-Neira, D. J. Sandin, and T. A. DeFanti, "Surroundscreen projection-based virtual reality: the design and implementation of the CAVE," in Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques, pp. 135–142, Anaheim, Calif, USA, August 1993.
 - [22] Jzau shsheng lin, win ching yang, "Wireless brain computer interface for electric wheelchairs with EEG and eye blinking signals" International journal of innovative computing, information and control, volume 8, number 9, September 2012, ISSN 1349-4198
 - [23] J.D.R. Millán, F. Renkens, J. Mouriño, and W. Gerstner, "Noninvasive brain-actuated control of a mobile robot by human EEG", IEEE Trans Biomed Eng, 51, Jun., 1026–1033, (2004).
 - [24] G. Pfurtscheller, T. Solis-Escalante, R. Ortner, and P. Linortner, "Self-Paced operation of an SSVEP-based orthosis with and without an imagery-based brain switch: A feasibility study towards a Hybrid BCI". IEEE Trans Neural Syst Rehabil Eng, 18(4), Feb., 409–414, (2010).
 - [25] R. Krepki, B. Blankertz, G. Curio, and K. Müller, "The Berlin Brain-Computer Interface (BBCI) – towards a new communication channel for online control in gaming applications", Multimedia Tools Appl, 33, 73–90, (2007).
 - [26] R. Scherer, A. Schloegl, F. Lee, H. Bischof, J. Jansa, and G. Pfurtscheller, "The self-paced Gra
 - [27] (2014) Sinya Electronics Company website. [Online]. [Cited January 5, 2014]. Available: http://sinya.en.seekic.com/product/integrated_circuits_ics/INA118P.html
 - [28] (2014) Unlocking the Secrets of Mind Website. [Online]. [Cited January 5, 2014]. Available: <http://teknomage.wordpress.com/2008/08/15/eeeg-schematic/>
 - [29] G.Rajendra Kumar, Dr.Samuel Vara Prasada Raju, D. Santhosh Kumar, "Classification Of Eeg Signals For Drowsiness Detection In Brain And Computer Interface", ISSN 1512-1232
 - [30] (2014) Active Electrodes for The Modular EEG website. [Online]. [Cited March 10, 2014]. Available: <http://open EEG.sourceforge.net/doc/hw/ae.html>
 - [31] (2014) Ambu website. [Online]. [Cited March 10, 2014]. Available: http://www.ambu.com/corp/products/clinical_studies/ambu%C2%AE_neuroline_eeg_cup.aspx
 - [32] Leo M. J. de Sonnevill, Stephan C. J. Huijbregts, Robert Licht, Joseph A. Sergeant and Francjan J. van Spronsen, "Pre-attentive processing in children with early and continuously-treated PKU. Effects of concurrent Phe level and lifetime dietary control" ISSN 1519-1239
 - [33] Zahir, Ebad, et al. "Device Control Using Brain Waves." International Journal of Research in Computer Engineering & Electronics 3.3 (2014).

Propagation Characteristics of Ultra-Wideband Pulse in Multilayered Human Chest Tissue

Nouroz Rahman Amon,^{1,*}, Ishmam Mahbub¹, and Pran Kanai Saha¹

¹Department of Electrical and Electronic Engineering,
Bangladesh University of Engineering and Technology, Dhaka-1205, Bangladesh
*nouroz_rahman@yahoo.com

Abstract—In this paper two major propagation characteristics such as reflection and absorption of Ultra-wideband (UWB) pulse in a human body have been analysed. A simplified multi-layered non-homogenous model of human tissue is considered here. S-parameters and path loss due to transmission of UWB pulse in the multi-layered tissue model are evaluated. Frequency dependent Electromagnetic properties of three layers (skin, fat and muscle) such as permittivity and conductivity of human tissues based on Cole-Cole model are used for the simulations. Specific Absorption rate (SAR), which is directly related to the internal electric field and the conductivity, is computed at a wide band of frequency ranging between 1-5 GHz.

Keywords—Ultra-wideband, Multilayered Human Tissue, Reflection, Absorption, Path loss, Specific Absorption Rate

I. INTRODUCTION

The ultra-wide band (UWB) using the frequency range from 3.1 to 10.6 GHz is a new promising technology suitable for high rate communications in small distances [1]. UWB is a very promising technology for in-body to on-body communication in medical applications. Two major properties in case of transmission of UWB pulse are reflection and absorption. A good amount of work to study the interaction of electromagnetic waves in human tissues has been done in [2-6]. But most of the work did not consider the frequency dependence of conductivity and relative permittivity. In [7] frequency dependence on EM properties were considered but thickness of skin layer was considered 10mm. Thickness for skin layer generally lies between 1-3mm. So, considering 10mm skin thickness is not realistic. In [7] no air gap between the antennas and model was mentioned. Again, in [8] frequency dependent properties were used and proper tissue modelling was done. But frequencies at which simulations were done were at frequencies 3GHz, 6 GHz and 8 GHz which are spaced far apart. Change in EM properties occurs significantly in short range of frequencies. In [3] only fat-muscle layer was considered and muscle layer was assumed very thick. In [9] S-parameters and SAR were estimated for separate layers but not as a multilayered model. From [2-8] Path loss due to transmission of EM waves was not analysed. In this paper frequency dependent - real and imaginary part of relative permittivity are calculated using Cole-Cole model. Tissue has been modelled anatomically. Realistic air gap between transmitter and tissue has been considered.

Section II describes the EM properties and propagation characteristics. Modelling of human tissue and simulation parameters are given in section III. Simulation results are presented with brief discussion in section IV. The section V concludes this paper.

II. PROPAGATION CHARACTERISTICS OF EM WAVES

In case of transmission of EM waves, both the dielectric property and the conductivity of a medium depend significantly on the frequency of the EM wave [7]. In all the tissues, these properties changes with frequency because of their difference in water content [7]. In a lossy medium:

$$\epsilon_r = \epsilon_r' - i\left(\frac{\sigma}{\epsilon_0\omega}\right) \quad (1)$$

Where, σ is the conductivity of the medium. ϵ_r' is the real part of the permittivity. Conductivity changes with frequency with following equation:

$$\sigma = |\epsilon_0\epsilon_r''| \times \omega \quad (2)$$

Where, ω =angular frequency of the incident signal, ϵ_r'' =complex permittivity in lossy medium and ϵ_0 = permittivity of free space.

Dispersive characteristics of biological tissues like dielectric constant variation with frequency can be expressed by well-known Cole-Cole model as [10]:

$$\epsilon = \epsilon_\infty + \frac{\epsilon_s - \epsilon_\infty}{1 + (i\omega\tau)^{1-\alpha}} \quad (3)$$

Where, ϵ_∞ and ϵ_s are the relative permittivity of material at infinite and zero frequencies respectively. And, α is actually a distribution parameter in the interval $0 \leq \alpha < 1$ [10].

When the transmitted power propagates through tissue medium, part of it is lost because of absorption. And some portion of it is available to the receiver. If S-parameter representation is used, for the two-port system, S_{11} is the input port voltage reflection coefficient. S_{11} is simply the reflection co-efficient of the two port network. However transmission coefficients and hence transmitted power can be estimated from using this also. The amount of power that is reflected back can be calculated from this. Now, reflected power:

$$R = \text{Input power} \times |S_{11}|^2 \quad (4)$$

Path loss is defined by the ratio of transmitted power to received power. It is often expressed in decibels.

$$\left| \frac{1}{S_{21}} \right|^2 = \frac{P_{tx}}{P_{rx}} \quad (5)$$

The equation illustrates the path loss of the tissue model. But more realistic value for path loss can be obtained by the equation:

$$PL = \frac{1 - |S_{11}|^2}{|S_{21}|^2} \quad (6)$$

Values from equation (5) and (6) will be substantially different when S_{11} will be close to unity. Equation (5) gives uncompensated path loss and equation (6) gives compensated values for path loss [11].

Specific absorption rate (SAR) is a measure of the rate at which energy is absorbed by the human body when it gets interacted with electromagnetic field. It is defined as the power absorbed per mass of tissue and has units of watts per kilogram (W/kg) [12]. Simplified formula for calculating SAR is:

$$SAR = \frac{\sigma \times |E|^2}{\rho} \quad (7)$$

Where σ is the conductivity of the medium, E is the RMS electric field and ρ is the density of the medium [12]. So SAR depends both on electric field and conductivity of the medium.

III. MODELLING AND SIMULATION

In order to inspect the propagation characteristics of model of human tissues, proper parameters are needed to be known. Three layers of human tissues: skin, fat and muscle have been modelled here. There is no gap between any two layers of the following model. Values of different parameters for modelling in Simulation tool, for the 3 layers of human tissue, were taken from CST Voxel family data. Thickness of different layers is not same. In this work the thickness has been set assuming that the tissues are from chest tissue of an adult male and thickness of the skin layer is 2mm, fat layer is 10mm and muscle layer is 28 mm which are anatomically right [13]. All the tissues are of equal surface of 20mm×15mm but different thicknesses. All the tissue samples have been modelled as brick shaped cubic box. So the multi-layered human tissue model is ready and looks like this in the simulation 3D view:

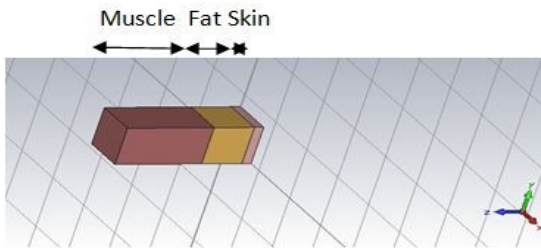


Fig.1: Rectangular box shaped tissue model for simulation.

To simulate the above model, CST microwave studio (CST MWS) has been used. It is mainly based on FDTD (Finite Difference Time Domain) method. Again in some manual it is written as FIT (Finite Integral Technique). But after looking closely the two methods are somehow similar. FIT is actually a Finite difference method based on a small scale interpretation of Maxwell's equations. It is a powerful tool for the 3D electromagnetic (EM) simulation of high frequency components. After properly modelling the multi-layered tissues, it is necessary to set simulation parameters. The simulation of the model has been done from 1GHz to 5 GHz. Gaussian pulse of 2nd order has been used as excitation signal from the transmitter end. Distance between the transmitter and the tissue model was 4mm. Input power from the implanted UWB slot antenna (transmitter) is 0.5 Watts. Mesh type is hexahedral and accuracy is set as -30 decibels. EM wave propagates through negative Z direction. Since the transmitter is implanted, so it meets the muscle layers first, then fat layer and lastly the skin layer. The receiver is also 4mm away from the model.

IV. RESULTS AND DISCUSSION

From FDTD simulation we get several graphs and distribution of 3D fields on 2D plane at different frequencies. These graphs show all S parameters of the two port network. Real and imaginary parts of relative permittivity, which is a strong function of frequency, were estimated using Cole-Cole model as well. 3D SAR distribution on 2D plane at 1GHz-5GHz on a difference of 1GHz are found. Lastly 3D SARs on 2D plane from the simulation post processing at 1 GHz interval are given with reference values.

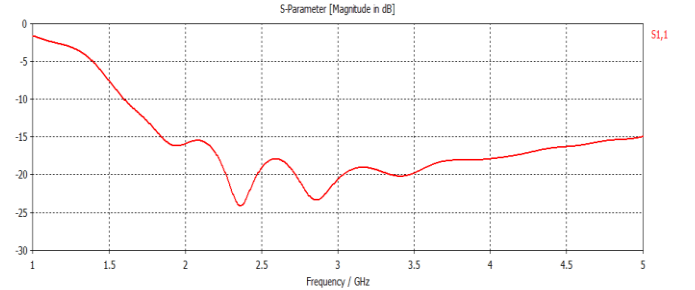


Fig.2: S_{11} vs frequency.

Figure 2 shows S_{11} . -10dB occurs at 1.6GHz which is like UWB slot antenna used in [1]. Impedance bandwidth is here about 68% in this 1-5GHz range. The large available bandwidth covers UTMS (1.9 GHz and 2.1 GHz), WLAN (2.45 GHz) and some parts of UWB (3.1-10.6 GHz) [14]. In [14] only low UWB band ranging from 3.1-5GHz was investigated.

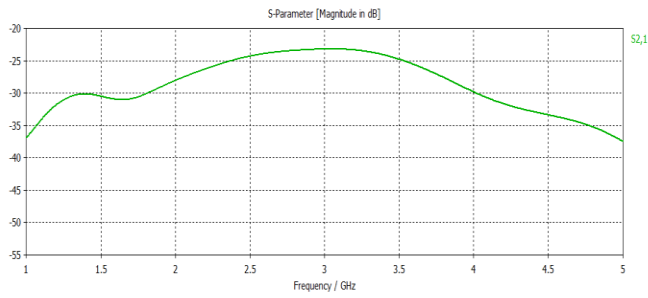


Fig.3: S_{21} vs frequency.

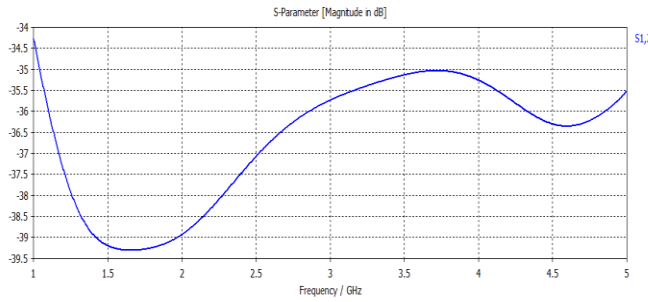


Fig. 4: S_{12} vs frequency.

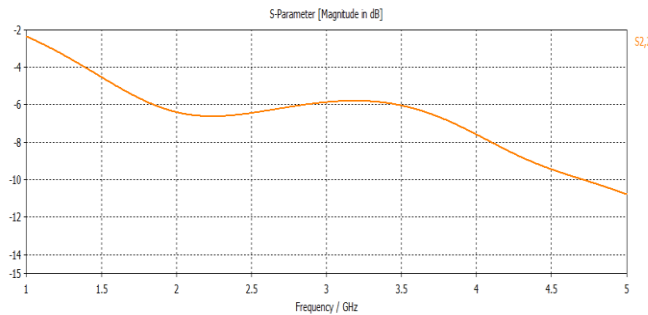


Fig.5: S_{22} vs frequency.

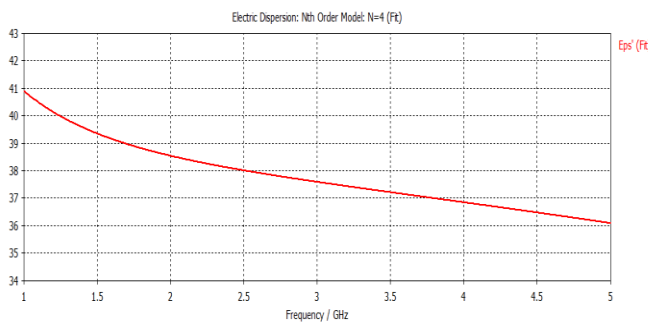


Fig.6: Real part of relative permittivity(at skin layer) vs frequency

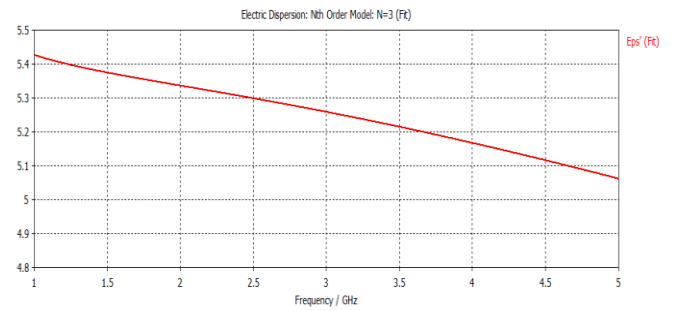


Fig.7: Real part of relative permittivity(at Fat layer) vs frequency.

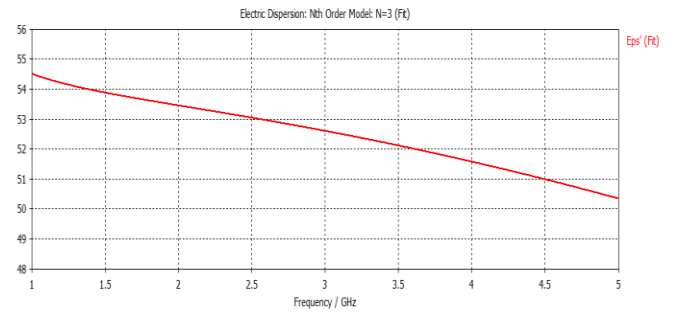


Fig.8: Real part of relative permittivity(at muscle layer) vs frequency.

Change in real part of permittivity is not so significant in all layers but considering this change with frequency ensures best results. In muscle layer the real permittivity is 54.5, 52.7 and 50.3 at 1 GHz, 3 GHz and 5 GHz respectively.

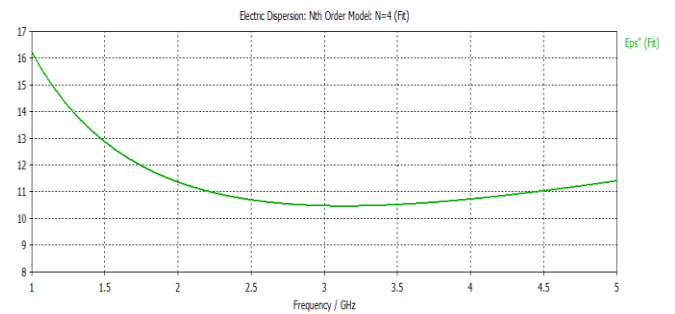


Fig.9: Imaginary part of relative permittivity (at Skin layer) vs frequency.

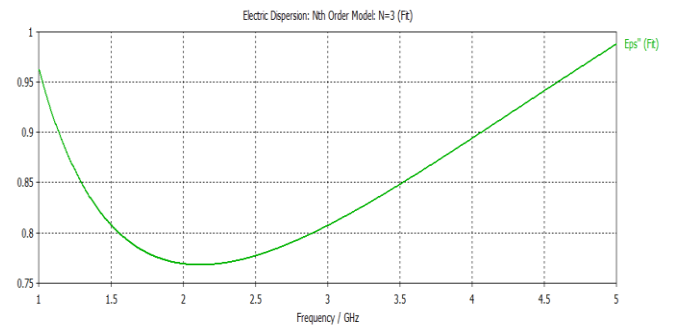


Fig.10: Imaginary part of relative permittivity (at Fat layer) vs frequency

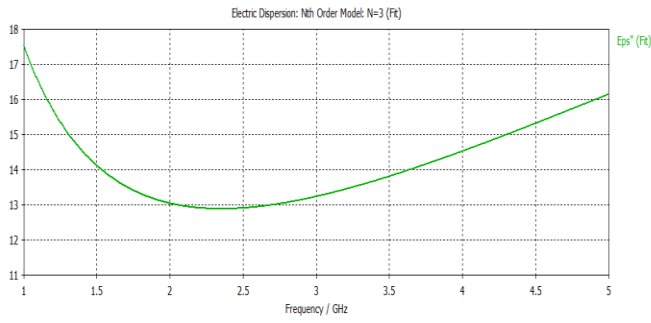


Fig.11: Imaginary part of relative permittivity (at muscle layer) vs frequency

Conductivity of tissues significantly changes with frequency. Using equation (1) conductivity in muscle layer at 1 GHz, 3GHz and 5GHz were 0.97S/m, 2.19S/m and 4.46S/m respectively.

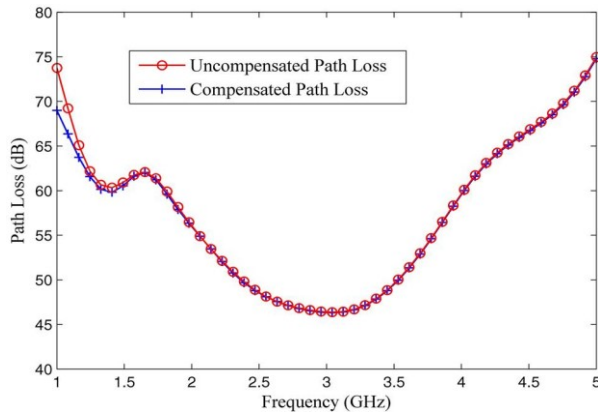


Fig.12: Uncompensated and compensated path loss by the tissue model vs frequency

Figure 6 shows the path loss. The large path loss of the UWB pulses is the main obstructing factor for wireless communication in the tissue medium [10]. In an experiment it was demonstrated that 1 Mbit/s link is possible with a BER of 10^{-2} at a maximum depth of 12 cm to on body receiver from implanted transmitter [15]. Here, the path loss is decreasing from 1-1.5 GHz then it again increases. From 1.7 to 3.1 GHz it decreases. Path loss around 2.7-3.2 GHz is minimum, around 46-47dB. After 4.5 GHz path loss is so enormous, after 5.5 GHz the path loss was more than 80dB.

The path loss values at 5cm depth and changes of path loss support the experimental results found in [10] which is claimed to be the first in body to on body path loss model. The simulation was restricted to 5GHz because since after that point the electric field is so large and conductivity is also huge that almost all transmitted power is absorbed by human tissue. Also the propagation losses of low-band UWB were also evaluated. In WBAN scenarios, low-band UWB is more advantageous than high-band UWB from the viewpoint of reducing propagation losses [16].

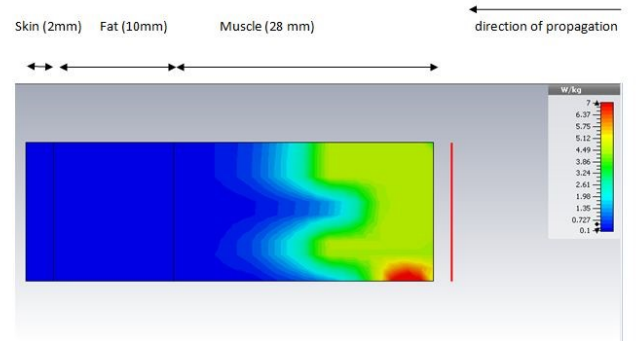


Fig.13: 3D SAR distribution on 2D plane at f= 1GHz.

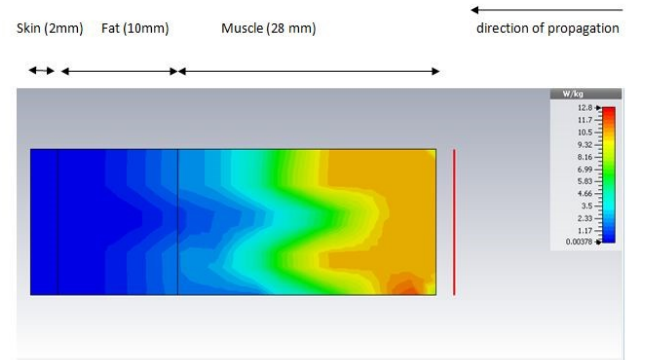


Fig.14: 3D SAR distribution on 2D plane at f= 2GHz.

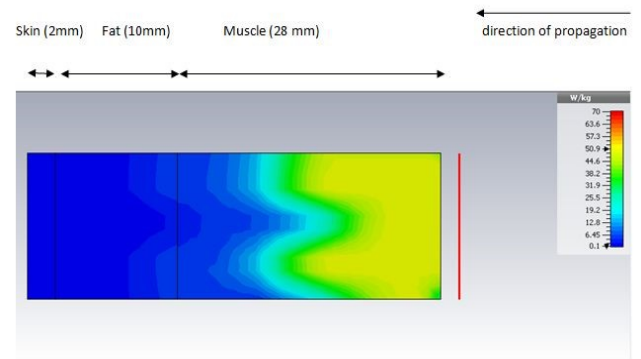


Fig.15: 3D SAR distribution on 2D plane at f= 3GHz.

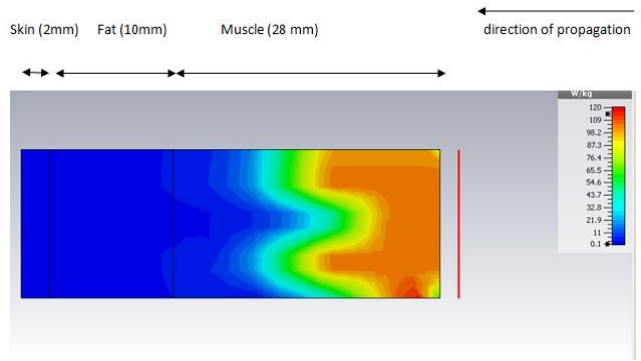


Fig.16: 3D SAR distribution on 2D plane at f= 4GHz.

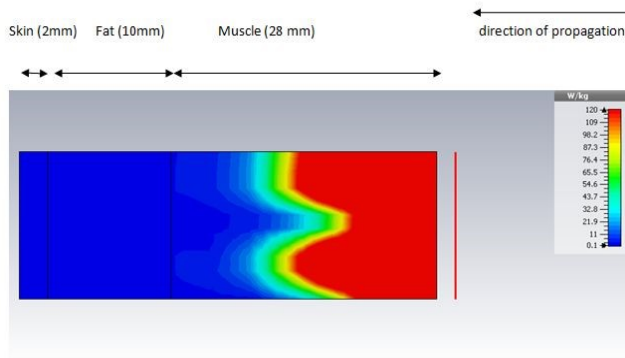


Fig.17: 3D SAR distribution on 2D plane at $f=5\text{GHz}$.

Figures 13-17 show SAR distribution in tissue layers. SAR is significantly increasing with the increase of frequency. For the muscle layer closest to the transmitter, SAR values are 4.5 W/kg, 9.75 W/kg, 44 W/kg, 89 W/kg and 120 W/kg at 1 GHz, 2 GHz, 3 GHz, 4 GHz and 5 GHz respectively. The SAR distributions are presented. SAR of Fat layer also increases but less significantly. Meanwhile, SAR for skin layer remains almost constant with frequency (almost 0.1 W/kg) since this layer is farthest of the three layers from transmitter.

V. CONCLUSIONS

S-parameters of the human tissue model, path loss and SAR distribution was presented at frequency range 1-5 GHz. Using S-parameter values we can estimate reflected power. Using path loss we can also estimate received power. Path loss was found to be around 47-65 dB at 3.1-4.5 UWB band. This is low UWB. In this range high speed rate transmission for implants is possible. After that range path loss increases significantly. So this low UWB (3.1-4.5) can be useful for in-body to on-body communication.

Specific absorption rate (SAR) distribution of different layers in human body in frequency range 1-5GHz was found. With the increase of frequency, SAR increases so fast due to increase of E-field value and conductivity increase. In 5 GHz in muscle interface SAR is even more than 120 W/kg which is too dangerous for health. Safety limit of SAR according to FCC regulations in that frequency range lies between 8-20 W/kg for per 10 gm of tissue in human body. Also that SAR decreases so rapidly with depth of penetration of the E-field, was also found from the 3D distribution on 2D plane. As the frequency increases SAR values and E-field magnitudes increase so strongly at the closest part of the transmitter which is muscle in this work.

ACKNOWLEDGEMENT

The author acknowledges the department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, for constant support and help throughout the research work.

REFERENCES

- [1] H. Y. Xu, H. Zhang, J. Wang, and L. X. Ma, "A new tapered slot antenna with symmetrical and stable radiation pattern," *Progr. Electromagn. Res. Lett.*, vol. 5, pp. 35-43, 2008.
- [2] H. P. Schwan, "Interaction of Microwave and Radio Frequency Radiation with Biological Systems," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-19, March 1971, pp.146-152.
- [3] Arthur W. Guy, "Electromagnetic Fields and Relative Heating Patterns Due to a Rectangular Aperture Source in Direct Contact with Bilayered Biological Tissue," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-19, March 1971, pp. 214-223.
- [4] H. P. Schawn and Geo. Morris Piersol, "The Absorption of Electromagnetic Energy in Body Tissues," in *Biological Effects of Electromagnetic Radiation*, IEEE Press, 1983, pp. 6-21
- [5] Curtis C. Johnson and Arthur W. Guy, "Nonionizing Electromagnetic Wave Effects in Biological Materials and systems," in *Biological Effects of Electromagnetic Radiation*, IEEE Press, 1983, pp. 47-73.
- [6] O. P. Gandhi, "State of the Knowledge for Electromagnetic Absorbed Dose in Man and Animals," in *Biological Effects of Electromagnetic Radiation*, IEEE Press, 1983, pp. 91-99.
- [7] Das, Ripan. *Transmission of Electromagnetic Power Through A Biological Medium*. Diss. Faculty of the Louisiana State University and Agricultural and Mechanical College in partial fulfillment of the requirements for the degree of Master of Science in Electrical Engineering in The Department of Electrical and Computer Engineering by Ripan Das B. Tech., Indian Institute of Technology, Kanpur, India, 2003.
- [8] Klemm, Maciej, and G. Troester. "EM energy absorption in the human body tissues due to UWB antennas." *Progress In Electromagnetics Research* 62 (2006): 261-280.
- [9] Akram, Gasmelseed, and Yunus Jasmy. "Lab VIEW-based Planar Multilayered Model for Estimation of the Absorbed Energy Inside Biological Tissues [EM Programmer's Notebook]." *IEEE Antennas and Propagation Magazine* 2.50 (2008): 152-158.
- [10] Christ, Andreas, et al. "The dependence of electromagnetic far-field absorption on body tissue composition in the frequency range from 300 MHz to 6 GHz." *IEEE transactions on microwave theory and techniques* 54.5 (2006): 2188-2195.
- [11] Floor, Pål Anders, et al. "In-Body to On-Body Ultrawideband Propagation Model Derived From Measurements in Living Animals." *IEEE journal of biomedical and health informatics* 19.3 (2015): 938-948.
- [12] J. A. Stratton, "Electromagnetic theory," McGraw Hill Book Company, New York, (1941).
- [13] V. Hebelka, "Planar antennas in proximity of human body models," *Elektrorevenue*, vol. 3, no. 4, pp. 10-13, Dec. 2012.
- [14] Nuno Goul'ao do Serro, "UWB Antenna for Portable Devices," M.S. thesis, Departamento de Engenharia Electrotécnica e de Computadores, Instituto Superior Técnico., Ist, Portugal, 2008.
- [15] Anzai, Daisuke, et al. "Experimental evaluation of implant UWB-IR transmission with living animal for body area networks." *IEEE Transactions on Microwave Theory and Techniques* 62.1 (2014): 183-192.
- [16] Yamamoto, Hironobu, and Takehiko Kobayashi. "Ultra-wideband Propagation Loss Around a Human Body in Various Surrounding Environments." *Ultra-Wideband, Short Pulse Electromagnetics 9*. Springer New York, 2010. 11-17.

Sentiment Analysis On Facebook Group Using Lexicon Based Approach

Sanjida Akter

*Lecturer in CSE dept.
Northern University Bangladesh
Dhaka, Bangladesh
Sanjida.akter@gmail.com*

Muhammad Tareq Aziz

*Software engineer
goBD
Dhaka, Bangladesh
tareq@gob.co*

Abstract—Internet is one of the primary sources of Big Data. Rise of the social networking platforms are creating enormous amount of data in every second where human emotions are constantly expressed in real-time. The sentiment behind each post, comments, likes can be found using opinion mining. It is possible to determine business values from these objects and events if sentiment analysis is done on the huge amount of data. Here, we have chosen *FOODBANK* which is a very popular Facebook group in Bangladesh; to analyze sentiment of the data to find out their market values.

Keywords— *social media, sentiment analysis, market bucket analysis, supervised learning, probabilistic classifier, lexicon based approach*

I. INTRODUCTION

Sentiment analysis is the concept: “how a person reacts on something or some events?”. It is always an important issue “what people think”. People’s thoughts are available in social medias to be used as data source of market basket analysis. Social media plays an important role here because we can have data from people even we don’t know. Such analysis can help to determine recent market trend or market value of a particular product or event. based on people’s interest. Here we have utilized FACEBOOK which is a very popular social media platform as our data source and. “FOODBANK” which a well-known FACEBOOK group.

For opinion mining or sentiment analysis some methods are applied like – Naive Bayes Machine Learning Classifier, Senti Word Net, Support Vector Machine. Here we applied both machine learning approach and Lexicon Based Dictionary. Most of the work on opinion mining is done with Naïve bayes. As our data set contains Bangla language; we built a Dictionary on top of the data set. We count the occurrence of sentimental words and featured words which are tagged as the target of calculated sentiment value. This approach can find 73% cases correctly whether post is positive or negative. After comparing two results, we found that lexicon based algorithm works better here.

At the end of analysis; we can determine recent trends and characteristics of people’s food habit. It also can be identified that “what will be the food habit in future” and “where investors should put their attention”.

II. SENTIMENT ANALYSIS

In recent time, sentiment analysis is a popular research topic because many real-life problems are part of this subject matter. It is also highly challenging as NLP research topic that covers many novel sub problems. Additionally, there was almost no significant research done before the 2000 in either NLP or linguistics. It is because of lack of availability of opinion or text in digital forms. Since the year 2000 the field has grown rapidly to become one of the most active research areas in NLP. It is also widely researched in data mining, Web mining, and information retrieval technologies. In fact, it has spread from computer science to management and business intelligence.

There are mainly three methods or level of research on Sentiment Analysis.

Document level [1]: analyze the overall sentiment expressed in the text and determine if the overall sentiment is positive or negative.

Sentence level [2] - examine the sentiment expressed in sentences and determines whether each sentence expressed be a positive, negative, or neutral opinion.

Aspect level [3] - aspect level performs better analysis when document level and the sentence level analysis do not find what people liked and did not like. Aspect level also called feature level.

Aspect level does not work on document, paragraph sentences or clauses, it directly finds out the sentiment and the target. The sentiment can be positive, negative or neutral towards that specific target or entity. Realizing the importance of opinion targets also helps us understand the sentiment analysis problem better. For example, “I love to eat chicken”.

Here we can see a positive sentiment and the target of this sentiment or the entity is “chicken”. There is another challenge which is that we can classify opinion in two types: Regular opinion and Comparative opinion [4].

III. LEXICON BASED

Undoubtedly, the most important indicator of sentiments are sentiment words. These are words that are commonly used

to express positive or negative sentiments. For example, good, wonderful, and amazing are positive sentiment words, and bad, poor, and terrible are negative sentiment words. Sentiment words are instrumental to sentiment analysis for obvious reason. A list of such words are called a sentiment lexicon (or opinion lexicon). Over the years, researchers have designed numerous algorithms to compile such lexicons.

Although sentiment words are important for sentiment analysis, only using them is not sufficient. The problem is much more complex.

Some issues related to lexicon based approach are given below:

- A positive or negative sentiment word may have opposite orientations in different application domains.

For example, “dangerous” is a negative word but it can also be highly positive when someone posts “dangerous tasty”.

- A sentence containing sentiment words may not express any sentiment. This happens frequently in question types of sentences.

For example, “Suggest a good coffee shop around gulshan n banani...”

- Many sentences without sentiment words can also imply opinion.

For example, “They took a lot of time to serve us”. This sentence has a negative sentiment. In recent time many researchers are working on lexicon based approach [5].

IV. SENTIMENT ANALYSIS ON SOCIAL MEDIA GROUP

It is possible to understand the real world movement by analyzing data of social media groups. [5] Sentiment analysis tools have been applied to examine the relationship between release of products, the ‘discussion’ online, and actual sales of products with the outcome being that such data can be used to predict sales volumes [6].

As we have mentioned that sentiment analyses are about opinion or expression or imply positive or negative sentiment or emotion towards an object. So the elements of Sentiment analysis are:

“Opinion” which are a judgment, viewpoint, or statement about matters commonly considered to be subjective.

“Entity” which are a product, service, topic, issue, person, organization, or event. It is the target of an opinion.

“Subjectivity and emotions” which are the state of mind of a person and instinctive responses.

Let us use a Facebook post on “FoodBank” group to describe opinion properly. The post is given below with the time when it was posted.

"Today we visit KFC. The decoration was awesome. We ordered Chicken Zinger burger. It was so tasty. But my sister think it contains too much calories."

Observation:

1. This post contains a number of opinion.
2. It has more than one target object or Entity.
3. It has both positive and negative opinion.
4. It has two opinion holders.
5. This post is holding a date.

V. SENTIMENT ANALYSIS USING MACHINE LEARNING APPROACH

almost 90% researchers work with different machine learning algorithms to detect sentiment or for opinion mining. Some common algorithms that are used for sentiment analysis are – Naïve Bayes, Maximum Entropy, support vector machine etc. [8]. Naive Bayes classifiers are highly scalable, requiring a number of parameters linear in the number of variables (features/predictors) in a learning problem. Maximum-likelihood training can be done by evaluating a closed-form expression, which takes linear time, rather than by expensive iterative approximation as used for many other types of classifiers.

VI. SENTIMENT ANALYSIS USING LEXICO BASED APPROACH

The lexical approach is a method of teaching foreign languages described by M. Lewis in the 1990s. At recent days researchers also working with lexicon based approach for sentiment analysis.[9][10] The basic concept on which this approach rests is the idea that an important part of learning a language consists of being able to understand and produce lexical phrases as chunks. Students are thought to be able to perceive patterns of language (grammar) as well as have meaningful set uses of words at their disposal when they are taught in this way.

Machine translation can use a method based on dictionary entries, which means that the words will be translated as a dictionary does i.e. word by word, usually without much correlation of meaning between them. Dictionary lookups may be done with or without morphological analysis or lemmatization. While this approach to machine translation is probably the least sophisticated, dictionary-based machine translation is ideally suitable for the translation of long lists of phrases on the sub sentential (i.e., not a full sentence) level, e.g. inventories or simple catalogs of products and services.

VII. OUR APPROACH AND EXPERIMENT RESULT

Apart from many processes, we have used two main approaches to analyze sentiments of Social Network Data. These are:

- Naïve Bayes Machine Learning classifier.
- Dictionary Based Approach.

A. Collecting Data

As per our interest we decided to do our sentiment analysis experiment on Social network data. But social network is full of chaos. So we needed a group where people would talk on specific topic.

We found there are groups like BDCyclists, FoodBank where people discuss on same kind of topic. we considered “FoodBank” for this work.

To collect Data, there is an API for developers in Facebook called Graph API. This API lets programmer to do different kinds of programmatic activities on top of Facebook data. So we wrote a C# console application to download all the posts that is available on FoodBank since the beginning.

B. Formatting Data

Before using the data in any of the sentiment analyzing process, we needed to do some formatting on the data. The data we downloaded were in JSON format.

1) *Formatting Data for Naïve Bayes Classification:* We write another c# console application that would extract and make a list of Messages of all posts for naïve Bayes classification. We write another Visual GUI application in C# that would let our team mates do the manual classification since all machine learning classifiers are heavily depend on trained data set.

2) *Formatting Data for Naïve Bayes Classification:* To do this, we tokenize each and every words. We made a list of unique words and there were 9000 words. Then we manually classified all these words based on sentiment rating like neutral, positive, negative, super positive, super negative, food name or location name etc.

C. Formatting Data

In machine learning, naïve Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes' theorem with strong (naïve) independence assumptions between the features. Since it's a probabilistic classifier, it heavily depends on the probability mathematics. We used Python programming language for this classification technique. We took about 3600 random statuses from the list and classified them manually based on human judgment.

Problems with Naïve Bayes Classifier:

In different scientific papers it has been found that Naïve Bayes machine learning classifier is extremely successful when applied on well-formed text corpuses. Like movie reviews or novels.

But in Facebook, Naïve Bayes has not performed well, at least according to our experimentation. When we gave a very close look to the poor performance of Naïve Bayes Classifier, we found that the Facebook Group Posts are extremely noisy. Since people post in random lengths with lots of spelling mistakes. For example:

Status	Problems
The delicious pizza. Bt must add. the mushroom was very nice...🍕🍕	No mention of what pizza and where
Ajke kheyte ashlam Comic Cafe er set menu no.81 OMG OMG !!! that was soooooo tasty !! Quantity etttoo beshi j eka kheyte shesh kora possible na. Amar shesh korte onkkk koshto hoise. 320/= ekdom usul hoy trust me!! Place- Comic Cafe Price- 10/10 (320 tk) Service- 9/10 (10mins wait korte hoise seat pawar jonno)	Bangla written in English form. Frequent in Social Media
Best place for dinner in Gulshan or may be Banani. The food must be real good! Urgent reviews needed! Tia	Nearly 50% of query is comprised of queries like this. As a result many of the sentences do not express exactly any sentimental value

Fig. 1. variation in post on social network

Since variation and combination of words in local Facebook groups are way too high; we needed a huge amount of training sample in order to reach to a satisfactory level of accuracy. But at this moment we have found simple counting positive and negative words. Specialized key words are giving better performance than Naïve Bayes classifier. So, for the rest of the sentiment analysis technique we emphasized on Lexicon Based approach.

D. Approach II (our proposed solution: Lexicon based Sentiment Analysis)

In a nutshell, it is nothing more than counting key positive and negative and neutral key words We found this approach is extremely efficient since ordinary people posts on this FoodBank and they are not comprising of complex sentence structures that would let us to satirical or ironic posts. Point to be noted here that people on FoodBank uses a useful pattern to review something. For example:

my review on #fazitas sub-sandwich nd nachos... #nachos.nachos were preety good... actually great grin emoticon quite similar like Mexican one.Enjoyed it smile emoticon taste:10/10 price:10/10 #sandwich:nacho meat nd sandwich meat was same colorthree emoticon they gave just a single slice of capsicum,from emoticon not enough ketchup cry emoticon (I think after getting such reviews on sub-sandwich of fazitas my expectations were high) my sis didn't like it at all colorthree emoticon taste:5.5/10 price:9/10	মাওয়া ঘাটের তাজা ইলিশ না খাইলে মিস্ 🤔 ফেরী ঘাটে যখন যাবেন খাবার নেওয়ার আগে দাম টা জিজ্ঞেস করে নিবেন । কারন ওখানে যার থেকে যেমন নিতে পারে তেমন নিয়ে যায়। সাধারণত বাছাই করে কাঁচা ইলিশ নিলে ওরা এক পিচ ইলিশ ৬০ টাকা করে নিয়ে থাকে । আবার কেউর কাছে ১০০, ১২০ তার বেশিও নিয়ে থাকে । আমি নিজের পছন্দ মত নিয়ে ছিলাম কাঁচা ইলিশের পিচ তারপর তাজা তাজা বেজে দিলো । কিন্তু আমার কাছ থেকে ৬০ টাকা করে নিয়েছে। smile emoticon Price-10/10 Taste-9/10 Service-8/10
---	--

Fig. 2. Rating pattern on FOODBANK

We can observe that sometime people rate foods or location on a scale of 10. This is an interesting keyword to determine what people think about certain feature words. But since people write such kind of rating in different manners; sometimes it is hard to find out actually what is rated and if there are multiple feature words. So, we consider such sentiment rating as a basis of judging the whole messages. Our

approach calculates the sentiment of the status in subjective manner rather than objective.

E. The proposed work

1. Download all the data from FoodBank.
2. Make a list of unique words both that are used as a whole in all of the statuses.
3. Noise out all the words that doesn't represent any sentimental value or feature value like location and food.
4. Rate the rest of the words that do represent the sentimental or feature value, make a wordlist out of it.
5. Check for the occurrence of words of our word list.
6. Tag each status with those words to each status, the sentimental values represent the sentiment of whole status. The food or location indicates to whom the sentiment is expressed.
7. When all the messages sentiment is detected, a macro level overview like food/ restaurant popularity, trends etc. is possible.

We have followed all these steps to analyze sentiment in this lexicon based technique.

VIII. APPLICATIONS OF SENTIMENT ANALYSIS

There are many applications of Sentiment analysis. A short list of application is given bellow.

1. Customer Opinion mining
2. Manipulation
3. Political Observation
4. Social Study
5. Crowd Sourcing
6. Artificial Intelligent Research
7. Business Insight

IX. ADVANTAGES OF SENTIMENT ANALYSIS USING LEXICON BASED APPROACH

Advantages of sentiment analysis on a social media are given bellow.

1. Making decision based on knowing others interest.
2. Low cost and time saving method of getting consumer insight.
3. A faster way of getting customer data.
4. Ability to act on customer suggestion.
5. Helps to Identify Organization's Strength, Weakness, Opportunities and Threats. (SWOT) analysis.

6. More accurate and insightful customer perception and feedback

X. CONCLUSION

In this paper, we have shown how to predict the sentiment behind a status post of Facebook which in nature of unstructured dataset, cross language domain and noisy. So, it can be determined that traditional opinion mining is not efficient enough to find out sentiments from social media's like FACEBOOK. But lexicon based dictionary approach works efficiently in such kind of works. As sentiment analysis has many practical advantages so a structured process need to be developed.

References

- [1] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan "Thumbs up? Sentiment classification using machine learning techniques" Proceedings of EMNLP, pp. 79-86, 2002.
- [2] Steven Bethard, Hong Yu, Ashley Thornton, Vasileios Hatzivassiloglou and Dan Jurafsky Wilson, Wiebe and Hwa, "Automatic Extraction of Opinion Propositions and their Holders" Computing Attitude and Affect in Text: Theory and Applications Volume 20 of the series The Information Retrieval Series pp. 125-141.
- [3] Wilson, Wiebe, Hwa "Just how mad are you? finding strong and weak opinion clauses" presented at the 19th national conference on Artificial intelligence, pp. 761-767.
- [4] Mingqing Hu and Bing Liu "Feature-based opinion mining and summarization" KDD '04 Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining pp.168-177
- [5] Nitin Jindal and Bing Liu "Mining Comparative Sentences and Relations" (2006), AAAI'06 proceedings of the 21st national conference on Artificial intelligence - Volume 2, pp.1331-1336.
- [6] Moreo A, Romero M, Castro JL, Zurita JM. "Lexicon-based comments-oriented news sentiment analyzer system", Expert Syst Apple 2012; pp. 66-80.
- [7] Walaa Medhat, Ahmed Hassan, Hoda Korashy "Sentiment analysis algorithms and applications: A survey", Ain Shams Engineering Journal, Volume 5, Issue 4, December 2014, pp. 1093-1113.
- [8] Bo Pang, Lillian Lee, Shivakumar Vaithyanathan "Sentiment classification using machine learning techniques" Proceedings of EMNLP, 2002, pp. 79-86.
- [9] Prabu Palanisamy, Vineet Yadav, Harsha Elchuri "Serendio: Simple and Practical lexicon based approach to Sentiment Analysis" presented at Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Seventh International Workshop on Semantic Evaluation (SemEval 2013), Atlanta, Georgia, June 14-15, 2013. pp. 543-548.
- [10] Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, Manfred Sted "Lexicon-Based Methods for Sentiment Analysis", Computational Linguistics Volume 37, Number 2, 28 September 2010, pp. 268-307.
- [11] Lincoln C. Wood, Torsten Reiners, Haris S. Srivistava "Expanding Sales And Operations Planning Using Sentiment Analysis: Demand And Sales Clarity Form Social Media" ANZAM 2013, December 4-6, 2013.
- [12] Mohammad Ali Abbasi, Sun_Ki Chai, Huan Liu, Kiran Sagoo "Real-World behavior analysis through a social media lens" SBP'12 Proceedings of the 5th international conference on Social Computing, Behavioral-Cultural Modeling and Prediction, pp. 18-26.
- [13] Bo Pang, Lillian Lee "Opinion mining and sentiment analysis", Foundations and Trends in Information Retrieval Vol. 2, No 1-2 (2008) pp.1-135.

Classification of Motor Imagery Hands Movement Using Levenberg-Marquardt Algorithm based on Statistical Features of EEG Signal

Md. Mamun or Rashid

Department of Biomedical Engineering,
Khulna University of Engineering & Technology
Khulna-9203, Bangladesh
almamun369@yahoo.com

Mohiuddin Ahmad

Department of Electrical and Electronic Engineering,
Khulna University of Engineering & Technology
Khulna-9203, Bangladesh
mohiuddin0ahmad@gmail.com

Abstract—Motor Imagery (MI) is a highly supervised method nowadays for the disabled patients to give them hope. This paper proposes a differentiation method between imagery left and right hands movement using Daubechies wavelet of Discrete Wavelet Transform (DWT) and Levenberg-Marquardt back propagation training algorithm of Neural Network (NN). DWT decomposes the raw EEG data to extract significant features that provide feature vectors precisely. Levenberg-Marquardt Algorithm (LMA) based neural network uses feature vectors as input for classification of the two class data and outcomes overall classification accuracy of 92%. Previously various features and methods used but this recommended method exemplifies that statistical features provide better accuracy for EEG classification. Variation among features indicates differences between neural activities of two brain hemispheres due to two imagery hands movement. Results from the classifier are used to interface human brain with machine for better performance that requires high precision and accuracy scheme.

Keywords— *Motor imagery (MI), discrete wavelet transform (DWT), brain machine interface (BMI), neural network (NN).*

I. INTRODUCTION

Patients with severe neuromuscular disorders and paralyzed from higher level spinal cord injury or last stage amyotrophic lateral sclerosis (ALS) patients are not capable of moving any voluntary muscle [1]. Their sensory and cognitive functions become only way to communicative with surroundings. Such patients can be motivated, trained for MI movement to execute real movement by incorporating with Brain Computer Interface (BCI) utilizing neural activity to do like cursor control, selection of letters or words, control of prosthesis, navigation of wheelchair etc.

Imagery movement of two hands can be distinguished according to the change of the electrical (neural) activity. MI is the procedure where imagining motor activity human observe activation of motor cortex with or without external trigger. Motor imagery movements mimic original motor act without even accomplishing any real motor activity. Motor performance, sensory provocation, and mental imagination modify the functional activity of the cortex and results in increasing or decreasing amplitude. Motor imagery

movements dramatically change the neural activity of the primary sensorimotor areas in a quite indifferent way as real performed movements [2].

Practical implementation of mind machine interface (MMI), brain machine interface (BMI) or direct neural interface (DNI) system requires a well organize brain signal processing method to extract responsible features and to perform classifications with better accuracy. Several methods have been reported to extract features in different domains where wavelet transformation (WT) based features extraction method is highly efficacious method, because it works greater with the random characteristics of EEG signals than other scheme. Wavelet-based features like wavelet entropy, wavelet coefficients and wavelet statistical features (mean, median, and standard deviations) most commonly used for normal EEG analysis in clinical applications as well as in research areas also [3].

In literature, two new features, (1) multifractal cumulants and (2) predictive complexity measure are introduced in addition with band power feature [4]. Linear Discriminant Analysis (LDA) classifier is used to classify and compares classification accuracy with each other (features). A multilayer perceptron neural network (MLPNN) based classification model is [3] proposed using EEG signal. DWT and MLPNN outcomes classification accuracy for five different experiments are above 95% for individuals. Authors in [5] proposed a scheme to calculate relative wavelet energy in terms of approximation coefficients and detailed coefficients. Extracted features are given as input to the SVM, K-nearest neighbor (K-NN), multi-layer perceptron (MLP) and Naive Bayes classifiers with classification accuracy 98.75%, 98.21%, 98.21%, 83.57% respectively. In [6] authors unite time and frequency features approach for the purpose of EEG signals differentiation between epileptic and healthy subject. Time domain and frequency domain features are extracted using the cross correlation (CC) and power spectral density (PSD) method respectively. The combination of the cross correlation and power spectral density method for EEG classification is very effectual. Stockwell Transform (ST) dependent study of EEG signals are introduced during performing different mental tasks or assignment [1].

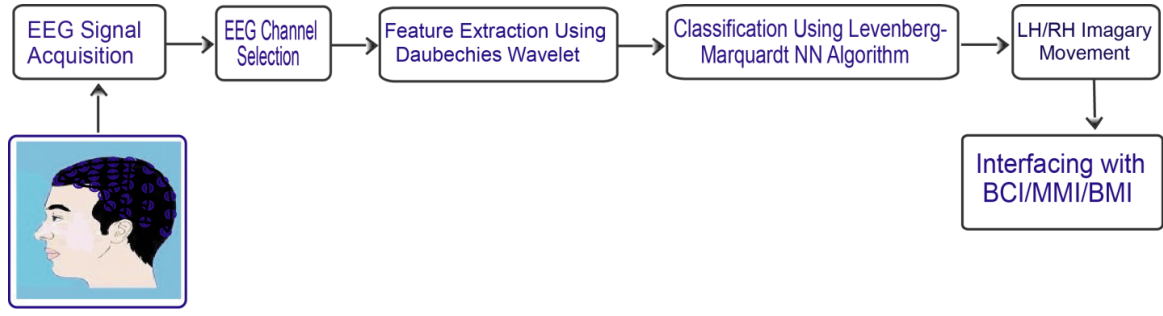


Fig. 1. Block diagram of the classification of imagery LH/RH movement.

kNN, LDA and SVM classifiers were employed to test the classification strength of the proposed features. Accuracy achieved to distinguish the multi-class problems (five mental tasks) is ranging between 84.72% and 98.95%. Wavelet pack entropy features are selected to classify a baseline task from a cognitive task with SVM classifier which provides overall 87.5–93 % accuracy [7]. This study has used a database where seven subjects performed five assign tasks—baseline (open eyes) task, multiplication, visual counting, mental letter composing, and geometric object rotation task.

Many wavelet based research work accomplished [5] except using wavelet statistical features. This paper proposes that wavelet statistical features can be used to classify EEG signal in addition with better classification accuracy. This paper also investigate the performance of the Levenberg-Marquardt algorithm to classify right and left hands imagery movement using Daubechies 4 wavelet.

This paper is organized as follows: Section II illustrates materials and proposed method. Section III states the results and discussion is in section IV. Finally section V concludes the paper.

II. MATERIALS & PROPOSED METHOD

A. EEG Data Acquisition & Description of Dataset

This dataset is collected from internet [4] which is taken at INRIA Renne-Bretagne Atlantique using the OpenVIBE, an open-source and free software [4]. This data set contains EEG signals of one subject performing left hand and right hand motor imagery. Five hundred sixty trials of motor imagery (280 trials per class) were captured over a 2 week period.

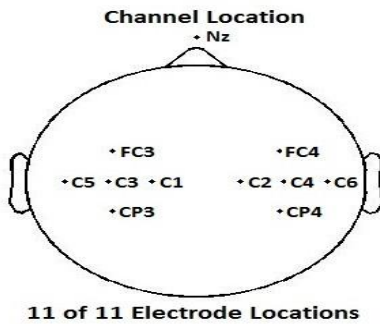


Fig. 2. Electrode placement system of the subject.

During data acquisition same experimental protocol is used as Graz BCI protocol [2], [4] to differentiate the two mental conditions. The experimental task is to imagine the movement of one hand (left/right) for specific time of each trial. Each trial lasts for 8 seconds including resting, imagery movements of either left or right hand and for feedback. Sampling rate of EEG data is 512 Hz and recorded from frontal, central and central frontal lobe using the following 11 electrodes: C1, C2, C3, C4, C5, C6, CP3, CP4, FC3, FC4 and a nose reference electrode Nz [4]. Such electrodes placement helps to employ a discrete Laplacian spatial filter [8] over C3 and C4 to collect better noise free signal, as recommended in [4]. All electrodes placement are depicted in Fig. 2, Nz is the reference electrode. Six electrodes from ear to ear band are very crucial for MI.

B. Channel Selection and Experimental Flowchart

Eleven channels have been used including reference electrode for data collection but our interested channels are above the sensorimotor cortex region. The motor imagery EEG signal is predominant in sensorimotor cortex, corresponding to electrode positions C3 and C4 [4]. In our dataset each electrode collects data of both L/R hands data. Based on the universal truth that left hemisphere controls our right side of the body and right hemisphere powers over left part of the body, we selected C3 channel data for right hand imagery movement and C4 channel data for left hand imagery movement. Figure 1 represents the complete work flow starting with taking data from subject and after classification of L/R hand; output result can be used to interface BCI or BMI.

C. Discrete Wavelet Transform (DWT)

DWT is a novel tool for decomposition and reconstruction filters to extract non stationary characteristics of EEG signal. In this work db4 wavelet is used to analyze signals into 5 levels of sub bands frequencies using continuous high pass and low pass filtering shown in Fig. 3. The high-pass filter, g is the discrete mother wavelet and h is the mirror version of low-pass filter. DWT decomposes a provided signal into increasingly finer detail depending on two sets of fundamental functions [9], the wavelets and the scaling functions of a signal $S(t)$ as follows:

$$S(t) = \sum_k 2^{j_0/2} a_{j_0}(k) \phi(2^{j_0 t - k}) + \sum 2^{j/2} d_j(k) \psi(2^j t - k) \quad (1)$$

Where $\phi(t)$ and $\psi(t)$ are the basic scaling and mother wavelet functions respectively.

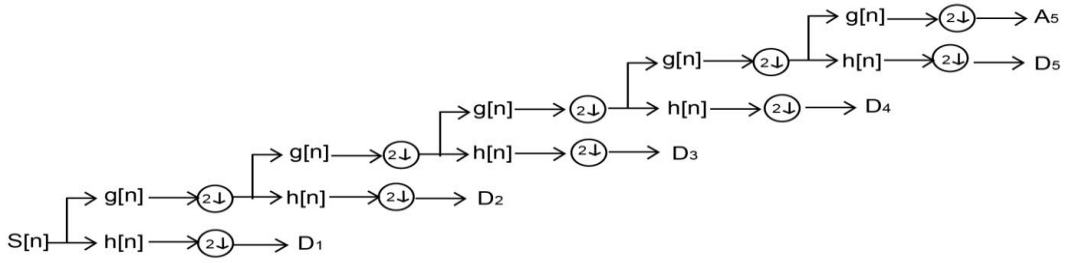


Fig. 3. Decomposition of EEG segmented signal using 5th order Daubechies wavelet.

The sampled output signals from the first filters are called first level approximation, A_1 and detail coefficients, D_1 of the original signal. Approximations and details are the high-scale, low-frequency components and the low-scale, high-frequency components of the signal respectively. The two filters are related to each other and they are defined as a quadrature mirror filter. Approximation and detail coefficients of subsequent level are estimated by using the approximation coefficient of the earlier level. The required decomposition level depends on the commanding frequency components of the signals [3]. After decomposition significant 10 features are collected to make feature vector for further processing to classify imagery LH/RH movement.

D. Neural Network (NN)

In ANNs, the intelligence based on the interrelation of weights between neurons to neurons used in the NN design. Therefore, training course of neurons is the compulsory, inseparable part of an ANN system because a well trained network performs well. After training completion, NN can be arranged to work perfectly for improved outcome.

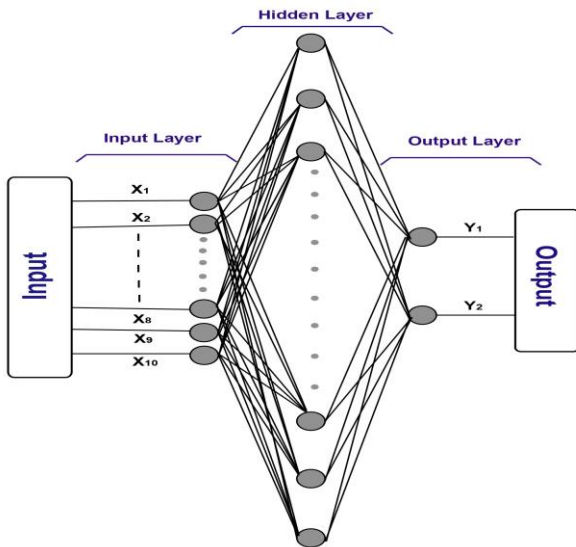


Fig. 4. LMA based NN design with 15 neuron in hidden layer.

Levenberg Marquardt Algorithm (LMA) is a fast and efficient algorithm for EEG signal classification which uses the standard numerical optimization technique. It has both property of local convergence of the Gaussian-Newton method and the global of the Gradient descent method. LMA increases the training speed, decreases the training steps as well as improves classification accuracy, though it's better to use trial and error method for better accuracy. The neural network used in this study is depicted in Fig. 4. This neural network designed with LMA where the number of neuron is chosen 15 into the hidden unit. The number of neuron is determined by doing trial and error depending on the performance accuracy of the network. 70 % of the total feature vector provided for training as good training determines outcomes well and from the rest of the data 15 % of the data used as input vector for validation and remnant 15% vectors used as testing input data.

III. RESULTS

A. Feature extraction results using DWT

Decomposition of segmented EEG signal upto level 5 is shown in Fig. 5 and in Fig. 6 for LH and RH data respectively to extract the statistical features. In both figures signal S decomposes into 6 sub-signals according to their frequencies and amplitudes. Sum of all the detail coefficients and approximation coefficient will be the equivalent of the S signal. Features for each segment data are mean, median, maximum, minimum, standard deviation, median absolute deviation, mean absolute deviation, L1 norm, L2 norm, max norm. For a segmented EEG data of left and right hand imagery movement taken from right hemisphere and left hemisphere are given in Table I.

TABLE I. EXTRACTED FEATURES OF A SEGMENTED EEG DATA FOR IMAGINATION LH/RH MOVEMENTS

Features	Left Hand	Right Hand
Mean	9263	4706
Median	9263	4706
Maximum	9275	4714
Minimum	9249	4697
Standard Dev.	4.514	2.865
Median Abs. Dev.	2.929	2.065
Mean Abs. Dev.	3.574	2.293
L1 Norm	7.123	3.619
L2 Norm	2.569	1.305

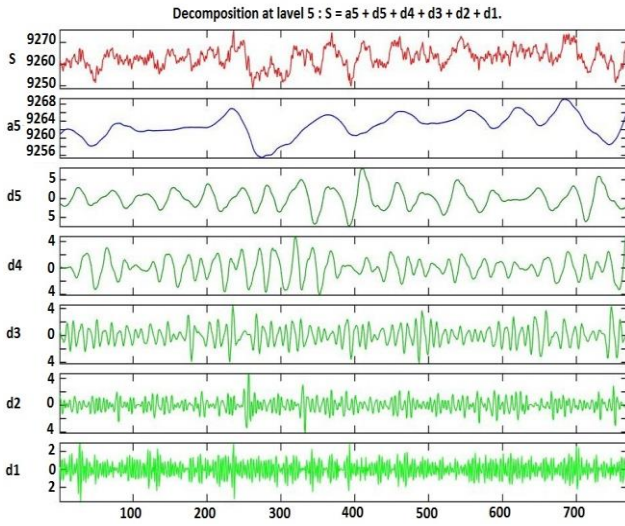


Fig. 5. Decomposition of imagery LH movement data.

B. Classification results using NN

Features collected from left and right hemisphere definitely assure differences for the imagery movement of the two hands. Now for the classification between two imagery hands movement (two classes), among 500 both hands trials, 450 trials (both hands) randomly has been used for classification. For training purposes 314 trials for training, 68 validation trials for and 68 trials for testing (both left and right) randomly used. Levenberg-Marquardt Backpropagation training algorithm used in this network because its training consumes less time. After classification confusion matrix of LMA based NN is shown in Table II. Among 100% (450) trials, for LH imagery 50% (225) trial, 42.7 % (192) trials are correctly classified. Same happens for RH trials where 49.3 % (222) trials are correctly classified from 225 trials (50%) of the total trials.

TABLE II. CONFUSION MATRIX OF NN

CONFUSION MATRIX		
Output/Desired	Left Hand Imagery Movement	Right Hand Imagery Movement
Left Hand Imagery Movement	192 42.7%	3 0.7%
Right Hand Imagery Movement	33 7.3%	222 49.3%

The performance of the classifiers was determined by the calculation of sensitivity, specificity and total classification accuracy are defined below and results shown in Table III.

$$\text{Sensitivity (True positive rate)} = \text{TP} / (\text{TP} + \text{FN}) \quad (2)$$

$$\text{Specificity (False positive rate)} = \text{TN} / (\text{TN} + \text{FP}) \quad (3)$$

$$\text{Total classification accuracy (Percentage of all samples correctly classified)} = (\text{TN} + \text{TP}) / (\text{TN} + \text{FN} + \text{TP} + \text{FP}) \quad (4)$$

Where,

Total number of correctly classified positive patterns = TP,

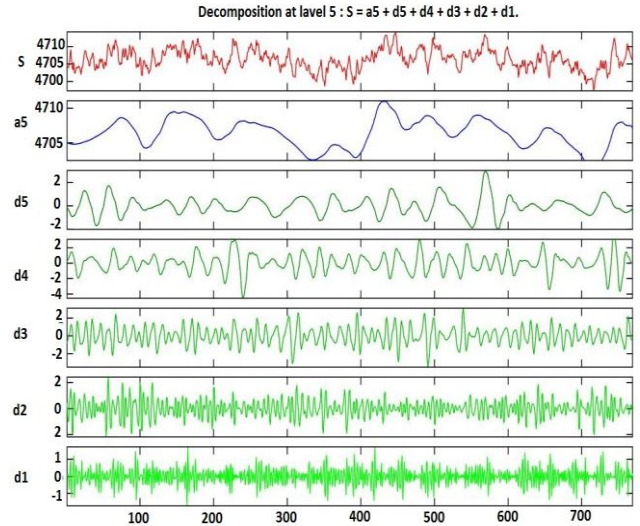


Fig. 6. Decomposition of imagery RH movement data.

Total number of actual positive patterns = TP+FN

Total number of correctly classified negative patterns = TN

Total number of actual negative patterns = TN + FP

Total number of correctly classified patterns = TN + TP

Total number of applied patterns = TN + FN + TP + FP.

TABLE III. CALCULATION OF STATISTICAL PARAMETER

Statistical Parameters		
Sensitivity (%)	Specificity (%)	Accuracy (%)
87.06	98.46	92.00

Performance curve represents the relationship among validation performance, train and test performance of a neural network with respect to the best. Validation curve determines network generalization, and to terminates training when generalization halts progressing.

The validation curves and test curves are quite analogous shown in Fig. 7. In the graph if the test curve increments meaningfully prior to the validation curve increases, possibly

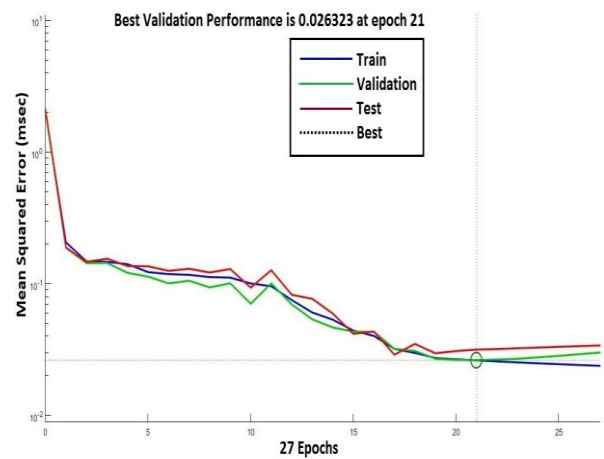


Fig. 7. Relationship between train, validation and test curve.

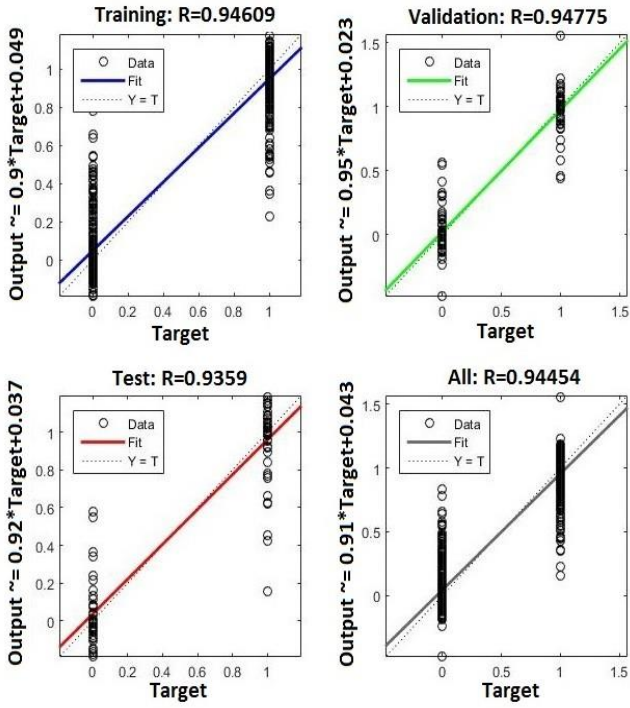


Fig. 8. Regression curve of train, validation, test and all feature vectors with 'R' value.

then some over fitting have occurred. Mean Squared Error predicts the average squared difference between outputs and targets provided in the designed NN network from the feature vectors. Lower values are acceptable and zero means no error at all. In the proposed method, validation performance is 0.026323 at epoch 21 which is shown in Fig. 7.

The four plots in Fig. 8 represent the regression, R values of training, validation, testing for all data and this pictorial view depicts the accuracy of the features vectors for train, test and validation data. The dashed line in Fig. 8 depicts the perfect result – outputs = targets. The solid line illustrates the finest fit linear regression line within outputs vectors and targets vectors. The R value demonstrates the interconnection between outputs and targets. If $R = 1$, this denotes that there is an exact linear relationship between outputs and targets feature vectors. If R is close to zero or zero, then there is no linear relationship. Here in the portrait of training, test and all data indicates that most of the data fit well. The validation results also show R values greater than 0.948. If the network performance on the validation vectors stops to increase or remains unchanged validation vectors are used to halt training early. Training ceases when the significant number of epochs are reached or the adequate amount of time is exceeded or even if the performance of network is attenuated to the target. Test vectors controls over the network generalization but do not have any influence on training.

IV. DISCUSSIONS

This proposed methodology with db4 and LMA promises satisfactory classification accuracy. Difference between the electrical activities in comparison with one hemisphere to another is expected because right half of our cerebrum controls the left side of the body and vice-versa. This assumption makes this study feasible.

We make a comparison list of some previous researches with our proposed method by comparing setup, methodology, and classification accuracy as shown in Table IV.

TABLE IV. COMPARISON OF METHODS AND CLASSIFICATION ACCURACY WITH PROPOSED METHOD

Researchers	Setup and Database	Feature Extraction And Classification Algorithms	Classification Accuracy
M. Hariharan et al. [1]	Keirn and Aunon database Seven subjects (five mental tasks).	S-transform based MSR features kNN, LDA and SVM	Between 84.72% and 98.95%
Solhjoo et al. [10]	Dataset I: EEG data set (Oxford University). Dataset2: Three channel EEG dataset (University of Technology Graz).	Mahalanobis distance, Gaussian classifiers, LDA classifiers using PSD method	63.1%, 65.4% and 65.5% respectively
M. H. Alomari et al. [11]	The EEG dataset (Physio Net). http://www.physionet.org/pn4/eegmmidb/ .	NN and SVM	89.8% and 97.1% respectively
S.-M. Zhou et al. [12]	Graz BCI data set	Higher order features for classifiers NN and LDA	90.00% and 89.29% respectively
Yue Chen et al. [13]	Graz data set B, BCI competition 2008	Levenberg-Marquardt algorithm with Multilayer feed forward neural network	78.2% and 87.1% respectively
Zhang et al. [14]	Keirn and Aunon database Four subjects with five mental tasks	High frequency power and Fischer's discriminant classifier	72.4–76.4 %
Xue et al. [15]	Six-channel, four subjects EEG data performing three mental tasks	Wavelet packet transform for feature extraction with RBF classifier	85.3 %
Proposed Method	OpenViBE / INRIA data, using the OpenViBE	Db4 wavelet and LMA based NN	92%

V. CONCLUSION

This study proved that statistical parameters (features) can be used to differentiate two imagery hands movement with standard accuracy. This proposed methodology provides convincing classification accuracy of 92 % using statistical parameters of EEG signal. In BCI or MMI application proposed method is highly applicable. In future, classification accuracy of imagery movements of hands and feet will be determined.

ACKNOWLEDGEMENT

This work was supported by Higher Education Quality Enhancement Project (HEQEP), UGC, Bangladesh; under Subproject "Postgraduate Research in BME", CP#3472, KUET, Bangladesh.

REFERENCES

- [1] M. Hariharan, V. Vijejan, R. Sindhu, P. Divakar, A. Saidatul, and S. Yaacob, "Classification of mental tasks using stockwell transform," *Computers & Electrical Engineering*, vol. 40, no. 5, pp. 1741–1749, July 2014.
- [2] G. Pfurtscheller and C. Neuper, "Motor Imagery and Direct Brain-Computer Communication," *Proceedings of the IEEE*, vol. 89, no. 7, pp. 1123–1134, July 2001.
- [3] U. Orhan, M. Hekim, M. Ozer, "EEG signals classification using the K-means clustering and a multilayer perceptron neural network model," *Expert Systems with Applications*, vol 38, issue 10, pp. 13475–13481, 15 September 2011.
- [4] N. Brodu, F. Lotte, and A. Lécuyer, "Exploring Two Novel Features for EEG-based Brain-Computer Interfaces: Multifractal Cumulants and Predictive Complexity," *Neurocomputing*, vol 79, pp. 87–94, 1 March 2012.
- [5] H. U. Amin, A. S. Malik, R. F. Ahmad, N. Badruddin, N. Kamel, M. Hussain and W. Chooi, "Feature extraction and classification for EEG signals using wavelet transform and machine learning techniques", *Australasian Physical & Engineering Sciences in Medicine*, vol. 38, issue 1, pp. 139–149, March 2015.
- [6] Z. Iscan, Z. Dokur, T. Demiralp, "Classification of electroencephalogram signals with combined time and frequency features," *Expert Systems with Applications*, vol. 38, issue 8, pp. 10499–10505, August 2011.
- [7] L. Zhiwei, S. Minfen (2007), "Classification of mental task EEG signals using wavelet packet entropy and SVM," *Electronic Measurement and Instruments, 2007. ICEMI '07. 8th International Conference on*, pp. 3-906–3-909, August 16 2007–July 18 2007.
- [8] B. Hjorth, "An on-line transformation of eeg scalp potentials into orthogonal source derivations," *Electroencephalography and Clinical Neurophysiology*, vol. 39, issue 5, pp. 526–530, November 1975.
- [9] D. Gajic, Z. Djurovic, S. D. Gennaro and F. Gustafsson, "Classification of EEG signals for detection of epileptic seizures based on wavelets and statistical pattern recognition," *Biomedical Engineering: Applications, Basis and Communications*, vol. 26, issue 2, April 2014.
- [10] S. Solhjoo and M. Moradi, "Mental task recognition: a comparison between some of classification methods", *Proceedings of BIOSIGNAL 2004 International EURASIP Conference*, pp. 24 – 26 , 2004.
- [11] H. A. Mohammad, A. Samaha, and K. AlKamha; "Automated Classification of L/R Hand Movement EEG Signals using Advanced Feature Extraction and Machine Learning," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 4, no. 6, 2013.
- [12] S. M. Zhou, J. Q. Gan, and F. Sepulveda, "Classifying mental tasks based on features of higher-order statistics from eeg signals in brain-computer interface", *Information Sciences*, vol. 178, issue 6, pp. 1629–1640, 15 March 2008.
- [13] Y. Chen, Z. Shaobai, "Research on EEG Classification with Neural Networks Based on the Levenberg-Marquardt Algorithm," *Communications in Computer and Information Science*, vol. 308, pp. 195–202, January 2012.
- [14] L. Zhang, W. He, C. He, P. Wang, "Improving mental task classification by adding high frequency band information," *Journal of Medical Systems*, vol. 34, issue 1, pp. 51–60, February 2010.
- [15] J-Z. Xue, H. Zhang, C-x. Zheng, X-G. Yan, "Wavelet packet transform for feature extraction of EEG during mental tasks," *International conference on machine learning and cybernetics*, vol. 1, pp. 360–363, November 2003.

Implementation and Comparison of Routing Protocol RIVER with other VANET Protocols

Md. Hasnat Riaz¹ and Dr. Mohammed Humayun Kabir²
Dept. of Computer Science and Telecommunication
Engineering
Noakhali Science and Technology University,
Noakhali-3814, Bangladesh.
Email: hasnat.riaz90@gmail.com, hkabir.269@gmail.com

Md. Fahidul Islam Shaon³ and Md. Javed Hossain⁴
Dept. of Computer Science and Telecommunication
Engineering
Noakhali Science and Technology University
Noakhali-3814, Bangladesh.
Email: fahidul.cste@gmail.com, javed.nstu@gmail.com

Abstract - In VANET, vehicles communicate with each other and possibly with a roadside infrastructure to provide a long list of applications varying from transit safety to driver assistance and Internet access. In these networks, knowledge of the real-time position of nodes is an assumption made by most protocols, algorithms and applications. Owing to rapid topology changes and frequent disconnection make it difficult to design an efficient routing protocol for routing data among vehicle to vehicle communication and vehicle to road side infrastructure. The existing routing protocols for VANET are not efficient to meet every traffic scenario. Thus design of an efficient routing protocol has taken significant attention so that routing protocol must adapt to the rapidly changing topology of vehicles in motion. In this paper we have analyzed the performance of a newly emerged routing protocol RIVER. We have also compared RIVER with some traditional VANET protocols AODV, DSR and GPSR.

Index Terms - VANET, AODV, DSR, GPSR, RIVER.

I. INTRODUCTION

Several routing protocols have been defined by many researchers for VANET. With the passage of time there is a need of having new protocols in order to have successful communication. The history of VANET routing begins with the traditional MANET [1] [2] routing protocols such as AODV and DSR. AODV and DSR have been considered efficient for Multi hop wireless and ad-hoc networks. However, several topology based routing protocols for MANET had been analyzed for VANET.

O. Abedi et. al [4], [5] enhanced traditional MANET routing protocol AODV to improve route stability and less overhead of network and makes it suitable for VANET, they named it as PAODV and DAODV [4], [5]. Their study showed that more appropriate routes can be found with and without mobility prediction. In their study, they selected fewer routes to overcome routing overhead on network and this effect overcomes the link breakage as compared to AODV.

Moreover, some position based routing protocols had been proposed that need geographic information for the selection of nodes. Karp and Kung [9] proposed a position based routing protocol i.e. greedy perimeter stateless routing (GPSR) that uses geographic information of the nodes which are close to destination in order to forward data packets and make communication successful. But this routing algorithm has some problems when it was evaluated for large city

environments because it uses direct communication and with the existence of obstacles (building) in the city environments may interrupt the communication.

Bernsen et al., proposed a routing protocol Reliable Inter-Vehicular Routing (RIVER) [6] which utilizes an undirected graph that represents the surrounding street layout. Unlike existing protocols, RIVER performs real-time, active traffic monitoring and uses this data and other data gathered through passive mechanisms to assign a reliability rating to each street edge. The protocol then uses these reliability ratings to select the most reliable route. Control messages are used to identify a node's neighbours, determine the reliability of street edges, and to share street edge reliability information with other nodes.

II. BACKGROUND

Ad-hoc On Demand Distance Vector Routing (AODV) [10] is reactive routing protocol. It belongs to multihop type of reactive routing. AODV routing protocol works on demand basis when it is required by network. Route discovery and route maintenance is also carried out on demand basis even if only two nodes need to communicate with each other. It maintains and discovers routes only when there is a need of communication among different nodes. It uses an efficient method of routing that reduces network load by broadcasting route discovery mechanism and by dynamically updating routing information at each intermediate node. AODV reduces several problems that occurred in proactive routing protocols. AODV provide support by reacting at on demand needs for communication for such ad-hoc network where large numbers of nodes and this can help when the sudden change in topology happens. It updates the information of active nodes in the routing table. This feature can help maintaining the routing tables with the related number of entries and nodes only have the information of currently active routes for communication. It reduces flooding of messages in the network as compared to proactive routing protocols so AODV reduces the network overhead. AODV also minimizes the route redundancy and large memory requirements.

Dynamic Source Routing protocol (DSR) [11], designed for multi-hop wireless ad-hoc networks. This protocol consists of two operations "Route Discovery" and "Route Maintenance" that makes it self-configuring and self-

organizing. DSR routing protocol manage the network without any centralized administrator or infrastructure. In route discovery this protocol discovers for the routes from source node to destination. In DSR, data packets stored the routing information of all intermediate nodes in its header to reach at a particular destination. Routing information for every source node can be changed at any time in the network and DSR updates it after each change occurs. Intermediate routers don't need to have routing information to route the passing traffic, but they save routing information for their future use.

An important property of DSR routing protocol is network type flexibility. A packet using DSR routing protocol can reach its destination even when the intermediate nodes are using different type of networks. DSR makes it possible that nodes with different networks can participate in ad-hoc network, DSR protocol considers them as Ad-hoc Network.

Greedy Perimeter Stateless Routing (GPSR) [9] is one of the best examples of position based routing. GPSR uses closest neighbour's information of destination in order to forward packet. This method is also known as greedy forwarding. In GPSR each node has knowledge of its current physical position and also the neighbouring nodes. The knowledge about node positions provides better routing and also provides knowledge about the destination. On the other hand neighbouring nodes also assists to make forwarding decisions more correctly without the interference of topology information. All information about nodes position gathered through GPS devices. GPSR combines the greedy forwarding with the perimeter forwarding to provide better routing decision on both full and planarized network graph by maintaining neighbour's information in the location table.

Reliable Inter-Vehicular Routing (RIVER) [6] is a position-based, greedy, vehicle-to vehicle routing protocol for vehicular ad-hoc networks. This protocol prefers transmitting messages using routes along paths it seems to be reliable through its traffic monitoring components. This traffic monitoring happens in real-time by actively sending probe messages along streets and by passively monitoring messages that are transmitted between adjacent intersections. Furthermore, RIVER takes traffic monitoring a step further by propagating reliability information within the network. Street reliability data is distributed in a more localized manner by piggybacking the information on routing messages, probes and beacons. This protocol allows routes to be recalculated dynamically at any anchor point during message transmission. The recovery option in RIVER is a reactive mechanism that engages when a forwarding node on a route encounters a network gap in the direction it intended to forward the data packet. When this happens, a new route is calculated and compared to the existing one. If the new route is estimated to be more reliable, the data packet's route header is rewritten, and the packet is forwarded along the new route. The recalculation option in RIVER is a proactive mechanism that evaluates a data packet's route when it is received by forwarding node that is within range of a vertex. If the forwarding node determines that a significantly more reliable

route is available, the route in the packet is overwritten, and the packet is forwarded along the new route. RIVER performance evaluation includes a *combined* strategy that proactively recalculates routes and also reacts with the recovery option if a network gap is encountered.

III. SIMULATION

For simulation the Operating system Ubuntu 12.04 LTS and NS-2 version 2.35 with CMU wireless extensions were used. The RIVER utilizes the Brad Karp's location database that is installed with GPSR and a few code extensions that are installed with GPSR [3]. Firstly GPSR code was installed so that RIVER may utilize locdbase.cc/h. For these simulations, an urban "Manhattan" street grid was used with 5 streets running in the horizontal and vertical directions spaced approximately 400 m apart for a total area of approximately 6.05 km². This simulation area was populated with varying traffic densities of 10 to 300 vehicular network nodes. Vehicles travelled in both directions along each street, and vehicles may turn at intersections. The connection pattern consisted of sender/receiver pairs using constant bit rate (CBR) data flows of 512 Kbps. That is, each sender transmitted a 512 byte packet and total 105 packets sent during each simulation. Each packet sent was offset by at least one second from the previous send to ensure that no senders transmitted simultaneously. We generated CBR traffic between five randomly selected vehicles. Total simulation time was 200 seconds. For comparison, protocols were determined per the criteria (ratio between the transmitted and received packets, amount of data per second, Number transmissions of data packets and loss of data). For graphical representation we have used MATLAB.

A. Data Packet Throughput

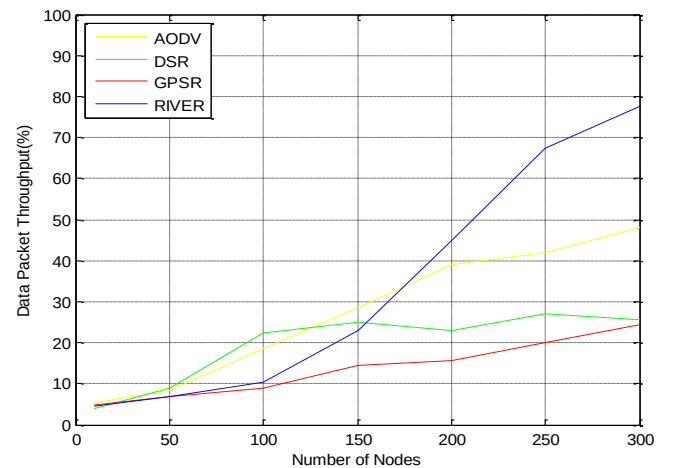


Fig. 1: Packet Delivery Ratio as a function of nodes.

Figure 1 shows that the reception of data packet in RIVER protocol is increasing according to the number of vehicles is increasing. Though GPSR is popular but it shows worse performance for these measurements. Protocol RIVER, among

the selected protocols, is successful in the delivery of data packets to the target vehicle. AODV protocol is the best for the smallest population of vehicles. For the protocol GPSR, it is necessary to increase the density of vehicles to improve the delivery of data packets.

B. The number of unnecessarily generated paths for routing

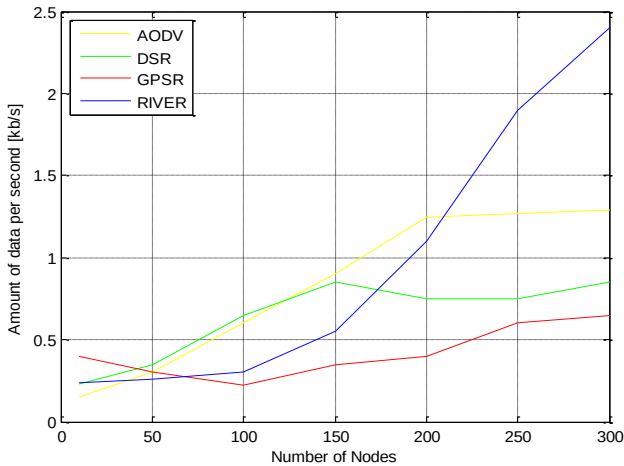


Fig. 2: Routing Overhead depending on the number of nodes

From figure 2 we can see that traditional MANET protocol (AODV and DSR) adapted for use in a network of vehicles that do not have a lot of route calculations. Their paths count only on request and with the help of this welcome message. GPSR protocol for each hole in the network, re-calculates the route to the destination node. Curve RIVER Protocol on the graph shows very promising results. Protocol RIVER, in a hopeless position again using Dijkstra’s algorithm, returns the packet to the last anchor point and from there it determines the path to the target node. However, we want that the network is not transmitting excessive amount of unnecessary data (which affects the utilization of transmission media) and this measurement protocol GPSR is worse.

C. The amount of data per second

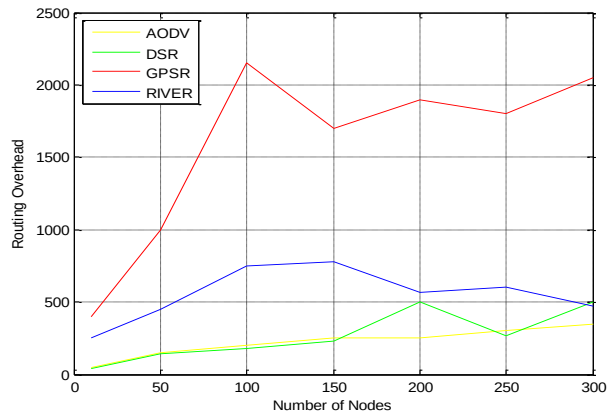


Fig. 3: Amount of data sent over the air depending on the number of nodes.

The figure 3 shows a considerable amount of data per second is transferred by protocol RIVER. This means greater accuracy by determining and computing the path to the target node. The GPSR protocol for these metric shows bad statements. GPSR, DSR and AODV show that even with increasing density of vehicles they wouldn’t improve the amount of data per second.

D. Loss of data packets

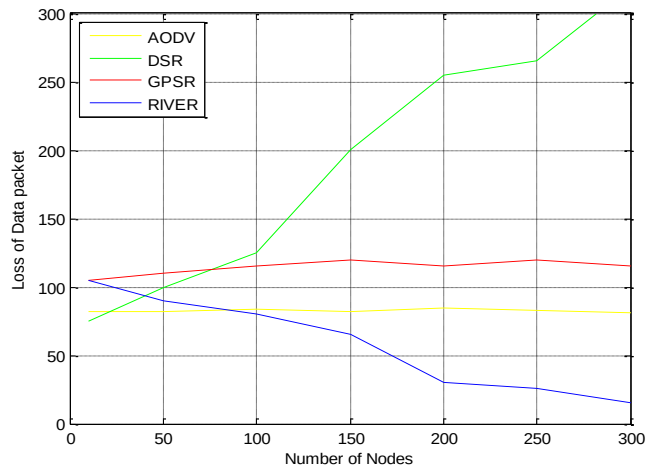


Fig. 4: Data Packet Loss as per number of vehicles.

DSR protocol for this metric shows worse performance. Protocol AODV and GPSR show satisfactory results. Protocol RIVER, with this metric shows good results for densely populated nodes.

E. RIVER Performance Analysis

We have analyzed the performance of RIVER by measuring data packet throughput, forwards per route and route transit time. All of them have included recovery and recalculation strategies.

E.1 Data Packet Throughput

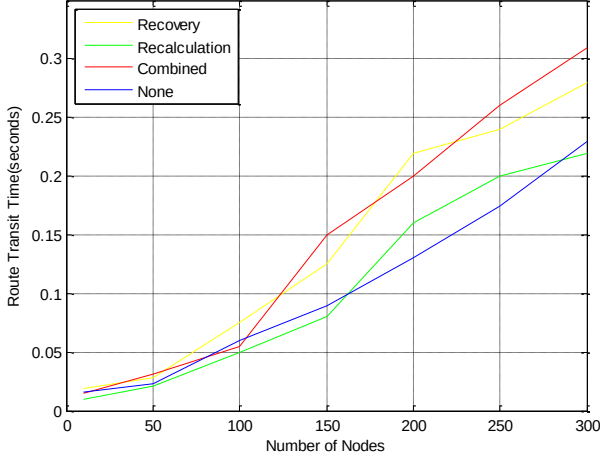


Fig. 5: Data Packet Throughput with Recovery and Recalculation Strategies

Data Packet Throughput represents the percentage of routed data packets that were successfully delivered. Figure 8 is the corresponding graph of throughput verses number of nodes. From figure 8 we find that recovery strategy delivers the best throughput, while combined strategy performs nearly as well. However, we also find that in high-density traffic scenario RIVER delivers best throughput and in low-density traffic scenario its throughput is low.

E.2 Forwards per Route

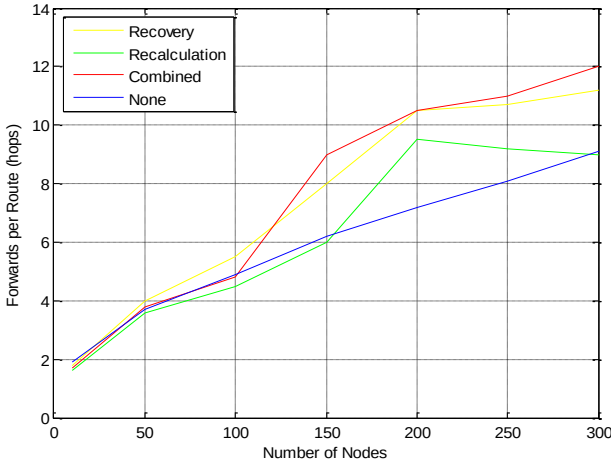


Fig. 6: Forwards per Route with Recovery and Recalculation Strategies

Forwards per Route represent the average hop count of a routing packet. In Figure 6, we find that the recalculation strategy requires the fewest number of hops per route (excluding none strategy). From figure 9 we also find that in low node density it requires few numbers of hops per route.

E.3. Route Transit Time

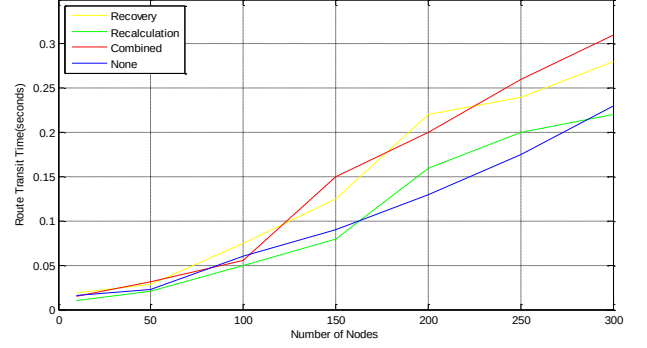


Fig. 7: Route Transit Time with Recovery and Recalculation Strategies.

Route transit time represents the number of seconds required to deliver a routing packet from its original sender to its final destination. In Figure 10, we observe that the recalculation strategy delivers route packets more quickly than either the recovery or combined strategies. From figure 10 we also find that in low-density traffic scenario the RIVER delivers route packets more quickly.

VII. CONCLUSION

In this paper we have examined several routing protocols strengths and weaknesses. Then we compared RIVER protocol with the protocols AODV, DSR and GPSR. The results of the comparisons showed that the RIVER due to its components traffic control is a great potential. The comparison also shows that the protocols at low traffic density are serving poorly. In case of RIVER from simulation we found that RIVER provides the highest throughput in most traffic densities when using its recovery strategy, but the recalculation strategy yields higher throughput in low traffic density with less overhead. We also found that RIVER's reliability distribution components perform the best in an average to high density scenarios. During the design of RIVER protocol its performance in low-density traffic areas was not considered, so this is an area where its performance could be enhanced.

REFERENCES

- [1]. R. Bai, M. Singhal, BRD: Bilateral Route Discovery in Mobile Ad Hoc Networks, in: Proceedings of the 6th International IFIP-TC6 Networking Conference (Lecture Notes in Computer Science Vol.4479), 2007.
- [2]. R. Bai, M. Singhal, DOA: DSR Over AODV Routing for Mobile Ad Hoc Networks, IEEE Transactions on Mobile Computing 5 (10) (2006) 1403–16.
- [3]. B. Karp, H. T. Kung, GPSR: Greedy Perimeter Stateless Routing for Wireless Networks, in: Proceedings of the 6th Annual International Conference on Mobile Computing and Networking (MobiCom '00), ACM, New York, NY, USA, 2000.
- [4]. O. Abedi, R. Berangi, and M. A. Azgomi, “Improving Route Stability and Overhead on AODV Routing Protocol and Make it Usable for VANET,” *Distributed Computing Systems Workshops, 2009. ICDCS Workshops '09. 29th IEEE International Conference on* pp. 464 - 467, 22-26 June 2009.
- [5]. O. Abedi, M. Fathy, and J. Taghiloo , “Enhancing AODV routing protocol using mobility parameters in VANET,” *Computer Systems and Applications, 2008. AICCSA 2008. IEEE/ACS International Conference on* pp. 229 - 235, March 31 2008-April 4 2008.
- [6]. J. Bernsen, A Reliability-Based Routing Protocol for Vehicular Ad-Hoc Networks, Master’s thesis, University of Kentucky, 2011.
- [7]. E.W. Dijkstra, A Note on Two Problems in Connexion with Graphs, *Numerische Mathematik* 1 (1959) 269–271, 10.1007/BF01386390. URL <http://dx.doi.org/10.1007/BF01386390>
- [8]. Dijkstra’s Algorithm, URL: http://en.wikipedia.org/wiki/Dijkstra%27s_algorithm
- [9]. Karp, B. and Kung, H. T., “GPSR: greedy perimeter stateless routing for wireless networks”, In *Proceedings of the 6th Annual international Conference on Mobile Computing and Networking* (Boston, Massachusetts, United States, August 06 - 11, 2000). MobiCom '00. ACM, New York, NY, pp. 243-254.
- [10]. C. Perkins and E. Royer, “Ad-hoc on-demand Distance Vector Routing,” *Proc. 2nd IEEE Wksp. Mobile Comp. Sys. App.*, Feb. 1999, pp. 90–100.
- [11]. David B. Johnson, David A. Maltz, and Josh Broch, “DSR: The Dynamic Source Routing Protocol for Multi-Hop WirelessAd Hoc Networks”, in *Ad Hoc Networking*, Editor: Charles E.Perkins, Chapter 5, pp. 139-172, Addison-Wesley, 2001

Abnormal Mass Classification in Breast Mammography using Rotation Invariant LBP

Haider Adnan Khan, Abdullah Al Helal, Khawza I Ahmed and Raqibul Mostafa

Department of Electrical and Electronic Engineering
United International University, Dhaka 1209, Bangladesh

Email: {hkhan, helal, khawza, rmostafa}@eee.uiu.ac.bd

Abstract—We present a novel approach for abnormal breast mass classification from digitized mammography images. The proposed framework exploits rotation invariant uniform Local Binary Pattern (LBP) as texture feature. These features are classified using Support Vector Machine (SVM). In addition, we take advantage of the breast mammograms taken from multiple views or angles. We classify breast scans from ‘cranial-caudal’ view and ‘mediolateral-oblique’ view separately, and combine these classification scores to make an improved diagnosis. This reduces the classification error, and achieves higher recognition rate than that of either views individually. The proposed computer aided diagnosis system was evaluated on DDSM (Digital Database for Screening Mammography) data set, and was able to achieve a classification accuracy of 74%.

Keywords—Local Binary Pattern, Rotation Invariant LBP, Breast Cancer, Mammography, Computer Aided Diagnosis

I. INTRODUCTION

Breast cancer is the most common form of cancer among women. At present, there is no known method to prevent breast cancer. However, early detection and diagnosis can greatly increase the chance of cure. The most clinically accepted imaging for breast cancer is screening mammography. Unfortunately, detection of suspicious abnormalities by screening mammography is not accurate. These errors are largely due to human factors involved in the visual inspection by radiologists. This motivates automated screening of mammograms or Computer Aided Diagnosis (CAD).

An excellent study of the CAD systems can be found in [1]. Sheshadri and Kandaswamy used a statistical measure for the texture of the breast lesions to obtain an accuracy of 80% for breast tissue classification [2]. Moayedi et al. exploited contourlet transform based features with Support Vector Machine (SVM) to obtain an accuracy of 96.6% in breast mass classification [3]. Verma et al. proposed a cluster based classification framework that used a direct learning classifier based on neural network [4]. Their framework demonstrated 97.5% accuracy in classifying benign and malignant pattern in digital breast mammography.

Mammograms are typically taken from two to four angles or views [5]. The two most standard mammographic views are ‘cranial-caudal’ or CC view, and ‘ediolateral-oblique’ or MLO view. In CC-view, breast scans are taken from above, while in MLO-view, scans are taken from side at an angle. Multiple view breast scanning is preferred as they complement each other, and is useful for better diagnosis. In fact, inspecting two views rather than a single view has been proven to improve the classification accuracy for CAD tools [6].

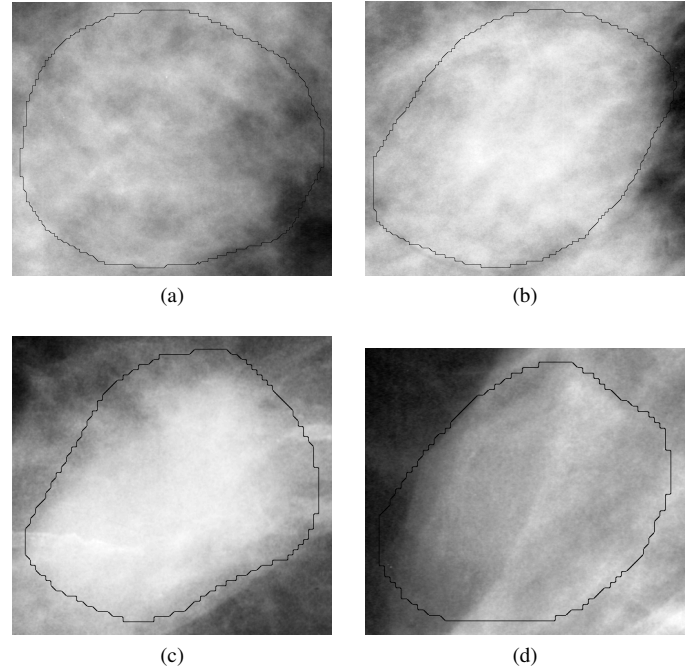


Fig. 1. Multiples views of mammographic scans, the top row (a) and (b) correspond to the same benign mass from CC-view and MLO-view respectively, the bottom row (c) and (d) correspond to the same malignant mass from CC-view and MLO-view respectively.

In this paper, we introduce a novel framework that exploits Local Binary Pattern (LBP) as a texture feature for abnormal breast mass classification. LBP is a texture descriptor that is widely used in many computer vision applications. We use rotation invariant LBP with uniform pattern for improved texture classification. To our best knowledge, this feature has never been used for breast cancer diagnosis. In addition, we reduce recognition error by taking advantage of the multi-view breast mammography. We classify breast scans from CC-view and MLO-view separately, and combine the classification scores to achieve an improved classification accuracy.

The rest of the paper is organized as follows: in Section II, we briefly describe LBP and few of its variants that are used in our framework. Section III depicts our proposed computer aided diagnosis system in detail, while experimental setup is explained in Section IV. In Section V, results are analyzed, and finally, concluding remarks are made in Section VI.

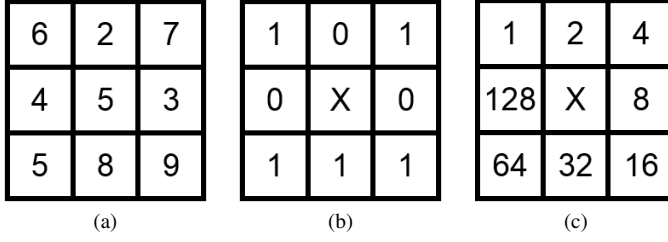


Fig. 2. Illustration of Local Binary Pattern (LBP) calculation (a) example image, (b) binary pattern by thresholding and (c) weight matrix.

II. LOCAL BINARY PATTERN

Local Binary Pattern (LBP) is a powerful texture descriptor [7]. It describes the local texture pattern, and can be used for texture classification. LBP is computationally simple, which makes it suitable for real-time applications. LBP is also robust against varying illumination conditions. Consequently, it has been used in a number of computer vision applications including face recognition [8], handwritten character recognition [9], human detection [10], moving object detection [11] and medical image analysis [12].

LBP compares each pixel with its neighbor pixels, and accordingly generates a binary pattern or code. Each comparison with a neighbor pixel results into a binary digit: '0' or '1', where '0' indicates that the gray scale value of the neighbor pixel is lower than that of the center pixel, while '1' indicates the contrary. Next, a bit stream of binary digits is concatenated by comparing all the neighbor pixels. Finally, this binary number is converted into a decimal number. This decimal number is called the LBP value of the center pixel. This procedure is explained with the following expressions:

$$s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \quad (1)$$

$$LBP = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p \quad (2)$$

where $s(x)$ is the threshold function, and g_c and g_p represent the gray-scale value of the center pixel and the p -th neighbor respectively. Hence, the binary value of the p -th neighbor is multiplied by its corresponding weight 2^p . LBP is computed by summing up these weighted binary values. This is illustrated in Fig. 2 which shows that the binary pattern 01110101 is converted to $LBP = 1 + 4 + 16 + 32 + 64 = 117$.

Once the LBP values for all pixels are computed, a histogram of these LBP values is generated. This histogram can be used as a local texture descriptor for texture matching or classification. Instead of computing one histogram from the entire image, the image is usually split into multiple smaller blocks. Then, the histogram for each block is computed separately. This ensures that the locality of the texture pattern is not lost.

In its most basic form, LBP can be computed in a 3×3 neighborhood with only eight adjacent pixels. This results into $2^8 = 256$ unique binary patterns. However, many different

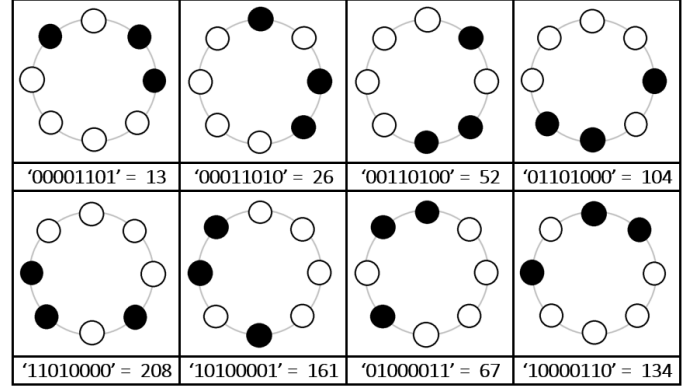


Fig. 3. Illustration of Rotation Invariant Local Binary Patterns

variants of LBP are possible by varying the radius (distance), orientation and the number of neighbors (samples). In the following section, we briefly describe a few variations of LBP that are exploited in our proposed method.

A. Uniform LBP

Uniform Local Binary Pattern is a popular extension of basic LBP. It is observed that some binary patterns are more frequent than the others [13]. Binary patterns with more than two transitions from '0' to '1' or '1' to '0' are extremely uncommon in texture images. These patterns are defined as 'non-uniform' patterns. Whereas, patterns with at most two transitions are defined as 'uniform' patterns. For instance, patterns such as '00000000' and '11111111' have no transitions. Likewise, '00000111' and '11110000' have only one transition. Similarly, '11100111' and '00110000' have two transitions. These are all examples of 'uniform' patterns. On the other hand, '00111001' and '01010011' have three and five transitions respectively. These are 'non-uniform' patterns. Out of 256 basic binary patterns, only 58 are 'uniform', and the rest are 'non-uniform', and appears scarcely in natural textures. It is argued that these 'non-uniform' patterns occur so rarely that their probabilities cannot be estimated accurately. Hence, in uniform LBP, only 'uniform' patterns are given unique labels whereas all 'non-uniform' patterns are tallied under the same label. This also leads to feature dimension reduction from 256 features to $58 + 1 = 59$ features.

B. Rotation Invariant LBP

In many pattern recognition problems, the target patterns may not have any specific unique orientation. On the contrary, the identical patterns may appear in different orientations. As such, the pattern descriptors must be robust enough so that the patterns can be identified irrespective of their orientation. This motivates rotation invariant local binary pattern [13]. Even though, there are 256 unique binary patterns in an eight neighbor basic LBP, many of these patterns can be replicated by simply rotating other patterns around the center. This is illustrated in the Fig. 3. Hence, from rotation invariance point of view, all such patterns should be considered and labeled as one identical pattern. In Rotation Invariant LBP, all patterns that can be generated by circular bit-wise right shift are labeled with the lowest LBP value of those patterns. This corresponds 13, in case of Fig. 3. Apart from achieving rotation invariance,

this approach also reduces the feature dimension significantly from 256 features to only 36 rotation invariant features.

C. Rotation Invariance with Uniform Pattern

It was observed that for a 3×3 neighborhood (or 8 adjacent neighbor) LBP, about 90% of the rotation invariant patterns were uniform pattern [13]. Hence, rotation invariance was combined with uniform pattern. This resulted into improved texture classification. This also leads to further feature dimension reduction. For a 3×3 neighborhood, there exists only 10 rotation invariant uniform patterns. So, the feature dimension is reduced to 10 instead of 256.

In the next section, we describe our proposed method for abnormal breast mass classification in mammographic screening, and explain how we exploit LBP in our proposed framework.

III. PROPOSED METHOD

In our proposed method, we use both CC and MLO views to classify abnormal breast masses. First, Rotation Invariant Uniform LBP features are extracted separately from both CC and MLO view scans. Next, these extracted features are classified using Support Vector Machines (SVM). This gives us two separate scores for CC and MLO views. In the final stage, we combine these two scores to predict the diagnosis of an abnormal breast mass more accurately. This framework is illustrated in Fig. 4. In the subsequent sections, we explain each block of this block diagram in further detail.

A. Feature Extraction

In our framework, we use rotation invariant LBP combined with uniform patterns. Rotation invariance is critical as breast masses do not have any preferred orientation. Hence, similar textures may appear in many different orientations, and we need to classify them as the same diagnosis. We also exploit the idea of uniform patterns. As mentioned in the previous section, uniform patterns has demonstrated improved results. It should also be noted that unlike usual practice, we did not split our breast scans into multiple smaller blocks for generating LBP histograms. Thus, we have only one histogram for the entire breast scan. In typical texture classification tasks, different parts of the image have different textures. For instance, in face recognition, eyes and lips have completely different textures. Hence, the image should be split into smaller blocks so that locality of the texture is not lost, that is eyes are matched with eyes and lips are matched with lips. However, in our case, the breast masses do not have any definitive location or position. Therefore, multiple smaller blocks did not have any advantage. In fact, using multiple blocks would lead to increase in the feature dimension (by a factor of the number of blocks), which in turn, may lead to over-fitting.

Another important issue is that, different breast masses have different sizes - from very small to very large. Hence, the ROI of the images can vary greatly in size. Therefore, many feature extraction schemes may end up with different number of features for different breast masses. This would cause a difficult problem for the classifier. However, using LBP histogram as the feature, we always generate a fixed number of features (in our case 10 features). This is another advantage of LBP.

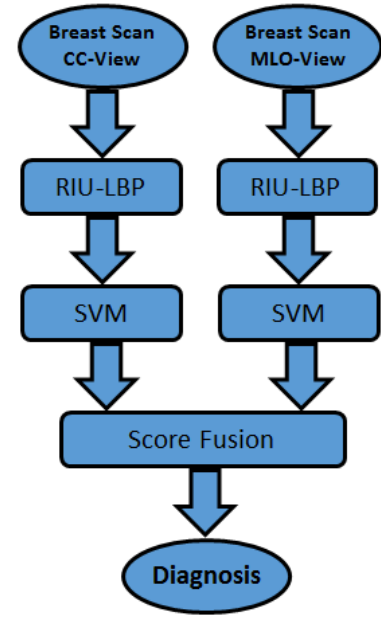


Fig. 4. Block diagram of the proposed computer aided diagnosis system.

B. Support Vector Machine

We use Support Vector Machine (SVM) as the classifier. SVM is a well-known classifier. It exploits kernels in order to implicitly map inputs into high dimensional feature space. This enables SVM to efficiently perform non-linear classification using linear decision boundaries. SVM is robust against noise [14] and outliers, which motivates us to exploit SVM for noisy mammograms. In our experiments, we have employed Radius Basis Function (RBF) as the kernel. This implements a Gaussian function to map the data samples non-linearly to a higher dimensional vector space in order to separate the classes with a linear decision boundary. Given an input vector X_i , a trained SVM classifier outputs a scalar y_i which can have values between 0 and 1. This is typically interpreted as class 0 or class 1 by applying a threshold at 0.5. However, in our case, we use this score in the next stage of score fusion.

C. Score Fusion

In the final stage of this computer aided diagnosis, we combine the two scores from two different scans of the same abnormal breast mass, namely CC-view and MLO view. Even though these two views share some degree of redundancy, they also reveal unique perspective, and can complement each other. Hence, exploiting multiple views for classification could be advantageous. Also, combining two classification scores somewhat safeguards against over-fitting.

$$Score = \frac{1}{2}(Score_{cc} + Score_{mlo}) \quad (3)$$

$$Diagnosis = \begin{cases} \text{Malignant} & \text{if } Score \geq 0.5 \\ \text{Benign} & \text{if } Score < 0.5 \end{cases} \quad (4)$$

For score combination, we use simple arithmetic average of the two scores - CC-view score and the MLO-view score as

shown in Eq. (3). Finally, this combined score is converted into benign (class-0) or malignant (class-1) diagnosis by applying thresholding. This is expressed in Eq. (4).

IV. EXPERIMENTS

A. Data Set

In order to evaluate our method, we used DDSM (Digital Database for Screening Mammography) dataset [15], [16]. DDSM is a popular resource used by the mammographic image analysis research community. This is the largest repository of mammographic digitized annotated images with pathology proven diagnosis whenever applicable. The database contains benign and malignant tumor scans with pixel-level location of ROI among other ground truths. Out of the 2165 breast mass scans, 1081 are benign masses and 1084 are malignant. These masses correspond to 1140 different patients (or cases). Most of patients have both CC-view scan and MLO-view scan, enabling us to combine them for more accurate diagnosis.

B. Preprocessing

The gray level images were acquired from different scanners. To remove the dependency and effect of the scanners, calibration was done by mapping the gray levels to optical density as per instructed in the database.

C. Segmentation

Each mammography image is accompanied with an overlay file that contains the metadata about the scan including diagnosis, Bi-RAD scores and the outlines of the breast mass marked by the radiologists. These outlines or contours are given in chain code format. We decoded these chain codes and mapped them into their corresponding ROI.

D. Train, Validation and Testing

We have randomly divided the data set into training, validation and test sets in an 80% – 10% – 10% fashion according to stratified sampling. We ensured that samples of the same patient did not fall into different sets. For classification, we implemented SVM with RBF as the kernel using LIBSVM. There are two model parameters in an RBF SVM: C and gamma. For the best model selection, i.e., parameter search, we follow the grid search suggested in [17].

V. RESULTS

Experimental results demonstrate that both CC-view scan and MLO-view scan achieve similar classification accuracy - 70% and 71% respectively. This is promising as this accuracy is achieved using only 10 features. We also observe a significant amount of increase in accuracy by combining two scores. This validates our assumptions of better classification through multi-view classification.

TABLE I. ABNORMAL BREAST MASS CLASSIFICATION

View	Accuracy
CC	70%
MLO	71%
Combined	74%

VI. CONCLUSION

We propose a framework that exploits rotation invariant LBP with uniform patterns for computer aided diagnosis of abnormal breast masses from mammography images. The results reveal the strength of Local Binary Pattern as a texture descriptor. It likely that higher accuracy can be achieved by combining LBP features with traditional statistical features or morphological features. Furthermore, we combine scores from different views or angles to reduce the classification error. In future, different ways for score fusion, such as weighted average or geometric mean, can be explored for achieving better results.

REFERENCES

- [1] J. Tang, R. M. Rangayyan, J. Xu, I. El Naqa, and Y. Yang, "Computer-aided detection and diagnosis of breast cancer with mammography: recent advances," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 13, no. 2, pp. 236–251, Mar. 2009. [Online]. Available: <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=4757273>
- [2] H. Sheshadri and A. Kandaswamy, "Experimental investigation on breast tissue classification based on statistical feature extraction of mammograms," *Computerized medical imaging and graphics*, vol. 31, no. 1, pp. 46–48, Jan. 2007. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0895611106001030>
- [3] F. Moayed, Z. Azimifar, R. Boostani, and S. Katebi, "Contourlet-based mammography mass classification using the SVM family," *Computers in Biology and Medicine*, vol. 40, no. 4, pp. 373–383, Apr. 2010. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0010482510000041>
- [4] B. Verma, P. McLeod, and A. Klevansky, "Classification of benign and malignant patterns in digital mammograms for the diagnosis of breast cancer," *Expert systems with applications*, vol. 37, no. 4, pp. 3344–3351, 2010. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0957417409008823>
- [5] S. W. Fletcher and J. G. Elmore, "Mammographic screening for breast cancer," *New England Journal of Medicine*, vol. 348, no. 17, pp. 1672–1680, 2003.
- [6] R. Warren, S. Duffy, and S. Bashir, "The value of the second view in screening mammography," *The British journal of radiology*, vol. 69, no. 818, pp. 105–108, 1996. [Online]. Available: <http://www.birpublications.org/doi/abs/10.1259/0007-1285-69-818-105>
- [7] M. Pietikäinen, A. Hadid, G. Zhao, and T. Ahonen, *Computer vision using local binary patterns*. Springer, 2011, vol. 40.
- [8] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face description with local binary patterns: Application to face recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 12, pp. 2037–2041, 2006.
- [9] T. Hassan and H. A. Khan, "Handwritten bangla numeral recognition using local binary pattern," in *Electrical Engineering and Information Communication Technology (ICEEICT), 2015 International Conference on*. IEEE, 2015, pp. 1–4.
- [10] X. Wang, T. X. Han, and S. Yan, "An hog-lbp human detector with partial occlusion handling," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 32–39.
- [11] M. Heikkilä and M. Pietikäinen, "A texture-based method for modeling the background and detecting moving objects," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 4, pp. 657–662, 2006.

- [12] L. Nanni, A. Lumini, and S. Brahnam, "Local binary patterns variants as texture descriptors for medical image analysis," *Artificial intelligence in medicine*, vol. 49, no. 2, pp. 117–125, 2010.
- [13] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 7, pp. 971–987, 2002.
- [14] H. Xu, C. Caramanis, and S. Mannor, "Robustness and regularization of support vector machines," *The Journal of Machine Learning Research*, vol. 10, pp. 1485–1510, Dec. 2009. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1755834>
- [15] K. Bowyer, D. Kopans, W. Kegelmeyer, R. Moore, M. Sallam, K. Chang, and K. Woods, "The digital database for screening mammography," in *Third international workshop on digital mammography*, vol. 58, 1996, p. 27.
- [16] M. Heath, K. Bowyer, D. Kopans, P. Kegelmeyer Jr, R. Moore, K. Chang, and S. Munishkumaran, "Current status of the digital database for screening mammography," in *Digital mammography*. Springer, 1998, pp. 457–460.
- [17] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A practical guide to support vector classification," Tech. Rep., 2003. [Online]. Available: <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>

Optimizing the Parameters of Plasmonic Metal Nanoparticles to Maximize the Energy Conversion Efficiency of Thin-Film Solar Cells

Saniat Ahmed Choudhury

Department of Electrical and Electronic Engineering
Independent University, Bangladesh
Dhaka, Bangladesh
Email: ac.saniat@gmail.com

Mustafa Habib Chowdhury

Department of Electrical and Electronic Engineering
Independent University, Bangladesh
Dhaka, Bangladesh
Email: mchowdhury@iub.edu.bd*

Abstract— Photovoltaic (PV) cells are important devices that can be used to convert photons into electrical energy. However, the cost of current PV devices needs to be significantly reduced and their efficiencies need to be substantially enhanced to enable large scale implementation of PV cells as sources of clean electrical energy. Surface plasmon polaritons (SPP) in metallic nanostructures in close proximity to a Si substrate can lead to enhanced optical absorption and improved energy conversion efficiency in amorphous silicon thin-film PV devices. Optimization of this phenomenon requires a detailed characterization and understanding of the parameters that govern the underlying physical mechanisms. This paper explores the relationship between energy conversion efficiency in thin-film Si solar cells and the type of metallic nanoparticle, the size of the metallic nanoparticles, and the distance between neighboring metallic nanoparticles in a periodic array of the metallic nanoparticles deposited on top of the Si substrate. Of the different metallic nanoparticle systems studied, it is found that silver nanoparticles with diameter of 100 nm and spaced 220 nm from neighboring nanoparticles in a periodic array of the metallic nanoparticles deposited on top of the Si substrate provides the maximum increase in the generated short circuit current density (J_{sc}).

Keywords—photovoltaics, plasmonics, thin-film solar cells, light trapping, nanoparticles, surface plasmons, silver nanoparticles.

I. INTRODUCTION

Photovoltaic (PV) cells are important devices that can be used to convert photons into electrical energy [1]-[4]. PV cells can thus be used to provide virtually unlimited amounts of electrical energy by efficiently converting sunlight into clean electrical energy [1]-[4]. However, the cost of current PV devices needs to be significantly reduced and their efficiencies need to be substantially enhanced to enable large scale implementation of PV as sources of clean and renewable electrical energy [1]. Amorphous thin-film Si solar cells can provide an opportunity towards realizing this goal because of their low materials and processing costs. Unfortunately, the materials quality and resulting energy conversion efficiencies of amorphous thin-film Si solar cells are still substantially lower than crystalline, wafer-based silicon [1]-[4]. The low conversion efficiencies result due to the large mismatch

between the electronic and photonic length scales in these devices [1]-[4]. If light absorption could be improved in ultra-thin layers of amorphous Si substrate, it could directly lead to higher energy conversion efficiencies.

Surface plasmon polaritons (SPP) in metallic nanostructures have generated significant interest in the past 30 years for a variety of applications involving the large electromagnetic fields created due to the presence of an SPP resonant excitation in the metal, the pronounced forward scattering of incident radiation at wavelengths (frequencies) in the vicinity of SPP resonance, or other related effects [5] – [9]. One recent study demonstrates the enhanced optical absorption and photocurrent generation in semiconductor photodiodes induced by scattering from SPP resonances in metallic nanoparticles deposited on the photodiode surface [6],[8]. Other studies also extend the application of this phenomenon to achieve improved energy conversion efficiency in amorphous silicon thin-film PV devices [4], [10]. Optimization of this phenomenon will require detailed characterization and understanding of the relevant underlying physical mechanisms and the parameters that govern such physical mechanisms. To this end, this study explores the relationship between energy conversion efficiency in thin-film Si solar cells and type of metallic nanoparticle, the size of the metallic nanoparticle, and the distance between neighboring metallic nanoparticles in a periodic array of the metallic nanoparticles deposited on top of the Si substrate.

II. SIMULATION

A. Setup

The computer system used to perform our simulations had the following specifications:

Operating system: Windows 7

Processor: Intel(R) Core(TM) i7-4770 CPU @ 3.40GHz

Memory: 8192MB RAM

GPU: Intel(R) HD Graphics 4600

* Corresponding Author

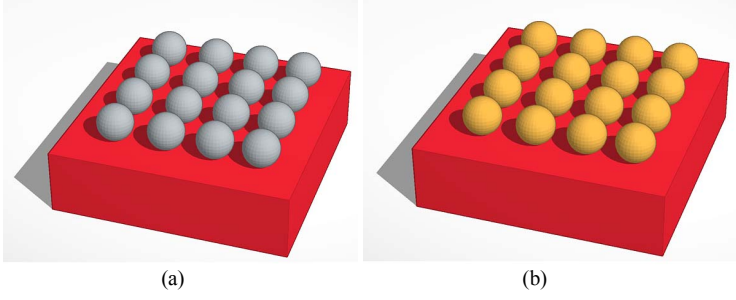


Figure 1. Schematic diagram showing: (a) Ag nanoparticle array on 2 μm thick Si wafer; and (b) Au particles on 2 μm thick Si wafer

The simulation software that we used to run our analysis is FDTD Solutions for the optical simulations, and DEVICE for the calculation of short circuit current density. Both solvers were developed by Lumerical Solutions Inc. FDTD Solutions utilize the finite difference time domain algorithm to accurately approximate the solutions to Maxwell's (differential) equations for a given system that involves electromagnetic radiation.

B. Simulation

The entire analysis has been subdivided into three major sections. All three sets of calculations were performed utilizing the standard solar spectral irradiance of AM1.5G [13].

For the first section, we generated the absorption, scattering and extinction spectra for different diameters of Ag and Au nanoparticles. The diameters that were of interest to us were $D = 50\text{nm}$, 100nm and 200nm . In order to achieve this, we placed field monitors on all four sides of the nanosphere, a total field scattered field (TFSF) source after that, and finally a second set of monitors beyond the source. This setup has been illustrated in Figure 2(a). The first set of monitors were used to calculate the absorption spectra, while the second set was used to generate the scattering spectra. The two were combined to calculate the extinction spectra. The analysis was performed over a wavelength range of 300nm to 1100nm . This range was chosen in order to allow us to analyze the spectra over a wide range and thus provide us with a better and more accurate picture of the plasmon resonance phenomenon.

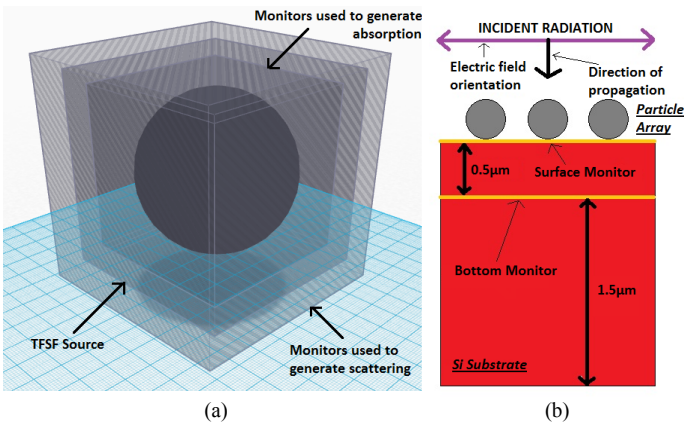


Figure 2. (a) Placement of monitors and source to generate absorption, scattering and extinction spectra, (b) system designed to simulate the nanoparticle array over the Si substrate

Next, we periodically placed nanospheres of Au and Ag respectively over a thin-film Si wafer of thickness $2\text{ }\mu\text{m}$ (the geometric illustrations are shown in Figure 1(a) and 1(b)). Monitors were placed on the surface of the substrate and $0.5\text{ }\mu\text{m}$ below the surface, as shown in Figure 2(b). Their purpose is to calculate transmission of the radiation at those regions to generate the absorption data. We initially kept the pitch, which is the distance from one nanoparticle to the next, constant at 300nm (center-center) while varying the diameter (50nm , 100nm and 200nm) of the nanospheres to generate the absorption enhancement graph of the Si wafer, and the short circuit current density (J_{SC}). The wavelength range was initially set to 400nm to 1100nm . We essentially defined the absorption enhancement factor (g) as the ratio of absorption in Si wafer with the metal nanoparticle array to the absorption of the bare Si wafer. The above procedure was then repeated for a shortened wavelength range of 400nm to 700nm to analyze any possible the variations in J_{SC} . We then changed the pitch to 220nm (center-center), and repeated the same process as with the previous analysis for both wavelength ranges of 400nm to 1100nm and 400nm to 700nm , respectively.

Finally, we generated near-field enhancement images for each of the respective structures to analyze the interaction of electromagnetic fields with the PV cell.

III. RESULTS AND DISCUSSION

The results have also been divided into three sections for each of the separate set of simulations, and the results are then related with the J_{SC} for all the systems analyzed.

A. Absorption, Scattering and Extinction Spectra

For the first set of simulations, we analyzed the extinction spectra of Ag and Au plasmonic nanoparticles of different diameters, specifically with diameter (D) of 50nm , 100nm and 200nm , respectively. Each of the particles were placed inside the simulation region, and the absorption, scattering and extinction spectra were obtained using the appropriately placed monitors, as shown in Figure 2(a). The resultant spectra from the simulations are shown in figures 3(a) through (f).

For all six particles analyzed, it can be seen that there is a resonant wavelength at which the maximum absorption, scattering, and thus extinction, occurs. For all three diameters of Ag, the resonant peak occurs between wavelengths 400nm and 500nm , while that of Au is between wavelengths 500nm and 600nm . This is in agreement with results that have been previously generated [6]. The amplitude of the peak of the absorption spectra for particles of both metal decreases as diameter increases. The opposite is seen to be true for variations in the peak of the scattering spectra. This is once again in agreement with previous results [14]. Furthermore, a visible red shift in the absorption, scattering and extinction spectra is observed for each successive increase in diameter. For Au, the fall in the absorption peak is much more pronounced. Additionally, the extinction peak broadens for increasing diameter of the nanoparticles.

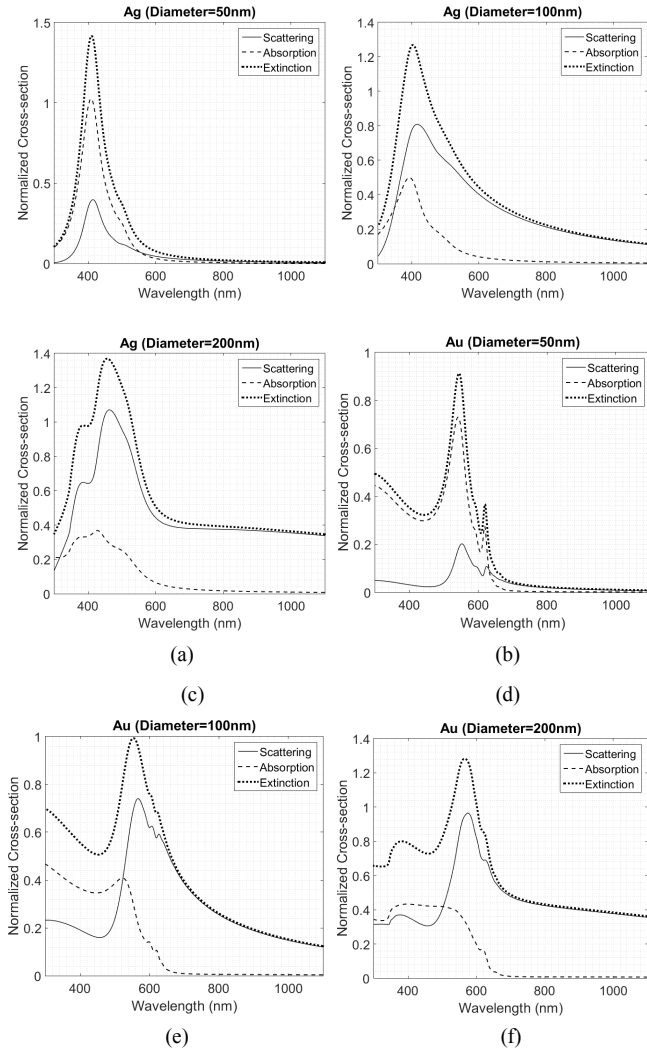


Figure 3. Absorption, scattering and extinction spectra for (a) Ag nanoparticle of D = 50nm (b) Ag nanoparticle of D = 100nm (c) Ag nanoparticle of D = 200nm (d) Au nanoparticle of D = 50nm (e) Au nanoparticle of D = 100nm (f) Au nanoparticle of D = 200nm

B. Absorption Enhancement Across Si Wafer Across the Entire Spectrum

The particles were then placed in a periodic orientation on top of the Si wafer with a pitch of 300nm (center-center), and the absorption enhancement factor across a wavelength of 400nm to 1100nm was calculated. In order to do so, the simulation region chosen only focused on one particle on top of the wafer, and boundary conditions of anti-symmetric for x-axis, and symmetric for y-axis were opted for in order to reduce the simulation time. The absorption enhancement factor (g) is then calculated by using the following formula:

$$g = \frac{\text{Absorption across Si with metal nanoparticles}}{\text{Absorption across bare Si}} \quad (1)$$

The absorption is calculated by placing monitors on the surface of the Si substrate, and 0.5 μ m below the surface (i.e., into the

Si substrate). The transmission of electromagnetic radiation from the bottom monitor is subtracted from the transmission of electromagnetic radiation from the surface monitor to obtain the individual absorption spectra. The resultant raw absorption spectra calculated for the Si substrate both with the plasmonic nanoparticles and without the plasmonic nanoparticles are applied to Eq. (1) above to obtain the wavelength-dependent absorption enhancement factor. The absorption enhancement factor gives us an appropriate mathematical framework to analyze. The resultant graphs are illustrated in Figures 4 and 5.

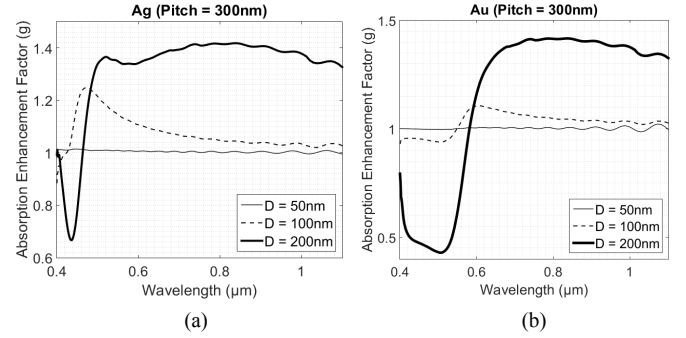


Figure 4. Absorption enhancement factor spectra for (a) Ag (with pitch = 300nm) and (b) Au (with pitch = 300nm)

The graphs show that there is an almost no absorption enhancement (i.e., absorption enhancement factor ~ 1) for Si wafers topped with Ag and Au particles of diameter D = 50nm. The absorption enhancement factor reaches a peak for D = 100 nm Ag at approximately $\lambda \sim 420$ nm and for D = 200 nm Ag at approximately $\lambda \sim 470$ nm, while the same for Au is around 560nm (D = 100nm) and no clear peaks for the D = 200nm nanoparticle, respectively. This can be related to the extinction spectra for both metals as shown in Fig. 3, as their resonant peaks lie approximately close to the same wavelengths.

The analysis is then repeated by changing the pitch to 220nm (center-center). The graphs obtained are shown in Figure 5. While the absorption enhancement factor for diameter D = 50nm is approximately the same as before for both Ag and Au, the peak g factor for diameter D = 100nm is higher than it was for the pitch of 300nm. Furthermore, it can be seen that the g factors for both Ag and Au with diameter D = 200nm are below 1 throughout the entire spectrum, from which it can be inferred that this should adversely affect the overall light coupling efficiency into the Si substrate. One possible explanation could be either increased absorption of the incident radiation by the Ag and Au nanoparticles and/or increased scattering of the incident radiation by the Ag and Au nanoparticles in a direction away from the Si substrate thus leading to decreased absorption by the Si substrate.

The conclusion that we can draw from the results shown in Figures 4 and 5 is that the plasmonic nature of the metals involved have an inherent effect on the light coupling efficiency of the thin-film Si substrate. Of the three diameters of metal nanoparticles analyzed, D = 100nm is seen to provide the largest absorption enhancement of the incident radiation into the Si substrate. Additionally, the variations in absorption and scattering spectra have an essential influence on the absorption enhancement as well. As previously shown, the

increase in diameter causes the absorption of the particle to decrease, while increasing the scattering efficiency of the nanoparticles.

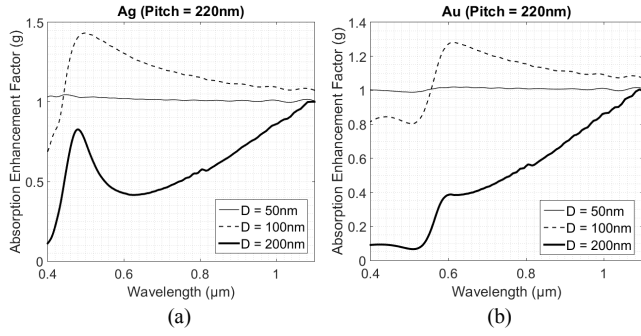


Figure 5. Absorption enhancement factor spectra for (a) Ag (with pitch = 220nm) and (b) Au (with pitch = 220nm)

This allows us to draw a correlation between the scattering of the plasmonic nanoparticles and absorption enhancement of the Si substrate. It is possible that for the $D = 200\text{nm}$ metal nanoparticles, increased scattering of the incident radiation by the Ag and Au nanoparticles may be in a direction away from the Si substrate thus leading to decreased absorption by the Si substrate.

C. Analysis of J_{sc}

The short circuit current density (J_{sc}) for all of the plasmonic nanostructures studied ($D = 50\text{nm}$, 100nm and 200nm for pitch = 220nm and 300nm) is calculated, and illustrated in tabular form in Tables I (a) and (b) and Tables II (a) and (b).

TABLE I. J_{sc} OF THE DIFFERENT STRUCTURES ANALYZED

(a)

Pitch = 300nm , $\lambda = 400\text{nm}$ to 1100nm

Structure	$J_{sc} (\text{mA}/\text{cm}^2)$	% change
No particles	76.2793	0
Ag (Diameter = 50nm)	77.3442	13.961
Ag (Diameter = 100nm)	89.7007	17.595
Ag (Diameter = 200nm)	77.0193	0.970
Au (Diameter = 50nm)	76.4398	0.210
Au (Diameter = 100nm)	77.0183	0.969
Au (Diameter = 200nm)	48.9927	-35.772

(b)

Pitch = 220nm , $\lambda = 400\text{nm}$ to 1100nm

Structure	$J_{sc} (\text{mA}/\text{cm}^2)$	% change
No particles	76.2793	0
Ag (Diameter = 50nm)	78.2283	2.555
Ag (Diameter = 100nm)	95.5242	25.230
Ag (Diameter = 200nm)	41.1047	-46.11
Au (Diameter = 50nm)	76.5712	0.383
Au (Diameter = 100nm)	76.6562	0.494
Au (Diameter = 200nm)	19.2681	-77.362

TABLE II. J_{sc} OF THE DIFFERENT STRUCTURES FOR WAVELENGTH BETWEEN 400nm AND 700nm

(a)

Pitch = 300nm , $\lambda = 400\text{nm}$ to 700nm

Structure	$J_{sc} (\text{mA}/\text{cm}^2)$	% change
No particles	64.1602	0
Ag (Diameter = 50nm)	65.1141	1.487
Ag (Diameter = 100nm)	75.6064	17.840
Ag (Diameter = 200nm)	59.2331	-7.679
Au (Diameter = 50nm)	63.8908	-0.420
Au (Diameter = 100nm)	59.8337	-6.743
Au (Diameter = 200nm)	26.0504	-59.398

(b)

Pitch = 220nm , $\lambda = 400\text{nm}$ to 700nm

Structure	$J_{sc} (\text{mA}/\text{cm}^2)$	% change
No particles	64.1602	0
Ag (Diameter = 50nm)	65.8854	2.689
Ag (Diameter = 100nm)	79.7938	19.691
Ag (Diameter = 200nm)	32.9415	-48.657
Au (Diameter = 50nm)	63.6557	-0.786
Au (Diameter = 100nm)	56.1374	-12.504
Au (Diameter = 200nm)	8.44858	-86.832

The largest increase in J_{sc} can be seen to be for Ag nanoparticles of diameter $D = 100\text{nm}$ and a pitch of 220nm , which is a 25.23% increase. The increase is not as significant for Au as it is for Ag, which supports evidence that has been previously provided [4]. Note that although the J_{sc} is not as adversely affected by Ag nanoparticles of diameter $D = 200\text{nm}$ at a pitch of 300nm , the J_{sc} significantly decreases for $D = 200\text{nm}$ Au at a pitch of 300nm (-35.77%). There is also a sharp decline in J_{sc} for Ag and Au particles of diameter $D = 200\text{nm}$ at a pitch of 220nm . This is in agreement with our analysis of the absorption enhancement factor, as at a pitch of 220nm the absorption enhancement factor was found to be less than 1 (Figure 5). Reducing the wavelength of interest region to $400\text{nm} - 700\text{nm}$ (Table 2) from $400\text{nm} - 1100\text{nm}$ (Table 1) leads to reduced J_{sc} . This indicates that the Ag nanoparticles lead to increased absorption into the Si substrate across the entire wavelength range of $400\text{nm} - 1100\text{nm}$.

D. Near Field Enhancement

In order to observe and analyze the transmission of electric fields within the cells themselves, we opted to generate near-field enhancement images of the structures in question. This would allow us to relate the optical and electrical enhancement of the cell due to the presence of the plasmonic metal nanoparticles with the interaction of the structure with the fields. Near field images have to be generated at a specific wavelength. For our purpose, we chose the wavelengths at which the particles exhibited plasmon resonance, i.e. $\lambda \sim 450\text{nm}$ for Ag particles of all three diameters, and $\lambda \sim 550\text{nm}$ for Au particles of all three diameters.

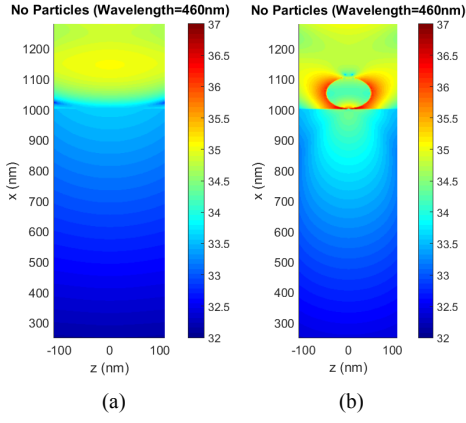


Figure 6. Near-field images for (a) Si substrate without metal nanoparticles (b) Si substrate with Ag nanoparticle of $D = 100\text{nm}$ at a pitch of 220nm .

For each structure of interest, we first generated the near-field image with no nanoparticles present, and then generated the same with the particles present. The results are shown in Figure 6. Finally, in order to generate the near-field enhancement profile, we divided the raw data from the second image with the same from the first image. The near field enhancement images for cells with Ag nanoparticles of diameters $D = 50\text{nm}$, 100nm and 200nm are shown in Figure 7. the color scale is in the log scale, and hence the areas which are dark red color have an enhancement of 1 in the log scale that corresponds to a near-field enhancement of $\times 10$ (10 times). The images demonstrate that the presence of the nanoparticles significantly enhance the intensity of the electromagnetic fields around the particle and within the cell substrate as well. A large portion of the area around the particle can be seen to be dark red in color, which shows that a considerable amount of near-field enhancement has occurred around that region. Furthermore, this enhancement has been carried into substrate directly below the nanoparticle itself. This conforms with the overall increase in optical absorption enhancement that we found and showed in Figures 4 and 5, and the increase in J_{SC} in Tables I and II. It is also found that there is a general increase in enhancement with respect to increasing diameter of the nanoparticles. Additionally, the width of electromagnetic enhancement within the substrate is smaller for $D = 200\text{nm}$ than $D = 100\text{nm}$, which conforms with the corresponding decrease in J_{SC} as shown previously.

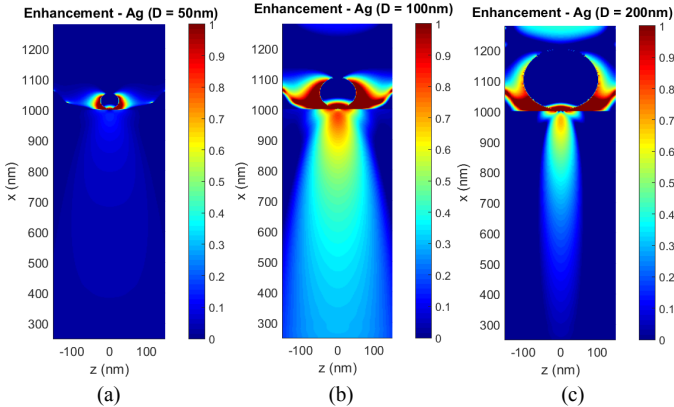


Figure 7. Near-field enhancement image for cell with Ag nanoparticles of diameter (a) $D = 50\text{nm}$, (b) $D = 100\text{nm}$, and (c) $D = 200\text{nm}$.

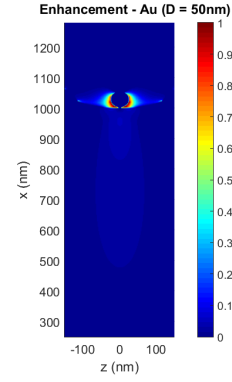


Figure 8. Near-field enhancement image of cell with Au nanoparticle of diameter $D = 50\text{nm}$.

To effectively comprehend the electromagnetic enhancement of the cell due to the presence of the Ag nanoparticles, we generated the near-field enhancement of the cell with Au nanoparticles of diameter $D = 50\text{nm}$ place at a pitch of 220nm , which is shown in Figure 8. It can be seen that the enhancement around the Au nanoparticle is lower than the corresponding Ag nanoparticle-based structure. This is once again in agreement to the absorption enhancement and J_{SC} analysis, which showed that cells with Au nanoparticles had lower increase in optical and electrical improvement than cells with Ag nanoparticles.

The relationship between increased absorption of the incident radiation by the Si substrate as a result of the influence of the plasmonic metal nanoparticles with increased short circuit current density (J_{SC}) needs to be further explored. It is not entirely clear from the data shown thus far that there is a direct one-to-one correlation between increased absorption of incident radiation by the Si substrate and increased J_{SC} generation. It is possible that the polarization of the absorbed light can directly affect the efficiency of generation of electron-hole pairs via the photo-electric effect. It is thus possible that dipolar or additional multi-polar scattered incident radiation by the larger nanoparticles can de-polarize the scattered incident radiation and thus possibly leading to less efficient J_{SC} generation. Hence, future computations will explore the nature of polarization of the absorbed radiation in the Si substrate.

IV. CONCLUSIONS

This study explores the relationship between energy conversion efficiency in amorphous thin-film Si solar cells and the type of metallic nanoparticle, the size of the metallic nanoparticles, and the distance between neighboring metallic nanoparticles in a periodic array of the metallic nanoparticles deposited on top of the Si substrate. The different configurations of metallic nanoparticle systems studied were as follows: $D = 50\text{ nm}$, $D = 100\text{ nm}$, and $D = 200\text{ nm}$ Ag nanoparticles spaced with a pitch of 300 nm (center-center) and a pitch of 220 nm (center-center), respectively, and $D = 50\text{ nm}$, $D = 100\text{ nm}$, and $D = 200\text{ nm}$ Au nanoparticles spaced with a pitch of 300 nm (center-center) and a pitch of 220 nm (center-center), respectively. Of the different metallic nanoparticle systems studied, it is found that Ag nanoparticles of $D = 100\text{ nm}$ and spaced 220 nm (center-center) from neighboring

nanoparticles in a periodic array of the metallic nanoparticles deposited on top of the Si substrate provides the maximum increase (i.e., 25%) in the generated short circuit current density (J_{SC}). There is a strong spectral correlation between the extinction and scattering spectra of the $D = 100$ nm Ag nanoparticles and the wavelength-dependent increase in incident radiation absorption into the Si substrate. It is also shown that decreasing the pitch size (from 300 nm to 200 nm) leads to a larger increase in the J_{SC} . This can be due to the increased interaction between the neighboring Ag nanoparticles in the shorter pitch case and thus the generation of a stronger SPP resonance condition. This strong spectral correlation suggests a significant role played by SPP resonance in the Ag nanoparticles and the increase in J_{SC} . However, our results also display that the $D = 200$ nm Ag nanoparticle array configuration with a pitch of 300 nm (center-center) shows appreciable increase in the absorption enhancement of the incident radiation into the Si substrate but not an increase in J_{SC} . The relationship between increased absorption of the incident radiation by the Si substrate as a result of the influence of the plasmonic metal nanoparticles with increase in the J_{SC} needs to be further explored. It is not entirely clear from the data shown thus far that there is a direct one-to-one correlation between increased absorption of incident radiation by the Si substrate and increased J_{SC} generation. It is possible that the polarization of the absorbed light can directly affect the efficiency of generation of electron-hole pairs via the PV effect. It is thus possible that dipolar or additional multi-polar scattered incident radiation by the larger metal nanoparticles ($D = 200$ nm) can de-polarize the scattered incident radiation and thus possibly leading to less efficient J_{SC} generation. Hence, future computations will explore the nature of polarization of the absorbed and/or scattered radiation in the Si substrate. In the wavelengths of interest investigated here, Au nanoparticles did not lead to an increase in the generated short circuit current density (J_{SC}) in the Si solar cells. This can be due to the optical properties of Au where Au displays relatively high absorption cross-sections for wavelengths in the 300 – 650 nm range.

ACKNOWLEDGEMENT

The authors would like to thank Dr. Khosru M. Salim for the lively discussions, encouragement and funding for the research assistant, and IUB for funding the research.

REFERENCES

- [1] S. Pillai, and M. A. Green, "Plasmonics for photovoltaic applications," *Solar Energy Materials & Solar Cells*, vol. 94, pp. 1481-1486, 2010.
- [2] J. Nelson, *The Physics of Solar Cells*. London, England: Imperial College Press, 2003.
- [3] P. Wüffel, *Physics of Solar Cells: Principles to New Concepts*. New York, NY: Wiley-VCH, 2004.
- [4] H. Zhou, Q. Dhen, G. Li, S. Luo, T. Song, H. Duan, Z. Hong, J. You, Y. Liu, and Y. Yang, "Interface engineering of highly efficient perovskite solar cells," *Science*, vol. 345 (6196), pp. 542-546, 2014.
- [5] M. C. Beard, J. M. Luther and A. J. Nozik, "The promise and challenge of nanostructured solar cells," *Nature Nanotechnology*, vol. 5, pp. 951-954, 2014.
- [6] S. H. Lim, W. Mar, P. Matheu, D. Derkacs, and E. T. Yu, "Photocurrent spectroscopy of optical absorption enhancement in silicon photodiodes via scattering from surface plasmon polaritons in gold nanoparticles," *J. Appl. Phys.*, vol. 101, pp. 104309-1-104309-7, 2007.
- [7] J. P. Kottmann, O. J. F. Martin, D. R. Smith, and S. Schultz, "Spectral response of plasmon resonant nanoparticles with a non-regular shape," *Opt. Express*, vol. 6, pp. 213-219, 2000.
- [8]
- [9] K. L. Kelly, E. Coronado, L. L. Zhao, and G. C. Schatz, "The Optical Properties of Metal Nanoparticles: The Influence of Size, Shape, and Dielectric Environment," *J. Phys. Chem. B*, pp. 668-677, 2003.
- [10] M. A. Green, and S. Pillai, "Harnessing plasmonics for solar cells," *Nature Photonics*, vol. 6, pp. 130-132, 2012.
- [11] C. F. Bohren, and D. R. Huffman, *Absorption and Scattering of Light by Small Particles*. New York, NY: Wiley, 1983.
- [12] X. Li, W. C. H. Choy, H. Lu, W. E. I. Sha, and A. H. P. Ho, "Efficiency Enhancement of Organic Solar Cells by Using Shape-Dependent Broadband Plasmonic Absorption in Metallic Nanoparticles," *Adv. Func. Mat.*, vol. 23 (21), pp. 2728-2735, 2013.
- [13] D. Derkacs, S. H. Lim, P. Matheu, W. Mar, and E. T. Yu, "Improved performance of amorphous silicon solar cells via scattering from surface plasmon polaritons in nearby metallic nanoparticles," *Appl. Phys. Lett.*, vol. 89, pp. 093103, 2006.
- [14] R. Hulstrom, R. Bird, and C. Riordan, "Spectral Solar Irradiance Data Sets for Selected Terrestrial Conditions," *Solar Cells*, vol. 15, issue 4, pp. 365-391, 1985.
- [15] E. T. Yu, D. Derkacs, P. Matheu, and D. M. Schaadt, "Plasmonic nanoparticle scattering for enhanced performance of photovoltaic and photodetector devices," *Plasmonics: Nanoimaging, Nanofabrication, and Their Applications IV*, 2008.

Predicting Video Conversion Time from Video Metadata and Conversion Parameters using Gradient Boosting Machine

Md. Toufique Ahmed¹ and Md. Taufeeq Uddin¹

Department of Computer Science and Engineering, International Islamic University Chittagong, Bangladesh
{toufique.h.ahmed, mohamed.taufeeq.uddin}@gmail.com

Abstract—Predictive analytics techniques can tremendously improve the performance of computing systems by optimizing energy, waiting time and throughput via predicting the execution time of scheduled jobs beforehand. As a consequence of the correlation between video conversion parameters and video conversion time, the conversion time is highly predictable from input video properties and conversion parameters. Hence, this paper proposes gradient boosting machine to predict the conversion time of videos using video metadata and conversion features with no detailed information about the applied codec. The evaluation results of the experiments conducted on benchmark Youtube video characteristics dataset showed that our model reduced the conversion time prediction error by as much as 4.03% over previously applied models. The proposed model also indicates that features about coding standard and codec allocated memory used for conversion, size, duration, bitrate and framerate of videos are crucial for conversion time prediction.

Keywords—Video Properties, Video Conversion Time Prediction, Gradient Boosting Machine, Quality of Services

I. INTRODUCTION

Video conversion or transcoding is the process of converting a video file from one format to another to make video files viable across different configurable platforms and devices. There are several major video conversion types such as resolution conversion, bitrate conversion, temporal conversion, container conversion, codec conversion, error reliance conversion, and any combination of these [1]. Recently, conversion of video content is becoming an increasing requirement for the digital media industry since the increasing number of devices and platforms, and the availability of numerous video formats with various platforms. As we know, users want to play a video in various formats on various devices and platforms with a bunch of properties. For instance, some users want high definition videos while others want low-resolution videos, but it is almost impossible in the realistic setting to store a single video in all possible formats to fulfill the end users needs. Fortunately, video conversion mechanism can lead us to get rid of these kinds of problems.

The major methods of video conversion are originally employed to reduce video file size, and to make a distribution on mobile networks easier. Surprisingly, it is increasingly used in recent time to make online media contents viewable across platforms and enable HTTP live streaming and adaptive bitrate delivery. However, video conversion is still a delicate and

time-consuming process while extremely useful in evolving streaming of nowadays multi-media marketplace. Moreover, almost every conversion results in some degree of loss, and subtle encoding nuances that can directly affect the video quality.

Note that the estimation of probable time of video conversion has a tremendous impact in many domains. For instance, distributed and multicore systems can conquer the problem of over-provisioning by estimating the resource requirement of each job. As we know that the unavailability of resource requirement in advance makes the run-time scheduling of conversion job in multicore and cloud environment much more challenging. Moreover, someone needs to depend on the worst case values for video conversion job since the resource requirement of video conversion job is mainly rely on video data and conversion parameters that lead to over-provisioning of resources to maintain a satisfactory quality of services (QoS).

Fortunately, every computing systems such as smartphones, tablets, laptops, servers, data centers, and clouds have resource management components which determine how to schedule the execution of different jobs over time [2] to make sure optimal system usages, efficient energy utilization, or program resources allocation e.g. memory, storage, bandwidth. In order to get optimal performance from management components, they need to assess how to perform an assigned task depending on its properties, and therefore, to learn to project for the best for the future. For instance, consider a cloud conversion service with a set of two kinds of conversion requests, fast conversion job in set S_1 , and slow conversion job in set S_2 . S_1 scheduler is often dealt with the decision of whether to run each set of different CPU resources, possibly taking longer for execution, or to interleave between the two sets and assign the jobs fairly and possibly executing the job much faster. The scheduler can make the wise decision that is it can return results faster, possibly minimize energy and maximize throughput, *when it can efficiently estimate the duration of each job for execution on a given platform*. Again, the proper job scheduling is really essential unless the skewed distribution of conversion time as shown in Figure 1 can tremendously affect the performance of the system [3]. Fortunately, low overhead, data preprocessing techniques, and highly accurate

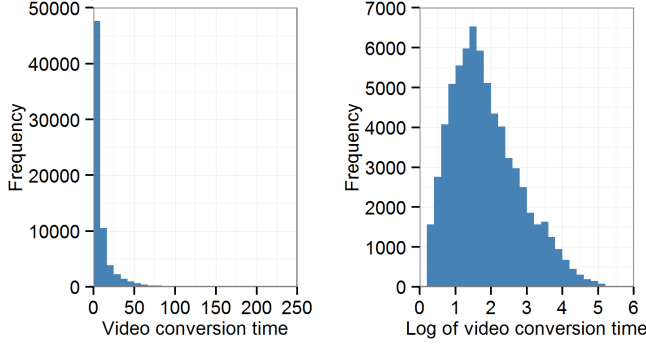


Fig. 1. From left to right, distribution of original conversion times, and logarithms of original conversion times of videos in Youtube video characteristics dataset [5]

machine learning algorithms can be applied to leverage such things.

In this paper, we apply a novel ensemble learning algorithm, namely gradient boosting machine [4] to predict the conversion time of videos using heterogeneous set of video metadata features such as duration, bitrate, height and width of videos and conversion features such as coding standard, output codec, and total allocated memory for conversion. As a result of the prediction of the conversion time of videos before scheduling for conversion, *scheduling decision can be made beforehand based on the predicted conversion time to obtain optimal usages of resources such as memory, bandwidth.*

In our experiments, we found that gradient boosting machine predicted the conversion time of videos with around **4.75%** MAPE error rate on publicly available Youtube video characteristics dataset that was as much as **4.03%** reduction over previous methods. In addition, gradient boosting machine measured that information about output width and output codec and total codec allocated memory used for conversion, bitrate, width, size, duration, and frame rate of videos, the number of frames in a video were crucial for predicting the conversion time of videos. To the best of our knowledge, this is the first time that gradient boosting machine with linear regression objective function is applied to the task of predicting the conversion time of videos.

II. RELATED WORK

Over the last couple of years, several researchers proposed numerous cloud-based distributed transcoding and resource allocation systems. For instance, Zhenhua Li et al. [6] proposed and implemented a cloud-based conversion system focusing on providing video conversion service rather than conversion job scheduling. In their system, video request contains a video link and the user-customized transcoding parameters by users who upload a video request rather than the video content parameters including format, resolution. Jokhio et al. [7] designed a scalable distributed message passing interface based on transcoder in which video was statically segmented at a group of pictures in three difference

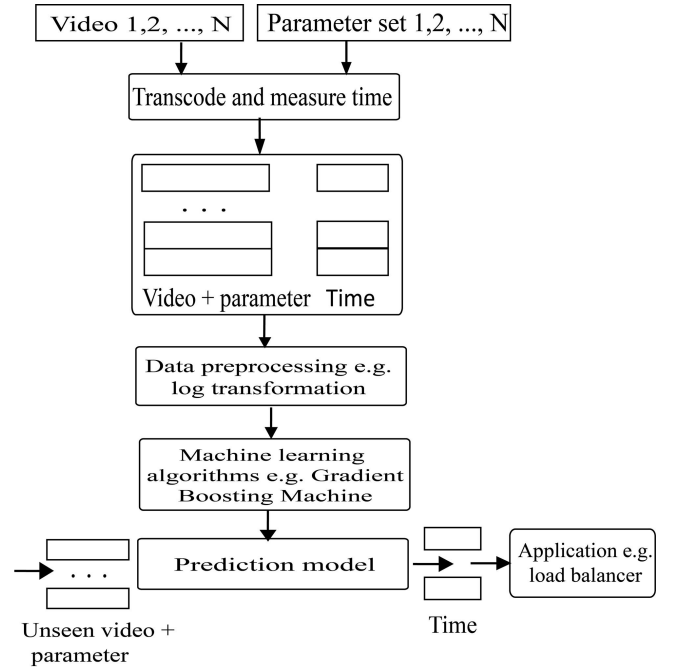


Fig. 2. A generic video conversion framework

ways. Their key contribution was video segmentation for conversion in distributed environment, and evaluation of the video startup time in such environment. Later Jokhio et al. [8] demonstrated a prediction-based dynamic resource allocation and deallocation algorithms to scale video conversion service in a cloud environment. Based on a regression model, they proposed an algorithm providing a mechanism to allocate and deallocate virtual machines that tracked and predicted the aggregate target conversion rate required by the service. Huang et al. [9] presented multilevel parallelized conversion framework for cloud-based video proxy to deliver converted videos for streaming.

As per our knowledge based on the literature review, there are only two similar studies available in the literature that also attempted to predict the conversion time to boost the performance of the scheduling tasks and to ensure QoS. First, Roitzsch et al. [10] developed a prediction model to estimate per-frame decoding times of modern decoding algorithms using expert selected metric to train and to predict decoding times of video encoded in MPEG-1, MPEG-2, and MPEG-4 part 2 algorithms. Their major limitation is that their method could not support a set of codecs from multiple codec families. Then, Deneke et al. [1] proposed support vector regressor (SVR) and multilayer neural network (MNN) based prediction model for predicting the conversion time of videos using video metadata and conversion parameters with no specific information about the used codec to design load balancing scheme to aid throughput and QoS.

III. VIDEO CONVERSION FRAMEWORK

The main focus of this work is to estimate the conversion time of videos using input video and conversion parameters

to automate conversion system. The predicted conversion time can then be utilized by the service provider for many purposes such as decision making for scheduling, load balancing, and QoS-predictions. Therefore, using this system, conversion service providers can predict the conversion time of videos, and can manage their conversion server more efficiently. The key components of the video conversion system are fundamental, relevant and highly predictive video metadata and conversion features e.g. codec, bitrate and framerate, data preprocessing techniques e.g. logarithmic transformation, and machine learning algorithms e.g. gradient boosting machine that can learn to measure the probable conversion time of videos using the video metadata and conversion features. Figure 2 displays the big picture of the video conversion system. Usually, conversion service providers possess a log regarding conversion parameters and actual video. Hence, we can create training dataset in supervised learning format based on the traces by including conversion parameters, actual video or its primary properties e.g. width, height, framerate and bitrate, and computed conversion time. The collected dataset is then preprocessed using various data preprocessing techniques e.g. logarithmic transformation. Several type conversion techniques are also applied to create final suitable and predictive training dataset. After that, machine learning methods e.g. gradient boosting machine (GBM) are applied on the preprocessed dataset to train to predict the conversion time of the seen and unseen videos. Finally, the predicted results are used for multiple purposes including load distribution across the nodes, and QoS.

IV. GBM FOR CONVERSION TIME PREDICTION

The goal of this work is to predict the conversion time of videos using GBM algorithm in the regression setting. In training step of this method, heterogeneous set of video metadata and conversion features are fed to GBM to build the desired prediction model. The trained GBM model is then used in prediction step to carry out conversion time prediction task of the unseen videos. The rest of this section covers a brief description of the key idea, algorithmic description, and regularization strategies of GBM for regression.

A. Key Idea

GBM [11] is an ensemble learning algorithm that is the combination of gradient-based optimization and boosting. GBM produces strong prediction model by combining multiple weak prediction models in which weak models are created by sequentially applying to the incrementally changed dataset. Optimization based on the gradient in GBM utilize the gradient computations to minimize the cost function of a model with respect to training dataset while boosting additively gathers an ensemble of weak models to build the prediction model for video conversion time prediction challenge. In short, the main idea beneath GBM is to build a series of simple and probably inaccurate decision trees or weak models successively from the prediction residuals of the preceding decision trees and

combine them to construct a final highly accurate prediction model.

B. Algorithmic Description

The technical description of the predictive GBM [4] [12] is given below:

- **Input:** GBM takes $(x_1, y_1), \dots, (x_n, y_n)$ as training dataset in which $x_i \in X$ are the set of features extracted from video contents and conversion properties and y_i is the required conversion time that is the prediction label. GBM also takes a differentiable loss function $L(y, F(x))$ and the number of gradient boosting iteration M as input.

- **Algorithm:**

I. Initialization of the model with a constant value

$$F_0(x) = \operatorname{argmin}_{\gamma} \sum_{i=1}^n L(y_i, \gamma) \quad (1)$$

II. For each gradient boosting iteration $m = 1, \dots, M$

1. For $i = 1, \dots, n$, compute pseudo-residuals

$$r_{im} = - \left[\frac{\delta L(y_i, F(x_i))}{\delta F(x_i)} \right]_{F(x)=F_{m-1}(x)} \quad (2)$$

2. For $j = 1, \dots, J_m$, GBM fits a regression tree to the labels y_i providing terminal region R_{jm} . J is the size of the tree that controls the level of feature interaction in the model.

3. For $j = 1, \dots, J_m$, GBM then computes the multiplier γ_m

$$\gamma_{jm} = \operatorname{argmin}_{\gamma} \sum_{x_i \in R_{jm}} L(y_i, F_{m-1}(x_i) + \gamma) \quad (3)$$

4. Finally, update the model

$$F_m(x) = F_{m-1}(x) + v \sum_{j=1}^{J_m} \gamma_{jm} I(x \in R_{jm}) \quad (4)$$

where v ($0 < v \leq 1$) is the shrinkage constant that regulates the learning rate of GBM by reducing the size of the incremental steps.

- **Output:** $F_M(x)$

C. Regularization

Generalization capabilities of GBM are one of the essential concern. A number of parameters can contribute towards reducing the effects of overfitting by controlling learning rate and/or by introducing randomness into GBM. For instance, the smaller values of learning rate v such as $v \leq 0.1$ can generally ensure better generalization and performance on the unseen dataset. Overfitting can also be eliminated by fitting weak model on a subsample or constant fraction e.g. $0.5 \leq \text{fractions} \leq 0.8$ of the training dataset at random with no replacement. Furthermore, to reflect generalization, we can use a large number of trees or the boosting iteration M , and control J , by picking the value of J in between 4 and 8.

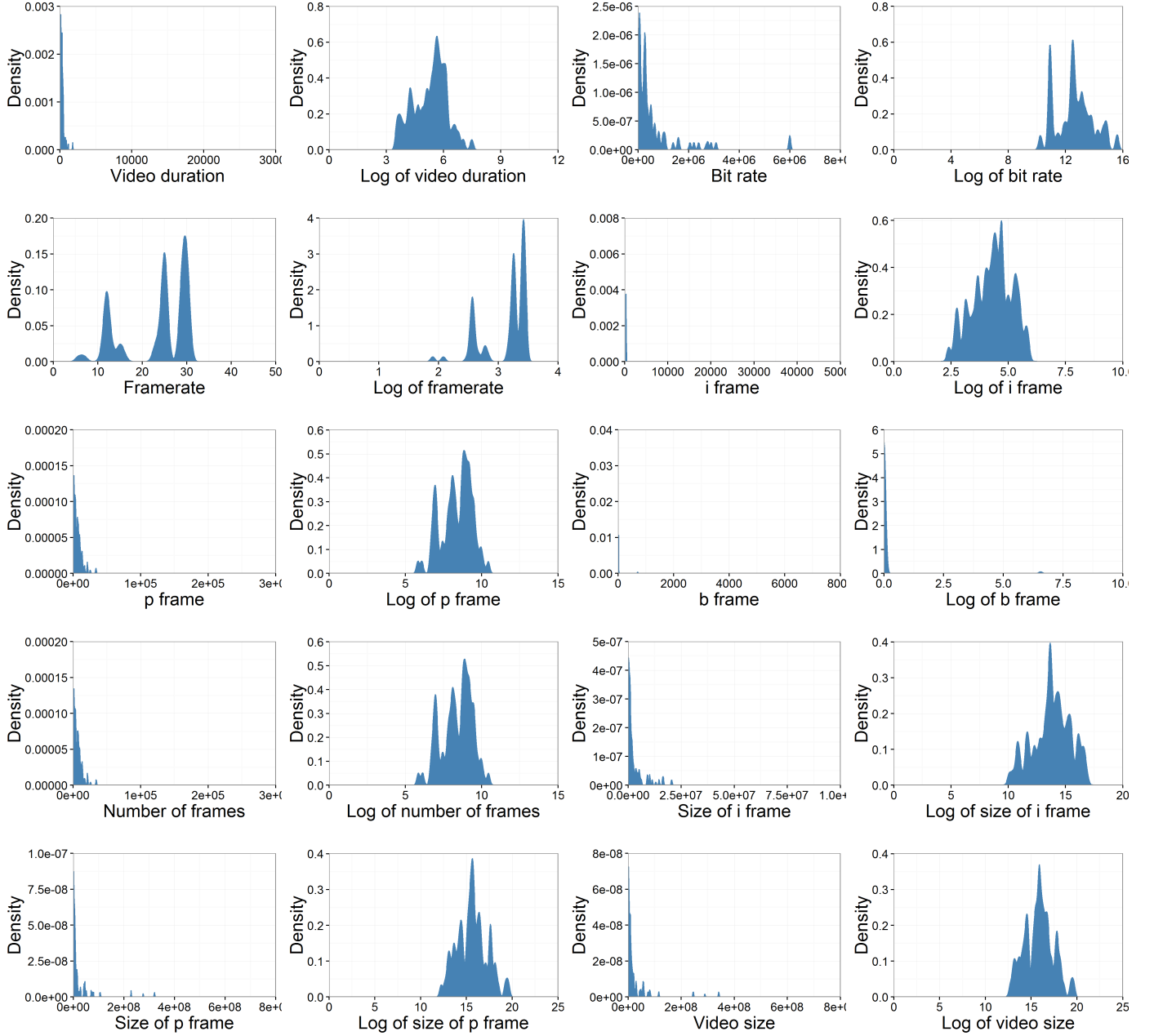


Fig. 3. Feature distribution of several original, and logarithm of original features of Youtube video characteristics dataset

V. YOUTUBE VIDEO CHARACTERISTICS DATASET

Online video characteristics and conversion time prediction (**Youtube video characteristics**) dataset [5] is publicly available at <https://archive.ics.uci.edu/ml/datasets/>. Youtube video characteristics dataset comprises of 68784 Youtube video instances including 10 fundamental video characteristics with video ID. Deneke et al. developed crawler using random prefix sampling method [13] to sample over videos because crucial statistics regarding fundamental video characteristics are not publicly available on Youtube. They then probed, analyzed, and collected elemen-

tary properties of each of the sampled video using Ffprobe to generate video transcoding parameters. They achieved actual conversion time of videos for each of the simulated conversion request through the measurement on a cloud instance from Amazon EC2. Youtube video characteristics dataset composes of 20 features including input and output video characteristics and their conversion time and memory resource requirements during the conversion of videos to different formats. These 20 features cover information about duration and bitrate of video, height and width of video in pixels, actual and estimated video frame rate, coding standard used for the video, YouTube video

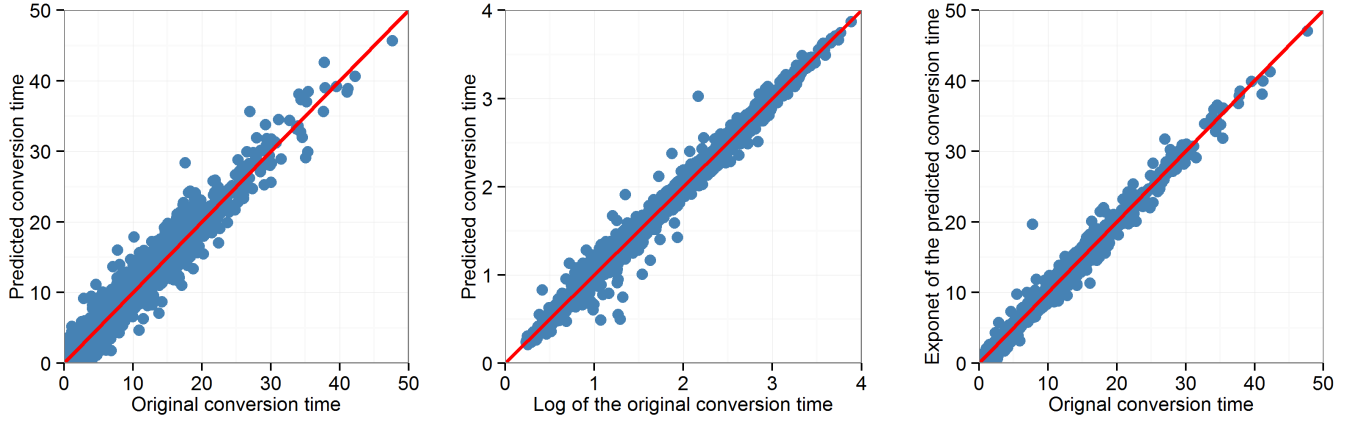


Fig. 4. From left to right, for GBM, scatter plots of original conversion time against predicted conversion time based on the model M_1 , logarithm of the original conversion time against predicted conversion time based on the model M_2 , and the original conversion time against the exponent of the predicted conversion time based on the model M_2 on the test subset of the Youtube video characteristics dataset

category, direct link to video with expiration date, number of i, p, and b frames in the video, number of frames in video, total size of i, p, and b videos in byte, total size of video, output codec, bitrate, framerate, width and height used for video conversion, and total codec allocated memory for conversion.

VI. EXPERIMENTS AND RESULTS

The proposed GBM prediction model has evaluated on Youtube video characteristics dataset in which we randomly split the entire dataset into training dataset (80%), and test dataset (20%). We performed 6-fold cross-validation on the training dataset and the final cross-validated model was tested using the test dataset. We benchmarked the proposed model against previously applied models such as SVR and MNN [1] using Youtube video conversion dataset. We display the obtained results in Table I. Note that each of the experiments was run for 20 times with different random seeds, and the results were obtained by averaging over 20 different experimental runs.

We preprocessed 10 skewed features e.g. duration, bitrate, framerate, i, p and b frames, the total number of frames, size of i and p frames, and total size of videos using logarithmic transformation. The distribution of the skewed features before and after the logarithmic transformation are shown in Figure 3. Note that we modeled the video conversion time prediction problem as regression problem as we aim to estimate the conversion time of videos in real time. Hence, we built two different GBM models in which one of them (M_1) was used original conversion times as the prediction label, and another one (M_2) was used the logarithm of the original conversion time as the prediction label. During the performance evaluation of the GBM models, we evaluated the models by measuring the mean absolute percentage error (MAPE) between the predicted times and the original conversion time (**GBM_MAPE**) for M_1 ; between the predicted times and logarithm of the original conversion time (**GBM_MAPE_LOG**) for M_2 ; and between the exponents of predicted conversion time and the original conversion time (**GBM_MAPE_EXP**) for M_2 .

TABLE I
VIDEO CONVERSION TIME PREDICTION BENCHMARK. THIS TABLE DISPLAYS THE MAPE AND STANDARD DEVIATION OF MAPE (**MAPE-SD**) GENERATED FROM TWO DIFFERENT MODELS

Setting	MAPE	MAPE-SD
GBM_MAPE	23.6657%	0.0317
GBM_MAPE_LOG	2.5959%	0.0086
GBM_MAPE_EXP	4.7490%	0.0131

For training and validating the proposed GBM, we set *linear regression* as objective function, the value of the learning rate $v = 0.1$ to reflect better generalization, the size of the tree $J = 3$, and subsample size *fraction* = 0.8. Finally, the number of boosting iteration M was selected using 6-fold validation strategy. Table I shows the performance of GBM during the prediction of video conversion time. From Table I, it can be outlined that GBM was better than SVR and MNN since MAPE generated from GBM was **4.75%** that was relatively smaller than the MAPE generated from SVR (18.64%) and MNN (8.78%). Another important point to note from Table I and Figure 4 is that training over the logarithm of the original conversion time is extremely useful to fit better GBM model as we can observe MAPE generated from GBM trained on the original conversion time was 23.67%, while MAPE generated from GBM trained on the logarithm of the conversion time and further transformation of the predicted conversion time by taking the exponent of the predicted conversion time was 4.75%. Finally, as we predicted the conversion time of videos, these prediction results can then be utilized for the performance optimization of the job scheduler, load balancer and QoS expectation of the users by the conversion service providers via estimating the video conversion cost of requests, and managing the conversion servers more wisely and effectively.

GBM uses *gain* [14] to estimate the contribution of each feature to the prediction model. GBM takes each gain of

TABLE II
TOP 14 FEATURES BASED ON THE IMPORTANCE SCORES OF
FEATURES, MEASURED VIA GBM, FOR VIDEO CONVERSION
TIME PREDICTION

Feature	Gain * 100
Output width in pixel	43.4686
Output codec	31.47
Total codec allocated memory	9.55
Video bit rate	7.7755
Width of video in pixel	2.95
Output bit rate	1.8163
Output frame rate	1.5384
Total size in bytes of p videos	0.527
Actual video frame rate	0.5267
Output height used in pixel	0.3413
Number of i frames in video	0.0095
Duration of video	0.0093
Total size of video	0.0088
Number of frames in video	0.0086

each feature of each tree and computes mean per feature to provide a vision of the entire prediction model. We measured the relative importance scores of features using GBM during the prediction of the conversion time of videos for which we trained GBM using entire Youtube video characteristics dataset. Table II highlights the relative importance scores or $Gain * 100$ of the top 14 features. In can be observed from Table II that features related to the information about output width and height in pixels, output codec, bitrate and framerate, total codec allocated memory, total size and duration of videos, and number of frames in videos had strong significance towards the prediction of the conversion or transcoding time of videos.

VII. CONCLUSION AND FUTURE WORK

This paper introduced and implemented GBM to tackle the challenge of predicting the conversion time of videos to boost the performance of conversion services using some metadata and conversion features e.g. bitrate, framerate, output codec. Our findings suggest that GBM is able to measure the video conversion time with a minimal error rate using these metadata and conversion features. More interestingly, we found that the strategy of training the GBM model on the logarithm of the conversion time is extremely useful for better model fitting. GBM also outlined relevant and influential metadata and conversion features e.g. output width and height in pixels, output codec, bitrate and framerate, total codec allocated memory,

total size and duration of videos, and a number of frames in videos computing their relative importance scores. Future work will include, first, the evaluation of the GBM model on more complex and different video streaming datasets. Second, the comparison of the GBM based video conversion time prediction model with many other state-of-the-art models.

REFERENCES

- [1] T. Deneke, H. Haile, S. Lafond, and J. Lilius, "Video transcoding time prediction for proactive load balancing," in *Multimedia and Expo (ICME), 2014 IEEE International Conference on*. IEEE, 2014, pp. 1–6.
- [2] M. Isard, V. Prabhakaran, J. Currey, U. Wieder, K. Talwar, and A. Goldberg, "Quincy: fair scheduling for distributed computing clusters," in *Proceedings of the ACM SIGOPS 22nd symposium on Operating systems principles*. ACM, 2009, pp. 261–276.
- [3] M. Harchol-Balter, "The effect of heavy-tailed job size distributions on computer system design," in *Proc. of ASA-IMS Conf. on Applications of Heavy Tailed Distributions in Economics, Engineering and Statistics*, 1999.
- [4] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Annals of Statistics*, vol. 29, pp. 1189–1232, 2000.
- [5] T. Deneke, "Online video characteristics and transcoding time dataset data set," last accessed: June 18, 2016. [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/Online+Video+Characteristics+and+Transcoding+Time+Dataset>
- [6] Z. Li, Y. Huang, G. Liu, F. Wang, Z.-L. Zhang, and Y. Dai, "Cloud transcoder: Bridging the format and resolution gap between internet videos and mobile devices," in *Proceedings of the 22nd international workshop on Network and Operating System Support for Digital Audio and Video*. ACM, 2012, pp. 33–38.
- [7] F. Jokhio, T. Deneke, S. Lafond, I. Porres, and J. Lilius, "Bit rate reduction video transcoding with distributed computing," in *Parallel, Distributed and Network-Based Processing (PDP), 2012 20th Euromicro International Conference on*. IEEE, 2012, pp. 206–212.
- [8] F. Jokhio, A. Ashraf, S. Lafond, I. Porres, and J. Lilius, "Prediction-based dynamic resource allocation for video transcoding in cloud computing," in *Parallel, Distributed and Network-Based Processing (PDP), 2013 21st Euromicro International Conference on*. IEEE, 2013, pp. 254–261.
- [9] Z. Huang, C. Mei, L. E. Li, and T. Woo, "Cloudstream: Delivering high-quality streaming videos through a cloud-based svc proxy," in *INFOCOM, 2011 Proceedings IEEE*. IEEE, 2011, pp. 201–205.
- [10] M. Roitzsch and M. Pohlack, "Principles for the prediction of video decoding times applied to mpeg-1/2 and mpeg-4 part 2 video," in *Real-Time Systems Symposium, 2006. RTSS'06. 27th IEEE International*. IEEE, 2006, pp. 271–280.
- [11] M. M. Cliff Click, Jessica Lanford and V. Parmar, "Gradient boosted models with h2o's r package," February 2015: Second Edition, published by H2O.ai, Inc. 2307 Leghorn Street Mountain View, CA 94043, Inc. Available: <https://leanpub.com/gbm/read>. [Online]. Available: <https://leanpub.com/gbm/read>
- [12] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*. Springer series in statistics Springer, Berlin, 2001, vol. 1.
- [13] J. Zhou, Y. Li, V. K. Adhikari, and Z.-L. Zhang, "Counting youtube videos via random prefix sampling," in *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*. ACM, 2011, pp. 371–380.
- [14] tianqi chen, kailong chen, and tong he, "Xgboost," 2015, <http://mloss.org/software/view/543/>.

Application of Computer Vision in Cricket: Foot Overstep No-ball Detection

AZM Ehtesham Chowdhury, Md Shamsur Rahim, Md Asif Ur Rahman

Department of Computer Science
American International University-Bangladesh
Dhaka, Bangladesh
{ehtesham, shamsur, asifur.rahman}@aiub.edu

Abstract — Cricket is a worldwide popular game where different technologies are being used to help the match umpires to make decisions. Often due to the human perception, deciding whether a bowled delivery is a no-ball or legal ball which causes controversy. As only a single ball can change the fate of the game, so it is obvious to make accurate decision regarding no ball. In this paper, we have divided the bowling crease into two regions and applied image subtraction method on both regions to find the change in pixel values. Later, we have applied our proposed method on real world video frames. Since our proposed method to detect the overstep no ball is grounded on pixel by pixel image subtraction, it eliminates the inadequacy of human perception.

Keywords—cricket; no-ball; popping crease; video feed; image subtraction; hough transform

I. INTRODUCTION

Cricket [1] is a game of bat and ball where a single delivery can turn the expected winner into loser. So, every delivery is counted as crucial moment for both teams. In every cricket match, umpires are responsible for deciding the approval of a ball bowled by a bowler. There are many scenarios when a delivery is disapproved by umpires. Some of the scenarios are declared as No Ball [2]. A No Ball can be declared for illegal actions by bowlers or fielders. As a consequence of a No Ball, the opposition team gets an extra run and delivery. Furthermore, batsman will not be given out except run out. One of the most common reasons is overstepping of popping crease by front foot during delivering the ball. But the examination of this kind of No Ball requires some minutes in certain cases using television replay. So umpires make their decisions on their perception. But human perception cannot be accurate all the time. Besides, it is not always possible to conclude the accurate judgment because of the limitations of existing technology [3]. In that case, the benefit of doubt [7] goes with the batting team. And this creates mass confusion and debates among the viewers and cricket lovers.

It is necessary to eradicate the debate to increase the acceptability of the decisions. If this can be achieved with the help of technology, then the game will be more fair and enjoyable. But it is challenging to develop such technology that can able to decide the overstepping No Ball in real time with higher accuracy. Previously proposed researches regarding this field have tried to solve the problem, but these

are infeasible, costly and prone to error because of the usage of sensors on field and bowlers. On the other hand, we have proposed computer vision based approach in this research paper which performs subtle pixel based distance analysis between feet and popping crease that will provide high degree of accuracy on decision. The proposed method does not need extra infrastructure because it will get video feed from broadcasting cameras. Our proposed method is expected to perform better and low cost in operating due to no infield sensor and other devices.

II. BACKGROUND STUDY

A. Cricket- The Game

Cricket is the game [1] where two teams play and each team has eleven players. In the center of the field, a 22 yards pitch is used to bowl and bat. One team is designated as the batting team after a toss and they attempt to score as many as possible runs. And the opponent team is designated as bowling team while they try to keep the run as less as possible. Each phase of the game is known as an innings. The winning team scores more runs than another team. In international cricket matches, two umpires provide infield decisions; the third umpire helps two field umpires on crucial decisions. There is a reserve umpire and match referee for managing and controlling the whole match.

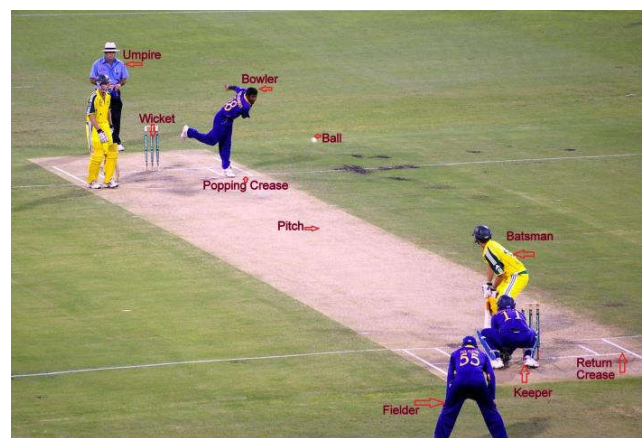


Fig. 1. A snippet of cricket match where one batsman is on striking position, another is on non-striking position on a pitch; a bowler is delivering ball;

wicket keeper and a slip fielder is to stop scoring along with more 9 fielders (not in picture); an umpire is standing behind non striking wicket to judge the ball.

B. Cricket Pitch

In cricket, a 22 yards pitch [4] is demarcated by some white marker lines painted or chalked on the middle of the field. These markers are called creases. These creases are used to determine the area of the operations of batsmen and bowlers. Four creases are drawn around a set of stumps on each side of a pitch; one popping crease, one bowling crease and two return creases (Fig. 2). The batsmen run between and play on popping creases of both ends.

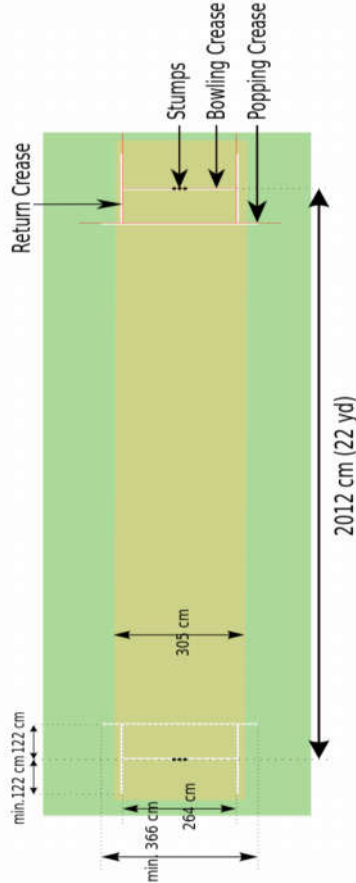


Fig. 2. A cricket pitch having types different creases [4].

The popping crease [6] determines the no ball for fielding side if the fielder stands up on the pitch or wicket keeper has moved in front of the wicket or the bowlers front foot oversteps the popping crease or behind the bowling crease or out of the return crease at the time of ball delivery. Sometimes the crease markers fade or vanish after 15-20 overs of bowling. Then it is difficult for the field umpires to determine whether a ball is a legal ball or no ball. Even the third umpire faces difficulties to judge the legality of the ball with the help of slow motion replay of several times because of error in the human perception [5].

C. Overstep No Ball

A delivery is considered overstepped no ball [2] if the heel of the front foot of bowler lands or in front of the popping crease. Fig. 2 shows the two states of heel bowler's front foot. In state 1, bowler's heel inside the no ball region and in others state bowler's some part of heel inside legal region.

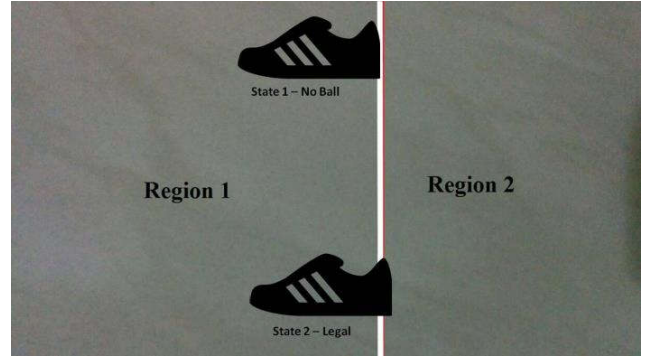


Fig. 3. A snippet of pitch having two states of bowler's heel; white bold marker is popping crease; red marker depicts the end of popping crease; from the red thin marker to front portion is region 1 or no ball region; backward portion is region 2 or legal region.

D. Related Works

From the last decade, various types of technological has been introduced in cricket. For instance, to detect events from broadcasting video, Rahish Tandon and Dr. Amitabha perform semantic analysis on captured videos involving the auxiliary cues [8]. Another advanced technology has been introduced named "Hawk Eye". It is used on cricket match as well in cricket coaching to detect the ball trajectory, which helps to analyze the bowling change, bowler's variations and help to decide "Leg Before Wicket (LBW)" out [9]. Another two technologies, Hot spot and Snicko-meter are using to detect the touch of ball with the bat. Cricket shot boundary detection and shot detection has been established based on multi-scale spatio-temporal analysis of color and optical flow features [10]. Another classifier has been established to classify different cricket shots using batsman's motion vectors during shot [11]. Comparing with other factors in cricket, very few researches have been performed for detecting no ball. Till now, two sensor based techniques have been proposed to detect no ball. Wazir Zada Khan, Mohammed Y Aalsalem and Quratul Arshad have used three sensors among them two are in the pitch and one in bowler's shoe [12] [13]. But it is not feasible to modify all of the shoes of bowler's to detect no ball. In another research, Mohanlal S Malu and Kailas G Dangi have proposed using the bowling crease and popping crease themselves as sensor by placing the sensors under the creases [14]. In this approach, the system will declare a legal ball as a no ball if the front foot is behind the popping crease. But it is not no ball according to the rule of cricket. Non striking batsman stands over and up of the popping crease at the other side of wickets. In this technique, this factor affects on detection as sensor will trigger on non-striker batsman's foot movement. In our system, we have removed the

dependence of sensors. We will use cameras that are used for broadcasting the match. The captured video feeds will be analyzed using computer vision based approach to detect the overstepped no ball. Various computer vision based approaches have been using from several decades to analyzing the image data for detection, recognition and tracking. David Lowe's SIFT features [15] and Timor Kadir's improvement [16] on SIFT have played a significant role on the field of computer vision based detection and analysis. Different extended tracking algorithms from Lucas Kanade's [17] [18] and Kalman filter [19] have played an important role on field of object detection, tracking as well as in action recognition. Subtracting background from MPEG video, a technique [20] has been proposed to detect objects and estimate motion of detected objects.

In this paper, we have proposed a method that will get video data of only bowler's front foot and popping crease during bowling. Our method has overcome the dependence on sensors as well as the effects of non-striking batsman on overstepped no ball detection and eradicate the error of umpire's perception on judging the legality of the delivery.

III. PROPOSED METHODOLOGY

In our research paper, we have proposed an image subtraction technique to analysis the video data to detect the overstep no ball. Our system model to detect the no ball is represented on following flow chart:

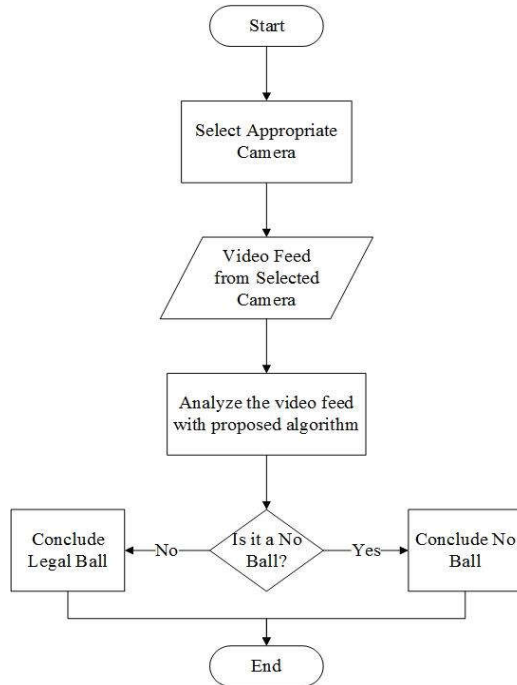


Fig. 4. System model of the No Ball detector

Our proposed system does not need any extra devices or sensors to detect the no-ball. In modern days, lots of high

definition cameras are set up on various angles to cover the match. At least 4 cameras are placed on both end of pitch aligning with the popping crease (see on Fig. 4). Our system will get video feed from these 4 cameras. At first, field umpires ask for the clarification from third umpire. The third umpire initiate system. He will select the appropriate camera based on the bowler's bowling end of pitch and side from wickets. Appropriate camera selection is important since the system will work on only bowler's front foot and bowling end popping crease. Then, the system has got the video feed from the selected camera for last ball delivery. A computer vision based algorithm has processed the frames of the video feed to detect the front foot positions of the bowler from the popping crease on the pitch.

The technique for no ball detection has been shown below:

A. Selection of Template or Base Frame

Before starting the game, a single frame is taken as the template or base frame for analysis. For every delivery the start frame is considered as base frame replacing previous base frame.

B. Marking Popping Crease

Popping crease is detected following the approach of line detection Hough transform which is proposed by Peter E. Hall and Richard O. Duda [21]. Then the white crease is marked. It stays static till end of the match.

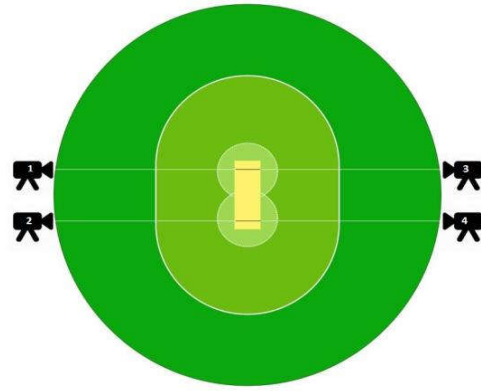


Fig. 5. Camera alignment for analysing crease related decisions.

C. Establishing Decision Regions

The crease marker divides the scene in two regions. For example, if the system get data from camera 2, all the pixel set from detected crease pixels to the left side will be considered as no ball region. Rest of the pixel set is considered as legal region (Fig. 3).

D. Comparing Template and Test Frames

When umpire asks for confirmation, the system will compare by subtracting base frame from all the frames. The change of pixel data can be detected for simple change in pixel data from changed frame to base frame. Mean is calculated for detected pixel changes for both regions.

E. Determining the threshold

To find out the threshold value of detector, at first, we have taken several lab generated images where foot heel of bowler is very much near right side or inside of decision marker regarding popping crease. Then, we have compared each of those images with template frame or image. Later, we have found the mean changes in pixels of region R2 for all of those images. The lowest mean change among all of the mean changes is considered as the threshold value of no-ball detector in our proposed algorithm. And the value is 0.53.

F. Decision of Legality

If the mean for region 1 is higher than R1_threshold and mean for region 2 is lower than R2_threshold, then system conclude the scene as no ball state. Otherwise, the scene is concluded as legal state.

Pseudo-Code for the no-ball detection algorithm (Considering Camera 2 and single frame comparison):

```

1: templateRGB, testRGB are arrays size 1280 x 720 x 3
2: templateFrame, testFrame are arrays of size 1280 x 720
3: creasePoint is the X co-ordinate or column number of
  the marked line or crease.
4: noBall = TRUE or FALSE
5: thresh = getThreshold();
6: templateFrame = rgb2gray(templateRGB)
7: testFrame = rgb2gray(testRGB)
8: creasePoint = lineHoughFunc(templateFrame) + 1
9: testFrame(1 to 720, creasePoint) = 1
10: templateFrame(1 to 720, creasePoint) = 1
11: changeInR1 = testFrame(1 to 720, 1 to creasePoint-1)
    - templateFrame(1 to 720, 1 to creasePoint-1)
12: changeInR2 = testFrame(1 to 720, creasePoint to 1279)
    - templateFrame(1 to 720, creasePoint to 1279)
13: meanR1 = mean(mean(changeInR1))
14: meanR2 = mean(mean(changeInR2))
15: if (meanR2 < thresh)
        noBall = TRUE
    else if (meanR2 > thresh)
        noBall = FALSE
    else
        noBall = FALSE

```

IV. EXPERIMENT RESULTS

To perform experiment on our proposed method, we have taken 7 different cases based on the distance between foot and popping crease under lab environment. For the three cases, the marker of popping crease is quite clear and for the rest the cases, the marker is slightly wiped out. To understand the error due to human perception, we have surveyed among 200 people with the 7 cases. Later on, we have applied our proposed method on the cases. From the survey we have found that above 90% people have judged delivery for case 1, 2, 3, 4 and 7 appropriately. On the other hand, because of the lacking in perception about 2% people can judge the delivery considering benefit of doubt [7] for case 5 and 6. Our proposed method has shown 100% accuracy rate on determining the legality of the delivery in case of overstepped no ball.

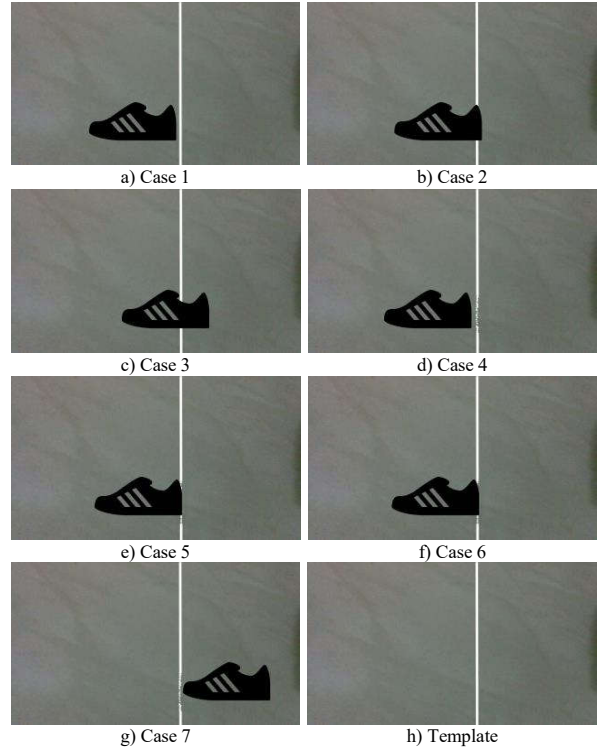


Fig. 6. Lab generated images of different cases of foot and popping crease and template used for proposed approach.

The table I represents the percentage of accuracy of no ball detection for human and proposed method.

TABLE I. RESULT ANALYSIS

Case Number	Actual Result	Human Accuracy (%)	Proposed Method's Accuracy (%)
Case 1	No Ball	93.5	100
Case 2	Legal	100	100
Case 3	Legal	100	100
Case 4	No Ball	95.5	100
Case 5	No Ball	1.5	100
Case 6	Legal	2	100

Case 7	Legal	100	100
--------	-------	-----	-----

V. FUTURE WORKS

In our approach to detect foot overstep no ball in cricket match; we have used image subtraction method which does not include sensors in cricket field. So, no distraction will be created for players. This analogy can be used for other no ball detection related to foot and crease and wide ball detection. Our definite aim is to develop a computer vision based automated cricket umpiring system.

VI. CONCLUSION

In this paper, we have analyzed and extensively implemented different computer vision based algorithms to detect overstepped no ball in cricket match.

In our proposed approach, firstly, we have established a template frame for each of the cameras where the marker of popping crease is detected using Hough transformation algorithm. Then, we have applied image subtraction analogy among the video frames and template frame from the umpire's selected camera. Our proposed method has eradicated the shortcoming of human perception to decide overstepped no ball on wiped out popping crease. To best of our knowledge and according to the several no ball detection approaches and their application, our approach can be effective and efficient.

References

- [1] H. De Sélincourt, *The cricket match*. London: Jonathan Cape Ltd, 1924.
- [2] "Law 24 (No ball)," *Lords.org*, 2016. [Online]. Available: <https://www.lords.org/mcc/laws-of-cricket/laws/law-24-no-ball/>. [Accessed: 23- May- 2016].
- [3] J. Reason, *Human error*. Cambridge [England]: Cambridge University Press, 1990.
- [4] "Cricket pitch," *Wikipedia*, 2016. [Online]. Available: https://en.wikipedia.org/wiki/Cricket_pitch. [Accessed: 24- May- 2016].
- [5] R. Steinmetz, "Human perception of jitter and media synchronization," *IEEE J. Select. Areas Commun.*, vol. 14, no. 1, pp. 61-72, 1996.
- [6] "Law 9 (The bowling, popping and return creases)," *Lords.org*, 2016. [Online]. Available: <https://www.lords.org/mcc/laws-of-cricket/laws/law-9-bowling-popping-and-return-creases/>. [Accessed: 24- May- 2016].
- [7] "Law 27 (Appeals)," *Lords.org*, 2016. [Online]. Available: <https://www.lords.org/mcc/laws-of-cricket/laws/law-27-appeals/>. [Accessed: 24- May- 2016].
- [8] R. Tandon, "Semantic Analysis of a Cricket Broadcast Video," pp. 1-9, 2009.[Online].Available:<http://www.cse.iitk.ac.in/users/rashish/FinalReport.pdf>
- [9] "Law 36 (Leg before wicket)," *Lords.org*, 2016. [Online]. Available: <https://www.lords.org/mcc/laws-of-cricket/laws/law-36-leg-before-wicket/>. [Accessed: 24- May- 2016].
- [10] D. S. Rughwani, "Shot Classification and Semantic Query Processing on Broadcast Cricket Videos Shot Classification and Semantic Query Processing on Broadcast Cricket Videos," no. September, 2008.
- [11] D. Karmaker, A. Chowdhury, M. Miah, M. Imran and M. Rahman, "Cricket shot classification using motion vector," *2015 Second International Conference on Computing Technology and Information Management (ICCTIM)*, 2015.
- [12] W. Z. Khan, M. Y. Aalsalem, and Q. A. Arshad, "The Aware Cricket Ground," *arXiv preprint arXiv:1109.6199* (2011), vol. 8, no. 4, pp. 269-273, 2011.
- [13] M. Y. Aalsalem, W. Z. Khan, "A General Architecture for Decision Making During Sports," *International Journal of Computer Science and Information Security* 9.8 (2011): 54.
- [14] M. S. Malu, "No Ball Detection," vol. 5, no. 3, pp. 111-113, 2015
- [15] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004.
- [16] T. Kadir, P. Hobson, M. Brady, and J. Close, "FROM SALIENT FEATURE TO SCENE DESCRIPTION," *Workshop on Image Analysis for Multimedia Interactive Services*. 2005, pp. 2-5
- [17] S. Baker and I. Matthews, "Lucas-Kanade 20 Years On: A Unifying Framework," *International Journal of Computer Vision*, vol. 56, no. 3, pp. 221-255, 2004.
- [18] B. Duvenhage, J. Delport, and J. de Villiers, "Implementation of the Lucas-Kanade image registration algorithm on a GPU for 3Dcomputational platform stabilisation," *Proceedings of the 7th*, 2010. [Online]. Available:<http://dl.acm.org/citation.cfm?id=1811172>.
- [19] D.a. Forsyth and V.O. Brien, "Computer Vision second edition," *Computer Vision: A Modern Approach* (2003): 88-101.
- [20] T. Yokoyama, T. Iwasaki and T. Watanabe, "Motion Vector Based Moving Object Detection and Tracking in the MPEG Compressed Domain," *2009 Seventh International Workshop on Content-Based Multimedia Indexing*, 2009.
- [21] R. Duda and P. Hart, "Use of the Hough transformation to detect lines and curves in pictures," *Communications of the ACM*, vol. 15, no. 1, pp. 11-15, 1972.

Capacity Optimization of Underwater Acoustic Communication System

A.Z.M. Imran*, Md. Murad Hossen*, Md. Tahidul Islam*

*Department of ETE,IIUC

azmimran28@gmail.com, muradhossen.ete92@gmail.com, tahidcce@gmail.com

Abstract—The underwater acoustics communication deals with few factors such as attenuation, multipath propagation, severely limited bandwidth and so on. It also has the little impact of regions sea environment as compared with terrestrial communication where the communication is going to be held. In this paper, it is evaluated the performance of underwater communication system with various parameters such as frequency, temperature and the optimal capacity was analyzed for non-coherent Differential Phase Shift Keying (DPSK) modulation technique. The comparative capacity analysis of Arabian Sea and Bay of Bengal also simulated for pointing the optimal frequency.

Keywords—DPSK, optimal capacity, optimal frequency, signal to noise ratio, under water communication.

I. INTRODUCTION

Underwater acoustic communication enhance a new era in the undersea research for the stability and the increasing demand of modern civilization. It deals oceanographic research, surveillance systems, navigation and the offshore exploration to disaster prevention and environmental monitoring [1,2] etc. And the arena of application of underwater communication system increases vying with time. The underwater environment is not suitable for radio frequency (RF) signal as the RF signal absorption high in water and also need larger antenna. So, Acoustics communication is the best option for undersea communication. The major challenges of underwater communication are: noise from multiple sources, attenuation, fading and availability of less bandwidth for communication. The low speed of sound in water (1500m/s) [3]. Thus, a wary design of underwater capacity has drawn well-deserved attention for the recent time. Several papers have investigated the capacity of underwater communication by considering varieties of factors. In [4], the impact of distance on capacity for underwater acoustic communication channel is observed, where transmission power is taken into account for different distances and frequencies. In order to estimate the channel capacity of shallow water, the broad band acoustic propagation modeling is applied in [5]. A. Radosevic et al. evaluated the capacity with different channel conditions for both single and multiple antennas [6]. Moreover full form Multiple Input Multiple Output (MIMO) based capacity analysis was accomplished by Ibrahim M. Alwan et al. [7] However, any of the paper did not emphasis optimal capacity exploration, which is substantially required for underwater communication system. An optimal frequency and furthermore the optimal capacity is a burning issue in the underwater acoustic communication system for future advancement. In this paper, we have evaluated perfor-

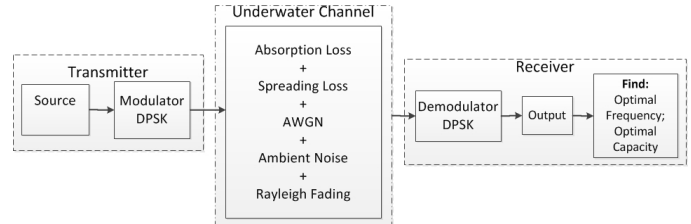


Fig. 1. System Block Diagram for Proposed UW Com.

mance of underwater communication in terms of signal to noise ratio (SNR) and capacity. The assessment of performance is accomplished by considering varieties of factors such as distance, frequency, temperature and depth. Moreover, the comparative optimal capacity of Bay of Bengal and Arabian Sea is presented for demonstrating the efficiency. sections of the paper contains acoustic syetem and propagation model , optimal pcapity anlysis ,simulation result and conclusion.

II. ACOUSTIC SYSTEM AND PROPAGATION MODEL

A. The System Model

There are three major blocks in the fig.1 proposed Under Water Communication System (UWCS) system diagram such as transmitter, receiver and the aquatic channel. Transmitter blocks contains data source, modulation amplifier and electrical to acoustic conversion using transducer. Receiver block contain acoustics to electrical conversion using, signal detection and amplification, demodulation and data storage or reuse. Aquatic channel consists of Absorption loss, Spreading loss, Additive White Gaussian Noise (AWGN), ambient noise, fading such as Rayleigh etc. Underwater Acoustic channels are volatile due to their dependence on ambient factors like temperature, salinity and acidity, leading to high local variances, large propagation delay and greater SNR and BER. In this paper we have considered the DPSK modulation due to simplicity and noncoherent detection as underwater acoustic have several constraints. Despite the incredible advances in methods for phase coherent processing over the past 15 years, incoherent systems still have their place and are still in use. Their biggest advantage is algorithmic simplicity, requiring modest processors and thus reducing cost and power. A number of commercial systems are available. Incoherent methods generally include frequency-hopping (with single or multiple tones transmitted in a given time instant), and M-ary frequency-shift keying using symbol

durations that are longer than the channel clearing time, and variations that employ alphabets of FM sweeps transmitted one at a time, or simultaneously in different bands.[8] DPSK modulation performs differential encoding and decoding of the message. Where the modulator makes decisions based on the difference of the two signals that's why the information data is encoded for the difference between adjacent signals. The encoding formula for DPSK modulation can be written as:

$$d_i = \overline{a_i \oplus d_i} \quad (1)$$

Where a_i is the raw message bit for i^{th} time and d_i is the encoded message. Also decoding formula:

$$a_i = \overline{d_i \oplus d_{i-1}} \quad (2)$$

If d_i and d_{i-1} are the same, then they represent a 1 of a_i . If they are different, they represent a 0 of a_i . This can be verified by comparing sequences $\{d_i\}$ and $\{a_i\}$. The sequence $\{d_i\}$ is modulated onto a carrier with phase 0 or π . In the absence of noise and other channel impairment, the demodulator output a_i is identical to the message sequence [9].

B. Channel Model

1) *Ambient Noise*: Underwater ambient noise represents any element or sound that tends to interfere with our ability to reliably transmit data from one point to another. Multiple human and natural factors contribute to the generation of noise in shallow water. Different sources of frequency-dependent noises in the underwater channel. Four main types of noise can be identified. They are turbulence, shipping, waves and thermal noise. The total power spectral density (PSD) of the ambient noise is the combination of its four components [10]:

$$N(f) = 10 \log_{10}(10^{0.1N_{th}} + 10^{0.1N_{turb}} + 10^{0.1N_w} + 10^{0.1N_s}) \quad (3)$$

2) *Attenuation*: Absorption Coefficient:

Attenuation occurs due to the conversion of acoustic energy to heat. The process of attenuation depends on frequency. So at higher frequency the absorption is higher that means attenuation. Several models are used to define the attenuation. Marsh and Schulkin empirical formula are considered here which is applicable for frequency range 3KHz and 0.5MHz. The absorption is stated as follows:

$$\beta(f) = 8.686 \times 10^3 * \left(\frac{S \alpha f_T f^2}{f^2 + f_T^2} + \frac{B f^2}{f_T} \right) * (1 - 6.54 \times 10^{-4} P) \quad (4)$$

Where f is in KHz, $\alpha = 2.34 \times 10^{-6}$, $B = 2.38 \times 10^{-6}$, S is the salinity (%), The hydrostatic $P = 1.01(1 + z * 0.1)$ in $[kg/cm^2]$, z is depth and $f_T = 21.9 \times 10^{\frac{6-1520}{T+273}}$ is the relaxation frequency (kHz).

Spreading loss:

Spreading loss is due to the ever-increasing area covered by the same amount of the sound signal energy, as a wave front moves outward from the source. It is given by

$$PL_{spread} = k * 10 \log(r) \quad (5)$$

Where the range(r) is in meters, and k is the spreading factor. When the medium in which signal transmission occurs is unbounded (spherical Spreading loss), the spreading factor is

$k = 2$, meaning that the source intensity decreases with the square of the distance(r). In the case of bounded spreading this factor takes different values. For example, $k = 1$ for a cylindrical boundary. Practical spreading coefficient of UWCS is $k = 1.5$ due to complex scenario of underwater. The overall attenuation is denoted by $A(X)$ and defined as [11]:

$$A(X, f) = X^k 10^{\frac{X \beta f}{10}} \quad (6)$$

In UWAC, SNR can be calculated based on signal attenuation and the noise PSD, specifically, the SNR observed at the receiver can be calculated by using the following equation:

$$SNR(X, f, T) = \frac{P_r}{A(X, f, T) * N(f)} \quad (7)$$

Where the $SNR(X, f, T)$ is over a distance (X) and a transmission center frequency f ; P is the signal transmission power.

III. OPTIMAL CAPACITY ANALYSIS

The Shannon capacity provides maximum data rate which can be transmitted through a channel. This quantity is defined as the maximum of the average mutual information $I(T_s R_s)$ between an input signal T_s and output signal R_s , where T_s is distributed over the channel. The Shannon Capacity formula is defined as:

$$C = B \log_2(1 + S/N) \quad (8)$$

The basic Shannon relationship shown in Equation (8) can be extended to be applicable in cases where the noise is dependent on frequency to take the form of:

$$C = B \log_2(1 + S(f)/N(f)) df \quad (9)$$

Thus for underwater communication the capacity is denoted by C' for x distance and frequency f ; and can be written as follows:

$$C' = f \log_2(1 + \frac{P_r}{N(f) * A(X, f)}) df \quad (10)$$

Where, P is the transmit power of the signal and can be written as:

$$P_r = \int_{-\infty}^{\infty} S(f) df \quad (11)$$

From (4) and (6), the equation can be written as:

See equation in page 3

The above underwater acoustic communication channel capacity determines the maximum data rate that can be supported by an acoustic channel for the different constraint. From the formula, in general, in order to obtain the maximum capacity, the attenuation and noise are required to minimum. Moreover, generally if the higher frequency is allocated, more data can be transferred, which intuitively augments the capacity. However, it is evident that for lower frequency, some noises are subtracted and for higher frequency more noises are added. Moreover, for higher frequency the absorption is higher and higher spreading loss for higher frequency, thus higher attenuation. As the noise and attenuation are dependent on several factors such as transmit power, frequency, distance

$$C(X, f) = f \log_2 \left(1 + \frac{P_r}{[8.686 * 10^3 * (\frac{S \alpha f_T f^2}{f^2 + f_T^2} + \frac{B f^2}{f_T}) * (1 - 6.54 * 10^{-4} P)] \times [10 \log_{10}(10^{0.1 N_{th}} + 10^{0.1 N_{turb}} + 10^{0.1 N_w} + 10^{0.1 N_s})]} \right) \quad (12)$$

between nodes, salinity of water and prevailed temperature of water, therefore all these factors are taken into account with a view to obtaining optimal capacity. Therefore, in order to obtain the maximum capacity it is required to select maximum frequency for which SNR should be maximum, which is an optimization issue. Thus the optimization formula for optimal capacity within distances X_1 to X_2 can be written as:

$$\begin{aligned} & \max_f \left(\frac{P_r}{A(X, f, T) * N(f)} \right) \\ & \text{Subject to} \\ & \{X_1 \leq X \leq X_2\}, \\ & \left\{ \frac{P_r}{N(f)} \times A(X, f, T) > \Phi \right\}, \\ & P_r \geq \rho_1 \end{aligned} \quad (13)$$

In the above optimization formula, the maximum transmit power is ρ_1 . It is essential to keep the SNR above threshold Φ . Moreover, for temperature ranges T_1 to T_2 the optimization formula can be stated as follows:

$$\begin{aligned} & \max_f \left(\frac{P_r}{A(X, f, T) * N(f)} \right) \\ & \text{Subject to} \\ & \{T_1 \leq T \leq T_2\}, \\ & \left\{ \frac{P_r}{N(f)} \times A(X, f, T) > \Phi \right\}, \\ & P_r \geq \rho_1 \end{aligned} \quad (14)$$

The maximum transmit power is ρ_1 for (14). In order to maximize capacity, the SNR above threshold Φ , the maximum frequency is investigated. In this case, the distance is remained constant.

IV. SIMULATION RESULT

In order to evaluate the optimal capacity, the simulation is performed by considering different parameter such as frequency, distance, depth, temperature and salinity etc. For each transmission distance, there exists an optimal frequency for which the maximum SNR is obtained. Also for different temperature, there exists an optimal frequency for which SNR is obtained. In general, the signal transmission power is assumed as $P_r = 20$ dB and the threshold is assumed as: $\frac{\Phi}{P_r} > 0$.

In fig.2, frequency versus SNR is demonstrated for different distances in $15^\circ C$. From the figure, it is depicted that SNR is better for lower distance between nodes. If the node distance decreases then a surging SNR is obtained. For instance, in 1 K. meter distance faces SNR below 0.1 dB, on the other hand 200 meter distance it is above 0.3 dB. Furthermore, in all cases, the optimal SNR is attained in the range of 2 KHz to 15 KHz for all distances.

The fig.3 describes effect of temperature fluctuation of the sea surface layers on SNR for different frequency for 1 Km

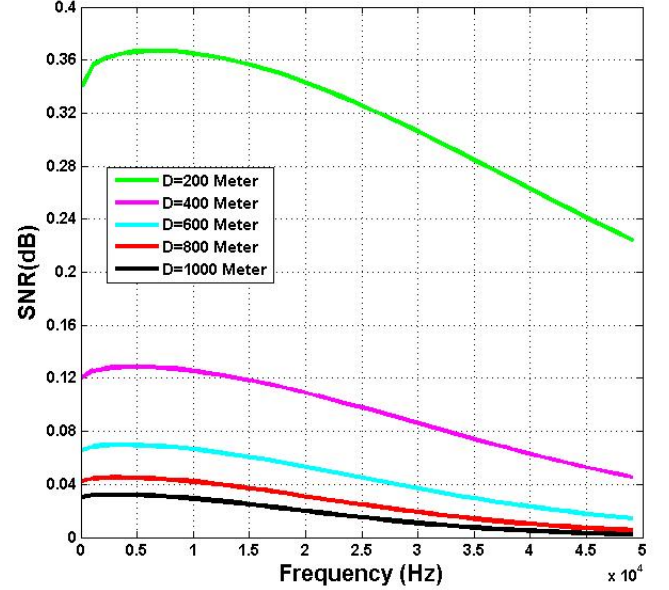


Fig. 2. Frequency versus SNR for different distances

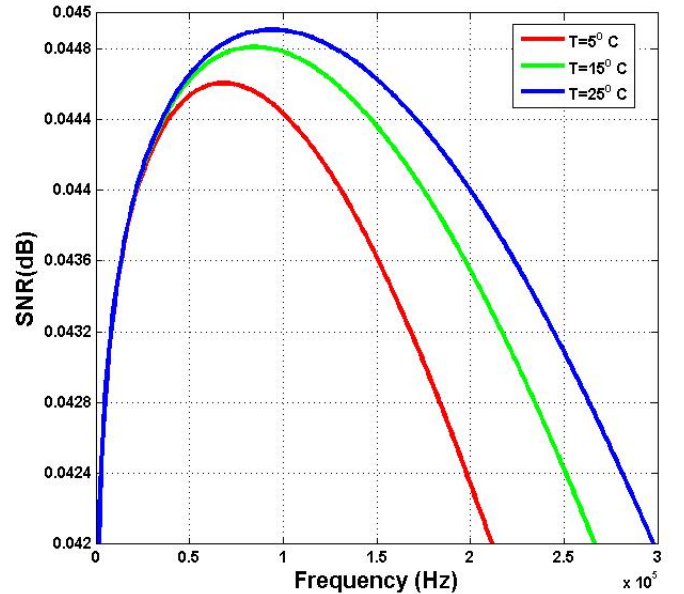


Fig. 3. Frequency versus SNR for different temperatures

distance. The highest SNRs are achieved in approximately 70 KHz, 90 KHz and 100 KHz for 5° , 15° $25^\circ C$ temperatures respectively. Also it is clear that the higher temperature have higher SNR. If we increase the distance between two

nodes the optimal frequency decreases and the temperature variation effect becomes less important.

In Fig.4, the capacity analysis is performed for different distances. In the figure, it is depicted that capacity is higher with lower distance and the capacity goes down at longer distance between sending and receiving point. For illustration, in lower distance 200 meter the capacity is the highest and approximately 800 Bits/sec in 50 KHz. As the distance increase the capacity decreases and also optimal frequency decreases. In 1 Km distance optimal frequency is the lowest (19 KHz approximately).

In Fig.5 the capacity variation for temperature are described. At 20°C temperature the capacity is the highest but it is noticeable that as the depth increases the temperature becomes lower than the SNR as well as capacity decreases. In Fig. 6,

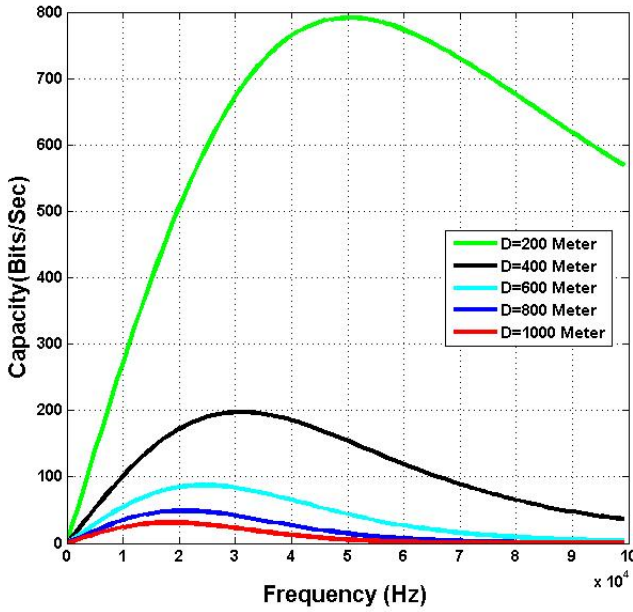


Fig. 4. Frequency versus capacity for different distances

the comparative analysis is done for two sea named Arabian Sea and Bay of Bengal for different depth as well as all channel factor. As the depth rises between two nodes, the temperature and salinity change due the environment of the Sea [12]. It is assumed that two nodes are in 100 meter distance. This provides a clear indication about optimum frequency for which capacity will be efficient. Also the Fig.6 depicts that the optimal capacity in between 30 to 50 KHz for 1 Km distance, for those both Seas. Within the optimal frequency range the optimal capacity for Bay of Bengal is lower than the Arabian Sea due to the rapid fluctuation of temperature and salinity of the Bay of Bengal. From the above discussion and graphs it is clear that optimal frequency in 200m distance for Arabian Sea is approximately 61 KHz and for Bay of Bengal it is less than 61 KHz. For 50 meters distance the optimal frequency is above 700 bits/sec for both sea. For Arabian Sea the optimal frequency is approximately 48 KHz but for Bay of Bengal it is lower than 48 KHz.

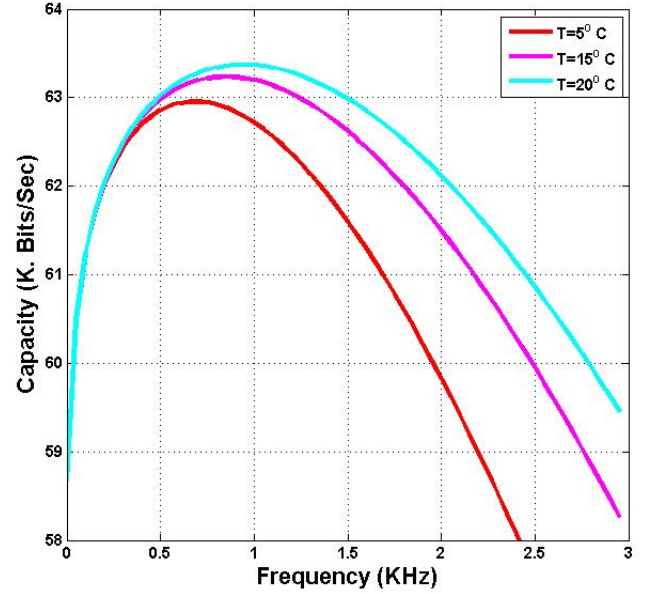


Fig. 5. Frequency versus capacity for different temperatures

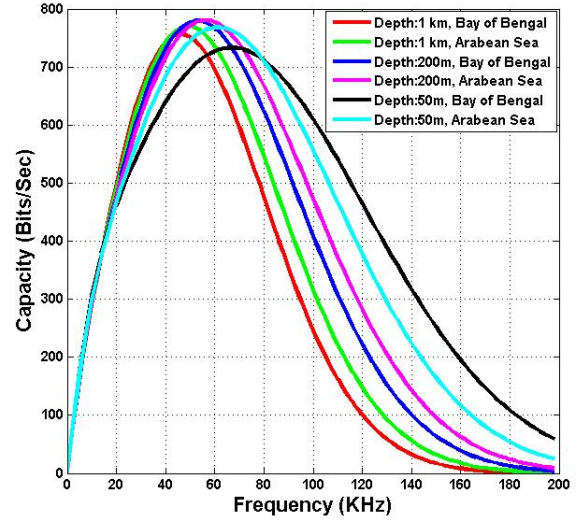


Fig. 6. Frequency versus capacity for Arabia Sea and Bay of Bengal

V. CONCLUSION

High frequency underwater signal are advantageous due to the high bandwidth but main drawback is the presence of attenuation factors that adversely affect the SNR and finally the bandwidth as well. In the paper optimal capacity analysis is performed for DPSK modulation and also the comparative analysis is simulated for the different sea as the acoustic communication defer with the environment of the place of interest. The paper also insisted on DPSK non-coherent modulation technique which provide simplicity in the detection process as the coherent system detection process is complex. The comparative capacity analysis shows that the capacity

fluctuation occurs more in Arabian Sea than Bay of Bengal.

REFERENCES

- [1] I. F. Akyildiz, "Underwater Acoustic sensor networks: research challenges", *Ad Hoc Networks* (Elsevier), 3(3):257–279, May 2005.
- [2] J. Heidemann et al, "Research Challenges and Applications for Underwater Sensor Networking", *Wireless Communications and Networking Conference, WCNC-2006*. IEEE, 3-6 April 2006.
- [3] M. Stojanovic, *Wireless Underwater Communications and Networks: Current Achievements and Future Research Challenges*, IEEE Oceanic Engineering Society Newsletter, vol. 41, no. 2, 2006.
- [4] M. Stojanovic, On the Relationship Between Capacity and Distance in an Underwater Acoustic Communication Channel, *Proceedings of the 1st ACM international workshop on Underwater networks*, September 25-25, 2006, Los Angeles, CA, USA.
- [5] T.J.Hayward et al, "Underwater Acoustic Communication Channel Capacity : A Simulation Study" *AIP Conference Proceedings*, pp.114 -121 2004.
- [6] A. Radosevic, Statistical characterization and capacity of shallow water acoustic channels, *OCEANS 2009-EUROPE*, 11-14 May 2009, IEEE.
- [7] Ibrahim M. Alwan et al, Capacity Analysis of MIMO Hydrophone in Underwater Acoustic Communication ,*International Journal of Computer Science and Engineering Technology (IJCSET)*, Vol. 4 No. 08 Aug 2013.
- [8] Mandar Chitre et al, Recent Advances in Underwater Acoustic Communications and Networking, *OCEANS 2009, MTS/IEEE Biloxi -Marine Technology for Our Future: Global and Local Challenges*, 26-29 Oct. 2009.
- [9] Couch II, L.W (1990)*Digital and Analog Communication Systems*, 3rd Edition New York.
- [10] Prof. Dr. I.D. Kraus, "Underwater Acoustics, D-28199 Bremen, Germany.
- [11] G. ZAIBI et. al, Survey of Temperature Variation Effect on Underwater Acoustic Wireless Transmission , *5th International Conference : Sciences of electronic, Technologies of Information and Telecommunications*, 2009.
- [12] A pocketbook on the ocean with special reference to the waters around India. http://www.nio.org/index/option/com_nomenu/task/show/tid/2/id/140/

Impact of Cyber-Attack on Isolated Power System

Md. Sahabuddin¹, Bilash Dutta², Mehedi Hassan³

Department of Electrical and Electronic Engineering
Jessore University of Science & Technology
Jessore, 7408, Bangladesh

E-mail: ¹shabuddin_jstu@yahoo.com, ²bilashdutta11@yahoo.com, ³mhj24.jsr@gmail.com.

Abstract -In this modern era of civilization demand of electricity from interconnected grid system is increasing day by day. That is why improving the reliability of the power system has become the main concern to the experts. This article investigates the impact of cyber-attack on an isolated power station in frequency disturbing aspects during sudden changes of load of the system. Stable limit of the speed regulation for load frequency control (LFC) and integral controller gain for automatic generation control (AGC) is derived from their characteristic equations. Depending upon the nature of cyber-attack (positive biased attack or negative biased attack) simulation curves are developed to show the frequency deviations and oscillations of the power system. Most severe attack is also characterized to protect power system from unstable situation.

Index Terms – cyber-attack; advanced control; cyber-security; LFC; AGC

I. INTRODUCTION

Modern power system is largely dependent on information and communication technologies and digital computer techniques [1], [2]. With the advancement of information and communication technologies (ICT) and digital computer advanced controllers are increasingly integrated into power generation and control system [3], [4]. These technologies greatly increase the dependency of the power system on ICT that paves the way for cyber-attack.

The cyber component of a power station is the advance control loop, communication networking and sensors [5], [6]. The data for different parameters are collected using sensors and then these data are sent to controller [5], [7], [8] for right operation of the system. The data sending action are performed using communication network. This communication network is one of the cyber vulnerable parts of the power system.

The attacker can get the data when sending from sensing section to control section in a power station. After accessing the confidential data the attacker can get control on the automatic controller of a power station. By getting control over the controller the attacker can change the controller action, this results in malfunctioning of controller. The malfunctioning leads to make the system unstable.

To ensure system stability the impact of cyber-attack should be analyzed properly. Therefore, the aim of this paper is to investigate the way as well as the impact of cyber-attack on power system. The most severe attack on isolated power station is also identified.

The rest of the paper is organized as follows. Cyber-attack on LFC is given in Section II, cyber-attack on AGC is given in Section III, and conclusions are given in Section IV.

II. CYBER-ATTACK ON LFC

Load frequency controller is an important part of a power station. For stable operation of power station proper operation of LFC is so important. But problem is that LFC is a great means of cyber-attack. Load of a power station is always varying in nature. It is known that with the change in load of a power station the frequency of the alternator changes. The change in frequency may cause the system unstable. Again it is not possible to make the load fixed and the only factor in hand is to control the speed of the prime mover. Here LFC plays a vital role. The frequency sensor senses the system frequency [9] and according to the signal from frequency sensor the LFC sets the prime mover speed to compensate the frequency. Depending on the governor speed regulation it controls the speed of the prime mover and the system frequency.

Malfunctioning of the governor speed regulator due to cyber-attack, the speed of the prime mover may not be under control and this result in making the system frequency unstable. If the control system is in healthy condition that means it is not attacked by unauthorized people then sudden load change (load increases) causes the system frequency falls below the nominal frequency for a short instant of time. The frequency sensor senses the fall of the system frequency and sends signals to the LFC for proper speed regulation. According to the signals from frequency sensor the LFC sets the governor speed regulation to compensate the prime mover speed and to system frequency. Due to cyber-attack the LFC may not be able to set the regulation properly and if this occurs then the system frequency will oscillate and this will make the system unstable.

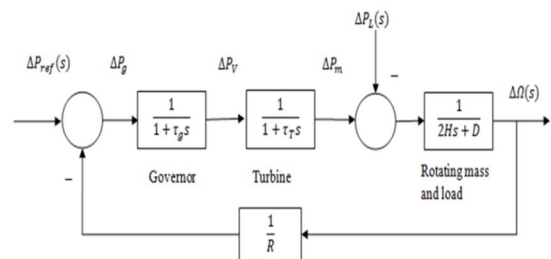


Fig. 1 LFC block diagram [9]

For stability of the system frequency speed regulation setting plays important role. Stable limit of the speed regulation (R) can be derived from the characteristic equation of the Load Frequency Control loop using Routh-Hurwitz array. The transfer function of the LFC is the following.

$$\frac{\Delta\Omega(s)}{-\Delta P_L(s)} = \frac{(1 + \tau_g s)(1 + \tau_T s)}{(2Hs + D)(1 + \tau_g s)(1 + \tau_T s) + 1/R} \quad (1)[3]$$

The following parameters of an isolated power system are assumed [9].

TABLE I. TABLE FOR ASSUMED PARAMETER

Parameter	Value
Turbine time constant (τ_T)	0.2 sec
Governor time constant (τ_g)	0.5 sec
Generator inertia constant (H)	5.0 sec

Load variation is assumed to be 0.8 percent for 1 percent change in frequency and turbine output power is 250MW at a nominal frequency of 50Hz with a sudden load change of 50MW [9]. From above equation using the assumed data by using Routh-Hurwitz array the limit of stability of speed regulation is $R > 0.0135$. To make stability visible, simulation is done in Matlab environment. From Fig. 2 it is seen that when $R=0.05$, the system frequency deviation is in stable condition. So for stable operation of this power station the set value of R (speed regulation) should be 0.05. Any deviation from this set value makes the system frequency deviation unstable. The instability of system frequency deviation means making the governor incapable of compensating the frequency deviation i.e. the system frequency will not be restored in case of sudden load change. The deviation of speed regulation from set value is due to cyber-attack on LFC (Unauthorized access to the control of LFC). Here the cyber-attack is classified into two types, first one is positively biased attack and the second one is negatively biased attack depending on the affected value of speed regulation(R).

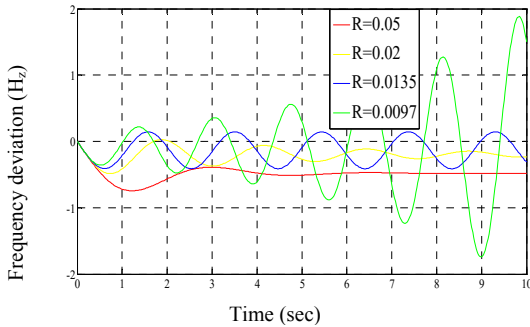


Fig. 2 Performance curve of LFC for different values of R

A. Result of positively biased attack

When the value of speed regulation increases due to cyber-attack, it is termed as positively biased attack. From the following figure it is clear that as the speed regulation increases from the value 0.05, the frequency deviation curve comes to stable condition quickly although the frequency falls down more. It will be possible for the governor to compensate the frequency deviation if the deviation comes to stable condition quickly. In case of positive bias the system frequency falls down but comes to stable condition quickly. But too much deviation of system frequency is not allowable. So it can be said that in case of positive bias the system frequency can be restored without disturbing the stability of the system frequency. But too much deviation of system frequency may allow the governor to take more time to set the desired position to restore the system frequency.

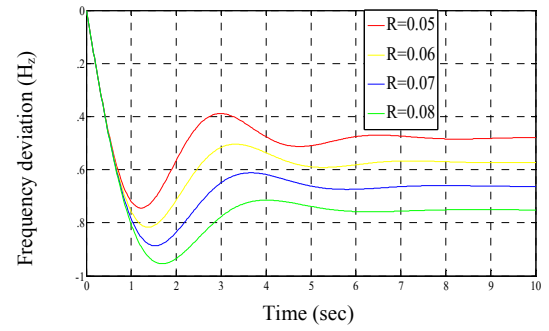


Fig. 3 Frequency deviation in case of positively biased attack.

B. Result of negatively biased attack

The decreasing of speed regulation value due to cyber-attack is termed as negatively biased attack. From the above discussion it is seen that positive biased attack is not so serious because after attack the governor is capable to restore the system frequency by increasing the fuel supply i.e. increasing the speed of turbine because the deviated curve comes to stable position after short time. In the following figure (Fig. 4) the red curve ($R=0.05$) is the set value of speed regulation. The yellow curve ($R=0.03$) takes more time to come to stable position than red one as the yellow curve's corresponding value is within the limit of stable condition ($R > 0.0135$) but close to oscillating value. For $R=0.03$ it is possible for the governor to restore the system frequency but it takes more time than the set value. For other two cases it is clear that the frequency deviation curve never comes to stable position. That is why it is impossible for the governor to set a fixed point to compensate the frequency deviation.

Due to attack if the speed regulation value falls below the stable condition ($R > 0.0135$) the frequency deviation will be oscillating in nature and also tends to increase the oscillation. For this instability of frequency deviation governor fails to restore the system frequency. So it is clear that negatively biased cyber-attack is the major challenge for the engineers to maintain the system frequency in stable condition during sudden change of load on the system. Negative biased attack is so serious in case of maintaining system stability and this type of attack may collapse the

whole system stability. Negative biased attack is the major concern to the engineers.

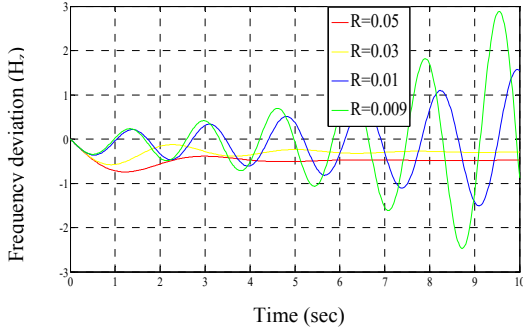


Fig. 4 Frequency deviation in case of negatively biased attack.

III. CYBER-ATTACK ON AGC

With the primary LFC loop, a change in the system load will result in steady state frequency deviation, depending on governor speed regulation [8]. The primary LFC loop takes a considerable time to restore system frequency that is not allowable. In order to reduce the frequency deviation to zero a modification is needed to LFC. The modification can be achieved by introducing an integral controller to act on the load reference setting to change the speed set point. The integral controller increases the system type by one which forces the final frequency deviation to zero [8]. The LFC system, with the addition of secondary loop (integral controller) is the Automatic Generation Control (AGC). The integral controller gain K_I must be adjusted for satisfactory transient response. In the figure (Fig. 5) various functional parts of AGC are showed. AGC is nothing but LFC with parallel combination of integral controller. The addition of integral controller allows the governor to set appropriate point to increase speed of turbine in case of sudden load rise and to restore the system frequency quicker than primary LFC loop. For proper operation of the AGC appropriate value selection of integral controller gain (K_I) is the most important task when working with AGC. In case of failure of the selection of appropriate value selection of K_I leads to malfunction of governor to set appropriate point to restore the system frequency. Here K_I is regarded as the vulnerable quantity due to cyber-attack. Any change of K_I due to unauthorized access to AGC loop may cause the system frequency oscillation which disturbs the system stability. The harmful effect of the attack on AGC is analyzed here.

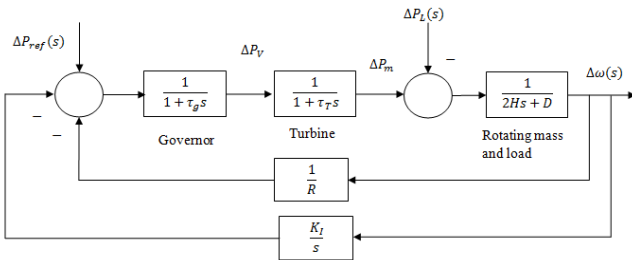


Fig. 5 Block diagram of AGC [9].

The closed loop transfer function of the control system shown in above figure is

$$\frac{\Delta\Omega(s)}{-\Delta P_L(s)} = \frac{(1 + \tau_g s)(1 + \tau_T s)}{(2Hs + D)(1 + \tau_g s)(1 + \tau_T s) + K_I + s/R} \quad (2) [9]$$

To determine the right value of K_I at which frequency deviation is zero, the frequency deviation step response is shown in the following figure.

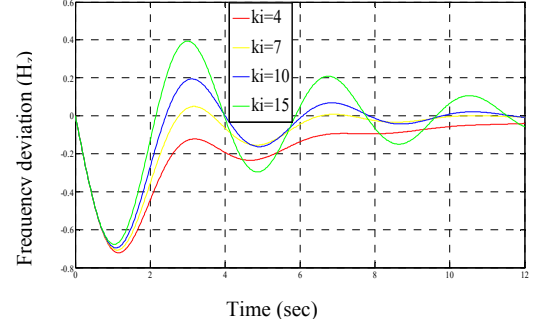


Fig. 6 Frequency deviation step response for different values of K_I .

A. Result of positively biased attack on AGC

Due to unauthorized access to AGC control system power system may lead to system frequency deviation and sometimes make the system frequency deviation oscillating in nature i.e. makes the deviation unstable. If the system frequency deviation is oscillating in nature, it will be impossible for governor to restore the system frequency. Moreover the attack on AGC may cause unwanted delay to restore system frequency. In the following figure (Fig. 7) the effect of positively biased attack on AGC is shown and the red curve is the set value under unaffected condition. In the figure the yellow curve means that there is slight frequency deviation from the set condition (Red curve). This leads to unwanted delay to restore the system frequency. The blue curve means that the frequency deviation is slightly oscillating in nature. The oscillation of the frequency deviation leads to the instability of system frequency. Due to this oscillation the governor will fail to set the appropriate point to restore the system frequency. The green curve means that the deviation oscillation is higher and this also indicates serious attack on AGC and the governor will no longer be able to restore the system frequency. From this discussion it can finally be said that more the positive biased attack, more serious impact on system frequency is appeared.

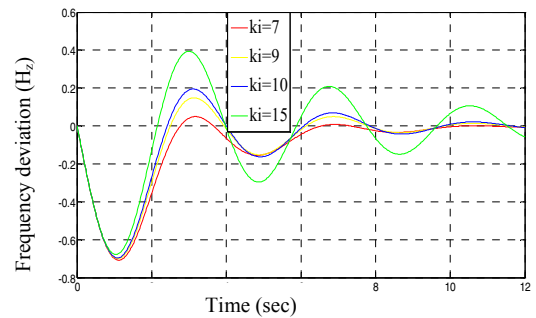


Fig. 7 Frequency deviation in case of positively biased attack.

B. Result of negatively biased attack on AGC

The seriousness of the attack on AGC depends on the type of attack. Negatively biased attack means the falling down of the value of K_I from the set value due to unauthorized access to AGC system. This attack may oppose the objectives of the use of integral controller with primary loop of LFC to obtain AGC system to reduce the frequency deviation to zero. The result of negatively biased attack on AGC is shown in the figure (Fig. 8). The primary LFC control loop is modified to form AGC to reduce the frequency deviation to zero. From the figure (Fig. 8) it is clear that due to negative biased attack, frequency is deviated from nominal frequency. So it can be said that the negative biased attack makes the integral controller ineffective and opposes the objective of the use of integral controller. The negative biased attack leads to the unwanted delay of restoring the system frequency. Negative biased attack on AGC is not so serious as like positive one because it does not make the frequency deviation oscillating in nature.

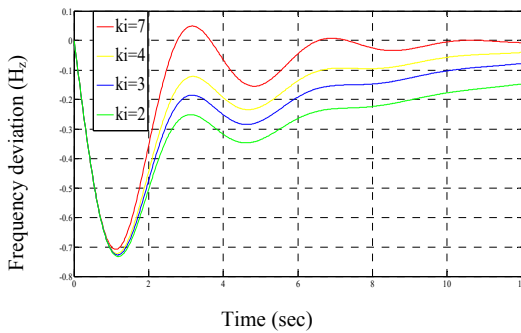


Fig. 8 Frequency deviation in case of negatively biased attack.

IV. CONCLUSION

This paper analyses the impact of cyber-attack on an isolated power station in frequency disturbing aspects during sudden changes of load of the system. To make the system stable, proper function of LFC as well as AGC should be ensured first which can be a suitable means of cyber-attack. Depending upon the nature of cyber-attack (positive biased attack or negative biased attack) it is observed that positive biased attack on LFC causes the system frequency falls down but comes to stable condition quickly whereas it develops serious oscillation in frequency on AGC that makes the governor unable to set the appropriate point to restore the system frequency. For AGC, negative biased attack is not as serious as like positive one but for LFC, frequency deviates highly with serious oscillation. To ensure cyber security it is important to know the impact of attack that is done in this work. Ensuring cyber security of the control system against unauthorized access and mitigating the seriousness of cyber-attack is the future aim of this work.

REFERENCES

- [1] Rahul Nema, Anurag Trivedi, "Load Frequency Control of a Small Isolated Power Station by Using Supercapacitor Based Energy Storage System," International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), vol.1, pp. 162-167, December 2012.
- [2] Bo Chen, Salman Mashayekh, Karen L. Butler-Purry, DeepaKundur, "Impact of Cyber Attacks on Transient Stability of Smart Grids with Voltage Support Devices," IEEE Power and Energy Society General Meeting (PES), pp.1-5, 2013.
- [3] Bo Chen, Karen L. Butler-Purry, DeepaKundur, "Impact Analysis of Transient stability Due to Cyber Attack on FACTS Devices," North American Power Symposium (NAPS), pp.1-6, 2013.
- [4] Yilin Mo, Tiffany Hyun-Jin Kim, Kenneth Brancik, Dona Dickinson, Heejo Lee, Adrian Perrig, Bruno Sinopoli, "Cyber-Physical Security of a Smart Grid Infrastructure," Proceedings of the IEEE, vol. 100, pp. 195 – 209, 2012.
- [5] Abdallah K. Farraj, Eman M. Hammad, Ashraf Al Daoud and DeepaKundur, "A Game-Theoretic Control Approach to Mitigate Cyber Switching Attacks in Smart Grid Systems," IEEE International Conference on Smart Grid Communications, pp. 958-963, 2014.
- [6] Mirjana Milosevic, Philippe Rosa, Marius Portmann, Goran Andersson, "Generation Control in Small Isolated Power Systems," IEEE Power and Energy Society General Meeting (PES), pp.1-8, 2007.
- [7] Abdallah Farraj, Eman M. Hammad, DeepaKundur, "On Using Distributed Control Schemes to Mitigate Switching Attacks in Smart Grids" IEEE 28th Canadian Conference on Electrical and Computer Engineering. Halifax, Canada, pp. 1578-1582, May 3-6, 2015.
- [8] Abdallah K. Farraj, Eman M. Hammad, Jin Wei, DeepaKundur, Karen L. Butler-Purry, "Performance of Flocking-Based Control Schemes in Smart Grid Applications," IEEE Global Conference on Signal and Information Processing for Energy Exchange and Intelligent Trading (GlobalSIP-14), pp. 233-237, 2014.
- [9] Hadi Saadat, Power – System – Analysis, Chapter – 12.

Comparative Analysis of PV Module Efficiency for Different Types of Silicon Materials Considering the Effects of Environmental Parameters

Md Faysal Nayan¹, S.M.Safayet Ullah², S. N. Saif³

Department of Electrical and Electronic Engineering, Ahsanullah University of Science and Technology, Dhaka, Bangladesh

¹faysal.nayan@gmail.com, ³aabeersaif@gmail.com

Department of Natural Sciences, Daffodil International University, Dhaka, Bangladesh

²s.m.safayet.ullah@gmail.com

Abstract— To meet the energy demand, solar photovoltaic has already emerged as a convincing renewable energy technology. Solar Photovoltaic (PV) modules are one of the most effective, sustainable, and eco-friendly systems in recent world. Different types of PV solar technologies like Mono crystalline silicon, polycrystalline silicon, amorphous silicon, thin film are the most popular technologies to produce electricity. Their module performance and efficiency depend on the electrical and environment parameters of the PV materials. This paper presents some theoretical evaluations for variation in the efficiency of energy for different types of photovoltaic (PV) module under changes of environmental parameters. A proper comparison is developed among the mono-crystalline, polycrystalline and amorphous material using the model parameters. This work also illustrates the variation of fill factor and efficiency of a specific photovoltaic module under real operating conditions. MATLAB is used to investigate the P-V characteristics to evaluate the maximum power point of solar photovoltaic cell considering the effect of temperature and solar radiation.

Keywords—Solar cell; solar radiation; PV array; temperature; amorphous; mathematical model; maximum power point; crystalline; fill factor; efficiency

I. INTRODUCTION

Now a day's energy shortage is worldwide concern issue. In order to overcome the obstacles of energy crisis, the sources of renewable energy have become a major issue in the recent years [1]. Renewable energy is used to generate electrical power from the solar energy using the concept of Photovoltaic (PV). It has come into view for its great advantages and less maintenance [2-3].

Among the solar technology, the photovoltaic array (PV) has been making a positive attention due to its capability of energy conversion without intermediate thermal process ([4]-[6]). If the solar energy is used in right track, it can meet the need of having sufficient energy of the world ([1],[7-9]). Photovoltaic solar cell systems always shows non-linear voltage current relationship which varies with different types of environmental and electrical parameters [1].

Most commonly used materials for Photovoltaic energy conversion are mono-crystalline, poly-crystalline silicon, amorphous silicon and thin Film technologies. Silicon crystalline have different types of structure's such as single crystalline silicon, multi- crystalline silicon and ribbon cast multi-crystalline silicon [8].

Mono-crystalline solar panels known for higher efficiency, space efficient and longevity but this technology is most expensive [9]. Now a days Poly-crystalline solar panels have lower efficiency but this technology is cost effective and most popular. Solar thin film technologies based on amorphous silicon is now a major research topic. Though their efficiency is not good as mono-crystalline and poly-crystalline silicon technology, their mass production is simple and high temperatures and shading have less impact on their performance. However, performances of these type of modules depend on various types of electrical and environmental parameters.

In this paper, our study mainly focuses on analysing three-type PV materials [5-7]. To this aim, three different PV systems consisting of mono-crystalline, poly-crystalline and amorphous silicon materials were considered to make a comparison in terms of the efficiency [5]. Environmental parameters including solar radiation and temperature are varied to observe the characteristics of different Si based solar PV technologies under different circumstances.

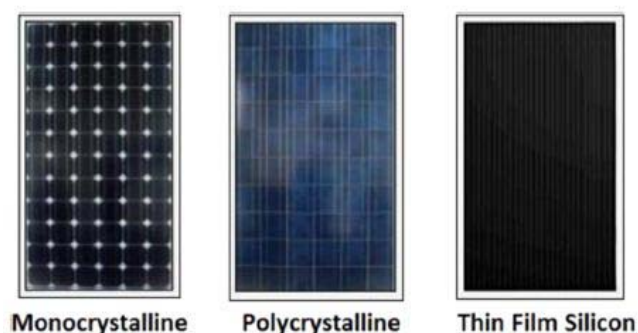


Figure 1: Different silicon technologies

II. PHYSICS OF SOLAR CELL

Photovoltaic systems (PV) are made from semi conducting materials and convert photon energy into electricity. When sufficient amount of sunlight reaches these materials, photons with a specific wavelength trigger electrons to flow through the materials to produce direct current (DC) electricity ([5],[10-14]). Most importantly, semiconductor materials are used to make the solar panels and among these materials, silicon is generally used. In order to achieve higher output

voltage, solar cells are connected in series along with the parallel connection for producing higher output current ([1-4],[11-17]).

Photovoltaic cells are generally connected in series configuration to form a PV module. The modules can be arranged to form large PV panel. Panels can be grouped to form large photovoltaic arrays [18-22]. The operating principle of solar cell may be described from a PN junction where there are drift and diffusion currents for forward and reverse polarization. The influence of electric fields separates these carriers within the junction. As a result, a current is produced which is proportional to solar irradiation incidence [7,10, 11].

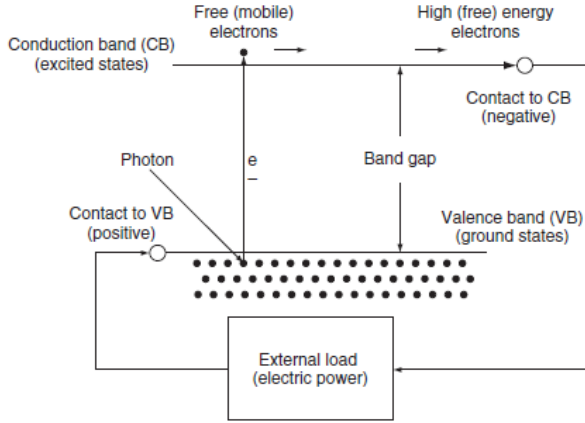


Figure 2. Physics of solar cell in terms of band theory.

III. MODELING OF PHOTOVOLTAIC ARRAY

Solar cell is one kind of device that is fabricated in a thin layer of semiconductor. A typical solar cell device converts light energy into electrical energy ([14]-[18]).

A. Equivalent circuit of solar cell

To analysis the characteristics of different types solar cell materials we have considered a single diode model equivalent circuit (Figure 3) to implement a solar cell model. We have derive the equations of output voltage, output current, Fill factor (FF) and Efficiency (Ef) in terms of various parameters ([1],[12],[17]-[19]).

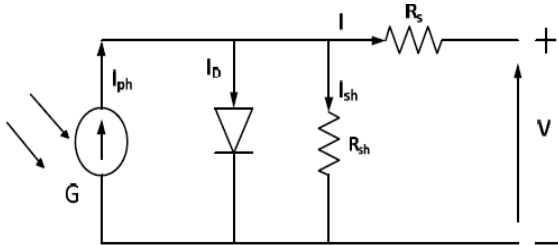


Figure 3. Equivalent circuit of solar cell

B. Fundamental Mathematical equations of Solar Cell

The basic equations that represent the operating principle of solar cell is based on the equivalent circuit are given below ([1],[11]-[17], [23-24]).

Applying the KCL on the equivalent circuit, we can write the equation of photocurrent which is given in equation (1).

$$I_{ph} = I_D + I_{sh} + I \quad (1)$$

In this equation, I_{ph} is Photo current generated in PV cell which depends on the solar radiation (G). I_D and I_{sh} are Diode current and shunt current respectively. I represent the Load current of the solar cell ([1],[11]-[17], [23-24]).

Semiconductor diode is a nonlinear device and the basic fundamental characteristic of semiconductor diode is defined by the following exponential equation ([1], [24]).

$$I_D = I_S \left[\exp\left(\frac{V_D}{V_T}\right) - 1 \right] \quad (2)$$

Here, I_D and V_D represent the diode current and voltage across the semiconductor diode. I_S signifies Temperature dependent reverse saturation current and V_T is Thermal voltage which depends on reference temperature (T_C) and ideality factor of diode (n).

$$V_T = \frac{nkT_C}{q} \quad (3)$$

Here, n= Diode ideality factor,
 k = Boltzmann constant ($1.3806503 \times 10^{-23}$ J/K)
 q = charge on electron (1.602×10^{-19} C)

If we operate the semiconductor diode in reverse bias region then reverse saturation current will be I_{SO} .

$$I_{SO} = \frac{I_{SC}}{[\exp(\frac{V_{OC}}{N_s V_T}) - 1]} \quad (4)$$

Where, V_{OC} is Open circuit voltage per cell in Volt, I_{SO} is Diode reverse saturation current, I_{SC} means Short circuit current per cell at reference temperature and solar radiation.

N_s and N_p represent the Number of solar cell connected in series and Number of solar cell connected in parallel respectively.

The following equation represents temperature dependence of diode the saturation current([1],[11]-[17], [23-24]).

$$I_S = I_{SO} \left(\frac{T_O}{T_C}\right)^3 * \exp\left[\frac{qE_g}{kn} \left(\frac{1}{T_O} - \frac{1}{T_C}\right)\right] \quad (5)$$

Photo current generated in PV cell mainly depends on solar radiation and temperature. The value of Photocurrent is linearly proportional to the amount of solar radiation. The equation of photo current based on temperature and solar radiation is represented by the equation (6) ([1],[11]-[17], [23-24]).

$$I_{ph} = [I_{SC} + K_{SC} (T_C - T_O)] * \left(\frac{G}{G_S}\right) \quad (6)$$

We can write the equation of load current from basic equation as follows:

$$I = I_{ph} - I_S \left[\exp\left(\frac{V + IR_s}{V_T}\right) - 1 \right] - \frac{V + IR_s}{R_{sh}} \quad (7)$$

$$I = N_p I_{ph} - I_s \left(\exp\left[\left(\frac{q}{nkT_o}\right)\left(\frac{V}{N_s}\right)\right] - 1 \right) - \frac{V}{R_{sh}} \quad (8)$$

Where, K_{sc} is Short circuit current temperature coefficient and E_g represent the Band gap energy of crystalline material in eV.

R_s =series resistance of the cell that represents the internal losses of solar cell and R_{sh} is Shunt resistance of the cell which is parallel with the diode that takes the leakage current to the ground.

G_s and G is solar radiation at standard test condition and operating Radiation condition respectively.

T_c is standard test condition temperature and T_o is operating cell temperature.

Efficiency in photovoltaic solar panels is measured by the ability of a panel to convert sunlight energy into DC energy output. Basic equation of Fill Factor (FF) and Efficiency (Ef) is given below ([1],[11]-[17], [23-24]).

$$FF = \frac{V_m I_m}{V_{oc} I_{oc}} \quad (9)$$

$$Ef = \frac{P_{max}}{P_{in}} \quad (10)$$

Where, maximum Power, $P_{max} = V_m I_m$ (11)

In this paper we used MATLAB software to investigate the maximum power point, fill factor and efficiency in different types of condition temperature and solar radiation. To perform this analysis we have use the data list of table 1. These data are based on 180W ZED fabric mono-crystalline PV solar panel [25].

TABLE I
PARAMETERS OF ZED FABRIC PHOTOVOLTAIC SOLAR

Parameters	Type/Value
Cell Technology	Si crystalline
Number of cells in series	72
Open circuit voltage (Total)	45 Volt
Short circuit current (Total)	5.25 Ampere
Voltage at maximum power	36.31 volt
Current at maximum power	4.98 ampere
Maximum system voltage	1000 volt
Maximum power	180 watt
Temperature coefficient	.0023 V/ ⁰ Celsius
Saturation current	1.6595 nA
Temperature range	-40 ⁰ C to 80 ⁰ C
Cell efficiency	15.2%
Module efficiency	15%
Standard test temperature	25 ⁰ C
Standard test radiation	1000 w/m ²

IV. COMPARISON OF EFFICIENCY BETWEEN DIFFERENT TYPES OF CRYSTALLINE MATERIAL

Efficiency of a solar cell depends on the spectrum and intensity of the incident sunlight and the temperature of the solar cell.

Table (II-VII) provides the relevant statistics based on maximum power, fill factor and efficiency of Mono-crystalline, Polycrystalline and Amorphous based solar cells respectively.

A. Variation of Efficiency of Different Types of Materials due to change of Solar Radiation

Table (II-IV) presents the variation of efficiency due to change of solar radiation.

1) Variation of Efficiency of Monocrystalline due to change of Solar Radiation

TABLE II. SIMULATION RESULT FOR MONOCRYSTALLINE DUE TO CHANGE OF SOLAR RADIATION

G	Pm	Vm	FF	Efficiency
200	35.5	36	0.150265	13.87772
400	73.84	38	0.31255	14.43282
600	113.7	38	0.48127	14.81593
800	154.3	39	0.653122	15.0798
1000	195.3	39	0.826667	15.2694

2) Variation of Efficiency of Polycrystalline due to change of Solar Radiation

TABLE III. SIMULATION RESULT FOR POLYCRYSTALLINE DUE TO CHANGE OF SOLAR RADIATION

G	Pm	V	FF	Efficiency
200	28.57	31	0.120931	11.16863
400	62.53	33	0.264677	12.22216
600	98.37	35	0.416381	12.81832
800	135.3	36	0.572698	13.22292
1000	173.1	37	0.732698	13.5337

3) Variation of Efficiency of Amorphous due to change of Solar Radiation

TABLE IV. SIMULATION RESULT FOR AMORPHOUS DUE TO CHANGE OF SOLAR RADIATION

G	Pm	V	FF	Efficiency
200	30.57	33	0.129397	11.95047
400	66	35	0.279365	12.90041
600	103	36	0.435979	13.42164
800	141	36	0.596825	13.77998
1000	179.9	37	0.761481	14.06536

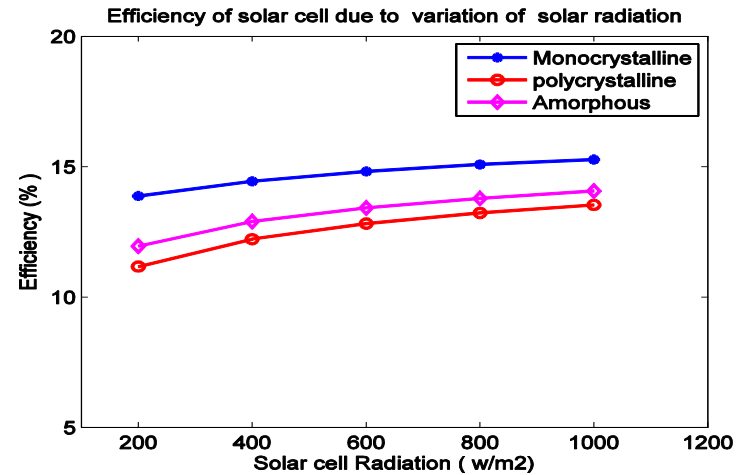


Figure 4. Efficiency of solar cell due to variation of Solar Radiation

Figure 4 demonstrate the variation of solar efficiency by increasing the solar radiation. It can be said that, Mono-crystalline materials show maximum efficiency in standard test condition of the solar model.

B. Variation of Efficiency of Different types of Materials due to change of Temperature

Performance of different solar materials in terms of efficiency due to the variation of temperature presents in table (V-VII).

1) Variation of Efficiency of Monocrystalline due to change of Temperature

TABLE V. SIMULATION RESULT FOR MONOCRYSTALLINE DUE TO CHANGE OF TEMPERATURE

T	Pm	V	FF	Efficiency
0	170.7	43	0.72254	13.34606
20	157.4	40	0.666243	12.30621
40	144.2	37	0.61037	11.27418
60	131.2	33	0.555344	10.25778
80	118.4	30	0.501164	9.257022

2) Variation of Efficiency of Polycrystalline due to change of Temperature

TABLE VI. SIMULATION RESULT FOR POLYCRYSTALLINE DUE TO CHANGE OF TEMPERATURE

T	Pm	V	FF	Ef
0	223.2	46	0.944762	17.45074
20	183.1	38	0.775026	14.31555
40	143.8	31	0.608677	11.2429
60	105.6	24	0.446984	8.256263
80	69.81	17	0.295492	5.458047

3) Variation of Efficiency of Amorphous due to change of Temperature

TABLE VII. SIMULATION RESULT FOR AMORPHOUS DUE TO CHANGE OF TEMPERATURE

T	Pm	V	FF	Efficiency
0	219	45	0.926984	17.12236
20	187.7	39	0.794497	14.6752
40	156.6	33	0.662857	12.24366
60	126.1	27	0.533757	9.859042
80	96.3	22	0.407619	7.529149

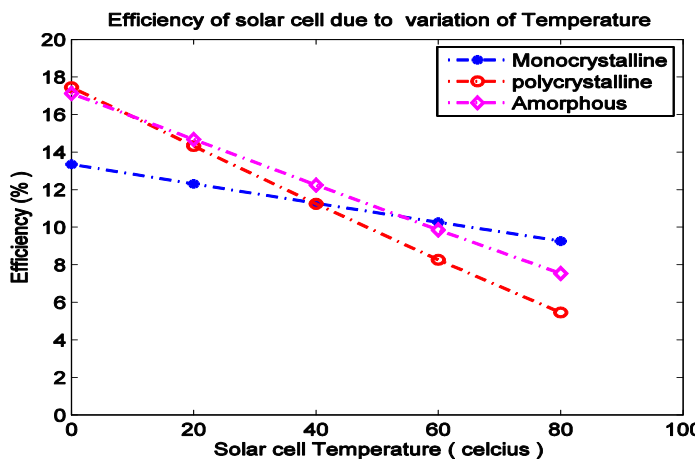


Figure 5. Efficiency of solar cell due to variation of Temperature

As can be seen from the tables (II-VII) and figures (III-IV), it can be clearly said that the maximum power (P_m) as well as efficiency are decreased when temperature of the amorphous cell is increased. However, not only the maximum power but also the efficiency is climbed up with the rise of radiation of the amorphous cell. Similarly, for Monocrystalline solar cell, the maximum power (P_m) is decreased with the increasing of temperature and simultaneously, the efficiency is reduced too. On the other hand, both the maximum power (P_m) and efficiency are hiked with the rise of the radiation. Furthermore, for Polycrystalline solar cell, the increasing growth of the maximum power (P_m) and efficiency can be observed with the gradual improvement of radiation. But with the gradual boom of the temperature, the maximum power (P_m) and efficiency are both declined.

V. P-V CHARACTERISCS OF MATERIALS DUE TO VARIATION OF SOLAR RADIATION

A. P-V characteristics for monocrystalline

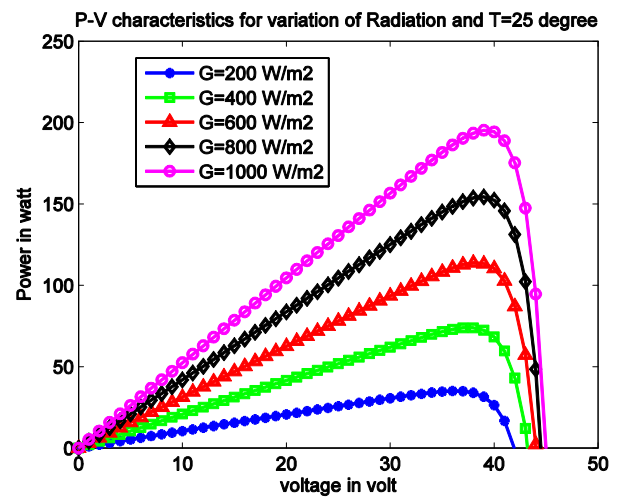


Figure 6. P-V Characteristics for mono-crystalline material due to variation of solar radiation

B. P-V characteristics for Polycrystalline

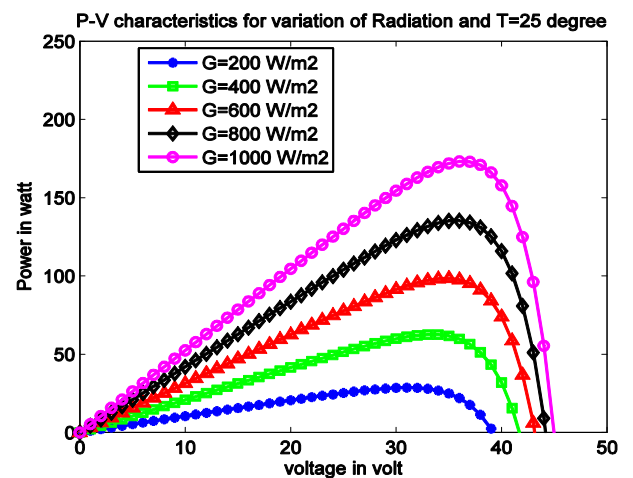


Figure 7. P-V Characteristics for Poly-crystalline material due to variation of solar radiation

C. P-V characteristics for Amorphous

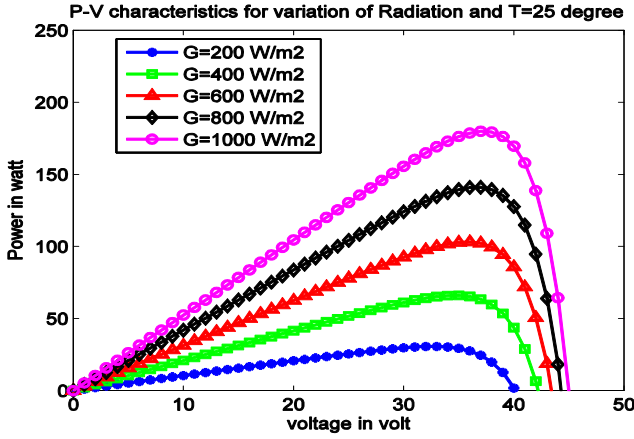


Figure 8. P-V Characteristics for amorphous material due to variation of solar radiation

VI. P-V CHARACTERISTICS OF MATERIALS DUE TO VARIATION OF TEMPERATURE

A. P-V characteristics for monocrystalline

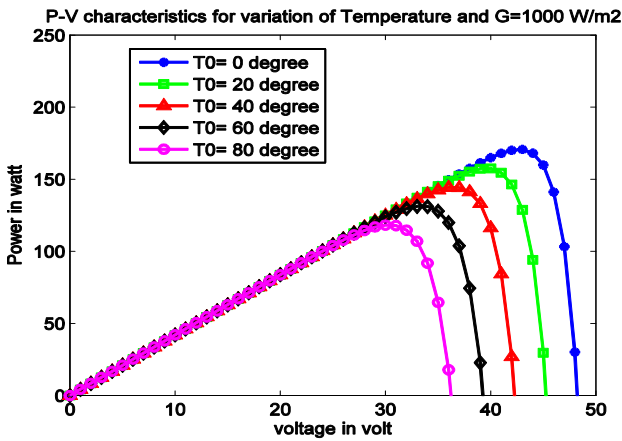


Figure 9. P-V Characteristics for mono-crystalline material due to variation of Temperature

B. P-V characteristics for Polycrystalline

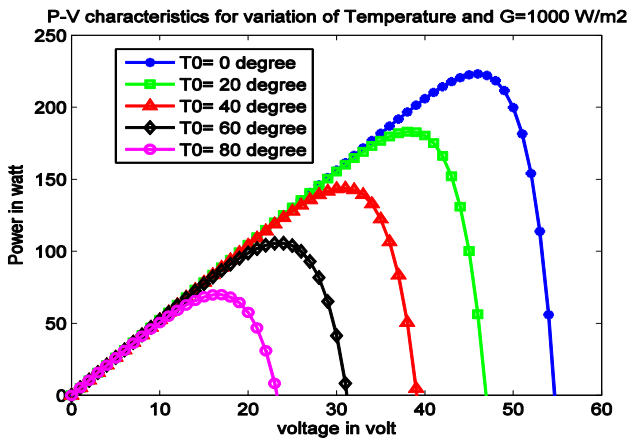


Figure 10. P-V Characteristics for poly-crystalline material due to variation of Temperature

C. P-V characteristics for Amorphous

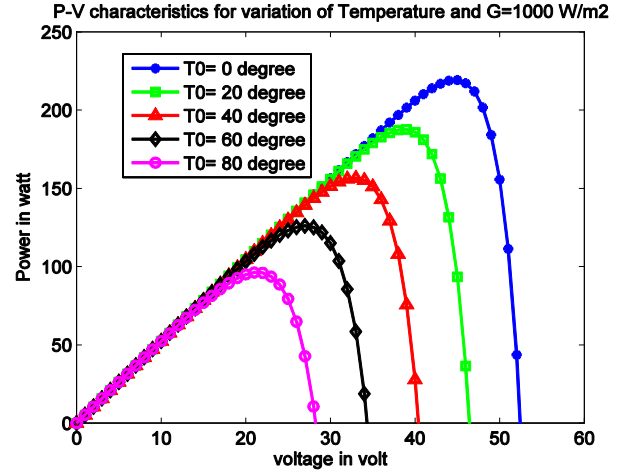


Figure 11. P-V Characteristics for amorphous material due to variation of Temperature

VII. CONCLUSIONS

The next century will witness a productive insight in terms of renewable energy sources, particularly solar energy and its derivatives. This renewable and enormous energy source offers a solution to the energy problem all over the world. Smart Grid technology can be a proper solution for overcome the power crisis problem. Renewable energy like solar energy is an added advantage to the Smart Grid ([1],[5-6],[12-15],[26-27]).

This paper investigated a specific PV model to observe the performance the Photovoltaic solar cell under different varying parameters such as solar radiation, ambient temperature. In this paper, A MATLAB model for the solar PV cell, modules and array was presented and analyzed through simulations using MATLAB. This model used the fundamental circuit equations of a solar PV cell taking into account the effects of electrical and environmental parameters. In our paper, we refer the temperature and solar radiation as environmental parameters. Ideality factor, series and shunt resistance, reverse saturation current considered as electrical parameters. In this paper, we emphasized on the evaluation of the effects ambient temperature and solar radiation on the efficiency of the solar panels. We have considered the mono-crystalline, Poly-crystalline and amorphous material for our evaluation.

From the analysis throughout the paper, we can say that performance of solar cell for any types of material depend on the standard conditions of solar cell. But Monocrystalline solar cell shows better performance than other type of material. Now a days, Commercially available PV materials are polycrystalline silicon, mono-crystalline silicon, amorphous silicon, and thin film technologies, such as cadmium telluride (CdTe), and copper indium diselenide (CIS). As from discussion, Si solar cells demonstrate high conversion efficiency, but in order to produce cheap and efficient PV devices, research on a large number of different thin film technologies is in progress.

REFERENCES

- [1] Nayan, Md Faysal, and S. M. Safayet Ullah. "Modelling of solar cell characteristics considering the effect of electrical and environmental parameters." *Green Energy and Technology (ICGET), 2015 3rd International Conference on*. IEEE, 2015.
- [2] S. Leva, D. Zaninelli, "Technical and Financial Analysis for Hybrid Photovoltaic Power Generation Systems", *WSEAS Transactions on Power Systems*, vol.5, No.1, May 2006, pp.831-838.
- [3] S. Leva, D. Zaninelli, R. Contino, "Integrated renewable sources for supplying remote power systems", *WSEAS Transactions Systems*, vol.2, no.2, February 2007, pp.41-48.
- [4] Wang Q., Qiu H. N., "Situation and outlook of solar energy utilization in Tibet, China", *Renewable and Sustainable Energy Reviews*, Elsevier, 2009, pp. 2181–2186.
- [5] Saban Yilmaz, Hasan Riza Ozcalik, Selami Kesler, Furkan Dincer, Bekir Yelmen, "The analysis of different PV power systems for the determination of optimal PV panels and system installation—A case study in Kahramanmaraş, Turkey" *Renewable and Sustainable Energy Reviews* (2015) in Elsevier.
- [6] S. Sheik Mohammad, "Modeling and Simulation of Photovoltaic module using MATLAB/Simulink", *International Journal of Chemical and Environmental Engineering*, Volume 2, No.5, October 2011.
- [7] European Photovoltaic Industry Association. Global Market Outlook for Photovoltaic's Until 2016 May 2012. P. 11.
- [8] S. Yilmaz, H. R. Ozcalik, S. Kesler, F. Dincer, B. Yelmen. "The analysis of different PV power systems for the determination of optimal PV panels and system installation—A case study in Kahramanmaraş, Turkey", *Renewable and Sustainable Energy Reviews*, Volume 52 Pages 1015-1024, December 2015
- [9] Online: <http://energyinformative.org/best-solar-panel-monocrystalline-polycrystalline-thin-film/>, last accessed August 2016
- [10] R. Bhol, A. Pradhan, "Environmental Effect Assessment On Performance of Solar Pv Panel", 2015 International Conference On Circuit, Power And Computing Technologies [ICCPCT].
- [11] J. S. Kumari and C. S. Babu, "Mathematical Modeling and Simulation of Photovoltaic Cell using Matlab-Simulink Environment", *International Journal of Electrical and Computer Engineering (IJECE)*, Vol. 2, No. 1, February 2012, pp. 26–34, ISSN: 2088-8708.
- [12] V. Khanna, B. K. Das, D. Bisht, "Matlab/Simelectronics Models Based Study Of Solar Cells", *International Journal Of Renewable Energy Research*, Vol.3, No.1, 2013.
- [13] M. A. Islam, N. Mohammad, P. K. S. Khan, "Modeling and Performance Analysis of a Generalized Photovoltaic Array in Matlab", *Proceedings of 29th annual IEEE Power Electronics Specialists Conference*, vol. 1, pp. 86-93, 1998.
- [14] S. Hegedus, and A. Luque, *Achievements and Challenges of Solar Electricity from Photovoltaics. Handbook of Photovoltaic Science and Engineering (Second edition)*. p. 4-5.
- [15] European photovoltaic industry association. Solar generation 6: solar photovoltaic electricity empowering the world, 2011. p. 10.
- [16] F.A. Salem, K.K. Matrawy and A-F. Mahrous, "Mathematical Modeling of PV Array With Different Performance Parameters", *International Journal of Control, Automation and Systems* Vol.4 No.2 April 2015
- [17] M. Bouzguenda, T. Salmi, A. Gastli and A. Masmoudi, "Evaluating Solar Photovoltaic System Performance using MATLAB", 2012 First International Conference on Renewable Energies and Vehicular Technology in IEEE
- [18] R. Krishan, Y. Raj Sood, B.U. Kumar, "The Simulation and Design for Analysis of Photovoltaic System Based on MATLAB" 2013 International Conference on Energy Efficient Technologies for Sustainability (ICEETS, IEEE 2013, 978-1-4673-6150-7)
- [19] Swapnil Dubey, Jatin Narotam Sarvaiya, Bharath Seshadri, "Temperature Dependent Photovoltaic (PV) Efficiency and Its Effect on PV Production in the World A Review", *PV Asia Pacific Conference 2012, Energy Procedia* 33 (2013) 311–321, Elsevier
- [20] P. Sivakumar, Abdullah Abdul Kader, Yogeshraj Kaliavaradhan, M. Arutchelv, "Analysis and enhancement of PV efficiency with incremental conductance MPPT technique under non-linear loading conditions", *Renewable Energy* 81 (2015) 543e550, Elsevier
- [21] T. Salmi, M. Bouzguenda, A. Gastli, A. Masmoudi, "MATLAB/Simulink Based Modelling of Solar Photovoltaic Cell", *International Journal Of Renewable Energy Research*, Vol.2, No.2, 2012
- [22] D. Bonkougou, Z. Koalaga, D. Njomo, "Modelling and Simulation of photovoltaic module considering single-diode equivalent circuit model in MATLAB", *International Journal of Emerging Technology and Advanced Engineering*, Volume 3, Issue 3, March 2013
- [23] Tanvir Ahmad, Sharmin Sobhan, Md. Faysal Nayan, "Comparative Analysis between Single Diode and Double Diode Model of PV Cell: Concentrate Different Parameters Effect on its Efficiency" *Journal of Power and Energy Engineering, ISSN Online: 2327-5901*, (Accepted in February, 2016).
- [24] R. Boylestad, L. Nashelsky, "Electronic Devices and Circuit Theory", Prentice Hall, Upper Saddle River, New Jersey Columbus, Ohio, 7th Edition.
- [25] (online) www.zedfactory.com/microgeneration.pdf, last accessed August 2015
- [26] F. Nayan, M. Islam and S. Mahmud, "Feasibility Study of Smart Grid in Bangladesh," *Energy and Power Engineering*, Vol. 5 No. 4B, 2013, pp. 1531-1535. doi: 10.4236/epe.2013.54B290
- [27] T. Hasan, M. F. Nayan, M. A. Iqbal, M. Islam, "Smart Solar Home System with Safety Device Low Voltage Alert", *UKSim 14th International Conference on Modelling and Simulation*, Pages 201-204, ISBN: 978-0-7695-4682-7

Enhancement of sensitivity and birefringence of a gas sensor on micro-core based photonic crystal fiber

Sayed Asaduzzaman, Bikash Kumar Paul and Kawsar Ahmed

¹Department of Information and Communication Technology (ICT),
Mawlana Bhashani Science and Technology University (MBSTU)

Santosh, Tangail-1902, Bangladesh.

Email: sayed.asaduzzaman.bd@ieee.org, bikash.k.paul@ieee.org, k.ahmed.bd@ieee.org

Abstract—This paper proposed a microstructure cored photonic crystal fiber for gas sensing applications. The key research vision is to identify hazardous and colorless gases over a larger range of wavelength band covering 1 μm to 2.2 μm . Numerical simulation on the proposed PCF is carried out by full vectorial finite element method (FEM) with perfectly matched layer. The proposed PCF core contains vertically arranged elliptical holes which enhances both relative sensitivity and birefringence. The analysis depicts that the proposed PCF shows high sensitivity of 43.7%, high birefringence of 8.62×10^{-3} and low confinement loss of 1.89×10^{-6} dB/m at the wavelength of 1.33 μm . The structural geometric parameters have tuned to optimize the parameters of proposed PCF.

Keywords—Birefringence; Confinement loss; Evanescent field; Gas sensor; Relative sensitivity

I. INTRODUCTION

A finite length of optical fiber is called PCF which consists of organized cylindrical air hole through the finite length of that fiber [1]. Photonic crystal fiber brings out a revolutionary change in modern hasty communication. It has also opened a new era in not only communication area but also in sensing applications [2]. PCF strongly holding it places as a sensor due to its structural geometric versatility with higher sensitivity.

The structural and manufacturing flexibility of PCF offers to model PCF based gas sensor compared with traditional optical fiber [3, 4, 5]. Enclosing much light into the core which is filling with gas samples considered as hollow core PBG-PCFs is proposed in [6]. O. Frazão et al [7] proposed a Suspended-core fibers sensing which working principle is based on total internal reflection. To reduce confinement loss and to exalt relative sensitivity Park et al. [8] proposed SiO₂ gas ring around the central air hole.

In the inner ring hexagonal holes are used instead of circular holes. As a result it tremendously increases the sensitivity. Analyzing the result of the paper confinement loss and relative sensitivity has investigated but it is not momentous for as novel gas sensor. In 2013 a new era has been opened by Zhang et al. [9] an evencent wave base PCF has been proposed. In 2015 M. Morshed et al. [10] proposed PCF that gives better relative sensitivity about 16.88% and confinement loss to 1.765×10^{-8} dB/m.

In this paper, a PCF is proposed and analyzed where cladding was hexagonal and core contains elliptical holes also

arranged in hexagonal manner. The relative sensitivity and birefringence increases by introducing elliptical holes. We have used elliptical holes in our proposed PCF which can be used as gas sensor whereas the elliptical holes were used previous PCF [11] which was used for chemical sensing.

II. GEOMETRICAL STRUCTURE OF THE PROPOSED PCF

Fig. 1 shows the cross section of the proposed PCF. The proposed PCF contains cladding with five layers of air holes which arranged in a hexagonal manner. The cladding holes diameter denoted by d . The cladding holes are kept same to avoid fabrication complexities. The outer ring holes located in the edges are missing as [12]. The core region is formulated with 12 elliptical air holes. The elliptical air holes are formed into a hexagonal manner where the central elliptical hole is missing. The length of the major and minor axis of ellipse are defined as d_a and d_b . The distance between centers of two adjacent holes is called the pitch. The pitch of cladding is denoted as Λ and the pitch of the core as Λ_1 . Elliptical holes are used to gain high birefringence as well as high sensitivity. The perfectly matched layer (PML) of the proposed PCF is set around 10% of the whole diameter of the PCF for absorbing the lightwave. As background material pure silica is used, its refractive index changes with respect to wavelength following Sellmier Equation [13].

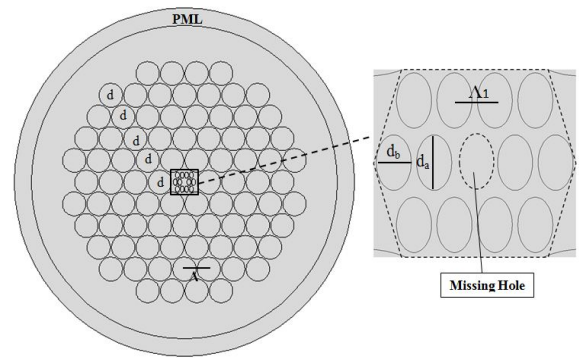


Fig. 1 Cross sectional view of the Proposed PCF.

III. NUMERICAL METHOD

The Finite Element Method (FEM) has been utilized for numerical simulation of the proposed gas sensor PCF. Two key parameters like Confinement loss and Sensitivity, are related to measuring the efficiency of a gas sensor PCF have been investigated using the Finite Element Method (FEM). The

guided light through the PCF pervade into the cladding region for a number of reasons; this phenomenon is known as confinement loss. The confinement loss L_c can be obtained through the imaginary part of refractive index n_{eff}

$$L_c = \frac{40\pi \text{Im}[n_{eff}] \times 10^6}{\lambda \ln(10)} \quad (\text{dB/m}) \quad (1)$$

Where, $\text{Im}[n_{eff}]$ is the imaginary part of the refractive index and λ is the wavelength of light. The relative sensitivity is computed from the interaction between light and the gas samples and it can be calculated through the equation (2)

$$r = \frac{n_r}{n_{eff}} f \quad (2)$$

Where, n_r represents the refractive index gas to be sensed and n_{eff} is the modal effective index. The f is the percentage ratio of the air hole power and the total power.

$$f = \frac{\int_{\text{sample}} \text{Re}(E_x H_y - E_y H_x) dx dy}{\int_{\text{total}} \text{Re}(E_x H_y - E_y H_x) dx dy} \times 100 \quad (3)$$

Where, E_x and H_x are transverse electric field and magnetic field, E_y and H_y are longitudinal electric field and magnetic field. Using finite-element method (FEM) the mode field pattern E_x , H_x , E_y , H_y and effective index n_{eff} has been obtained. Birefringence is the difference of the refractive indices between x-polarization and y-polarization of the effective mode index which can be defined as,

$$B = |n_x - n_y| \quad (4)$$

IV. NUMERICAL RESULTS AND DISCUSSION

The modal intensity distribution of the proposed PCF at the 1.33 μm operating wavelength has been shown in Fig. 2. From the Fig. 2 it is clear that the evanescent field strongly interacts with the samples located at the core region. The x-polarized and y-polarized mode is tightly bounded at the core due to high index contrast.

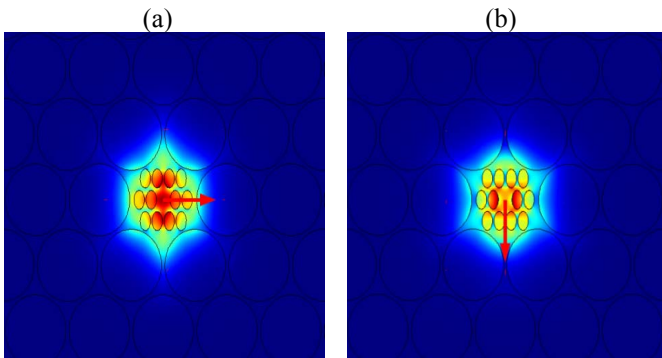


Fig. 2 Modal Intensity distribution of the proposed PCF along (a) X-polarization and (b) Y-polarization at the wavelength $\lambda = 1.33 \mu\text{m}$.

The variation of the relative sensitivity of x-polarization and y-polarization with wavelength has been shown in the Fig. 3. The figure shows the relative sensitivity increases simultaneously with the wavelength. The figure also shows

that the high sensitivity can be found at x-polarization of the proposed PCF.

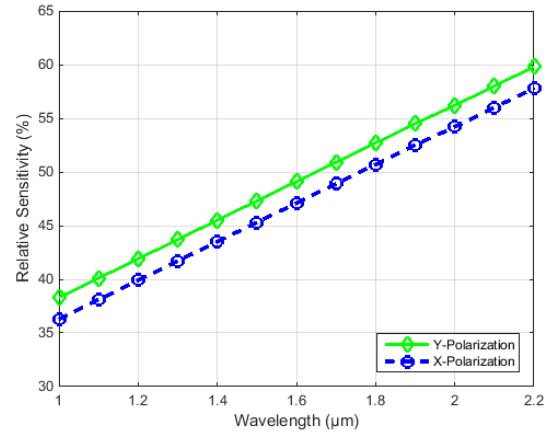


Fig. 3 Variation of Relative sensitivity as a function of wavelength for the optimized parameters of proposed PCF.

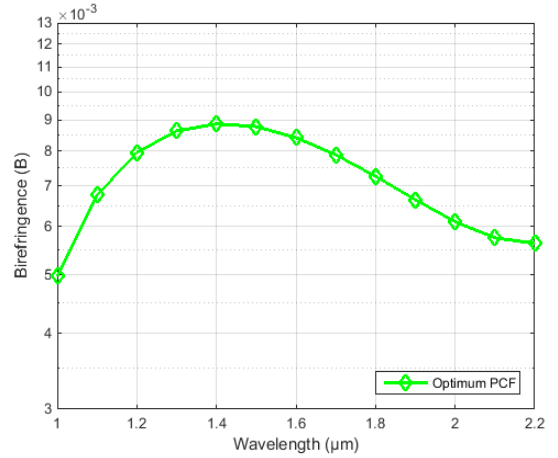


Fig. 4. Variation of birefringence as a function of wavelength for the optimized parameters of proposed PCF.

Fig. 4 shows the variation of birefringence with the wavelength for the optimized parameters of proposed PCF. Fig. 4 also illustrates that the proposed PCF shows high birefringence which varies with the wavelength. The highest birefringence can be found at the wavelength of 1.4 μm . From wavelength 1 μm to 1.4 μm birefringence increases at a steady rate and then it decreases sharply from wavelength 1.5 μm to 2.2 μm .

Fig. 5 depicts the variation of relative sensitivity as a function of wavelength for different diameters of the holes of cladding. The figure shows that the high relative sensitivity can be achieved for $d = 1.52 \mu\text{m}$. The effect of change of pitch of the cladding is shown in Fig. 6. The figure depicts that there is a very small effect of relative sensitivity for variation of the pitch values whereas there is a tremendous effect on confinement loss for the various pitch values. If the pitch increases that means the distance between holes of the cladding increases results a higher confinement loss.

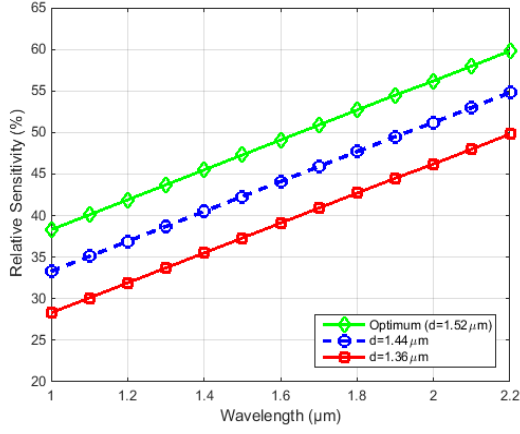


Fig. 5. Variation of relative sensitivity as a function of wavelength for $d = 1.52 \mu\text{m}$, $d = 1.44 \mu\text{m}$ and $d = 1.36 \mu\text{m}$ when the other optimized parameters were kept constant.

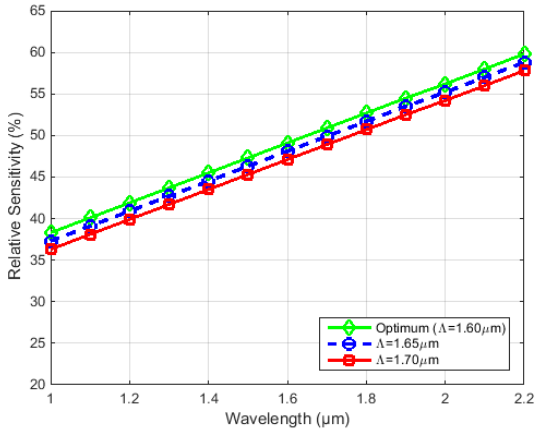


Fig. 6. Variation of relative sensitivity as a function of wavelength for $\Lambda = 1.6 \mu\text{m}$, $\Lambda = 1.65 \mu\text{m}$ and $\Lambda = 1.70 \mu\text{m}$ when the other optimized parameters were kept constant.

Fig. 7 shows the effect of various ratio of major and minor axis of the diameters of the elliptical holes of the core. The figure clearly describes that for higher ratio the higher sensitivity can be gained. The proposed PCF shows higher sensitivity when $da/db = 0.65$. During the fabrication process variation of the global parameters may be occurred. But $\pm 1\%$ - $\pm 2\%$ is tolerable. We have tuned the global parameters $\pm 1\%$ to $\pm 2\%$ from the optimized parameters. Fig. 8 shows that the relative sensitivity does not affect more from its optimized structure.

Table 1. shows the comparison among the proposed PCF and the prior PCFs. From the table it is cleared that the proposed PCF shows higher relative sensitivity as a gas sensor. The structural differences are also shown in the table. As we showed all the results for the wavelength $1.33 \mu\text{m}$ it is the absorption line of the two hazardous gases methane (CH_4) and hydrogen fluoride (HF).

Fabrication is a challenging issue for any kind of PCF. The proposed PCF is a dual shape mixing PCF where the core

holes are elliptical and cladding holes are circular. Two different fabrication processes can be used to fabricate the proposed PCF. Dual shape mixing PCFs can be fabricated by [14] and elliptical hole based PCF can be fabricated by [15]. Recently sol-gel technique [16] can be used to fabricate the PCFs with diverse hole size and different shapes.

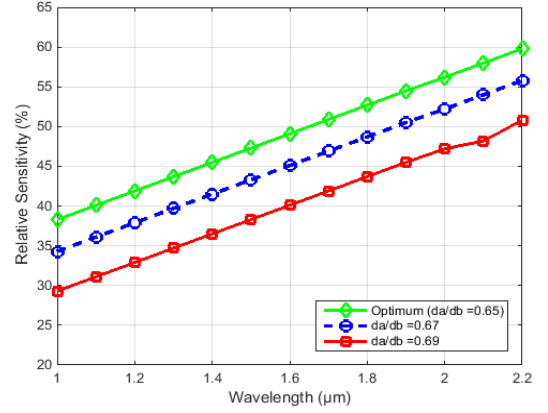


Fig. 7. Variation of relative sensitivity as a function of wavelength for $da/db = 0.65$, $da/db = 0.67$ and $da/db = 0.69$ when the other optimized parameters were kept constant.

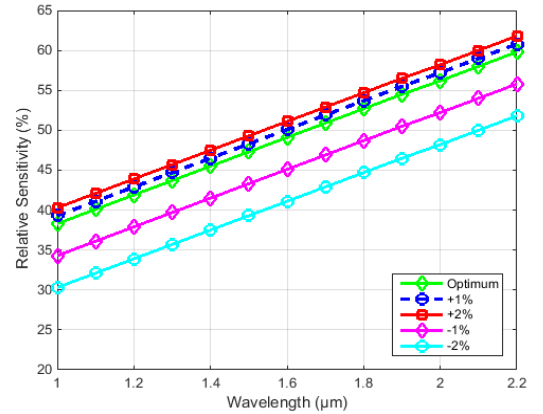


Fig. 8. Variation of relative sensitivity as a function of wavelength among the optimized PCF and global variation of the parameters.

TABLE I. COMAPRISION OF SIMUTATED RESULT AND STRUCTURE SHAPE AMONG PROPOSED PCF AND PRIOR PCFS AT $\lambda = 1.33 \mu\text{m}$.

PCFs	Sensitivi ty (r%)	Con. Loss (dB/m)	Birefringe nce (B)	Structural Shape Core Cladding
Prior PCF ₁ [6]	13.3%	3.77×10^{-6}	-	Defected core
Prior PCF ₂ [10]	16.88%	1.765×10^{-8}	-	Defected core with non-circular holes
Proposed PCF	43.7%	1.89×10^{-6}	8.62×10^{-3}	Elliptical holes in hexagonal manner

V. CONCLUSION

In conclusion, a PCF is proposed which has been numerically investigated by FEM shows high relative sensitivity, birefringence and low confinement loss

simultaneously. The proposed PCF can be used as a gas sensor to detect toxic and colorless gases whose absorption line lies between the wavelengths 1 μm to 2.2 μm . In core hexagonally arranged elliptical holes increases both sensitivity and birefringence. The arrangement of the cladding holes interacts lights along the core and decreases the leakage loss or confinement loss. The whole simulation process is done by varying the global parameters of both core and cladding. The proposed PCF is simple in design than the previous PCFs and also shows better results.

ACKNOWLEDGMENT

The authors are grateful to those who participated in this research work.

REFERENCES

- [1] P. St J Russell, "Photonic-Crystal fibers," *Journal of Lightwave Technology*, vol. 24, no. 12, pp. 4729–4749, Dec. 2006.
- [2] H. Ademgil, "Highly sensitive octagonal photonic crystal fiber based sensor," *Optik - International Journal for Light and Electron Optics*, vol. 125, no. 20, pp. 6274–6278, 2014.
- [3] M. N. Petrovich *et al.*, "Photonic crystals and Photonic crystal fibers for sensing applications," pp. 15–60050, Nov. 2005.
- [4] T. Ritari, J. Tuominen, H. Ludvigsen, J. Petersen, T. Sørensen, T. Hansen, and H. Simonsen, "Gas sensing using air-guiding photonic bandgap fibers," *Optics Express*, vol. 12, no. 17, pp. 4080–4087, Aug. 2004.
- [5] J. M. Fini, "Microstructure fibres for optical sensing in gases and liquids," *Measurement Science and Technology*, vol. 15, no. 6, pp. 1120, 2004.
- [6] S. Olyaei and A. Naraghi, "Design and optimization of index-guiding photonic crystal fiber gas sensor," *Photonic Sensors*, vol. 3, no. 2, pp. 131–136, Jan. 2013.
- [7] O. Frazão, R. M. Silva, M. S. Ferreira, J. L. Santos, and A. B. Lobo Ribeiro, "Suspended-core fibers for sensing applications," *Photonic Sensors*, vol. 2, no. 2, pp. 118–126, Apr. 2012.
- [8] J. Park, S. Lee, S. Kim, and K. Oh, "Enhancement of chemical sensing capability in a photonic crystal fiber with a hollow high index ring defect at the center," *Optics express*, vol. 19, no. 3, pp. 1921–1929, 2011.
- [9] L. Zhang, G. Ren, and J. Yao, "A new photonic crystal fiber gas sensor based on evanescent wave in terahertz wave band: Design and simulation," *Optoelectronics Letters*, vol. 9, no. 6, pp. 438–440, Oct. 2013.
- [10] M. Morshed, M. I. Hasan, and S. A. Razzak, "Enhancement of the sensitivity of gas sensor based on microstructure optical fiber," *Photonic Sensors* 5, pp.312–320 ,2015.
- [11] H. Ademgil and S. Haxha, "PCF based sensor with high sensitivity, high birefringence and low confinement losses for liquid Analyte sensing applications," *Sensors*, vol. 15, no. 12, pp. 31833–31842, Dec. 2015.
- [12] M. Morshed *et al.*, "Microstructure core photonic crystal fiber for gas sensing applications," *Applied Optics*, vol. 54, no. 29, pp. 8637–8643, Oct. 2015.
- [13] G. Ghosh, "Sellmeier coefficients and dispersion of thermo-optic coefficients for some optical glasses," *Applied Optics*, vol. 36, no. 7, pp. 1540–1546, Mar. 1997.
- [14] A. Argyros *et al.*, "Ring structures in microstructured polymer optical fibres," *Optics Express*, vol. 9, no. 13, pp. 813–820, Dec. 2001.
- [15] N. A. Issa, M. A. van Eijkelenborg, M. Fellew, F. Cox, G. Henry, and M. C. J. Large, "Fabrication and study of microstructured optical fibers with elliptical holes," *Optics Letters*, vol. 29, no. 12, pp. 1336–1338, Jun. 2004.
- [16] R. T. Bise and D. J. Trevor, "Sol-gel derived microstructured fiber: Fabrication and characterization," *OFC/NFOEC Technical Digest. Optical Fiber Communication Conference, 2005*.

Impact of User Mobility on the Performance of Downlink Resource Scheduling in Heterogeneous LTE Cellular Networks

¹Abdullah Bin Shams, ²Syed Rafiee Abied, ³Md. Ashraful Hoque

Department of Electrical and Electronic Engineering,

Islamic University of Technology, Gazipur 1704, Bangladesh.

Email: ¹abdullahbinshams@gmail.com, ²rafiee.abied@gmail.com, ³mahoque@iut-dhaka.edu

Abstract— In LTE, resource scheduling algorithm is one of the key parameters that determine the overall system and UE performances in terms of throughput and fairness respectively. Multiple scheduling techniques are available that emphasizes on optimizing any one of the key system performance parameters. Among them the most potential algorithms are proportional fair (PF) and round robin (RR). PF optimizes UE throughput with a fairly high fairness index whereas RR provides the best fairness among the UEs but with a comparatively lower throughput. Moreover the scheduler performance is affected by the UE mobility and UE density. This can significantly degrade the overall system efficiency. In this paper we have analyzed and compared several LTE downlink performance parameters between PF and RR in heterogeneous network (HetNet) to determine which scheduling scheme is convenient for users with high mobility.

Keywords—UE mobility; UE density; Proportional Fair; Round Robin; Heterogeneous Network

I. INTRODUCTION

The main goals of LTE are to provide high speed data, better spectral efficiency and lower latency. As it is evolved based on UMTS, LTE uses orthogonal frequency division multiple access (OFDMA) for downlink transmissions which reduces inter symbol interference (ISI) [1]. To improve throughput LTE supports multiple MIMO techniques namely transmit diversity (TxD), closed loop spatial multiplexing (CSLM), open loop spatial multiplexing (OLSM) etc. For a 20MHz channel a maximum data rate of 326.4Mbps and 86.4Mbps can be achieved in the downlink and uplink respectively [2]. This results in better usage of the overall spectrum increasing the spectral efficiency. To improve the performance of the UEs LTE introduced many features in its recent releases. One of many potential approaches is femto cells [3]. These are low power short ranged base stations (BSs) used in conjunction with the existing macro cell BSs forming a HetNet [4]. The deployments of femto cells are random and are used in coverage depleted areas. This new feature improves network availability to the UEs, shares the traffic load and, with the additional BSs inside the existing macro cell coverage area, improves UE experience in terms of data rate and better spectral efficiency. In LTE, the available bandwidth is divided into multiple carriers or resource blocks

(RBs). Each RB has a frequency band of 180 kHz and contains 12 subcarriers (SB). Data is transmitted from the BS to UEs through RBs. The code rate of a UE is determined from its channel quality indicator (CQI) at the BS [5]. After determining the data rate the UEs they are queued by the MAC layer of the BS in order to be served. The types of queuing processes have a significant impact on both the UE and system performances. Several schemes are designed to optimize any one of the performance metrics e.g. high throughput, fairness, low drop rate etc. Many schedulers take the UE CQI into account and a change in UE CQI will change its priority in the scheduling process. Therefore mobility of the UE has a significant impact on the outcome of the resource scheduler and on the overall system performance.

A significant amount of research work has focused on the performance of various scheduling schemes and their results indicate PF to dominate over other schedulers in terms of achieving a balance between high UE throughput and better fairness [6-9]. But most of them did not consider the mobility of the UEs. Moreover PF algorithm is a technique which was developed assuming stationary UEs. In practice UEs within a cell can move at high velocities which would affect the overall performance of a resource allocation scheme. Therefore UE mobility must be taken into account for a fair and realistic comparison between different schedulers.

In this paper, we have examined and compared several LTE downlink performance parameters in HetNet for a wide range of UE velocities under PF and RR scheduling schemes to assess which technique is suitable for high mobility users within a cell and at cell edges. At cell edges, where the received SINR by the UE is generally low, a seamless network service is required by the UEs even at high velocities to avoid call drops.

The remainder of the paper is organized as follows. In section II the HetNet layout, PF and RR scheduling schemes with various key performance metrics are presented. Simulation setup and analysis of the performance metrics between the two scheduling techniques are discussed in section III. Results and discussions are described in section IV and finally the conclusions are drawn in section V.

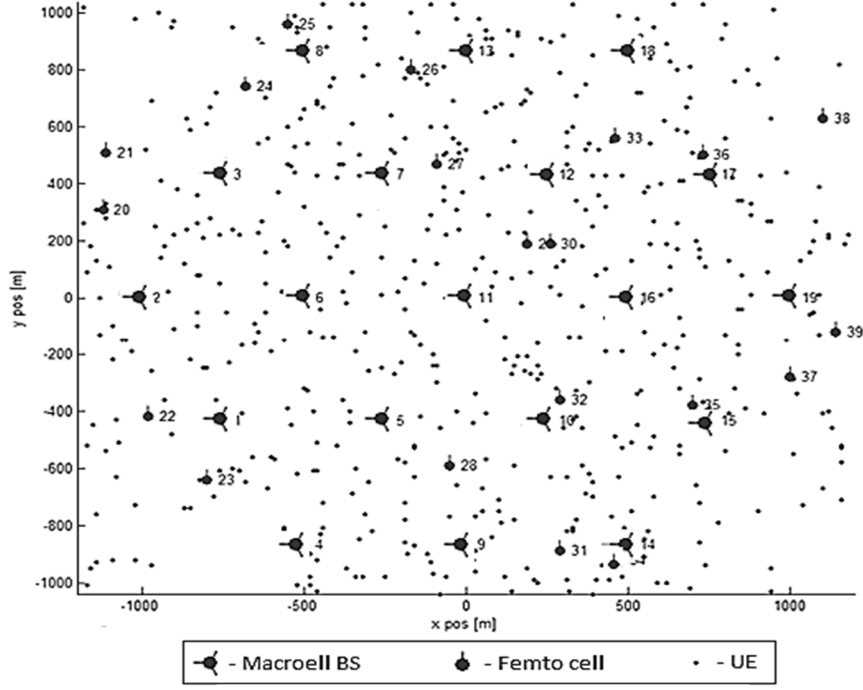


Fig. 1: Heterogeneous network (HetNet) deployment.

II. SYSTEM MODEL

A. Network model

In our simulation, to compare the performance of PF and RR, HetNet layout was considered. For forming the heterogeneous network, macro cells with femto cells were deployed where the macro cells consisting of 19 BSs are placed hexagonally and having the inter BS distance of 500 m. Femto cells with single antenna are placed in coverage depleted area and 20 femto cells are placed surrounding the UEs. Figure 1 illustrates three-sector antenna BSs as well as UEs.

B. Resource Scheduling

Proportional fair (PF) scheduler is used for allocating resources among the UEs which helps to attain a high UE data rate and good fairness index. Depending upon UE priority, this scheme helps to gain a maximum total data rate and also provides a preferable fairness index to all the UEs. In the PF scheme, each UE is considered as the priority coefficient of the priority function:

$$P = \frac{T^\alpha}{R^\beta} \quad (1)$$

Here for a time interval, feasible throughput is considered as T and average data rate as R where α and β are the parameters to adjust the fairness. For PF technique, $\alpha \approx 1$, $\beta \approx 1$ should be considered [10]. Round robin algorithm is one of the simplest scheduling techniques as it does not account for the channel condition of a UE. It treats all the UEs under a BS with equal

priorities and allocates system resources to all the UEs in a cyclic manner. This results in achieving the best fairness among all the UEs at a cost of lower UE throughput. The fairness parameters of the RR algorithm are $\alpha = 0$ and $\beta = 1$.

C. Key Performance Indicators

Average UE throughput: UEs in a heterogeneous network are positioned arbitrarily relative to the BSs. Hence, the distances from UE to BS are not equal for all the UEs. Depending upon the distance, UEs located close to the BS have high SINR values while UEs that located far from BS have a reduced SINR. As a result, UEs under a BS have a wide range of SINR and throughput values. Therefore, average UE throughput (T_{avg}) can be considered as a key performance parameter for understanding the performance of the network, and defined as:

$$T_{avg} = \frac{\sum_{k=1}^n T_k}{n} \quad (2)$$

where T_k is defined as the total throughput for k^{th} user and n is the total number of users.

Cell edge throughput: UEs located in the cell edge having a inter cell interference from the neighboring cell and signal strength is weak due to the distance from the BS. As a result, UEs located in the edge of the cell have a very poor data rate and this degrades the performance of the network. For attaining continuous coverage and to avoid call drop on the duration of handover from cell to cell it is necessary to keep a minimum data rate. It is considered as the 5th percentile of the UE throughput ECDF.

Spectral efficiency: Spectral efficiency is a key performance parameter that depends upon both on bandwidth and data rate. For the increase in bandwidth, data rate will also be increased. Spectral efficiency is defined as the data transmission rate over a specific bandwidth which followed by this equation:

$$S = \frac{\sum_{k=1}^n T_k}{BW} \quad (3)$$

where T_k is denoted as throughput for k^{th} user and BW is the system bandwidth.

Fairness index: Fairness index is used to determine how the system resources are shared among the UEs. The most popular Jain's fairness index is used [11] to assess the fairness among the UEs. Fairness index will be 1 if all the UEs have same share of the resources. For n UEs, Jain's fairness index is computed from the following equation:

$$J(T) = \frac{[\sum_{k=1}^n T_k]^2}{n[\sum_{k=1}^n T_k^2]} \quad (4)$$

where T_k is the average throughput of k^{th} user.

III. SIMULATION

A LTE system level simulator [12] was used to evaluate the performance of a HetNet for a 2x2 MIMO system. Transmission diversity mode (TxD) was selected because it performs better in various fading conditions. Mutual information based effective SINR mapping (MIESM) technique was adopted to predict the channel quality indicator (CQI) of the UEs. An urban environment was considered and the macroscopic propagation model used for both macro cell BS and femto cell is given by the following equation:

$$L = 40(1 - 40 * 10^{-3} h_{BS}) \log_{10}(R) - 18 \log_{10}(h_{BS}) + 21 \log_{10}(f) + 80dB \quad (5)$$

where R is the BS-UE separation in kilometers, f is the carrier frequency in MHz and h_{BS} is the height of BS in meters. For UE mobility we have considered the random walk model. Additional simulation parameters are presented in Table-I. To compare the performance of PF and RR scheduling schemes under user mobility, simulation was realized for 570 UEs placed randomly within the HetNet and for a UE velocity range of 0-125kmph. Then to analyze the effect of UE density on the scheduler performance number of UEs was varied from 10-100 per cell and to observe a cumulative effect of UE velocity and UE density three velocity were considered namely 3kmph for pedestrians, 60kmph for medium UE velocities and 120kmph for high UE velocities. Finally a simulation was carried out for both schedulers under all the

TABLE I – SIMULATION PARAMETERS FOR HETEROGENEOUS NETWORK

Simulation parameters	
Channel model	winner+
Frequency	900MHz
Bandwidth	10MHz
No. of transmitter	2
No. of receiver	2
Transmission mode	Transmit diversity (TxD)
BS height	20m
BS power	45dB
Receiver height	1.5m
Adaptive RI	2
Antenna azimuth offset	30°
Antenna Gain	15dBi
BS Transmitter power	45dBm
Femto cell transmission power	10 (Watts)
Simulation time	50 TTI

system bandwidths (1.4MHz, 3MHz, 5MHz, 10MHz, 15MHz and 20MHz) supported by LTE for three velocity cases mentioned above to determine how effectively the schedulers utilizes the available spectrum to transmit data under user mobility.

IV. RESULTS AND DISCUSSION

The average UE throughput decreases with an increase in UE velocity under both scheduler techniques, as depicted in Fig. 2(a). For low velocity PF performs better but as the velocity increases RR dominates PF and approaches a minimum UE throughput whereas the latter experiences constant performance degradation. As RR does not consider the channel conditions of the UEs therefore for rapid changes in UE link quality at high velocities it outperforms PF.

The performance of both the schedulers in terms of cell edge throughput, as shown in Fig. 2(b), also undergoes a declining trend with increase in UE velocity. In contrast to UE throughput, at low UE velocities or at stationary RR provides better data rate than PF at cell edges. This is because PF takes UE channel conditions into account and therefore as cell edge UEs are located distant away from the BS their CQI feedback will be low due to high path loss. A low CQI means PF allocates least resources to UEs at cell edges. As the UE velocity increases cell edge throughput under PF technique falls rapidly and reaches zero. On the other hand RR faces a slow decline but still is able to support a low data rate at high velocities.

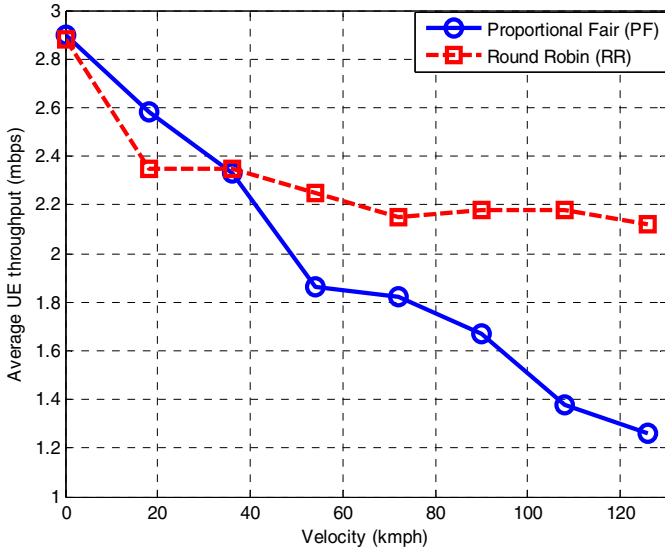


Fig. 2(a): Average UE throughput vs. UE velocity under PF and RR scheduler algorithm.

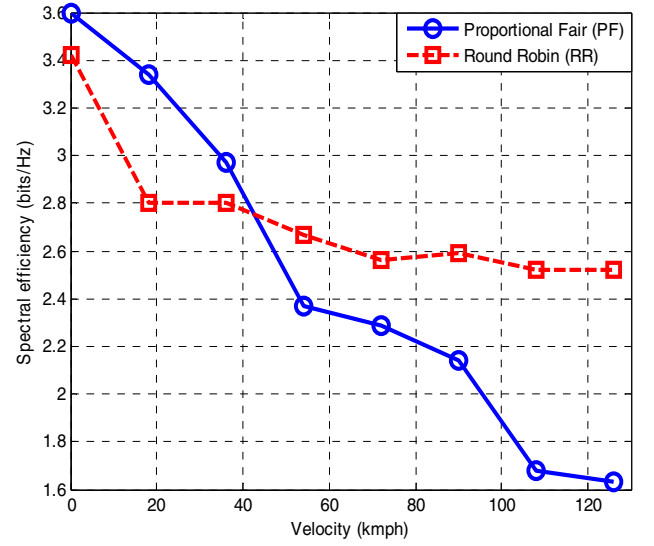


Fig. 2(c): Spectral efficiency vs. UE velocity under PF and RR scheduler algorithm.

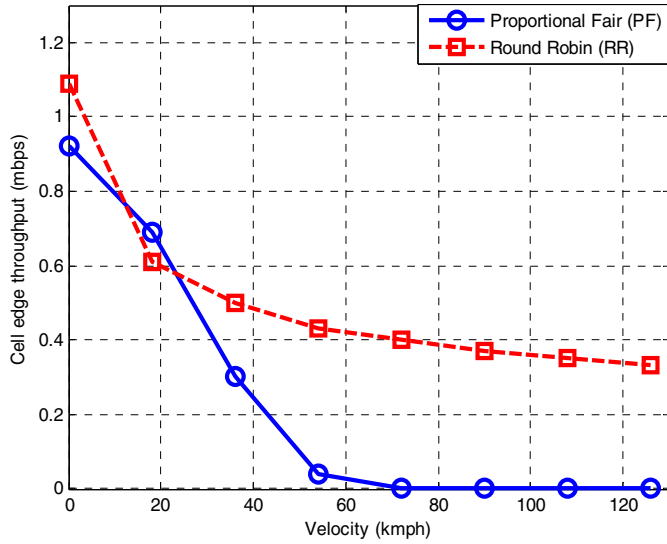


Figure 2(b): Average cell edge throughput vs. UE velocity under PF and RR scheduler algorithm.

Fig. 2(c) illustrates that as the velocity of the UE increases then for a given system bandwidth the spectrum is less efficiently utilized. As mentioned above that for low UE velocity PF delivers higher average UE throughput thereby achieving a better spectral efficiency than RR. But at high velocities RR provides better throughput therefore achieves a higher spectral efficiency.

The results depicted in Fig. 2(d) shows that although fairness deteriorates as the velocity of the UE increases the cyclic nature of UE serve in RR always delivers a higher fairness than PF.

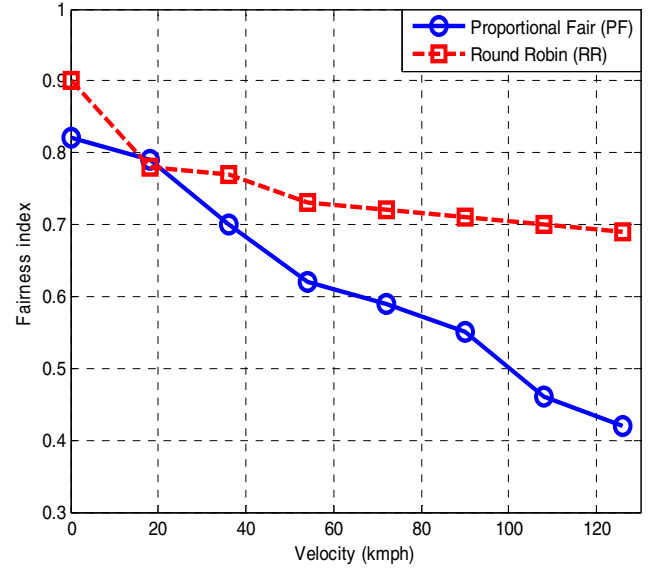


Figure 2(d): Fairness index vs. UE velocity under PF and RR scheduler algorithm.

Fig. 3(a) illustrates that with increase in UE density the average UE throughput provided by either PF or RR decreases rapidly and becomes constant at high velocities. The results also show that for high UE densities velocity has a minor impact on the performance of either scheduler in delivering resources to the UEs. The average cell edge throughput decreases with increase in UE densities as shown in Fig. 3(b) but the effect is more significant on PF for low UE velocity. For medium and high velocity cases PF fails to deliver any data rate to cell edge UEs. On the other hand RR provides similar cell edge performances for high UE densities.

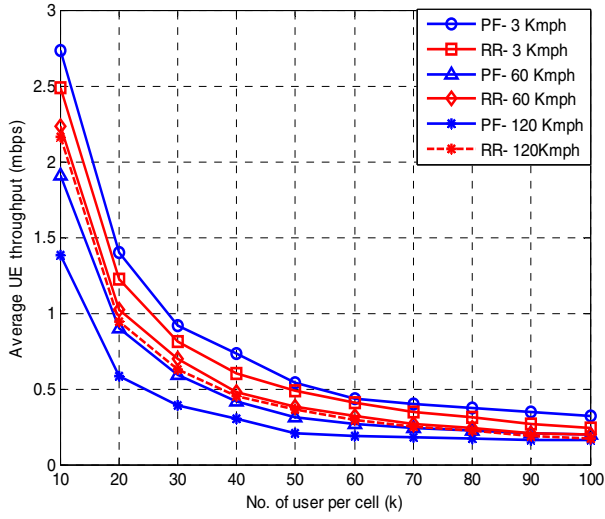


Fig. 3(a): Average UE throughput vs. UE density under PF and RR scheduler algorithm for low, medium and high velocity UE cases.

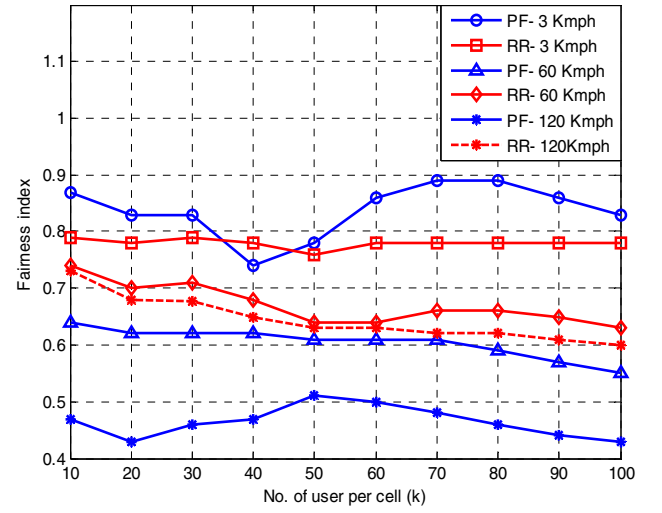


Fig. 3(d): Fairness Index vs. UE density under PF and RR scheduler algorithm for low, medium and high velocity UE cases.

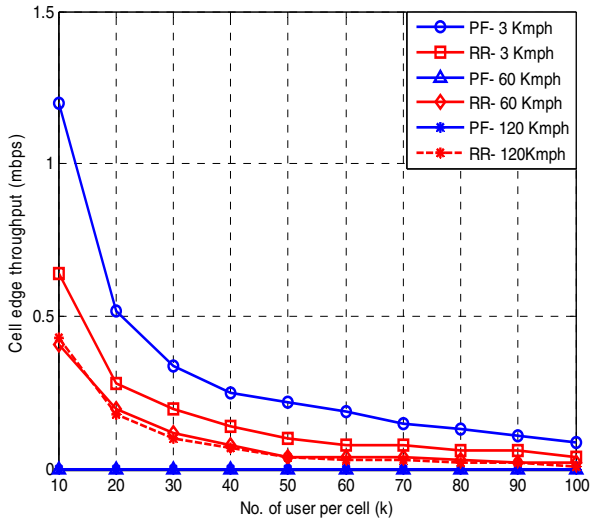


Fig. 3(b): Average cell edge throughput vs. UE density under PF and RR scheduler algorithm for low, medium and high velocity UE cases.

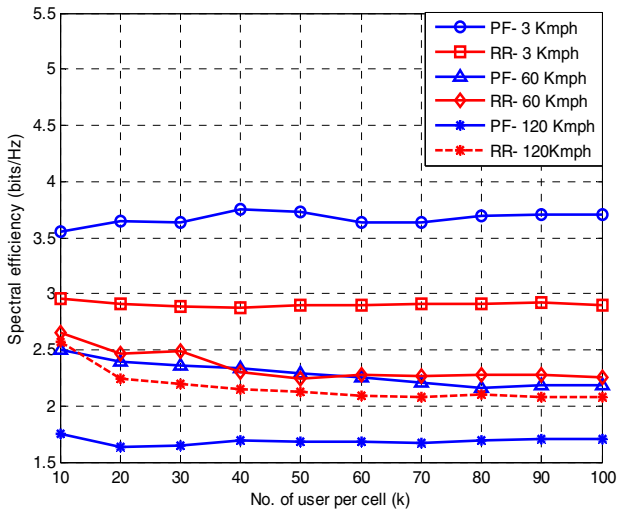


Fig. 3(c): Spectral efficiency vs. UE density under PF and RR scheduler algorithm for low, medium and high velocity UE cases.

UE density does not have a significant effect on the performance of PF or RR in terms on spectral efficiencies, shown in Fig. 3(c) and fairness indexes, depicted in Fig. 3(d). The impact of UE density on these metrics is not as severe as UE velocity.

Fig. 4 illustrates a better understanding on the utilization of the system bandwidth by the schedulers under different UE velocities. It is observed in Fig. 4(b) that for RR, the ratio of the average UE throughputs with various system bandwidths is almost constant indicating that similar spectral efficiencies are obtained by RR at different system bandwidths which is not much affected by UE velocity. Fig. 4(a) shows the decline in spectral efficiencies for various system bandwidths with increase in UE velocity under PF algorithm.

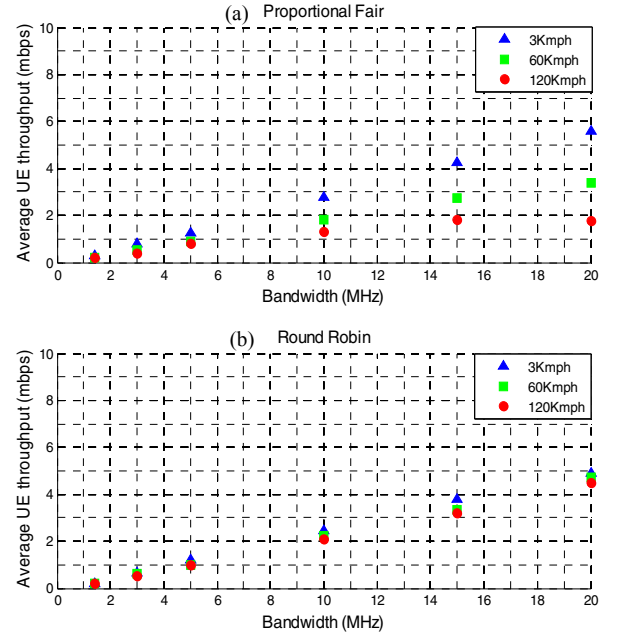


Fig. 4: Average UE throughput vs. system bandwidth under PF and RR scheduler algorithm for low, medium and high velocity UE cases.

V. CONCLUSION

According to our simulation results RR outperforms PF at high UE velocity and provides higher UE throughput and better fairness. Moreover at cell edges and at high velocity RR provides a low throughput which is very important to maintain a seamless network connection especially during UE handovers to prevent call drops. On the other hand PF fails to serve any UEs, moving at high velocity, at cell edges. RR also achieves higher spectral efficiency and better fairness at high UE velocity. Moreover our results show that the average UE throughput and cell edge throughput are sensitive to UE density and both the metrics decrease with an increase in UE density. In contrast spectral efficiency and fairness index are not significantly affected by a change UE density. Therefore for UEs moving at high velocities resource scheduling should be done using RR algorithm to achieve higher performances.

REFERENCES

- [1] Tara Ali-Yahiya, "Understanding LTE and its performance," *Springer* 2011.
- [2] Morten Tolstrup, "Indoor Radio Planning: A Practical Guide for 2G, 3G and 4G," 3rd edition, *Wiley* 2015.
- [3] 4G Americas, "Mobile Broadband Evolution Towards 5G: Rel-12 & Rel-13 and Beyond," June 2015.
- [4] D. Calin, H. Claussen, and H. Uzunalioglu, "On femto deployment architectures and macrocell offloading benefits in joint macro-femto deployments," *IEEE Commun. Magazine*, vol. 48, no. 1, pp. 26 –32, Jan.2010.
- [5] Evolved Universal Terrestrial Radio. "Physical layer procedures." (2008).
- [6] Y.Barayan and I.Kostanic, "Performance Evaluation of Proportional Fairness Scheduling in LTE," *World Congress on Engineering and Computer Science (WCECS)*, 2013.
- [7] M . H. Habaebi, J. Chebil, A.G. Al-Sakkaf and T. H. Dahawi, "Comparison between Scheduling techniques in long term evolution," *IJUM Engineering Journal*, Vol. 14, No. 1, pp 66-75. 2013.
- [8] S. A. AlQahtani and M.AlHassany "Performance Modeling and Evaluation of Novel Scheduling Algorithm for LTE Networks," *JSAT* 2014.
- [9] M. T. Kawser, H. M. A. B. Farid, A. R. Hasin, A. M. J. Sadik, and I. K. Razu, "Performance Comparison between Round Robin and Proportional Fair Scheduling Methods for LTE," *International Journal of Information and Electronics Engineering*, Vol. 2, No. 5, September 2012.
- [10] Y. Ji, Z. Yifan, W. Ying, Z. Ping , "Average rate updating mechanism in proportional fair scheduler for HDR," *IEEE Global Telecommunications Conference*, 2004, pp.3464–3466, doi:10.1109/GLOCOM.2004.1379010.
- [11] B. Chisung, and C. Dong, "Fairness-Aware Adaptive Resource Allocation Scheme in Multihop OFDMA System," *Communication letters IEEE*, 11 (2007) 134-136.
- [12] LTE System Level Simulator v1.6r885 by Institute Of communication and Radio Frequency Engineering, Vienna University of Technology, Vienna.

A New Method of Reversible Data Hiding Based on Compressed Gray Level Histogram Shifting

Tanwi Biswas, Md. Mehedi Hasan, Tanoy Debnath
Department of Electronics and Communication Engineering
Khulna University of Engineering & Technology
Khulna 9203, Bangladesh
tanwibiswas.tb@gmail.com, mehedi_ece@yahoo.com, kuettanoydebnath@yahoo.com

Abstract—Histogram shifting is one of the most useful techniques of Reversible Data Hiding (RDH). In histogram shifting based data hiding technique, lower distortion and improved capacity can be achieved. In this paper, a new method of reversible data hiding based on compressed gray level histogram shifting has been proposed which provides a high capacity of data hiding and also maintains a good image quality. In the proposed technique of data hiding, the range of the image pixel values is compressed by taking the square root of these pixel values. Considering the rounded pixel values increases the amount of peak pixel and thus also increases the data hiding capacity greatly. After data insertion, pixel values are squared and round values of the pixels are considered to form the image. At the receiver the hidden data is extracted and the actual cover image is regained by following the reverse technique of data insertion. The proposed scheme has been applied to various images to check its generality. The bit-rate and PSNR value obtained by the proposed method have been compared to a reference method and it was found that the proposed method provides a higher capacity than the reference method but PSNR value is a little bit lower in some cases.

Keywords—*histogram shifting; reversible data hiding; compressed gray level; PSNR; bit-rate;*

I. INTRODUCTION

Because of the large increase in the number of computer and network technology, the transmission of data over the internet has become a part of everyday life. However, this data transmitted over the internet need to be secured. The best way to provide the security is to hide the secret data. Data hiding refers to the technique of concealing secret data i.e., information into a covering media to protect it from unauthorized access [1]. The covering media can be image, audio, video files etc. Data hiding and transmitting using images is the most commonly used technique. Generally, data hiding scheme can be categorized into two main types, such as, irreversible data hiding scheme [2] and reversible data hiding technique [3,4]. In case of irreversible data hiding, only actual data can be regained but not the actual image. On the contrary, reversible data hiding technique provides the facility to recover the secret data and to regain the actual image.

Nowadays most of the applications such as medical imagery [5], military imagery, law enforcement etc. need to utilize both the secret message and the actual recovered image. So the reversible technique is getting more popular.

There are several techniques to achieve the reversible data hiding scheme such as integer-to-integer Wavelet Transform [6], Histogram Modification [7] and Difference Expansion [8]. Histogram shifting based data hiding is one of the most popular and effective way of reversible data hiding technique. Histogram Shifting based data hiding also supports multilayer data embedding [9]. The basic scheme of reversible data hiding based on histogram shifting was first proposed in 2006. Here, at first the peak and zero points of the histogram of image were searched. The secret data was carried out only by the peak pixel values and the pixels in between the peak and zero points were modified during data embedding but carried no data. In the receiver the pixels were modified in the reverse way of data inserting procedure to regain the secret data and to recover the actual image. Here, the hiding capacity was limited by the number of peak pixel values in the image.

Different techniques have been proposed later to increase the hiding capacity before shifting the histogram of the image. An updated technique of reversible data hiding based on histogram shifting with increased capacity was proposed by Piyu in 2009 [10]. Here, the uniformity of adjacent pixels in the images was searched by using the prediction method and to hide the secret data the residual histogram of the computed errors of the original image was used. Moreover, to further increase the hiding capacity overlapping of the peak and zero pairs was utilized. Here, the image was divided into small blocks and capacity of image was different for different block size.

Another improvement of Reversible Data Hiding technique based on Histogram Shifting was introduced by Wen-Chuan in 2012 [11]. This technique combines two remaining methods for embedding more data. According to this method the image blocks are at first categorized into smooth or complex ones

Here, the proposed method describes a reversible data hiding scheme based on compressed histogram shifting. The data hiding capacity of the image has increased by increasing peak pixel values due to the compression technique. The next parts of this paper are arranged like: Section 2 describes the basic histogram shifting based reversible data hiding method. Section 3 shows the proposed scheme with description. Section 4 includes experimental analysis and the paper is concluded with Section 5 and 6 which include conclusion and future plan respectively.

At the receiver, the concealed data is regained and the actual image is regenerated. Here, each pixel can be processed in two different ways. If the considering pixel value x is not in between the pair of zero and peak points, it is not checked. If x

At the receiver, the image and the pair of zero and peak points (4,6) are needed to extract the hidden data and to recover the original cover image. Each pixels of the stego-image is checked from left-to-right and from top-to-bottom. For peak pixel value i.e., pixel having value 6, 1-bit data with numeric value at 1 is derived and pixel value is kept unchanged. For considering pixel value being 5, 1-bit secret data 0 is derived and the pixel is adjusted by 1 toward the peak pixel i.e., 5 is replaced by 6. If the considering pixel value is 4, it is also adjusted by 1 closer to the peak pixel i.e., it is replaced by 5 and there is no derived data. Ultimately, the extracted data after completing all these process is $(101100)_2$

Fig. 2: Example of data hiding in image: (a) stego-image
(b) regenerated image.

and the actual cover image is reformed as shown in Fig. 2(b). The data hiding capacity of this proposed method depends on the number of peak pixels of an image. To increase the hiding capacity, it is also necessary to increase the number of peak pixels before inserting the secret data.

III. THE PROPOSED METHOD

To provide increased hiding capacity with keeping good quality of the transmitted-image is the main objective of the proposed method. For increasing the capacity to conceal the data in histogram-based scheme, computing the square root of the pixel value is employed here. This compresses the range of image pixels and increases the amount of peak pixel values in the image histogram. With increasing the peak pixel values, the bit rate i.e. capacity also increases. The steps for embedding data in the image using the proposed method are shown in Fig. 3.

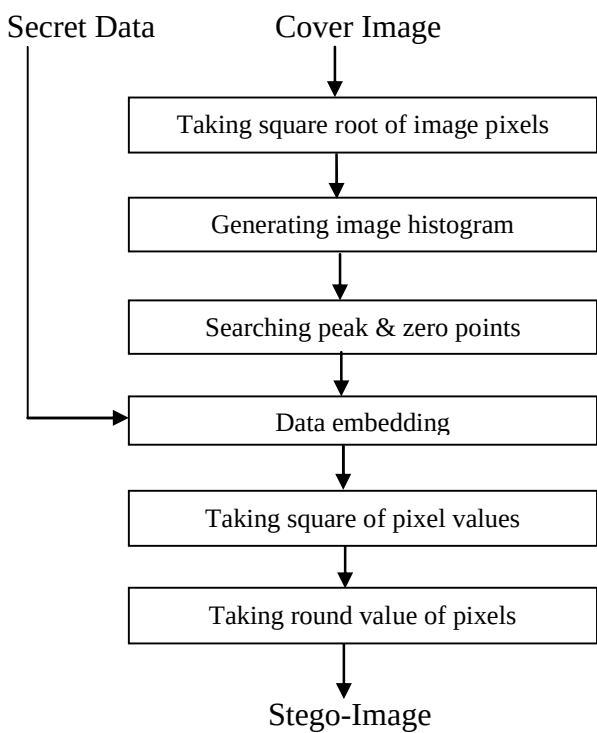


Fig. 3: Steps of the proposed embedding method for data embedding.

2.2361	2.4495	2.4495	2.4495	2.6458
2.2361	2.2361	2.4495	2.4495	2.4495
1.7321	1.7321	2.2361	2.2361	0
1.0000	1.4142	1.4142	1.4142	1.4142
1.0000	1.4142	1.7321	1.7321	1.0000

Fig. 4: Root of image pixels

Here a gray scale image is considered for embedding the data to be hidden. At first the square root of the pixels of image is computed i.e. the range of gray-scale pixel values of the image is compressed. Then the pair of zero and peak point is searched from the histogram of the round value of these pixels. The secret data is then embedded. Three types of changes of the image pixels occur during embedding this secret data. Firstly, all the pixels value amongst the pair of zero and peak points are moved by 1 toward the zero points. Secondly, if the secret data is 1 the peak pixel value is kept unchanged and thirdly for the secret data being 0, the pixel value is adjusted by 1 toward the zero point. After completing the insertion of the secret data, the pixels of the image are squared and after taking the round value of the pixels the image is then ready for transmission.

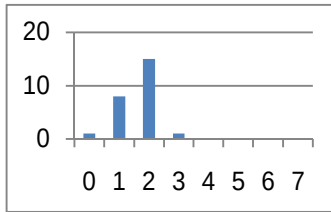
With the following example the method can be easily understood: Reconsider the matrix shown in Fig. 1(a) as the cover image and Fig. 1(b) shows the corresponding image histogram. After taking the square root of the pixel values the matrix shown in Fig. 4 is obtained.

For searching the pair of zero and peak points of the image histogram first of all the round value of pixels are considered which is shown in Fig. 5(a) and from the image histogram of these round value as displayed in Fig. 5(b), the zero and peak points are obtained. Here the peak point is considered as 2 and the zero point is considered as 4. After that the data is inserted into the image.

Suppose, secret data need to be embedded into the image is $(101100101011001)_2$. The image after inserting the secret data is shown in Fig. 6. From the figure it is seen that the pixel value at point (0,4) has increased by 1 i.e. adjusted to the zero point by 1. When the secret data is 1, the pixels are unchanged at points (0,0), (0,2), (0,3), (1,2), (1,4), (2,1),(2,2) and (4,3). For the secret data 0, the pixels have increased by 1 i.e. adjusted by a value closed to the zero point by 1. In this matrix the pixels at points (0,1), (1,0), (1,1), (1,3), (2,0), (2,3) and (4,2) have increased by 1 for secret data 0. After inserting this data square value of the image pixels are considered which is shown in Fig.7. Finally the round value of the image pixels are considered to form the image as shown in Fig. 8. The image and the pair of zero and peak points are then transmitted.

2	2	2	2	3
2	2	2	2	2
2	2	2	2	0
1	1	1	1	1
1	1	2	2	1

(a)



(b)

Fig. 5: Example of the proposed method (a) Rounding square value (b) Histogram of rounded pixels

2.2361	3.4495	2.4495	2.4495	3.6458
3.2361	3.2361	2.4495	3.4495	2.4495
2.7321	1.7321	2.2361	3.2361	0
1.0000	1.4142	1.4142	1.4142	1.4142
1.0000	1.4142	2.7321	1.7321	1.0000

Fig. 6: After inserting the data

5.0000	11.8990	6.0000	6.0000	13.2915
10.4721	10.4721	6.0000	11.8990	6.0000
7.4641	3.0000	5.0000	10.4721	0
1.0000	2.0000	2.0000	2.0000	2.0000
1.0000	2.0000	7.4641	3.0000	1.0000

Fig. 7: After squaring the pixel values

5	12	6	6	13
10	10	6	12	6
7	3	5	10	0
1	2	2	2	2
1	2	7	3	1

Fig. 8: Transmitting image

After receiving the image, the secret data is derived and the image is recovered. The flowchart for extracting data according to the proposed method is shown in Fig. 9. The stego-image and the pair of zero and peak point of the image histogram are needed at the receiver for deriving the hidden data and regaining the actual image. At first square root of the image pixels are taken. Each pixel is checked from left to right and top to bottom. Data is extracted by the reverse technique of insertion. After data extraction, pixel values are squared. Then considering the round value of pixels, original cover image is regenerated.

For better understanding this extraction procedure, suppose, the image shown in Fig. 8 is received. The square root of the pixel value of the image is first computed which is shown in Fig. 10. Each pixels of the image is then checked from left-to-right and top-to-bottom. If pixels with peak value are faced, 1 bit data with value 1 is stored and the pixel is kept unchanged. For pixel values having difference 1 from the peak point, a 1 bit data 0 is derived and pixel value is moved by 1 at a value closer to peak point. The pixels amongst the zero and peak points are adjusted by 1 to a value closer to the peak point and no data is stored. After performing this procedure on Fig. 10, the image shown in Fig. 11 is obtained and the data $(101100101011001)_2$ is extracted. After extracting the data steps are followed to reform the original image. To obtain this the pixels of the image displayed in Fig. 11 are squared and the image shown in Fig. 12 is found. Since fractional values are not valid for the image pixels, the rounded value of the image pixels are considered which is shown in Fig. 13. From Fig. 13, it is clear that this image is similar to the actual cover image. So it can be said that the proposed method is able to extract data and to recover the original image without any loss.

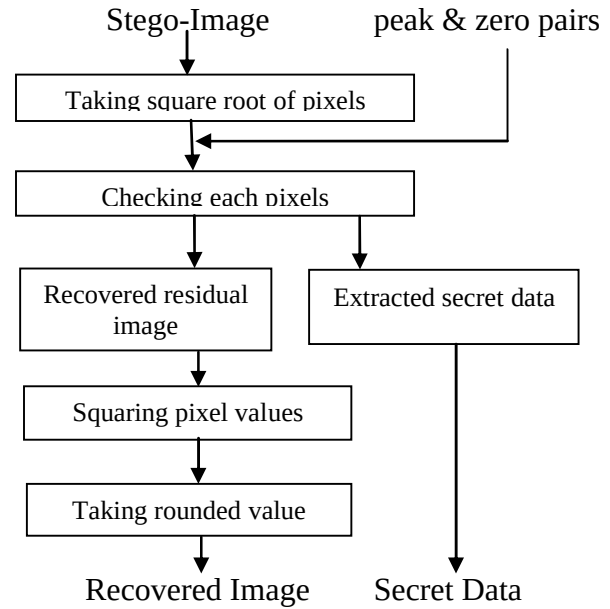


Fig. 9: Flowchart of the proposed method for extracting data and recovering image.

2.2361	3.4641	2.4495	2.4495	3.6056
3.1623	3.1623	2.4495	3.4641	2.4495
2.6458	1.7321	2.2361	3.1623	0
1.0000	1.4142	1.4142	1.4142	1.4142
1.0000	1.4142	2.6458	1.7321	1.0000

Fig. 10: Taking root of received image pixels

2.2361	2.4641	2.4495	2.4495	2.6056
2.1623	2.1623	2.4495	2.4641	2.4495
1.6458	1.7321	2.2361	2.1623	0
1.0000	1.4142	1.4142	1.4142	1.4142
1.0000	1.4142	1.6458	1.7321	1.0000

Fig. 11: After data extraction

5.0000	6.0718	6.0000	6.0000	6.7889
4.6754	4.6754	6.0000	6.0718	6.0000
2.7085	3.0000	5.0000	4.6754	0
1.0000	2.0000	2.0000	2.0000	2.0000
1.0000	2.0000	2.7085	3.0000	1.0000

Fig. 12: Squaring image pixels

5	6	6	6	7
5	5	6	6	6
3	3	5	5	0
1	2	2	2	2
1	2	3	3	1

Fig. 13: Recovered Image

IV. RESULT ANALYSIS

In this research, the proposed method was applied to several gray scale images such as Lena, Peppers, Man, Barbara, Boat, Elaine, Baboon and MRI which are shown in Fig. 14. In all cases, (a) represents the original image and (b) represents the images after data insertion. All the images are of size 512×512. After comparing the actual image and the data inserted image, it is proved that the image quality after data insertion is quite good. To evaluate the proposed method's performance, Embedding Capacity (EC) and Peak Signal-to-Noise Ratio (PSNR) were calculated for the images mentioned above. Embedding Capacity or Bit-rate can be defined as the ratio of the number of bits of data an image can carry to the total number of pixels of that image.

EC = No. of bits carried by an image/ total no. of pixels

The PSNR value determines the quality of the image after data is inserted. PSNR is measured in dB. The value of PSNR depends on Mean Squared Error (MSE). Suppose, I is an $m \times n$ noise-free image and the noisy approximation of I is given by K . Now MSE is defined by the equation (1) as follows:

$$MSE = \frac{1}{m \times n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2 \dots\dots(1)$$

The PSNR can be defined by equation (2) as follows:

$$PSNR = 10 \log_{10} \left(\frac{MAX^2}{MSE} \right) \dots\dots\dots(2)$$

Here, MAX is the maximum possible value of the image pixels.

Here, we have compared the result of the proposed method to the method of Piyu [10]. Since the basic technique of data hiding and extracting is similar to that of the Piyu's method [10], we have chosen it for comparing the result. For this, both EC and PSNR have been calculated for the images mentioned above also by applying the method of Piyu [10]. Table 1 represents the number of bits an image can carry and the bit-rate or embedding capacity of the image both for the proposed method and the method of Piyu [10]. It is seen from the table that for every image the bit-rate of the proposed method is greater than that of the reference method [10].

Table 2 represents the comparative PSNR performance of the proposed method to the method of reference [10]. Here, the proposed method due to carrying a larger number of data, provides a low PSNR value compared to the method of Piyu [10].

Bit-rate vs PSNR graph represents how the PSNR of an image changes with each bit insertion into the image. Fig. 15 shows the bit-rate vs PSNR graph of an 225×225 image. Here, bit-rate of the proposed method is greater but the PSNR value is lower than the reference method [10].



Fig. 14: Images of Lena, Peppers, Man, Barbara, Boat, Elaine, Baboon, MRI from top to bottom respectively (a) Original image (b) Transmitted image.

TABLE I. BIT-RATE OF VARIOUS IMAGES ACCORDING TO THE PROPOSED METHOD AND PIYU'S METHOD [10]

Name of images	No of bit carry (Piyu's Method [10])	Bit-rate (Piyu's Method [10]) bpp	No of bit carry (proposed method)	Bit-rate (Proposed method) bpp
1.Lena	37146	.1419	54832	.2092
2.Peppers	31674	.1208	44840	.1711
3.Man	33845	.1291	48748	.1860
4.Barbara	28135	.1073	36066	.1376
5.Boat	25096	.0957	99808	.3807
6.Elaine	23401	.0893	51249	.1955
7.Baboon	12935	.0493	53916	.2057
8.MRI	106408	.4059	111655	.4259

TABLE II. PSNR VALUES OF VARIOUS IMAGES ACCORDING TO THE PROPOSED METHOD AND PIYU'S METHOD [10]

Name of images	PSNR of Piyu's method [10] dB	PSNR (proposed method) dB
1. Lena	50.02	24.87
2.Peppers	49.81	23.83
3.Man	50.16	24.81
4.Barbara	50.40	23.76
5.Boat	49.84	25.47
6.Elaine	49.52	24.98
7.Baboon	49.92	27.18
8.MRI	50.65	53.62

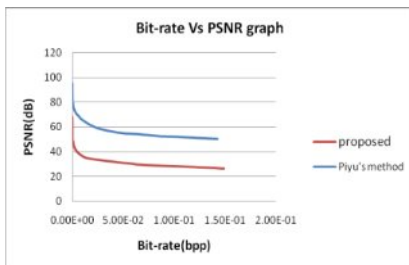


Fig. 15: Bit-rate vs PSNR graph

V. CONCLUSION

In this paper, we have proposed a new method of reversible data hiding which is based on the basic histogram shifting based data hiding scheme. The proposed method compresses the gray scale of the images by taking square root of the image that drastically increases the number of peak pixels of an image. Thus the number of secret data that an image can carry also increases largely. But due to large amount of data insertion the PSNR performance provided by the proposed method is little low.

VI. FUTURE WORK

Though the proposed method maintains good visual quality of images, the PSNR value which mathematically determines the quality of images after insertion of data is a little bit low for some images. So in future we will try to improve the PSNR value i.e., the image quality with more mathematical analysis. We will also compare the result of the proposed method with some other methods of reversible data hiding.

References

- [1] W. Zeng, "Digital watermarking and data hiding: technologies and applications," in *Proc. Int. Conf. Inf. Syst., Anal. Synth.*, vol. 3, pp. 223–229, 1998.
- [2] Q. Cheng, T.S. Huang, "An additive approach to transform-domain information hiding and optimum detection structure", *IEEE Transactions on Multimedia* 3 (3) pp. 273–284, 2001.
- [3] J. Tian, "Reversible data embedding using a difference expansion", *IEEE Transactions on Circuits and Systems for Video Technology* 13 (8) pp. 890–896, 2003.
- [4] K. Yamato, K. Shinoda, M. Hasegawa, S. Kato, "Reversible data hiding based on two-dimensional histogram and generalized histogram shifting", *IEEE international conference on image processing*, pp 4216–4220, Oct. 2014.
- [5] R. Norcen, M. Podesser, A. A. Pommer, H. Schmidt and A. Uhl, "Confidential storage and Transmission of Computers in Biology and Medicine 33, pp. 277–292, 2003.
- [6] S. Lee, C. D. Yoo, T. Kalker "Reversible Image Watermarking Based on Integer-to-integer wavelet transform", *IEEE trans. on info. Forensics and security*, vol. 2, pp. 321–330, Oct. 2007.
- [7] Zhenfei Zhaoa, Hao Luoc, Jeng-Shyang Pand "Reversible data hiding based on multilevel histogram modification and sequential recovery" *International journal of electronics & communications*, 2010.
- [8] J. Tian, "Reversible data embedding using a difference expansion," *IEEE Trans. Circuit syst. Video tech.*, vol-13, no 8. pp 890–896, aug-2003.
- [9] Z. Pan, S. Hu, X. Ma, L. Wang, "Reversible data hiding based on local histogram shifting with multilayer embedding", *Elsevier, Journal of visual com. & image representation*, vol. 31, pp. 64–74, August 2015.
- [10] Piyu Tsai, Yu-Chen Hu, Hsiu-Lien Yeh, "Reversible image hiding scheme using predictive coding and histogram shifting" *Elsevier, Signal Processing* 89 pp. 1129–1143, 2009.
- [11] Wen-Chuan Wu, Yu-Chieh Wu, "Improved Reversible Data Hiding Based on Residue Histogram Shifting" *Sixth International Conference on Genetic and Evolutionary Computing*, pp. 23–26, 2012.
- [12] Zhicheng Ni, Yun-Qing Shi, Nirwan Ansari, and Wei Su, "Reversible data hiding" *IEEE Transactions on Circuits and Systems for Video Technology* 16 (3), 354–362, 2006.

An Approach of Cardiac Disease Prediction by Analyzing ECG Signal

Tahmida Tabassum, Monira Islam

Department of Electrical & Electronic Engineering, KUET, Khulna-9203, Bangladesh

e-mail: tahmidkuet@gmail.com, monira_kuet08@yahoo.com

Abstract—Electrocardiogram (ECG) gives useful information about morphological and functional details of heart which is used to predict various cardiac diseases. In this paper a method of detecting cardiac diseases using support vector machine (SVM) is proposed. In this proposed method diseases are modeled using the time domain features of ECG signal which are extracted using BIOPAC AcqKnowledge software. Raw ECG signal contains these useful features which can be used to detect cardiac arrhythmia. The various ECG parameters like heart rate, QRS complex, PR interval, ST segment elevation, ST interval of ECG signal are used for analysis. Based on these parameters of ECG signal, different heart disease like atrial fibrillation, sinus tachycardia, myocardial infarction and apnea are detected. The individual accuracy of tachycardia arrhythmia, MI arrhythmia, atrial fibrillation arrhythmia and apnea proposed by SVM are 83.3%, 86.4%, 88% and 85.7% respectively.

Keywords— ECG signal; QRS complex; ST segment; heart rate; Support Vector Machine (SVM)

I. INTRODUCTION

Humans are most commonly affected by cardiac diseases nowadays. Many people are suffering from heart attack due to stress and some of them had to take up coronary artery bypass surgical treatment for their cardiac problem. To prevent this unwanted heart attack, early detection and timely treatment of arrhythmia is necessary. This would slow down the progression of heart failure and prevent the loss of life [1].

The graphical recording of the electrical signals generated by the heart is known as ECG which provides diagnostic information about the cardiac condition of a patient. P wave is the first electrical positive signal in the normal ECG. It has positive polarity and duration is less than 120 milliseconds. The largest part of the ECG signal is the QRS complex which is the result of contraction of the ventricles where:

- Q wave is the first negative or downward deflection.
- R wave is always the first positive deflection,
- S wave is the negative deflection followed by the R wave [2].

The difference in amplitude and interval of different wave provides a measure for detecting various cardiac abnormalities. The most distinguishable time domain features of ECG signal are QRS-complex, RR interval, ST segment etc [3]. Disturbance of this electrical function is common in cardiac disease. The abnormalities in heart is found by the

doctors by observing the deviation of P, QRS and T signal from the normal signal in terms of time duration and amplitude[4].

ECG is a bipolar low frequency weak signal and the normal range of this signal is 0.01-150Hz. Its amplitude ranges from 0.05-3mV and normal heart rate is 60-100 BPM [5]. A general ECG signal and its associated fiducial points are shown in Fig. 1.

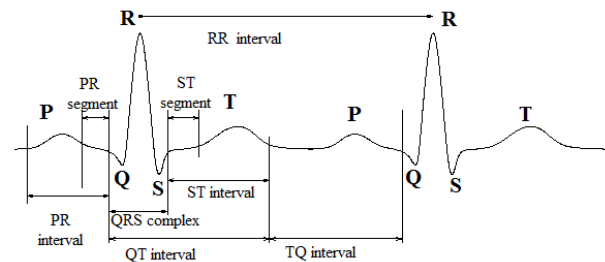


Fig. 1. A real time ECG signal

As computer can measure the features of ECG signal more precisely than human eye, therefore an automatic ECG analysis is more preferable in cardiac monitoring and detection of cardiac condition [2]. Normal sinus rhythm refers to any cardiac rhythm where contraction (depolarization) of the cardiac muscle which begins at the sinoatrial node [4]. A normal sinus rhythm shows the following five characteristics

- normal P wave
- constant P-P and R-R interval
- Constant P wave configuration in a given lead
- P-R interval and QRS interval within normal limit
- Heart rate between 60 to 100 beats/min

Cardiac arrhythmias refer to any abnormality or perturbation in the normal activation sequence of the myocardium [5]. Depending upon R-R interval arrhythmias are of two types named as bradycardia and tachycardia. Different cardiac diseases are: myocardial infarction (MI), hypertrophy, atrial fibrillation, apnea etc. Myocardial infarction (MI), commonly known as heart attack, is due to the occlusion of one of the coronary arteries or one of its smaller branches. The MI evolves through the sequence:-

1. T-wave peaking followed by T-wave inversion
2. ST-segment elevation
3. Appearance of pathological Q-wave

In case of cardiac disease, the ECG features vary in different ways such as absence of P wave, wider QRS

complex, absence of one complete PQRS complex or irregular RR interval etc. Change in different features for normal ECG signal and cardiac disease are shown in TABLE I [6].

TABLE I. VARIATION IN DIFFERENT FEATURES OF ECG SIGNAL SHOWING DIFFERENT CARDIAC DISORDERS

Heart condition	Rhythm	Heart beat (BPM)	PR interval (sec)	QRS complex (sec)	P wave (mV)	R height (mV)	ST segment
Normal sinus rhythm	Regular	60-100	0.12-0.2	0.04-0.12	<0.25	2.5-3.0	Normal in baseline
Atrial fibrillation	Irregular	400-600	Not visible	-	absent	-	Normal in baseline
Sinus tachycardia	Regular	>100	0.12-0.2	0.04-0.12	<0.25	-	Normal in baseline
Myocardial infarction	Regular	-	-	-	-	-	Elevated from baseline
Apnea	Regular	60-67	0.12-0.2	0.04-0.12	-	-	Normal in baseline

To detect these features of ECG signal various methods can be used. Authors in [7] used multi-scale discrete wavelet transform based peak detection method to detect the QRS complex for identifying the peak and extracting the fetal features of ECG. Fetal heart rate data and its variation have been measured to define the diagnostic indices for fetal growth monitoring [7]. Authors in [8] applied empirical mode decomposition (EMD) method for diagnosing cardiovascular diseases. Using time domain analysis different time domain features like RR intervals, standard deviation (SDNN) of RR intervals, variance and differences of root mean square (rms) values of successive RR intervals have been calculated which has been used to detect different cardiovascular disease more accurately. By analyzing ECG signal cardiovascular abnormalities and ischemia has been detected in [3]. A neural-fuzzy approach is used in [9] for classifying myocardial Infarction (MI) arrhythmia with back propagation neural networks (BPNN) classifier. In this case the ST elevation is considered as an input parameter. Authors in [10] proposed LABVIEW virtual instrumentation system for heart disease detection which has the advantage of automatic noise removal and filtration of acquired raw ECG signal. Authors in [11] proposed an artificial neural network (ANN) based cardiac arrhythmia diagnosis system which mainly classifies disease in normal and abnormal classes. In this paper, we have considered support vector machine (SVM) to classify arrhythmia cases into normal and abnormal as well as into the types of diseases possessed. The data set used for experimental purpose contains the time domain attributes of ECG signal.

The remainder of the paper is organized in the following sections: proposed method is described in Section II and experiments and analysis are given in Section III. Section IV concludes the paper.

II. METHODOLOGY

This method works in two stages. In first stage a subject is classified as sound or patient and the type of disease is

distinguished in second stage. Fig. 2 shows the main steps of the proposed method.

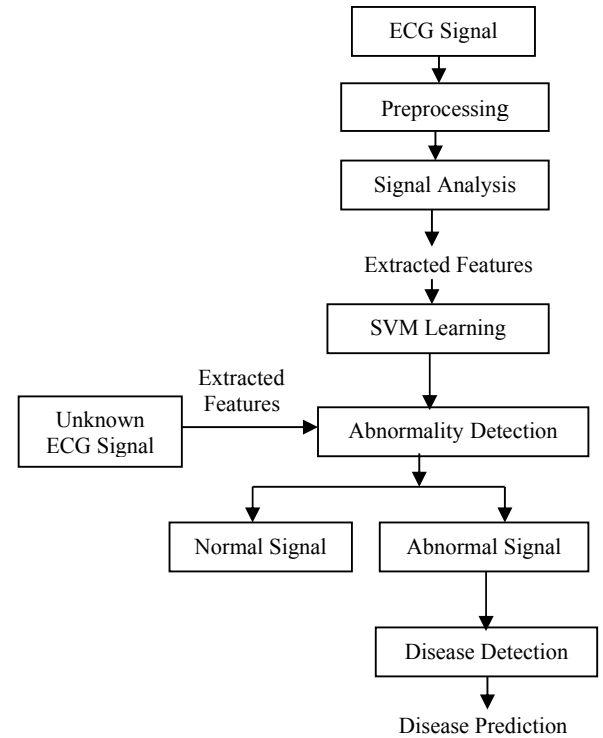


Fig. 2. Proposed method of disease detection from ECG signal

In this proposed method, the input is the ECG signal itself. BIOPAC data acquisition unit and AcqKnowledge® 4.0 Software are used to collect data, analysis, store and recover in the research [12]. In this work, we used data collected from 70 different subjects.

In case of analyzing ECG signal the sampling rate used was 320 samples per seconds. For sufficient resolution in

capturing the abrupt changes of ECG signal the digitization rate was chosen as 1-KHz. ECG signal was collected for classification from different subjects in laboratory. The pictorial view of ECG data collection in laboratory is shown in Fig. 3.



Fig. 3. Pictorial view of ECG signal extraction

A. Preprocessing

The quality of ECG signal is affected by the motion artifact, electrical interference or baseline wandering. The signal was filtered by a band pass filtering to remove these noises which has a high pass cut off frequency of 2 Hz and a low pass cut off frequency of 320 Hz. Then a low pass filter was used to remove the high frequency noise. The raw signal was smoothed to identify and reject the undesirable noise shown in Fig. 4.

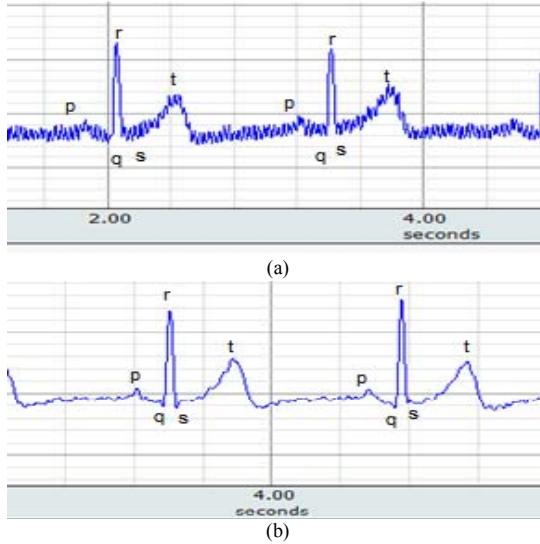


Fig. 4. ECG signal (a) raw signal and (b) filtered signal

The smoothing function is one type of transformation that computes the moving average of a series of data points and replaces each value with a mean value. The mean and median value smoothing formula is given in equation (1) and (2) respectively.

$$(X_{ECG}) = \sum_{i=n-\frac{y}{2}}^{i=n+\frac{y-1}{2}} \frac{f_1(ECG)}{y} \quad (1)$$

$$f(X_{ECG}) = \text{median}(n - \frac{y}{2}; n + \frac{y}{2}) \quad (2)$$

Where y is the number of points and n is the sample numbers, $f(X_{ECG})$ and $f_1(X_{ECG})$ are the input and output signal respectively.

B. Signal Analysis

The ECG signal was then analyzed using AcqKnowledge software and different time domain features of that signal had been extracted. Here different peaks and intervals such as R-peak, P wave, ST interval, QRS complex, RR interval etc of ECG signal had been extracted. In this paper, we had considered heart rate, PR interval, QRS complex, ST interval and ST segment elevation. These features have the prime importance in finding out the patient's heart condition. Normally one ECG cycle represents one beat and heart rate is calculated by considering the number of cycles in 60 sec.

$$\text{HeartRate} = \frac{360}{RR_interval} \times 60 \quad (3)$$

This analyzing process using BIOPAC AcqKnowledge helps to find out each peak like P, Q, R, S and T waves as well as the durations in each cycle of ECG signal. These duration and amplitude of different waves and intervals are calculated as follows:

R peak: It is the first positive deflection due to ventricular depolarization.

RR interval: By considering the time difference of two consecutive R peaks RR interval is calculated.

PR interval: It is calculated by considering the time difference between P and Q wave, if Q wave is absent then it is the time difference between P and R wave.

QRS complex: The duration of QRS complex must be less than 0.1 sec and measured by finding the time difference between the beginnings of Q wave to the end of S wave.

ST interval: It is calculated by measuring the time difference from the end of QRS complex to T wave.

Isoelectric line and ST segment location: It is the almost zero amplitude level or baseline. Normally ST segment lies in the ECG baseline.

Elevation of ST segment is measured by calculating the slope of the wave from the baseline. All the amplitudes of ECG signal are measured from the baseline and if there is any baseline shift, the value of shift has been considered. The features extracted for seven arbitrary subjects used in this research are given in TABLE II.

C. Abnormality Detection Based on SVM

Support vector machine is a popular machine learning techniques which uses a hyper-plane (linear and nonlinear mapping) applied for various purposes like ECG signal classification based on the features extracted. The classification process involves separating data into two sets

such as training and testing sets. Suppose training set of instance-label pairs $(x_i, y_i); i=1, \dots, l$ where $x_i \in R^n$ and $y_i \in \{1, -1\}$, the support vector machines (SVM) requires the solution of the following optimization problem

$$\begin{aligned} \min_{\bar{w}, b, \xi} \quad & \frac{1}{2} \bar{w}^T \bar{w} + C \sum_{i=1}^l \xi_i \\ \text{Subject to} \quad & y_i (\bar{w}^T \phi(\bar{x}_i) + b) \geq 1 - \xi_i \\ & \xi_i \geq 0 \end{aligned} \quad (4)$$

Where $\bar{x}_i$ are data samples, $\phi(\bar{x}_i)$ is the feature map, y_i are class labels, $C > 0$ is a positive regularization parameter, ξ_i are slack variables, $\bar{w}$ is the norm vector of the separating hyper-plane and b is the bias [13]. The following dual problem is usually has to be solved due to possible high dimensionality of the vector variable w .

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \alpha^T Q \alpha - e^T \alpha \\ \text{Subject to} \quad & y^T \alpha = 0 \\ & 0 \leq \alpha_i \leq C \end{aligned} \quad (5)$$

Where $i = 1, \dots, l$ and $e = [1, \dots, 1]^T$ is the vector of all ones, Q is an l by l positive matrix

$$Q_{ij} = y_i y_j k(x_i, x_j)$$

And the kernel function is defined by

$$k(x_i, x_j) \equiv \phi(\bar{x}_i)^T \phi(\bar{x}_j) \quad (6)$$

Support vector machines (SVM) have high generalization capabilities in terms of detecting heart condition. In SVM, the learning and classification component consists of two modules: a training module and a classification module for different heart condition. In this work, SVM classification was performed through MATLAB as it is simple to implement and gives output in an understandable form. For better rate of detection, SVM modeling was implemented through RBF Kernel function which is defined as-

$$k(x_i, x_j) \equiv \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0 \quad (7)$$

Where γ is a kernel parameter.

The extracted features are classified using SVM technique to detect the cardiac condition of a person. If the person's ECG signal is found abnormal, then his/her ECG features are further classified to identify the cardiac arrhythmia.

III. RESULT AND DISCUSSION

The magnitude and duration of different peak and intervals of ECG signal changes with the cardiac condition for different subjects. The features extracted are analyzed using SVM techniques. The changes in different features and inference are shown in TABLE II. From the extracted features the normal or abnormal ECG signal is identified using SVM analysis technique.

TABLE II. EXTRACTED FEATURES FROM RAW ECG SIGNAL FOR DIFFERENT SUBJECTS

	RR interval (sec)	Heart Rate (BPM)	PR interval (sec)	QRS complex (sec)	ST interval (sec)	ST segment elevation	SVM output
S_1	0.2725	220.18	0.1355	0.094	0.36	0	Abnormal
S_2	0.745	80.53	0.17	0.11	0.27	0	Normal
S_3	0.74	81.1	0.18	0.11	0.285	0	Normal
S_4	0.735	79.47	0.07	0.34	0.24	0.15	Abnormal
S_5	0.775	77.42	0.18	0.105	0.28	0	Normal
S_6	0.45	185.1852	0	0.11	0.15	0	Abnormal
S_7	0.588	62.04082	0.14	0.32	0.238	0	Abnormal

The plot of extracted time domain features for different subjects shows that there is variability in case of abnormal cardiac condition. Heart rate for normal patient is in the range of (60-100) BPM and usually it is constant. This feature varies widely for abnormal heart condition. Fig. 6 shows the plot of

heart rate for three different subjects. This attributes remain almost constant throughout all the cycles of the ECG signal in case of sound subjects. But it varies from the normal range while the subject possesses any kind of heart disease.

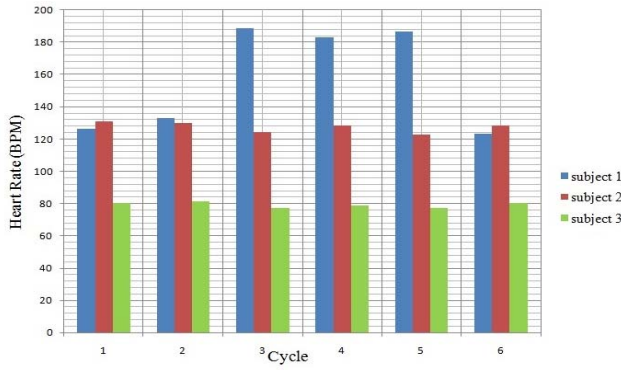


Fig. 6. Plot of heart rate for three arbitrary subjects

The abnormal signals were further classified using multiclass SVM classifier to predict the type of disease. The detected diseases are shown TABLE III for cardiac patients.

TABLE III. RESULT AND ACCURACY OF SVM CLASSIFIER FOR CLASSIFICATION OF DISEASES

	SVM classifier output		Accuracy
	Cardiac Condition	Disease	
S_1	Abnormal signal	Sinus tachycardia	83.3%
S_4	Abnormal signal	Myocardial infarction	86.4%
S_6	Abnormal signal	Atrial Fibrillation	88%
S_7	Abnormal signal	Apnea	85.7%

IV. CONCLUSION

ECG signal plays a vital role in case of cardiac disease detection as it contains all functional information about cardiac condition of a person. In this paper, we proposed a cardiologist friendly approach for cardiac disease prediction using SVM classifier. The time domain features of ECG signal are capable of discriminating healthy signal and patient signal as well as can differentiate different types of cardiac diseases. The classification accuracy for cardiac disease detection is found 84.6% which can help to a more accurate early diagnosis of different cardiac

diseases. As time domain features are used for the detection of cardiac diseases, this method would be very helpful for the medical professionals and will be a useful component for clinical decision support system to detect cardiac arrhythmia more precisely.

- [1] <https://www.google.com>
- [2] S. Jayalalitha, D. Susan, Shalini Kumari and B. Archana, "K-nearest Neighbour Method of Analysing the ECG Signal (To find out the Different Disorders Related to Heart)", Journal of Applied Sciences, 14: 1628-1632
- [3] Malgina, O., Milenkovic, J.; Plesnik, E.; Zajc, M.; Tasic, J.F., "ECG Signal Feature Extraction and Classification Based on R Peaks Detection in the Phase Space", pp. (381-384), February 2011.
- [4] G. K. Sahoo, S. Ari and S. K. Patra, "ECG signal analysis for detection of cardiovascular abnormalities and ischemia episodes", pp. (1055-1059), April 2013.
- [5] ABM Abdullah, "ECG in Medical Practice", 4th edition
- [6] L. N. Sharma, R. K. Tripathy and S. Dandapat, "Multiscale Energy and Eigenspace approach to Detection and Localization of Myocardial Infarction", IEEE Trans. J. Magn, vol. 62, Issue 7, pp. (1827-1837), June 2015.
- [7] A. M. Al-Busaidi, L. Khriji, "Digitally filtered ECG signal using low-cost microcontroller", pp. (258-263), May 2013.
- [8] Desai K.D., Sankhe M.S., "A Real Time Fetal ECG Feature Extraction Using Multiscale Discrete Wavelet Transform", pp. (407-412), October 2012.
- [9] Islam M.R., Ahmad S., Hirose K. and Molla M.K.I, "Data Adaptive Analysis of ECG signal for cardiovascular Disease Diagnosis", pp. (2243-2246), June 2010.
- [10] M. Arif, I. A. Malagore, F. A. Afsar, "Automatic Detection and Localization of Myocardial Infarction using Back Propagation Neural Networks", pp. (1-4), June 2010.
- [11] Poonam Kaur, R.K. Sharma, "Labview Based Design of Heart Disease Detection System", pp. (1-5), May 2014.
- [12] Shivajirao M., Jadhav M. Sanjay L.N., Ashok A. Ghatol, "Artificial Neural Network Based Cardiac Arrhythmia Disease Diagnosis", pp. (1-6), July 2011.
- [13] Reference manuals of AcqKnowledge software guide and MP system hardware guide, (2008). BIOPAC Systems Inc., 42 Aero Camino, Goleta, CA-0067, USA.
- [14] S. Yu, A. Daemen, L. C. Tranchevent, J. A. Suykens, B. De Moor, "L2-norm multiple kernel learning and its application to biomedical data fusion", BMC Bioinformatics, vol.-11, 2010.

PI Controlled Bi-Directional DC-DC Converter (BDDDC) and Highly Efficient Boost Converter for Electric Vehicles

Remon Das

Department of Electrical and Electronics Engineering
International Islamic University Chittagong
Shitakundu, Chittagong, Bangladesh
remondas4@gmail.com

Md. Ashraf Uddin Chowdhury

Department of Electrical and Electronics Engineering
International Islamic University Chittagong
Shitakundu, Chittagong, Bangladesh
ashraf.ctg7@gmail.com

Abstract - Due to the constraint of the fuel, the electric vehicle is designed for the modern world. Bidirectional DC to DC converters (BDDDC) has significantly enlarged the interest in Electric vehicle applications. A bidirectional DC to DC converter for Electric Vehicle (EV) for driving a DC motor is the aim of this present work. In a bi directional converter (conventional) PWM (Pulse Width Modulation) pulse is castoff for the generation of triggering pulse of the switches. We use PI and PID controller for triggering the IGBTs of DC to DC converter. A comparison is done between the three types of bidirectional dc to dc converter. Then a boost converter is designed over conventional boost converter by using parallel combinations of IGBT and MOSFET for electric vehicle applications.

Index Terms - PI Controller, Bi-Directional DC-DC Converter, Mixed parallel operation, Regenerative braking.

I. INTRODUCTION

Recently in most of the electric vehicle batteries are used as primary energy-storage devices including solar cars. There are so many advantages such as: noble transient operation, certain load levelling, nearly zero discharge and energy recovering during braking operation electric drive by battery fed are used for electric vehicle application. A highly efficient Bi directional converter with power flow capabilities in both way and a boost converter are mandatory to connect the accumulator (battery) of the motor drive system to the dc link [1].

The block of electric vehicle is as follows:

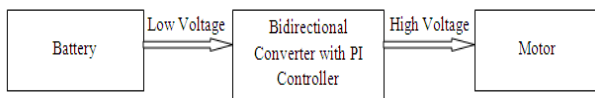


Fig 1: Battery Discharging Mode (Normal or Motoring Operation).

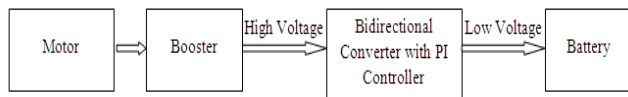


Fig 2: Battery Charging Mode (Regenerating Operation).

II. MODELING OF A BOOST CONVERTER

Getting higher efficiency is a race for improving the product. For innovative power conversion products, innovative topologies play an important role for the development. By using two novel properties of IGBT and MOSFET are low ON state Voltage and fast switching characteristic respectively, a new device is developed which is achieved by the parallel

combination of an IGBT and a MOSFET. It is also known as the “Mixed Parallel Operation” [2]. Here IGBT reduces the turn on loss of MOSFET and MOSFET reduce the turn-off loss of IGBT. The efficiency of the new device is highly comparable with the typical boost converter.[3]

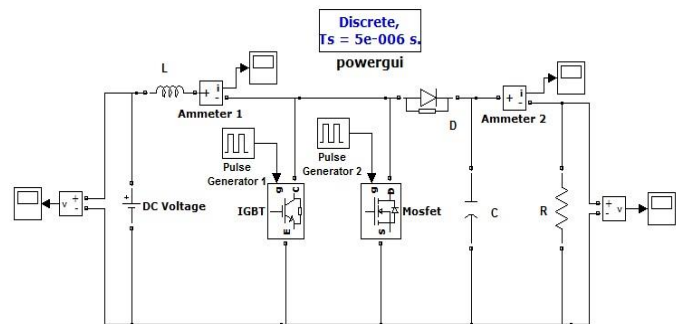


Fig 3: Circuit diagram of parallel IGBT-MOSFET switch boost converter.

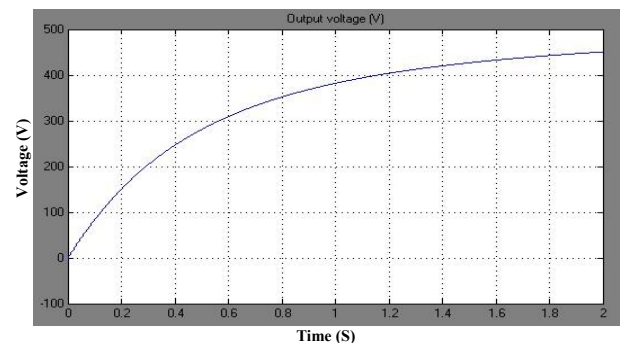


Fig 4: Output Voltage Curve.

Figure 4 shows the output voltage curve of the projected boost converter for 48 volt input. The peak output voltage is 450 volt. The curve is shown for 2s. In 2s the curve goes into the stable position and at that time the voltage is 450 volt. The proposed boost converter is taken 2s for going its stable condition.

TABLE I
ASSESSMENT BETWEEN TYPICAL AND PARALLEL IGBT-MOSFET SWITCH (PROPOSED) BOOST CONVERTER

Boost Converter	Input Voltage (V)	Input Current (A)	Output Voltage (V)	Output Current (A)	Efficiency (%)
Conventional	48	139.6	333	11.1	55%
	24	69.71	166.3	5.543	55%
Proposed	48	113.5	450.7	9.014	74%
	24	56.08	222.7	4.453	74%

From the above table it is seen that, for both 24V and 48V input voltage the conventional boost converter efficiency is 55 % and for proposed boost converter the efficiency is 74%. That means higher efficiency is obtained for the boost converter by mixed parallel operation then the typical boost converter.

III. MODELING OF A BI-DIRECTIONAL DC-DC CONVERTER

A BDDDC is a dc to dc converter where the power can flows in both the directions as supply end to load end and also load end to supply end. It has two modes of operation, buck mode & boost mode [4].

A. Conventional BDDDC

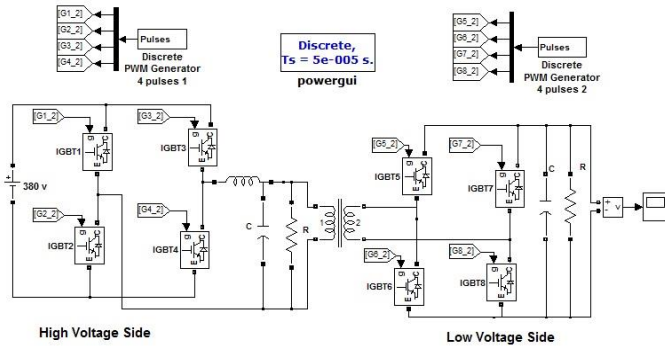


Fig 5: Conventional Bi-Directional DC-DC Converter.

Figure 5 represents a typical/conventional BDDDC. Basically there are two parts of the conventional BDDDC. One is high power side and another is low power side. A transformer between high power side and low power side which is transform power from high power to low power side and vice versa. In buck operation receiving end voltage became lower than the supply end voltage. And in boost operation receiving end voltage become higher than supply end voltage. The triggering pulse of IGBTs is supplied from two different PWM generators. The output voltage curve is shown below for both boost mode and buck mode.

1) Boost Mode:

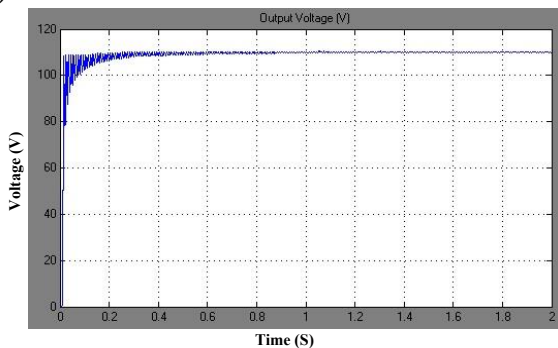


Fig 6: Output Voltage Curve for Boost mode.

Figure 6 shows the the Output waveform of conventional BDDDC (boost mode) where the output voltage is 110 Volt DC for 48 volts input. The output voltage curve is shown for 2s. From starting to 1s some noise are added with the curve. It is able to update the output curve with less noise by using more stage RC filter.

2) Buck Mode:

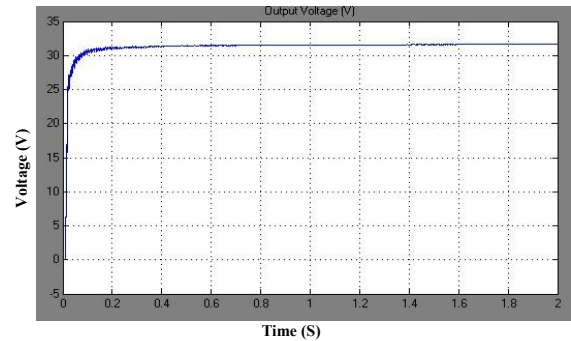


Fig 7: Output Voltage Curve for Buck mode.

Figure 7 shows the Output waveform of conventional BDDDC (buck mode) where the peak voltage is 32 volt DC for 380 volts input. The curve is shown for 2s. Before going to the 1s the curve is gone in stable condition.

B. PI Controlled BDDDC

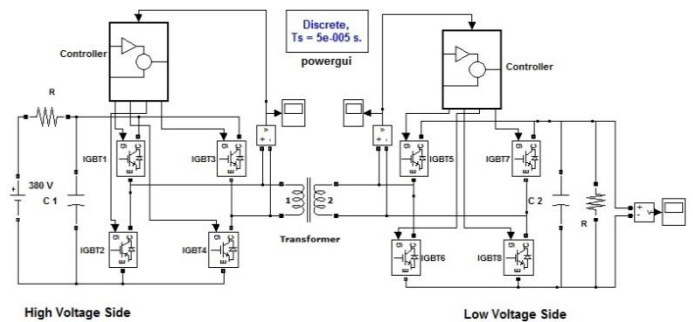


Fig 8: PI Controlled Bi-Directional DC-DC Converter.

Figure 8 represents the BDDDC controlled by PI controller. In this converter there is also two parts, high power (HP) side and low power (LP) side. Transformer between the HP and LP side transform the power from HP side to LP side and vice versa. Controlled pulses from PI controller are supplied to the IGBT instead of PWM generator used in conventional one. The proposed converter is also worked in two mode, boost mode and buck mode. In buck operation receiving end voltage became lower than the supply end voltage. And in boost operation receiving end voltage become higher than supply end voltage.

1) PI Controller:

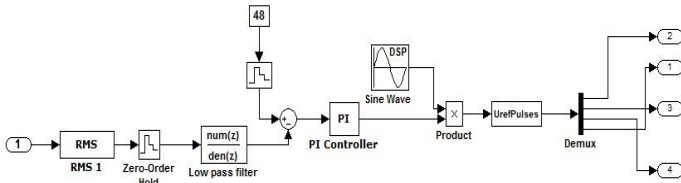


Fig 9: PI Control Circuit.

For increasing the working efficiency PI controller is used. The PI controller is used to provide the pulse to the IGBT's. Firstly PI controller takes the value of output voltage. Secondly compare it with the required output voltage. Thirdly Set the value and provides trigger signal to obtain a fixed value of output voltage through the input voltage of the converter is changed.

2) Boost Mode:

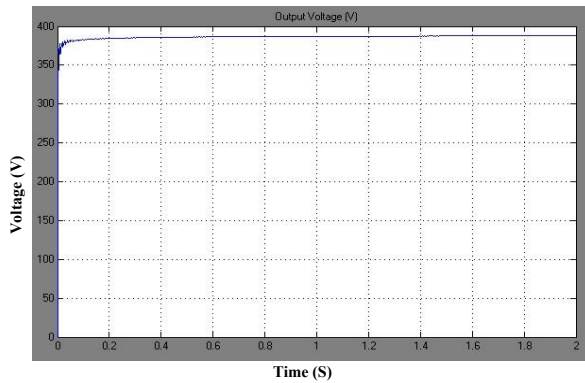


Fig 10: Output Voltage Curve for Boost mode.

Figure 10 shows the Output voltage waveform of PI controlled BDDDC (boost mode) where the output voltage is 380 Volt DC for 48 volts input. The curve is shown for 2s. The curve has gone to the stable condition very quickly and less noise is visualized in the curve.

3) Buck Mode:

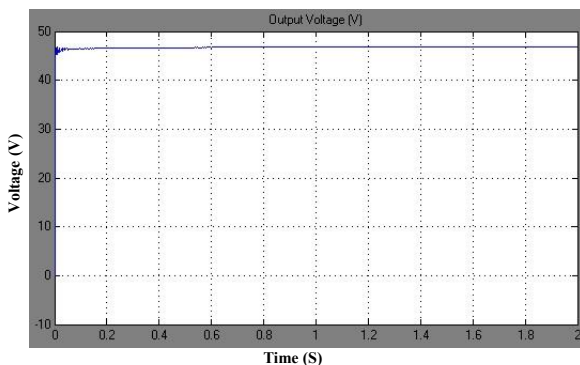


Fig 11: Output Voltage Curve for Buck mode.

Figure 11 shows the Output waveform of PI controlled BDDDC (buck mode) where the output voltage is 48 Volt DC for 380volts input. The curve is shown for 2s. It is also seen that the output curve is more stable than the conventional BDDDC and less noisy voltage curve is also observed.

C. PID Controlled BDDDC

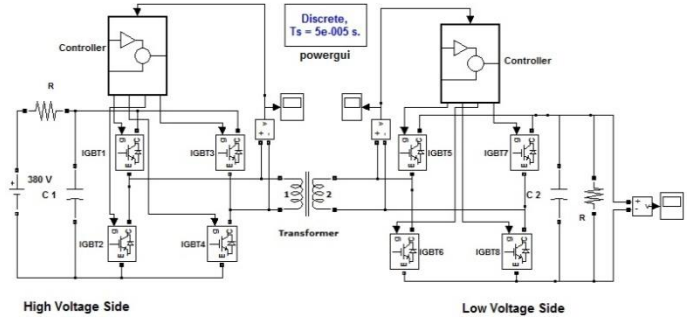


Fig 12: PID Controlled Bi-Directional DC-DC Converter.

Figure 12 represents the PID controlled BDDDC. In this converter there is also two parts, HV side and LV side. Transformer between the HV and LV side transform the power from HV side to LV side and vice versa. Controlled pulses from PID controller are supplied to the IGBT instead of PI controller and PWM generator. The proposed converter is also worked in two mode, boost mode and buck mode. In buck operation receiving end voltage became lower than the supply end voltage. And in boost operation receiving end voltage become higher than supply end voltage.

1) Boost Mode:

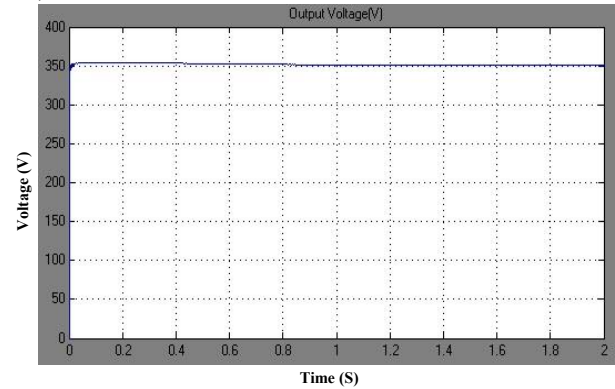


Fig 13: Output Voltage Curve for Boost mode.

Figure 13 shows the output curve of boost condition for 48 volt input. The Output voltage is 350 volt. The curve is shown for 2s. The curve has gone to the stable condition very quickly but output voltage is less than the PI controlled BDDDC.

2) Buck Mode:

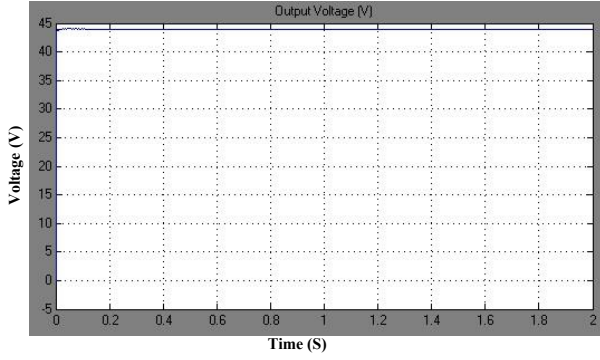


Fig 14: Output Voltage Curve for Buck mode.

Figure 3.15.4 shows the output curve of buck operation, for 380 volt input. The output voltage is 44 volt. The curve is shown for 2s. Initially the curve is more trembling.

IV. ASSESSMENT BETWEEN CONVENTIONAL, PI AND PID CONTROLLED BI-DIRECTIONAL DC-DC CONVERTER

TABLE II

ASSESSMENT BETWEEN CONVENTIONAL, PI CONTROLLED AND PID CONTROLLED BDDDC (BOOST MODE) FOR DIFFERENT INPUT VOLTAGES

Input (V)	Bi- Directional DC-DC Converter		
	Conventional Output (V)	PI Controlled Output (V)	PID Controlled Output (V)
45	105	379	340
48	110	380	351
50	118	381.5	360

TABLE III

ASSESSMENT BETWEEN CONVENTIONAL, PI CONTROLLED AND PID CONTROLLED BDDDC (BUCK MODE) FOR DIFFERENT INPUT VOLTAGES

Input (V)	Bi- Directional DC-DC Converter		
	Conventional Output (V)	PI Controlled Output (V)	PID Controlled Output (V)
375	30	46	43
380	32	47.5	44
385	33.5	48.5	46

From the above tables, for the different input voltages less variation in output voltage is found in case of PI controlled BDDDC. Which output voltages are effectively comparable with the output voltage of PID controlled BDDDC. And also the output voltages are higher in case of PI controlled bi-directional DC-DC converter compare to other two.

V. APPLICATION

Now this parallel IGBT-MOSFET switch boost converter and the proposed PI controlled BDDDC are used for the electric vehicle. Basically three different modes namely: Normal Mode, Acceleration Mode and Regenerative Mode. The power flows from battery to motor during acceleration and normal mode but during regenerative mode the kinetic energy of the motor converted into electrical energy and charged the battery. [5]

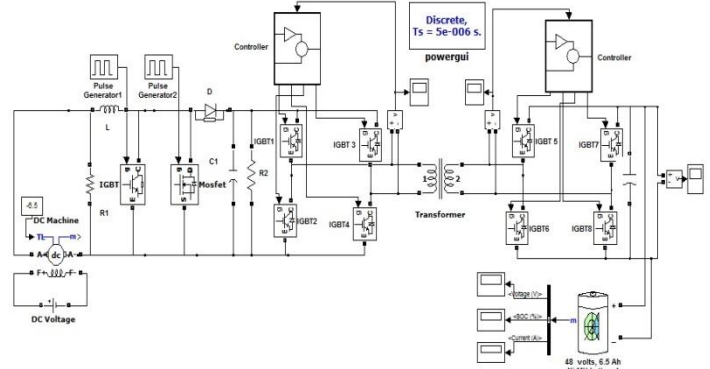


Fig 15: Regenerative Mode of Electric Vehicle.

Figure 15 represents the regenerative mode of electric vehicle where both proposed boost converter and PI controlled BDDDC is used. Firstly here DC machine is used as a DC generator (in the time of regenerative mode). But it creates a low voltage. In the second stage this low voltage are boosted by the proposed boost converter. Then in the next stage the boosted power are converted to a nominal level (buck mode) by which voltage the battery will be charged.

1) Battery Discharging Mode or Normal Mode

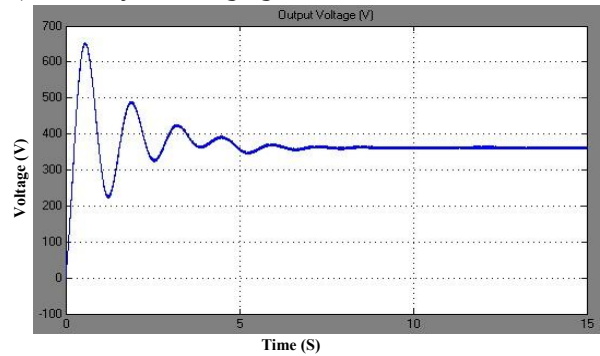


Fig 16: Output Voltage Curve of Normal Operation.

Figure 16 shows the Output waveform during motor is on load where the motor is running for 48 volts input. The battery supply 48 V then the BDDDC in boost mode which is boosted the low voltage to the higher voltage. Again the boost converter

enhanced the output of BDDDC voltage to a great extent. Then the motor will run.

2) Battery Charging Mode or Regenerative Mode

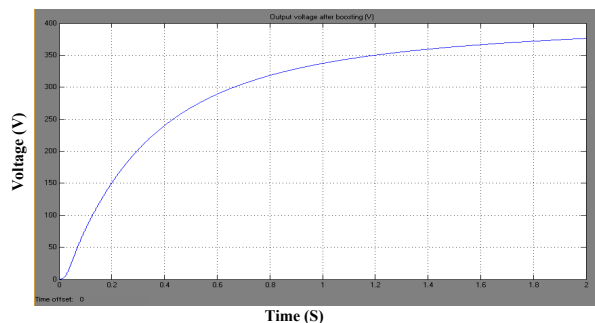


Figure 17: Output Voltage Curve of Boost Converter.

Figure 17 shows the Output voltage curve of boost converter. It steps up the voltage and this voltage generated by the regenerative action of motor during regenerative braking. The peak input voltage is 380 volt DC.

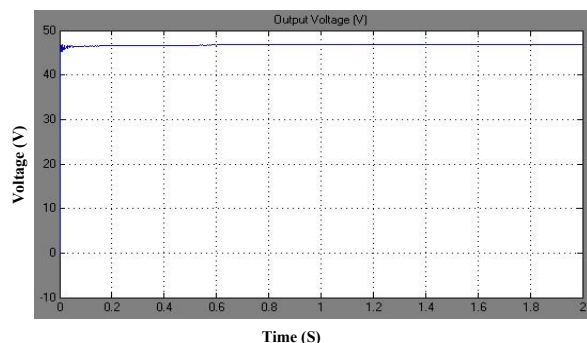


Fig 18: Output Voltage Curve of Regenerative Operation

Figure 18 shows the Output voltage curve of BDDDC for regenerative mode where the output voltage is 48 Volt DC. This output shows for 2s. This voltage further recharges the battery. This voltage supply 48V stable voltage.

IV. CONCLUSION

The main treats of environment and human life is pollution and a large extent of pollution is done by the use of huge number of automobile in the world. Especially global warming and the rapid depletion of the earth's petroleum resources are in extreme level. In the near future conventional vehicle will be replaced by electric vehicle. So that we aims to design and efficient PI controlled BDDDC and high efficient boost converter for Electric Vehicle. There are some reasons for selecting the PI controlled BDDDC. Firstly, less noisy output curve are found in PI controlled BDDDC rather than conventional and PID controlled converter. Secondly PI BDDDC provides more high voltage than conventional and PID controlled BDDDC.

ACKNOWLEDGMENT

Firstly, we express our gratefulness to almighty God as we have completed our dissertation successfully by His mercy. we also would like to express our endless gratitude to our supervisor Engr. Md. Mizanur Rahman for his enormous support and careful guidance.

REFERENCES

- [1] Chan, C. C., Alain Bouscayrol, and Keyu Chen, "Electric, hybrid, and fuel-cell vehicles: Architectures and modeling," IEEE Transactions on Vehicular Technology, 59, no. 2 (2010): 589-598.
- [2] Heeral P. ShethAnkit M. Adesara, "Paralleling of IGBT and MOSFET for High Power Applications," ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 2, Issue 12, December 2012.
- [3] Klaus F. Hoffmann, Jens Peter Kiirst, "High Frequency Power Switch-Improved Performance by MOSFETs and IGBTs Connected in Parallel". Annual IEEEConference and Exposition on Applied Power Electronics, pp. 1079-1084, ISBN 978-1-4244-1873-2 Austin, Texas, USA, February 24-28, 2011.
- [4] Zhang, Junhong. "Bidirectional DC-DC Power Converter Design Optimization, modeling and Control." Diss. Virginia Polytechnic Institute and State University," 2008.J.-G. Lu, "Title of paper with only the first word capitalized," J. Name Stand. Abbrev., in press.
- [5] Emadi, Ali, Young Joo Lee, and KaushikRajashekara, "Power electronics and motor drives in electric, hybrid electric, and plug-in hybrid electric vehicles," IEEE Transactions on Industrial Electronics, 55, no. 6 (2008): 2237-2245.

Efficient and fast convergent consensus algorithms for faulty nodes tracking in distributed wireless sensor networks

Rajkin Hossain
Computer Science and
Engineering Department
Brac University
Dhaka, Bangladesh
Email: rajkin69@gmail.com

Dr. Md. Muhidul Islam Khan
Computer Science and
Engineering Department
Brac University
Dhaka, Bangladesh
Email: muhidul.khan@bracu.ac.bd

Abstract—One of the challenging issues in a distributed computing system is to reach on a decision with the presence of so many faulty nodes. These faulty nodes may update the wrong information, provide misleading results and may be nodes with the depleted battery power. Consensus algorithms help to reach on a decision even with the faulty nodes. Every correct node decides some values by a consensus algorithm. If all correct nodes propose the same value then all the nodes decide on that. Every correct nodes must agree on the same value. Faulty nodes do not reach on the decision that correct nodes agreed on. Binary consensus algorithm and average consensus algorithm are the most widely used consensus algorithm in a distributed system. We apply binary consensus and average consensus algorithm in a distributed sensor network with the presence of some faulty nodes. We evaluate these algorithms for better convergence rate and error rate.

Index Terms- Wireless sensor networks, Consensus algorithm, Distributed systems, Convergence Rate, Faulty Node Tracking, Binary consensus, Average consensus.

I. INTRODUCTION

The miniaturization of electronic devices provide the impact on the development of wireless sensor network (WSN). WSNs consist of so many tiny sensor nodes capable of sensing, transmitting and receiving signals. Distributed sensor nodes are deployed in an ad hoc manner and responsible for various applications such as monitoring, surveillance, security and health care and load balancing [1,7] Nodes need to perform some tasks at each time step and report a central base station. For a large distributed sensor networks, it is difficult to deploy so many central base stations in the surveillance area. But with the central base stations, task management becomes centralized. We apply binary and average consensus algorithms in a distributed sensor network. No study has analyzed the convergence rate for an irregular network topology with the faulty nodes. Hence, in this paper, we investigate the proposed algorithm in irregular network topology, such as C-shaped, I-shaped and O-Shaped topology with faulty nodes. An irregular sensor model is introduced to evaluate the robustness of the algorithm. We apply segment tree data structures to reach consensus decisions at faster time. Actually we use our own

system model to apply binary consensus and average consensus algorithm in a distributed network for reaching a consensus decision. Our algorithm runs for a huge network. Initially we don't have any idea of the whole network besides there can be many faulty nodes. Our main goal is to reach consensus decision as fast as possible. In few cases, random chosen nodes will run faster but we are ignoring random choices because random choices [8] might get huge amount of time to reach a consensus decisions.

The rest of this paper is organized as follows. Section II discusses related work. Section III explains our network model. In Section IV we present background and our proposed method. Section V is our proposed algorithms and Section VI discusses simulation results and analysis for distributed sensor networks. Section VII concludes this paper with a brief summary.

II. RELATED WORKS

The widely used consensus algorithm to reach a decision in a distributed environment is gossip algorithm [4]. In gossip algorithm, pair of nodes are chosen randomly to exchange information and update their values. It is easy to implement. But the random information exchange creates more overhead in the network and it takes more time to reach consensus. Connectivity between nodes also affect the consensus value for the gossip algorithm. Another widely used consensus algorithm is binary majority consensus [6]. This algorithm terminates after a particular amount of time even the consensus is reached or not. Time limitation leads to lower efficiency and higher sensitivity to disturbances. The average consensus algorithm [5] helps to reach the consensus by updating their local values using average values of their neighbor's values. This consensus algorithm does not provide guaranteed consensus for the additive noise in the network.

III. NETWORK MODEL

The considered network model is undirected. The network model may have some faulty nodes but we don't have any

IV. BACKGROUND AND PROPOSED METHODS

A. Segment Tree

In Computer Science, a segment tree is a tree data structure for storing intervals, or segments. The nodes of a segment tree take a given priority based value from a specific range. Segment tree has a single root/top node and every node of a segment tree except leaves has 2 children which are left and right. Left child's id is $TreeNodeID*2$ and right child's id is $TreeNodeID*2+1$. As segment tree works in recursive process so every node except leaves go to its both left and right child recursively. It stops calling recursively when it reached leaf nodes. Generally a segment tree has 3 methods which are build, query and update. The depth of a segment tree is $\log(\text{no. of nodes})$. The general 3 methods of a segment tree are described below.

a) *Build*: We can build a segment tree from the given set of elements in an array. While building it will start from top node and go down in top-down processing. So from top node it goes to the leaf nodes saves the corresponding array value. The leaf nodes actually take same range value of the array index. For example: leaf node 7 which take the value of ranges [4-4] saves edge array value 4 in tree index 7. Then it goes to the parent node by backtracking and update its value depending on its child's by given priority. That's how Segment tree build process works. The complexity of build process is $O(\text{nodes} * \log(\text{nodes}))$ as we updated every leaf nodes individually.

b) *Query*: It finds out the priority-based value by given range. Its complexity is $O(\log(\text{nodes}))$ as trees maximum height is $\log(\text{nodes})$.

c) *Update*: While backtracking it updates priority-based tree nodes value where it visits. Its complexity is also $O(\log(\text{nodes}))$ as tree's maximum height is $\log(\text{nodes})$.

B. Creating Two Weighted Graphs

Now we need to create two weighted graph one for binary consensus and another for average consensus both are independent. Table 1 basically table based representation of both graphs where both are independent.

a) *Binary Consensus*: When we try to reach a consensus decision in a distributed network, every nodes of the network can hold initially one of the two values: zero and one. When two nodes communicate and run the updating protocols, they compare their current state and then each assume a new state based on what they have seen. When binary consensus algorithm is running, a node may be in one of the five states which can be described informally as:

0 - The node believes the majority opinion is most likely false. 1 - The node believes the majority opinion is most likely true. e0 - The node believes the majority opinion might be false. e1 - The node believes the majority opinion might be true. F - The node is a faulty node. Convergence occurs when all nodes have states 0,e0 or 1,e1.

Updating Protocol for Binary Consensus: For updating we

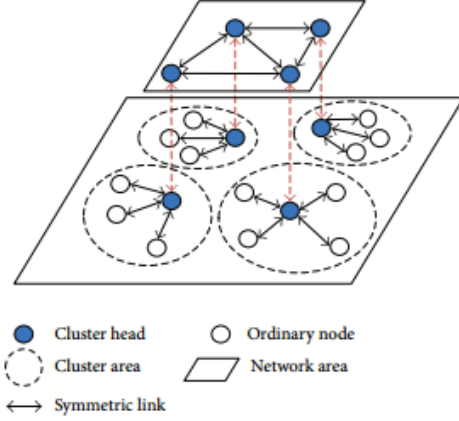


Fig. 1. Sample Network Model

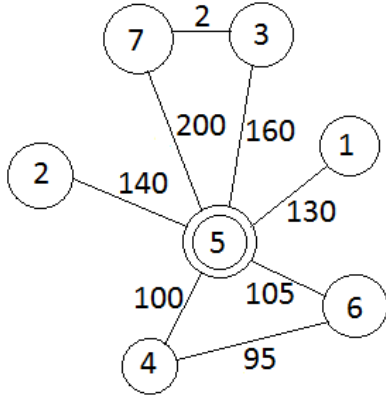


Fig. 2. Sample Network Model in Graph Representation

idea about that at first. We consider a clustered network with a cluster head for each cluster. Cluster heads can communicate with all the nodes in their respective clusters. Cluster head of a particular cluster can communicate with the cluster heads in the other clusters. Cluster heads also work as normal nodes with an additional characteristic is that they can be faulty. We assume the connection between the cluster heads form a complete graph and the graph is completely connected. Each node of the sensor network hold 3 classes of values which are energy, binary states and average states. As we are working on both binary and average consensus we need to use it. When a node's energy becomes zero or less than it becomes a faulty node and can not communicate throughout the network. Figure 1 defines a sample of network model and Figure 2 defines a sample network model in graph representation. Here edge defines physical distances and double circles define cluster heads.

have to follow updating protocols [3]. The protocols are given below with an example:

- 1) $(0,1) \rightarrow (e1,e0)$
- 2) $(e0,1) \rightarrow (1,e1)$
- 3) $(0,e1) \rightarrow (e0,0)$
- 4) $(e0,e1) \rightarrow (e1,e0)$
- 5) $(0,e0) \rightarrow (e0,0)$
- 6) $(e1,1) \rightarrow (1,e1)$
- 7) $(s,s) \rightarrow (s,s)$, for $s = 0,1,e0,e1$
- 8) $(s,F) \rightarrow (s,F)$ here F indicates faulty node

We introduce a new protocol here which is $(s,F) \rightarrow (s,F)$ as mentioned at number 8 above. We must give priority to these protocols. We assign priority based on decreasing order serial from the updating protocols. We propose serial no.1 as highest priority.

b) *Average Consensus*: For average consensus every states initially take some value and here energy plays very important role. While pair wise updating in average consensus the new states value for both pair is its average value. Besides we propose its weight which is below:

$A = \text{Energy}(u) - \text{distance}$ and $B = \text{Energy}(v) - \text{distance}$
 $\text{weight}(u,v) = A + B$ if A and $B \geq 0$
otherwise $\text{weight}(u,v) = 0$

Updating Protocol for Average Consensus: There is only one updating protocol for average consensus which is simple its average by consideration of number of iterations which is given below:

$$x_i(k) = x_j(k) = [x_i(k-1) + x_j(k-1)]/2 \quad (1)$$

TABLE I
TABLE REPRESENTATION FOR BOTH GRAPHS

Sensor Id	Binary States	Average States	Energy
1	Faulty	Faulty	0
2	e1	30	250
3	e0	65	200
4	0	45	120
5	e1	65	500
6	1	56	129
7	1	45	300

C. Edge Array and Edge Graph

We created two edge array for both binary and average consensus and a edge graph for tracking a sensor node's neighbor edges.

V. PROPOSED ALGORITHMS

Here we applied our proposed algorithm for consensus in two different independent segment trees. We first build two segment tree from corresponding two edge array and by the help of edge graph we track down both tree leaf index id in STreeNode, S2TreeNode from edge arrays. We did it for future bottom up processing and avoid unnecessary recursion.

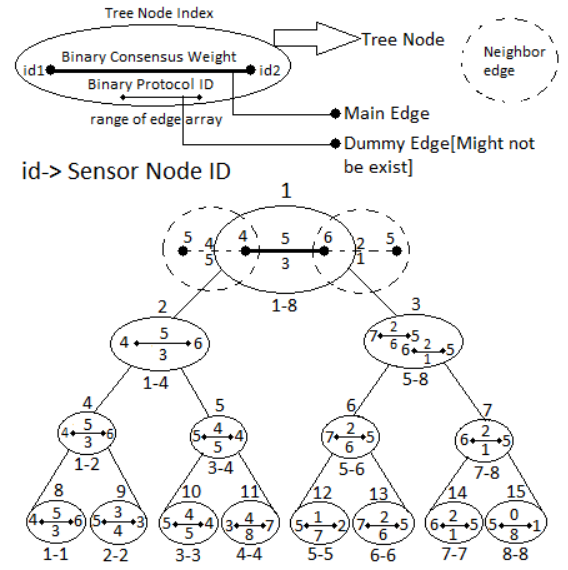


Fig. 3. Segment Tree for Binary Consensus

A. Segment Tree for Binary Consensus

Now we have to make a segment tree for the whole part of the edge array. Here in every node of the segment tree can take 4 elements which are main-Edge, dummy-Edge, weight and size. While considering the left child and right child for the upper nodes we will choose the highest priority based edge as main-Edge and dummy-Edge will be NULL but here if highest priority based edge exist more than one and all of them follow protocol id [1 and 6 both] then we should take two edges one for main-Edge and another for dummy-Edge. Besides if all of their weights follow same protocol (1 or 6 not both) then we save any single edge to tree nodes main-Edge and dummy-Edge will be NULL.

We consider every time top node from the segment tree and need to change its two endpoints value and update its information according to the updating protocols. Here we might have to change many edges information's according to updating protocols. Because if the edges end points connected to another edges then we should change their information also. By using edge graph we can track down its two sided neighbors and update their information and by bottom up processing until we reach the root we can re-settle every nodes information. While updating we must also minimize energy from that edge two pointer node. If segment trees topNode contains faulty nodes of the network so it follows lowest priority based updating protocols. So consensus will never occur we can claim that easily. Besides Binary consensus has zero percent error rate. Figure 3 defines binary consensus algorithm simulation for a single iteration and Algorithm 1,2 and 3 is for binary consensus. In algorithms mBX means mainEdges binary state X, BPID means binary protocol ID.

Algorithm 1 Binary Consensus Algorithm

```
1: Binary Consensus Algorithm ()
2: Build Segment Tree for Binary Consensus
3: Energy of all sensor nodes stores in Energy Array(Energy)
4: topnode = 1
5: while True do
6:   edge = STree[topnode].mainEdge
7:   if STree[topnode].dummyEdge == NULL and
      OneOrSix(STree[topnode].BPID) then
8:     Binary Consensus Reached..break
9:   else if STree[topnode].Bweight == 1 then
10:    Binary Consensus Reached.. break
11:  else if edge.Bweight == 0 then
12:    Binary Consensus Will Never Reached..break
13:  else
14:    distance = physicalDistance(edge.u, edge.v)
15:    EnergyU = Energy[edge.u] - distance
16:    EnergyV = Energy[edge.v] - distance
17:    if EnergyU < 0 or EnergyV < 0 then
18:      Make just STree[STreeNode[edge.id]] as a
      faulty edge by updating.
19:      UPDATE(STreeNode[edge.id])
20:    else
21:      update both energy array to EnergyU and En-
      ergyV
22:      Change STreeNode[edge.id] mainEdge by fol-
      lowing binary updating protocols.
23:      UPDATE(STreeNode[edge.id])
24:      for Each edge  $e \in EG.Adj[edge.u]$  do
25:        treeID = STreeNode[e.id]
26:        if  $e.u == edge.u$  and  $e.v != edge.v$  then
27:          STree[e.id].mBU = edge.BU
28:          Change STree[treeID] information of
          edge-weight and protocol id
29:          UPDATE(treeID)
30:        else if  $e.v == edge.u$  and  $e.u != edge.v$ 
then
31:          STree[e.id].mBV = edge.BU
32:          Change STree[treeID] information of
          edge-weight and protocol id
33:          UPDATE(treeID)
34:        end if
35:      end for
36:      for Each edge  $e \in EG.Adj[edge.v]$  do
37:        treeID = STreeNode[e.id]
38:        if  $e.u == edge.v$  and  $e.v != edge.u$  then
39:          STree[e.id].mBU = edge.BV
40:          Change STree[treeID] information of
          edge-weight and protocol id
41:          UPDATE(treeID)
42:        else if  $e.v == edge.v$  and  $e.u != edge.u$ 
then
43:          STree[e.id].mBV = edge.BV
44:          Change STree[treeID] information of
          edge-weight and protocol id
45:          UPDATE(treeID)
46:        end if
47:      end for
48:    end if
49:  end if
50: end while
```

Algorithm 2 Tracking Highest Priority Node(For Binary Consensus)

```
1: HIGHEST(SNode l, SNode r)
2: Initialize a SNode(Node) which keep a mainEdge, a
   dummyEdge and weight of them which are same
3: rBPID = r.mainEdge.BPID
4: lBPID = l.mainEdge.BPID
5: if r.Bweight == l.Bweight then
6:   if rBPID == 1 and lBPID == 6 then
7:     Node.mainEdge = r.mainEdge
8:     Node.dummyEdge = l.mainEdge
9:   else if lBPID == 1 and rBPID == 6 then
10:    Node.mainEdge = l.mainEdge
11:    Node.dummyEdge = r.mainEdge
12:   end if
13: else if r.Bweight > l.Bweight then
14:   Node=r
15: else Node=l
16: end if
17: return Node
```

Algorithm 3 Updating Segment Tree (For Binary Consensus)

```
1: UPDATE(node-id)
2: i = node-id
3: while ( $i = i >> 1$ )! = 0 do
4:   Left = i * 2
5:   Right = left + 1
6:   STree[i] = HIGHEST(STree[Left], STree[Right])
7: end while
```

B. Segment Tree for Average Consensus

Now we have to make another segment tree for the edge array. Here in every node of the segment tree can take only single elements which we name treeEdge. To build a segment tree as it goes from top to bottom so when we reach the leaf nodes of the segment tree we mapped the leaf node number to its corresponding edge to S2TreeNode-Array (have to do it for future purposes) and while considering the left child and right child for the upper nodes in segment tree we will choose the highest weight based edge for average consensus. This algorithm runs for edges and implemented in another Segment tree data structure. For every time we will consider the top node from the segment tree and as initially we have highest max weight edges based average consensus weight in the top node. When Top node Average consensus weight is not zero we will continue our average consensus algorithm otherwise we should stopped. Figure 4 defines average consensus algorithm simulation for a single iteration and Algorithm 4,5 and 6 is for average consensus. Average consensus must have some error rate. The formula for calculating error rate is given below:

$$\frac{||x(k) - \bar{x}||}{||x(0) - \bar{x}||} \quad (2)$$

We will calculate this error rate after every single iterations. If we reached in highest possible average consensus point we cant reduce error rate more. Few shortcuts tAX means treeEdges average state X.

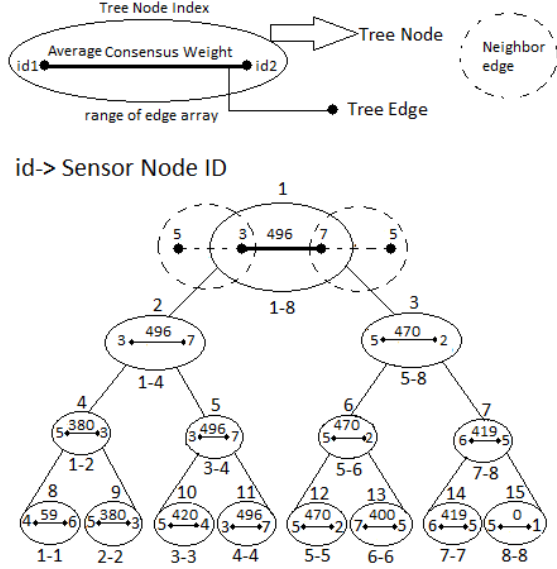


Fig. 4. Segment Tree for Average Consensus

Algorithm 4 Tracking Highest Priority Node(For Average Consensus)

```

1: HIGHEST2(S2Node l, S2Node r)
2: Initialize a S2Node (Node) which keep a treeNode
3: if (l.AstateU == l.AstateV) then Node=r
4: else if (r.AstateU == r.AstateV) or (l.Aweight >
   r.Aweight) then Node=l
5: else Node = r
6: end if
7: return Node

```

Algorithm 5 Updating Segment Tree(For Average Consensus)

```

1: UPDATE2(node-id,j)
2: i = node-id
3: if j == 1 then
4:   au = S2Tree[i].treeEdge.AU
5:   av = S2Tree[i].treeEdge.AV
6:   S2Tree[i].treeEdge.AU = (au + av)/2
7:   S2Tree[i].treeEdge.AV = (au + av)/2
8: end if
9: while (i = i >> 1) != 0 do
10:   Left = i * 2
11:   Right = left + 1
12:   treeLeft = S2Tree[Left]
13:   treeRight = S2Tree[Right]
14:   S2Tree[i] = HIGHEST2(treeLeft, treeRight)
15: end while

```

Algorithm 6 Average Consensus Algorithm

```

1: Average Consensus Algorithm ()
2: Build Segment Tree for Average Consensus
3: Energy of all sensor nodes stores in Energy Array(Energy)
4: topnode = 1
5: while True do
6:   edge = STree[topnode].treeEdge
7:   if edge.AU == edge.AV then
8:     Highest Average Consensus Reached..break
9:   else if edge.Aweight == 0 then
10:    Average Consensus Reached.. break
11:   else
12:     distance = physicalDistance(edge.u, edge.v)
13:     EnergyU = Energy[edge.u] - distance
14:     EnergyV = Energy[edge.v] - distance
15:     if EnergyU < 0 or EnergyV < 0 then
16:       Make just S2Tree[S2TreeNode[edge.id]] as a
       faulty edge by updating.
17:       UPDATE2(S2TreeNode[edge.id],0)
18:     else
19:       Update both energy array to EnergyU and
       EnergyV
20:       Change S2TreeNode[edge.id] treeEdge by fol-
       lowing average updating protocols.
21:       UPDATE2(S2TreeNode[edge.id],1)
22:       Change current selected sensor nodes LEFT-
       SIDED neighbor edges information by the help of Edge
       graph
23:       Update each neighbor edge ids tree edge aver-
       age U,V state to its neighbor edge U state
24:       UPDATE2(corresponding treeID,0)
25:       Change current selected sensor nodes RIGHT-
       SIDED neighbor edges information by the help of Edge
       graph
26:       Update each neighbor edge ids tree edge aver-
       age U,V state to its neighbor edge V state
27:       UPDATE2(corresponding treeID,0)
28:     end if
29:   end if
30: end while

```

VI. RESULTS AND ANALYSIS

We used Java programming language to simulate our results and for graphical representation of sensor nodes.

A. Random Topology

We also implemented our algorithm in random topology graph where sensor nodes are also selected randomly. Here we have selected 300-1800 sensor nodes. Table 2 represents for random topology.

B. Various Topology

We have implemented our algorithm in C[90 sensor nodes],O[130 sensor nodes],I[80 sensor nodes].We represent their iteration vs error-rate plots with their accurate graphical representations.

TABLE II
TABLE REPRESENTATION FOR RANDOM TOPOLOGY

Node	300	600	900	1200	1800
I(A)	5202	9015	17859	19363	33782
I(B)	756	1608	2105	3250	4846
E(A)	0.75	0.78	0.69	0.75	0.71
E(B)	0.00	0.00	0.00	0.00	0.00

C. Analysis

In various topology BCP means Binary consensus reaching point and HACP means Highest possible average consensus reaching point. It can be seen that BCP is always less than HACP as we try to reach highest point for average consensus but for binary consensus, reaching consensus point is actually highest point of it. Besides error rate after HACP point is fixed that defines we cant reach lesser error rate than HACP point for average consensus.

In random and various topology IA,IB means no. of iterations for average and binary consensus and EA,EB means error rate for average and binary consensus and error rate for binary consensus is always zero.

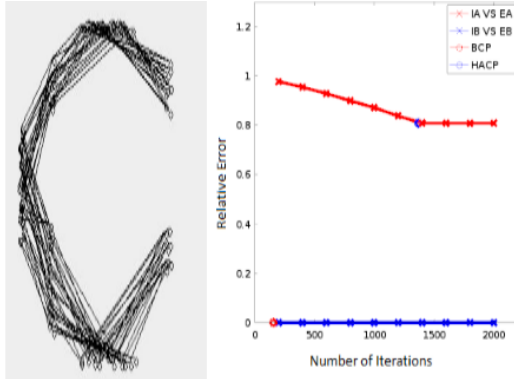


Fig. 5. C Topology

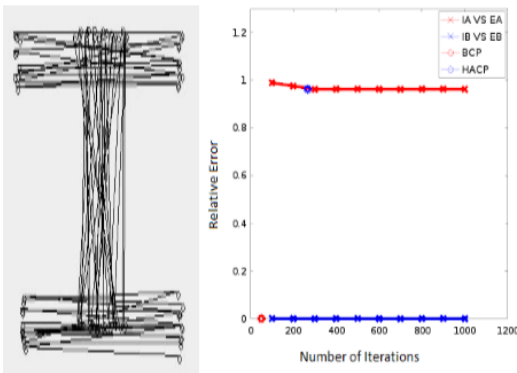


Fig. 6. I Topology

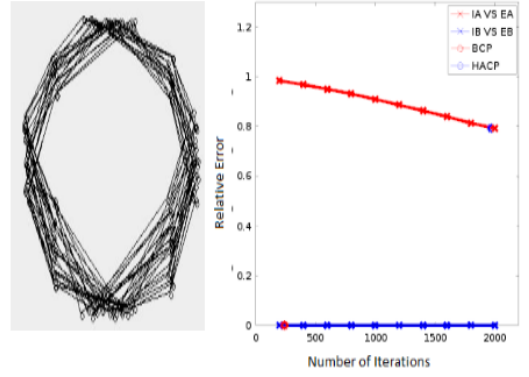


Fig. 7. O Topology

VII. CONCLUSION

In a distributed sensor network, where robustness is an important issue for finding out the faulty nodes efficiently. Consensus algorithms help to reach on a decision by which we are able to track the faulty nodes. Accuracy of knowing the right information about the faulty nodes is very challenging in a distributed environment. We consider different shapes of topology for applying the binary and average consensus algorithm. We also apply binary and average consensus algorithms in a random graph. We observe that average consensus algorithm provides faster convergence by giving less number of iterations. On the other hand, binary consensus is more accurate to find out the faulty nodes. So, using the appropriate consensus algorithm depends on the topology we are using and depends on the application.

In future, we will consider the robustness issue in details considering some noisy nodes in the environment which provide misleading value. We will choose a discrete event simulator for our work for making the simulation more cognitive.

REFERENCES

- [1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor network: a survey," Computer Network, 2002.
- [2] F. Benzet, A. G. Dimakis, P. Thiran, Vetterli, and M. Gossip, "Along the way: order-optimal consensus through randomized path averaging," in Proceeding of the Allerton Conference on Communication, Control, and Computing, Allerton, USA, September 2007.
- [3] M. Draief and M. Vojnovic, "Convergence speed of Binary interval consensus," in proceedings of annual joint conference of the IEEE computer and communications societies (INFOCOM 2010), San Diego California, 2010.
- [4] A. D. G. Dimakis, A. D. Sarwate, and M. J. Wainwright, "Geographic gossip: efficient averaging for sensor networks," IEEE Transactions on Signal Processing, 2008.
- [5] Yanwei Li, Zhenyu Zhou, and Takuro Sato, "A Cluster-Based Consensus Algorithm in a Wireless Sensor Network" International Journal of Distributed Sensor Networks Volume 2013, Article ID 547124.
- [6] Noor Al-Nakhala, Ryan Riley, and Tarek M. Elfouly, "Binary Consensus in Sensor Motes" Wireless Communications and Mobile Computing Conference (IWCMC), 2013 9th International.
- [7] D. Culler, D. Estrin, and M. Srivastava, "Overview of sensor networks" Computer, vol. 37, no. 8, 2004.
- [8] Alexander Gogolev, Lucio Marcenaro, "Randomized Binary Consensus with Faulty Agents" Entropy ISSN 1099-4300, 2014

Fabrication Of A Low Cost Microcontroller Based Spray Pyrolysis System For Thin Film Deposition

MD. Sifatul Muktadir

Department of Electrical, Electronic and Communication Engineering, Military Institute of Science and Technology
Mirpur Cantonment, Dhaka-1216, Bangladesh

Abstract— Spray pyrolysis technique confers a spontaneous way to deposit thin film on flexible substrates with minimum consumption of desired material simply by adding it with the spray solution. The constituent units of the simplified spray pyrolysis technique are spray-dry chamber, a temperature controlling unit, a spray-dry controlling unit and an exhaust gas emission unit. The aim of this work is to simplify the fabrication method of the system and minimize the cost. On this regard, a handmade wooden chamber, an electrical heater with temperature controller and exhaust fans are used in this system which are available in local market. A servo motor interfaced with a microcontroller and a LCD display is used for the spray-dry unit where the time is set by using a matrix keyboard. Eventually, Zinc Oxide thin film is obtained on a glass substrate by using this system and its optical property is observed in a Spectrophotometer after completing the fabrication of the system. An easy fabrication process with the user friendly equipment make this system simple, easy to handle and maintain with less mechanical hazards.

Keywords—spray pyrolysis; zinc oxide thin film; microcontroller; low cost

I. INTRODUCTION

The necessity of renewable energy source is increasing day by day as the conventional resources of energy like fossil fuel are facing number of challenges including high purchasing price, limited number of suppliers having adequate ability to provide and issues with environment pollution as well. However, Sun being the source of renewable energy is employed to power generation industries by means of photovoltaic cell which converts photon energy into electrical energy. For the production of the photovoltaic cell Silicon is the most favorite choice though it has high production cost. Thin film materials which have direct band gap termed as semiconductors are produced to overcome the limitations of silicon as they have low production cost [1]. Because of this property, their light absorption co-efficient is very high and thinner film can be deposited [1].

There are various ways applying which thin films can be deposited on a substrate. Among these processes spray pyrolysis is the most convenient way for depositing thin films in a large scale in normal open air condition without any complicated vacuum. [2, 3]. Besides it provides a very high quality thin film and the quality of the film can be controlled by the process parameters [4]. Spray Pyrolysis has a great scope of industrial implementation as it can produce a large number of thin film in a simple way [1]. However for the industrialization of spray pyrolysis and to make it user friendly its construction, operation and cost need to be optimum without degrading the quality of the thin film. In the proposed model this things are optimized by the use of microcontroller and

different market available materials but the quality of the thin film is kept to its standard value.

Metal oxide thin films are becoming the point of interest day by day as they are used in different industrial products such as solar cells, gas sensors, LEDs, LCDs, heat mirrors etc. ZnO is one of them which is increasingly used in different industrial aspects as it has a great semiconductor property with a direct band gap of 3.2-3.37 eV with a high binding energy of 60 eV [4]. ZnO is chemically and mechanically stable, low cost, abundant in nature and nontoxic. ZnO thin films can be achieved by any process that fulfils the conditions to produce high quality thin films among these techniques CVD, RF sputtering, sol gel, hydrothermal, pulsed laser deposition, spray pyrolysis are noteworthy.

A good quality thin film depends upon the conditions upon which it is deposited. In case of spray pyrolysis one of the most important parameters is the substrate surface temperature. In this process by controlling the spray and dry time the surface temperature can be controlled and so as the properties of the thin film. A heater with a temperature controller and a servomotor driven spray-dry controlling system is developed in this setup. For controlling the servo motor microcontroller unit has been included where the spray-dry can be set up by simply using the matrix keyboard. Besides there is a LCD keyboard for the user to see the set value and a spray gun is used for spraying the precursor to the substrate. The whole set-up is covered by a wooden box and there is no mechanical arrangement attached with this. Using this system ZnO thin film is deposited in this work and its optical property and indirect band gap was observed as well.

II. SPRAY PYROLYSIS METHOD

Spray pyrolysis is a process in which a thin film is deposited by spraying a solution on a heated surface, where the constituent reacts to form a chemical compound. The chemical reactants are selected such that the desired compound are volatile at the temperature of deposition. The coating is applied at elevated temperatures by spraying droplets of liquid precursor onto hot substrates. Fig. 1 demonstrates the schematic diagram of spray pyrolysis technique having three major controlling units such that: temperature controlling unit, spray-dry time controlling unit and the pressure controlling unit.

All three units are confined in a wooden box which have exhaust fans for taking out the exhaust gases from the experimental enclosure. The process of spray pyrolysis technique can be divided into three parts and atomization is the first step of them. Depending on the liquid properties and

operating condition the atomizer is selected for the process of atomization. There are several types of atomizers are available but for the ease of operation air blast atomizer is used for the system. However this process has limitations in producing droplets of micrometer and in controlling the distributions of the droplets [4]. Ultrasonic sprayer is the remedy to it but keeping the cost in consideration air blast atomizer is used to make the system user friendly.

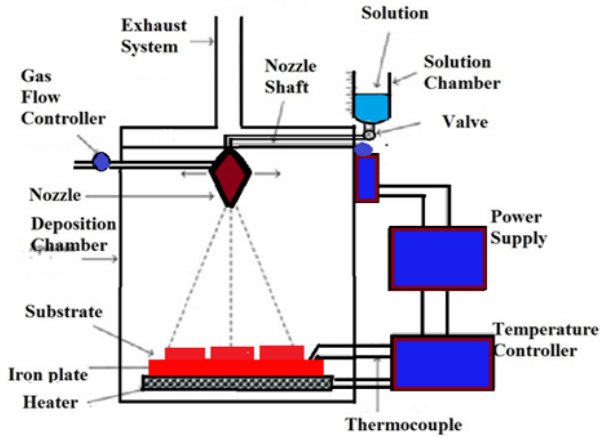


Fig. 1. Schematic diagram of Spray Pyrolysis system

The control over spray and dry timing is the next step. The amount of droplets that reach to the substrate depends on the spray time if the spray time is much then the amount of droplet is increased. Besides the dry time fixes the vaporization time for the solvent to vaporize. If the spray time is larger, then more droplets come before the vaporization. For the precise control over the timings, microcontroller unit as well as servo motors are used. The third and the last part of the process is the temperature control. A normal 1000 watts heater coil coated with asbestos layer are the constituents for the heater. When the droplets tend to touch the heated plated on the heater the solvent evaporates and the remaining chemical compound drops on the surface of the heater.

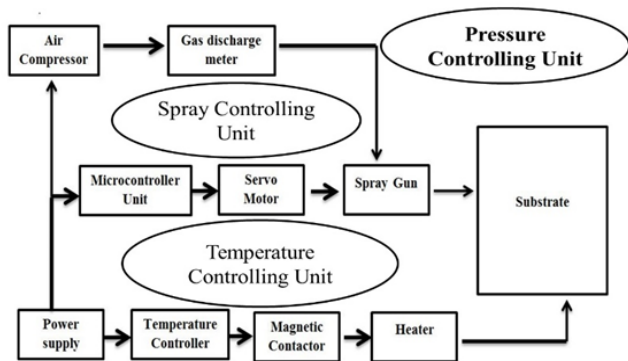


Fig. 2. Block diagram of Spray Pyrolysis system

From Fig. 2 it can be noticed that Power supply unit initiates power to the air compressor and the air pressure is controlled by the Gas discharge meter. The pressurized air is then supplied to the spray gun. The microcontroller unit is used to control the servo motor which pulls the trigger of the spray gun at the given instruction from the matrix keypad. In

fine, Substrate is heated by a heater whose temperature is controlled by the temperature controller which senses the temperature of the surface by a thermocouple. The Total system is contained in a wooden chamber where exhaust fan is used for bringing out the gasses produced during the process.

A. Pressure Controlling Unit

Pressure atomizer is used instead of ultrasonic atomizer to simplify the process [5]. The purpose of the pressure controlling unit is to supply pressurized air for the purpose of spraying the solution on the substrate surface. The pressure controlling unit consists of two components. Such as the air compressor and the gas discharge meter. The air compressor produces high pressured air and the gas discharge meter restrain the air pressure to accede the pressure limit. This compressed air is then supplied to the spray gun where it does the following works

- Splits the droplets to small grains
- Distributes the droplets to the surface of the substrate.
- Creates impact with the surface [3]

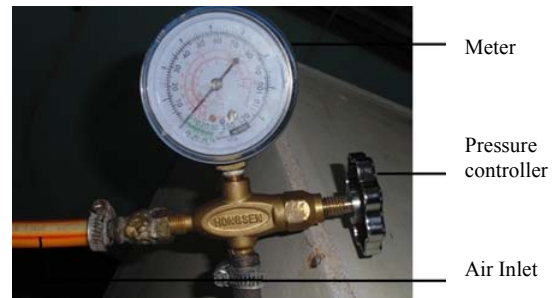


Fig. 3. Pressure Controlling Unit: Gas Discharge Meter

The air compressor stores and supplies compressed air and by the gas discharge meter depicted in Fig.3. the air pressure is controlled to a desired level.

B. Temperature Controlling Unit

Temperature is an important factor of spray pyrolysis system. It indeed needs a proper control of temperature because when the temperature is too low, the droplets does not get sufficient thermal energy for vaporization.

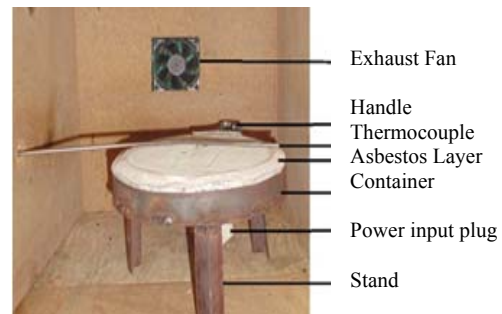


Fig. 4. Substrate Heater

Inversely if the temperature is too high, then because of sublimation at a sufficient distance which is away from the substrate condensation reaction takes place in the absence of a deposition surface, thereby producing oxide particles that form away from the substrate. Temperature being one of the key facts for the better control of temperature this unit comprises with a heater, thermocouple and temperature controller. At Fig.4. the temperature controlling unit is shown. The heater is made of Tungsten coil of 1000W and covered with an asbestos layer. Moreover thermocouple senses the heat of the surface and informs temperature controller where the temperature is set. If the temperature of the heater accedes the set value of temperature controller magnetic switch turns off the power supply to the heater.

C. Spray-Dry Controlling Unit

Spray-dry control unit acts as the brain of the total system because function of all units are controlled from this switch box. This unit comprises with a switch box, a servo motor and a spray gun. Fig.5. indicates a plastic box upon which the keypad and LCD display is attached.



Fig. 5. Spray-dry control unit box

Fig. 6 indicates the PCB circuit with the ATmega32 microcontroller which gets input from a 4x4 matrix keypad attached with the switch board. The spraying and drying time is set from the keyboard and in the LCD display spraying and drying time as well as the repetition time is displayed. The shaft of the servo motor undergoes angular displacement of 90° to a certain angle and remains fix at that position until the set spray time gets over.

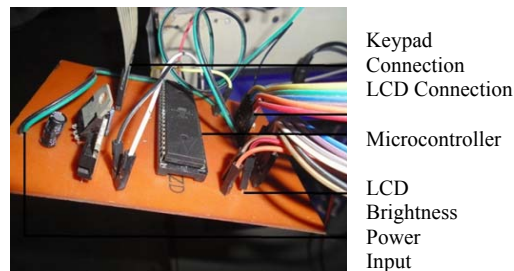


Fig. 6. Internal Circuit Diagram of Spray-dry Unit

During this angular displacement the shaft of the motor pulls the trigger of the spray gun. When the trigger of the sprayer is pressed liquid precursor mixes with the pressurized air coming from the air compressor and sprayed on heated

substrate placed on the surface of the heater. However, when the spray time is over the shaft comes to its 0° position and leaves the trigger of spray gun. With this the dry time starts and the calcining process starts up to the time set by the user and continues up to repetition time.

D. Complete Setup

Units discussed above are contained in a wooden box mounted in a steel stand whose dimension is 120×35 cm². Fig.7 indicates the complete set up of the system from the front side. Moreover it includes exhaust fans which elicit the chemical vapor during the process.

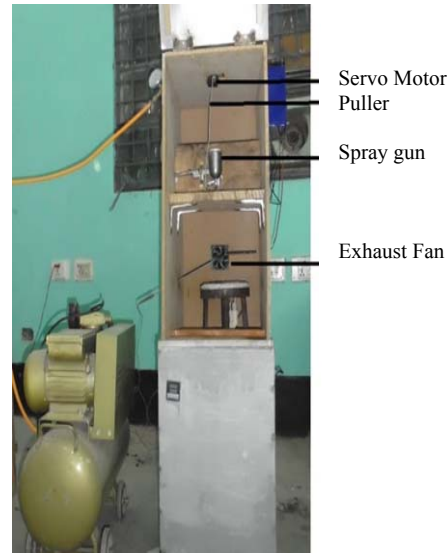


Fig. 7. Complete Setup

E. Preparation of Precursor Solution

In this work, ZnO films were deposited from zinc acetate & Ethanol solution at a concentration of 1.0 mol/L. At first, 5.488g of zinc acetate was dissolved in 25 ml of absolute ethanol to yield 1.0 mol/L concentration of precursor solution and then was magnetically stirred for 30 min. A milky solution was obtained where with a molar ratio of DEA/ZAD as 1:1; 2.61 ml of DEA was added into the emulsion. The solution was again stirred and a clear solution was obtained after a certain period.

III. RESULTS AND DISCUSSION

At the beginning temperature was set according to the following table then the pressure was set to 90mmHg. Afterwards the precursor solution was prepared according to the process discussed above the solution was sprayed on the glass substrate of dimension 2.5×1.5 cm² setting the spray time at 2s and dry time at 21s. The distance between spray nozzle and glass substrate was 14 inch.

TABLE I. CONDITION TABLE

Spray time	Dry time	Repetition	Substrate temperature	Room temperature	Distance	Solution
2s	21s	5	420°C	23°C	14inch	5ml

Fig. 8 depicts two glass substrates where the upper one is uncoated and the lower one is coated with ZnO thin film obtained using the system. The samples were taken to the

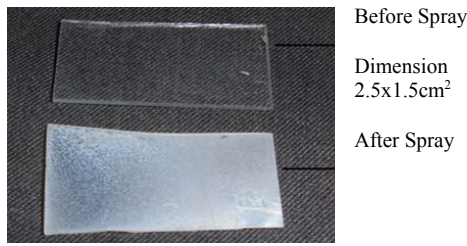


Fig. 8. ZnO thin film coated glass substrate

Spectrophotometer for testing their optical property In Fig. 9 the optical property of the thin film is graphically expressed in terms of transmittance vs wavelength

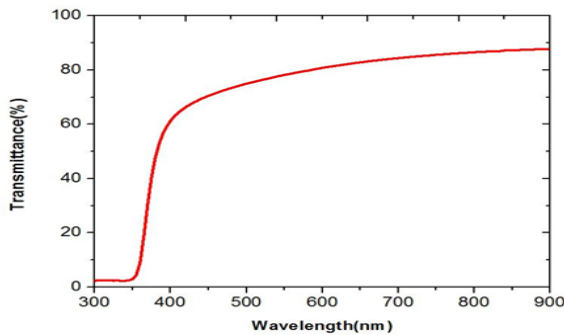


Fig. 9. Transmittance vs Wavelength curve of ZnO Thin Film

It can be seen that with the decreasing wavelength of light the transmittance is decreased. At 350nm wavelength the transmittance has suddenly fallen this means that in this range the absorption of photon energy is maximum. The visible spectra of light is in between 300nm to 400nm and if the absorption of the film lies between this range that means it can absorb solar light. Moreover in Fig.10. optical band gap

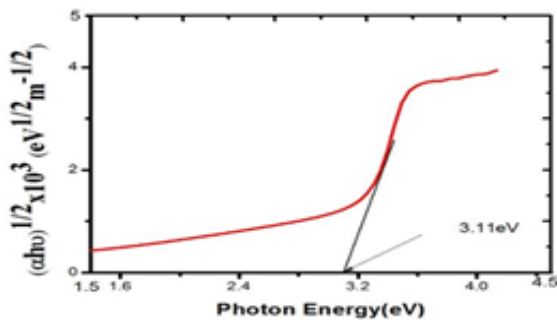


Fig. 10. Bandgap vs Photon Energy Curve

of the thin film is observed and the indirect band gap is 3.11eV as it is seen from the curve which is nearly equal of the indirect band gap of ZnO which is 3.1 eV[6].

v. COST ANALYSIS

One of the noteworthy aim of the work was to reduce the cost as well as to make the system user friendly. The used equipment for the system are easy available in the market and as there is no mechanical arrangements, the system is very user friendly, comfortable to use and easy to maintain. Present systems in the market are very costly which makes the production cost of the thin film much higher. However the proposed system provides lower production cost. The total cost of the system is 20000 BDT which is more than 50% reduced cost than present systems in the market. Nevertheless, this system lacks the computer interfacing in terms of controlling the spray time as well as temperature, horizontal and vertical movement of the spray gun and the quality of the film may differ but further study can remove these constraints without significant increase in the production cost.

IV. CONCLUSION

In conclusion, it can be said that the aim of the project was successfully done by fabricating a cost effective system furthermore depositing ZnO thin film by it. In the Nanotechnology and Photovoltaics laboratory, thin films of other materials like TiO_2 , CdS, Al doped ZnO are successfully deposited also. This works are kept in reserve for the further research work with the system. Moreover interfacing with the computer will increase its accuracy and control and will rectify its limitations. In near future the thin film produced by this system will be used for solar cell fabrication after further study of the properties.

ACKNOWLEDGMENT

This research work was conducted in the Nanotechnology & Photovoltaics Laboratory, RUET under the supervision of Prof Dr. Md. Faruk Hossain. I want to give thanks to him for his idea and proper guidance.

REFERENCES

- [1] B. Godbole, N.Badera, S.B. Shrivastav and V. Ganesan, "A simple chemical spray pyrolysis apparatus for thin film preparation", JI. of Instrum. Soc. of India Vol. 39 No. 1 March 2009
- [2] M. A. Munjer, M.F. Hossain, M. H. Rahman, M. Z. Mahmud "Fabrication of a cost effective automatic controller for spray-pyrolysis technique to deposit thin films", 7th International Conference on Electrical and Computer Engineering, pp.78-81, 20-22 December, 2012.
- [3] A.Nakaruk, C.C.Sorrell, "Conceptual model for spray pyrolysis mechanism: fabrication and annealing of titania thin films", Journal of Coatings Technology and Research, Volume 7, pp 665-676, September 2010.
- [4] A.S. Enigochitraa, P. Perumala, C. Sanjeevirajab, D. Deivamania, M. Boomashria, "Influence of substrate temperature on structural and optical properties of ZnO thin films prepared by cost-effective chemical Spray Pyrolysis Technique, Superlattices and Microstructures, Volume 90, pp.313-320, February 2016.
- [5] Lado Filipovic, Member, IAENG, Siegfried Selberherr, Giorgio C. Mutinati, Elise Brunet, Stephan Steinhauer, Anton K'ock, Jordi Teva, Jochen Kraft, J'org Siegert, and Franz Schrank Modeling, "Modeling Spray Pyrolysis Deposition", Proceedings of the World Congress on Engineering 2013 Vol II, July 3-5, 2013
- [6] Wasan R. Saleh, Nada M. Saeed*, Wesam A. Twej, Mohammed Alwan, "Synthesis Sol-Gel Derived Highly Transparent ZnO Thin Films for Optoelectronic Applications", Advances in Materials Physics and Chemistry Vol.2, Article ID:17891, November 16, 2011.

Design and Measurement of a flexible Implantable Stripline-Fed Slot Antenna for Biomedical Applications

Tajwar Abrar Aleef, Akash Biswas ; American International University-Bangladesh; tajwar.aleef@gmail.com

Abstract—A novel design of a stripline-fed flexible implantable antenna is proposed in this paper. The specialty of this antenna is its small design compared to the previous antennas of similar type. The antenna presented in this paper is designed to operate in the Industrial, Scientific and medical (ISM) band, (2.4-2.4835 GHz). The thickness of the antenna is only 0.887mm, meaning this antenna can easily be subjected to bent conditions. The whole strip line fed antenna is embedded inside a silicon layer of 0.443mm, resulting in an overall packed dimension of 10.125mm $\times$ 8.728mm $\times$ 1.505mm. The antenna was then placed inside a human tissue model before simulating it in CST Microwave studio. All the performance parameters of the antenna was measured and evaluated in both planar and bent conditions. Finally, the antenna was fabricated and SMA female connector was used as the port of the antenna. Performance measurements were then taken using a Network analyzer and were compared with the simulated results.

Keywords -ISM Band, Stripline-fed, Flexible Antenna, Implantable Antenna, Human Tissue Model, Fabrication, Network Analyzer.

I. INTRODUCTION

In recent times, in the field of biomedical technology, the significance of microwave applications is growing rapidly. Some examples of these applications are: monitoring blood pressure and temperature, tracking dependent people, wirelessly transferring diagnostic information from an electronic device implanted in the human body for human care and safety, such as a pacemaker, to an external RF receiver [1]. The non-ionizing radiation pattern that ensures no harm to human cell is the key cause of the ever increasing demand of this technology in this sector. Hence, small implantable antenna is being developed to be used in patient's health monitoring devices in order to track the abnormalities. One of the outstanding benefits of wireless implanted devices are the possibility of observing the patient's vital signs as well as manipulating the device's control variables, despite the great distance between medical center (base station) and the patient[2]. As a result, it can give doctors the data to diagnose the problem more quickly compared to more traditional methods, where the patient himself/herself have to physically visit the hospital to get a checkup. This not only reduces the patient's cost but also provides a more efficient and quicker response to treatments.

A significantly small sized antenna for the miniaturization of overall implantable device along with its biocompatibility is the main challenge while designing an implantable antenna. The relative permittivity and conductivity of different human tissue layers have to be taken into account in order to calculate biocompatibility ensuring the minimum loss in radiation. The operational frequency bands for medical applications according to FCC regulation are for Medical Device Radio communications Service (MedRadio), i.e. from 401 MHz to 406 MHz and for Industrial, Scientific and Medical Radio (ISM), i.e. from 2.4 GHz to 2.5 GHz respectively [3]. Although signal attenuation in lossy human tissues decreases with the frequency, design of antenna at ISM band is preferred, since a considerably small size of the antenna for miniaturization of the overall implanted devices could be achieved due to its small wavelength [2].

The on-going research scenario on implantable antenna propagation shows a good potential in medical diagnosis and imaging purpose. A flexible folded slot dipole implantable antenna operating in the Industrial, Scientific, and Medical (ISM) band (2.4–2.4835 GHz) for biomedical applications is presented in [4] where the dielectric constant of the human muscle tissue was taken into account while testing the sensitivity of the antenna. In [5], a novel differentially fed dual-band planar antenna operating at the medical implant communication service (MICS) band (402-405 MHz) and the industrial, scientific and medical (ISM) band (2400-2480 MHz) is presented. Again a small-size dual medical implant communications service (MICS) (402-405 MHz) and industrial, scientific, and medical (ISM) (2.4-2.48 GHz) band implantable antenna for continuous glucose-monitoring applications is presented in [6] where development of skin mimicking gel is also described.

The biggest motivation behind this paper is to introduce a reasonable design with practical dimensions that can be truly considered for implantation inside a living body. With the sizes of antenna available today, practical implantation becomes unrealistic due to the size constraint. Hence, this paper demonstrates size miniaturization and a novel design of a stripline-fed implantable antenna. This design reduces the dimension while keeping the performance similar or in some cases better compared to previous versions of similar

antennas. The organization of this paper is summarized as follows. In Section II, the design and manufacturing process of the proposed antenna is discussed, where details to the geometry of the design and its simulation environment is discussed. In Section III, the simulations taken from the antenna are analyzed and discussed for both the bent and planar conditions. In Section IV, the fabrication process and measurements taken from the antenna inside a human mimicking gel is analyzed and compared with the simulation data. Lastly, in Section V, the conclusion is summarized.

II. DESIGN AND MANUFACTURING PROCESS OF THE BIOCOMPATIBLE STRIPLINE ANTENNA

The antenna presented in this paper is a flexible stripline-fed antenna, submerged inside a biocompatible silicon block. As this antenna has been designed for implantable purposes, its physical dimension plays a vital role. The antenna was initially designed in free space, using reference from an existing stripline-fed implantable antenna [2] with CST Microwave Studio. The proposed antenna is designed to operate in the ISM band since ISM band has been considered more suitable for biomedical applications as its high bandwidth, which allows making the wavelength smaller as well to obtain higher bit rate after implantation [7-8]. Design was altered through iteration technique to accommodate overall size reduction while keeping the antenna's operating resonant frequency inside the ISM band. The patch and ground plane material is cooper of thickness 0.175mm is fabricated on substrate Rogers RT 6010 with the dielectric constant of 10.2, $\tan \delta$ of 0.0023 and thickness of 0.887 mm. This tiny thickness means the antenna was not only reduced in terms of dimension, but this also made the antenna more flexible meaning the antenna can be curve against the normal curvature of a living body. In Table I, all the dimension values of the designed antennas are given. Through iteration, this dimension values were obtained, where the best performance was observed at the lowest possible dimension. In order to make the antenna suitable for implantation, biocompatible superstrate silicon block was used to encapsulate the overall design. The superstrate layer is mainly used to shield the antenna from direct contact of the living tissue. Also, the superstrate acts as a buffer between the metal radiator and human tissues by reducing RF power at the locations of lossy human tissues [9]. The overall dimension of the antenna was $10.125\text{mm} \times 8.728\text{mm} \times 1.505\text{mm}$. This meant the proposed antenna has an overall size reduction of 30% compared to the reference antenna.

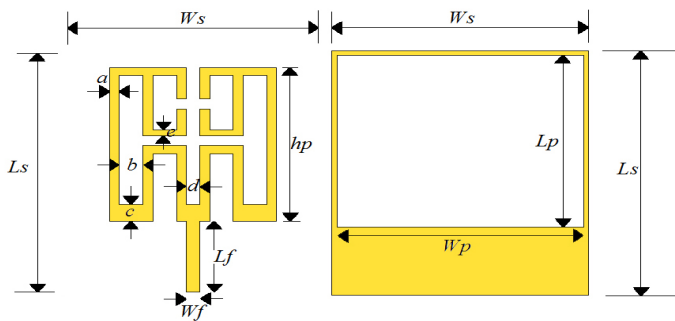


Figure 1. Model of the Antenna

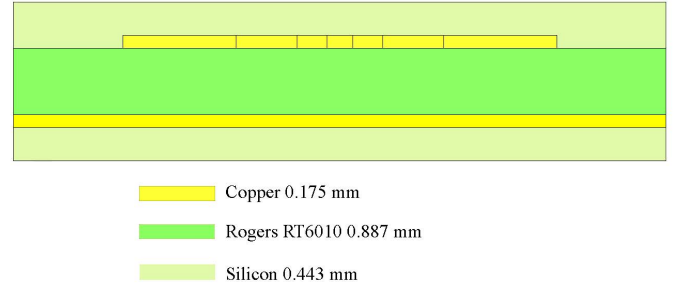


Figure 2. Frontal view of the proposed antenna

In Figure 1, both the top and bottom geometry of the antenna is shown along with the size parameters of the antenna. Different slots have been cut to reduce the return loss and tune the antenna to the ISM band. In Figure 2, the thickness of the different materials is shown from a frontal view. It can be seen from the figure that the overall antenna was encapsulated inside the biocompatible silicon superstrate and its overall thickness summed up to 0.887mm.

TABLE I. Antenna Size Parameters

Parameter	Size (mm)
W_s	8.728
L_s	10.125
W_p	8.379
L_p	7.122
W_f	0.350
L_f	3.010
h_p	6.500
a	0.349
b	0.817
c	0.698
d	0.454
e	0.244

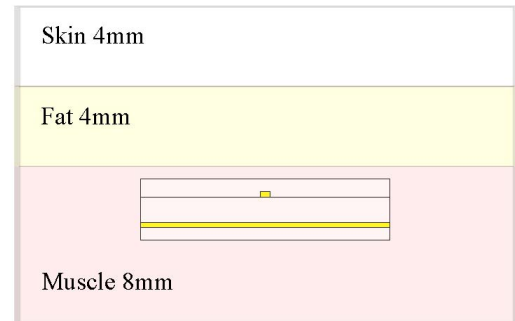


Figure 3. Three Layer Geometry Model

Since the antenna is primarily designed for implantable use, the simulation couldn't be carried out in free space. Hence, for simulation, a simple cubic three layer geometry model consisting of 8mm layer of muscle, 4 mm layer of Fat and 4 mm layer of skin was chosen. The antenna with the silicon block was then inserted inside this model before carrying out any further simulations.

The permittivity and conductivity of the tissue for the one layer tissue model are from those of dry skin, and the dielectric properties of tissue vary with frequency. The numerical values of permittivity and conductivity for all tissues used in our models at our operating frequency are listed in Table II, which can be found in [10]

TABLE II. Dielectric Values of Human Muscle Tissue at Different Frequencies

Frequency (GHz)	ϵ_r	σ S/m
2.0	53.29	1.45
2.4	52.79	1.70
2.45	52.72	1.73
2.5	52.66	1.77
3.0	52.05	2.14

III. SIMULATION AND COMPARISON ANALYSIS

Various simulation environments are seen when simulating implantable antennas. In the case of the proposed antenna, directly a 3-layer tissue model was used and the thickness was determined from [4]. Figure 4 shows the antenna inside the three layers Human Phantom Model. The S11 parameter represents how much power is reflected from the antenna and it is known as the reflection coefficient. From Figure 5, it can be seen that the proposed antenna has a resonant frequency of 2.48GHz, which falls in the ISM Band. The return loss of the antenna can be again seen from Figure 5 which is at -30.9dB. The lower the return loss the higher is the signal radiated rather than being reflected. The bandwidth of the antenna is calculated by drawing a straight line across the -10 dB mark and is found to be 490MHz.

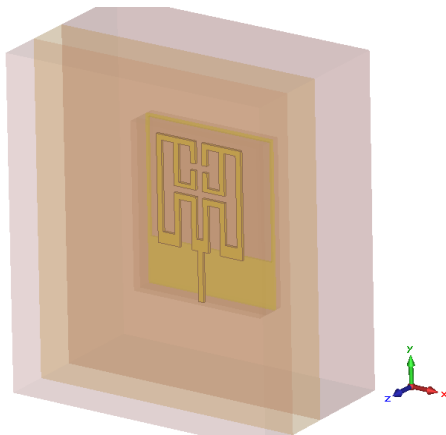


Figure 4. Antenna inside Human Phantom Model

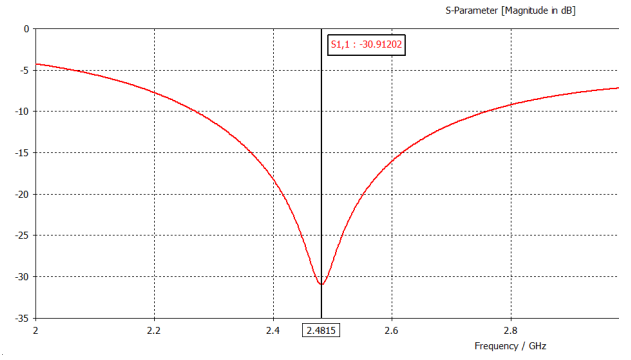


Figure 5. S11 Parameter of the antenna inside Human Phantom model

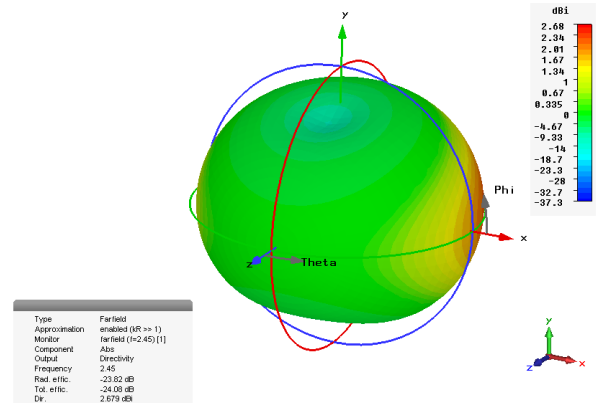


Figure 6. 3D view of Far-Field Region of the antenna inside Human Phantom Model

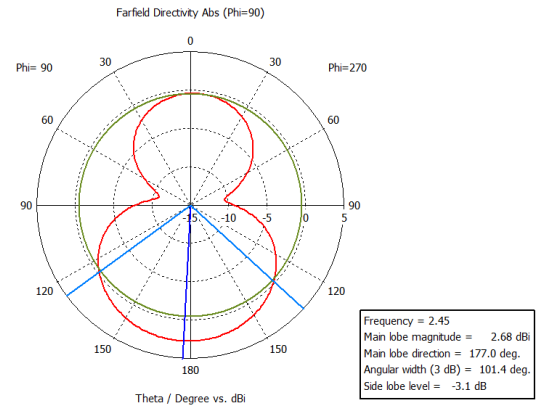


Figure 7. Polar view of Far-Field Region of the antenna inside Human Phantom Model

Using CST Microwave Studio simulations, the radiation characteristics of the antenna inside the human phantom model are determined. From Figure 7, it can be observed that the antenna has a directivity of 2.679 dBi. From Figure 8, it can be observed that the main lobe magnitude is 2.68 dBi which is centered at 177 degree with an angular width of 101.4 degree. The maximum gain of the antenna is -10.75 dBi at $\theta = 75$ degree and $\phi = 90$ degree.

As this antenna is intended for implantable use, the antenna is expected to flex along with the curvature of a human body. In order to make sure that the antenna has a similar performance

when subjected to such curvature, the antenna was bent around the x-axis around a 4cm diameter cylinder as shown in Figure 9. The thin dimension of the overall design allows easy flexing making the antenna easier to wear.

The antennas reflection coefficient was then compared with the reflection coefficient in its normal planar state. Figure 10 shows the S11 parameter graph of both bent and planar condition. From the graph it can be observed that there is a slight shift of the resonant frequency from 2.4815 GHz to 2.4905 GHz. The return loss is also seemed to have decreased from -30.9dB to -23.73dB. Overall, the resonant frequency was still in the ISM band and its return loss was well below the -10dB mark. This meant that the antenna when susceptible to duress can still perform accordingly.

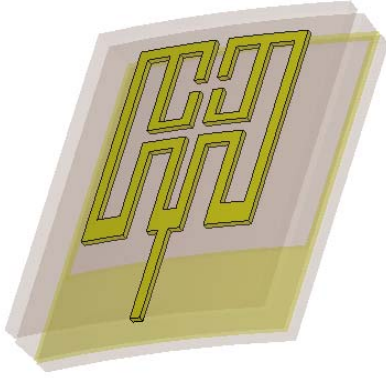


Figure 8. Slot Dipole Rectangular Antenna Undergoing Bent Test

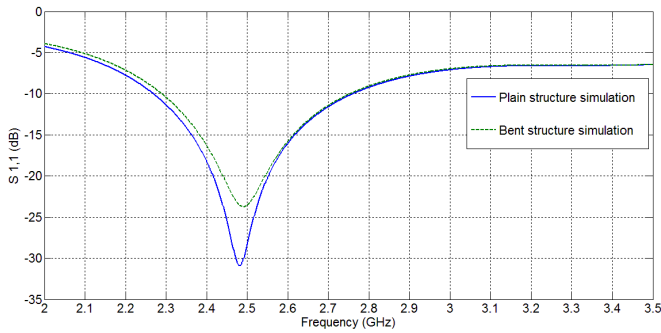


Figure 9. S11 Parameter of Bent Antenna inside Human Phantom Model

IV. FABRICATION AND IN-HUMAN PHANTOM GEL MEASUREMENTS

In order to check if the antenna performance is similar in the real world compared to the simulated results, the antenna must be fabricated. Hence the newly designed antenna was fabricated by etching copper off a double sided Rogers RT6010 board. Female SMA Connector was used as the port for the final fabricated design. The antenna is placed at the center of a plastic container of dimensions 80 cm x 50 cm x 20 cm filled with 30 liters of the Human Muscle Tissue liquid MSL2450 [11]. This liquid mimics the dielectric characteristics of human muscle tissue at 2.45 GHz, standardized in [12] to be $\epsilon_r=52.7$, $\sigma=1.73\text{S/m}$, $\rho=1000$

kg/m^3 . Dielectric values of the liquid at 2.45 GHz measured by the manufacturer result to be $\epsilon_r=50.8$, $\sigma=2.01\text{S/m}$, $\rho=1.030 \text{ kg/m}^3$ [11]. Agilent technologies E5071C Network analyzer was used to measure the antennas performance while embedded inside the human phantom liquid. Figure 11 shows the fabricated antenna and the measurement setup used. The resulted S11 parameter as seen from Figure 12 had a slight shift to the left compared to the simulated S11 parameter. The measured resonant frequency was at 2.38 GHz and its return loss was at -27 dB. This discrepancy is due to the tolerance introduced in the fabrication process.

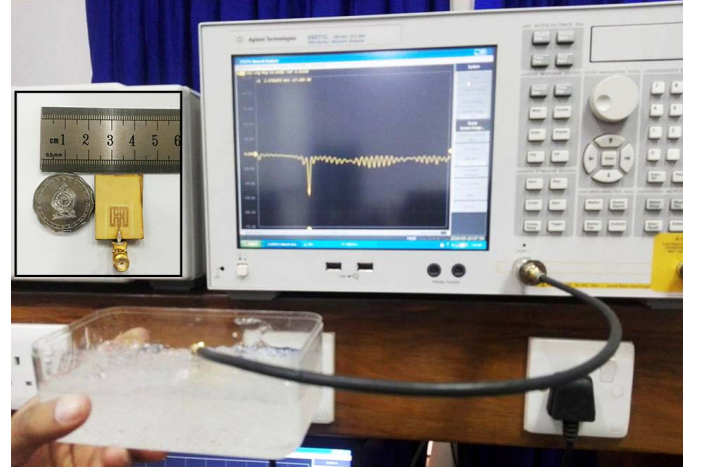


Figure 10. Fabricated Strip line Antenna and the measurement setup

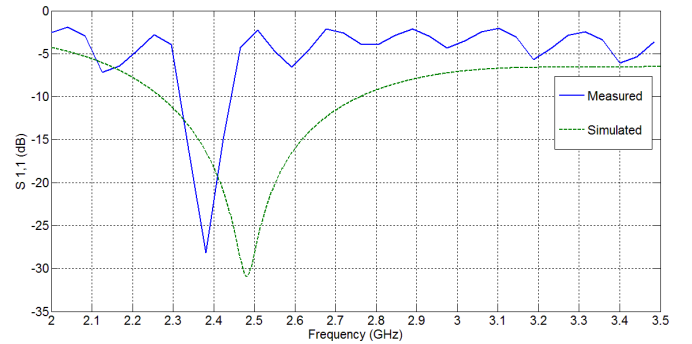


Figure 11.S11 parameter comparison between measured and simulated results

V. CONCLUSION

A flexible stripline-fed implantable antenna was designed in this paper. After the design process, the antenna was encapsulated inside a silicon block superstrate and was then inserted inside a 3 layer geometry model. Different performance parameters were simulated first such as the resonance frequency, the return loss, the bandwidth, the far field radiation, the antennas susceptibility to bend. Later, the antenna was fabricated to compare measured data with the simulated ones.

Some of the important points that can be drawn for the designed antenna are-

- Novel design and 30% size reduction compared to its previous reference antennas.
- Thin dimension which allows the antenna to flex along the curvature of the body without high performance loss.
- Higher bandwidth and good return loss.
- Very low discrepancy between measured and simulated data.

By looking at the results, this proposed antenna solves the drawbacks that many previous implantable antennas had, that are the overall dimension. This antenna exhibits proper performance all while having an overall dimension reduction. Future works should cover an integration of the fabricated antenna with a total biomedical system or device. In short, seeing the overall performance, it can be said that this antenna is ready to be implanted.

REFERENCES

- [1] B. M. Steinhaus, R. E. Smith, & P. Crosby, "The role of telecommunications in future implantable device systems," *Proc. 16th IEEE EMBS Conf.*, Baltimore, MD, vol. 2, pp. 1013–1014, Nov. 1994.
- [2] Xuyang Li, Malyhe Jalilvand, Wendy You, Werner Wiesbeck, Thomas Zwick, "An implantable stripline-fed slot antenna for biomedical applications" *Radar Conference, IET International*, Xi'an, pp.1-5, 14-16 Apr. 2013.
- [3] "FCC rules: CFR-47, part-15, ISM" FCC, Washington, DC, Oct. 2005. [Online]. Available: <http://wireless.fcc.gov/services/index.htm>
- [4] Maria Lucia Scarpello, Divya Kurup, Hendrik Rogier, Dries Vande Ginste, Fabrice Axisa, Jan Vanfleteren, Wout Joseph, Luc Martens & Gunter Vermeeren "Design of an Implantable Slot Dipole Conformal Flexible Antenna for Biomedical Applications," *IEEE Transactions on Antennas and Propagation*, vol. 59, no. 10, pp. 3556--3564, Oct. 2011.
- [5] Yijun Liu, Yifan Chen, Haili Lin & Filbert Juwono, "A Novel Differentially Fed Compact Dual-Band Implantable Antenna for Biotelemetry Applications," in *IEEE Antennas and Wireless Propagation Letters*, vol. PP, no. 99, pp. 1, 01 Mar. 2016.
- [6] T. Karacolak, A. Z. Hood & 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, Apr. 2008.
- [7] H. Usui, M. Takahashi, & K. Ito, "Radiation characteristics of an implanted cavity slot antenna into the human body," in *Proc. IEEE Antennas and Propagation Society International Symposium*, Albuquerque, NM, pp.1095–1098, Jul. 2006.
- [8] W. Xia, K. Saito, M. Takahashi, & K. Ito, "Performances of an implanted cavity slot antenna embedded in the human arm," *IEEE Trans. Antennas Propag.*, vol. 57, no. 4, pp. 894–899, Apr. 2009.
- [9] J. Kim & Y. Rahmat-Samii, "Implanted antennas inside a human body: Simulations, designs, and characterizations," *IEEE Transactions on Microwave Theory and Techniques*, vol. 52, no. 8, pp. 1934–1943, Aug. 2004.
- [10] S. Gabriel, R. W. Lau, & C. Gabriel, "The dielectric properties of biological tissues," *Physics in Medical and Biology*, pp. 2231–2293 [Online]. Available: <http://niremf.ifac.cnr.it/tissprop/>, Oct.2004, 1996.
- [11] [Online]. Available: <http://www.speag.com/measurement/liquids/>
- [12] C. Gabriel, S. Gabriel, & E. Corthout, "The dielectric properties of biological tissues: I. Literature survey," *Phys. Med. Biol.*, vol. 41, no.11, pp. 2231–2249, 1996.

License Plate Detection and Character Recognition System for Commercial Vehicles based on Morphological Approach and Template Matching

Animesh Chandra Roy, Muhammad Kamal Hossen, Debashis Nag

Dept. of Computer Science & Engineering
Chittagong University of Engineering & Technology
Chittagong, Bangladesh

animesh_roy09cse@yahoo.com; kamal_cuet@yahoo.com

Abstract—Though license plate detection and recognition system has a lot of applications but a very little work has been done on Bangla license plate recognition. Variations among the license plate patterns and complex background of Bangladeshi license plates make it more difficult to use the existing algorithms. Because of this, we propose a solution for Bangla license plate detection and recognition. Firstly, the position of the vehicle is determined. In our country, commercial License Plates have the unique color (green) of its own. That's why; we select the portions of green color with the matching RGB intensity of the plate. A boundary based contour algorithm, area and aspect ratio have been proposed to track down the license plate in the vehicle region. License plate rows containing registration information have been separated by using horizontal projection with threshold. The characters and the digits of the rows have been segmented using vertical projection with threshold value. Finally, template matching has been used for recognizing the characters and the digits of the Bangla license plate. We tested our algorithms for over 180 still images captured from the road. We achieve 93% success in license plate detection, 98.1% success in segmentation and 88.8% success rate in Bangla license plate recognition.

Keywords—image processing; license plate detection; extraction; contour algorithm; segmentation; classification.

I. INTRODUCTION

License Plate Detection and Recognition is the first step of LPR technology to automatically read license plate characters. It is the foremost modules of Intelligent Transportation System (ITS). A fully operative LPR system is consists of three major steps such as, Detection of plate region, Segmentation and classification of characters and digits.

LPR system is well-known over the world but it is very new for our country. Bangladesh is one of the most densely populated countries in the world and since the number of automobiles increases, it means the number of Plates delivered by BRTA (Bangladesh Road Transport Authority) also increases. The fraudulent Plates also give opportunity to increase vehicles related crimes. Thus recognition is the only solution to investigate the unique number plates.

We can indicate some significant applications of LPR system, such as electronic payment systems, finding hijacked vehicles, catching the over speedy drivers by analyzing the average speed on the road, resolve which vehicle belongs to a parking garage and which is not etc. On account of being a fundamental element of Intelligent Transportation System (ITS), contribution of an LPR system on daily purpose is massive. It can be used to bring automation in toll collection systems for highways, flyovers, and bridges. By using an LPR system toll can be collected without manual intervention that reduce time and traffic congestion. LPR system can be used by metropolitan police department for effective law enforcement, effective enforcement of traffic rules. It can also bring highest flexibility for border control systems.

But there are also some difficulties that interrupt the successful detection of a license plate, such as the resolution of an image can be poor or blurry, brightness and contrast can also be very poor, body of an object overlays the plate region, there can be some dirt on the plate which hides some of the portion of the plate.

Through this paper, we prefer a technique that matches RGB color intensity with contour detection and then localize the plate region by using aspect ratio of the region. Before extracting the plate region we have to correct the skewed image. We separated the rows containing information in the license plate. We segmented the Bangla characters and digits by using horizontal and vertical projections. Finally, we recognized the extracted characters and digits by using Template Matching.

The remaining portion of the paper is formulated as follows: Section II shows a description of relevant works. Section III introduces our prospective license plate detection and recognition scheme. Section IV gives the tentative result of our proposed algorithms and correlates the performance with well-known methods. Finally, section V completes the paper with conclusion.

II. RELATED RESEARCH

The combination of morphological operations and edge like structures are used in [1], [7] and [9]. Brightness difference is denser in the plate region compared than other portions in the image. This has been used as a distinguishing property to find the region of interest (plate region). Firstly, the computation of the local variance and gradient magnitude of the edges is done. Mostly, License plates are located based on the higher magnitude of the gradient and the variance. Morphological operations, such as area, aspect ratio (AR), orientation are also introduced with binary image processing algorithms in [11] to find the plate in an image. These criteria are used to clarify candidate license plate regions, which are in fact binary objects. Binary objects whose area and aspect ratio do not fall into the respective reasonable ranges are eliminated. Gabor filters are used in [13] along with aspect ratio, size and orientation to extract the number plate region. The Gabor Filters have accepted considerable attention because the frequency and the orientation representations of human visual system can be easily approximated by these filters. In [16], an adaptive technique with labeling and filtering capability is used to select the required threshold value from automatic threshold and the HSI color model shows the detection of number plate region. Hough Transform (HT) is also used in literature [4] where the boundary line scan be detected of the plates. For extracting the structure here the Hough transform successfully finds out both the perfect and imperfect objects between a given shapes. Many algorithms [3, 10] based on color processing are introduced in the literature for detection of the number plate. By using color processing method, the main idea of the detection of the plate region is that, plate region(background) and its characters (foreground) contains the different color and these color is unique, and mostly the plate region consists this unique combination rarely in other regions. This color combination is searched pixel-by-pixel through the image and the matching region is the required region of interest (number plate). However, color processing techniques have some disadvantages. When the lighting or brightness condition changes the color also differs. With different brightness, different colors can appear on the same object. That's why; license plate detection rate may vary in different illumination conditions and also color based techniques are country specific because license plate colors vary from one country to another country. An image separation technique named Sliding Concentric Windows (SCW) is used for license plate detection in [6, 14]. The SCW algorithm is based on the local characteristics such as mean and variance. This technique uses two concentric windows where A, B are of different sizes $X1 \times Y1$ and $X2 \times Y2$ respectively, where $X1 > X2$ and $Y1 > Y2$.

Computing the horizontal projection for row and vertical projections for column is the most elementary method to segment characters [8, 9]. Usually the projections are computed using a binary image. The horizontal projection is used to separate the rows of the license plate and the characters and digits of the rows are segmented by the vertical projection. Hidden Markov chain [12] is used to prototype a model of relationship between an input image and the segmented characters. Research work [1] presented a character

segmentation algorithm based on mathematical morphology. After converting a binary image from the captured image, this method reduces the noises by using morphological operations, such as thickening and pruning. After removing noises, the vertical projection is used based on the minimum values are computed. Thus, some characters might be divided into multiple segments and two or more characters might be put into one segment.

For decomposing the candidate region the image must be aligned without any rotation since the captured image can be skewed. Skewed image detection and correction of this problem are the prerequisite before the segmentation of components and classifying those for further processing. The least square fitting with perpendicular offsets (LSFPO) is used in [6], the VLP region is matched to a straight line.

To classify the characters and digits, multilayer feed-forward neural networks have been introduced in many research works [3, 4, 7]. To achieve better classification rate over some test data, Support Vector Machine (SVM) was proposed in [12] which recognizes the characters in Korean license plates. This method used four SVM based character recognizer in their system to recognize the characters and digits in four different parts of the license plate, such as upper character, upper numbers, lower character, and lower numbers. Template matching algorithm which widely used for finding segmented parts of an image that matches template image. This is an efficient technique where a predefined template is created for the recognition of fixed-size single or double font and non-rotated characters. Template matching is successfully implemented in [2, 15].

III. PROPOSED METHOD

A. Overview

It is obvious that a study and effective LPR system is required for Bangladeshi commercial license plates. Though according to some research work in [7, 8, 10] have been done for recognizing bangla license plate, but no effective LPR system has been developed yet for Bangladeshi commercial license plates which are actually in green color. In this paper, an LPR system for Bangladeshi commercial license plates is presented. This LPR system is comprised of mainly three stages, such as (i) detection,(ii) segmentation, and (iii) recognition. In the first stage, the position of the vehicle in the input image is located by using color segment and contour algorithm. The tilt of the plate is corrected by using geometrical transformation. Then, the extraction of the required number plate is done. In the second stage, characters and digits are segmented using horizontal and vertical projection. In the third stage, these characters and digits are recognized by using template matching.

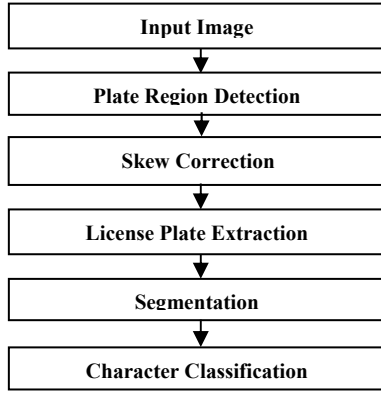


Fig. 1. Proposed method.

B. Plate Region Detection

Bangladeshi commercial License Plates have the unique color (green) of its own. That's why; at first the portions of green color is selected with the matching RGB intensity of the plate. The RGB value of a substance in image depends on a brightening condition. It is difficult to calculate the illumination condition from an image of unknown scene. However, the license plate color is known since the color is usually defined by the law. In color judgment method, an illumination condition can be estimated from the license plate intensity. The vehicle color is judged by using color sample according to the illumination condition. Then color samples are statistically created which have correspondences of colors and RGB values in advance. Each color has multiple RGB values, although an RGB value only has one color. The color sample is considered supposing different illumination conditions which make the plate intensity differently.

The vehicle color estimation technique is carried out as following procedure:

- 1) *Selection of a color sample: A color sample is created according to the license plate intensity.*
- 2) *Judge the color of each pixel in the area: The nearest color of the pixel from selected color sample is considered by calculating the RGB color scale.*



Fig. 2. (a) Input image; (b) Binary image

After taking those selected color area, the image is converted into binary form. The input and binary image is shown in Fig. 2(a) and Fig. 2(b).

Then the closing operation is processed to find the connected regions. The closing of image f is done by a structuring element s , denoted as simply a dilation followed by erosion. This operation $(f.s)$ mainly used to eliminate the holes and to fill gaps in the contour.

$$\text{Where, } f.s = (f \oplus s) \ominus s \quad (1)$$

And, s = structuring element

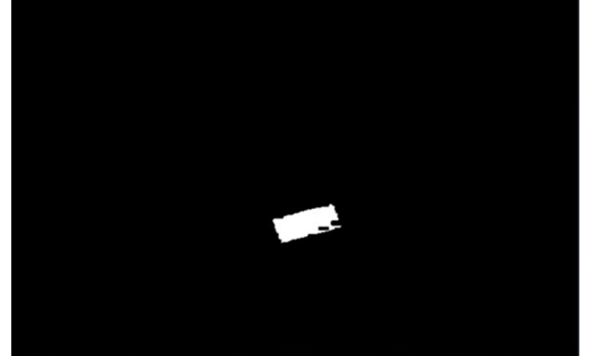


Fig. 3. Binary image after closing operation

For detecting the regions of plate-candidate object, contour algorithm is used to find out the closed boundary objects. A number of candidate judgment algorithms are processed on the binary image. One of the contour algorithms named square tracing algorithm is used to separate plate objects. This algorithm works as follows; in order to extract the contour of the pattern, after finding a black pixel, move left and every time standing on a white pixel, turn right, until encounter the start pixel again. The black pixels walked over will be the contour.

Next approach was to draw rectangles over the boundary regions based on their maximum external edge values. These rectangles contain the center, the height and the width of these regions. To locate the plate region finally the aspect ratio and area were judged over the selected rectangles. The ratio between height and width is almost same and it is 1:2. The rectangle that satisfies the aspect ratio is the Region of Interest (ROI).



Fig. 4. (a) Selected Contour(marked by Black color); (b) Rectangle over the Contour(marked by Green color).

C. Skew Correction and Plate Extraction

To correct the skewed image, at first, rotation angle of the image have to find. This can be done by using the

selectedRect.angle() function. Then, to rotate the image, a rotation matrix is needed to multiply with the image. With the angle the rotation matrix can be generated. Then the plate region is cropped from the rotated image and is resized in a fixed size for further processing.

Here, the *getRotation* matrix is defined by,

$$\begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix} \begin{bmatrix} (1 - \alpha).center.x - \beta.center.y \\ \beta.center.x + (1 - \alpha).center.y \end{bmatrix}$$

Where,

$$\begin{aligned} \alpha &= scale.cos(angle) \\ \beta &= scale.sin(angle) \end{aligned}$$

Now, rotate the image using the rotation matrix where the formula is,

$$dst(x, y) = src(M_{11}x + M_{12}y + M_{13}, M_{21}x + M_{22}y + M_{23})$$

Where,

src = input image

dst = output image

M = (2 × 3) rotation matrix

Fig. 5(a) and Fig. 5(b) show the overall process.



(a)



(b)

Fig. 5. (a) Tilt corrected image; (b) Extracted & Tilt corrected license plate.

D. Segmentation

As mentioned before, Bangla characters in Bangladeshi license plates are very difficult to segment because of their various complexities. To deal with these complexities, an algorithm is introduced that uses horizontal and vertical projections with threshold to segment Bangla characters and digits efficiently. The segmentation process has two steps. In

the first step, separation of the rows containing license information is done. Then, segmentation of the characters and digits is done in the second step.

1) *Separating the rows*: To extract the segmented bangla characters and words from the plate, at first the image is converted into gray scale image and then Robert cross operator is used for edge detection. Since, Bangladeshi number plates contains more than one lines mainly two lines of text; horizontal projection can be done by calculating total number of row pixels which segments the rows in two different lines. The row with minimum values of horizontal pixels are the start or the end of a row in plate. Fig. 6(a), Fig. 6(b) and Fig. 6(c) show the overall process.

Formula for converting into gray scale image is,

$$Y = 0.2126R + 0.7152G + .0722B \quad (2)$$

For edge detecting, Kernels of the Robert Cross operator are,

+1	0	0	+1
0	-1	-1	0

The gradient magnitude is given by:

$$|G| = \sqrt{Gx^2 + Gy^2} \quad (3)$$

Although typically, an approximate magnitude is computed using:

$$|G| = |Gx| + |Gy| \quad (4)$$

For a given binary image $\{BI(i,j); 1 \leq i \leq H, 1 \leq j \leq W\}$, its horizontal is computed by using (5).

$$fh(i) = \sum_{j=1}^W BI(i, j) \quad (5)$$



(a)

(b)



(c)

Fig. 6. (a) Gray Scale image; (b) Edge detection; (c) Horizontal Projection.

Due to some obstacle barrier like screws, stuck characters and dash (-) character in plates, difficulty in Bangla character segmentation becomes more complex. That's why, before segmenting the characters and digits Otsu binarization method is used at each projected rows for better output.



Fig. 7. Segmented rows after Otsu method.

2) *Segmenting the characters and digits*: Vertical projection is used to separate each bangla characters and words. During vertical projection, dash (-) character and screws are removed which contains very small frequency. After implementing this process there 9 top picks are found where there are 7 characters and 2 words from the number plate. By removing the redundant area from each output of the segmentation process the required characters and digits can be found.

For a given binary image $\{BI(i,j); 1 \leq i \leq H, 1 \leq j \leq W\}$, its vertical projection is computed by using (6).

$$fv(i) = \sum_{j=1}^H BI(i,j) \quad (6)$$

E. Character Classification

In the case of template matching, template images of each character in each font that become the references for the comparison are created and stored in a database. Then, the template matching method made continuously search to find whether a similar template exists within the area. Generally, to make the template matching usable in real practice, the size of the candidate images is normalized to a predefined dimension, which is exactly the same with that of the template images.



Fig. 8. Segmented bangla words and digits.

If Character Image is T and Target image from search image set is I, then for each point in character image (T) and search image (I), find normalize cross correlation using (7),

$$R(x,y) = \frac{\sum_{x',y'} T(x',y') \cdot I(x+x',y+y')}{\sqrt{(\sum_{x',y'} T(x',y')^2) \cdot (\sum_{x',y'} I(x+x',y+y')^2)}} \quad (7)$$

Maximum value for search image returned as matched character.



Fig. 9. License Plate Recognition.

IV. EXPERIMENT RESULTS

In this system, a digital camera is used to capture at 5 to 10 meters distances from camera to vehicle with different background. These algorithms are applied on 180 color images with the size of 640×480 pixels of Bangladeshi commercial vehicles snapped under different illumination conditions, such as night, sunny, and cloudy. Those are still images captured from vehicle but images from moving vehicle would give less accurate result. The parameters that are used in the filtering operation to detect ROI from the candidate list mentioned in Table I. Table II shows the experimental result of the system.

TABLE I. FILTERING PARAMETER

Filtering Parameter	Candidate Region
Aspect Ratio	[1.0, 2.0]
Area	[2000, 8000]

TABLE II. EXPERIMENTAL RESULT

Step	No. of Instances	No. of success	Success Rate (%)
Detection	180	165	93
Segmentation	165	162	98.1
Recognition	162	145	88.8

In the presence of complex background and highly variable license plate patterns in the input images, the average detection rate of our algorithm is 93%, segmentation rate 98.1% and recognition rate 88.8%. From Table III, Table IV, and Table V, we can see that the successful license plate detection rates of the algorithms [4] and [8] are 85%, and 92%, the successful segmentation rates of the algorithms [4] and [8] are 93.75%, and 95%, the successful classification rates of the algorithm [15] is 85%, respectively. It proves that the performance of our algorithm is better than that of the other algorithms in recognition Bangla license plate.

TABLE III. COMPARISON OF LICENSE PLATE DETECTION STAGE

Algorithms	No. of Images	License plate detection	Success rate (%)
Proposed Scheme	180	165	93
Duan et al.[4]	80	68	85
Joarder et al. [8]	100	92	92

TABLE IV. COMPARISON OF SEGMENTATION STAGE

Algorithms	No. of LP	Correct Segmentation	Success Rate (%)
Proposed Scheme	165	162	98.1
Duan et al. [4]	80	75	93.75
Joarder et al. [8]	100	95	95

TABLE V. COMPARISON OF RECOGNITION STAGE

Methods	Template Matching	No. of Instances	Correctly Classified	Success Rate (%)
Proposed Method	Character classification	160	142	88.8
Huang et al.[15]	Character classification	120	107	85

V. CONCLUSION

In this paper, we present a set of algorithms for the detection, segmentation and recognition of license plate number written in Bangla. In each stage, we developed our techniques and algorithms suitable for recognizing Bangla license plates. We detected the vehicle in the image by using the property of color and contour. We took 180 images containing vehicle under different conditions such as sunny day, cloudy day and at night. We tested our algorithm for these inputs and we found that our proposed scheme outperform existing algorithms in the recognition of Bangla license plate.

REFERENCES

- [1] J.W. Hsieh, S.H. Yu, and Y.S. Chen, "Morphology based license plate detection from complex scenes," in Proc. of 16th Int. Conf. on Pattern Recognition, vol. 03, 2002.
- [2] S.L. Chang, L.S. Chen, Y.C. Chung, and S.W. Chen, "Automatic license plate recognition," IEEE Trans. on Intelligent Transportation Systems, vol. 05, no. 01, 2004.
- [3] C.E. Anagnostopoulos, I.E. Anagnostopoulos, V. Loumos, and E. Kayafas, "A license plate recognition algorithm for intelligent transportation system applications," IEEE Trans. on Intelligent Transportation Systems, vol. 07, no. 03, pp. 377-382, 2006.
- [4] T.D. Duan, T.L.H. Du, T.V. Phuoc, and N.V. Hoang, "Building an automatic vehicle license-plate recognition system," Int. Conf. in Computer Science-RIVF, pp. 59-63, February 2005.
- [5] C.N.K. Babu, and K. Nallaperumal, "An efficient geometric feature based license plate localization and recognition," Int. Journal of Industrial and Systems Engineering, vol. 02, no. 02, January 2008.
- [6] K. Deb, A. Vavilin, J.-W. Kim, T. Kim, and K.-H. Jo, "Projection and least square fitting with perpendicular offsets based vehicle license plate tilt correction," in Proc. of IEEE Conf. on the Society of Instrument and Control Engineers, pp. 3291-3298, August 2010.
- [7] A.K. Ghosh, S.K.D. Sharma, M.N. Islam, S. Biswas, and S. Akter, "Automatic license plate recognition (ALPR) for Bangladeshi vehicles", Global Journal of Computer Science and Technology, vol. 11, no. 21, ver. 1.0, pp. 68-73, December 2011.
- [8] M.M.A. Joarder, K. Mahmud, T. Ahmed, M. Kawser, and B. Ahamed, "Bangla automatic number plate recognition system using artificial neural network," Asian Trans. on Science & Technology, vol. 02, no. 01, March 2012.
- [9] S. Ozbay, and E. Ercelebi, "Automatic vehicle identification by plate recognition," in Proc. of Word Academy of Science Engineering and Technology, vol. 09, no. 04, pp. 222-225, November 2005.
- [10] M.S. Mashuk, M.A. Majid, N. Basher, and T.R. Rahman, "Automatic detection of Bangla characters in Bangladeshi car registration plates," in Proc. of Sec. Int. Conf. on Computational Intelligence, Modelling and Simulation, pp. 166-171, 2010.
- [11] G. Cao, J. Chen, and J. Jiang, "An adaptive approach to vehicle license plate localization," in Proc. of 29th Industrial Electronics Conf., pp. 1786-1791, 2003.
- [12] P. Wu, H.H. Chen, R.J. Wu, and D.F. Shen, "License plate extraction in low resolution video," in Proc. of 18th Int. Conf. on Pattern Recognition, pp. 824-827, 2006.
- [13] R. Zunino and S. Rovetta, "Vector quantization for license-plate location and image coding," IEEE Trans. on Industrial Electronics, vol. 47, no. 01, pp. 159-167, 2000.
- [14] F. Kahraman, B. Kurt and M. Gkmen, "License plate character segmentation based on the Gabor transform and vector quantization," in Proc. of 18th Int. Symposium on Computer and Information Sciences, pp. 381-388, 2003.
- [15] Y.P. Huang, S.Y. Lai, and W.P. Chuang, "A template-based model for license plate recognition," in Proc. of IEEE Int. Conf. on Networking, Sensing and Control, pp. 737-742, 2004.
- [16] M. K. Hossen and K. Deb, "Vehicle license plate detection and tilt correction based on HSI color model and SUSAN corner detector," Smart Computing Review, vol. 4, no. 5, October 2014.

Toward Letter Recognition System: Determination of Best Wavelet and Best Rhythm using EEG

Shabnam Wahed, Masud Rana, S. M. Kamrul Hasan, and Mohiuddin Ahmad

Department of Electrical and Electronic Engineering
Khulna University of Engineering & Technology (KUET)
Khulna, Bangladesh
shabnam.wahed@gmail.com

Abstract—The paper describes the application of various wavelet analysis together with machine learning algorithm for the recognition of English alphabet from EEG signal. Decision making was executed in two stages. At first important features such as maximum, minimum, delta value, moment, kurtosis, skew, median, mean and standard deviation at various sub-bands are calculated using wavelet functions —Daubechies 8, Coiflet 6, Biorthogonal 4.4, Symlet 4. Finally, a learning-based algorithm like support Vector Machine (SVM) classifier is implemented for classifying letters. From the analysis, Daubechies 8 is found the most appropriate wavelet function among the wavelet families in the proposed research for precise recognition of different letters. So the focus of the work is to recognize different letters through SVM classifier. In this analysis, among different rhythms of EEG signal delta rhythm shows best performance in recognizing letters and the recognition rate is 80%.

Keywords— Letter recognition, Discrete Wavelet Transform (DWT), Electroencephalography (EEG), machine learning, feature extraction.

I. INTRODUCTION

Direct communication with the outside world using human brain activity is the most thought-provoking area in the present research world. Amyotrophic Lateral Sclerosis (ALS), muscular dystrophies, brainstem stroke, cerebral palsy, multiple sclerosis, brain or spinal cord injury and other different disorders can damage the neuromuscular networks [1]. Human body has billions of nerve cells and the brain alone has about 100 billion neurons on an average. The movement of electrons in these neurons in human brain produce Electroencephalogram (EEG) signal. EEG generally detects the signal of task performed by the particular brain region over a short period of time (10–20 minutes) where the electrodes are positioned. EEG signal has basically 4 frequency bands as shown in Table I [2].

Since the early days of automatic EEG processing, representations based on Fourier transform have proved beneficial for various EEG characterizations. Fourier transform delivers frequency information that can only be extracted for the whole duration of signal. But it gives no information about the local discrepancies in time for non-stationary signal. In late 1980s wavelet transform was proposed to perform time scale analysis as a powerful method. This method with its capability to perform local time-frequency decomposition is really able to measure the time dependent variations of instantaneous frequency contents of a non-stationary signal.

TABLE I. BASIC FREQUENCY BANDS OF EEG SPECTRAM

Rhythm	Frequency (Hz)	General functionality
Delta (δ)	0.5-4	Prominent during some continuous attention tasks
Theta (θ)	4-8	Found in adults, children during drowsiness/ arousal condition
Alpha (α)	8-13	Dominant during relaxed/ reflecting condition
Beta (β)	13-30	Appears in active, anxious thinking or concentration

Although wavelet analysis is used as the most significant technique of time frequency analysis of EEG signal, the researchers are still working to find out the most appropriate wavelet for analysis of non-stationary EEG signals. For signal representation there are a number of wavelet families and for signal analysis selection of suitable wavelet is very vital. Authors in [3] found that using the Mean Absolute Value (MAV) of the Coiflet wavelet coefficients as inputs to Neural network the optimum classification performance was 89.11% for the purpose of imagined fists and feet movements. Authors in [4] made a comparative study among different wavelet analysis and used PNN for classification purpose and found out Coiflet 1 is the best wavelet family for signal analysis. Author in [5] presented a wavelet-chaos methodology for analysis of EEGs and delta, theta, alpha, beta, and gamma sub-bands of EEGs for detection of seizure and epilepsy. The methodology was applied to three distinct groups of EEG signals: a) healthy subjects, b) epileptic subjects at the time of a seizure-free interval, c) epileptic subjects at the time of a seizure. For higher frequency beta and gamma sub-bands, correlation dimension (CD) differentiated among the three groups, however for the lower frequency alpha sub-band, the largest Lyapunov exponent (LLE) distinguished among the three groups.

II. PROPOSED METHODOLOGY

A. Proposed Method

The aim of the current work is to find out which is the most appropriate wavelet function. The wavelet function then can be used to extract prominent features from brain signal to design a proficient system for recognition of English alphabet accurately. BIOPAC data acquisition unit (MP36) and AcqKnowledge® 4.0 Software are used to collect data, analysis, store, and recover in the research. The MP System, a computer-based data acquisition system performs many of the same functions as a chart recorder or other data viewing device

[6]. The key phases of the proposed method are presented in Fig. 1.

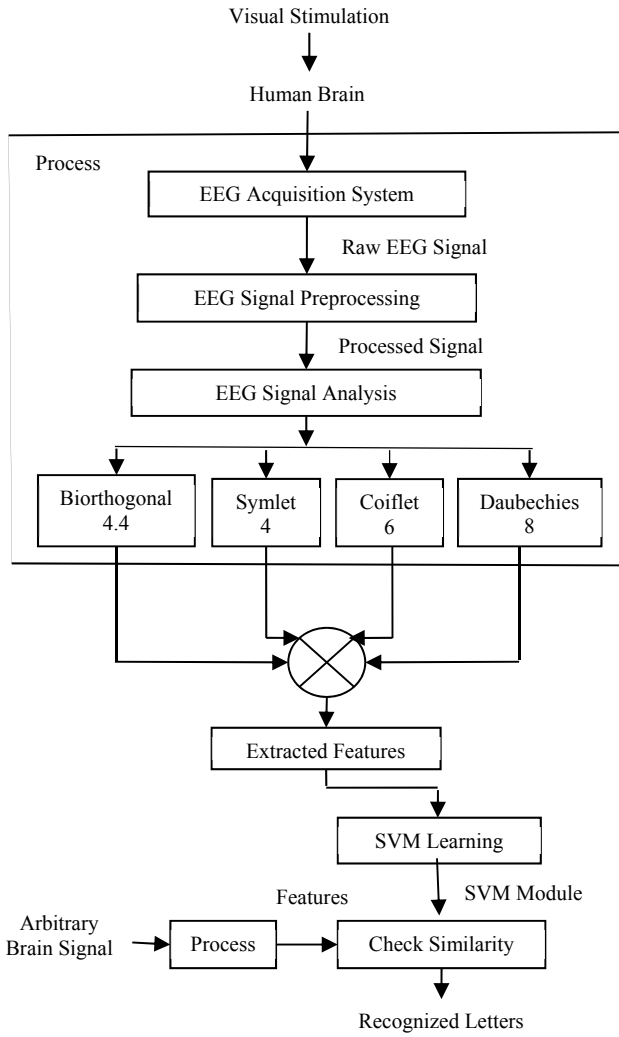


Fig. 1. Proposed flow diagram of methodology

B. Subject Selection

Several healthy students from the Khulna University of Engineering & Technology (KUET), Bangladesh were enrolled in this study. The subjects had not taken any caffeinated or alcoholic drinks before the study. They had not taken any meditation and were in good physical and mental state during experimental sessions. They had normal vision and did not have any problem in communicating. The experiment was approved by local ethics committee.

C. EEG Signal Acquisition

The study was performed in Biomedical Signal Processing Laboratory (BSPL), Department of Biomedical Engineering of this university. In the experiment, EEG recording was performed with electrode lead set connected to data acquisition unit by following standard procedures as shown in Fig. 2. For best electrode connection, the electrodes were placed on the scalp at least 5 minutes before the start of the experiment.

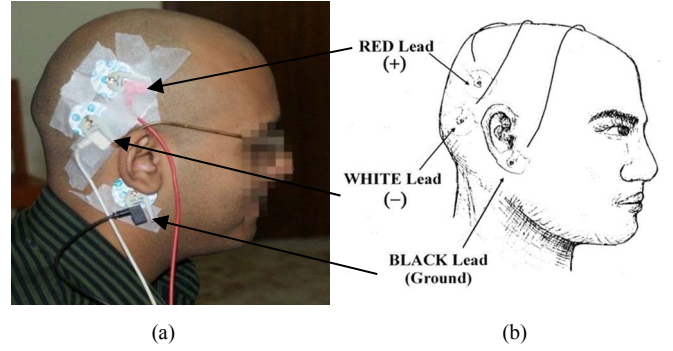


Fig. 2. Electrode Placement Position in (a) Practical Approach and (b) Theoretical View

EEG signal as well as its four spectral bands were recorded from the subjects by showing A to E letters in power point presentation as shown in Fig. 3. The power point slide show for each letter was presented for 5 sec and then there was a gap for 10 sec, wherein the subjects were asked to take rest in sitting position with eyes closed. A single experimental session normally covered 5 trials of complete display of slide show of letters.

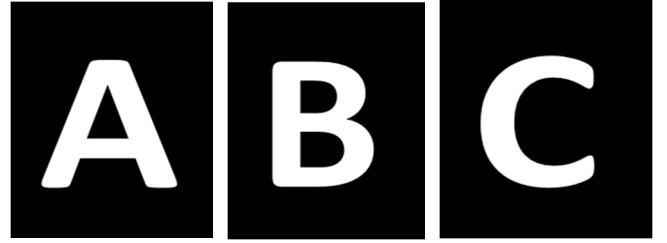


Fig. 3. Snap shot of the power point slide show

D. EEG Signal Preprocessing

Raw EEG signal contains high frequency noise. For the reason signal preprocessing is necessary. Noises such as 50 Hz frequency component can be removed by filtering. Mean value smoothing technique is used here to find smoothed EEG signal. The formula is shown in (1), where the number of points in the window is 'y' and the sample number is 'z'.

$$X_S = \sum_{p=z-\frac{y}{2}}^{p=z+\frac{y-1}{2}} \frac{X_i(p)}{y} \quad (1)$$

E. EEG Signal Analysis

The Wavelet transform has become one of the most significant and powerful tools of signal representation. The wavelet, a smooth, fast vanishing oscillating mathematical function has good localization in both frequency and time. A wavelet family $\varphi_{f,s}(t)$ is a set of fundamental function produced by scaling and translations of a unique mother wavelet, $\varphi(t)$.

$$\varphi_{f,s}(t) = \frac{\varphi\left(\frac{t-s}{f}\right)}{\sqrt{|f|}} \quad (2)$$

Where $f, s \in \mathbb{R}, f \neq 0$ and ' f ', ' s ' are the dilation and translation parameter respectively, ' t ' is the time. The dilation parameter, ' f ' decides the frequency of oscillation and the wavelet length

whereas the translation parameter, 's' decides the shifting position [4].

SVM classifier is a machine learning algorithm which is used for letter recognition from human body signal. SVM task usually involves producing a model based on the training data which estimates the target values of the test data given only the test data features. Multiclass linear kernel is used here which is an auto scaling process and sigma value is one. The concept of MKL is at first recommended in SVM classification, where the original goal is as follows,

$$\begin{aligned} \min_{\bar{w}, b, \xi} \quad & \frac{1}{2} \bar{w}^T \bar{w} + C \sum_{i=1}^l \xi_i \\ \text{Subject to} \quad & y_i (\bar{w}^T \varphi(\bar{x}_i) + b) \geq 1 - \xi_i \\ & \xi_i \geq 0 \end{aligned} \quad (3)$$

Where $\bar{x}_i$ are data samples, $\varphi(\bar{x}_i)$ is the feature map, y_i are class labels, $C > 0$ is a positive (+ve) regularization parameter, ξ_i are slack variables, $\bar{w}$ is the norm vector of the splitting hyper plane, and b is bias [7].

III. RESULTS AND DISCUSSION

The recorded EEG signal with its four frequency bands for a typical subject is shown in Fig. 4.

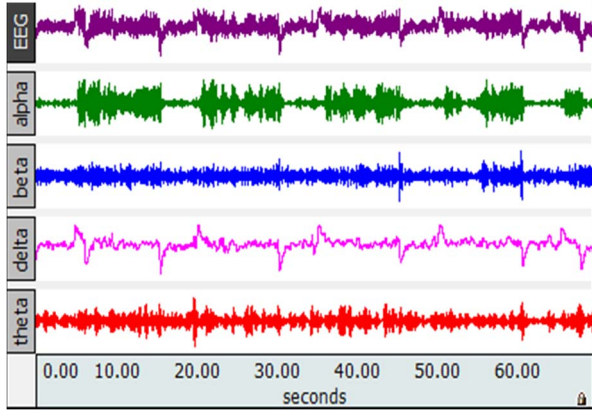


Fig. 4. Raw EEG signal with its different bands

Here, different wavelet methods are used for the purpose of decomposition of the raw EEG signal for extracting features. Then the signals are classified by using SVM. The extracted features are the best characteristics of each letter. These characteristics of each letter are compared with testing signal to categorize unknown letters. The number of decomposition levels which play a significant role in optimization process are selected to be four taking principal frequency components of EEG signal as base. From the plot of different features of 4th wavelet co-efficient sub-band (D₄) Daubechies 8, it can be seen that there exists distinctive pattern for each letter. Other wavelet coefficient sub-bands D₁, D₂, D₃ and A₄ showed overlapping problems. Among the features kurtosis and amplitude show the largest dispersion in magnitude which vary with observation of different letters as shown in Fig. 5 (a). Moreover, from Fig. 5 (b) we can find that mean and skew vary in the range of -0.25 μ V to 1.22 μ V and -13.35 μ V to 8.91 μ V respectively distinguishing each letter successfully.

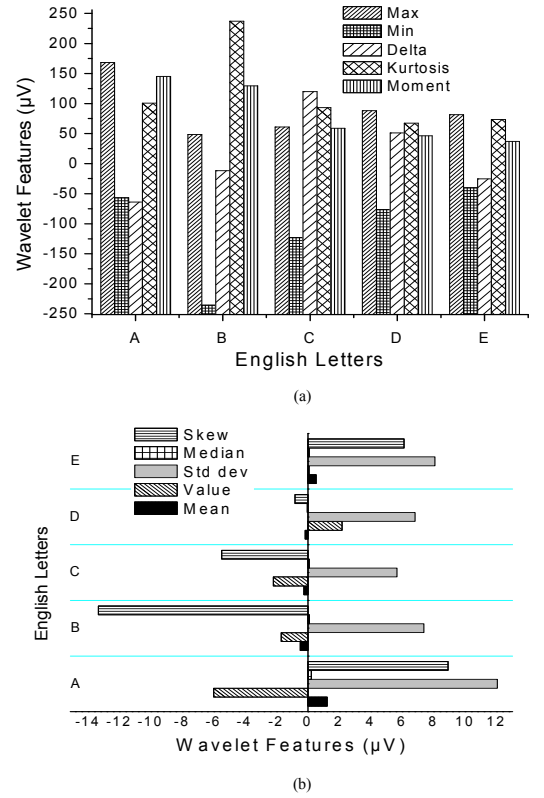


Fig. 5. Deviations in different features (a) amplitude, delta, kurtosis, moment and (b) mean, skew, value, std dev and median over the set of detailed wavelet coefficients at the fourth-decomposition level (D₄) for Daubechies 8

The extracted features from different wavelet analysis have been used for SVM classifier as a training data set to find accuracy rate of different letters as shown in Table II.

TABLE II. ACCURACY OF DIFFERENT WAVELET ANALYSIS USING SVM

Letter	Coiflet analysis		Symlet analysis		Biorthogonal analysis		Daubechies analysis	
	Individual Accuracy (%)	Average accuracy (%)	Individual Accuracy (%)	Average accuracy (%)	Individual Accuracy (%)	Average accuracy (%)	Individual Accuracy (%)	Average accuracy (%)
A	45	78.45	55.56	84.21	50	77.78	60	90
B	53.04		60		66.67		62.5	
C	61		66.67		75		71.43	
D	71		50		60		75	
E	82.48		81.25		83.33		85.71	

For Coiflet and Biorthogonal analysis overall accuracy is 78.45% and 77.78% respectively. Daubechies analysis shows highest overall accuracy and it is 90%. Letter E shows highest recognition rate for all wavelet analysis. In our work Daubechies wavelet analysis has been proved as the best analysis method.

As maximum value gives better performance in separating letters from Fig. 5, the feature is used to find the most active rhythm for the proposed system. The existence and amplitudes

of delta rhythms are highly variable within and between individuals. While focusing on electrical brainwave activity, as recorded by EEG and analyzed for α , β , δ and θ rhythms for single subject as shown in Fig. 6, the letters can be recognized with high degree of accuracy by delta rhythm. It is clear from the figure that other brain rhythms (alpha, beta, theta) failed to differentiate the letters.

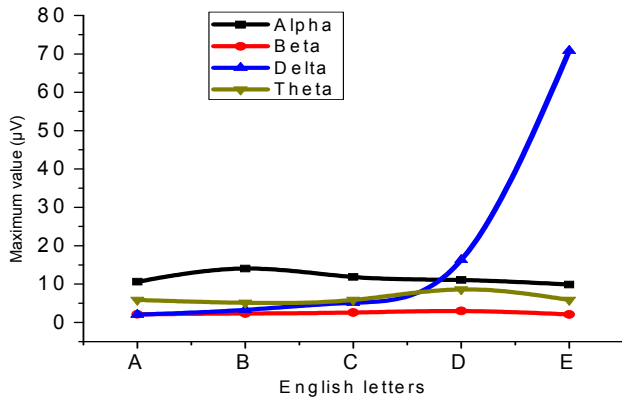


Fig. 6. Plot of peak value of wavelet analyzed signal of α , β , δ and θ rhythm

In Table III, we have tried to find the superior rhythm for the proposed method using SVM classifier. DWT analyzed signals for different rhythms have been investigated for this purpose. Two states have been specified here for letter recognition. If the letter is recognized, then it is in state 1 and if not, then it is in state 0. From Table I, it is known that the alpha action is inversely proportional to the awareness of an individual. The information of Table III indicates that when the subjects were shown the letters, the recognition rate is only 20% for alpha rhythm and 60% for beta rhythm. As previously mentioned, the delta rhythm appears in signal at time of continuous attention task and looking at the table the desired result is found. For delta rhythm the accuracy rate is 80% which is better than the recognition rate for other rhythms.

TABLE III. OVERALL ACCURACY OF EEG RHYTHMS USING SVM

Rhythm	English Alphabet					Overall accuracy
	A	B	C	D	E	
Alpha	0	0	0	0	1	20%
Beta	1	0	1	0	1	60%
Delta	1	1	0	1	1	80%
Theta	0	1	0	0	0	20%

IV. CONCLUSION

With the analysis of brain signal, the most suitable wavelet was found out for letter recognition purpose. This leads to the enhancement of performance of brain signal analysis in comparison with arbitrary selection of wavelet analysis. From the sub-bands ten features are extracted and then for classification purpose SVM is used. Based on the classification accuracy, Daubechies 8 is found to be the best wavelet family for EEG signal analysis with overall accuracy 90%. Besides among various brain rhythms delta rhythm has best recognition rate (80%). So it can be concluded that Daubechies 8 is the most efficient analysis method for precise classification of EEG signal for letter recognition system.

ACKNOWLEDGEMENT

This work was partially supported by Higher Education Quality Enhancement Project (HEQEP), UGC, Bangladesh under Subproject "Postgraduate Research in BME", CP#3472, KUET, Bangladesh.

REFERENCES

- [1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clinical Neurophysiology*, vol. 113, pp. 767–791, 2002.
- [2] S. Rao, B. Gawal, S. C. Mehrotra, P. Rokade, and R. Deore, "Number recognition system using electroencephalogram (EEG) signals," *Advances in Computational Research*, vol. 4, pp. 66–68, 2012.
- [3] Mohammad H. Alomari, Emad A. Awada, Aya Samaha, and Khaled Alkamha, "Wavelet-Based feature extraction for the analysis of EEG signals associated with imagined fists and feet movements," *Computer and Information Science*, vol. 7, no. 2, 2014.
- [4] Tapan Gandhi, Bijay Ketan Panigrahi, and Sneh Anand, "A comparative study of wavelet families for EEG signal classification" *Neurocomputing*, ELSEVIER, vol. 74, pp. 3051–3057, 2011.
- [5] Hojjat Adeli, Samanwoy Ghosh-Dastidar, and Nahid Dadmehr, "A Wavelet-Chaos Methodology for Analysis of EEGs and EEG Subbands to Detect Seizure and Epilepsy," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 2, 2007.
- [6] Reference manuals of Acqknowledge software guide and MP system hardware guide. (2008). BIOPAC Systems Inc., 42 Aeri Camino, Goleta, CA-0067, USA.
- [7] S. Yu, T. Falck, A. Daemen, L. -C. Tranchevent, J. A. Suykens, B. De Moor, Y. Moreau, "L2-norm multiple kernel learning and its application to biomedical data fusion," *BMC Bioinformatics*, vol. 11, 2010.

Geometrical Approach for Determining the Throwing Destination of an Automatic Throwing Robot

Nahid Uzzaman*, Sanzida Hossain†, Altab Hossain‡

*†Department of Mechanical Engineering, Military Institute of Science and Technology (MIST)

‡Department of Nuclear Science and Engineering, Military Institute of Science and Technology (MIST)
Dhaka, Bangladesh

*nahiduzzaman.eefat@gmail.com, †sanzidahossain92@gmail.com, ‡altab76@gmail.com

Abstract—To accomplish a task automatically without any human intervention, a robot's first step is to collect necessary information from the surroundings as input. Depending on the application, in some cases a robot may need to perform throwing action automatically. For an accurate throw, an unerring destination detection should be executed. In this paper, a method for determining the throwing destination of a robot is presented. Regarded that the robot works in an open terrain and is designed to consider any physical object found within a certain range around it as a potential target. A relative geometrical analysis between throwing arm's rotational axis position, sensor position and sensor data gave certain equations in different cases to directly determine the polar coordinates of the throwing location relative to the arm. This gives the flexibility of placing the sensor in any suitable place. To substantiate the obtained equations, a system for determining the throwing position was developed using the derived method with same preconditions. Satisfactory performance of the system with good experimental results was found, which upholds the validity of the derived method.

Keywords—throwing robot, geometrical analysis, throwing destination, polar coordinate.

I. INTRODUCTION

Nowadays a robot can perform almost any job assigned to it. Robots are used in industries for handling material, welding, cutting, sealing with different types of control [1]. Not only in industries, robots are also being used for military [2], gaming [3], study [4] and as well as household activities [5]. For doing these works robots nowadays should be modernized with various features. Such a feature is throwing capability of a robot [6-7]. A throwing robot can be used for several purposes. In can be used in military purpose for throwing bombs or transportation of goods [8-9]. Can be useful in disasters for throwing useful things to some place where a human cannot be physically present.

If a robot is to perform throwing action automatically, i.e., deciding where to throw the concerning object and throwing accurately to that position, the robot must be incorporated with a predefined throwing position locator technique and a precise throwing mechanism. This paper presents a method for

determining the throwing destination. This method is valid for a throwing robot which works in an open terrain. The robot is designed to throw towards any object found within a certain range around the robot. Keeping in mind such considerations, the system is designed and destination detection method is developed. Another method that can be used for such detection is visual feedback [10]. Though our method is designed only for a particular scenario, it uses only a single sensor for determining 2-D location in polar coordinates. Which makes it cost effective for that particular use.

II. DESIGN AND WORKING PRINCIPLE OF THE ROBOT

The design of the robot is the most prior thing to be concerned with. All other considerations come in accordance with the design. Based on the functional necessity and working environment a robot should be designed. Necessary calculations are made according to this, because the considerations depend on the geometry of the system. So, while designing the robot, every aspect should be kept in mind.

For our consideration we had to design the robot so that we can perform our analysis based on its geometry. Fig. 1. This study emphasizes more on the working principle. Fig. 2. Material property and mechanical design is not matter of concern. It is a stationary robot. The robot has a base arm that holds throwing arm and can move with the help of a base servo motor. The throwing arm comprises of two servo motors. One gives the thrust for throwing action and another is used for controlling the grip of the throwing object.

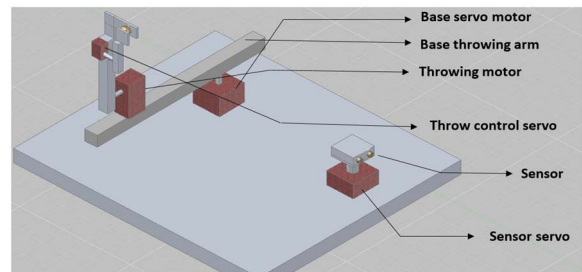


Fig. 1. Robot structure and mechanisms showing different components.

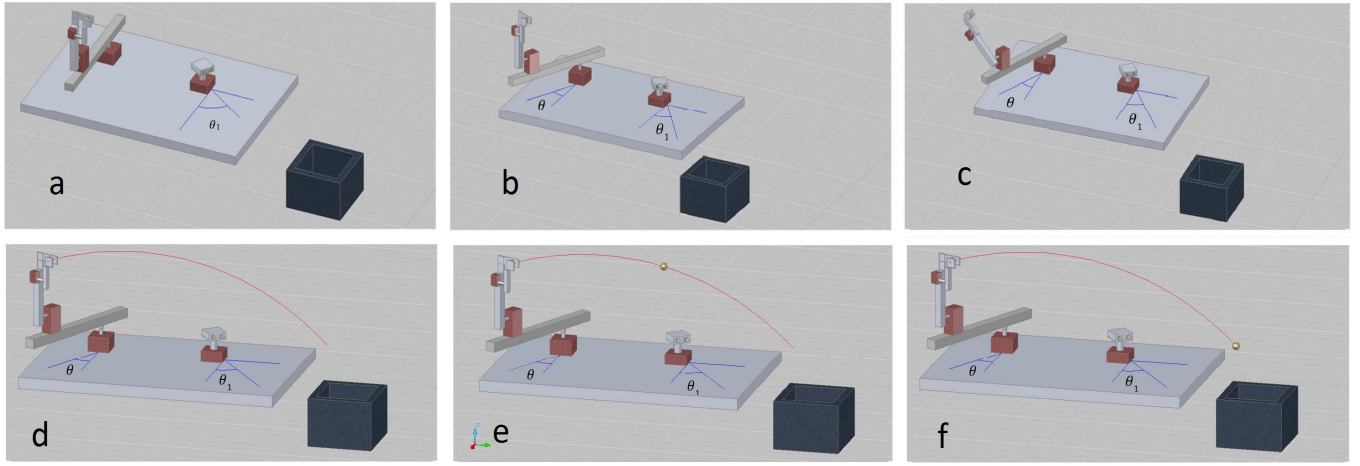


Fig. 2. Steps of throwing of the robot (a) destination detection by the sensor, (b) base arm taking required position, (c) preparing for throw (d) projectile of the throw (e) throwing object traveling towards the destination (f) arriving at the destination

Our design also incorporates another servo motor and a distance measuring sensor, which are used to determine the location. Whenever the sensor measures distance less than a certain value it implies that something is present within the range which is considered as throwing destination. The system then calculates the distance of the object from the throwing arm and the angle that the arm needs to rotate from the reference angle. Though initially the distance and angle is measured with respect to sensor, it is converted into required values with relative geometrical analysis.

The sensor servo causes the sensor to rotate all the time. That allows detecting any object found within the 180° angle. As can be seen from Fig. 2. If the sensor finds any object around, the controller immediately calculates the distance of the object from the sensor and records the angle of servo at which the object is found. This distance and angle is then used to calculate the distance of the object from the throwing arm, and the angle at which the arm needed to position itself for throwing in that location.

III. GEOMETRICAL ANALYSIS FOR LOCATING THE THROWING DESTINATION

If the sensor detects any obstruction in the travel path of the ultrasonic sound wave it sends the signal to the controller unit. In the meantime the sensor servo motor rotates randomly with successive angles. But the angle found is the displacement angle of the sensor. But the throwing arm is not in the same position that of the sensor. These data can be used to measure the angle at which the lower arm need to rotate.

The lower arm can only move in a circle. So to throw in various position the line of throwing should be the tangent of the circle in which the lower arm moves.

Keeping in mind all these consideration we can start our analysis to determine the angle of rotation of lower arm. The analysis is made for two different conditions. One is when the throwing destination is located in the left side of robot and the other is when the location is at right.

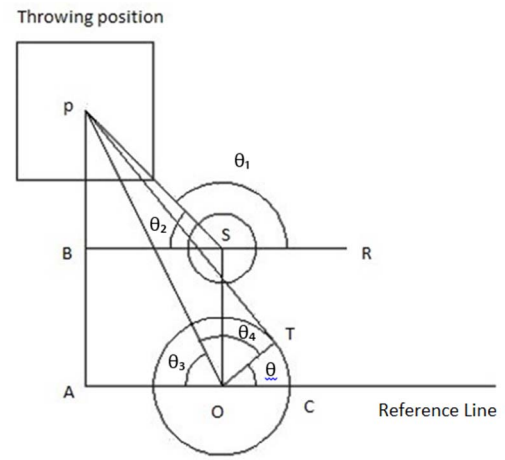


Fig. 3. When the throwing destination is located at the left side of robot.

Now, for the left side position, from Fig. 3.

P = location of the throwing point, T = throwing mechanism, S = sensor, SR = reference line of the sensor servo, OT = r = lower arm length, θ = angle at which lower arm need to rotate, θ_1 = angle of the sensor servo motor, PS = a = distance measured by sensor, AB = x = perpendicular distance between sensor and base servo.

$$PB = PS \cdot \sin\theta_2 = a \cdot \sin\theta_2 \quad (1)$$

$$PB = a \cdot \sin(180 - \theta_1)$$

$$\text{So, } PB = a \cdot \sin\theta_1 \quad (2)$$

$$BS = a \cdot \cos(180^\circ - \theta_1)$$

$$BS = -a \cdot \cos\theta_1 = AO \quad (3)$$

$$\text{Now, } PA = PB + AB$$

$$\text{So, } PA = a \cdot \sin\theta_1 + x$$

From ΔAOP ,

$$PO = \sqrt{PA^2 + AO^2}$$

$$\text{So, } PO = \sqrt{(a \sin \theta_1 + x)^2 + (-a \cos \theta_1)^2} \quad (4)$$

$$\theta_3 = \tan^{-1} \frac{PA}{AO}$$

$$\theta_2 = \tan^{-1} \frac{a \sin \theta_1 + x}{-a \cos \theta_1} \quad (5)$$

Now, from ΔPOT ,

$$PT = \sqrt{PO^2 - OT^2}$$

$$\text{Or, } PT = \sqrt{(a \sin \theta_1 + x)^2 + (-a \cos \theta_1)^2 - r^2} \quad (6)$$

Here, PT = Throwing distance = d

$$\theta_4 = \cos^{-1} \frac{OT}{PO}$$

$$\theta_4 = \cos^{-1} \frac{r}{\sqrt{(a \sin \theta_1 + x)^2 + (-a \cos \theta_1)^2}} \quad (7)$$

So, the required angle,

$$\theta = 180^\circ - (\theta_3 + \theta_4)$$

$$\theta = 180^\circ - \tan^{-1} \frac{a \sin \theta_1 + x}{-a \cos \theta_1} - \cos^{-1} \frac{r}{\sqrt{(a \sin \theta_1 + x)^2 + (-a \cos \theta_1)^2}} \quad (8)$$

For the right side position, from Fig. 4.

From ΔPSB ,

$$SB = PS \cdot \cos \theta_1$$

$$SB = a \cos \theta_1 = AO \quad (9)$$

$$PB = a \sin \theta_1 \quad (10)$$

$$PA = PB + AB$$

$$PA = a \sin \theta_1 + x \quad (11)$$

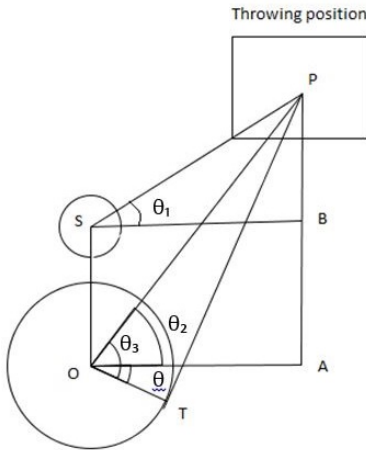


Fig. 4. When the throwing destination is located at the right side of robot.

Now from ΔPOA ,

$$PO = \sqrt{PA^2 + AO^2}$$

$$PO = \sqrt{(a \sin \theta_1 + x)^2 + (a \cos \theta_1)^2} \quad (12)$$

$$\theta_2 = \tan^{-1} \frac{PA}{OA}$$

$$\theta_2 = \tan^{-1} \frac{a \sin \theta_1 + x}{a \cos \theta_1} \quad (13)$$

From ΔPOT ,

$$\theta_3 = \cos^{-1} \frac{OT}{PO}$$

$$\theta_3 = \cos^{-1} \frac{r}{\sqrt{(a \sin \theta_1 + x)^2 + (a \cos \theta_1)^2}} \quad (14)$$

Again,

$$PT = \sqrt{PO^2 - OT^2}$$

$$PT = \sqrt{(a \sin \theta_1 + x)^2 + (a \cos \theta_1)^2 - r^2} = d \quad (15)$$

Required angle,

$$\theta = \theta_3 - \theta_2$$

$$\theta = \cos^{-1} \frac{r}{\sqrt{(a \sin \theta_1 + x)^2 + (a \cos \theta_1)^2}} - \tan^{-1} \frac{a \sin \theta_1 + x}{a \cos \theta_1} \quad (16)$$

Here, (6) and (15) are the same equations. So polar coordinates of the throwing location can be directly determined by (6), (8) and (16).

IV. CONSTRUCTING THE MEASURING SYSTEM

After An actual system with all the necessary conditions was made to realize the effectiveness of the derived equations. The system is shown in Fig. 5. As the design demands, a distance measuring sonar sensor is placed on the rotating head of the servo motor. The sensor and servo motor is controlled by an Arduino Uno controller unit. After reading the sensor data and servo motor position the controller calculates the position of the required destination by using (6), (8) and (16). Determined values by the controller can be seen on the serial monitor of Arduino software.

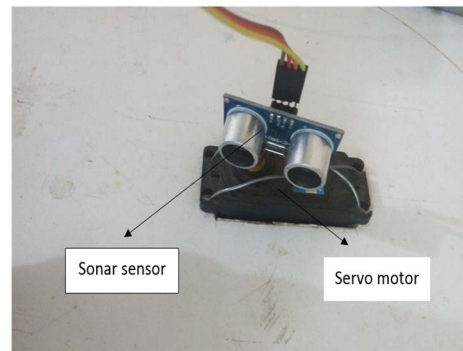


Fig. 5. Measuring system for determining distance and angular position of throwing location

TABLE I. SYTEM MEASURED AND MANUALLY MEASURED DISTANCE AND ANGLE

Ob s. No.	Distance measured			Angle measured		
	By sensor, a (meter)	By the system, d (meter)	Manual (meter)	By sensor, θ_1 (degree)	System, θ (degree)	Manual (degree)
1	1.15	1.378793	1.39	50	23.7907	22
2	1.36	1.540602	1.51	35.5	38.0581	38
3	2.21	2.480865	2.52	67	15.6915	14.5
4	1.06	1.340038	1.31	102	17.8501	19
5	1.32	1.546005	1.56	131	41.1179	39
6	0.9	1.085083	1.08	144	51.7362	55.5

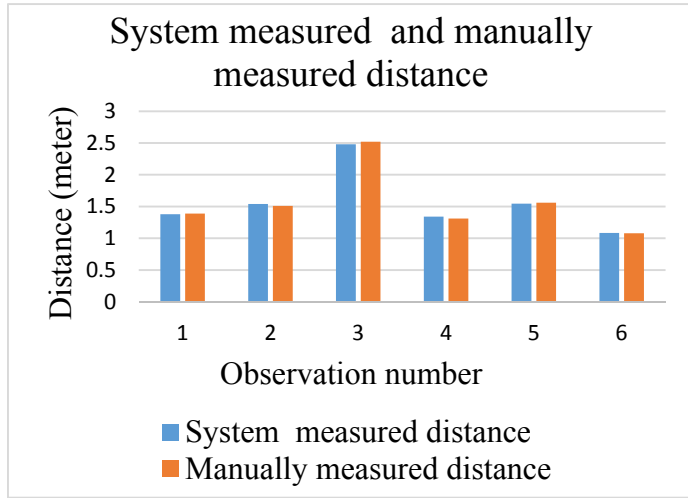


Fig. 6. Comparison between system measured and manually measured distance.

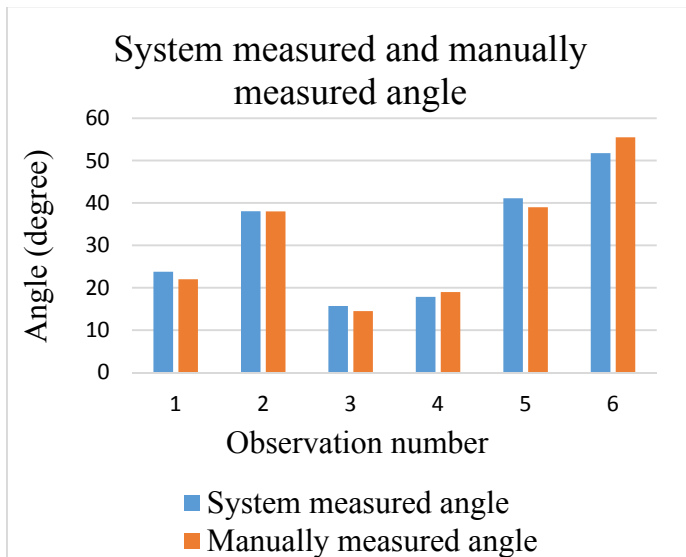


Fig. 7. Comparison between system measured and manually measured angle.

V. RESULTS OF PERFORMANCE TEST AND DISCUSSION

For the experiment, an object within the range of the system was set. The controller calculates throwing distance and angle directly from the sensor data and sensor servo angle using (6), (8) and (16). These data are collected from the Arduino serial monitor. Manually measured data was then compared with the system data.

As can be seen from TABLE I, Fig. 6. And Fig. 7. System measured and manually measured data are almost same. Only a little difference is found in all the observations. The absolute error in throwing distance from observation 1-6 is 0.011207, 0.0306, 0.039135, 0.03004, 0.013995, 0.00508 meters respectively. Also, the absolute error in throwing angle is 1.7907, 0.05811, 1.19152, 1.149904, 2.11792, 3.763795 degree respectively.

For throwing distance the percent error is within 2% for all the observations. However, for angle calculation it is slightly more for some of the observations. Observation 1 and 3 incorporates 7% of percent error and for observation 6 it is 8%. Though there are notable errors in angle calculation, these errors will not have significant domination in the actual robot. The error in data are due to the incorporated faults in the manual measurements. The manual measurement was not completely accurate and some error in data are included there. Specially measuring angle accurately was very difficult. Because it is not possible to locate the exact point precisely the sensor is taking data from. Also, precise positioning and efficiency of the sensor is a factor that causes error. Thus, these errors will not be dominant in a robot with accurate actuation. Because the robot will consider the same point sensor is locating. Whereas for a human being it is very difficult to find out that point correctly, which causes deviation of the results.

VI. FUTURE UPGRADE AND RECOMMENDATION

In our study we tried to work with an open terrain scenario. However this arrangement can further be improved to work in a more complex environment with many other object other than the throwing target. Image processing, fuzzy clustering algorithm [11] etc. can be used to recognize the target object in a more complex orientation. Once the target is recognized, our method can be used to determine the location and execute a throw.

For achieving better performances sensors, actuators and controller should have fast response capability. Otherwise substantial error will result.

VII. CONCLUSION

For an automatic throwing robot to function, target destination detection is the first step to begin with. We have designed a stationary throwing robot and performed some geometrical analysis to determine the throwing destination for it. An actual measuring system that uses the equations derived in the analysis was developed for the concerning robot. Performance test of the system was carried out, that produced

data for different object position. Comparison of manually measured value of the variable of interest with the system measured values shows satisfactory results. The design of the system can be improved in future by further research to handle more complex situations and to hold good in real time activities.

REFERENCES

- [1] T. Thomessen, L. Nagy and K. Olah, "Remote control of industrial robot systems using cognitive info communication," *Cognitive Infocommunications (CogInfoCom)*, 2011 2nd International Conference on, Budapest, 2011, pp. 1-6.
- [2] S. Naskar, S. Das, A. K. Seth and A. Nath, "Application of Radio Frequency Controlled Intelligent Military Robot in Defense," *Communication Systems and Network Technologies (CSNT)*, 2011 International Conference on, Katra, Jammu, 2011, pp. 396-401.
- [3] E. Park, K. J. Kim, S. S. Kwak and A. P. del Pobil, "Robot as teammate vs. opponent in video gaming," *Humanoid Robots (Humanoids)*, 2011 11th IEEE-RAS International Conference on, Bled, 2011, pp. 47-52. doi: 10.1109/Humanoids.2011.6100811
- [4] R. Radharamanan and H. E. Jenkins, "Robot Applications in Laboratory-Learning Environment," 2007 Thirty-Ninth Southeastern Symposium on System Theory, Macon, GA, 2007, pp. 80-84.
- [5] N. Suzuki *et al.*, "System assisting walking and carrying daily necessities with an overhead robot arm for in-home elderlies," *Engineering in Medicine and Biology Society, 2000. Proceedings of the 22nd Annual International Conference of the IEEE*, Chicago, IL, 2000, pp. 2271-2274 vol.3
- [6] H. Frank, A. Mittnacht, T. Moschinsky and F. Kupzog, "1-DOF-robot for fast and accurate throwing of objects," 2009 IEEE Conference on Emerging Technologies & Factory Automation, Mallorca, 2009, pp. 1-7.
- [7] H. Frank, T. Frank, A. Mittnacht and C. Sichau, "A bioinspired 2-DOF throwing robot," *AFRICON, 2011*, Livingstone, 2011, pp. 1-6.
- [8] H. Frank, "Throwing of Axial-Symmetric Objects in Production Systems," *Computer Modelling and Simulation (UKSim)*, 2010 12th International Conference on, Cambridge, 2010, pp. 1-1.
- [9] H. Frank, "Throwing of Objects: A New Technology for Factory Automation," *Computational Intelligence, Communication Systems and Networks, 2009. CICSYN '09. First International Conference on*, Indore, 2009, pp. 4-5.
- [10] J. S. Hu, M. C. Chien, Y. J. Chang, S. H. Su and C. Y. Kai, "A ball-throwing robot with visual feedback," *Intelligent Robots and Systems (IROS)*, 2010 IEEE/RSJ International Conference on, Taipei, 2010, pp. 2511-2512.
- [11] A. Del Amo, D. Gomez and J. Montero, "Spectral fuzzy classification system for target recognition," *Fuzzy Information Processing Society, 2003. NAFIPS 2003. 22nd International Conference of the North American*, 2003, pp. 495-499.

A Comparative Analysis Between Light Dependent and Ultrasonic Method of Gesture Recognition

Nahid Uzzaman^{*}, Sanzida Hossain[†], Rezaur Rashid[§], Altab Hossain[‡]

^{**†}Department of Mechanical Engineering, Military Institute of Science and Technology (MIST)

[§]Department of Computer Science and Engineering, BRAC University (BRACU)

[‡]Department of Nuclear Science and Engineering, Military Institute of Science and Technology (MIST)
Dhaka, Bangladesh

^{*}nahiduzzaman.eefat@gmail.com, [†]sanzidahossain92@gmail.com, [§]aadrick045@gmail.com, [‡]altab76@gmail.com

Abstract—In this paper, two methods of gesture recognition are discussed and a comparison between these methods is carried out. These are light dependent and ultrasonic method of gesture recognition. The light dependent method was developed in an early study, where a control panel detected some predefined gesture commands and a robotic arm was controlled with those commands. In this new study the same gesture commands are used to control the robotic arm. But this time they are determined by both light dependent and ultrasonic method. All the procedure and arrangements are kept unchanged so that a comparative performance test for these two methods can be carried out. Experimental results and cost analysis shows that the light dependent method is more convenient for indoor application of the system. Whereas, for outdoor application, the ultrasonic method is preferable.

Keywords—gesture recognition, comparison between gesture recognition methods, light dependent method, ultrasonic method, HMI.

I. INTRODUCTION

Human-machine interaction (HMI) is the process of communication between a machine and human being. A great number of researches are done on HMI [1-2]. A machine that needs operator control generally needs HMI. It allows the operator to give specific signals to do certain works. There are various methods of HMI such as switch controlled, remote controlled [3-4], gesture controlled [5-6], speech controlled [7-8], brain controlled [9] etc. Gesture recognition is a kind of HMI that reads the movement of the body part of the operator and recognizes a specific predefined signal. There are various methods for gesture recognition on which researches have been done, such as vision based gesture recognition [10], accelerometer-based gesture recognition [11] and gesture recognition based on fusing frame images [12].

In our previous work we have studied light dependent method of gesture recognition [13]. In this paper another method is presented and relative comparison of performance and effectiveness between these two methods are discussed with experimental evidence. To perform this current study the same setup and working procedure of our previous work is used. Both the methods have their own technique of recognizing the gesture. The light dependent method is more

suitable for indoor application. Because it works better in a fixed lighted environment, whereas the lighting condition of outside varies due to the effect of sun light. On the other hand, the ultrasonic method holds good for both indoor and outdoor conditions because it does not depend on the lighting conditions.

However for a fixed lighted environment, performance of the light dependent method is quite satisfactory. So for a fixed lighted environment any of the mentioned methods can be used. But one important factor is the cost of using the method in the concerning environment. As flexibility of using any of the methods is found from the experiment, the method that is preferred by the cost analysis should be selected.

II. METHODOLOGY FOR GESTURE RECOGNITION

Methodology of gesture recognition with light dependent and ultrasonic method is completely different. Light dependent method uses the concept of light intensity variation to determine gesture indicated. Whereas the ultrasonic method uses variation in travel time of the ultrasonic sound reflection to determine the indication. Both of these methods are discussed below along with their working principles.

A. Light dependent method of gesture recognition

In the preceding study [13] we used light-dependent resistor (LDR) sensor. It is a photo resistor whose resistance decreases with increasing incident light intensity. Thus the voltage drop across the sensor is increased by reducing the light intensity and vice versa. When the operator hand is casted, light intensity changes which causes voltage variation. With the varying voltage the controller predicts that something is going on. This phenomenon is used to determine the position of the hand of the operator ahead of the sensor panel. Fig 1.

For gesture recognition first the voltage drop (V_o) is measured with open sensor. Then hand is positioned with the same distance from the sensor in which the gesture pattern will be indicated and the voltage drop (V_h) reading is taken. The reference voltage (V_{ref}) is calculated from average of these two values.

$$V_{ref} = \frac{V_o + V_h}{2} \quad (1)$$

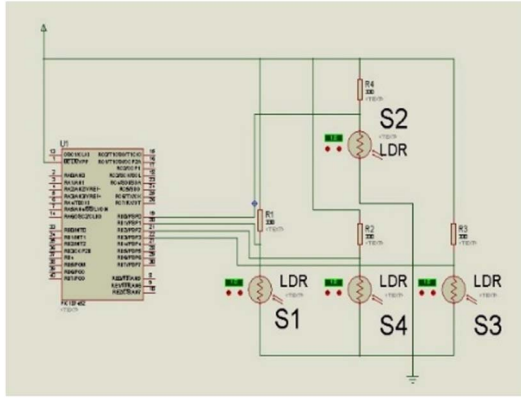


Fig. 1. Circuit connection of LDR sensor.

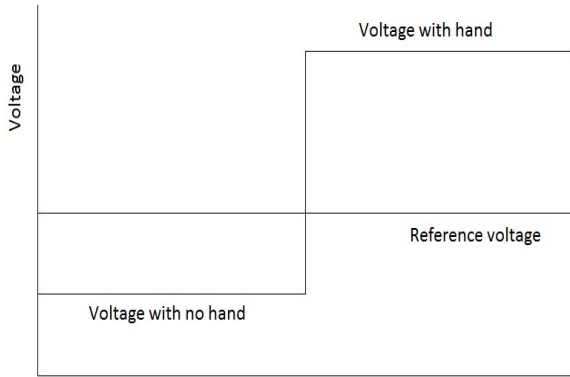


Fig. 2. Voltage variation with hand casting across the reference value.

During operation, the sensor will have quite a constant voltage output all the time. But when the operator will cast his hand on the sensor, voltage will increase. The controller compares the reference voltage with the sensor voltage output as shown in Fig. 2. If the current sensor output voltage is greater than the reference voltage the controller decides that the operator hand is casted.

B. Ultrasonic method of gesture recognition

For this method sonar sensor is used. A sonar sensor converts ultrasonic sound into electrical signal or vice versa. In our application the sonar sensor is used to determine distance of the operator hand. The travel time of the ultrasonic sound is converted into distance. Thus sensor data continuously gives the distance (d_{sensor}) of any object found in front of the sensor. With this phenomenon our gesture recognition method is designed. Fig. 3.

Likewise the light dependent method, a reference value is used. But this time the value is distance. It is better to take reference distance (d_{ref}) as double of the intended distance (d_{sensor}) from the sensor. With this reference value the operator need not be very accurate with the distance.

$$d_{\text{ref}} = 2 \times d_{\text{sensor}} \quad (2)$$

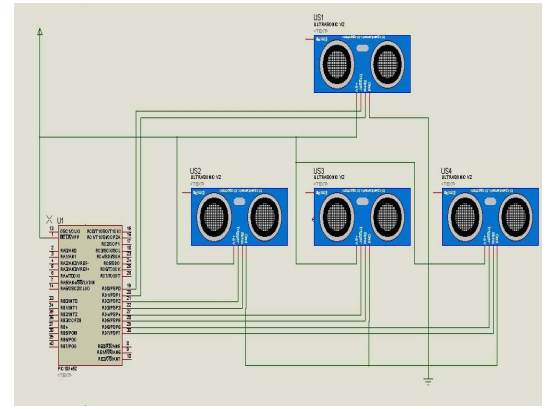


Fig. 3. Circuit connection of Sonar sensor.

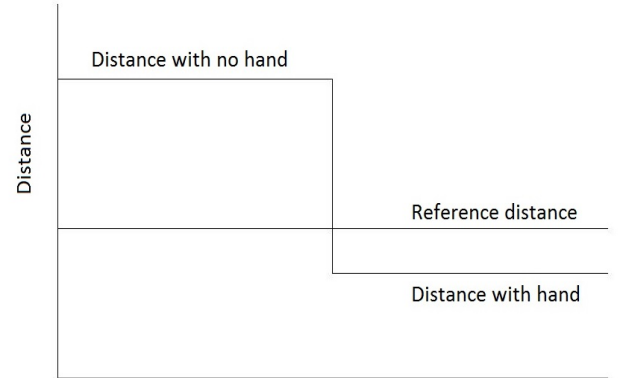


Fig. 4. Distance variation with hand casting across the reference value.

During operation, the sonar sensor continuously gives a distance value (d_{sensor}). When the operator cast his hand this value will be reduced. The controller compares this value with the reference value (d_{ref}). If the d_{sensor} is found less than d_{ref} operator hand will be considered at the location of the sensor by the controller as shown in Fig. 4.

For light dependent method current voltage has to be greater than reference voltage to locate any hand casting. Because the voltage increases with hand casting. But for ultrasonic method current output distance has to be smaller to locate hand casting. Because the sensor always measures a greater distance. When hand is casted the distance will be smaller.

III. PATTERN RECOGNITION

For our analysis we chose some simple and predefined gesture patterns and followed the procedure of pattern recognition of foregoing study [13]. We used a gesture detection panel that incorporates 4 sensors in it. The gesture detection panel is shown in Fig. 5(a). The sensors are named s-1, s-2, s-3 and s-4.

With these 4 sensors in the panel a system is developed to determine 6 different gesture patterns. The patterns are clockwise-anticlockwise, upward-downward and right-left movement of the hand, shown in Fig. 5(b), 5(c) and 5(d).

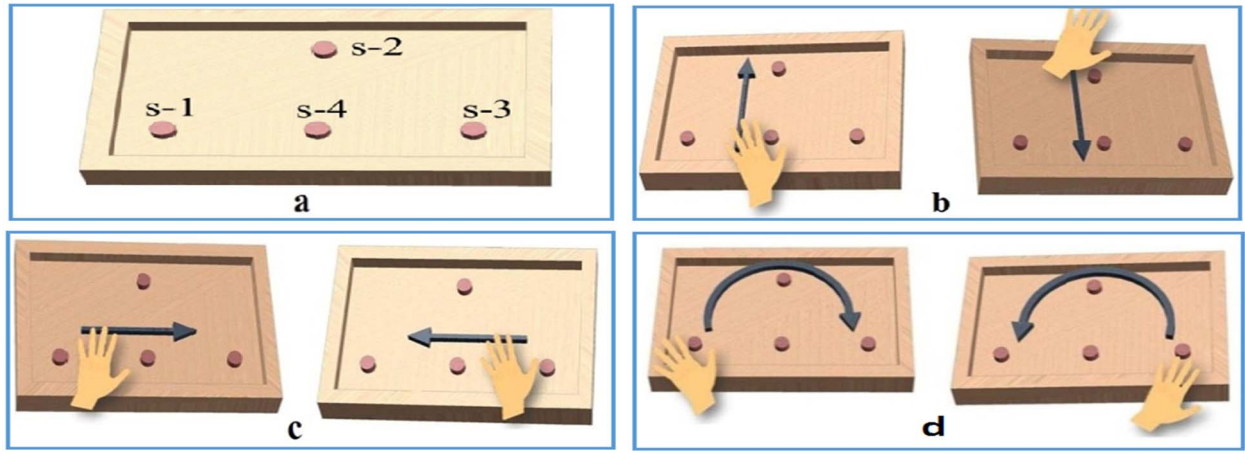


Fig. 5. (a) Arrangement of four sensors, Gesture pattern indication by moving the hand in directions: (b) upward-downward, (c) right-left directions and (d) clockwise-anticlockwise.

This particular design and gesture patterns are used for both light dependent and ultrasonic method. Whenever the operator cast any pattern indication in the detection panel, controller will detect hand comparing sensor data and reference values. The movement of hand to indicate pattern will cause different sensors to locate hand at different times. The controller knows the location of the concerning sensors. Thus by recording the consecutive sequence of sensor number in which sensor value was found less than reference, the controller can map the gesture pattern. After recognizing the command the controller can execute necessary action that was predefined for this pattern. For our study this particular arrangement of sensors will remain fixed. However, using our described procedure any gesture pattern can be identified with different arrangement of the sensors.

To have a visible outcome of the recognition system, it was connected with a robotic arm [13] or lifter. The robotic lifter executes different movements according to the gesture commands. Fig. 6. Motor m-1 of the robotic lifter gives clock wise and anticlockwise movements, motor m-2 gives up-down motion and motor m-3 opens and closes the grip with right left gesture commands.

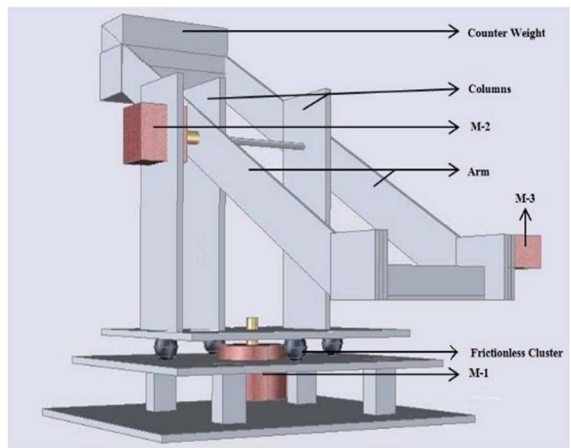


Fig. 6. Robotic lifter.

This lifter has no impact on the methods. It only gives indication whether a successful gesture recognition is carried out or not. Though malfunction in the lifter can occur. As a successful gesture recognition is understood by the functioning of the lifter, when any malfunction occurs it will be difficult to determine whether there is a problem in gesture recognition or in the lifter.

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

For light dependent method, Output voltage of the sensors varies with lighting conditions. So for outdoor and indoor conditions, variation of voltage with varying hand distance is measured in our previous experiment [13]. Variation of the voltage in daylight (0.5V-1.15V) was less than that of indoor fixed lighting condition (3.1V-4.95V). The large difference in operating voltages makes the selection of reference voltage easier for the indoor condition. Because the reference value is the average of highest and lowest values. If there is no significant difference between these values the system will get confused and will not work correctly.

However, the ultrasonic method can be used both in outdoor and indoor environments. Because it directly measures the distance of the hand of the operator from the sensor. But with the increasing distance the system faces difficulties in recognizing the gesture patterns. Because when the distance of the hand is significantly high from the control panel it becomes difficult to cut the travel path of the ultrasonic wave. If the ultrasonic wave misses the hand it will not be reflected and the controller will not identify anything.

To understand the effect of hand distance in the success of recognition for both the methods, the system was commanded with gesture patterns randomly. After indicating any gesture pattern the lifter is checked to determine whether the intended task is performed or not. Thus, for expected actuation of the lifter, the gesture recognition is considered successful. For a particular hand distance, 30 such observation are recorded and percentage of successful gesture recognition is calculated. Gradually the distance is varied and the performance is observed in the same manner.

TABLE I. PERCENTAGE SUCCESS OF GETURE RECOGNITON WITH VARYING HAND DISTANCE

Hand distance (inches)	Percentage success of light dependent method at indoor condition (%)	Percentage success of light dependent method at outdoor condition (%)	Percentage success of ultrasonic method (%)
3	93.33	53.33	100
6	80	3.33	86.67
9	56.67	0	50
12	13.13	0	23.33

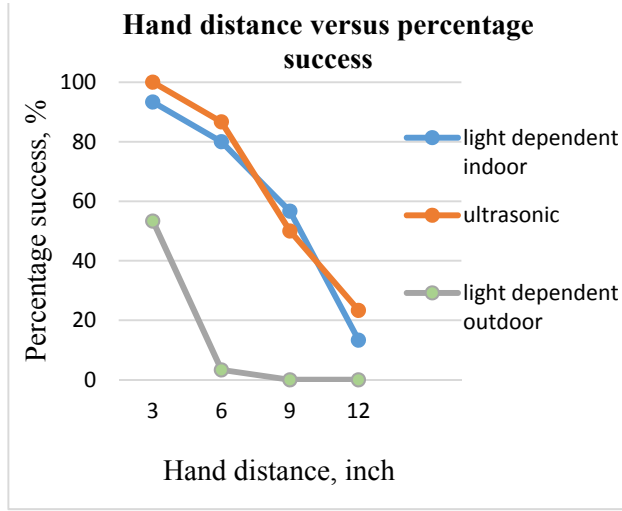


Fig. 7. Hand distance versus percentage success.

The pattern indication may not be recognized by the controller if the pattern is indicated too fast. But it should not take too much time as well. For our system in each of the observations, patterns are indicated within 2 seconds.

From Fig. 7, it is clear that, performance of light dependent method at indoor condition and performance of ultrasonic method is quite similar. The success rate for both those condition is almost same. Both has a very high success rate for gesture pattern indication from a small distance. Whereas outdoor application of light dependent method has a very poor success rate. For a very small gap between the hand and control panel it has only 53.33% success rate and with increasing distance, it becomes completely nonfunctional. Thus for indoor application either light dependent or ultrasonic method can be used depending on which is preferred by the cost analysis. But for outdoor application only ultrasonic method will be able to function properly.

From market investigation it is found that on average the cost of an ultrasonic sensor is 0.99\$. Whereas the cost of a light dependent LDR sensor is 0.05\$ on average. As the controller and other elements of the system remain same for both the methods, cost of them is not considered.

TABLE II. %SUCCESS TO COST RATIO OFGETURE RECOGNITON WITH VARYING HAND DISTANCE

Hand distance, inch	Cost to %success ratio for light dependent method at indoor, \$	Cost to %success ratio for light dependent method at outdoor, \$	Cost to %success ratio for ultrasonic method, \$
3	0.000535714	0.0009375	0.0099
6	0.000625	0.015	0.011423077
9	0.000882353	∞	0.0198
12	0.00375	∞	0.042428571

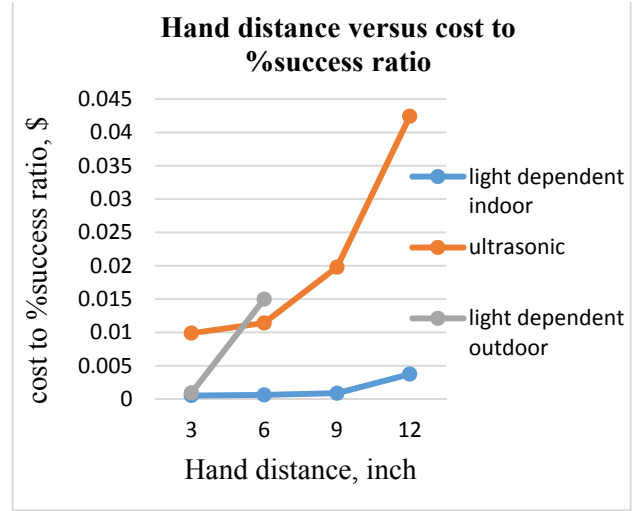


Fig. 8. Hand distance versus cost to %success ratio

Now, percentage success rate from the Table I. and above cost estimation is used to determine the cost to success ratio. This ratio indicates the cost incorporated by the methods for unit success rate. Obviously the lower value of this cost to success ratio will be preferred.

From Fig. 8, we can understand the amount of cost needed to achieve corresponding success with successive hand distance. The cost of light dependent method at indoor for unit success rate is lower than that of ultrasonic method. However for outdoor, the cost of light dependent method for unit success becomes infinity with increasing distance.

The light dependent method is not effective at both indoor and outdoor conditions. As can be seen from Fig. 7, its performance is good for certain range of operator hand distance for indoor use only. For outdoor use it has a very poor performance. So it should be used only for indoor use. However, from Fig. 7, it can also be seen that the ultrasonic method also has a very good performance for certain range of operator hand distance. Moreover, the ultrasonic method has similar performance for both indoor and outdoor application. So it can be said that ultrasonic method should be used for both indoor and outdoor and no need to be concerned with light

dependent method. However, from Fig. 8, it can be seen that for indoor application light dependent method has a lower cost than ultrasonic method. So for indoor use it is better to use light dependent method. For outdoor application ultrasonic method is better in both performance and cost effectiveness. Thus makes it preferable for outdoor use.

V. CONCLUSION

In this study we have compared a new method of gesture recognition with a previously experimented method. For the analysis, preceding procedure and equipment are used. Two concerning methods are light dependent and ultrasonic method of gesture recognition. Percentage of success of gesture recognition with respect to the distance from which the operator can indicate pattern was calculated. Also cost accompanying with the successful gesture recognition for those two methods are calculated. From these experiments, it has been found that the light dependent method is more convenient for indoor application. But, for outdoor use the ultrasonic method is more convenient.

REFERENCES

- [1] C. Zhuo and Z. Xiang, "Research and Development on Locomotive HMI Display Interface," *Intelligent Systems Design and Engineering Applications, 2013 Fourth International Conference on*, Zhangjiajie, 2013, pp. 433-435.
- [2] Naohisa Hashimoto, Shin Kato and Sadayuki Tsugawa, "HMI for elderly drivers in an off-vehicle parking assistance system," *Intelligent Vehicles Symposium, 2008 IEEE*, Eindhoven, 2008, pp. 482-487.
- [3] T. Zhang, Q. Li and F. Ma, "Remote control system of smart appliances based on wireless sensor network," *2013 25th Chinese Control and Decision Conference (CCDC)*, Guiyang, 2013, pp. 3704-3709.
- [4] K. Kobayashi, Y. Nakagawa, S. Yamada, S. Nakagawa and Y. Saito, "Rebo: A remote control with strokes," *RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication*, Toyama, 2009, pp. 751-756.
- [5] É S. Silva and M. A. F. Rodrigues, "A Gesture Control System for Aiding Surgical Procedures," *Virtual and Augmented Reality (SVR), 2014 XVI Symposium on*, Piata Salvador, 2014, pp. 287-296. doi: 10.1109/SVR.2014.34
- [6] G. C. Luh, H. A. Lin, Y. H. Ma and C. J. Yen, "Intuitive muscle-gesture based robot navigation control using wearable gesture armband," *Machine Learning and Cybernetics (ICMLC), 2015 International Conference on*, Guangzhou, 2015, pp. 389-395.
- [7] B. H. Fan and K. Y. Li, "The Speech Control System of Intelligent Robot Prosthesis," *2010 Second WRI Global Congress on Intelligent Systems*, Wuhan, 2010, pp. 407-409.
- [8] Z. Xinfeng, Y. Diange and L. Xiaomin, "Speech Control based on Command Matching in Vehicle," *Vehicular Electronics and Safety, 2006. ICVES 2006. IEEE International Conference on*, Beijing, 2006, pp. 419-423.
- [9] K. A. Moxon, "Brain-control interfaces for sensory and motor prosthetic devices," *Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01). 2001 IEEE International Conference on*, Salt Lake City, UT, 2001, pp. 3445-3448 vol.6.
- [10] M. K. Ahuja and A. Singh, "Static vision based Hand Gesture recognition using principal component analysis," *MOOCs, Innovation and Technology in Education (MITE), 2015 IEEE 3rd International Conference on*, Amritsar, 2015, pp. 402-406.
- [11] A. Akl, C. Feng and S. Valaee, "A Novel Accelerometer-Based Gesture Recognition System," in *IEEE Transactions on Signal Processing*, vol. 59, no. 12, pp. 6197-6205, Dec. 2011.
- [12] T. Zhang and Z. Feng, "Dynamic Gesture Recognition Based on Fusing Frame Images," *Intelligent Systems Design and Engineering Applications, 2013 Fourth International Conference on*, Zhangjiajie, 2013, pp. 280-283.
- [13] N. Uzzaman, S. Hossain, A. Hossain, "Controlling an Industrial Robotic Arm Using Light Dependent Gesture Recognition", AIP conf. proc., International Conference on Mechanical Engineering (ICME), 2015, Bangladesh, in press.

Study of Molybdenum Sulphide as a Novel Buffer Layer for CZTS Solar Cells

Mrinmoy Dey^{1,2*}, Maitry Dey³, Samina Alam³, Nipu Kumar Das^{1,2}, M. A. Matin^{1,2} and Nowshad Amin^{4,5}

¹Department of Electrical and Electronic Engineering, Chittagong University of Engineering and Technology, Chittagong-4349, Bangladesh

²Renewable Energy Laboratory (REL), Chittagong University of Engineering and Technology, Chittagong-4349, Bangladesh

³Department of Electrical and Electronic Engineering, Premier University, Chittagong-4203, Bangladesh

⁴Solar Energy Research Institute (SERI), The National University of Malaysia, 43600 Bangi, Selangor, Malaysia

⁵Department of Electrical, Electronic and Systems Engineering, The National University of Malaysia, 43600 Bangi, Selangor, Malaysia

*mrinmoycuet@gmail.com

Abstract—The absorber layer of CZTS solar cell is a compound semiconductor which has favourable optical and electrical properties. Researchers are highly interest to investigate the CZTS solar cells for its earth abundant, non – toxic and low cost feature. Consequently, the absorber layer of CIGS is replaced to CZTS absorber layer. In this research work, the potentiality of Molybdenum Sulphide (MoS₂) as buffer has been investigated to explore the higher performance of CZTS solar cells. SCAPS 2802 simulator was used to evaluate the performance of CZTS solar cells with prospective MoS₂ buffer layer. The possibility of ultra-thin CZTS solar cells was examined and the higher efficiency of 17.03% ($J_{sc} = 29.42 \text{ mA/cm}^2$, $V_{oc} = 1.01 \text{ V}$ and $FF = 0.574$) was achieved for 1 μm thickness of absorber layer. In addition, the proposed CZTS solar cells had better thermal stability at higher operating temperature also.

Keywords— CZTS solar cell, MoS₂, Earth abundant, SCAPS 2802.

I. INTRODUCTION

The Sun is the ultimate source of clean generation of energy which is capable to produce GW range of power through solar cell to make sustainable development in the world. It is predicted by the researchers that 20% solar energy will be generated out of total power supply in 2050 and this magnitude would be more than 50% in 2100. The rapid rise of thin film PV technologies suppressed the production of c-Si technology due to high production cost. The feature of low efficiency and instability are both responsible for lower manufacturing of Si thin film solar cell.

The scarcity and expense Ga, In and Te as well as the effect of toxicity in the environmental created by Cd are encouraged the research work on CZTS solar cells for sustainable and large scale production of solar cells. CZTS is a quaternary semiconductor material of group I-II-IV-VI having high absorption coefficient of 10^4 cm^{-1} . The direct band gap energy

of CZTS is 1.4 - 1.5 eV which is very promising for photovoltaic absorber layer material [1]. As an absorber layer material, CZTS is a semiconductor compound capable to make low cost and high efficiency solar cells. It is considered as earth abundant photovoltaic material which is non-toxic also. Recently, researchers are highly interested to do research on CZTS solar cell. Typical fabrication techniques of CZTS solar cells are electroplating [2], thermal evaporation [3], co-sputtering [4].

The material properties of CZTS solar cell are quite similar to CIGS solar cells though CZTS is favourable to the researchers for its rare metal free feature. The prospects of MoS₂ buffer layer is investigated by forming p-n junction with p-CZTS absorber layer. CZTS solar cell achieved 8.4% efficiency with the cell structure of Al/ZnO:Al/ZnO/CdS/CZTS/Mo/SLG [5] and 11.1% [6] by Se-containing Cu₂ZnSn(S,Se)₄. In 2012, 1.85% efficiency of CZTS solar cell with In₂S₃ buffer layer discussed by V. G. Rajeshmon et al [7]. According to photon balance calculations of Shockley-Queisser, CZTS is expected to have theoretical efficiency more than 30% [8]. The challenges of thin film solar cells are the reduction of absorber layer thickness, better cell stability and finally the performance of these solar cells for optimum use of photovoltaic material for sustainable development in the arena of harvesting the solar energy.

In this study, the prospects of CZTS solar cell with MoS₂ buffer layer is investigated utilizing SCAPS 2802 simulator to design low cost, high efficiency and stable solar cells by improving the performance of cell parameters. The effect of transparent conducting oxide (TCO) layer, thickness reduction of absorber layer, variation of absorber layer doping concentration, absorber layer deep level defect, effect of cross section as well as cell stability were investigated to design low cost high performance CZTS solar cell.

II. MODELING AND SIMULATION

The numerical modeling of this research work is done by SCAPS 2802 to build a reasonable physical model for testing feasibility and numerical simulations to predict any change in cell performance resulting from the modified reasonable parameters. In the early 1980s, the numerical modeling for the research work on thin film PV devices was started. The complex structure formation in thin film solar PV devices is considered as the driving force to model an efficient PV cell.

In this research work, numerical modelling has been done to explore the hidden potentiality of CZTS solar cells with promising MoS₂ buffer layer. According to the Poisson equation and hole and electron continuity equation, the steady-state band diagram, recombination profile and carrier transport are calculated. The proposed structure of CZTS solar cell has three layers which are 30 nm thickness of i-ZnO as TCO layer, MoS₂ with thickness of 40 nm as buffer layer and 1 μ m thickness of absorber layer. Fig. 1 illustrates the cell structure of proposed CZTS solar cell with ultra-thin range of absorber layer.

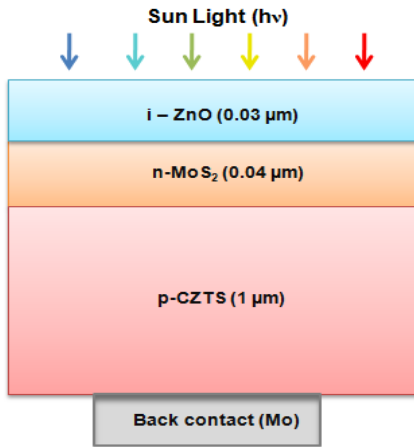


Fig.1 The proposed structure of CZTS solar cell

Table I shows the materials parameters used to design CZTS solar cells which were selected based on experimental data, literature values, theory, or in some cases reasonable estimations [9].

TABLE I MATERIAL PARAMETERS USED FOR NUMERICAL ANALYSIS OF THE PROPOSED CZTS SOLAR CELLS

Parameters	i-ZnO	n-MoS ₂	p-CZTS
W (μ m)	0.02	0.04	0.5-3
ϵ/ϵ_0	9	4	10
μ_c (cm ² /Vs)	100	100	100
μ_p (cm ² /Vs)	25	150	25
n, p (cm ⁻³)	1×10^{19}	1×10^{21}	1×10^{14}
Band gap, E _g (eV)	3.3	1.29	1.5
N _c (cm ⁻³)	2.2×10^{18}	7.5×10^{17}	2.2×10^{18}
N _v (cm ⁻³)	1.8×10^{19}	1.8×10^{18}	1.8×10^{19}
χ (eV)	4.6	4.2	4.5

In this study, the effect of TCO layer and the variation of thickness of absorber layer with different doping concentration and absorber layer deep level defect are analyzed. Consequently, the cell stability and the impact of capture cross section are also studied to model the high performance CZTS solar cell.

III. RESULT AND DISCUSSION

A. Effect of the presence of TCO layer

In this research work, the presence of intrinsic ZnO (i-ZnO) as TCO layer in the proposed CZTS solar cell with prospective MoS₂ buffer layer has been investigated to design an efficient solar cell. Table II shows the comparative study to evaluate the performance of CZTS solar cell with and without TCO layer. It is evident from Table II that the cell structure (i-ZnO/MoS₂/CZTS) shows the better performance of proposed CZTS solar cell.

TABLE II EFFECT OF TCO LAYER IN CZTS CELLS

Parameters	With TCO	Without TCO
V _{oc} (V)	1.01	0.91
J _{sc} (mA/cm ²)	29.42	18.85
FF	0.574	0.642
Eff. (%)	17.03	10.96

B. Effect of absorber layer thickness

In order to preserve the absorber layer material in case of designing the efficient solar cells it is mandatory to investigate the performance of solar cell with the variation of absorber layer thickness. Numerical analysis has been done to evaluate the performance of CZTS solar cell structure (i-ZnO/MoS₂/CZTS). The simulation results for thickness variation of absorber layer from 0.5 μ m to 3 μ m are shown in Fig. 2.

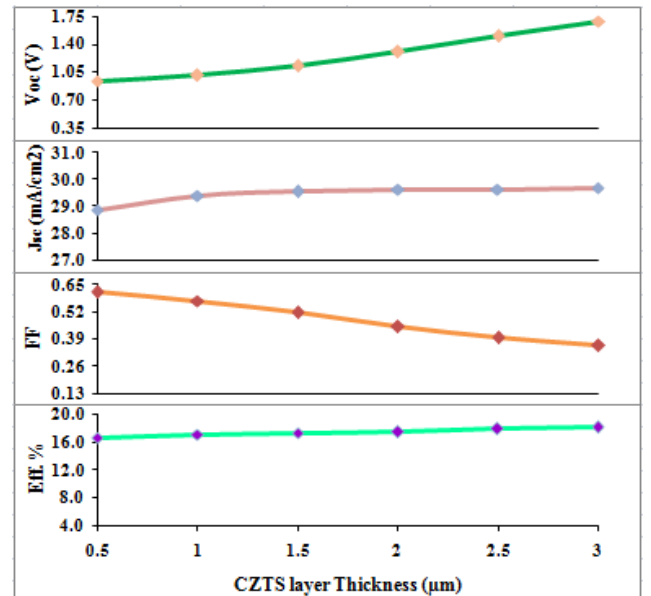


Fig.2 Effect of CZTS thickness variation for the proposed cell

Fig. 2 shows performance of CZTS solar cell with different cell performance parameters for the variation of absorber layer. The cell performance parameter V_{oc} is increased almost linearly with the increase of absorber layer thickness and J_{sc} shows increasing trend but approximately saturated after 1 μm thickness of absorber layer. The magnitude of FF is linearly in decreasing trend with the increase of absorber layer thickness. The highest conversion efficiency 17.03% ($FF = 0.574$, $V_{oc} = 1.01$ V and $J_{sc} = 29.42$ mA/cm^2) was achieved for only 1 μm thickness of absorber layer in the proposed CZTS solar cell. These results of cell performance are agreeable to the related published work [9]. The conversion efficiency is improved about 1% for the increase of absorber layer thickness from 1 μm to 3 μm . This research work indicated that the preservation of CZTS absorber layer material was possible around 67% considering the reduction of conversion of 1 % efficiency.

C. Effect of Absorber layer Doping Concentration

The performance of solar cells depends on the amount of doping concentration. In this study, the doping concentration p- type absorber layer is varied from 1×10^{10} to $1 \times 10^{14} \text{ cm}^{-3}$ to evaluate the performance of CZTS solar cell with potential MoS_2 buffer layer. Fig. 3 depicts the performance of CZTS solar cell under different level of practically feasible doping concentration. The cell parameter current density is almost constant in response to the variation of absorber layer doping concentration whereas the open circuit voltage cell parameter is increasing trend. The FF is in decreasing trend after a certain doping concentration. The performance of cell conversion efficiency is increased to 17.03% at highest doping concentration ($1 \times 10^{14} \text{ cm}^{-3}$) of CZTS absorber layer.

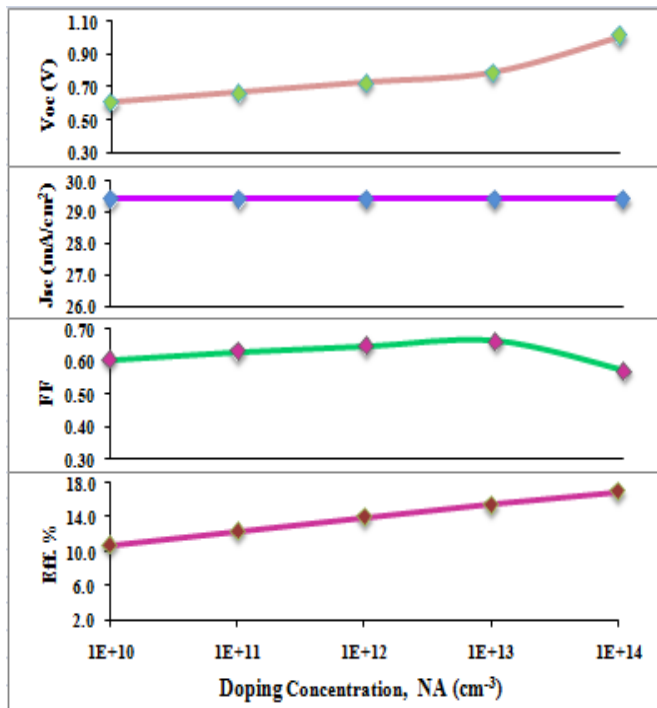


Fig.3 Effect of the CZTS doping concentration on cell performance

D. Effect of Deep level Defect

In semiconductors, deep level defect is considered as undesirable types of electronic defects. This effect is caused by the trap to the valence or conduction band is much larger than the characteristics thermal energy. The performance of CZTS solar cells are investigated considering the deep level defects.

In this study, the performance of CZTS solar cell is examined for both uniform and Gaussian energetic distributions which are shown in Table III. Table III shows that the cell configuration (i-ZnO/ MoS_2 /CZTS) of CZTS solar cell having best cell conversion efficiency for both uniform and Gaussian distribution.

TABLE III DEEP LEVEL DEFECTS IN CZTS SOLAR CELLS

Parameters	Energ. Dis.	Voc (V)	Jsc (mA/cm^2)	FF	Eff. (%)
With TCO	Uniform	1.01	29.42	0.574	17.02
Without TCO		0.90	18.84	0.642	10.94
With TCO	Gaussian	1.01	29.42	0.575	17.01
Without TCO		0.90	18.84	0.642	10.94

E. Effect of Capture Cross Section

In this study, the performance of CZTS solar cells are investigated based on the variation of hole capture cross section with the constant magnitude of electron capture cross section. Fig. 4 shows the performance of different cell parameters of CZTS solar cell after the variation of hole capture cross section. The conversion efficiency is almost constant with the increase of the order of capture cross section up to $1 \times 10^{-13} \text{ cm}^2$. The performance of CZTS solar cell is sharply in decreasing trend in response of hole capture cross section after $1 \times 10^{-11} \text{ cm}^2$.

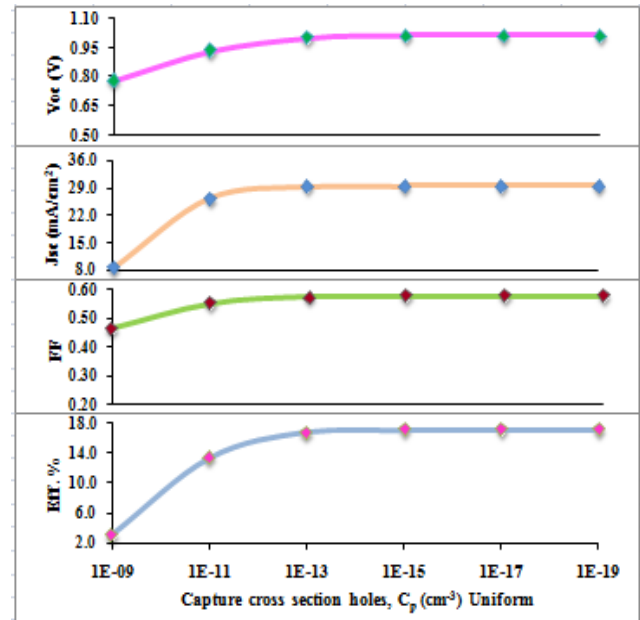


Fig.4 Effect of capture cross section of hole on cell performance

F. Effect of Temperature

The performance of solar cells is affected by the operating temperature and it is considered as important issue to test the designed cell for better thermal stability. To investigate the cell stability, it is mandatory to examine the performance of designed cell with higher operating temperature. For this purpose, the proposed cell is examined with higher operating temperature ranging from 25°C to 165°C which is shown in Fig. 5. The cell conversion efficiency is decreased with the increased of operating temperature as expected.

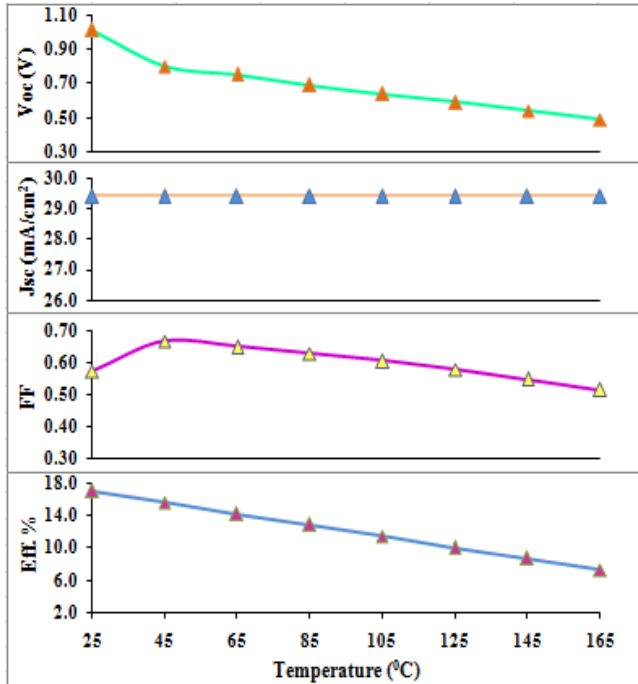


Fig.5 Effect of temperature on cell performances

The normalized efficiency of solar cell is decreased with the increase of operating temperature. It is evident from Fig. 6 that the normalized efficiency of the proposed CZTS cell is linearly decreased with the increase of operating temperature at a temperature coefficient (TC) of -0.41%/°C. This TC indicates better stability of the cells at higher operating temperature.

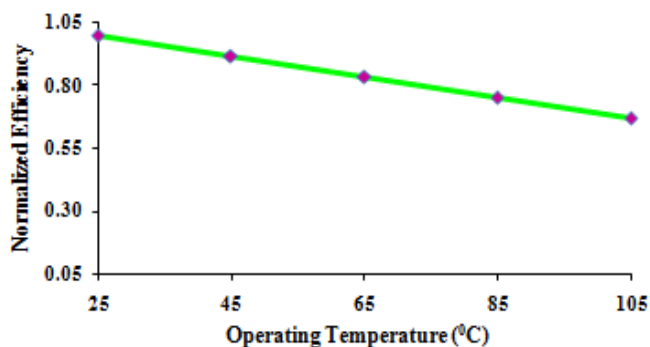


Fig.6 Effect of operating temperature on normalized efficiency

IV. CONCLUSION

The prospect of CZTS solar cell with potential MoS₂ buffer layer was investigated to design low cost, highly efficient and stable cell structure. The highest efficiency was achieved 17.03% ($J_{sc} = 29.42 \text{ mA/cm}^2$, $FF = 0.574$, $V_{oc} = 1.01 \text{ V}$) for 30 nm i-ZnO as TCO layer, 40 nm MoS₂ as buffer layer and only 1 μm CZTS absorber layer material. The thermal stability of the proposed cell showed better stability at higher operating temperature.

ACKNOWLEDGMENT

This work has been supported by the Department of Electrical and Electronic Engineering (EEE) and Renewable Energy Laboratory (REL), Chittagong University of Engineering and Technology (CUET), Bangladesh. The authors would like to thank the Electrical and Electronic Engineering department of Premier University Chittagong.

REFERENCES

- [1] H. Wei, W. Guo, Y. Sun, Z. Yang, and Y. Zhang, "Hot-injection synthesis and characterization of quaternary Cu₂ZnSnSe₄ nanocrystals," *Materials Letters*, vol. 64, pp. 1424-1426, 2010.
- [2] A. Ennaoui, M. Lux-Steiner, A. Weber et al., "Cu₂ZnSnS₄ Thin Film Solar Cells from Electroplated Precursors: Novel Low-Cost Perspective," *Thin Solid Films*, vol. 517, no. 7, pp. 2511-2514, 2009.
- [3] K. Wang, O. Gunawan, T. Todorov et al., "Thermally Evaporated Cu₂ZnSnS₄ Solar Cell," *Applied Physics Letters*, vol. 97, p. 143508, 2010.
- [4] H. Katagiri, K. Jimbo, W.S. Maw et al., "Development of CZTS-Based Thin Film Solar Cells," *Thin Solid Films*, vol. 517, no. 7, pp. 2455-2460, 2009.
- [5] B. Shin, O. Gunawan, Y. Zhu et al., "Thin Film Solar Cell with 8.4% Power Conversion Efficiency Using an Earth-Abundant Cu₂ZnSnS₄ Absorber," *Prog. Photovolt: Res. Appl.*, vol. 21, no. 1, pp. 72-76, 2013.
- [6] T. K. Todorov, J. Tang, S. Bag et al., "Beyond 11% Efficiency: Characteristics of State-of-the-Art Cu₂ZnSn(S, Se)₄ Solar Cells," *Adv. Energy Mater.*, vol. 3, no. 1, pp. 34-38, 2012.
- [7] V. G. Rajeshmon, N. Poornima, K. C. Sudha et al., "Modification of the Optoelectronic Properties of Sprayed In₂S₃ Thin Films by Indium Diffusion for Application as Buffer Layer in CZTS Based Solar Cell," *J. Alloys Comp.*, vol. 553, pp. 239-244, 2013.
- [8] W. Shockley, H. J. Queisser, "Detailed Balanced Limit of Efficiency of p-n junction solar cells," *J. Applied Physics*, vol. 32, pp. 510-519, 1961.
- [9] Mrinmoy Dey, Maitry Dey, Tama Biswas, Samina Alam, N. K. Das, M. A. Matin and Nowshad Amin, "Modeling of Cu₂ZnSnS₄ Solar Cells with Bismuth Sulphide as a Potential Buffer Layer," in *Proceedings of the 5th International Conference on Informatics, Electronics & Vision*, Dhaka, Bangladesh, May 2016.

CWT based improved approach to Wideband Spectrum Sensing for Cognitive Radios

Manobendu Sarker

Department of EEE, University of Asia Pacific, Bangladesh

rsarker.eee@uap-bd.edu

Abstract—Cognitive radio has become into a matter of great interest now-a-days for researchers as it opens a wide horizon of opportunities for maintaining the exponentially increasing bandwidth demand. In cognitive radio networks, Spectrum sensing is a major concerning issue for exploring dynamic spectrum management. This paper introduces an improved but simple continuous wavelet transform(CWT) based approach for efficient spectrum sensing of wideband channel, based on edge detection in which soft de-noising is utilized to suppress the noise level. One special feature of our approach is to use single scale transformation unlike most of the edge detection approach involves multi-scale transforms or product to suppress the noise as well as to detect edges properly which indicates frequency boundaries and identification of these edges is shown through simulation results. Performance of detection in low SNR region by our proposed method has also been improved than the conventional methods which is also illustrated through simulation result.

Index Terms—Cognitive radios, Edge detection, soft de-noising and Continuous Wavelet transform.

I. INTRODUCTION

The quick proliferation of up-growing RF spectrum demand has caused researchers to take initiatives on the fifth generation (5G) cellular communication systems which have gained accelerating momentum globally. In next generation (5G) communication, cognitive radio is considered as one of promising technology for tackling the challenges of heavy RF traffic with utilizing limited and shrunk resources [1]. We know that a cellular system is only able to operate in licensed band within fixed limited bandwidth under the current spectrum regulation policy. While the cellular bands are profoundly utilized and exploited, studies have found that some licensed frequency bands are surprisingly underutilized. Cognitive radio provides opportunity for a cellular network to effectively share the underutilized frequency bands without creating any harmful interference to the licensed users. The cost of purchasing a licensed band is expected to be much expensive than leasing the spectrum because of its opportunistic access and utilization. Therefore, cognitive radio permits a cellular network to widen its spectrum on demand at a relatively low cost, thereby proving a way out to cope with uncertain heavy data traffic.

Cognitive radio systems vastly depends on sensing the radio environment which targets to sense a bandwidth which is greater than the coherence bandwidth of the channel. For a

cognitive radio system, the most important problem is to find out the existence of primary users over wide range of spectrum. This process faces many difficulties due to detect various primary users which have different modulation schemes, data rates and transmission powers in presence of variable propagation losses, interference generated by other secondary users as well as thermal noise. This is correct for the case of broadcast TV channels, where the receivers are passive and impossible to detect the presence of a nearby receiver. Moreover, the sensing process needs to be very quick in order to complete scanning the entire wideband without significant delay. For sensing wideband frequency bands, energy and wavelet based detection is best fitted than other detection techniques. Wavelet transform is a powerful mathematical tool widely used for analyzing singularities and irregular structures. It can also be used in wideband spectrum detection by identifying the frequency locations of non-overlapping spectrum bands. Based on characteristics of CR sensing, researchers assume the entire wideband as a train of successive frequency sub-bands where the power spectral nature is continuous within each sub-band but exhibits a discontinuities change between adjacent sub-bands. Such changes leads to irregularities in PSD which can give us vital information on the locations and intensities of spectrum holes and these features can be extracted by wavelet transform.

An improved and easy approach is presented in this paper for collecting information effectively like edge detection using continuous wavelet transform which can create impact on the performance of a detector for wideband spectrum sensing. We thoroughly study about extraction of received signal from noisy environment by using common threshold technique (soft) along with some other processing s and show noticeable performance improvement by using our proposed method with help of the simulation result which is the main contribution of this paper. In our previous work in [2], same processing is adapted to suppress the noise level but for detecting the frequency edges Hilbert transform is used. False edge can be generated after using wavelet transform if the received signal is not processed correctly so its really very important to process the received signal. Finally, better performance of our proposed method than conventional method in low SNR is also manifested through Receiver's operating characteristics (ROC) curve.

The remainder of this paper is organized as follows. Related works are described in Section II. Section III discusses the

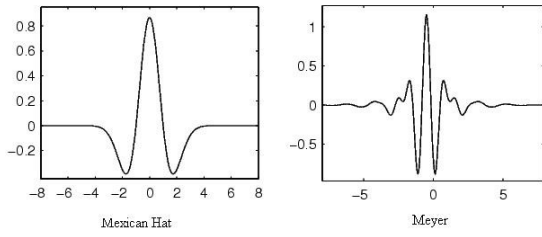


Fig. 1: Wavelet examples

background of wavelet, continuous wavelet Transform and de-noising in brief. System description for wavelet-based spectrum sensing using CWT in section IV. In section V, proposed method for edge detection is introduced. Section VI shows the simulation results of the proposed scheme with explanation, and Section VII draws the conclusion for this work.

II. RELATED WORKS

A considerable researches were done on wavelet as well as wavelet multi-scale edge detection. It has been heavily employed in feature detection of edges, de-noising and adjusting of both one dimension and two-dimension signals. In [3], the authors proposed a continuous wavelet transform (CWT) based method where the signals' PSD is determined to spot the irregular points which leads to find spectrum holes of the targeted frequency band through searching the local extreme of the first derivative or zero-crossings of the second derivative of wavelet modulus. Multi-scale wavelet transform is adapted to find edges and multi-band detection is also introduced using comparison of energies in [4] for sensing wide frequency range. They also mention as well as discuss about some new metrics to enhance the performance of receiver. Said E. El-Khamy et al. [5] propose a sensing technique based on the information found from multiscale wavelet transform to characterize singularities to increase the performance of the wavelet transform multiscale modulus maxima algorithm. Besides, they claim to develop some new performance metrics for evaluating the performance of their method at medium-to-high noise power. Authors in [6] presented a new method for spectrum sensing is deduced from differential and operational algebra as well as noncommutative ring theory. By giving emphasis on "spike-like" parts of the received noisy amplitude spectrum, their proposed algebraic based algorithm can successfully identify points for spectrum sensing. In [7] OFDM based cooperative spectrum sensing is studied using the energy and wavelet based detector scheme under AWGN as well as Rayleigh fading channel and for identifying the sharp and blurred peaks, they use change in slope of the power spectral density.

III. BACKGROUND ON WAVELET TRANSFORM AND DE-NOISING

A. Wavelet fundamental

A wavelet is a small wave which has energy concentrated within time. It is compactly supported and has finite energy

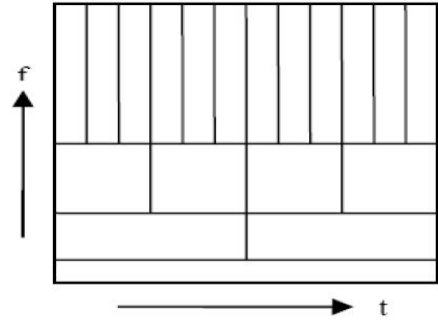


Fig. 2: Resolution of time and frequency in Wavelet Transform

function. It can satisfy admissibility condition and could be used for multiresolution analysis. The wavelet transform is nothing but a mathematical tool for breaking up a signal into a set of orthogonal waveforms localized both in time and frequency domains. It is a decomposition of the signal as a combination of a set of basic functions, obtained by means of dilation and translation of a single prototype wavelet; there are several wavelet functions (mother wavelet with different properties) like the Morlet or Mexican Hat wavelets or orthogonal Meyer wavelets as in Fig. 1.

Providing opportunity for analyzing in both frequency and time domain is one of the major advantages of wavelet method as in Fig. 2, it permits signals to be stored more efficiently than by the Fourier transform. It can direct to better approximation of real-world signals and also well-fitted for approximating data with sharp discontinuities. The wavelet transforms ignore polynomial components of the signal up to the approximation order of the basis, which is a disadvantage of wavelet method.

B. Equation for Continuous Wavelet transform

Wavelet analysis is conducted by the splitting up of a signal into a shifted and scaled version of the original wavelet. A Continuous wavelet transform can be defined as the sum of overall time of the signal multiplied by a scaled and shifted version of the wavelet function. The greater the scale factor is, the wider the basis function is and consequently, the corresponding coefficient gives information about lower frequency components of the signal and vice versa. The wavelet transform equation is given by

$$CWT_x^\psi(\tau, s) = \psi(\tau, s) = \frac{1}{\sqrt{|s|}} \int x(t) \psi^*\left(\frac{t - \tau}{s}\right) dt$$

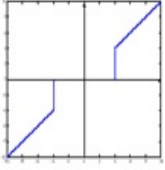
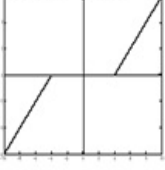
Where $x(t)$ = given signal, τ =translation parameter, s = scaling parameter $=1/f$ and $\psi(t)$ = mother wavelet.

C. Wavelet de-noising

Wavelet thresholding is nothing but a signal estimation process that applies the potentials of wavelet transform for

signal de-noising. It can decrease noise by avoiding coefficients that are not significant compare to some threshold. There are a number of different techniques for choosing de-noising parameters but unfortunately "best" universal threshold determination process can't be found. There are two types of global thresholding- soft and hard. The differences are as follows:

TABLE I: Difference between Hard and Soft Thresholding

Hard Thresholding	Soft Thresholding
It's operator is defined as: $D(U, \lambda) = U$ for all $ U > \lambda$	It's operator is defined as: $D(U, \lambda) = \text{sgn}(U)$ $\max(0, U - \lambda)$
Avoid coefficients greater than the absolute value of threshold.	Reduce coefficients greater than the absolute value of threshold.
The transfer function is shown below 	The transfer function is shown below 

IV. SYSTEM MODEL

A CR system usually checks the wireless channel for identifying white spaces or spectrum holes. We assume that N spectrum bands are engaged or occupied by the received signal and also we need to detect and identify their frequency position as well as PSD levels. The frequency range is between f_0 and f_N and their location of boundaries for frequencies are at $f_0 < f_1 \dots < f_N$. A wideband signal's PSD structure of n -th band is defined by $B_n : \{f \in B_n : f_{n-1} \leq f < f_n\}, n = 1, 2, \dots, N$, shown in Fig. 3. The following assumptions are considered in this work:

- (1) The spectrum boundaries f_0 and f_N are known to the CR system and we only concentrate on interested spectrum, i.e. between f_0 and f_N , although the observed signal may contain a wider band for the CR system.
- (2) The CR system doesn't know about N spectrum bands and their positions for $f_1, \dots, f_{N-1}$. We also consider that these environment parameters have varying nature but unchanged within a time burst.
- (3) There is significant discontinuity at the edge of each band so that we can identify these edges.
- (4) In this work we consider the additive white Gaussian noise (AWGN) with zero mean and two-sided PSD, defined by $S_r(f) = \frac{N_0}{2}$.

We can write the PSD of the received signal $r(t)$ of the CR system as

$$S_r(f) = \sum \alpha_n^2 S_n(f) + S_w(f), f \in [f_0, f_N] \quad (1)$$

where the n -th signal spectrum is $S_n(f)$ and the signal power density within the n -th band is α_n^2 . $\mathcal{F}\{R_r(\tau)\}$ is the

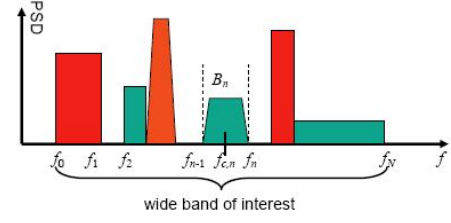


Fig. 3: PSD of N frequency bands

PSD $S_r(f)$, where the autocorrelation function is defined as $R_r(\tau) = E\{r(t)r(t+\tau)\}$ and $\mathcal{F}\{\cdot\}$ is Fourier Transform. The time-domain signal $r(t)$ can be represented as

$$r(t) = \sum_{n=1}^N \alpha_n x_n(t) + w(t) \quad (2)$$

where the time-domain signal is $x_n(t)$ with PSD of $S_n(f)$ and $w(t)$ is the AWGN with PSD of $S_w(f)$. So, we can express the n -th band occupied signal $x_n(t)$ as

$$x_n(t) = \sum_{k=-\infty}^{\infty} s_k p(t - kT_s) e^{j2\pi f_{c,n} t} \quad (3)$$

where the k -th modulated symbol is s_k , a pulse shaper, $p(t)$ with the bandwidth of frequency difference between f_n and f_{n-1} and the center frequency of the n -th band is $f_{c,n} = \frac{f_{n-1} + f_n}{2}$. The PSD $S_n(f)$ is equal to $|\mathcal{F}\{p(t)\}|^2$.

We need to estimate the parameters, i.e. N, f_n, α_n^2 which is considered as major wideband spectrum sensing problem for the CR system. Now, the Fast Fourier transform (FFT) is a simple and straightforward process to determine PSD. Let, the size of FFT is N , which implies that the whole observed frequency band is sampled for having N frequency points and they are called subcarriers. We define the PSD at subcarrier n be $P(n)$ where, $n = 0, 1, \dots, N-1$ and $P = [P(0) P(1) \dots P(N-1)]^T$. Then the estimation of PSD is presented as $\hat{P} = \frac{1}{M} \sum_{i=0}^{M-1} X_i \odot X_i^*$ where $M = \lceil \frac{N}{N_s} \rceil$, $\odot$ is the piecewise product and $*$ is the conjugate, X is the Fourier transform of receive signal x and it will be zero padded if necessary. There may be too many fake edges in the received signal PSD, hence we need to set a threshold for selecting local maximum points which are larger than the threshold so that setting threshold is a crucial job.

V. PROPOSED METHOD FOR EDGE DETECTION

The gist of our proposed method is illustrated in Fig. 4. As discussed earlier, there is a necessity to choose threshold to avoid fake edges. It is done with two stages. First, we propose to process the signal or received PSD by averaging it and after that we set the threshold. Let, Δ be the mean of the PSD, that

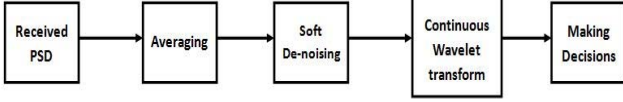


Fig. 4: Block diagram of proposed method for edge detection

is,

$$\Delta = \frac{1}{N} \sum_{n=0}^{N-1} P(n), \Delta > 0 \quad (4)$$

Then the PSD after averaging will be,

$$\rho = \frac{|P(n)|}{\Delta} \quad (5)$$

We can see that our received PSD is mixed with noise. So after averaging, we use soft de-noising to reduce noise. After applying continuous wavelet transform(CWT) on de-noised signal, the change of Wavelet transformed signal's direction reflects changing a edge as for example in Fig.5 where wavelet transform is applied on square wave. We can separate falling edge (dark red colored box) and rising edge (yellow colored box) by observing peak value of the transformed signal which is evident from this figure.

In our scenario, we can utilize this property to detect changing edges. So we can find local maxima and minima points which will lead to the edge frequencies. We have also no necessity to search the type of edge both in the beginning and end of the received PSD whether this is continuation of falling or rising edges for the generally used wavelet transform based estimator so that it can reconstruct the PSD completely. Our next step is to further classify the bands as occupied or vacant. We store the frequency points of local maxima or minima in μ , that is, $\mu = \{n \text{ is local maxima or minima point of } P(n)\}$. Let, $l_k (k = 1, 2, \dots, K)$ are elements of μ in ascending order. So we can easily determine the vacant or occupied bands by measuring the average PSD between $[l_k, l_{k+1}]$ band, that is,

$$\Omega(k) = \frac{1}{l_{k+1} - l_k} \sum_{n=l_k}^{l_{k+1}} P(n) \quad (6)$$

After calculating $\Omega(k)$, we can arrange bands by two categories with higher values and very small values. So bands with higher values can be called occupied and rest of them can be called vacant. However, the power of noise is uncertain and can be varied with time and location but through our proposed method, the noise power can be approximated by calculating average PSD of all detected vacant bands. The average SNR of each occupied band can be approximated with the help of the noise power.

VI. PERFORMANCE EVALUATION

In this section, simulation results are presented to evaluate the merits of our proposed method. All the simulations are done in the MATLAB. As this work is kind of extension of our previous work [2] so all the parameters used here are as

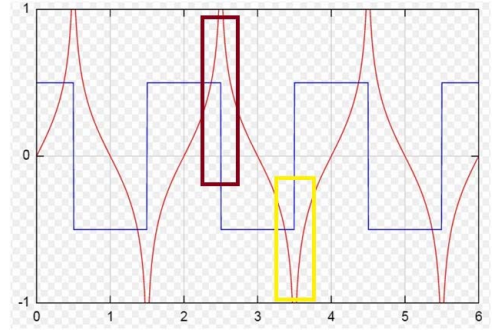


Fig. 5: Wavelet transform of square wave

same as used in [2] and for this reason some results have also similarities.

A. Process of Edge detection

Here, we consider [50, 300] MHz as band of interest for our simulation. Fig. 6 shows the observed PSD $S_r(f)$ by a CR like considered in [3]. We also consider the value of noise floor $S_w(f)$ is 200 . In Fig. 6, there are total $N = 6$ bands B_n , with frequency boundaries at $\{f_n\}_{n=0}^6 = [50, 120, 190, 225, 250, 260]$ MHz. In Fig. 6 we can see that among those bands B_1, B_3 and B_5 have relatively high signal PSD while B_2, B_4 and B_6 has low PSD. So they are viewed as white space or spectral hole because at that moment they are not utilized.

This received PSD is averaged on the basis of (5) as per our proposed method and the received signal PSD after averaging is shown in Fig. 7. For reducing the noise level, soft de-noising is applied on averaged signal which is shown Fig. 8. Here in our simulation, "dB4" wavelet is utilized for de-noising process which functioned in 3rd level. The Daubechies wavelets are familiar as a family of orthogonal wavelets and they are characterized by a maximal number of vanishing moments for some given support. An orthogonal multi-resolution analysis is produced by scaling function for each wavelet type of this class. In our simulation, universal threshold $\text{sqrt}(2 * \log(.))$ is used. The noise level is significantly decreased which is illustrated in Fig. 8. Now time has come to apply CWT on the de-noised signal and it is evident that the edges can be easily identified by its maxima and minima shown in Fig. 9 and Fig. 10. In the simulation, "Haar" wavelet is used for CWT process along with two dyadic scales $s = 2^j$; $j = 2, 3$. The Haar wavelet is a sequence of rescaled square-shaped functions which together form a wavelet family or basis. We can find the frequencies where edge is changed by finding the maxima or minima points. For negative peak, it indicates a rising edge and vice versa. In Fig. 9 we can observe that CWT signal with dyadic scale 2 and 3 and also in Fig.10 their absolute values are shown. Here frequency bands can be found and categorized to vacant or occupied with two simple steps. First, we need to take absolute value of the PSD like in Fig.10 so that there will be no negative peaks. Second, this unipolar signal is sampled and we can find the edge frequencies by finding the sample

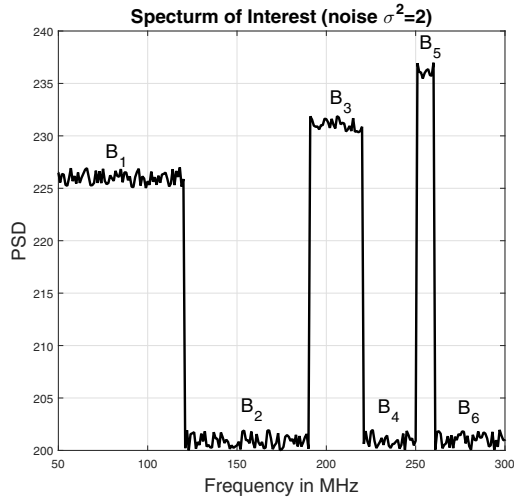


Fig. 6: Interested spectrum of PSD

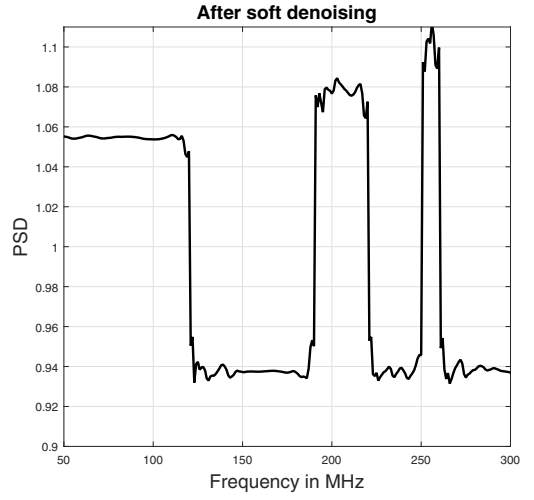


Fig. 8: Soft de-noised signal of averaged signal

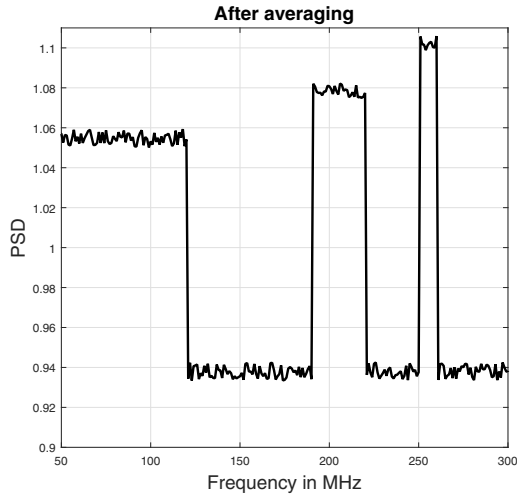


Fig. 7: Averaged PSD of received signal

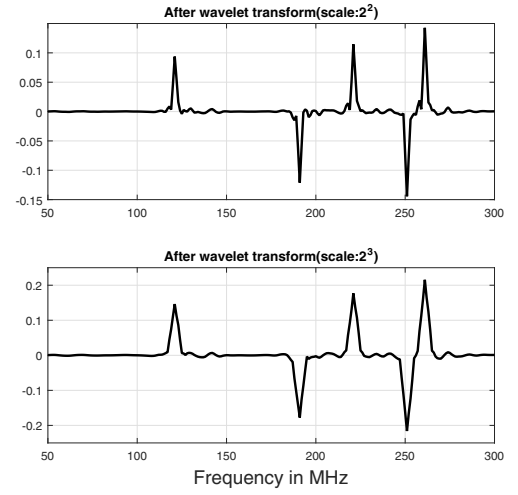


Fig. 9: Signal after Wavelet transform for level 2^2 and 2^3 respectively

(peaks) where it is greater than its previous and next sample. Now we have information about the frequencies of edges but unable to categorize the bands. Before that we need to know the slop of the first peak of Fig. 9 because by knowing its nature we can map the PSD and then easily classify the bands by measuring and comparing PSD of in between bands which is described in previous section.

The major contribution of our work lies in Fig. 9 and Fig. 10 because we don't use multi-scale product to suppress the noise. We just use CWT of any one scale (2 or 3) on de-noised signal following by averaged version of the received signal. So by observing these figures, we can easily understand that the noise level is significantly reduced for any one of the scale. So this method can be an alternative method for edge detection without using multi-scale product.

B. Impact of low SNR

The comparison between our proposed method and conventional method for low SNR i.e. $\gamma = -10dB$ is illustrated through the ROC curve which is shown in Fig. 11. It is found by changing probability of false alarm, P_f in Monte-Carlo simulation using 10,000 iterations and finding probability of detection, P_d for it [8]. Here, $\frac{Q^{-1}(P_f)}{\sqrt{L}} + 1$ is applied for threshold from [9] for both results, where L is signal length or size and $Q^{-1}()$ is the inverse Q function. It is significant from the Fig. 11 that our proposed method has better performance than the conventional method in low SNR described in [9]. It is desired to keep P_d high for lower P_f for best performance for spectrum sensing in low SNR. We can say that the ROC curve proves that our proposed method has such kind of ability.

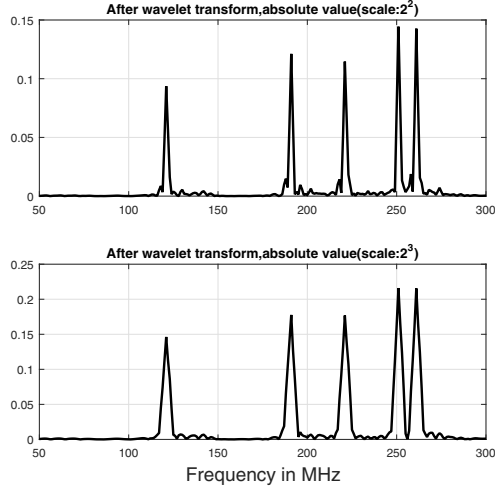


Fig. 10: Signal after Wavelet transform of absolute value for level 2^2 and 2^3 respectively

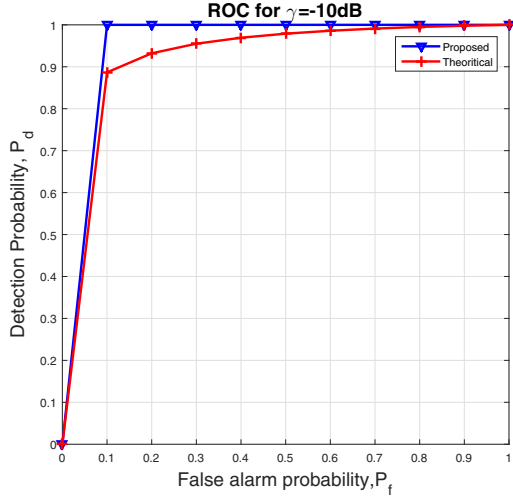


Fig. 11: ROC curve for proposed and conventional method [9], $\gamma = -10dB$

VII. CONCLUSION

An easy and improved method which completely avoids multi-scale product has been proposed in this paper for sensing wideband frequency spectrum. In CR system, classification of the bands into vacant or occupied is really very hard due to the noise power which usually has uncertainty and changing nature with time and location, our method functions well without any prior information of noise power to separate the bands and also does not need any attempts to search types of end edges which is a crucial pitfall with conventional edge detection methods. Besides in low SNR, our proposed method performs better than the conventional method which portrayed through the simulation result. Analytical analysis by applying the proposed method for finding mathematical expression for P_d or P_f as well as setting of threshold under different fading channel and

observing its sensing performance may be the possible future drive for research.

REFERENCES

- [1] R. Chvez-Santiago, M. Szydeko, A. Kliks et al., "5G: The convergence of wireless communications," *Wireless Pers. Commun.*, vol. 83, no. 3, pp. 1617-1642, 2015.
- [2] M. Sarker, "Hilbert transform based edge detection approach for Wideband Spectrum Sensing in Cognitive Radios," *International Conference on Wireless Communications, Signal Processing and Networking (WiSP-NET)*, 23- 25 March, Chennai, India, 2016.
- [3] Z. Tian and G. B. Giannakis, "A Wavelet Approach to Wideband Spectrum Sensing for Cognitive Radios," *International Conference on Cognitive Radio Oriented Wireless Networks and Communications*, Greece, 2006.
- [4] Y. Zeng, Y.-C. Liang and M. W. Chia, "Edge based wideband sensing for cognitive radio: Algorithm and performance evaluation," *Proc. IEEE Int. Symp. New Front. Dyn. Spectrum Access Netw.*, pp.538 -544, 2011.
- [5] S. E. El-Khamy, M. El-Mahallawy and E. Youssef, "Improved wideband spectrum sensing techniques using wavelet-based edge detection for cognitive radio," *International Conference on Computing, Networking and Communications (ICNC)*, pp. 418-423, 2013.
- [6] W. Guibene, M. Turki, A. Hayar and D. Slock, "A Complete Framework for Spectrum Sensing based on Spectrum Change Points Detection for Wideband Signals," *IEEE 75th Vehicular Technology Conference: VTC2012-Spring*, 6-9 May 2012, Yokohama, Japan, 2012.
- [7] P. R Parekh, M. B Shah, "Spectrum Sensing in Wideband OFDM based Cognitive Radio," *International Conference on Communications and Signal Processing (ICCSP)*, pp. 1476-1481, 2014.
- [8] M. Sarker, "Energy detector based spectrum sensing by adaptive threshold for low SNR in CR networks," *24th Wireless and Optical Communication Conference (WOCC)*, pp. 118 - 122, 2015.
- [9] Y. C. Liang, Y. Zeng, E. C. Y. Peh, and A. T. Hoang, "Sensing-throughput tradeoff for cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 4, pp. 1326-1337, 2008.

A Routing Protocol for Improving Energy Efficiency in Wireless Sensor Networks

Md. Taybur Rahama, Monir Hossen, and Md. Muminur Rahman

Department of Electronics and Communication Engineering

Khulna University of Engineering & Technology

Khulna, Bangladesh

Sagor.ece@gmail.com, mnrhossen@yahoo.com

Abstract— Wireless sensor network will get wide applicability and acceptance if improved energy efficient routing protocol is implemented. Most of the cluster based hierarchical routing protocols use base station (BS) as a main controller to select cluster heads (CHs) which increases energy consumption and round trip time delay. In this paper, we propose an improved energy efficient routing protocol to maximize the network life time and reduce data delay time. In this proposed protocol, new CH is selected by the current CH instead of the BS based on the maximum residual energy and minimum distance between the CH and a node among all the nodes in a cluster. Unlike the modified LEACH (MLEACH), in the proposed protocol re-clustering is occurred only when the residual energy of the current CHs lower than a threshold value. From the comparison of the simulation results it is proved that the proposed protocol provides lower cluster formation time and network failure with better network fault tolerance and packet delivery ratio than the MLEACH.

Keywords— Energy efficiency; cluster head selection; modified LEACH; clustering technique; routing protocol.

I. INTRODUCTION

Usually, a small sized, and low-power sensor nodes have been deployed for the development in sensor network technology. Recently, the wireless technology have been advances and prompted the deployment of sensor nodes extensively for various applications [1]. The modern wireless sensor network (WSN) also consists of tiny, low power, small memory, and low computation capable sensor nodes. All these sensor nodes of WSN are distributed in random or deterministic manner in a remote places or modern cities. The main function of a sensor node is measurement of any change in the sensing area and pass the sensed data to the sink node. As sensor nodes have smaller power source it requires energy efficient routing protocol to reduce energy consumption and improve WSN lifetime.

In a WSN, entire nodes consist of data processing and communication capabilities. Usually, the sensor nodes forward the sensed data to a sink either directly or through one or more intermediate nodes in hop-by-hop manner. Routing protocols play a vital role to transmit and receive the data among the source and sink nodes. For efficient data transmission and reception processes cluster based hierarchical routing protocols are used in most of the modern WSNs.

This cluster based routing protocols provide efficient data aggregation that can significantly minimize the energy consumption and improve the network lifespan. In the cluster based WSN, a small number of sensor nodes form a group and the whole network is divided into multiple groups these are called clusters [1, 2]. Then in each cluster, two types of nodes are found, i.e., cluster head (CH) and member nodes in the cluster. The member nodes of a cluster sense the physical parameters and forwarded the sensed data to CH. Then the CHs of a network aggregate all data of corresponding member nodes with their own data and forward to the base station (BS). Re-clustering is done only when the remaining energy of any CH is fallen below the certain value.

In the LEACH [3], the CH selection of any cluster is done in random manner that does not consider the remaining energy of nodes. Moreover, in the LEACH, new cluster is formed in each round that is the main reason that a low energy node may become the CH. That is why, this protocol does not provide fault-tolerant property. The nodes lose more energy due to random election of CH and that causes very frequent network failure problem.

In the modified LEACH (MLEACH) [4], the CH selection is done by the BS in every re-clustering period. Still the CH selection by the BS in the MLEACH consumes more energy and increases cluster formation time as well as round trip time. It can reduce the network failure rate than the LEACH protocol. However, it consumes more energy and increase the computation time. The main purpose of this research is to modify the MLEACH protocol to improve the energy efficiency and round trip time.

This paper proposes an improved energy efficient routing protocol for the cluster based WSN. We noticed that the data communication between two nodes consumes more energy than the sensing and computation period of a node. This energy consumption is also increased with the distance between the source and sink nodes. That's why, this paper proposed a routing protocol for the cluster based WSN where the current CH select the next CH only if the remaining energy of the current CH is lower than certain value. This scheme also reduces the data transmission distance during the next CHs election as it elects the next CHs by the current CHs instead of the BS. Based on this scheme we have analyzed different performance parameters in terms of the dead nodes, packet delivery ratio (PDR), average residual energy and total energy consumption for different packet sizes. From the

numerical results we found that the proposed routing protocol outperforms the existing MLEACH protocol based on dead nodes, PDR, and average remaining energy in the nodes.

The paper is organized in five sections: next section explains some existing works. In section III we explain about the network model of the proposed protocol. Section IV presents the performance parameters and simulation results, and section V concludes this paper.

II. RELATED WORKS

In the recent few years, a good number of cluster based hierarchical routing protocols have been proposed. Cluster based routing protocol is widely used for its improved performance.

The LEACH is the most popular hierarchical routing protocol for WSNs [3]. In the LEACH, clusters formation is done depending on the strength of received signal of any node. Each cluster consists of a CH that routes aggregated data of the whole cluster to the destination. Energy consumption is reduced in the LEACH because only CHs transmit the collected data to the BS rather than all the sensor nodes in the cluster. Data processing task, i.e., data gathering and aggregation, in a cluster are performed by the CH. Even though the LEACH is very popular cluster based routing protocol it has some disadvantages such as random selection of CHs in each round. As the selection of CH is random and it is not concerned about the energy strength of the electing CH, the low energy node may be selected as a CH that causes packet loss and network failure.

The MLEACH [4] is the modified version of the LEACH proposed for removing the constraint of random CH selection. In this protocol, the CHs for current round are elected based on the energy strength of the nodes. Only the nodes having higher energy that is above a threshold value may be selected as CHs and thus the low energy node is prevented to become CH. In this way the MLEACH reduces packet loss as well as network failure rate. This process results in an increase of network lifetime.

Most of the leading routing protocols, e.g., LEACH, MLEACH, PEGASIS [5], HEED [6], and Threshold sensitive energy efficient sensor network protocol (TEEN) [7], have not been given attention for reduction of energy consumption during selection of CHs. One thing that motivates us a lot is the relationship between the energy consumptions of data computation and transmission. The relationship in energy consumption during the data transmission and computation is highly dependent on the particular hardware that is used in a node. Normally, 1 nJ energy is required by a sensor processor to compute an instruction whereas Bluetooth transceiver is required about 100 nJ for single bit transmission. It can easily be shown that, for transmitting a data packet of 1 KB over 100 m distance consumes about equal energy to compute three million instructions [8]. That is why, in the proposed scheme, we have focused on developing a new communication protocol to make the cluster based WSN more energy efficient.

III. NETWORK MODEL

The main problem in most of the existing protocols of clustered WSN is the efficient selection of the CHs. In most of the protocols, the CHs are selected by the BS except in the case of LEACH. In this protocol, CHs are selected in a random manner. Due to the selection of the CHs randomly, a low energy node may become the CH of a cluster that reduces network fault tolerance capability. In all other existing protocols, in each re-clustering phase all of each cluster nodes send control packets directly to the BS and the BS selects one node as a CH which has minimum distance and maximum energy among the nodes of a cluster. This process is repeated continuously in every round. However, the distance between each CH and the BS is a substantial factor that requires more energy to exchange data packets between the CHs and BS.

The main idea of the proposed protocol is to avoid the energy consumption due to data transmission by the BS to all the nodes in a cluster. Here, the existing CH of each cluster performs the next CH selection process instead of the BS. The proposed protocol effectively reduces the energy consumption because it avoids longer distance data transmission to the BS by all the nodes in a cluster.

In this proposed protocol, we assume that the sensor network model should have the following properties:

- BS is located at the middle of the network having enough memory size and battery power.
- All sensor nodes are homogenous
- Both the nodes and BS are static.
- Member nodes of a cluster send the sensed data to the CH.
- A node is considered as a dead node only once battery power of that node is lost.

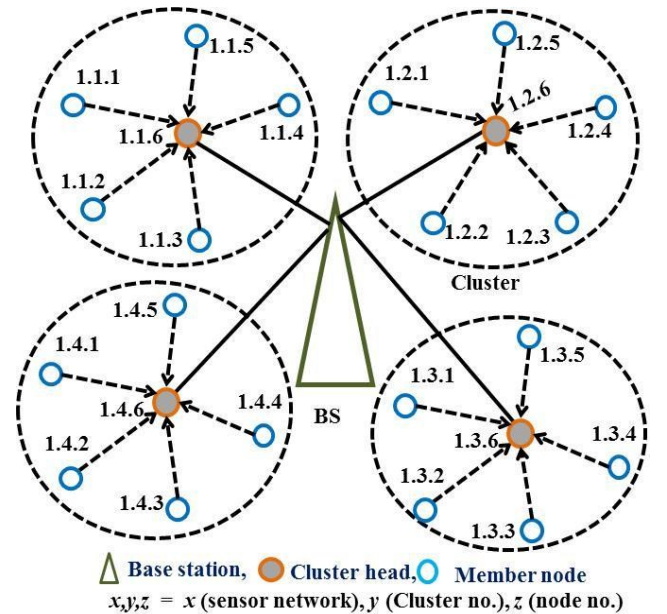


Fig. 1. Proposed network model.

Fig. 1 shows the network architecture of WSN based on the proposed routing protocol. For simplicity, one BS, four clusters, five member nodes and one CH in a cluster that are shown in the figure. The BS can be inside or outside of the sensing area. Fig. 2 shows the flowchart of the proposed protocol. The operation of the proposed protocol also can be described easily by dividing into following three phases.

A. Setup Phase of Cluster

In the setup phase, the BS sends a request to all the sensor nodes to transmit their own locations and energy levels. Each node calculates its distance from the BS depending on the received signal strength indicator (RSSI) and transmits the control packet to the BS. The nodes which have minimum distance from the BS and maximum energy level are elected as CHs. The BS forms the cluster by the help of CHs.

B. Steady State Phase

Once completing the 1st phase of cluster the steady state phase is started. Primarily, the CH allocates time division multiple access (TDMA) scheme to its member nodes for sending the sensed data of all the member nodes of a cluster. The member nodes of a cluster strictly follow their corresponding time slots to transmit data packets. After gathering data packets of all the member nodes, the CH performs data aggregation with its own data and forwards this aggregated data packets to the BS according to code division multiple access (CDMA) principle. Thus one round is completed.

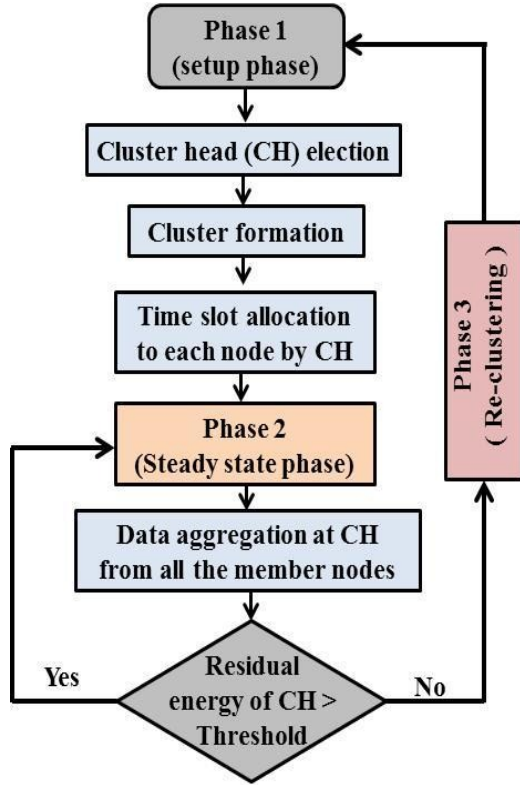


Fig. 2. Flowchart of the proposed routing protocol.

C. Re-clustering Phase

Normally, in the LEACH, the second round will start with re-clustering where new CHs are elected by the BS. In contrast, in the proposed protocol, the current CH checks its own residual energy and if it is still greater than the threshold value [4] it will be continued as a CH for the second round and the cluster performs the steady state phase. Here, the threshold is the predefined average energy of a node and it can be calculated by the following (1),

$$T_j = \frac{1}{m} \sum_{i=1}^m E_i \quad (1)$$

where, T_j is the threshold value of each node, E is the initial energy of each node, and m is the number of nodes in a cluster.

If the residual energy of the current CH is smaller than the threshold value, then it compares the energy level and distances of other sensor nodes in the cluster as all the sensor nodes sent their location, distance and energy level along with the sensing data. The CH of each cluster computes energy level of each member node to find out the node containing higher residual energy with lower distance. As a result time delay and energy consumption for selecting the next CH by the BS is eliminated which improve the energy efficiency as well as network lifetime in the proposed protocol.

IV. PERFORMANCE EVALUATION AND RESULTS

In this section, simulation parameters and performances of the proposed and existing MLEACH protocols have been described. The simulations have been done using computer programs. In order to simulate both the proposed and existing protocols, we have used a single cluster consisting of 10 sensor nodes and they are deployed in a deterministic manner in distances of 200, 80, 110, 100, 160, 150, 140, 120, 70, and 60 m, respectively, from the BS shown in Fig. 3. The simulated results are compared with the results of MLEACH in terms of network lifetime, packet delay, residual energy, and PDR. We also have calculated the number of dead nodes, total energy consumption, and PDR for different packet sizes. In the set up phase, the first CH is selected by the BS and from the second round the next CH is selected by the current CH of each cluster. Table I shows the summary of simulation parameters.

The radio energy model [9, 10] shown in Fig. 4 is used to calculate the energy consumption for both the transmission and reception of data packets between the sources and sink nodes.

For sending k bits for a distance of d , energy consumed by a member node is given by (2),

$$E_{Tx}(k, d) = k(E_{elec} + E_{amp}d^2) \quad (2)$$

where, E_{elec} is the electronics energy, E_{amp} is the amplification energy, $E_{Tx}(k, d)$ is the transmission energy, and k is the number of bits.

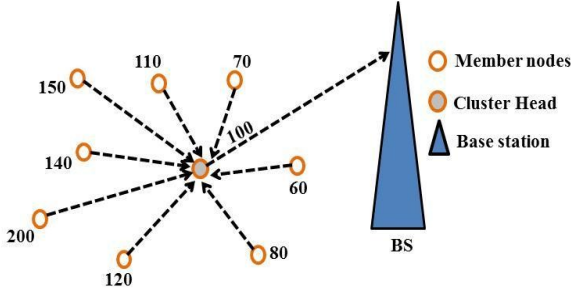


Fig. 3. Network model for simulation.

TABLE I. SIMULATION PARAMETERS

Notation	Description	Value
Channel	Wireless channel	
Propagation type	Two way propagations	
Antenna type	Omni directional	
$M \times M$	Area	200×200
N	Number of the sensors	10
E	Initial energy	1.5J
E_{elec}	Electronics energy	50nJ/bit
E_{DA}	Data aggregation energy	5nJ/bit
L_{Data}	Data packet size	500 Bytes
$L_{control}$	Control Packet size	25bytes
E_{amp}	Amplification energy	100pJ/bit/m ²
Simulation Round		80

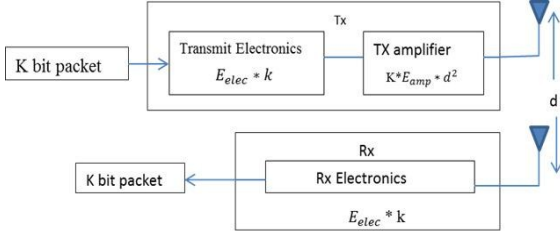


Fig. 4. Radio energy model of sensor nodes.

Again, for sending k bits to a distance of d , the total energy consumption by a CH node is,

$$E_{Tx}(k, d) = k(E_{elec} + E_{DA} + E_{amp}d^2) \quad (3)$$

where, E_{DA} is data aggregation energy.

For receiving the k bits, total energy consumed by both the CH and member nodes in a cluster can be calculated by (4),

$$E_{Rx} = E_{elec} \times k \quad (4)$$

where, E_{Rx} is the reception energy, E_{DA} is the energy of data aggregation.

For different packet sizes, we have calculated PDR, number of dead nodes, average residual energy, and total energy consumption.

A. Packet Delivery Ratio

It is defined as the ratio of the number of received packets to the number of packets sent. The formula for calculation of PDR is given below:

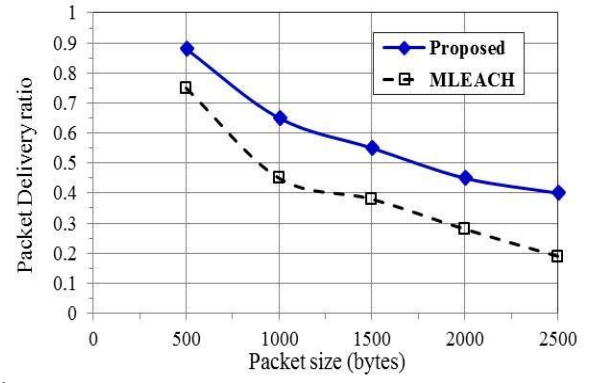


Fig. 5. PDR for different packet sizes.

$$PDR = \frac{\text{No. of packets received}}{\text{No. of packets sent}} \quad (5)$$

PDR for different packet sizes has been calculated for both the proposed and MLEACH protocols. Fig. 5 shows the comparison of PDR between the proposed and existing MLEACH protocols. From the above result, it is found that the PDR of the proposed protocol is higher than that of the MLEACH protocol. With the increase of packet size, the PDR decreases for both the protocols still the proposed protocol provide the higher PDR than the MLEACH protocol. At the maximum packet size of 2500 bytes the proposed protocol provides almost double PDR than the MLEACH.

The main reason of this better PDR in the proposed protocol is that it selects the CHs based on the higher residual energy. That is why, the data forwarding ability of the CH is increased which reduces the packet loss and enhances the PDR of the networks. Since energy consumption during CH selection in the MLEACH is greater than the proposed protocol, the energy losses of the CH and member nodes is faster resulting enough energy is not remain for data transmitting so the node dies and PDR is reduced. With the increases of the packet size the residual energy of the nodes in the MLEACH decreases faster than the proposed protocol.

B. Average Residual Energy

Remaining battery power of a sensor node is one of the most important parameters in the WSN. The throughput, PDR, network lifetime is depended significantly on this parameter. If the total residual energy of all the sensor nodes in a cluster is high then the lifetime of the network will be longer and the end-to-end packet delay will be reduced. Fig. 6 shows the comparison of average residual energy vs packet sizes for both the MLEACH and proposed protocols. From the figure, it is clear that the residual energy of the proposed protocol is higher than the MLEACH because in the proposed protocol all the member nodes should transmit the control packets to relatively shorter distance than the MLEACH, i.e., to the CH and not to the BS. Due to this larger transmission distance of control packets, i.e., to the BS not to the CH, in the MLEACH the energy consumption is high that lead lower residual energy to the member nodes of a cluster. However, as shown in the figure, it is confirmed that the proposed protocol provides

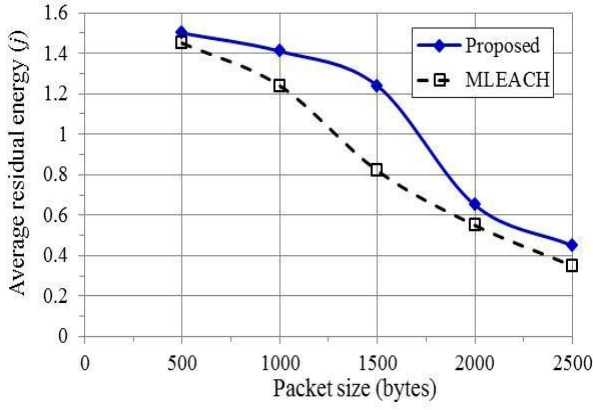


Fig. 6. Average residual energy for different packet sizes.

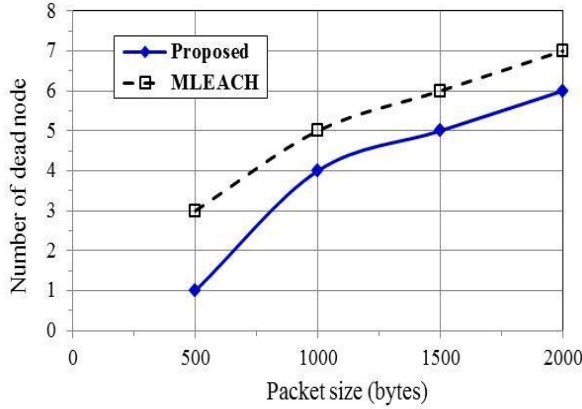


Fig. 7. Comparison of number of dead nodes vs packet sizes.

optimized performance for the packet size of 1500 bytes in terms of residual energy.

C. Number of Dead Nodes

We have calculated the requirement of minimum energy for sending a byte of data packet to the CH. This calculated minimum energy is taken as threshold and below this threshold nodes are unable to send data packet to the CH. Consequently, these nodes are considered as dead nodes. Though these nodes can sense environmental changes to generate data packets but these packets are lost due to energy constrained. For the proposed and existing MLEACH protocols, the same number of rounds for data packet transmission is performed for different packet sizes which are shown in Fig. 7. From the figure, it is seen that the number of dead nodes is higher in the MLEACH than the proposed protocol. This is because of energy loss rate of the MLEACH is higher than the proposed protocol. Consequently, the number of dead nodes in the proposed protocol is decreased which increased the PDR, network lifetime and also fault tolerance capability of the network.

D. Total Energy Consumption

Energy consumption in a WSN is the most important factor. The protocols in WSN should consider this factor. The protocol which has less total energy consumption provides

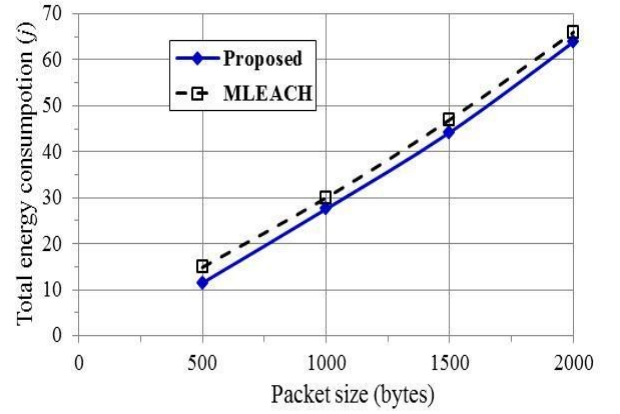


Fig. 8. Comparison of total energy consumption for different packet sizes.

longer network lifetime. Fig. 8 provides comparison of total energy consumptions between the proposed and MLEACH protocols. As we have seen from the figure that the total energy consumption in the proposed protocol is less than that of the MLEACH for any value of the packet size. So, again it can be concluded that the proposed protocol provides less energy wastage to select the CHs than the MLEACH.

V. CONCLUSION

In this paper, we have introduced a new protocol which improves energy efficiency and reduces number of dead nodes in the large WSN. The performance of the proposed protocol has been evaluated and compared to the existing MLEACH protocol in terms of the PDR, average residual energy, number of dead nodes and total energy consumptions. Unlike the existing MLEACH routing protocol, the proposed routing protocol can effectively reduce the total energy consumption and number of dead nodes while increases the PDR and network lifetime for any value of the packet size. This protocol can be applied to large WSNs. Here, the performance of the proposed protocol is evaluated for only ten nodes; practically 100 to 1000 nodes are deployed that is why further analysis is required to evaluate the performance of the real WSNs scenarios.

REFERENCES

- [1] M. Hossen, K. D. Kim, and Y. Park, "Efficient clustering algorithm for delay sensitive PON-based wireless sensor network" in Proc. International Conference on Optical Internet (COIN), Jeju Island, Korea, Jul. 2010.
- [2] M. Hossen, K. D. Kim, and Y. Park, "A PON-based large sensor network and its performance analysis with sync-LS MAC protocol" Arabian Journal for Science and Engineering, vol.38, no. 8, pp. 2115-2123, Aug. 2013.
- [3] H. Balakrishnan, A. Chandrakasan, and W. Heinzelman, "An application-specific protocol architecture for wireless microsensor networks," IEEE Trans. Wireless Commun., vol.1, no.4, pp.660-670, oct. 2002.
- [4] V. K. Subashree, C. Tharini, M. S. Lakshmi, "Modified LEACH: QoS - aware clustering algorithm for wireless sensor network", in Proc. International Conference on Commun. and Network Technologies (ICCNT), pp 119-123, Dec. 2014.

- [5] S. Lindsey and C. S. Raghavendra, "PEGASIS: Power-efficient gathering in sensor information systems," in Proc. Areospace Conference 2002, vol. 3, pp. 1125-1130, Mar. 2002.
- [6] O. Younis and S. Fahmy, "HEED: A hybrid, energy efficient, distributed clustering approach for ad hoc sensor networks," IEEE Trans. on Mobile Computing, vol. 3, no. 4, pp. 366-379, 2004.
- [7] A. Manjeswar, D. P. Agrawal, "TEEN: A protocol for enhanced efficiency in wireless sensor networks," in Proc. International Parallel and Distributed Processing Symposium, San Francisco, pp. 23-27, Apr. 2001.
- [8] G. J. Pottie and W. J. Kaiser, "Embedding the internet: Wireless integrated network sensors," in Proc. Commun. of the ACM, vol.43, no.5, pp. 51-58, 2000.
- [9] Z. Abolfazli, and M. Mahdabi, "A homogeneous wireless sensor network routing algorithm: An energy aware cluster based approach," in Proc. Iranian Conference on Electrical Engineering (ICEE 2014), May, 2014.
- [10] O. Buyanjargal, Y. Kwon, "An energy efficient clustering algorithm for event-driven wireless sensor networks (EECED)," in Proc. International Joint Conference on INC, IMS and IDC, 2009.

Design and Simulation of Maximum Power Point Tracking of Photovoltaic System Using ANN

Md. Samiul Haque Sunny, Abu Naim Rakib Ahmed, and Md. Kamrul Hasan

Department of Electrical and Electronic Engineering (EEE)

Khulna University of Engineering & Technology (KUET)

Khulna-9203, Bangladesh

E-mail: m.k.hasan@eee.kuet.ac.bd

Abstract—Now a days in power generation, renewable energy plays a vital role in which photovoltaic energy generation placed top in the list of the renewable energy because of the easy process of generation. The photovoltaic energy depends on the solar irradiance and the temperature. To get the maximum power from the PV panel, the idea of Maximum Power Point Tracking (MPPT) is arrived. Too many algorithms and controllers have been considered in the past to track the maximum power and to reduce the tracking time and also to improve the efficiency of PV panel. In this paper, Artificial Neural Network (ANN) techniques is proposed to track the maximum power. The proposed method has been evaluated by simulation in MATLAB environment. The simulation results show the effectiveness of the proposed technique and its ability to track the maximum power of the PV panel.

Keywords—Artificial Intelligence(AI); Photovoltaic System; Artificial Neural Network (ANN); Maximum Power Point Tracking (MPPT); PV Panel; MATLAB Simulator.

I. INTRODUCTION

In modern era, the main resources of power generation at the developing country like Bangladesh are oil, coal, and gas. But the present condition of this resource is not good. The decreasing rate of the sources is increasing. Oil, gas, coal and nuclear power energy are unfriendly to the nature. They spread heavily polluted gas and hazard. There are many alternative of the source. Main of them is solar energy, wind mill, water tidal energy, hydroelectric energy etc. This type of renewable energy is quite friendly to the environment. Solar energy obtained from the sunlight is energy in which light energy is converted to thermal or electrical energy. Solar energy is available to everywhere as well as the wind by which wind mill energy can be produced. Those types of energy have no limitation and can be stored easily. Only the initial cost is very high. Then the maintenance cost is very low comparatively to the oil, coal and gas energy station. Among all the renewable energy solar type is very much cheaper, easier and available.

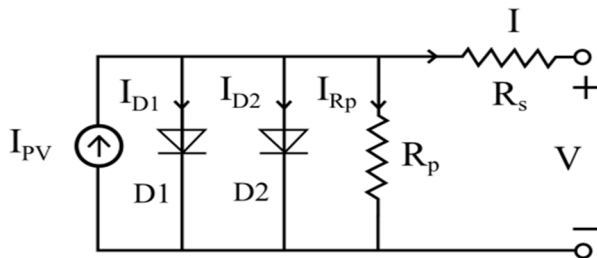


Fig. 1. Two diode solar cell model

The double diode equation is to be considered for the ideality factor n which is the function of voltage across the cell [1].

$$I = I_L - I_{01} \left[e^{\frac{V+IR_s}{a_1 V_{t1}}} - 1 \right] - I_{02} \left[e^{\frac{V+IR_s}{a_2 V_{t2}}} - 1 \right] - \left(\frac{V + IR_s}{R_p} \right) \quad (1)$$

Where, I_L is the illumination current i.e. the current induced by the cell.

$$I_L = I_{L0} \times \frac{I_r}{I_r} \quad (2)$$

I_r is the irradiance (light intensity) in W/m^2 falling on the cell. I_{L0} is the measured illuminated or photo-generated current for the standard irradiance I_{r0} .

I_{01} and I_{02} are the reverse saturation currents of diode1 and diode2 respectively. V is the voltage across the solar cell electrical ports and I is the output current of the cell. V_{t1} is the thermal voltage of the PV module having N_s cells connected in series.

q is the electron charge having value of $1.6 \times 10^{-19} C$ and K is the Boltzmann constant having value of $1.38 \times 10^{-23} J K^{-1}$. T is the measurement temperature parameter value. a_1 And a_2 are the ideality factors of diode1 and diode2 respectively.

Solar cell can produce a lot more power than other renewable sources and create no pollution. The heart of any PV system is the photocell, which converts the light energy from the sun into electrical energy. For practical reason, a large number of such cells are normally connected together to form a PV panel or module in order to produce more output. Panels are then connected in series-parallel combinations to obtain the required terminal voltage and output power. Fig. 1 represents the two diode solar model.

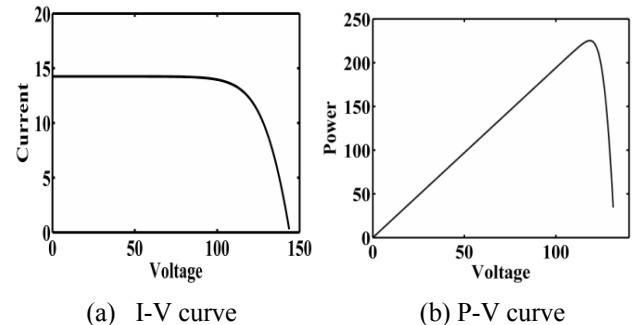


Fig. 2. Ideal I-V and P-V curve of PV system

Fig. 2 shows the ideal characteristics of the solar panel module which are the P-V and I-V curve. The significance of the I-V (Current-voltage) curve of a PV panel is fruitful which describe the conversion ability of the solar panel using existing irradiance (light intensity) and temperature.

II. THEORETICAL BACKGROUND

A. Artificial Neural Network

An Artificial Neural Network (ANN), usually called neural network (NN), is a mathematical model or computational model that is inspired by the structure and functional aspects of biological neural networks [2]. Artificial Neural Networks have been applied to the field of medicine, science, manufacturing, pattern recognition, classification, renewable energy, sports, business, management and finance. Fig. 3 shows the basic three layer neural network architecture. There are several types of architecture of NNs. However, the most widely used NN is feed forward network. In a feed forward network, information flows in one direction along connecting pathways, from the input layer via the hidden layers to the final output layer. In MATLAB we can design and train the Artificial Neural Network [3].

B. Maximum Power Point

When a PV panel receives solar irradiance, it produces power, the product of current and voltage. The operating point on a Power versus Voltage curve that has the highest value of the product of its corresponding voltage and current, or the highest power output varying the irradiance and the temperature. Maximum power of the photovoltaic panel depends on the atmospheric conditions such as ambient temperature and irradiance. In Fig. 4 maximum power point is indicated by the arrow on the P-V curve of the photovoltaic solar model [4-5]. There are many classical methods for maximum power point tracking [6]. Maximum power point tracking (MPPT) is a technique that charge controllers use for PV solar systems to employ and maximize power output. Maximum Power Point Tracking is a digital electronic technique. After getting the output from the panel the charge controller compares it to the battery voltage. From the compared value it finds out what is the best power for which panel can put out to the charge the battery. With this value and converting it to the best voltage we get the maximum ampere into the battery.

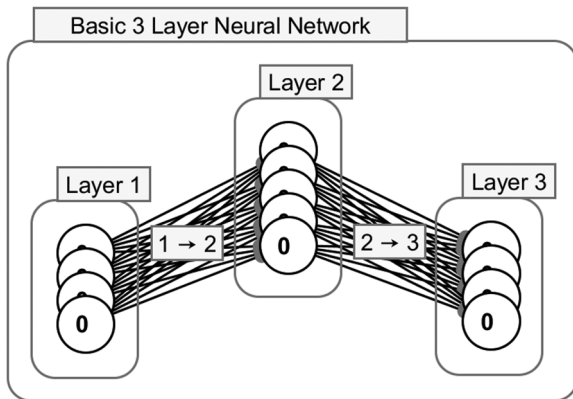


Fig. 3. Representation of basic three layer neural network

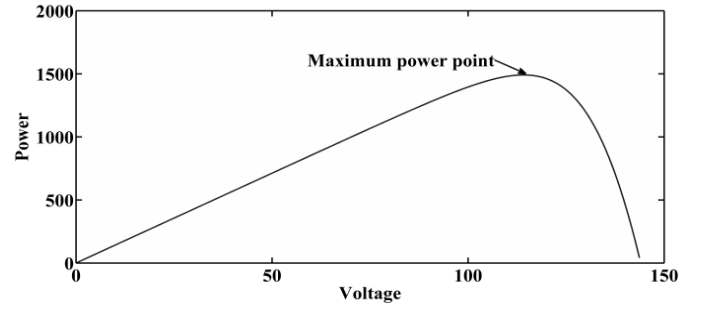


Fig. 4. Maximum power point of a PV curve

III. MAXIMUM POWER POINT TRACKING WITH NEURAL NETWORK

The main advantage of the neural network is it can learn from its training. And during training it updated its weights according to the desired output. A network can be developed with the characteristics of the Photovoltaic system to track the maximum power. A perfectly trained neural network can provide the maximum power at various temperature and irradiance.

A. Proposed Algorithm

To track the optimum power of the Photovoltaic system, artificial neural network method is proposed. ANN method is used before to track the maximum power of the photovoltaic system. But these [7-9] networks were trained with 2 input features such as Voltage and Current or Temperature and Irradiance. 3 hidden layers are used to train in [9]. But a Neural Network can be trained more efficiently if we use more feature in the input. That's why we are proposing a Neural Network with 4 input features and 4 hidden layers for training and optimization. With this method we provide the solar data as input for training the neural network. The trained network will be capable of tracking the maximum voltage and the maximum power of the photovoltaic system in any temperature and irradiance.

B. Schematic Diagram

Fig. 5 represents the schematic diagram of the proposed method. For the irradiance and the temperature solar panel produce voltage and current as the output of the solar panel. This output along with the power works as the inputs of the untrained neural network. After training maximum voltage and maximum power will be the output of the trained network.

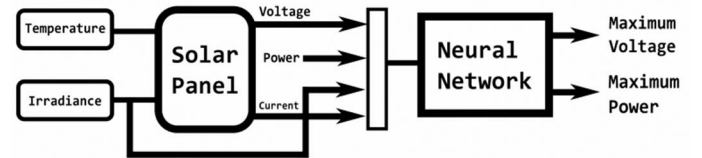


Fig. 5. Schematic diagram of the proposed method

IV. RESULT & DISCUSSION

The simulation of the proposed artificial neural network method for tracking the maximum power of the photovoltaic system is evaluated in the MATLAB platform. For simulation, one day data of a 40 watt solar panel is used as input samples for

training of the neural network. After simulation, the trained network converted into a Simulink model block using gensim.

A. Network Structure

A feed forward neural network is built in custom process for the training. The network has 4 hidden layers. Tansig transfer function is used in the hidden layers and linear function in the output layer. Fig. 6 is the architectural view of the network.

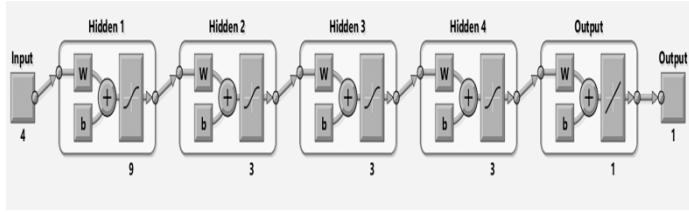


Fig. 6. Architectural view of the network

B. Network Training

The network is trained in a Neural Network (NN) training tool in MATLAB environment. The performance is based on the value of Mean Squared Error (MSE) and the training method is Levenberg-Marquardt. From the Performance, Training state, Error Histogram and Regression it can be observed that the network is trained or not. The Fig. 7 shows the training session of the network. The performance goal was set to zero. Training performance of the network of our proposed model was 1.08×10^{-23} where the performance in [9] is only 5.735×10^{-5} . This result is the identity of the betterment of our research work.

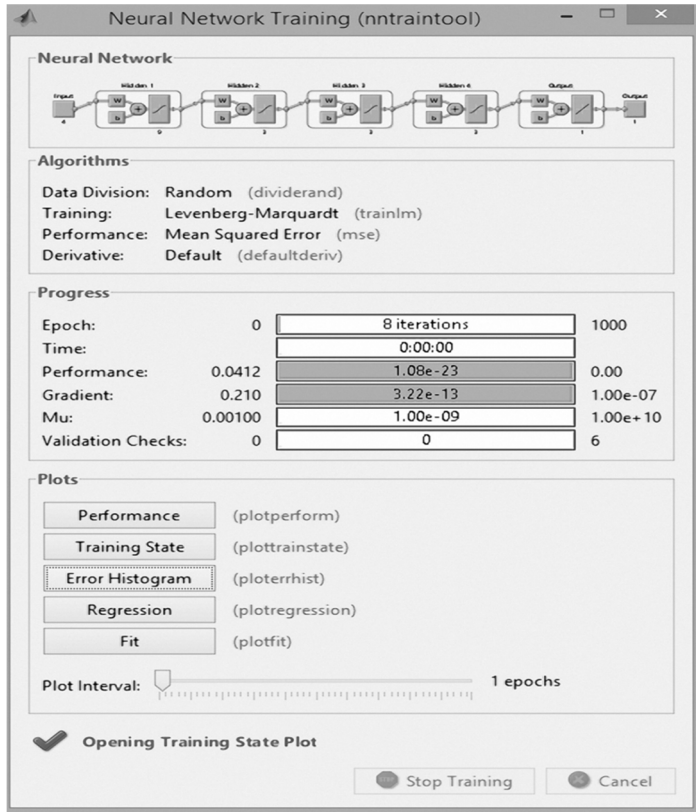


Fig. 7. Training session of the neural network

C. Simulation Results

After the training of the network the results below is observed. These results define the quality of the network. Fig. 8 represents the performance of the neural network. This is a curve of the mean squared error versus the iteration. From the figure we can see that the best performance is 9.2512×10^{-24} when the goal is zero. The best performance is observed at the iteration number 513. Fig. 9 is the regression plot of the network training. The R value indicates the relation between the outputs and the target data of the network training. From the plot we can see that the value of R is 1. That says that there is an exact linear relationship between the outputs and the target. So the training is successfully done. In Fig. 10 the training state of the network at the point of the best performance is described through the gradient, mu and the validation check of the value. From the training state we can see that the value of the gradient is 3.613×10^{-10} , the mu value is 1×10^{-11} and no value fail at the validation test up to the best performance iteration.

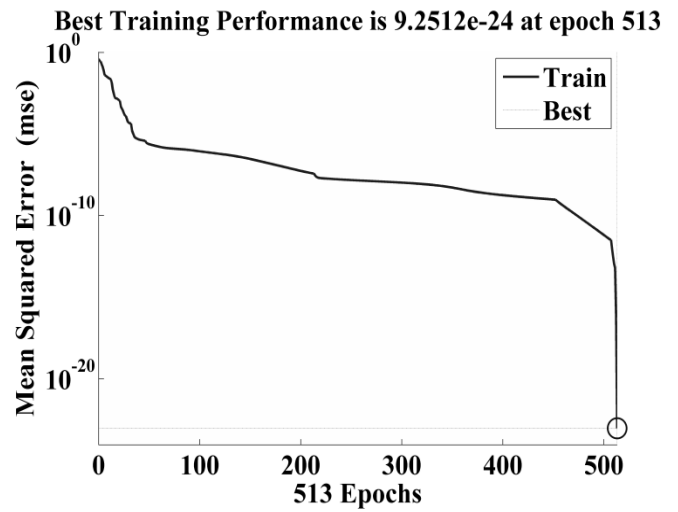


Fig. 8. Performance curve of the network

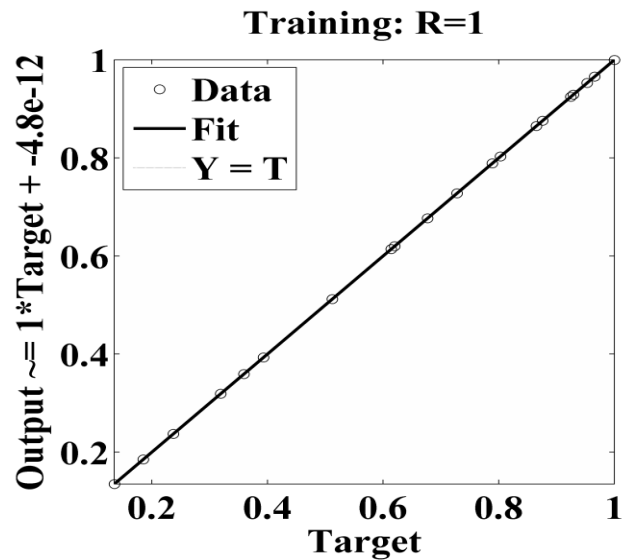


Fig. 9. Regression of the network training

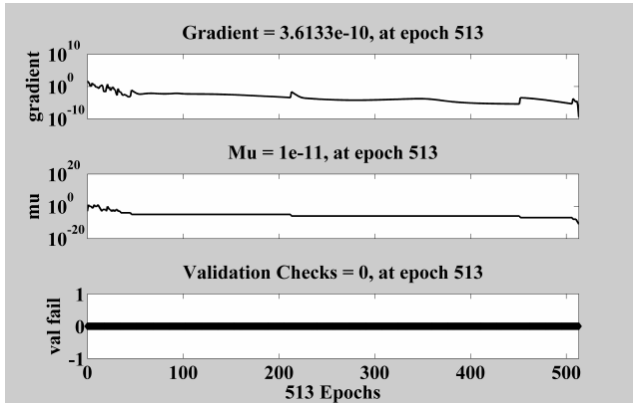


Fig. 10. Neural Network Training State

Error histogram of the training, validation and the test data compared to the zero error is presented in the Fig. 11. Almost all the data is in the zero error line indicates the perfect training of the network.

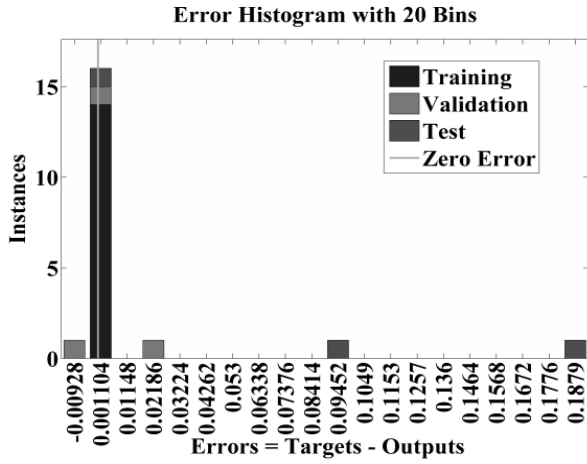


Fig. 11. Error histogram of the training

After training the trained network converted into a simulink block model using the gensim function of the MATLAB. The simulink model of the neural network looks like Fig. 12.

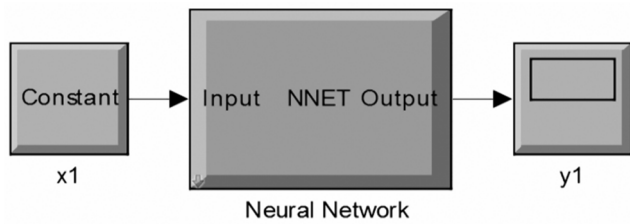


Fig. 12. Simulink block model of the neural network

Fig. 13 gives an idea about the layer arrangement with the weights inside the musk of the block of the neural network. A solar panel model has been built in Simulink platform of 40 watt to simulate the trained neural network. In the solar panel temperature and the irradiance is given as input. Irradiance varied randomly from 89 W/m² to 586 W/m². And the temperature varied from 25° C to 35° C. The trained neural network is added to the solar panel output and simulate with

MPPT algorithm to track the maximum power and the maximum voltage. Fig. 14 and Fig. 15 represent the outputs of the simulation of the neural network with the solar model. Comparing the output results of voltage and power with the result of [7-8] we can see that there is less oscillation and much quicker to track the maximum power and voltage.

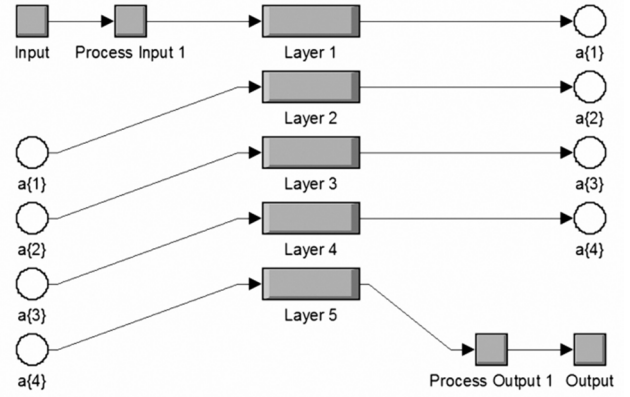


Fig. 13. Layer arrangement inside the musk

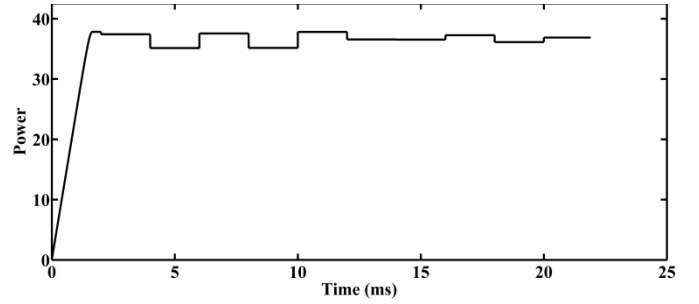


Fig. 14. Maximum power output of the solar panel

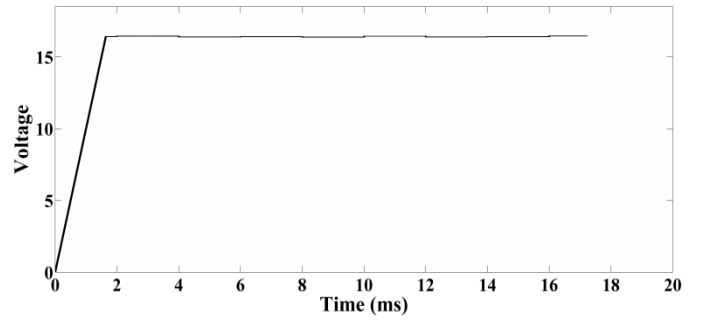


Fig. 15. Maximum voltage output of the solar panel

From the simulation result we get that the maximum power output of the 40 watt solar panel is around 39.62 watt and the maximum voltage output is 17 volt. Only few milliseconds took to track these maximum voltage and corresponding maximum power of the solar panel.

D. Comparative Analysis

There are many algorithms and methods [10-12] for tracking the maximum power point of the photovoltaic system. Comparing efficiency and tracking time with the proposed

neural network methods we can see that there is a high efficiency and lower tracking time of the neural network.

TABLE I. COMPARISON AMONG THE MPPT METHODS

MPPT Tracking Methods	Comparison terms	
	<i>Efficiency</i>	<i>Tracking time</i>
Conventional P&O	97.52%	0.040s
Open circuit Voltage	92.00%	0.080s
Short circuit Current	96.00%	0.150s
Incremental Conductance	97.96%	0.039s
Fuzzy Logic Controller	98.02%	0.033s
Proposed method based on Artificial Neural Network	99.05%	0.002s

The tracking speed of the proposed method is fast and tracking accuracy is very high about 2 ms and 99.05% respectively though the complexity of the algorithm is hard. Its dynamic response is pretty good and steady-state oscillation is less nearly negligible. And the efficiency of our proposed method of MPPT is higher than any other methods.

V. CONCLUSION

For maximum use of solar energy a method has been developed and Simulink model has been made after training. Advanced artificial neural network technology is used to predict the maximum power considering the variation in input levels of solar irradiation and temperature. Artificial neural network performance was 9.2512×10^{-24} when goal was zero. The maximum power point of solar panel is tracked in 2 ms using the proposed algorithm. The algorithm applied to a 40 Watt PV panel and from this the network tracks around 39.62 Watt as maximum power at different irradiance and temperature of 30°C. Tracking time is less than the conventional methods and there are no perturbations and small oscillations.

REFERENCES

- [1] N. Anani, M. Shahid, O. Al-kharji, J. Ponciano, "A CAD Package for Modelling and Simulation of PV Arrays under Partial Shading

Conditions," ELSEVIER Energy Procedia Vol. 42, pp. 397-405, 2013. ISSN 1876-6102.

- [2] M. S. Chauhan, M. Dutta and A. Tiwari, "A Comparative Study of Data Mining Tools on Parkinson's disease," International Journal of Advanced Research in Computer and Communication Engineering, Vol. 5, Issue 1, January 2016. DOI 10.17148/IJARCCE.2016.5162
- [3] M. H. A. Shamisi, A. H. Assi, and H. A. N. Hejase, "Using MATLAB to Develop Artificial Neural Network Models for Predicting Global Solar Radiation in Al Ain City – UAE," Engineering Education and Research Using MATLAB, Oct. 2011.
- [4] H. L. Tsai, C. S. Tu, and Y. J. Su, "Development of generalized photovoltaic model using MATLAB/SIMULINK," Proceedings of the World Congress on Engineering and Computer Science (WCECS), October 22 - 24, 2008, San Francisco, USA.
- [5] M. Villalva, J. Gazoli, and E. Filho, "Comprehensive Approach to Modeling and Simulation of Photovoltaic Arrays," IEEE Transactions on Power Electronics IEEE Trans. Power Electron., vol. 24, no. 5, pp. 1198–1208, 2009.
- [6] F. Dkhichi, and B. Oukarfi, "Intelligent control in photovoltaic systems by neural network," Intelligent Systems and Computer Vision (ISCV), pp. 1 – 5, 25-26 March 2015. DOI: 10.1109/ISACV.2015.7106181
- [7] R. Khanaki, M. A. M. Radzi, and M. H. Marhaban, "Comparison of ANN and P&O MPPT methods for PV applications under changing solar irradiation," IEEE Conference on Clean Energy and Technology (CEAT), pp. 287 – 292, 18-20 Nov. 2013. DOI: 10.1109/CEAT.2013.6775642
- [8] N. Khaldi, H. Mahmoudi, M. Zazi, and Y. Barradi, "Modelling and Analysis of Neural Network and Incremental Conductance MPPT Algorithm for PV Array Using Boost Converter," Advances in Environmental Technology and Biotechnology (AETB), ISBN: 152 978-960-4.
- [9] L. Sindhura, and K. Chaudary, "Artificial Neural Network Implementation for Maximum Power Point Tracking of Optimized Solar Panel," International Journal of Computer Applications (IJCA), vol. 78, no. 10, pp. 1–6, 2013.
- [10] J. S. Kumari, and C. S. Babu, "Comparison of maximum power point tracking algorithms for photovoltaic system," International Journal of Advances in Engineering & Technology (IJAET), Vol. 1, Issue 5, pp. 133-148, 2011. ISSN: 2231-1963.
- [11] L. Bouselham, B. Hajji, and H. Hajji, "Comparative study of different MPPT methods for photovoltaic system," 3rd International Renewable and Sustainable Energy Conference (IRSEC), pp. 1-5, 10-13 Dec. 2015.
- [12] R. Sankarganesh, and S. Thangavel, "Maximum power point tracking in PV system using intelligence based P&O technique and hybrid cuk converter," International Conference on Emerging Trends in Science, Engineering and Technology (INCOSSET), pp. 429 – 436, Tiruchirappalli, Tamilnadu, India, 13-14 Dec. 2012.

Drowsiness Detection for the Perfection of Brain Computer Interface Using Viola-jones Algorithm

Md. Kamrul Hasan*, S. M. Hasnat Ullah, Shantanu Sen Gupta, and Mohiuddin Ahmad
Department of Electrical and Electronic Engineering (EEE)
Khulna University of Engineering & Technology (KUET)
Khulna-9203, Bangladesh
*E-mail: m.k.hasan@eee.kuet.ac.bd

Abstract—Security and reconnaissance applications are prominent BCI paradigms which are less complex and sophisticated if there is no contamination in Electroencephalogram (EEG) signal. The better the quality of EEG signal ensures the better the performance (better Information Transfer Rate (ITR), high Signal to Noise Ratio (SNR), high Bandwidth (BW), and so on) of BCI paradigms. Drowsiness is one of the major contamination in EEG signal that hampers the operation of modern BCI paradigms. In this research, a non-intrusive machine vision based concept is used to determine the drowsiness from the patient which ensure the drowsy free EEG signal. In this proposed system, a camera which placed in a way that it records subjects (BCI Users) eye movement in every time as well as it can monitor the open and close state of eye. Viola-jones Algorithm is applicable for the detection of face as well as state of eye (Open, closed or semi-open) which is the key concern for the detection of drowsiness from the patient's EEG signal. After detecting this drowsiness, decision can be easily made for the perfect operation of BCI.

Keywords—Brain-computer Interface (BCI); BCI Paradigms; Drowsy EEG Signal; Electroencephalogram (EEG); Viola-jones Algorithm.

I. INTRODUCTION

In the modern field of Engineering, brain-computer interfaces (BCI) based on non-invasive scalp Electroencephalography (EEG) have become a progressively active research area [1]-[2]. There are lot of methodology for the monitor or investigation of brain function such that Electroencephalography (EEG), Magneto-encephalography (MEG), functional magnetic resonance imaging (f-MRI), and Positron Emission Tomography (PET). The first technique is too much easier and less costly than other kinds of investigation of brain function. Hence, EEG is the optimal choice for BCI application in BCI paradigm [3]. In EEG based BCI, Steady-state Visual Evoked Potential (SSVEP) acts as a vital raw signal which is processed and extracted for the desirable features. The perfection of any BCI devices depends on the perfection of EEG viz SSVEP signal. A typical SSVEP based healthy EEG signal in frequency (Hz) domain [4] is shown in the Fig. 1. Drowsiness is the state of the person which feels them peculiarly sleepy during the day or night at working time. The presence of drowsiness is the main contamination in EEG signal and also hampers the perfection of BCI. Drowsiness in the EEG signal is the momentary state between consciousness and unconsciousness. Estimation or prediction of drowsiness level

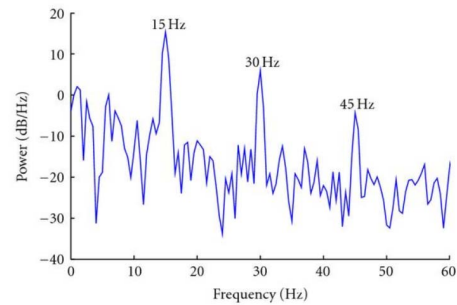


Fig. 1. Typical representation of SSVEP signal in frequency (Hz) domain

in EEG signal for perfect BCI operation using EEG signals is the key research field in the biomedical engineering (BME). The person who contains drowsy may fall asleep in unfortunate situations at any unsuitable time. There are several reasons of drowsiness of the people [5] such as-

- I. Continuous pain
- II. Diabetes
- III. Working long time without taking rest
- IV. Variations in blood sodium density
- V. Medicines of the sleep and antihistamines
- VI. Sleeping time is not sufficient and sleep disorder
- VII. Too much calcium in blood and underactive thyroid

Whatever the cause of the drowsiness, our main target is to determine the level of drowsiness from the patient's EEG signal. Once we detect the drowsiness level then the operation of Brain-computer Interface be like that behavior of EEG signal of the patients of interest.

The basic block diagram of the Human-computer Interaction (HCI) or Brain-computer Interface (BCI) is shown in the Fig. 2. The raw EEG is processed for the suitable features extraction before that raw signal should be filtered for the removing of unwanted portion of the raw signal. But the drowsiness is the prominent contamination in the EEG signal which degrades the BCI or HCI performance. In this research, the proposed work for the drowsiness detection is mentioned using Viola-jones Algorithm which is depicted in Fig. 3. The drowsiness should be removed from the EEG signal of patients as the perfection of BCI depends on the perfection of SSVEP signal.

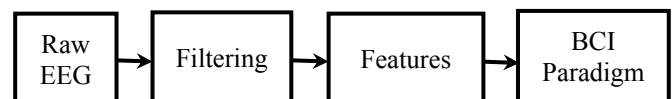


Fig. 2. Generalized block diagram for HCI or BCI representation

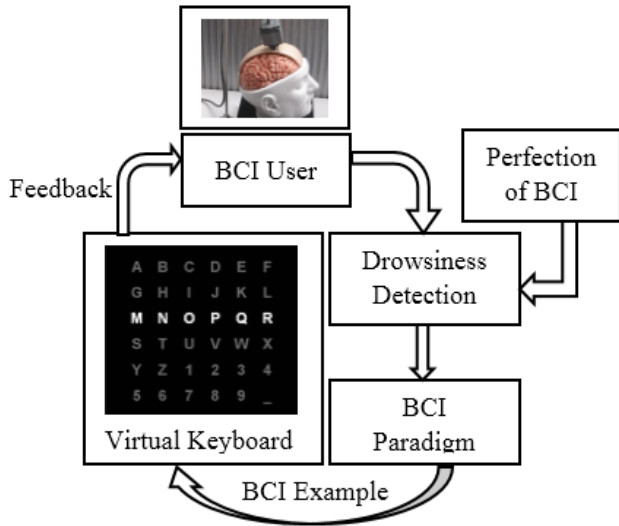


Fig. 3. Concept representation for drowsiness level detection

II. RESEARCH FRAMEWORK

The overall research framework is described in several sections which are noted below:

A. Subject Set-up

The proposed experimental set up is shown in Fig. 4. A monitoring camera is placed in such a way that it can get the whole face of the BCI users whose drowsy level is to be detected. From the captured image, eye can be detected and separated by using Viola-Jones algorithm. After analyzing this detected eye, we can come to a conclusion that the user's eye is close, semi open or fully open and so on. The condition of eye (close, semi open or fully open) is the key concern of the drowsy level detection.

B. Overall Framework

An overall framework is shown in Fig. 5. Image of the BCIs users can be read from the camera. Using Viola-Jones algorithm, firstly face detection and then eye detection have been done with degree accuracy. Then the eye has been cropped and converted it into binary image. Comparing this image with the ideal template the drowsiness level of the BCI users can be estimated. Then it can be decided that the level of drowsiness is acceptable or not.

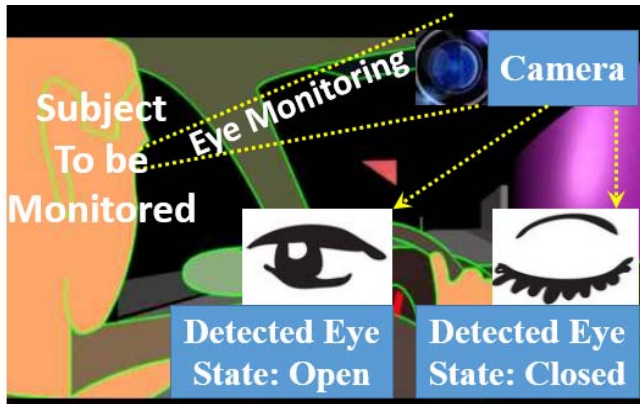


Fig. 4. Monitoring of subject for eye detection

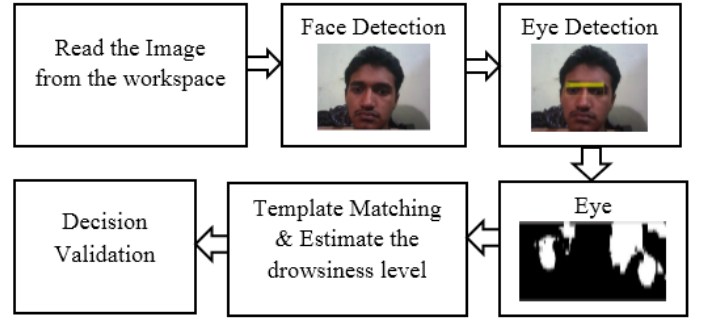


Fig. 5. Representation of overall proposed work

C. Flow Diagram of the Framework

The flow diagram of this framework is shown in Fig. 6. Using a preset camera image of the BCI users have been got. After taking the image, face of the users have been detected and annotated using Viola-jones object detection algorithm. Then again by using Viola-jones algorithm eye has been detected and separated. Comparing the eye with an ideal template which predefined, a decision comes that the eye is closed, semi open or fully open. If it is fully open then it is acceptable and the process can continue. If eye is semi open, the condition is tolerable but an alarm should be provided. If the eye is closed, it will be decided as the abnormal condition and process should stop there.

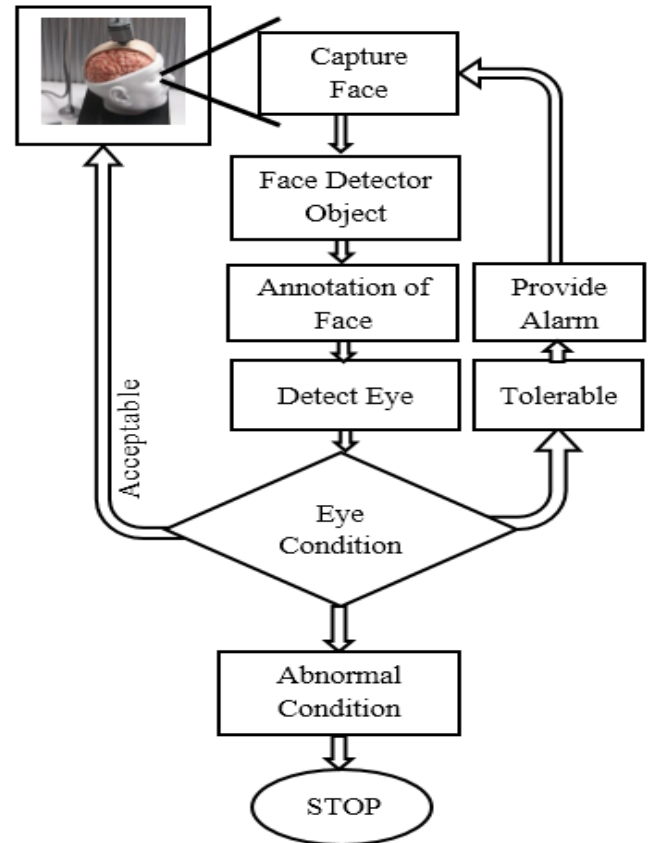


Fig. 6. Flow diagram of proposed work

III. MATHEMATICAL FORMULATION

A. Viola-jones Algorithm

The Viola-jones algorithm is the algorithm for the detection of the object from the framework which is the first object detection framework in the real time detection. In 2001, Paul Viola and Michael Jones proposed this sophisticated algorithm which is primarily used for face detection [6]-[8]. There are some vital advantages of this algorithm which are mentioned below in some points-

- Provides high computation speed
- Most effective feature selection
- Scaling of the features
- Very high detection rate

The features laboring by the uncovering framework generally include the sum of image pixels within rectangular areas. As such, they bear some resemblance to Haar basis functions, which have been used previously in the realm of image-based object detection [9]. Feature types used by Viola and Jones are shown in the Fig. 4. The learning algorithm of Viola and Jones is represented below in step by step-

The speed of evaluated features may not have adequately compensated for their number. Various types of learning algorithm namely Adaptive boosting shortly Ada-Boost is employed by the object detection framework for the finding of best features as well as to train classifiers.

The algorithm used in this manuscript constructs a strong classifier as a linear combination of weighted simple weak classifiers.

$$h(x) = \text{sign} \left(\sum_{i=1}^N a_i h_i(x) \right)$$

Where N is the total number of features used in classifier and a_i is the co-efficient vector.

Here, each and every weak classifier should have threshold function depending on the features vector f_i

$$h_i(x) = \begin{cases} -s_i & \text{if } f_i < \theta_i \\ s_i & \text{Otherwise} \end{cases}$$

The threshold value of θ_i , a_i and $s_i \in \pm 1$ should be clarified during the training of the classifier.

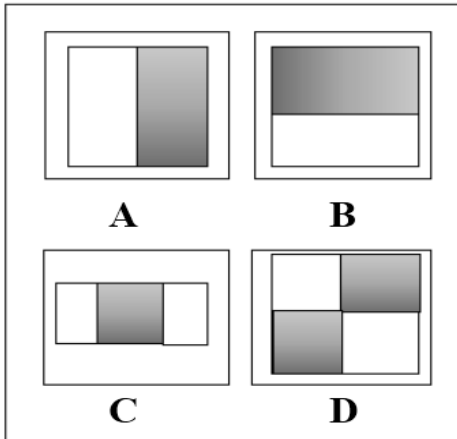


Fig. 7. Visualization of the five Haar-like features i) two rectangle features (A & B), ii) three rectangle features (C) and iii) four rectangle features (D)

Now, the overall algorithm which is primarily used in this research is briefly presented below-

At first, take a value of M, the positive and negative training image which should have labels (x^i, y^i) . Now, set the value of y^i depending on i .

$$y^i = \begin{cases} +1 & \text{if } i \text{ is a image} \\ -1 & \text{if } i \text{ is not a image} \end{cases}$$

- Initialize the value of weight, $w_1^i = \frac{1}{M}$ for any image.
- For each and every features f_i with $i = 1, \dots, N$.
 - To sum the weight to one the weight should be normalized.
 - To minimize the weighted error, the features should be applied to the image for finding the optimum threshold and polarity θ_i and s_i respectively.

$$\theta_i, s_i = \arg \min_{\theta, s} \sum_{i=1}^M w_i^j \varepsilon_i^j$$

$$\text{Where } \varepsilon_i^j = \begin{cases} 0 & \text{if } y^j = h_j(x^j, \theta_i, s_i) \\ 1 & \text{Otherwise} \end{cases}$$

- The weight which is inversely proportional to the error rate should be assigned from a_i to h_i .
 - The weight for w_{i+1}^j , should be reduced for the image which is correctly classified.
- c. At last, set the final classifier to

$$h(x) = \text{sign} \left(\sum_{i=1}^N a_i h_i(x) \right)$$

B. Mathematical Background for Filtering

In signal processing, a filter is a maneuver which processes the signal and eliminates the unwanted component or feature from the signal as shown in Fig. 5.



Fig. 8. Functional block diagram of Filter

In other scene, filtering is one kind of signal processing that completely or partial suppression of some undesirable part of the signal. There are four types of basic filter used in signal processing and electronic devices. These are-

1. Low Pass Filter (LPF)
2. High Pass Filter (HPF)
3. Band Pass Filter (BPF)
4. Band Stop Filter (BSF)

A BPF is the filter that allows signals between two specific frequencies to pass, but that discriminates against signals at other frequencies. The frequency response of an ideal BPF [6] is mention in Eq. (1)

$$G(j\omega) = K \frac{1}{1+\tau_1 j\omega} \frac{\tau_2 j\omega}{1+\tau_2 j\omega} \quad (1)$$

The double cut off frequency of the BPF can be determined where the magnitude falls to the value of $1/\sqrt{2}$. Hence, from the Eq. (1) we can write

$$\frac{K^2}{2} = \frac{K^2 \tau_2^2 \omega^2}{(1+\tau_1^2 \omega^2)(1+\tau_2^2 \omega^2)} \quad (2)$$

If τ_1^{-1} and τ_2^{-1} are separated then the solutions to Eq. (2) will be close to τ_1^{-1} and τ_2^{-1} which are cut-off frequencies of the low pass and the high pass filters respectively.

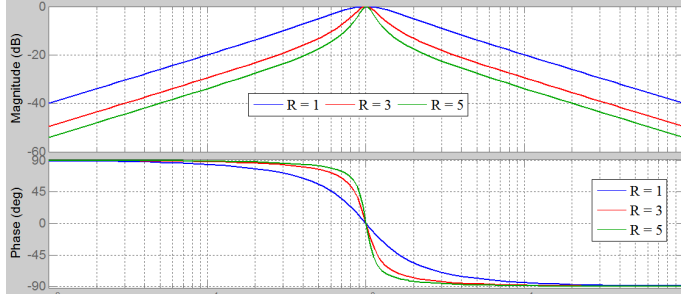


Fig. 9. Magnitude response (Up) and Phase response (Down) of a BPF

The mathematical expressions for the magnitude Eq. (3) and phase Eq. (4) of the frequency response function are mentioned below.

$$|G(j\omega)| = \frac{|K\tau_2\omega|}{\sqrt{1+\tau_1^2\omega^2}\sqrt{1+\tau_2^2\omega^2}} \quad (3)$$

$$\begin{aligned} \text{Phase } G(j\omega) &= \text{Phase}(j\tau_2\omega) - \text{Phase}(1 + j\tau_2\omega) \\ &\quad - \text{Phase}(j\tau_1\omega) \\ &= \frac{\pi}{2} - \tan^{-1} \tau_2\omega - \tan^{-1} \tau_1\omega \end{aligned} \quad (4)$$

The magnitude and phase response of a Band Pass Filter (BPF) at different frequency is shown in Fig. 6. From the Fig. 6. It is noticeable that the greater the value of the resistive part of the filter (BPF) the magnitude response is too flat. The corresponding pole-zero response, overshooting and nyquist frequency response is depict in the Fig. 7. For $R = 3 \text{ Ohms}$, $C = 2 \text{ Farad}$ and $L = 2 \text{ Henry}$ the overshooting time is less and system is more stable as mention in the Fig. 7.

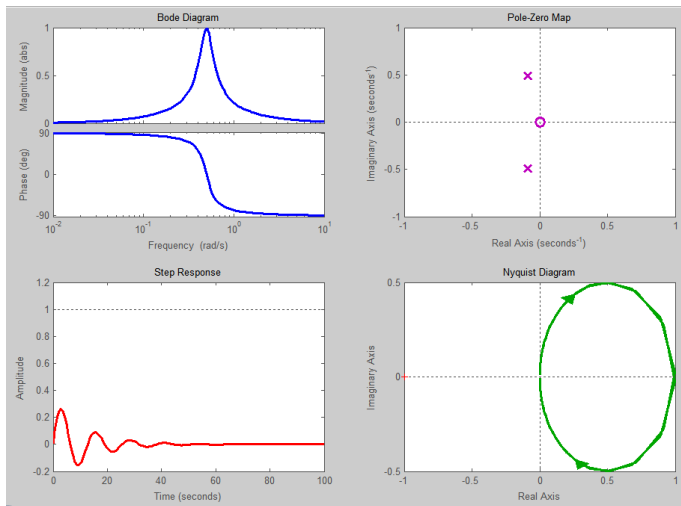


Fig. 10. Pole-zero, overshooting and nyquist response of BPF

IV. SIMULATION RESULTS AND DISCUSSION

The face (Output from the camera) taking from the workspace is loaded into the MATLAB. After taking this face, eye is detected using Viola-Jones algorithm. A typical detected eye is shown in the Fig. 11. Fig. 11(a) is for the closed eye of the BCI performers (Users) and Fig. 11(b) for the open eye of the BCI performers (Users).

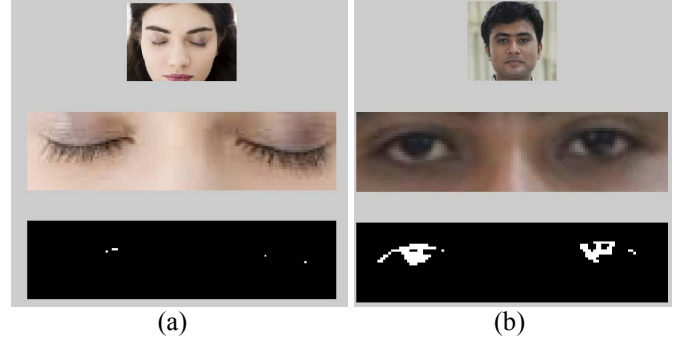


Fig. 11. Detected eye using Viola-jones Algorithm a) eye closed state b) eye open state

The drowsiness level due to various eye state (Open, Close, Partially open or close, medium open) is shown in the Fig. 12. From the Figure, it is noticeable that during the fully eye close state there is huge amount of drowsiness in the EEG signal comparing fully eye open state. There is a significant amount change in drowsiness from partially eye close state to fully eye close state.

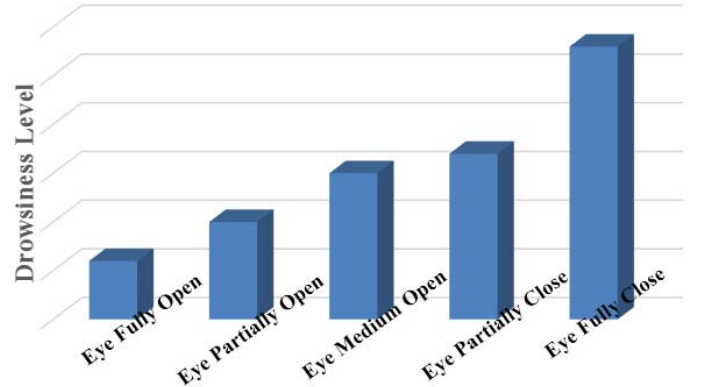


Fig. 12. Drowsiness level due to various eye states

The relaxation level on alpha and beta band (after filtering the raw EEG signal) of BCI performer is shown in the Fig. 13 with respect to drowsiness level. The more power of alpha band and less power of beta band indicates that subjects are in more relax state [10]. From the Fig. 13, it is also clear that at beginning of the BCIs performance, the relaxation is at same point but after time when drowsiness increases due to tiredness the slope of alpha band relaxation level is sharper than the slope of beta band relaxation level. It indicates that the alpha band power is more increased than beta band power. As a consequence the BCI performers feel more relax and ultimately the performance of BCI paradigms fall down.

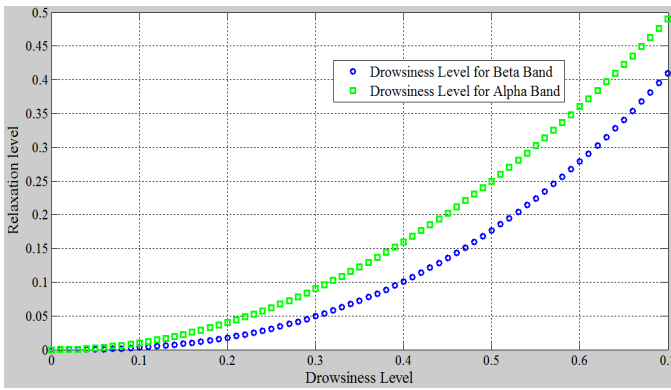


Fig. 13. Relaxation level on EEG signal band (Alpha and Beta) due to drowsiness level of BCI performers.

V. CONCLUSION AND APPLICATIONS

The drowsiness is one of the most dominated contamination which degrades the quality of EEG signal as well as it will degrade the perfection of BCI paradigms. In this manuscript, a simple technique for the drowsiness detection is proposed which is based on the eye detection technique based on the Viola-Jones algorithm. From the simulation output of this proposed system, the efficient detection of face as well as the level of drowsiness is focused. The power level is more for the alpha band than beta band of EEG signal for the increasing value of drowsiness. The more power of the alpha band of EEG indicates the more relaxation of the BCI performers. The more the relaxation of the BCI performers there is a quality degradation of SSVEP based EEG signal. Since SSVEP depends on the visualization effects on the eye. Our proposed system can easily applicable to the Intelligent Transportation System (ITS) because by using this system the fatigue condition can be monitored which can reduce the road accident all over the world.

REFERENCES

- [1] M. K. Hasan, R. Z. Rusho, and M. Ahmad "A Direct Noninvasive Brain Interface with Computer Based On Steady-State Visual-Evoked Potential (SSVEP) With High Transfer Rates" International Conference on Advances in Electrical Engineering (ICAEE), 2013, IUB, Dkaka, Bangladesh. DOI: 10.1109/ICAEE.2013.6750360.
- [2] Y. Wang, R. Wang, X. Gao, B. Hong, and S. Gao, "A Practical VEP Based Brain-Computer Interface", IEEE Transactions On Neural Systems and Rehabilitation Engineering, vol.14, no. 2, pp. 234-240, June 2006. DOI: 10.1109/TNSRE.2006.875576.
- [3] M. K. Hasan, R. Z. Rusho, T. M. Hossain, T. K. Ghosh, and M. Ahmad, "Design and Simulation of Cost Effective Wireless EEG Acquisition System for Patient Monitoring", International Conference on Informatics, Electronics and Vision (ICIEV), 2014, DU, Dhaka, Bangladesh. DOI: 10.1109/ICIEV.2014.6850797.
- [4] Available at: www.hindawi.com/journals/cin/2010/702357/fig2/
- [5] Drowsiness. Available at: www.nlm.nih.gov/medlineplus/ency/article/003208.htm
- [6] Viola-Jones object detection framework. Available at: en.wikipedia.org/wiki/Viola%E2%80%93Jones_object_detection_framework
- [7] P. Viola, M. Jones, "Rapid object detection using a boosted cascade of simple features", Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2001. CVPR 2001. DOI: 10.1109/CVPR.2001.990517.
- [8] P. Viola, M. Jones, "Robust Real-time Object Detection", International Journal of Computer Vision (IJCV), May 2004, Volume 57, Issue 2, pp 137-154. DOI: 10.1023/B:VISI.0000013087.49260.fb.
- [9] C. P. Papageorgiou, M. Oren, and T. Poggio, "A General Framework for Object Detection", 16th International Conference on Computer Vision, Bombay. DOI: 10.1109/ICCV.1998.710772.
- [10] M. K. Hasan, M. S. Hossain, T. K. Ghosh, and M. Ahmad, "A SSVEP Based EEG Signal Analysis to Discriminate the Effects of Music Levels on Executorial Attention", American Journal of Bioscience and Bioengineering, Science Publishing Group, vol. 3 , issue 3-1, pp. 27-33 , May, 2015. DOI: 10.11648/j.bio.s.2015030301.15.

A Polarization Maintaining Hexagonal Photonic Crystal Fiber for Residual Dispersion Compensation

M. R. Hasan^{1,*}, M. S. Anower¹, M. M. Rashid¹, S. Ali²

¹Department of EEE, Rajshahi University of Engineering & Technology, Rajshahi-6204, Bangladesh

²School of Computer and Communication Engineering, Center of Excellence, Advanced Communication Engineering Cluster, Pauh Putra Campus, Arau 02600, Perlis, Malaysia

*Email: ruetrabu@gmail.com

Abstract— A defected core hexagonal photonic crystal fiber is proposed to simultaneously obtain large negative dispersion and high birefringence by using finite element method. Simulation results demonstrate that by properly tuning the dimension of the central elliptical air hole, it is possible to achieve very high negative and flattened dispersion characteristics over a wide range of wavelengths. The proposed PCF exhibits large average negative dispersion of -531.3 ps/(nm.km) over 205 nm (1360-1565 nm) and -454.4 ps/(nm.km) over 205 nm (1420-1625 nm) wavelength bands for the fast and slow axis modes, respectively. In addition to this, the proposed fiber also shows high level of birefringence about 0.01168 at the excitation wavelength of 1.55 μm . Other important modal properties such as confinement loss, splice loss and effective area are also reported thoroughly.

Keywords— Photonic crystal fiber; finite element method; residual dispersion compensation; birefringence

I. INTRODUCTION

In recent years, photonic crystal fibers (PCFs) have been extensively studied due to their flexibility of tailoring unusual optical characteristics including ultrahigh birefringence, large nonlinear coefficient, ultra-flattened dispersion, low confinement loss and so on [1]. Among those, ultra-flattened dispersion and ultra-high birefringence are the most interesting features. Ultra-high flattened dispersion PCF can be used as a potential means for residual dispersion compensator in wavelength division multiplexing (WDM) system [2]. On the other hand, highly birefringent PCF is necessary for high-data rate optical communications especially in 40 Gbps systems as polarization maintaining fiber (PMF) [3]. In contrast to pure silica, other materials such as different compositions of soft glasses can be a suitable choice for engineering dispersion properties. Soft glasses are also known as compound glasses that exhibit comparatively lower melting temperatures than those of silicon [4]. Their typical melting temperature ranges over 200-1000°C, which is much lower than that of silica (1800-2300°C). This makes the fabrication process potentially easier [4]. Moreover, the high refractive index of soft glass provides large index contrast that permits precise control of chromatic dispersion.

A long haul data transmission system uses single mode fiber (SMF) that exhibits nonzero flattened dispersion around 12 to 22 ps/(nm.km). To compensate this dispersion over a long distance an efficient way is to use of dispersion compensating fiber (DCF) [5]. Unfortunately, a DCF can perfectly compensate the accumulated positive dispersion at a single

wavelength, thus there always remains some dispersions, which is known as residual dispersion. Therefore, another dispersion management technique must be used to compensate this residual dispersion. Although several techniques are available regarding this purpose but residual dispersion compensating fiber (RDCF) has gained increased attention since they are easy to machine. A good RDCF should have the characteristics of large negative and flattened dispersion over a wide range of wavelengths [2]. The length of the RDCF should be kept smaller in order to reduce both insertion loss and manufacturing cost. With this in mind, several research outcomes have been reported in the literature. A PCF proposed by Franco *et al.* [6] showed low average dispersion of -179 ps/(nm.km) with dispersion variation (ΔD) around 2.1 ps/(nm.km) over S + C + L + U wavelength bands. Tee *et al.* [7] has reported a PCF structure having average dispersion of -457.4 ps/(nm.km) with ΔD around 12.7 ps/(nm.km) over E + S + C + L + U wavelength bands. However, the main drawback of this structure is the high sensitivity to the dimension of the central defect core that needs additional care in the fabrication process. Apart from the hexagonal structure, few PCFs have been proposed with octagonal, decagonal and spiral symmetry. An equiangular spiral type PCF have been reported in [8] that offers high average dispersion of -393 ps/(nm.km) with ΔD around 10.4 ps/(nm.km) over the wavelength range 1.35 - 1.65 μm . This PCF also shows high birefringence of 0.0278 at 1.55 μm . Due to the complexity of this structure, it is almost impractical to fabricate. An elliptical shaped core octagonal PCF has reported in [9] that showed high negative dispersion of -465.5 ps/(nm.km) with ΔD around 10.5 ps/(nm.km) and high birefringence of 0.0268 at 1.55 μm . However, fabrication of this PCF is not easy because of using different sized air holes in the same ring. Most recently, a decagonal PCF has been proposed by Selim *et al.* [10] having very large average dispersion of -558.96 ps/(nm.km) with ΔD around 9.7 ps/(nm.km) over S + C + L + U wavelength bands. The fabrication of this PCF is too difficult due to the presence of six elliptical air holes in the core region. On the other hand, in Refs. [6-7] the birefringence property has not been studied. Besides, ultra-flattened negative dispersion in either fast or slow axes mode has been discussed for almost all structures and the dispersion phenomena in both axes mode has been disregarded so far.

In this paper, we report on a defected core soft glass hexagonal photonic crystal fiber (HPCF) and its guiding characteristics are numerically investigated by using finite element method. It

is demonstrated that with optimal design parameters, the proposed structure shows ultra-high and flattened dispersion over a wide range of wavelength and high level of birefringence about 0.01168 at 1.55 μm . Other important parameters such as splice loss, confinement loss and fabrication feasibility are also discussed. The excellent guiding properties of the proposed fiber make it a suitable candidate for effective residual dispersion compensation and fiber sensing applications.

II. DESIGN METHODOLOGY OF THE PROPOSED HPCF

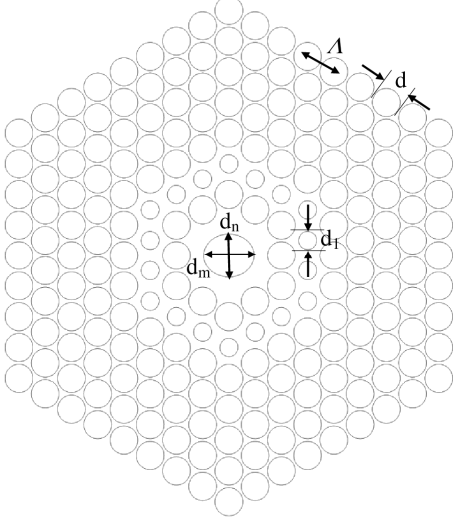


Fig. 1. Transverse cross-section of the proposed HPCF with optimized design parameters of Λ , d , d_1 , d_m and d_n .

The cross-sectional view of the proposed HPCF with optimized design parameters of Λ , d , d_1 , d_m and d_n is shown in Fig. 1. It contains seven air-hole rings ($N_r=7$). The number of air hole selected this way offers low confinement loss. The background material is chosen as soft glass (commercially known as Schott BK7) having refractive index of 1.52. The proposed structure is formed by removing all air holes from first ring of the conventional hexagonal symmetry. Thus, the second ring acts as the immediate first ring. The cladding is formed by regular triangular lattice of air holes. An elliptical air-hole has inserted in the core region in order to achieve high birefringence and large negative dispersion [8]. By varying the dimensions of semi-axis length of elliptical air-hole, it is possible to obtain large negative dispersion with small dispersion variation [8]. The center-to-center spacing between two neighboring air holes is denoted by the common pitch Λ . During the simulation, the value of Λ has kept constant for whole structure. The air-hole diameter of 1st, 3rd, 4th, and 5th ring has the identical value of d . The diameter of 1st air-hole ring is made larger in order to achieve large birefringence and tight electromagnetic field confinement [3]. The air-hole diameter of 2nd ring d_1 is kept relatively smaller than the other air-hole rings in order to improve the flatness of negative dispersion and control the residual dispersion slope [11]. Here, d_m and d_n are major and minor axes of the elliptical air hole at the center of the core, respectively.

III. SIMULATION RESULTS AND DISCUSSIONS

The commercially available state-of-the-art COMSOL v.5.0 software has been used to design and characterization of the proposed HPCF. This software is based on finite element method (FEM). A circular perfectly matched layer (PML) boundary condition has applied to account for the confinement loss. The complete mesh contains 49,104 triangular elements, 4,736 edge elements and 592 vertex elements. The PML thickness has been chosen about 15% of the total fiber radius. The chromatic dispersion (D), birefringence (B) and confinement loss (L_c) can be expressed by [2]

$$D = -\frac{\lambda}{c} \frac{d^2 \text{Re}[n_{\text{eff}}]}{d\lambda^2} \text{ ps/nm/km} \quad (1)$$

$$B = |n_{\text{eff}}^x - n_{\text{eff}}^y| \quad (2)$$

$$L_c = 8.686 \times k_0 \text{Im}[n_{\text{eff}}] \times 10^3 \text{ dB/km} \quad (3)$$

where λ is the operating wavelength, c is the speed of light, $\text{Re}[n_{\text{eff}}]$ and $\text{Im}[n_{\text{eff}}]$ are the real and imaginary part of the effective refractive index, and k_0 is the free space wave number.

Fig. 2 shows the dispersion curve of the proposed HPCF with optimized design parameters of $\Lambda=0.80 \mu\text{m}$, $d/\Lambda=0.95$, $d_1/\Lambda=0.60$, $d_m=0.67 \mu\text{m}$ and $d_n=0.57 \mu\text{m}$. As seen from the figure, an ultra-flattened negative dispersion of -531.3 ps/(nm.km) with ΔD of 11.5 ps/(nm.km) over 1360-1565 nm (shown in solid red line) and -454.4 ps/(nm.km) with ΔD of 9.4 ps/(nm.km) over 1420-1625 nm wavelength bands (shown in solid green line) are achieved for the fast and slow axis modes, respectively. To obtain these optimum design values a simple strategy has been applied. As seen from the Fig. 1, there are five degrees of freedom namely Λ , d/Λ , d_1/Λ , d_m and d_n . All these parameters are separately adjusted to achieve the optimum values.

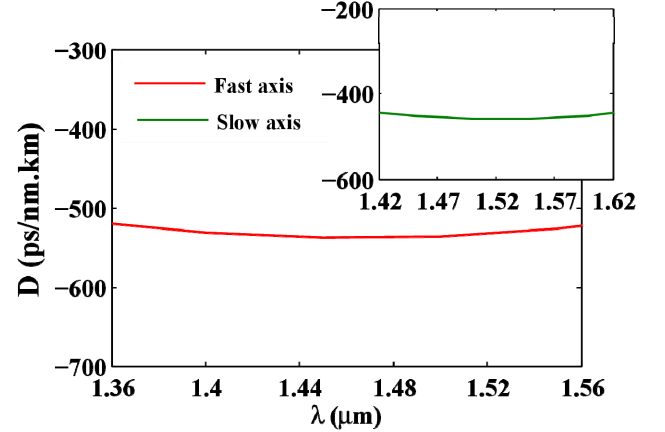


Fig. 2. Dispersion of the proposed fiber as a function of wavelength with optimum design parameters of $\Lambda=0.80 \mu\text{m}$, $d/\Lambda=0.95$, $d_1/\Lambda=0.60$, $d_m=0.67 \mu\text{m}$ and $d_n=0.57 \mu\text{m}$ for fast axis (indicated by red solid line) and slow axis (indicated by green solid line) modes.

First, the influence of varying air-hole diameter d/Λ on the dispersion characteristics has been analyzed. In this study, we only change the value of d/Λ to 0.89, 0.92 and 0.95 while

fixing the other parameters to $\Lambda=0.80 \mu\text{m}$, $d_1/\Lambda=0.60$, $d_m=0.67 \mu\text{m}$ and $d_n=0.57 \mu\text{m}$. The solid line indicates the fast axis and dotted line indicates the slow axis modes, which is illustrated in Fig. 3. For fast axis the corresponding average dispersions are -512.60, -522.48 and -531.30 ps/(nm.km), respectively in the wavelength bands ranging from 1360 to 1565 nm. The ΔD s for fast axis are 18.33, 16.47 and 11.50 ps/(nm.km), respectively. On the contrary, for slow axis the corresponding average dispersions are -438.48, -447.94 and -454.40 ps/(nm.km), respectively in the wavelength bands ranging from 1420 to 1625 nm. The ΔD s for slow axis are 25.30, 17.63 and 9.40 ps/(nm.km), respectively. When d/Λ is 0.89 and 0.92, ΔD for each case is higher for both axes modes. Therefore, the optimum design value of d/Λ has considered as 0.95 due to small ΔD for both axes.

The influence of the geometric parameter d_1/Λ on the dispersion properties is shown in Fig. 4. In the simulation, d_1/Λ is chosen as 0.57, 0.60 and 0.63 while rest of the parameters are fixed to $\Lambda=0.80 \mu\text{m}$, $d/\Lambda=0.95$, $d_m=0.67 \mu\text{m}$ and $d_n=0.57 \mu\text{m}$. In the figure, dispersion characteristics of the fast and slow axis modes are shown by solid and dotted lines, respectively. For fast axis the corresponding average dispersions are -563.32, -531.30 and -504.60 ps/(nm.km), respectively covering the 1360 to 1565 nm wavelength bands. The ΔD s for fast axis are 19.70, 11.50 and 15.80 ps/(nm.km), respectively.

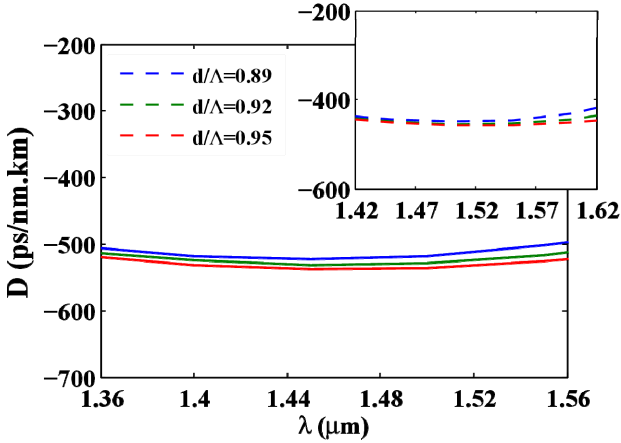


Fig. 3. Illustration of varying d/Λ on the dispersion characteristics for fast axis (indicated by solid line) and slow axis (indicated by dotted line) modes.

The ΔD s for slow axis are 20.10, 9.40 and 10.43 ps/(nm.km), respectively. Here, it can be seen that with d_1/Λ of 0.60 the average dispersion is moderate with lowest ΔD . With of d_1/Λ 0.57, average dispersions of both fast and slow axes increase but at the same time ΔD s are also increased. On the other hand, when d_1/Λ is 0.63, average dispersion is reduced significantly although flatness is improved. Therefore, considering both average dispersion and ΔD the optimum value of d_1/Λ is chosen as 0.60.

The effect of changing Λ on the dispersion has been studied considering other geometrical parameters with $d/\Lambda=0.95$, $d_1/\Lambda=0.60$, $d_m=0.67 \mu\text{m}$ and $d_n=0.57 \mu\text{m}$. With Λ of 0.79, 0.80 and 0.805 μm the resulting dispersion variation are shown in the Fig. 5 where dispersion characteristics of the fast and

slow axis modes are shown by solid and dotted lines, respectively.

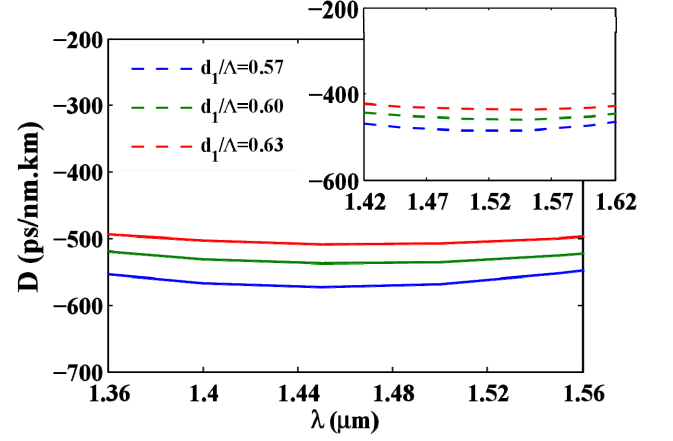


Fig. 4. Illustration of varying d_1/Λ on the dispersion characteristics for fast axis (indicated by solid line) and slow axis (indicated by dotted line) modes.

For fast axis the calculated average dispersions are -567.54, -531.30 and -510.72 ps/(nm.km), respectively in the wavelength bands ranging from 1360 to 1565 nm. The calculated ΔD s in fast axis are 19.50, 11.50 and 20.27 ps/(nm.km), respectively. For slow axis the calculated average dispersions are -480.76, -454.40 and -438.56 ps/(nm.km), respectively in the wavelength bands ranging from 1420 to 1625 nm. In slow axis, ΔD s have found about 22.60, 9.40 and 12.20 ps/(nm.km), respectively. Notice that with Λ of 0.79 μm the resulting average dispersion in both fast and slow axis modes are higher although flatness is low. On the other hand, with $\Lambda=0.805 \mu\text{m}$ results a significant reduction of average dispersion with increase of ΔD . But setting $\Lambda=0.80 \mu\text{m}$ can ensure the moderate average dispersion and improved flatness so the optimum Λ is chosen as 0.80 μm .

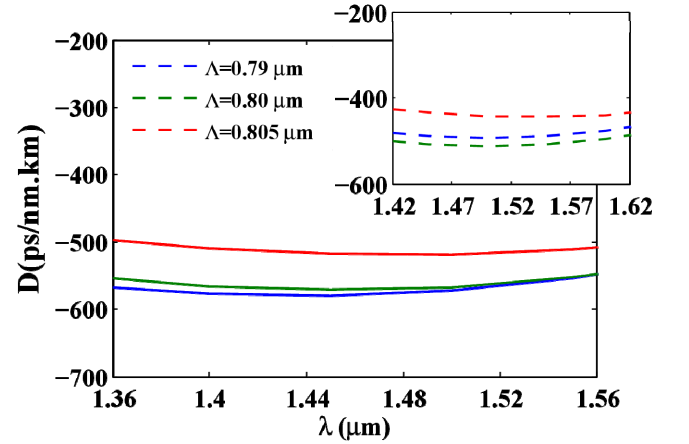


Fig. 5. Illustration of varying Λ on the dispersion characteristics for fast axis (indicated by solid line) and slow axis (indicated by dotted line) modes.

Fig. 6 depicts the effect of changing d_m on the dispersion properties with other geometrical parameters of $\Lambda=0.80 \mu\text{m}$, $d/\Lambda=0.95$, $d_1/\Lambda=0.60$ and $d_n=0.57 \mu\text{m}$. Now, d_m is varied from 0.64, 0.67 and 0.70 μm , the resulting dispersion is shown by solid line for fast axis and dotted line for slow axis modes. For

fast axis the calculated average dispersions are -474.38, -531.30 and -594.14 ps/(nm.km), respectively in the wavelength bands ranging from 1360 to 1565 nm. The calculated values of ΔD in fast axis are 22.50, 11.50 and 27.30 ps/(nm.km), respectively. In slow axis mode, average dispersions are have found about -428.96, -454.40 and -478.50 ps/(nm.km), respectively in the wavelength bands from 1420 to 1625 nm. The ΔD s for slow axis are 13.20, 9.40 and 14.23 ps/(nm.km), respectively. In this case, with d_m of 0.64 μm the resulting average dispersion significantly falls in both fast and slow axis and ΔD increases at the same time. However, with $d_m = 0.70 \mu\text{m}$ results an increase of average dispersion in both axis modes but ΔD goes higher. On the other hand, with $d_m = 0.67 \mu\text{m}$ offers large average dispersion with smallest value of ΔD . Therefore, in this study we select $d_m = 0.67 \mu\text{m}$ as the optimum value.

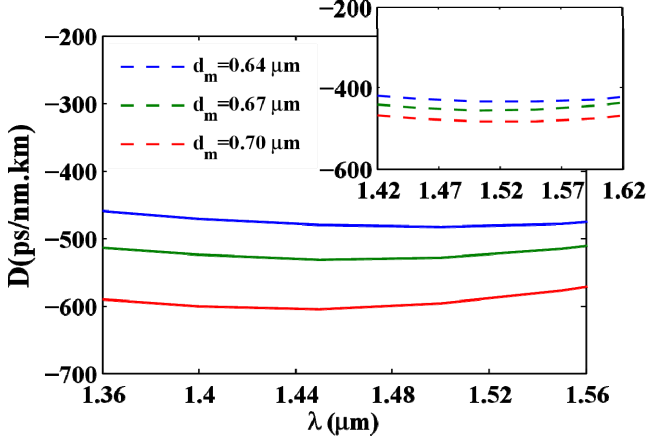


Fig. 6. Effect of changing d_m on the dispersion characteristics for fast axis (indicated by solid line) and slow axis (indicated by dotted line) modes.

Fig. 7 reveals the influence of varying d_n on the dispersion characteristics while other parameters set to $\Lambda = 0.80 \mu\text{m}$, $d/\Lambda = 0.95$, $d_1/\Lambda = 0.60$ and $d_m = 0.67 \mu\text{m}$. Here, d_n is chosen as 0.54, 0.57 and 0.60 μm , the resulting dispersion is shown by solid line for fast axis and dotted line for slow axis modes. For fast axis the calculated average dispersions are -500.66, -531.30 and -562.18 ps/(nm.km) respectively in the wavelength bands from 1360 to 1565 nm. In fast axis mode, the calculated dispersion variations are 19.20, 11.50 and 18.64 ps/(nm.km) respectively. For slow axis the calculated average dispersions are -407.46, -454.40 and -501.88 ps/(nm.km) respectively in the wavelength bands from 1420 to 1625 nm. The ΔD s for slow axis are 9.44, 9.40 and 11.30 ps/(nm.km) respectively. It is observed that with $d_n = 0.54 \mu\text{m}$ the average dispersion of both fundamental mode modes are low and ΔD increases for fast axis mode. When $d_n = 0.60 \mu\text{m}$, resulting average dispersion of both fast and slow axis modes is higher but ΔD is also high. When $d_n = 0.57 \mu\text{m}$, the average dispersion is moderate and the value of ΔD is lowest. Thus, $d_n = 0.57 \mu\text{m}$ is selected as the optimum value while others parameters are kept constant.

Based on the discussions, the optimum design parameters of the proposed HPCF are found as $\Lambda = 0.80 \mu\text{m}$, $d/\Lambda = 0.95$, $d_1/\Lambda = 0.60$, $d_m = 0.67 \mu\text{m}$ and $d_n = 0.57 \mu\text{m}$. It has been already demonstrated that the proposed fiber supports ultra-flattened

dispersion in fast and slow axis modes and both modes cover the most important C band (1530-1565 nm). Although in [2] average dispersion of -369.10 ps/(nm.km) has been reported for fundamental slow axis mode covering 1580 to 2110 nm but it missed out the most important communication band (C band).

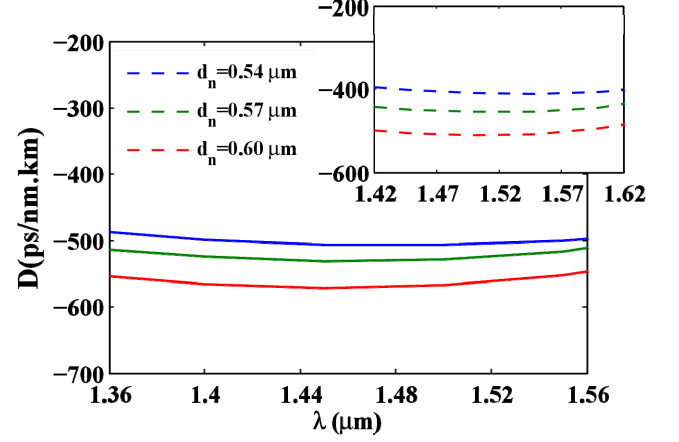


Fig. 7. Effect of changing d_n on the dispersion characteristics for fast axis (indicated by solid line) and slow axis (indicated by dotted line) modes.

In order to confirm the accuracy of the design it is necessary to investigate the tolerance of the fabrication. It is already known that $\pm 1\%$ variations of the fiber's global parameter may happen during the standard fabrication process [12]. Therefore, $\pm 2\%$ variations of the global parameter may require ensuring the average dispersion and dispersion flatness. Fig. 8 shows the effect of $\pm 1\%$ and $\pm 2\%$ variations of all global parameters (d , d_1 , d_m and d_n) on the dispersion magnitude and flatness around their optimum values. Here, the increment and decrement of all global parameters have been indicated by solid and dotted line, respectively. Fig. 8 (a) represents the variations in the fast axis mode at where it has seen that $\pm 1\%$ and $\pm 2\%$ variations of (d , d_1 , d_m and d_n) induce $\pm 2.84\%$ and $\pm 6.18\%$ changes from the optimum dispersion. On the contrary, in slow axis mode $\pm 1\%$ and $\pm 2\%$ variations of (d , d_1 , d_m and d_n) cause $\pm 3.03\%$ and $\pm 5.32\%$ changes from the optimum dispersion which is shown in Fig. 8 (b).

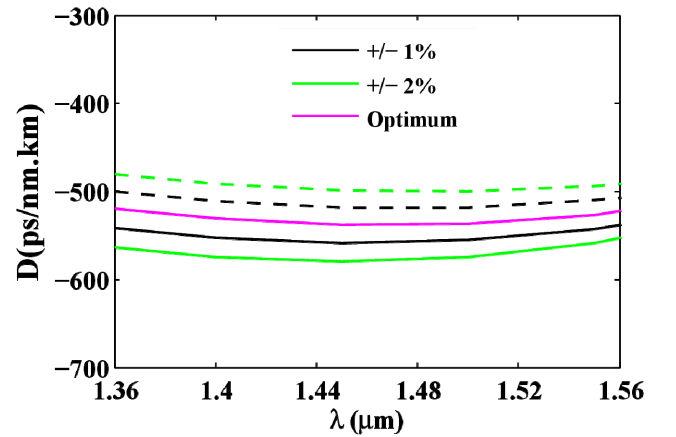


Fig. 8 (a). Effect of $\pm 1\%$ and $\pm 2\%$ variations of global parameters (d , d_1 , d_m and d_n) around the optimum values for fast axis mode.

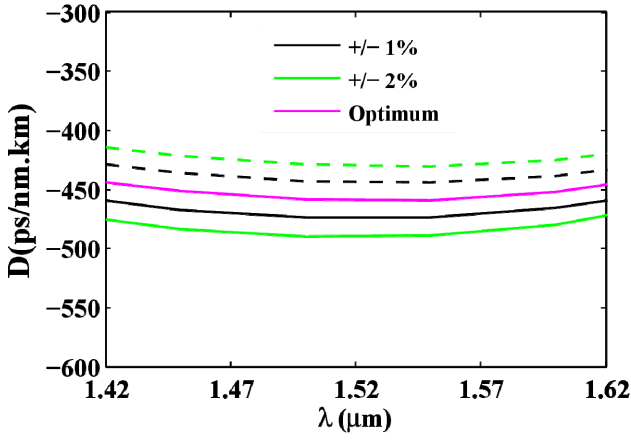


Fig. 8 (b). Effect of $\pm 1\%$ and $\pm 2\%$ variations of global parameters (d , d_1/d_m and d_n) around the optimum values for slow axis mode.

The proposed HPCF shows high birefringence due to the presence of elliptical air hole in the core region. The elliptical air hole is responsible for creating asymmetry in the core therefore high level of birefringence is induced. Modal birefringence as a function wavelength for optimal design parameters of $\Lambda=0.80 \mu\text{m}$, $d/\Lambda=0.95$, $d_1/\Lambda=0.60$, $d_m=0.67 \mu\text{m}$ and $d_n=0.57 \mu\text{m}$ is shown in Fig. 9. As illustrated from the figure, birefringence monotonically decreases with increasing of wavelength. The calculated birefringence at $1.55 \mu\text{m}$ has found about 0.01168. The conventional polarization maintaining (PM) fiber shows a modal birefringence of 5×10^{-4} thus the proposed fiber exhibits more than 21 times higher birefringence values [13]. Due to high birefringence, the proposed fiber is suitable for eliminating the effect of PMD in high bit-rate transmission systems and numerous sensing applications [5].

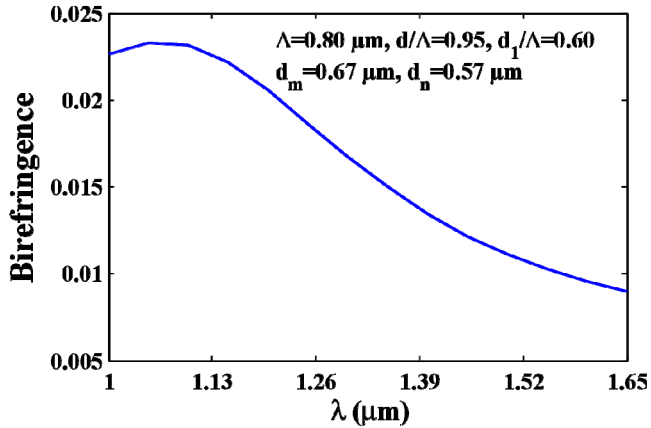


Fig. 9. Birefringence of the proposed fiber for changing wavelength with optimum design parameters of $\Lambda=0.80 \mu\text{m}$, $d/\Lambda=0.95$, $d_1/\Lambda=0.60$, $d_m=0.67 \mu\text{m}$ and $d_n=0.57 \mu\text{m}$.

Confinement loss of the HPCF is another key parameter for light guidance since it limits the total transmission distance in real fiber communication systems. A true RDCF should have very small confinement loss over a wide range of wavelengths [12]. Confinement loss of the proposed HPCF as a function of wavelength for optimal design parameters of $\Lambda=0.80 \mu\text{m}$, $d/\Lambda=0.95$, $d_1/\Lambda=0.60$, $d_m=0.67 \mu\text{m}$ and $d_n=0.57 \mu\text{m}$ is shown in Fig. 10. This figure represents confinement loss for fast axis

and slow axis modes with $N_r=7$, where N_r is the number of air-hole rings surrounding the core. It can be observed that confinement loss increases with increasing of wavelength. This is because at longer wavelength, fundamental mode field loosely bounds between elliptical air hole and first air hole ring hence mode field spreads outside the core region. As a result, confinement loss increases. The confinement loss has found at $1.55 \mu\text{m}$ about 0.2823 and 0.1 dB/km for fast axis (shown in solid line) and slow axis (shown in dotted line) modes, respectively. The confinement loss can be further reduced by inserting more air-hole rings in the cladding region, where the additional air-hole ring has insignificant effect on both average dispersion and birefringence [9].

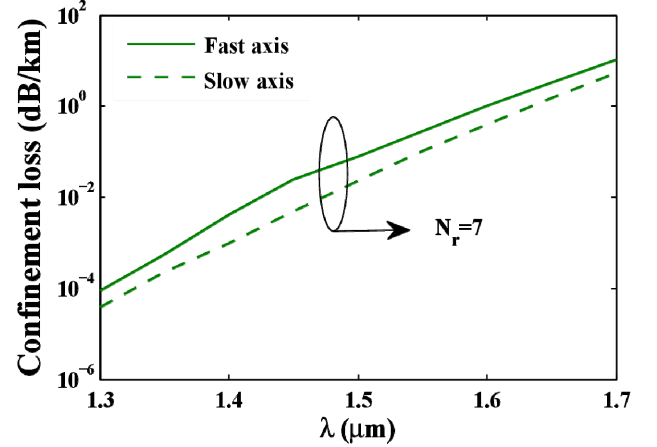


Fig. 10 Confinement loss as a function of wavelength for optimum design parameters of $\Lambda=0.80 \mu\text{m}$, $d/\Lambda=0.95$, $d_1/\Lambda=0.60$, $d_m=0.67 \mu\text{m}$ and $d_n=0.57 \mu\text{m}$.

Effective area (A_{eff}) and splice loss (L_s) for fast axis mode of the proposed HPCF for varying wavelength is shown in Fig. 11. It is evident that A_{eff} increases with increasing of operating wavelength. At longer wavelength, electromagnetic field spreads towards the cladding region. As a consequent, A_{eff} increases monotonically with wavelength. With optimal design parameters, A_{eff} of the proposed HPCF has found about 3.34 and $3.21 \mu\text{m}^2$ at $1.55 \mu\text{m}$ for x- and y-polarization, respectively. Due to the high Kerr non-linear coefficient of Schott BK7 the induced nonlinear coefficient might be higher that may create significant penalties on input and output coupling of light. A viable solution of this problem has been demonstrated in ref. [14].

For residual dispersion compensation, the dispersion slope of the RDCF should be as small as possible [2]. Simulation result reveals that the proposed fiber shows very small dispersion slope in both fundamental modes. The calculated dispersion slope varies between -0.36 to 0.041 ps/(nm².km) and -0.11 to 0.027 ps/(nm².km) in fast and slow axis mode, respectively.

The simulation results of the proposed have been validated using different element sizes of the mesh. In the simulation, we have adopted normal element size, fine element size, finer element size and extremely finer element size. It has found that changing different element sizes of the mesh do not influence the dispersion characteristics. Therefore, it can be concluded that the obtained results are valid.

Management of L_S is an important issue when a RDCF needs to be attached with a SMF for the purpose of dispersion compensation. Splice loss creates due to mode field diameter (MFD) mismatches between RDCF and SMF. From Fig. 11, the splice loss at 1.55 μm has found about 8 dB. Most telecommunication-based SMF is generally made of silica whereas the background material of the proposed RDCF is Schott BK7. As silica and soft glass have different melting temperatures, splicing between RDCF and SMF may exhibit higher coupling loss. The coupling loss can be reduced by using special fusion splicing techniques demonstrated in [15-16].

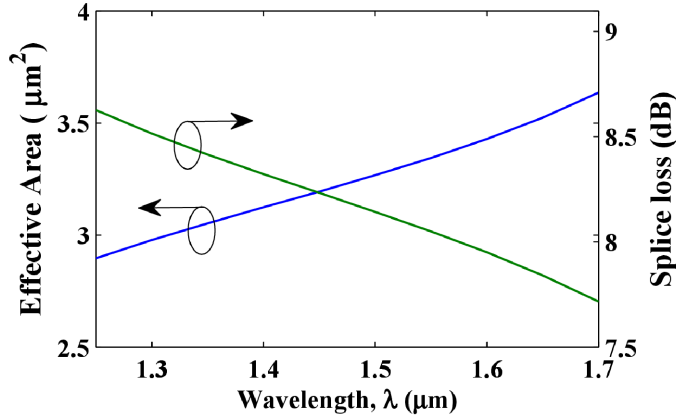


Fig. 11. Effective area and splice loss for fast axis mode of the proposed HPCF as a function of wavelength for optimum design parameters of $\Lambda=0.80 \mu\text{m}$, $d/\Lambda=0.95$, $d_1/\Lambda=0.60$, $d_m=0.67 \mu\text{m}$ and $d_n=0.57 \mu\text{m}$.

Finally, we consider the possible fabrication feasibility of the proposed HPCF. Due to the presence of elliptical air hole, fabrication of this PCF may not be so easy. However, production of the proposed HPCF is not beyond possibility from the available fabrication methods. Fabrication of soft glass PCF using the well-known extrusion technique has been already reported in [17]. The die cast process demonstrated in [18] can also be used which allows fabrication of the most commercial glasses. This method provides additional flexibility for fabrication of elliptical air holes.

IV. CONCLUSION

In summary, we have presented numerical modeling of a defected core hexagonal photonic crystal fiber and its guiding properties are investigated for variation of different geometric parameters. We have demonstrated that using a central elliptical air hole and adjusting its dimensions very high negative and flattened dispersion about has been obtained in both polarization modes over a wide range of wavelengths. Moreover, by taking the advantage of creating artificial asymmetry in the core high level of birefringence has been achieved at 1.55 μm . To ensure the possibility of practical fabrication, tolerance of the fiber parameters has been studied. Due to excellent guiding characteristics, this PCF can be considered as a good residual dispersion compensator.

REFERENCES

- [1] M. R. Hasan, M. S. Anower, and M. I. Hasan, "A polarization maintaining single mode photonic crystal fiber for residual dispersion compensation," *IEEE Photon. Technol. Lett.*, vol. 28, no. 16, pp. 1782-1785, 2016.
- [2] M. I. Hasan, S. M. A. Razzak, and M. S. Habib, "Design and characterization of highly birefringent residual dispersion compensating photonic crystal fiber," *J. Lightw. Technol.*, vol. 32, no. 23, pp. 4578-4584, Dec. 1, 2014.
- [3] M. I. Hasan, M. S. Habib, and S. Habib, and S. M. A. Razzak, "Highly nonlinear polarization maintaining dispersion compensating fiber for high-speed transmission system," *Opt. Eng.*, vol. 52, no. 11, pp. 116112, Nov. 2013.
- [4] X. Jiang, T. G. Euser, A. Abdolvand, F. Babic, F. Tani, N. Y. Joly, J. C. Travers, and P. St.J. Russell, "Single-mode hollow-core photonic crystal fiber made from soft glass," *Opt. Exp.*, vol. 19, no. 16, pp. 15438-15444, Aug. 1, 2011.
- [5] M. I. Hasan, M. S. Habib, and S. M. A. Razzak, "An elliptical-shaped core residual dispersion compensating octagonal photonic crystal fiber," *IEEE Photon. Technol. Lett.*, vol. 26, no. 20, pp. 2047-2050, Oct. 15, 2014.
- [6] M. A. R. Franco, V. A. Serrao, and F. Sircilli, "Microstructured optical fiber for residual dispersion compensation over S+C+L+U wavelengthbands," *IEEE Photon. Technol. Lett.*, vol. 20, no. 9, pp. 751-753, May 1, 2008.
- [7] D. C. Tee, M. H. A. Bakar, N. Tamchek, and F. R. M. Adikan, "Photonic crystal fiber in photonic crystal fiber for residual dispersion compensation over E + S + C + L + U wavelength bands," *IEEE Photon. J.*, vol. 5, no. 3, art. ID 7200607, Jun. 2013.
- [8] M. A. Islam and M. S. Alam, "Design optimization of equiangular spiral photonic crystal fiber for large negative flat dispersion and high birefringence," *IEEE J. Lightw. Technol.*, vol. 30, no. 22, pp. 3545-3551, Nov. 2012.
- [9] M. S. Habib, R. Ahmad, M. S. Habib, and M. I. Hasan, "Residual dispersion compensation over the S+C+ L +U wavelength bands using highly birefringent octagonal photonic crystal fiber," *Appl. Opt.*, vol. 53, no. 14, pp. 3057-3062, 10 May. 2014.
- [10] M. S. Habib and E. Khandker, "Highly birefringent photonic crystal fiber with ultra-flattened negative dispersion over S+C+L+U bands," *Appl. Opt.*, vol. 54, no. 10, pp. 2786-2789, Apr. 2015.
- [11] M. R. Hasan, M. I. Hasan, and M. S. Anower, "Tellurite glass defectcore spiral photonic crystal fiber with low loss and large negative flattened dispersion over S + C + L + U wavelength bands," *Appl. Opt.*, vol. 54, no. 36, pp. 9456-9461, 2015.
- [12] M. R. Hasan, M. S. Anower, and M. I. Hasan, "Polarization maintaining highly nonlinear photonic crystal fiber with closely lying two zero dispersion wavelengths," *Opt. Eng.*, vol. 55, no. 5, p. 056107, May 2016.
- [13] M. S. Habib, S. Habib, and S.M.A. Razzak, "Proposal for highly birefringent broadband dispersion compensating octagonal photonic crystal fiber," *Opt. Fiber Technol.*, vol. 19, no. 5, pp. 461-467, Oct 2013.
- [14] T. A. Birks, G. Kakarantzas, and P. S. J. Russell, "All-fibre devices based on tapered fibres," in *Proc. Opt. Fiber Commun. Conf.*, Feb. 2004.
- [15] M. R. Lange, E. Bryant, M. J. Myers, and J. D. Myers, "High gain coefficient phosphate glass fiber amplifier," *NFOEC*, paper no. 126, 2003.
- [16] D. Paul, R. Biswas, and N. S. Bhattacharyya, "Splicing hetero-core fibers in perspective of different material compositions," *Adv. Opt. Sci. Eng.*, vol. 166, pp. 209-214, June 3, 2015.
- [17] V. Kumar, A. K. George, W. H. Reeves, J. C. Knight, P. S. Russell, F. G. Omenetto, and A. J. Taylor, "Extruded soft glass photonic crystal fiber for ultrabroad supercontinuum generation," *Opt. Exp.*, vol. 10, no. 25, pp. 1520-1525, 2002.
- [18] Z. Guiyao, H. Zhiyun, L. Shuguang, and H. Lantian, "Fabrication of glass photonic crystal fibers with a die-cast process," *Appl. Opt.*, vol. 45, no. 18, pp. 4433-4436, Jun. 2006.

Effect of variation of different generator parameters on the stability of a small power plant: An Analysis Through ETAP Software

Md. Shazzadul Islam
Email-shazzadulislam89@gmail.com
United International University
Department of EEE

Md. Sahidul Islam
Email-sahid.eee1@gmail.com
United International University
Department of EEE

Muhammed Fayyaz Khan
fyk@eee.uiu.ac.bd
United International University
Department of EEE

Abstract—Transient analysis plays an important role in power system. As far as the stability of a plant is concerned, transient stability is very critical for small power plants. Transient analysis is aimed at determining whether a system will remain in synchronism with other generating plants after disturbances. In this paper, transient stability for a small power plant (80MW) has been studied through ETAP software. The plant is operating as IPP (Independent Power Producer) and owned by United Power Generation & Distribution Co. Ltd.(UPGD) under United Group, Bangladesh. There are in total 12 generators running in parallel. Sometimes, the generators run on Islanding mode and sometimes on grid mode. For both modes of operation, stability study was made to predict the behaviours of the generators under different abnormal situations. The impact of fault clearing time and inertia of the generator on transient stability of the power plant due to variation of different parameters of the generator is critically studied. Our main aim is to vary the different parameters of the generator and visualize impact on transient stability. In this paper, transient stability analysis is carried out by simulating symmetrical and unsymmetrical faults in the system under ETAP software environment.

Keywords— ETAP, Disturbances, Equilibrium State, System Faults, System Stability, Transient Stability Analysis.

I. INTRODUCTION

The ability of the engineer in providing reliable and uninterrupted supply to loads results in the successful operation of a power system. Power system stability is a condition where the synchronous machines of the system remain in synchronism. Generally, stability problems are divided into two major categories [1][5] as Steady-state stability and Transient stability. Steady-state stability refers to the ability of the power system to regain synchronism after small and slow disturbances whereas, transient stability is the ability of power system to survive the transition following a large disturbance and reach an acceptable operating condition. The vital aim of transient stability is to determine whether the system returns to a steady state value after the clearance of disturbance[2]. The recovery of a power system subjected to a severe large disturbance is of interest to system planners and operators. In grid mode, the power generation station or power plant

is in parallel with the grid. As the total plant is connected to grid, the frequency & voltage is uncontrollable and depends on the grid voltage & frequency. Our main concern is to observe impact of on grid mode operation under faulty condition. On Island mode, there is no grid connection and only local loads are connected to the plant. This is a very critical situation for generators, in case of fault generators automatically disconnect from the system. In case of unsymmetrical fault the impact is low it can be easily recovered. In this paper, we analysed the effect of change of inertia of the generator and clearing time of fault on the stability of the plant.

II. LAYOUT OF THE POWER PLANT

For the stability analysis, as already mentioned a small plant of 80 MW has been considered with twelve generators running in parallel. The plant is named as DEPZ plant located at EPZ area SAVAR. The plant is owned by United Group and the plant operates under BOO (Build-Operate & Own) mode. In the plant has 12 generators connected to two separate buses known as WARTSILA and Rolls Royce bus respectively. Four generators are connected to WARTSILA bus each is rated as 8.73MW and five generators connected to Rolls Royce bus each rated as 9.34MW. There are three small generators known as MTU connected to WARTSILA bus.

TABLE -I
Specifications for Wartsila Generators:

Output	8.73 MW
voltage	11/11000V
current	573A
speed	750 rpm
Over speed	900 rpm
Weight	27000 kg
Inertia	2150 kgm^2
p.f	.80
Number of gen	4

TABLE-II
Specifications for Rolls Royce:

islanding mode. Major parameters that were changed for the study were change of inertia and fault clearing time for simulation of faults at different points of the plant.

Grid mode operation:

- Fault on 11kv WARTSILA bus
- Fault on 11kv Rolls-Royce bus
- Fault on 33kv bus

Island mode operation

- Fault on 33kv bus

In island mode operation effect of fault is same if fault occur anywhere of the plant. That's why we consider only 33kv bus.

Parameter variation

- Inertia constant (MW-S/MVA)
- Clearing time (Sec)

The above mentioned faults were sequentially simulated and the curves depicting fault currents and stability of the system are shown below. During simulation under grid mode operation, the two interconnectors namely ic1 and ic2. Were kept open to connect both WARTSILA and Rolls Royce buses with their respective generators. We have also seen the stability of power plant by varying the value of machine inertia and fault clearing time.

VI. RESULTS AND DISCUSSION

Case1:

At first, we created fault on 33kv bus (grid mode) then analyse the stability of the system. Fault occurs at .5s and clearing time of fault is .6s. From fig 2 and fig 3, variation of system voltage and frequency can be observed. Voltage drops down to zero on faulty bus and when fault is cleared then voltage regains. Effect of voltage on another bus is little but still it goes down and get stable after fault is cleared as shown in fig3.

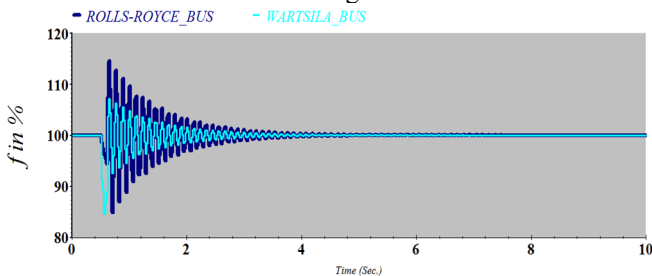


Fig 2: Frequency vs time

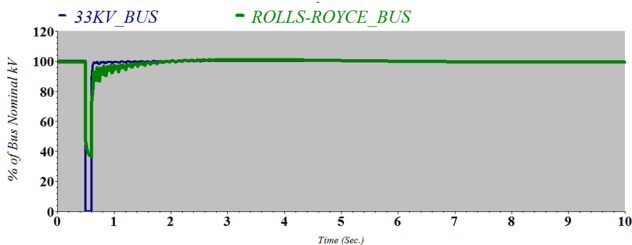


Fig 3: Bus voltage of 33kv and 11kv Rolls-Royce Bus

Case 2: Effect of fault clearing time

In this section, effect on fault clearing time on system stability has been highlighted. Fault (33kv bus) occurs at 0.5s and cleared at 0.51s. It is seen from fig.4 & 5 respectively that each generator goes into the stable condition. Long fault clearing time has been intentionally chosen to see the effect of fault on the generators and the stability of the system. It has been found that effect is very less in 11kvRolls-Royce bus because this generators are far away from the fault point and have also higher inertia.

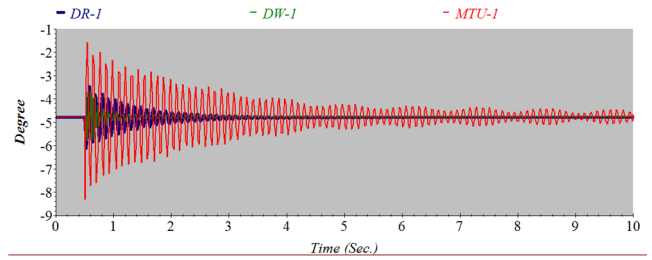


Fig 4 : Absolute power angle

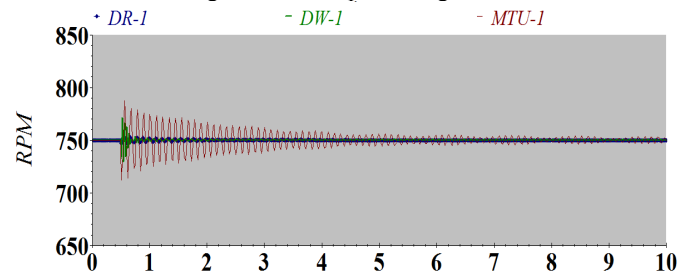


Fig 5: Speed of Rolls-Royce, Mtu, Wartsila

But when fault is cleared at .6s the WARTSILA and MTU generator lose synchronism very early and will automatically be disconnected from the system as shown in fig 6.

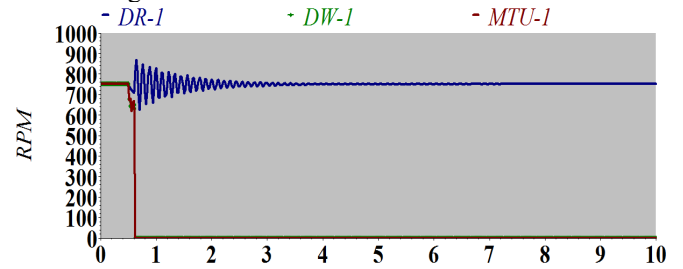


Fig 6: Speed of Rolls-Royce, Mtu, Wartsila

If clearing time is .7s, every generator loses synchronism and disconnects from the system. This is shown in fig 7.

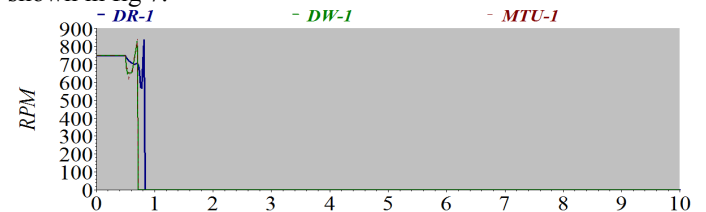


Fig 7: Speed vs time

Case 3:

By changing the machine inertia we can observe the effect on stability of the system. In case 2 all generators disconnected from the system due to low inertia. By increasing the inertia of the generators this problem can be minimize. Now If the clearing time of fault is increases (.7s) the generators not disconnected from the system. This is shown in fig 8

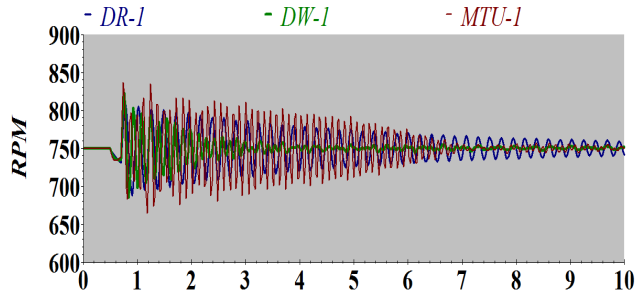


Fig 8: Speed

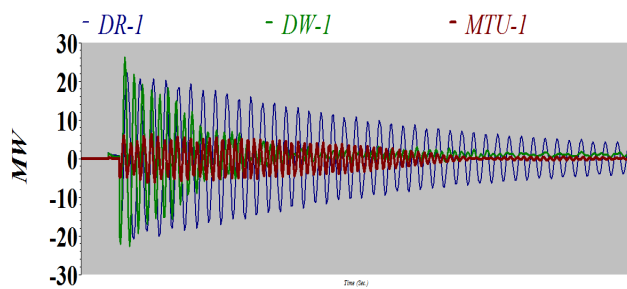


Fig 9: Electrical power output

Case 4:

Fault on 11kv WARTSILA bus. This time WARTSILA and MTU disconnect from the system very early because of very low inertia but the effect is very small in Rolls Royce generator. Effect is less because ic1 and ic2 are open. This is shown in fig 10

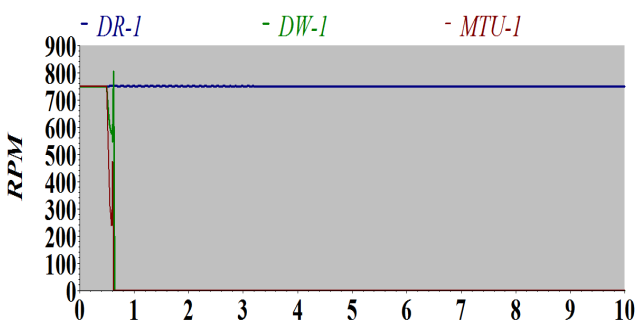


Fig 11: Speed of Wartsila, Mtu, and Rolls Royce

By increasing inertia of the MTU and WARTSILA generators we can prevent this problem. This is shown table:III and in fig 12.

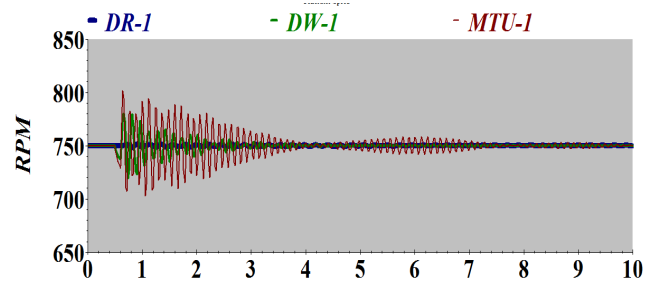


Fig 12: Speed of Wartsila, Mtu & Rolls Royce

Table : III

Table for Inertia

GEN	ORIGINAL	NEW
WARTSILA	.027	.4
ROLLS-ROYCE	.127	.5
MTU	.012	.27

Case 5:

Fault on 11kv Rolls Royce bus. No generator disconnect from the system, but effect is high in Rolls-Royce generator. Shown in fig 13.

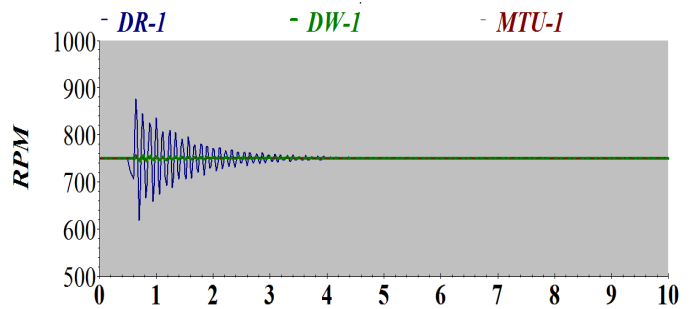


Fig 13: Speed of Wartsila, Mtu & Rolls-royce

Case 7:

Effect of change of inertia under Island mode operation Fault occurs in any bus of the system generators lose power. This happen though fault is clear very early. This shown fig 14

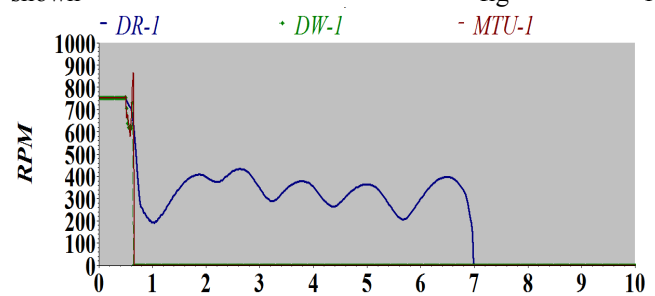


Fig 14: Speed of Wartsila, Mtu & Rolls-Royce

All generator disconnect from the system automatically because they have low. To recover this problem need high inertia shown in table-IV. shown in fig 15

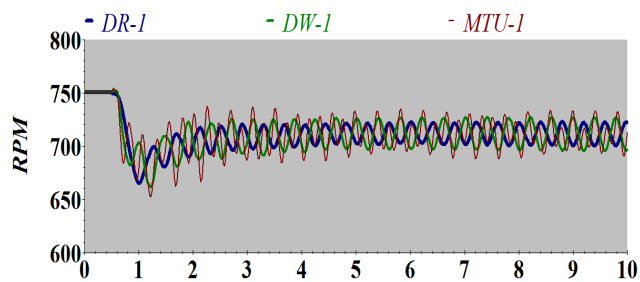


Fig 15: Speed of Wartsila, Mtu & Rolls-Royce
Table:IV

Gnerator Inertia

GEN	ORIGINAL	NEW
WARTSILA	.027	3
ROLLS-ROYCE	.127	5
MTU	.012	2

In islanding mode though inertia is very high but generators still take a lot of time to regain stable condition. That's why faults in islanding mode operation is very critical.

Case 7:

Unsymmetrical fault

In case of LG fault generator are not easily disconnect from the system but they lose synchronism for some times. If the clearing time of fault increases system regain its stability. The effect is shown in fig16.

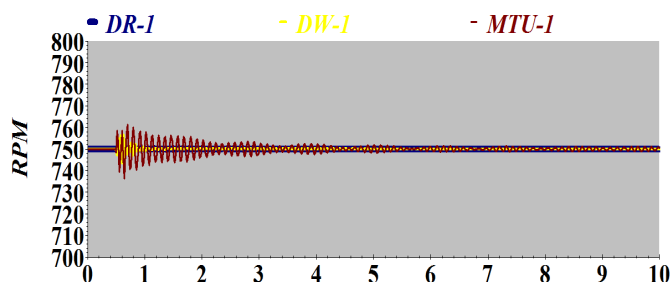


Fig 16: Speed of Wartsila, Mtu & Rolls-Royce

Case 8:

Difference of generator terminal current during symmetrical and unsymmetrical fault. In symmetrical fault current is higher than unsymmetrical fault. That's why effect of symmetrical fault is very high. This is shown in fig17 & fig18..



Fig 17: Terminal current for symmetrical fault

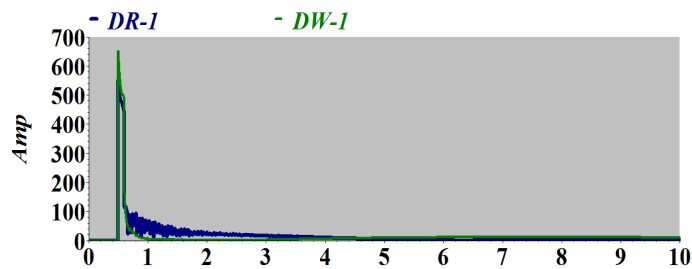


Fig 18: Terminal current for unsymmetrical fault

Case9:

At last case we analysed the effect fault on other parameters of the generators. Exciter voltage, MW, MVAR all are changes. Shown in fig 19, 20& 21.

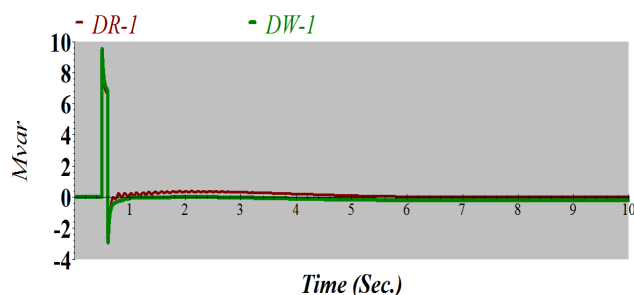


Fig 19: Reactive power

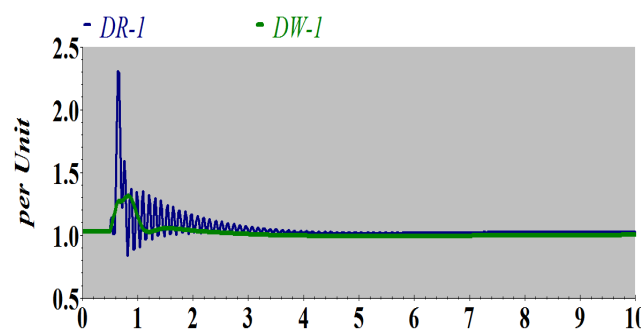


Fig 20: exciter voltage

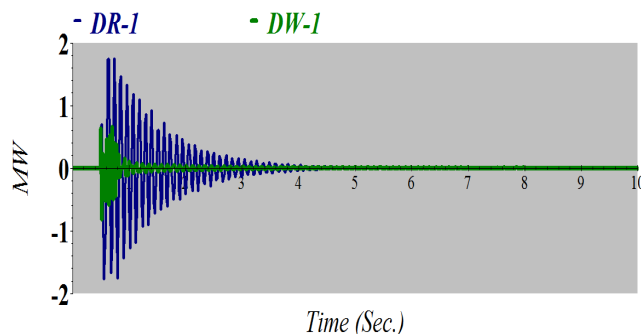


Fig 21: Electrical power output

VII. CONCLUSION

In this paper, the analysis of the system is a gas engine power plant (UPGD). Transient stability analysis have been simulated on the test system. The rate of change of system frequency, power angle and the voltage levels of the buses are considered for analysis. By changing inertia and clearing time of fault we observed the transient stability of this plant. For a stable power plant need to chose this parameters very carefully. In grid mode analysis we can recover stability by changing fault clearing time but in island mode operation clearing time of fault is not enough need to consider high inertia to recover stable condition. The dynamic performance of the proposed system was analysed using ETAP software.

REFERENCES

- [1]. Kundur. P, Paserba. J, Ajarapu. V, Andersson. G, Bose. A, Canizares. C, Taylor. C, Cusum. T and Vittal. V, "Definition and classification of power system stability," IEEE Transactions on Power Systems, vol.19, No. 2, pp. 1387–1401, May 2004
- [2]. Swaroop Kumar. Nallagalva, Mukesh Kumar Kirar, Dr.Ganga Agnihotri, "Transient Stability Analysis Of the IEEE 9-Bus Electric Power System," International Journal of Scientific Engineering and Technology ,Volume No.1, Issue No.3,pg : 161-166, July 2012
- [3]. Hadisadat, " Power system analysis ," McGraw-Hill International Editions, 1999.John J Grainer and WillamD Stevenson, "Power system analysis," McGraw-Hill International Editions, 1999.
- [4]. John J Grainer and WillamD Stevenson, "Power system analysis," McGraw-Hill International Editions, 1999.
- [5]. I. J. Nagrath and D. P. Kothari, "Modern Power system analysis," Tata McGraw-Hill, 2003.
- [6]. Pushp raj, "Load Flow and Short Circuit Analysis of 400/220 Kv Substation," International Journal of creative Research Thoughts, vol. 1 ,Issue.5 April 2013.
- [7]. P. M. Anderson and A. A. Fouad, "Power System. Control and Stability". New York: IEEE Press, 1994
- [8]. N. Perumal, A. C. Amran, " Automatic load Shedding in power system", Power Engineering Conference, 2003. . PECon 2003. Proceedings. National, vol., no., pp. 211- 216, 15-16 Dec.2003
- [9]. P.K.Iyambo, R. Tzonova, " Transient Stability Analysis of the IEEE 14- Bus Electrical Power System", IEEE Conf. 2007.
- [10]. Liang Wang, LiLi, ShanshanShi, Yiwei Zhang, Zongxiang Lu, Junliang ZhangG. Eason, B. Noble, and I. N. " Stability and Security Assessment for an Industrial Electric Grid with Enterprise-owned Power Plants", DRPT2008 6-9 April 2008 Nanjing China 1563.
- [11]. Lo KL, Tsai RJY. Power system transient Stability analysis by using modified Kohonen network. In: Proceeding of IEEE international conference on neural networks; 1995. p. 893–8.
- [12]. N.I. Abdul Wahab, A. Mohamed, A. Hussain, "Transient stability assessment of a power system using PNN and LS-SVM methods," Journal of Applied Sciences, vol. 7(21), pp.32083216, November 2007.

Design & Implementation of an Efficient Windmill Anemometer for Wind Speed measurement Using Microcontroller

Sree Bash Chandra Debnath

Department of Engineering and Computer Science
University of Trento
Trento, Italy
sreebashcuet@gmail.com

Rezaul Karim

Department of Electrical & Electronic Engineering
Premier University
Chittagong, Bangladesh
rezauleepuc@gmail.com

Biki Barua

School of Engineering (Telecommunication)
Université du Québec en Abitibi-Témiscamingue
Quebec, Canada
biki.barua@uqat.ca

Muslim Uddin Faisal

Department of Electrical & Electronic Engineering
Premier University
Chittagong, Bangladesh
muslimuddinfaishal@yahoo.com

Abstract—Direction & speed, these are the important parameters to characterize the wind condition of the environment. In our paper, we tried to build a microcontroller based efficient anemometer. Though, in the market, there are several models is existed for anemometer, in our design we have used microcontroller PIC 18F2550 to control the circuit that's making the design more efficient & cost efficient along with the model is a windmill type. A wind sensor is used to sense the wind speed and it is fixed to the shaft with a segmented wheel. When the wind blows, the wind sensor sense the wind and the shaft start to rotate continuously relative with wind speed. In the practical environment, our design shows the improved & efficient result.

Keywords—anemometer; microcontroller; windmill; error.

I. INTRODUCTION

The demand for anemometers in different applications, like: industrial, weather forecasting is growing up with the advancement of the day. Either in land or sea, the weather station generally observes the atmosphere with measuring instruments & equipment's in order to provide the information for weather forecast & to study the weather & climate. In this contrary, wind measurements are taken as free of other obstructions as possible [1]. Leon Batista Alberti is the person who first gave the idea about anemometer around 1450 [2]. An anemometer is a meteorological device that is used to measure wind velocity, where different sensing principles such as: thermopiles, resistors, diodes are already used for low power consumption, high airflow rate range, sufficient flow rate sensitivity and compatibility [3].

Anemometers can be partitioned into two classes [4]: those that quantify the wind's pace, and those that gauge the wind's weight; but as there is a close connection between the pressure

and the speed, an anemometer designed for one will give information about both. Traditional anemometers are cup anemometer, hot wire anemometer, windmill anemometer, Laser-Doppler anemometer, and sonic anemometer [5], [6]. The main problems among all of these types are in accuracy & long term management [7]. Here we have used windmill anemometer & it is basically a rotation based anemometer, which has a fixed axis upon which propeller is mounted.

The accessible anemometers in the business sectors of Bangladesh are very costly and exactness and precision are additionally less. So, in our work we have tried to design a windmill anemometer with a new method of circuit configuration, that particularly suitable for the precision measurement and monitoring of time related quantities and cost effective. In this system, the time calculated by counting. A count is loaded into a register & the count is incremented or decremented at every clock cycle. We have used a nobility method that contains PIC microcontroller (PIC 18F2550), the liquid crystal display, the operational amplifier, the crystal oscillator and proximity sensor in order to increase the accuracy & management system compare to the conventional methods. In section II we will discuss about the system design & section III about result & analysis of our system.

II. METHODOLOGY

A. Theory

When wind forced the shaft of the anemometer, it will start to rotate & each rotation will generate the single pulse [8]. Moreover, each pulse is counted as single frequency. In order to count this rotation, proximity sensor is used. So, from mathematical term we can write:

$$\omega = 2\pi N/60$$

$$2nf = 2nN/60$$

$$f = N/60$$

B. Hardware Part

1) *IR LED*: An infrared light-emitting diode (LED) also known as IR transmitter is a type of electronic device that emits infrared light not visible to the naked eye. An infrared LED operates like a regular LED, but may use different materials to produce infrared light like all LEDs, a type of diode, or simple semiconductor. We used this device as sensor.

2) *Operational Amplifier*: Operational amplifiers can be used to perform mathematical operations on voltage signals such as inversion, addition, subtraction, integration, differentiation, and multiplication by a constant. Here in our implementation, we use op-amp (LM324), which works as comparator. The LM324 series are low-cost, quad operational amplifiers with true differential inputs. They have several distinct advantages over standard operational amplifier types in single supply applications.

3) *Voltage Regulator*: A voltage regulator is designed to automatically maintain a constant voltage level. Electronic voltage regulators are found in devices such as computer power supplies where they stabilize the DC voltages used by the processor and other elements. We have used this device in our device in order to sustain the voltage level.

4) *Crystal Oscillator*: A crystal oscillator is an electronic oscillator circuit that uses the mechanical resonance of a vibrating crystal of piezoelectric material to create an electrical signal with a very precise frequency. This frequency is commonly used to keep track of time (as in quartz wristwatches), to provide a stable clock signal for digital integrated circuits, and to stabilize frequencies for radio transmitters and receivers. The most common type of piezoelectric resonator used is the quartz crystal, so oscillator circuits incorporating them became known as crystal oscillators, but other piezoelectric materials including polycrystalline ceramics are used in similar circuits. In our thesis we used 8MHZ crystal oscillator.

III. SYSTEM DESIGN

Outline system of our anticipated work is partitioned into two steps. These are simulative methodology and usage approach. In this section, both area has been covered.

A. Block Diagram

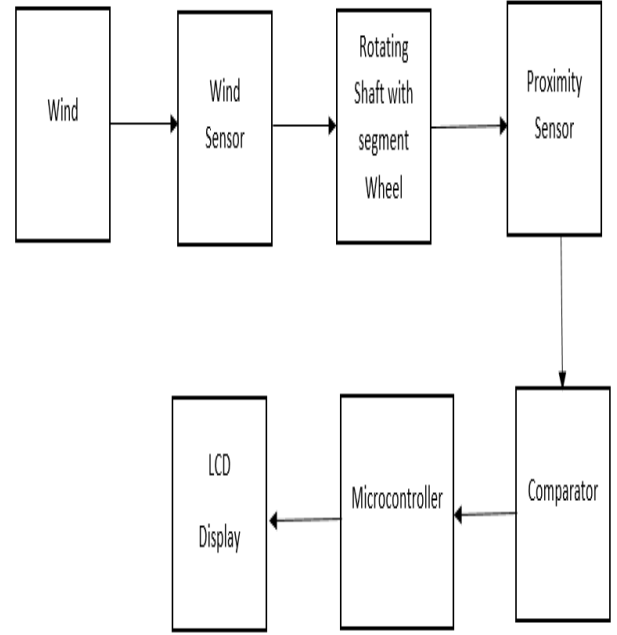


Fig. 1. Block diagram of proposed windmill anemometer.

The figure 1 shows the basic working procedure of our proposed anemometer. When the wind blows, the wind sensor sense the wind and it starts to rotate continuously relative with wind speed. To wind sensor, the shaft is connected through a bearing. A fix segmented wheel attached onto the wind sensor shaft. Thus the rotation of wind sensor the shaft also rotating with a segmented wheel. A proximity sensor has kept up beside of the rotating segmented wheel. It is a combination of both infrared and photodiode. A proximity sensor is a sensor able to detect the presence of nearby objects without any physical contact. A proximity sensor emits a beam of electromagnetic radiation (infrared) and absorbs this light by a photodiode. When the wind sensor shaft rotates, the proximity sensor generates pulses with the help of segmented wheel and it triggers the operational amplifier. The operational amplifier (comparator) amplifies the output signal of the proximity sensor. Then, the operational amplifier output or pulses fed to the microcontroller.

These pulses are counted by the timer/counter module of microcontroller for a minute. This counted pulse for a minute is proportional to the RPM of wind sensor and this RPM is relative to the wind speed. By this way, the wind speed has been calculated by the microcontroller. The computed estimation will be shown on the LCD. This procedure rehashes after a minute over and over. The LCD used is the 32 characters and 2 lines type. Finally, it has been displayed as the speed of the wind in KPH unit.

B. Circuit Diagram

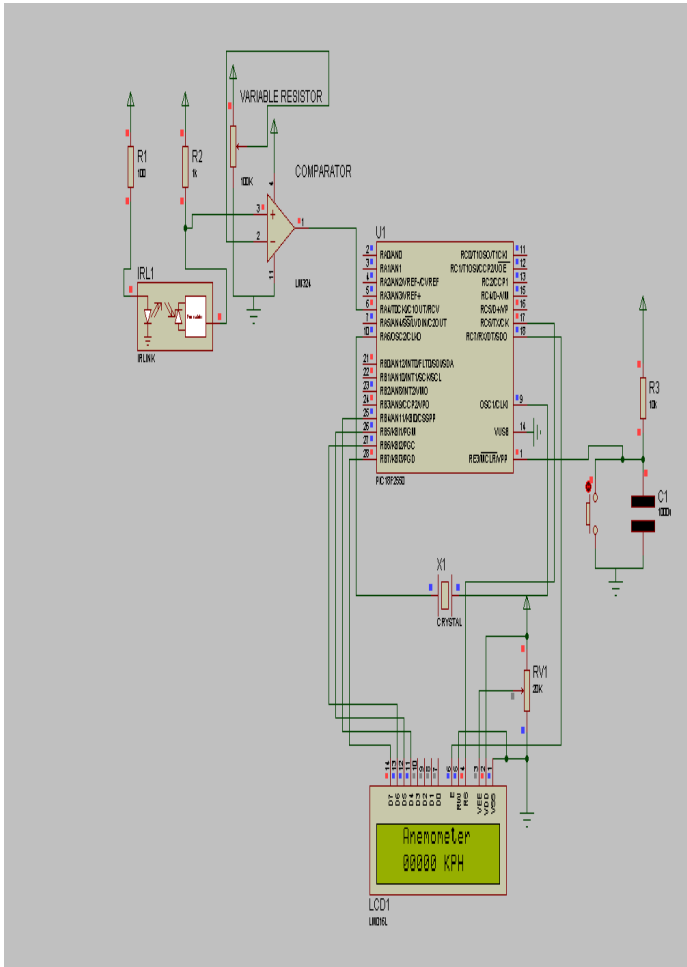


Fig. 2. Circuit Diagram.

The simulated diagram of circuit has shown in figure 2. And this simulation has performed by using proteus (VSM7.1 SP4) software.

C. Algorithm

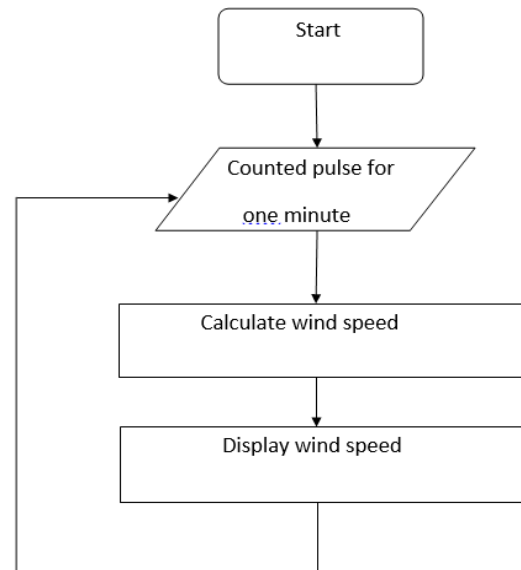


Fig. 3. Flow chart of programming

The algorithm of this system has shown in figure 3. Program start to count pulses for one minute, then counted pulses data are stores in memory. It forms this information and sends signs to LCD as it can show wind speed. This process has to be done repeatedly.

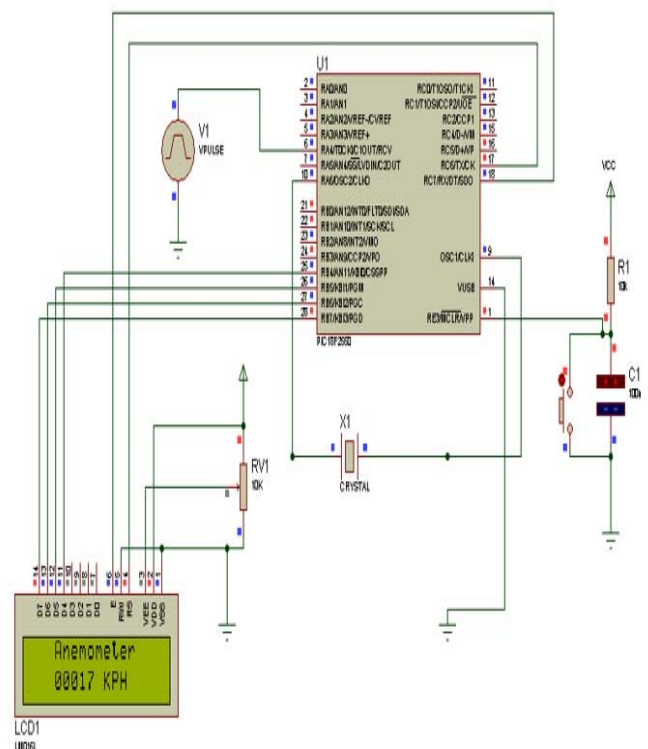


Fig. 4. Simulation output

D. Implementation

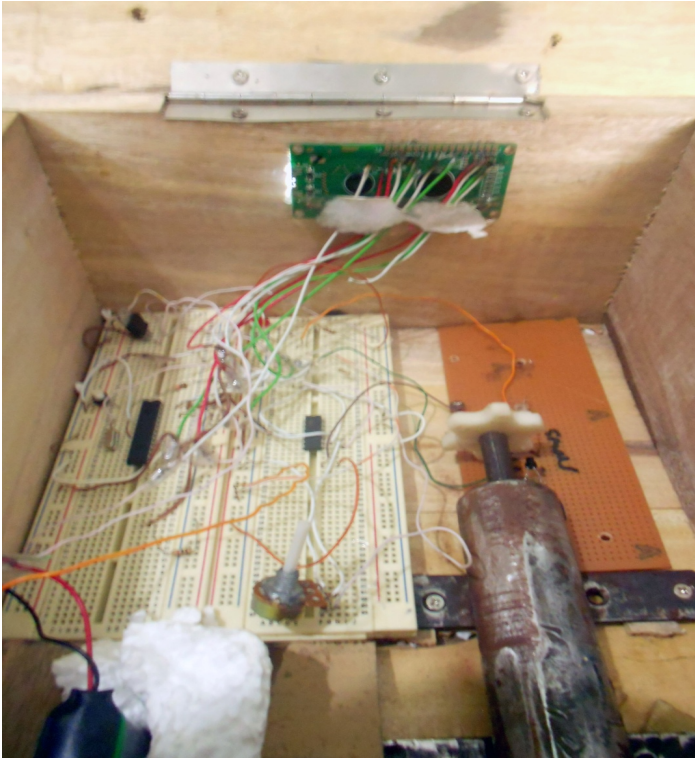


Fig. 5. Image of internal circuitry and shaft.

Figure 4 & 5 is showing the simulation circuit & implementation of our design.

IV. RESULT & ANALYSIS

The anemometer that has been designed in our laboratory is for speed of wind in Kilometer per hour unit, but meteorological department anemometer has been designed for speed of wind in Knots unit. So therefore wind speed output will be different from our anemometer output. In this context, the calculation of the conversion factor to convert in Kilometer per hour unit has been carried out.

TABLE I. PERFORMANCE ANALYSIS

Anemometer	Wind Speed (KPH)	Wind Speed (Knots)
Our System	8.51	4.6
Meteorological department's	9.25	5

A. Error Measurements

If we consider the meteorological departments anemometer is ideal anemometer, then the difference between two anemometers gives us the measurements of error.

Where, error in Knots= 0.4%

The measurement of wind speed and direction is highly sensitive to flow distortion by obstacles. So, some standard error correction factors have been maintained here accordingly.

TABLE II. INSTALLATION CORRECTION FACTOR

Height (m)	Correction Factor (%)
3-4	20
5-7	10
8-13	No Correction Needed
14-22	Subtract 10
23-42	Subtract 20
43-90	Subtract 30

B. Practical Data

Practice data has been measured in 2nd October 2015, at Patenga Sea Beach, Chittagong, Bangladesh.

TABLE III. MEASUREMENT DATA

Date	Time	Wind Speed (KPH)	Wind Speed (Knots)
02-10-2015	1:03 pm	7.00	1.94
02-10-2015	1:04 pm	7.50	2.08
02-10-2015	1:05 pm	6.75	1.87
02-10-2015	1:06 pm	8.50	2.36
02-10-2015	1:07 pm	10.25	2.80
02-10-2015	1:08 pm	7.50	2.08
02-10-2015	1:09 pm	7.00	1.94
02-10-2015	1:11 pm	9.50	2.63
02-10-2015	1:12 pm	9.50	2.63

TABLE IV. COMPARISON

Conventional Design	Our Design
1. These are mainly programming based. Example: LC handheld anemometer.	1. Only Microcontroller programming is needed.
2. High cost.	2. Low cost.

The average cost (market cost) for conventional anemometer is around 60\$ to 300\$ with different function any attachment. But on that contrary our design cost 30\$, that's more cheaper than market prices.

V. CONCLUSION

In this paper, the ultimate target that has been met is to build an efficient & cost effective windmill type anemometer. Our anticipated anemometer is easier to use than their ancient counterparts and also much more convenient and versatile. Additionally, the proposed design in this paper shows much more accuracy than the previous forms of the tool which displays the wind speed readings on an LCD screen and the rate of calculated error is low that makes the implemented device more applicable for the practical application with very low cost.

REFERENCES

- [1] G. K. K. Reddy, S. V. Reddy, and T. K. Ramkumar, "Development of Hall sensor propeller anemometer for measuring wind speed using embedded controller," in Communications, Devices and Intelligent Systems (CODIS), 2012 International Conference on, 2012, pp. 484-487.

- [2] Dines, William Henry. Anemometer. 1911 Encyclopedia Britannica.
- [3] J. Laconte, B. Rue, J.P Raskin, and D. Flandre, "Fully CMOS-SOI compatible low-power directional flow sensor", *Sensors*, 2004.Proceedings of IEEE, 2004, vol.2, p 864 – 867.
- [4] S.M. Ferdous, Muhammad Rokonzaman, Syedus Salehin, Arifa Ferdousi, Tanvin Hossaine "Design and Implementation of a Linear IC Based Low Cost Digital Tachometer for Laboratory and Industrial Application (MIEI0-104)" International Conference on Mechanical, Industrial and Energy Engineering 2010, 23-24 December, 2010, Khulna, Bangladesh.
- [5] M. Prokin, "Extreamly wide range speed measurement using a doublem butTered method," *IEEE Trans. Ind.Electron.*, vol.41, pp. 550-559, Oct. 1994.
- [6] R. J. Goldstein, *Fluid Mechanics Measurements*. Washington, DC: Hemisphere Pub. Corp., 1983.
- [7] G.-M. Ma, C.-R. Li, J. Jiang, J.-Y. Liang, Y.-T. Luo, and Y.-C. Cheng, "A passive optical fiber anemometer for wind speed measurement on high-voltage overhead transmission lines," *Instrumentation and Measurement, IEEE Transactions on*, vol. 61, pp. 539-544, 2012.
- [8] T. Ali, S. Nayeem, M. O. Faruk, M. Shidujaman, and S. Ferdous, "Design & implementation of a linear IC based low cost anemometer for wind speed measurement," in *Informatics, Electronics & Vision (ICIEV)*, 2012 International Conference on, 2012, pp. 99-102.

Small-signal Stability Evaluation of a Non-autonomous Microgrid System

A.H.M.A. Rahim

M.M. Atiqur Rahman

M. Ashraf Ali

Dept. of Electrical & Electronic Engineering
Ahsanullah University of Science & Technology
Dhaka, Bangladesh

Department of Electrical Engineering
K.F. University of Petroleum & Minerals
Dhahran, Saudi Arabia

Abstract: A microgrid comprised of wind, PV, and fuel cell generation systems along with a micro-alternator and a storage system is considered in this study. The small signal behavior of the system is investigated considering the real part of the dominant eigenvalues. The impact of the participation of the distributed generation sources on the small-signal stability has been investigated. While it is essential that participation of the alternator should be a minimum from economic and environmental considerations, it has been observed that its participation is essential for satisfactory operation of the grid in terms of maintaining frequency of the system as well as from general stability considerations. Too little contribution from the micro-alternator may drive the system unstable even under small disturbance conditions. Penetrations from wind and PV systems help stability, while the fuel cell tends to have destabilizing effect at heavier loads. Eigenvalue based small-signal stability of the system has been verified through response obtained by the dynamic model.

Keywords: Microgrid, renewable energy, eigenvalues, DG's, system stability.

I. INTRODUCTION

Widespread use of renewable energy sources and their integration at local distribution levels have contributed to the development of microgrids. Microgrid is a low voltage electrical network with distributed generation (DG) sources, energy storage elements and controllable loads [1]. An electrical microgrid can be viewed as a subsystem with its associated loads and generation units operating either autonomously or in connection with main utility grid (non-autonomous). It can have several DG units producing electrical power operating in synchronism with each other in order to supply local load or deliver power over the tie line. Microgrids are treated as small distribution systems that connect multiple customers to multiple distributed sources of generation and storage systems interfacing through power electronic converters [2].

A microgrid in general, has two main types of DG units, inertial or rotating type DG units and stationary or inverter type DG units which require power electronic converters for their connection to the main grid [3]. Under the category of inertial DG units, the most commonly used are micro

alternator, gas turbines, internal combustion engines etc. The photovoltaic systems, fuel cell systems and wind generation systems require power electronic devices for integration and hence they fall under the category of inverter type DGs. The small scale power production units employed in electrical microgrids are usually in the range of 4kW-10MW [4]. In this work we consider the most commonly used technologies for DGs i.e., the wind generation, fuel cell and photovoltaic systems.

In an electrical microgrid involving renewable sources as generation units, the power output is very uncertain. As the power participation from renewables is increased there arise several problems which disturb the system stability and give rise to critical issues [5]. During a short-circuit fault on a micro source, the mechanical input torque remains the same whereas the output electrical power and electromechanical torque are reduced due to diminishing terminal voltage. Owing to excess mechanical input, the turbine and generator accelerate and absorb more reactive power from the system, thus hindering the terminal voltage recovery. To enhance the dynamic performance under such phenomena various techniques have been proposed in literature focusing on voltage recovery and speed control. Several factors are responsible for abnormalities in the operation of microgrids under both the autonomous and grid connected mode. Disturbances, overloading of DGs, generation of switching harmonics, power quality issues, are some of the problems encountered in microgrids due to the presence of electronically coupled DG sources. Owing to these issues the safe operation and meeting the load demand under off grid mode is extremely challenging task requiring extensive control system to be incorporated. Unlike conventional power generation systems, there is no frequency response problem as such seen in microgrids as a result of load changes or generation change due to absence of rotational parts. The frequency response or droop characteristics can be obtained by controlling the interfacing inverter to emulate the synchronous alternator. In this manner, the power sharing mechanism can be established in microgrids with generations from electronically coupled sources [6]. Because of the prevalence of electronic components in a microgrid, the associated eigenvalues of the system are the sources of small-signal oscillations or instability. A number of studies have been reported over the

years on small signal stability both for autonomous and non-autonomous systems.

Eigen value based stability analysis for a small-signal microgrid model is reported in [7]. The authors have considered constant power loads and carried out sensitivity analysis in order to enhance the dynamic stability by regulating the voltage and current controller parameters. Small signal stability analysis of a microgrid through two-point estimate based method incorporating various uncertainties is reported in [8]. Power oscillation damping was investigated by Shagafi et al. in standalone microgrids using an integrated series compensator [9]. It was shown that by introduction of a series voltage, the dependence of power/frequency can be minimized. Robust PI Controller was employed to control the frequency of islanded microgrid using energy storage system [10]. To investigate microgrid dynamic stability, a small-signal model of a typical microgrid containing asynchronous generator based wind turbine, synchronous diesel generator, power electronic based energy storage and power network was used in [11]. Eigenvalue distributions of the microgrid system under certain steady operating status were identified to indicate the damping of the oscillatory terms and its effect on system stability margin. From the various studies it is apparent that the small-signal models are necessary to carry out the eigen value analysis in order to identify the threshold limit of generation mix from various DGs without hampering the stability. Small-signal state space model of micro generator, photovoltaic system, fuel cell system and wind generation system will constitute the small-signal model of the microgrid.

The dependence of small-signal stability of a microgrid on the extent of participation of the various DGs is investigated in this study. The relative stability is measured depending on the real part of the dominant eigenvalues of the system. Variation of the participation of the different components shows that too little participation of the micro-alternator can pose stability problems

II. THE MICROGRID SYSTEM

Fig. 1 shows a schematic diagram of the microgrid considered in this study. The microgrid has four DG units connected to the main grid along with integrated load at point of common coupling (PCC) bus. In order to make provision for sub-synchronous frequency ranges, the transient models of filters and connecting transmission links have also been incorporated. The first distributed generator in the diagram represents a small scale conventional generation system which may be a gas-turbine or micro generation system and it is considered to be equipped with an excitation control. A VSC along with associated power electronic filter and coupling line inductance connects photovoltaic system and fuel cell system to the main grid. A variable speed permanent magnet synchronous generator (PMSG) wind system driven by a horizontal axis wind turbine is assumed to be interfaced through appropriate converters with an additional filter circuit at the grid side. Energy storage device interfaced through a VSC functions as a central controller administering the system operation. The dynamic model of the system includes the transient models for each of the component DGs along with the interconnecting power electronic converters and the central controller.

The dynamic model of the system includes the transient models for each of the component DGs along with the interconnecting power electronic converters and the central

controller. A detailed layout of the microgrid architecture under study is shown in Fig. 2. The various power electronic converters along with interconnecting coupling lines and filter circuits are also displayed. The storage device is considered to be provided with a P-Q decoupled control strategy.

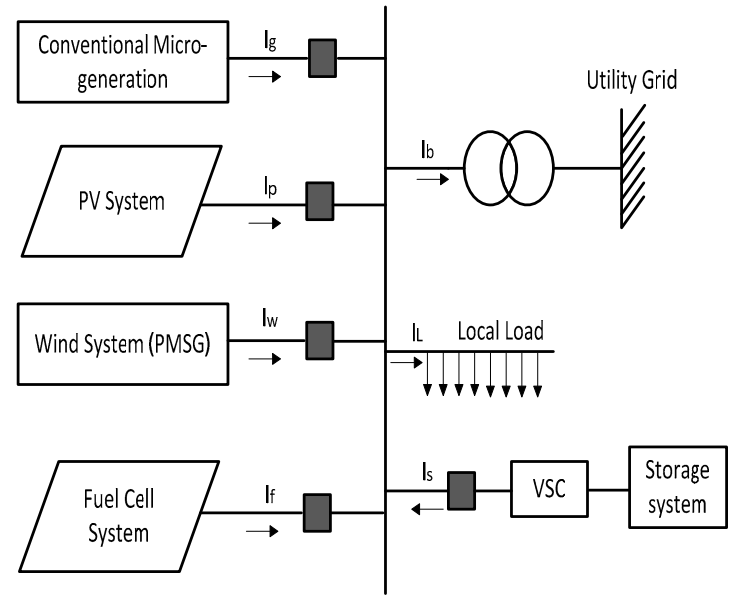


Fig. 1 Single line diagram of microgrid

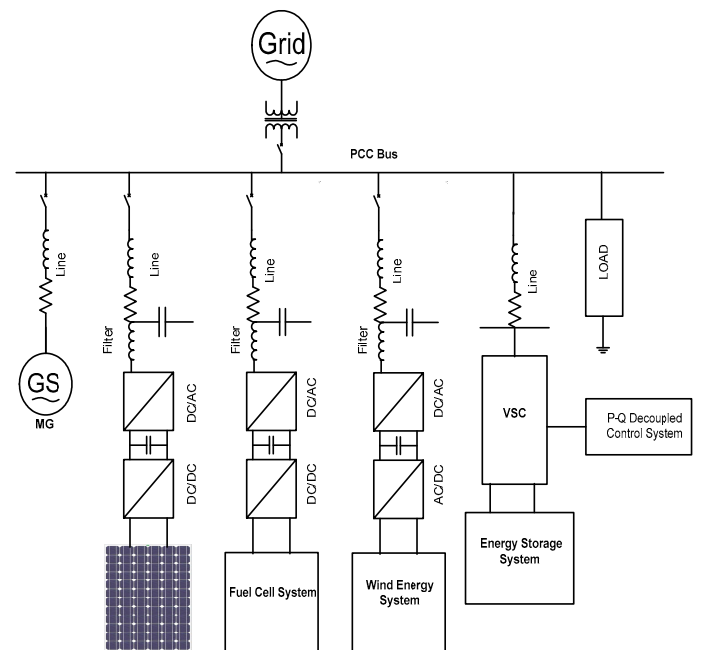


Fig. 2 Detailed layout of microgrid under investigation

Micro-alternator model

The conventional generation system can be modeled as a synchronous alternator driven by a gas-turbine or a diesel engine, and there by generating electrical power of fundamental frequency which makes possible for direct connection to the utility grid. A 3rd order model consisting of the swing equation and internal voltage equation of an alternator is being used as a dynamic model for synchronous generator [12]. The synchronous generator is considered to be equipped with a voltage regulator and using IEEE ST-1A exciter in the field.

The PV system

In a photovoltaic system, the fundamental block is a solar cell, which is a semiconductor device behaving as a current source when solar radiation is incident on it. It assumes the characteristic of an ideal diode when there is no illumination. The incident solar irradiation, array short circuit current, array open circuit voltage, cell temperature and load connected to the array determines the operation characteristics of a solar array [13]. The irradiance applied to a photovoltaic unit dictates the output current and power. There exists a nonlinear I-V and P-V characteristics for a solar array which has dependency of cell temperature and radiation intensity. The relationship between PV output current and temperature is very weak. The temperature has very little effect on PV current, but increased temperatures cause the PV cell to conduct at a low voltage, thus diminishing the power output [14].

The interconnection of photovoltaic system to utility grid requires the use of various power electronics devices. The fundamental electronic circuitry essential for a PV system is called a power conditioning unit. The DC/DC converter, DC link capacitor, DC/AC inverter and filter circuit constitutes the major components of a power conditioning unit. Since the power output of the solar array is DC type of low magnitude, a DC/DC converter is necessary in order to transform the output power from lower level to higher voltage level. Interfacing of DC power producing PV device to main utility grid requires a DC/AC inverter. The occurrence of harmonics and power quality problems due to the inclusion of power electronic devices can be smoothened by employing a LC filter at the output terminals of inverter. In order to mitigate the sub synchronous resonance a resistance is used in series with the filter capacitor.

The wind generator

A horizontal axis wind turbine drives the rotor of a permanent magnet synchronous generator through a gearless drive train model is considered. Field excitation to the generator is obtained through permanent magnets mounted on the generator rotor. Power electronic converters connected between the stator of PMSG and the grid provides an AC voltage of constant grid frequency to be able to feed the grid.

The wind turbine, the gearless drive train, the PMSG, the converter circuits and transmission line connected at grid side constitute the dynamic model of the wind generation system. The IGBT switch, controlled through PWM a method, forming an important component of power electronic converter system, connects directly with the stator of PMSG system and receives an AC power at variable frequency. Two voltage source converters (VSC) connected as a back-to-back converter system through a DC-link constitutes the power electronic circuitry. Generator side converter acts as a rectifier, while the grid side converter acting as an inverter maximizes the power injected into the DC-link. The dynamic model of each of the component involved with the generation system is dealt in the following section.

Fuel cell

Electrochemical devices working on the principle of electrolytic chemical reaction converting the chemical energy of a fuel directly into electricity are called Fuel cells. Only

water and heat are rejected to the surroundings making the fuel cell energy eco-friendly and environmentally useful. High efficiency, low pollutant emissions, high reliability and little maintenance are some of the characteristics of fuel cell power systems [15]. To drive the load connected at the terminals of a fuel cell system, power electronic interface is needed as the obtained power from fuel cells is an unregulated dc voltage. Fuel cells being, low voltage high current devices, require the use of DC boost converter and other associated power electronic circuitry in order to be integrated with the utility grid.

In order to achieve design, control and simulation of modern fuel cell power systems, an accurate model is necessary depicting the dynamic and steady state performances of such systems to a larger extent. Fuel cells of variety of configurations requiring variant fuel combinations have been reported in the literature, but the most commonly used is the proton exchange membrane fuel cell (PEM-FC). Chemical, Experimental and Electrical are the three important model types being reported for fuel cells. Each of this model type has certain limitations due to which they are not suitable for the simulation of modern fuel cell power systems. The generic model, combining the features of chemical and electrical models proposed in [15], is a new approach requiring only a few parameters provided in the manufacturers datasheet, is employed in this work. Such a model is capable of reflecting the effect of operating parameters on the fuel cell and can be easily simulated in the simulation programs.

III. SIMULATION STUDY

The small-signal model for the system is derived by taking small perturbations of the system nonlinear equations. A 36th order state model was employed to give a reasonably detailed dynamics of the system. The breakdown of the component equations is shown in the following table.

TABLE I
The components in the dynamic model

DG	# of Diff. Eqns.	Component Names
Micro-alternator	4	Rotor dynamics, internal voltage equation, exciter
PMSG Wind Generator	13	Turbine-generator drive train, generator stator, generator and grid side converters, LC filter, coupling line
Photo-voltaic system (PV)	8	DC/DC boost converter and inverter, DC link capacitor, LC filter, coupling line
Fuel Cell System	8	Fuel cell stack, electronic circuitry similar to that in PV
STATCOM Storage System	3	Voltage source converter, DC coupling capacitor

The small-signal model of the system is written by perturbing the nonlinear equations around the nominal operating condition and expressing the differential equations in the form,

$$\dot{x} = Ax + Bu \quad (1)$$

Here, x is a 36×1 vector of the states, A is a 36×36 system matrix, u is a vector of the various controls available in the different component models. Sample values of the eigenvalues for micro-alternator input of 0.3 pu, PV input 0.3 pu, wind input 0.5 pu, fuel cell input of 0.3 pu are:

Table II
Eigenvalues of the microgrid system

$-404.4 \pm j3209.9$	$-47.1 \pm j2286.8$	$-94.1 \pm j378.2$	-0.2
$-151.7 \pm j3055.2$	$-497.5 \pm j1065$	-99	$-1.1 \pm j3.6$
$-47.1 \pm j3040.7$	$-59.5 \pm j379.6$	$-0.5 \pm j34.1$	$-0.3 \pm j3.7$
$-191.1 \pm j216.9$	$-4.6 \pm j377$	$-0.1 \pm j9.2$	$-0.1 \pm j4.1$
$-152 \pm j2218.1$	$-94.2 \pm j377.7$	$-1 \pm j0.1$	

A single aggregate load of 1.4 pu is considered in the study. The grid is almost floating drawing a nominal power of 0.0246 pu only.

A closer look at the eigenvalues reveals that about half of them are lightly damped. These eigenvalues are mostly contributed by the electronic circuits and the associated filters. Frequency response and stability of the system are affected by the amount of penetration from the different DG's in the system. Fig.3 shows the variation of real part of the most dominant eigenvalue as the amount of micro-alternator input is varied. Plots are recorded for different extent of penetration from the wind system. Increasing the amount of wind input tends to make the system more stable. From the plot it can be observed that the amount of alternator input is an important factor. At very light alternator penetration, the system is lightly damped, and even unstable. Increased alternator participation tends to stabilize the system more. However from economic and environmental considerations, the alternator input should be a minimum. The input from fuel cell and PV sources are kept at 0.3 pu, each.

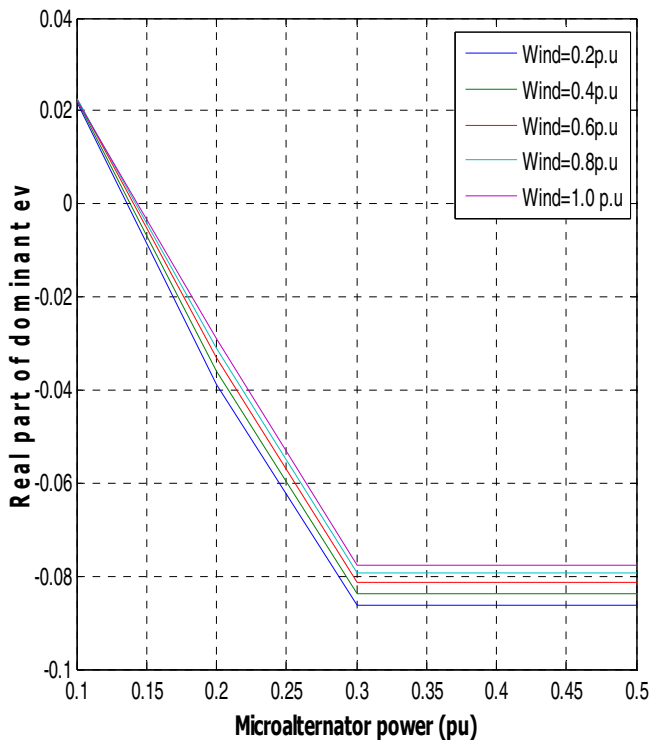


Fig. 3 Real part of dominant eigenvalue against alternator power for different levels of wind penetration

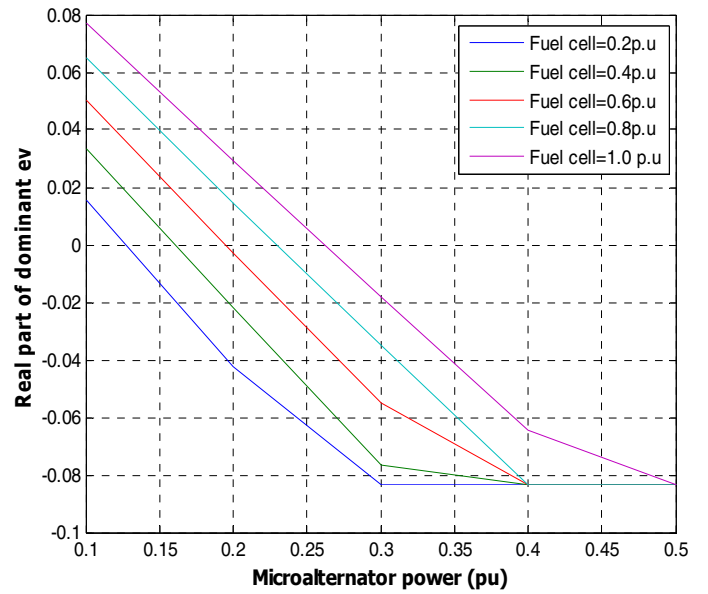


Fig. 4 Real part of dominant eigenvalue against micro-alternator input for various fuel cell power penetrations

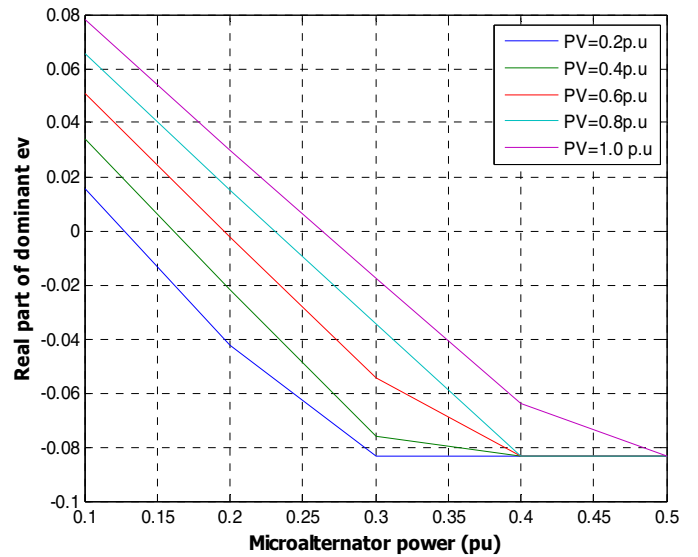


Fig. 5 Real part of dominant eigenvalue vs. alternator power output for different amounts of PV penetrations

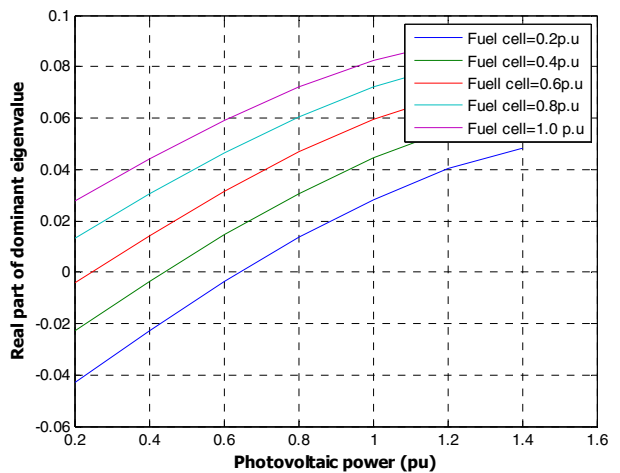


Fig. 6 Real part of dominant eigenvalue against PV power for different fuel cell inputs

Figs. 4 and 5 show the variation of the real parts of the dominant eigenvalues against alternator power input variations for different levels of fuel cell and PV power penetrations,

respectively. It can be observed in both the cases that too much penetration from fuel cell or PV sources reduces the small-signal stability margin. This is particularly so when the micro-alternator participation is small. While participation of alternator is an important item, care should be taken about the share of the other DGs also. Figure 6 shows how the relative values of PV and fuel cell participation can affect the small-signal stability.

Fig.7 shows the simulated speed variation response of the micro-alternator when subjected to a very small disturbance of 10% torque pulse for 0.2 sec duration. The response is obtained from simulation of 36th order nonlinear model of the microgrid dynamics. The responses are shown for 2 micro-alternator loading conditions, 0.2pu and 0.4pu. It can be observed that while the transient response exhibits stable scenario for 0.4pu generation, it is unstable for 0.2pu. This verifies the small-signal stability response obtained through eigenvalue analysis demonstrated earlier. The generations from other sources in both the cases are, wind 0.5pu, PV 0.3 and fuel cell 0.3pu.

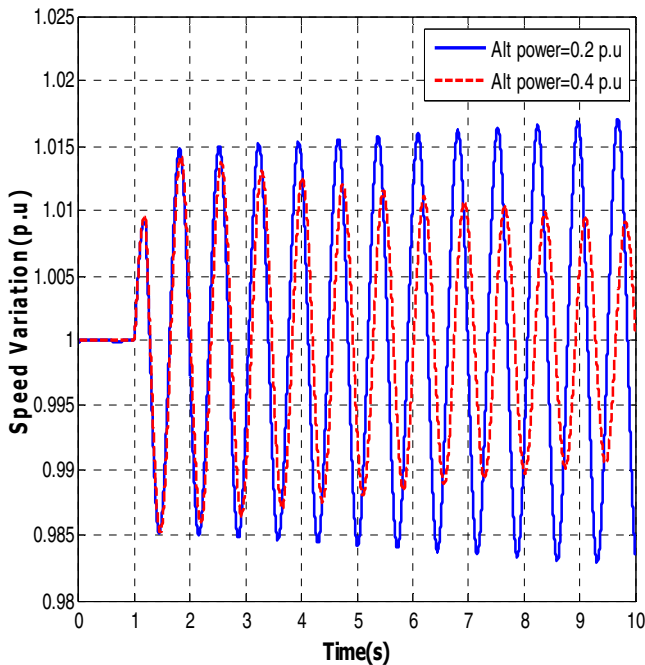


Fig. 7 Speed variation of the alternator at two lightly loaded conditions

IV. CONCLUSIONS

The small-signal behavior of a microgrid when subjected to different extents of participation by the DGs is investigated. The relative stability of the system is ascertained by the contribution of the real part of the dominant eigenvalue. While from economic and environmental considerations alternator input should be small, it has been observed that too little micro-alternator participation can cause relative instability in the system. Participation from wind and PV seems to help system stability when the alternator input is small.

Relatively large fuel cell input tends to have destabilizing effect on the microgrid system. This study indicates that the participation from different DG's should be carefully designed for smooth operation of the microgrid system.

ACKNOWLEDGMENT

This research work is supported by Ahsanullah University of Science and Technology (AUST) and King Fahd University of Petroleum and Minerals (KFUPM). The authors acknowledge the facilities provided by both the institutions.

REFERENCES

- [1] A. Rezvani, M. Izadbakhsh, M. Gandomkar, "Enhancement of microgrid dynamic responses under fault conditions using artificial neural network for fast changes of photovoltaic radiation and FLC for wind turbine", *Energy Systems*, Volume 6, Issue 4, pp. 551-584, 2015.
- [2] M. Hassan and M. Abido, "Optimal design of microgrids in autonomous and grid-connected modes using particle swarm optimization," *IEEE Transactions on Power Electronics*, vol. 26, no. 3, pp. 755-769, 2011.
- [3] W. Deng, X. Tang, and Z. Qi, "Research on dynamic stability of hybrid wind/PV system based on micro-grid", *International Conference on Electrical Machines and Systems, ICEMS 2008*, pp. 2627-2632, 2008.
- [4] T. M. Hussain, "Modeling and control of a microgrid including photovoltaic and wind generation," M.Sc. Engineering Dissertation, King Fahd University of Petroleum and Minerals, 2012.
- [5] M. Ashraf Ali, "Control of a microgrid through energy storage devices using evolutionary and neuro-fuzzy methods", M.Sc. thesis, King Fahd University of Petroleum & Minerals, May 2013.
- [6] Y. Mohamed and E. F. El-Saadany, "Adaptive decentralized droop controller to preserve power sharing stability of paralleled inverters in distributed generation microgrids," *IEEE Transactions on Power Electronics*, vol. 23, no. 6, pp. 2806-2816, Nov. 2008.
- [7] D. Ariyasinghe and D. Vilathgamuwa, "Stability analysis of microgrids with constant power loads," in *2008 IEEE International Conference on Sustainable Energy Technologies*, pp. 279-284, 2008.
- [8] X. Xu, T. Lin, and X. Zha, "Probabilistic analysis of small signal stability of microgrid using point estimate method," in *2009 International Conference on Sustainable Power Generation and Supply*, pp. 1-6, 2009.
- [9] H. Shagafi, H. Karshenas, A. Bakhshai, "Application of integrated series compensator in damping power oscillations in standalone microgrids", *Canadian Journal of Electrical and Computer Engineering*, Volume 38, Issue 1, Article number 7072605, pp. 2-9, 2015.
- [10] N. Srivatchan, P. Rangarajan, S. Rajalakshmi, "Robust PI controller for frequency stabilization of islanded microgrid operation using battery Energy Storage System", *Indian Journal of Science and Technology* Volume 8, Issue 29, Article number 78562, 2015.
- [11] X. Tang, W. Deng, Z. Qi, Investigation of the dynamic stability of microgrid, *IEEE Transactions on Power Systems*, Volume 29, Issue 2, Article number 6655978, pp. 698-706, 2014.
- [12] Y. Yu, *Electric Power System Dynamics*. Academic Press Inc, 1983, p. 256.
- [13] H. Tsai, C. Tu, and Y. Su, "Development of generalized photovoltaic model using MATLAB / SIMULINK," *Proceedings of the World Congress on Engineering and Computer Science 2008*, pp. 0-5, 2008.
- [14] V. J. Fesharaki, M. Dehghani, and J. J. Fesharaki, "The Effect of Temperature on Photovoltaic Cell Efficiency," *Proceedings of the 1st International Conference on Emerging Trends in Energy Conservation - ETEC Tehran*, Tehran, Iran, pp. 20-21, November, 2011.
- [15] S. M. Njoya, O. Tremblay, and L. -a. Dessaint, "A generic fuel cell model for the simulation of fuel cell vehicles," *2009 IEEE Vehicle Power and Propulsion Conference*, pp. 1722-1729, Sep. 2009.

Electrical Properties of Conductive Polymer Composites Containing Silver-coated Glass Particle and Multi-walled Carbon Nanotube Hybrid Fillers

Al Basir

School of Engineering and Materials Science
Queen Mary University of London
United Kingdom
E-mail: al.basir005@yahoo.com

Md. Imran Khan

Institute of Information and Communication Technology
Bangladesh University of Engineering and Technology (BUET)
Dhaka-1000, Bangladesh
E-mail: imrankhankuet@gmail.com

Abstract—Hybrid conductive polymer composites were prepared by conventional melt mixing method employing silver-coated glass particles and multi-walled carbon nanotubes (MWNT) as hybrid fillers and high density polyethylene (HDPE) as polymer matrix. The percolation behavior of HDPE/20 wt. % silver-coated glass particle/MWNT and HDPE/25 wt. % silver-coated glass particle/MWNT hybrid composites were examined and it was observed that both types of polymer composites clearly displayed percolated behavior. Experiments showed that the room temperature resistivity of hybrid composites decreased with increasing MWNT content and it was observed that higher loading of silver-coated glass particles content reduce the resistivity of hybrid composites. It was also found that hybrid composites containing 20 wt. % silver-coated glass particles with different weight percentages (0.5 wt. %, 0.55 wt. % and 0.6 wt. %) of MWNT and composites having 25 wt. % silver-coated glass particles with different weight percentages (0.4 wt. %, 0.5 wt. % and 0.6 wt. %) of MWNT exhibited positive temperature co-efficient (PTC) effect that occurred around the melting temperature of composites due to expansion of polymer matrix.

Keywords—Conductive polymer composite, multi-walled carbon nanotube, silver-coated glass particles, high density polyethylene, percolation behavior, positive temperature co-efficient effect.

I. INTRODUCTION

Polymer based positive temperature co-efficient (PTC) materials are extensively used in variety of industrial applications, such as self-regulating heaters, shielding structures and switching sensors [1-10]. Currently, a good number of researches are going on to develop PTC materials that will have lower filler content. Development of PTC materials with lower conduction thresholds not only reduce the manufacturing costs but also increase their processability. Basically conductive fillers play one of the most vital roles for the conductivity of the polymer composites. The conductivity of conductive polymer composites (CPCs) relies on various criteria of conductive fillers such as: type of conductive fillers, filler size, filler shape, amount of fillers used in CPC etc [11]. Moreover the dispersion of conductive fillers, formation of conductive network by fillers and disruption of this conductive network due to expansion of polymer matrix has indomitable influence on positive temperature coefficient effect of CPC.

Metal, ceramic and carbon based conductive fillers are mostly used by the researchers in order to examine the PTC behavior of polymer composites. In this research, hybrid composite materials were fabricated by conventional melt mixing of high density polyethylene (HDPE) with multi-walled with carbon nanotubes (MWNT) and silver-coated glass particles. This paper is focused on research of the electrical properties and PTC behavior of MWNT and silver-coated glass particles filled HDPE hybrid polymer composites.

This paper is organized as follows, section II part A describes about materials, part B and part C shows preparation of composite and composites characterization respectively. Three consecutive parts A, B, C of section III discussed about electrical properties, PTC behavior and morphology of HDPE/silver-coated glass particle/MWNT composites respectively. Finally section IV is for conclusion.

II. EXPERIMENTAL SECTION

A. Materials

Distrupol supplied their HDPE under the trade name Rigidex HD5218EA. The melt flow index, density and the melting temperature of Rigidex were 18g/10 min, 0.952 g/cm³ and 131° C respectively. Multi-walled carbon nanotubes (MWNT) were used as conductive fillers were supplied by Nanocyl with the trade name NC7000. The average diameter of MWNT was about 9.5 nm whereas the average length was about 1.5µm. Silver-coated glass filler particles used in the experiment were spherical in shape and was supplied by Potters Industries LLC, USA under the trade name TP50S06. In this experiment, silver-coated glass particles were used with multi-walled carbon nanotubes (MWNTs) as hybrid fillers.

The statistical analysis was carried out on silver-coated glass particles to find out their size distribution. The particle size distribution of silver coated glass is shown in figure 1. D10, D50 and D90 particle size values of silver coated glass fillers were obtained from data analysis. The values of D10, D50 and D90 are listed in table 1. Here D50 defined that 50% particles in the sample are larger than 47.81 µm and other 50% particles are smaller than 47.81 µm. On the other hand, D90 illustrated that 90% of the distribution lies below 55.89 µm

whereas the value of D10 described that 10 % of the particles lies below 39.94 μm .

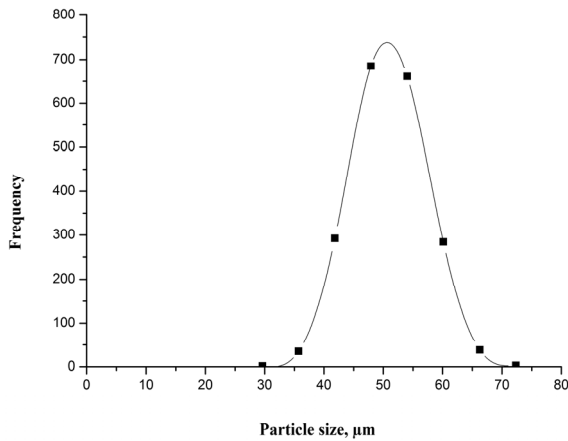


Fig. 1. Distribution of size for silver coated glass particles (TP50S06)

TABLE I. D10, D50 AND D90 PARTICLE SIZE DISTRIBUTION (SILVER COATED GLASS PARTICLE-TP50S06)

D10	D50	D90
39.94 μm	47.81 μm	55.89 μm

B. Preparation Of Composite Specimens

MWNT and silver-coated glass particles filled HDPE hybrid conductive polymer composites were prepared by conventional melt mixing method in twin screw mini-extruder (DSM Micro 15) employing a screw speed of 50 rpm at 170°C for 5 minutes. Before that, 5 wt% MWNT masterbatch and masterbatches containing different weight percentages of silver-coated glass particles were prepared. The screw speed, temperature and mixing time followed to prepare HDPE /MWNT masterbatch was 50 rpm, 170°C and 3 minutes respectively. On the other hand, HDPE/Silver-coated glass particle masterbatch was prepared using a screw speed of 50 rpm at 170°C for 15 minutes. HDPE/MWNT and HDPE/Silver-coated glass particle masterbatches were diluted by the addition of pure HDPE to form the desired concentrations. The hybrid blends were then pelletized and hot pressed at 200°C and 60 bars pressure to get samples with 23.90 mm $\times$ 3.80 mm $\times$ 1.63 mm dimensions.

C. Composites Characterization

In order to investigate the electrical properties, silver paste was applied at the ends of each sample surface and the reason of doing so was to reduce the contact resistance between sample and electrodes. Keithley picoammeter and multimeter was used to measure the resistance along the sample length. An applied voltage of 2V was employed in steps of 0.2V to conduct electrical tests at room temperature. The electrical

resistivity of each sample was calculated by employing equation (1).

$$\text{Resistivity, } \rho = R \frac{A}{l} (\Omega \cdot \text{cm}) \quad (1)$$

Here, R = Electrical resistance along the sample length (Ω).

A = Cross sectional Area of sample (cm^2). Cross sectional area was measured from the width (w) and thickness (t) of the sample as schematically shown in figure 2.

l = Distance between electrodes after clamping to the ends of the sample (cm).

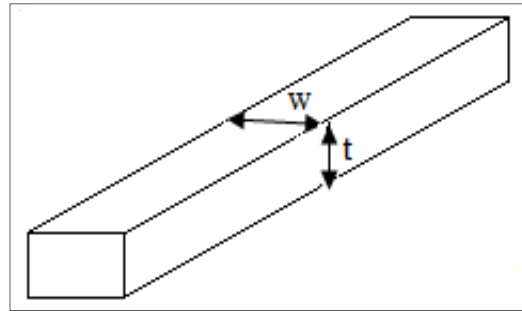


Fig. 2. Schematic diagram of fabricated sample.

The volume resistivity of the samples was measured as a function of temperature by placing the samples in a heating chamber and then clamping the copper electrodes to their ends. In order to reduce the contact resistance between sample and the electrodes, silver paint was applied at ends of the samples before clamping them with electrodes. The technique followed to investigate the PTC characteristics was to employ applied voltage 2V while continuously measuring the current of heated specimen from room temperature to maximum 200°C.

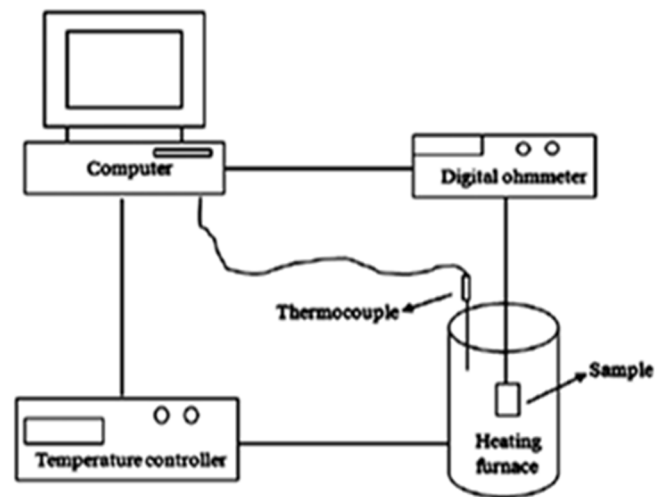


Fig. 3. Schematic diagram of testing the change in resistivity with temperature.

Whenever temperature reached the desired value, the voltage was switched off and the tested samples were allowed to cool to room temperature. A schematic of experimental set up is shown in figure 3. The PTC intensity of the samples was calculated by the formula given in equation (2).

$$\text{PTC intensity, } I_{PTC} = \log\left(\frac{\rho_{max}}{\rho_{rt}}\right) \quad (2)$$

Here, ρ_{max} = Maximum resistivity ($\Omega \cdot \text{cm}$),

ρ_{rt} = Resistivity at room temperature ($\Omega \cdot \text{cm}$).

The scanning electron microscope (FEI Inspector-F, Netherlands) was employed to understand the morphology of the fabricated samples. The hybrid composites were fractured at atmosphere of liquid nitrogen and after that the fractured surface was coated with a thin layer of gold in order to prevent charging before doing the imaging.

III. RESULTS AND DISCUSSION

A. Electrical Properties Of HDPE/Silver-coated Glass Particle/MWNT Composites

Figure 4 and 5 show the volume resistivity as a function of MWNT content for HDPE/20 wt. % silver-coated glass particle/MWNT and HDPE/25 wt. % silver-coated glass particle/MWNT hybrid composites respectively. Making HDPE composites by implementing 20 wt. % silver-coated glass particle with different weight percentages (0.5 wt.%, 0.55 wt. %, 0.6 wt. %, 0.7 wt. % and 1.0 wt.%) of MWNT displayed percolated behavior as shown in figure 4. Even a more reduction in resistivity was observed while hybrid composites were prepared by using 25 wt. % silver-coated glass particles with different weight percentages (0.4 wt. %, 0.5 wt. %, 0.6 wt. % and 0.7 wt. %) of MWNT which is shown in figure 5. Basically MWNTs in the range between 0.4 and 1.0 wt. % were used with silver-coated glass particles of two different weight percentages (20 wt. % and 25 wt.%) as hybrid fillers to prepare different composites and it was evident that employing MWNTs of less than 1 wt. % is inadequate to form a conductive path by MWNTs itself [12]. It was found that percolation threshold for MWNT filled low melt flow index (LMFI) HDPE composites to be 1.9 wt. % MWNT [12]. Moreover, percolation threshold 2 wt. % MWNT was found for HDPE/MWNT composites prepared by solution precipitation method whereas threshold 2.8 wt. % MWNT was observed for composites made by melt mixing method [9,13]. Employing highly conductive MWNTs with silver-coated glass particles as hybrid fillers lead MWNTs in the formation of network structure of conductive particles with silver-coated glass filler particles which lower the resistivity of hybrid composites at room temperature and this is comparable with the observation of Lee *et al.* who conducted experiment on HDPE/CB/MWNT hybrid composites [14]. From figure 4, a drop in resistivity was observed between hybrid composites containing 20 wt. % silver-coated glass particles with 0.5 wt. % MWNT and 20 wt. % silver-coated glass particles with 1.0 wt. % MWNT. This shows that,

resistivity decrease with higher loading of conductive fillers due to an increase in conductive paths within polymer matrix.

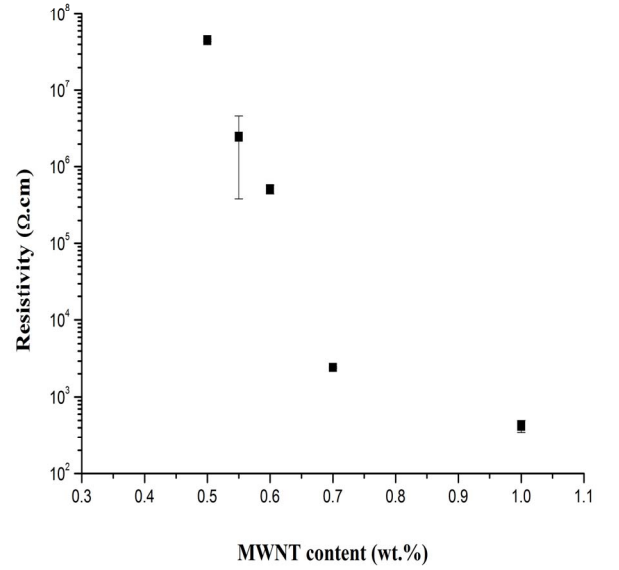


Fig. 4. Electrical volume resistivity of HDPE filled with 20 wt. % silver-coated glass particle with different weight percentages (0.5 wt. %, 0.55 wt. %, 0.6 wt. %, 0.7 wt. % and 1.0 wt. %) of MWNT.

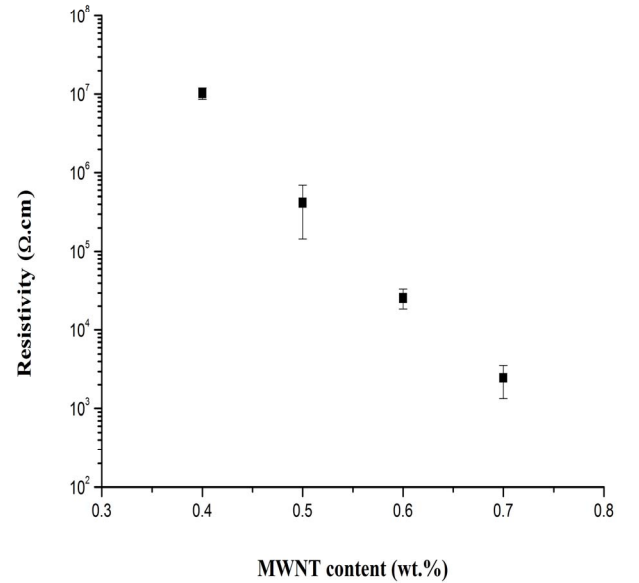


Fig. 5. Electrical volume resistivity of HDPE filled with 25 wt. % silver-coated glass particle with different weight percentages (0.4 wt. %, 0.5 wt. %, 0.6 wt. % and 0.7 wt. %) of MWNT.

Similar reason can be applied to justify figure 5 where a dramatic decrease in resistivity between HDPE/ 25 wt. % silver-coated glass particle/0.4 wt. % MWNT and HDPE/ 25 wt. % silver-coated glass particle/0.7 wt. % MWNT

composites were found. It was observed from figure 4 and 5 that that hybrid composites containing 25 wt. % silver-coated glass particles with different weight percentages (0.4 wt. %, 0.5 wt. % and 0.6 wt. %) of MWNT displayed lower resistivity in comparison to hybrid composites containing 20 wt. % silver-coated glass particles with different weight percentages (0.5 wt. %, 0.55 wt. % and 0.6 wt. %) of MWNT. It is evident that the loading of fillers in hybrid composites containing 25 wt. % silver coated glass particle with different weight percentages of MWNT is higher than composites filled with 20 wt. % silver coated glass particle with different weight percentages of MWNT. It indicates that the higher loadings of fillers increase the conductive pathways within polymer matrix and consequently decrease the resistivities.

B. PTC Behavior Of HDPE/Silver-coated Glass Particle/MWNT Composites

It was observed that all the hybrid composites used for the investigation of PTC behavior displayed PTC effect. As a demonstration, PTC effect of composite containing 25 wt. % silver-coated glass particle and 0.7 wt % MWNT hybrid fillers is shown in figure 6. Here polymer matrix starts to expand around the melting temperature of hybrid composite (131°C). This expansion of matrix disrupts the conductive network constructed by silver-coated glass particles and MWNTs hybrid fillers and the consequence is the increase in resistivity to exhibit PTC effect.

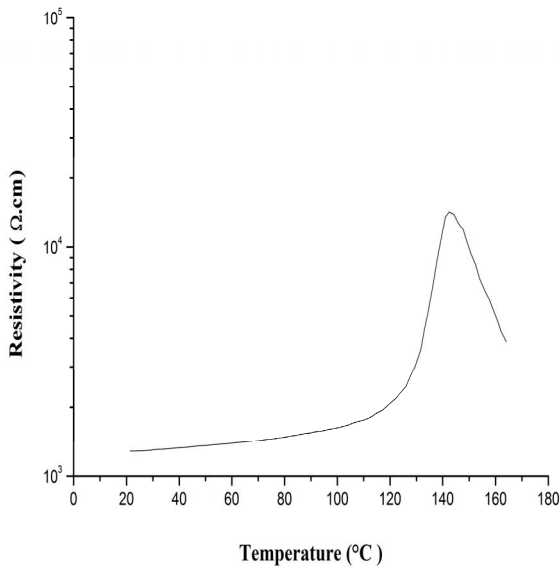


Fig. 6. PTC curve of HDPE containing 25 wt. % of silver-coated glass particle and 0.7 wt. % MWNT hybrid fillers.

A sudden drop in resistivity after the melting point can be observed from figure 6 which denotes that composite prepared with MWNTs and silver-coated glass particles hybrid fillers displayed negative temperature co-efficient (NTC) effect. PTC intensity (I_{PTC}) is calculated from the logarithm ratio of maximum resistivity (ρ_{max}) and resistivity at room

temperature (ρ_{rt}) as given in equation (2) and the calculated value of PTC intensity for HDPE /25 wt. % silver-coated glass particle/2 wt.% MWNT hybrid composite was approximately 1.2.

C. Morphology Of HDPE/Silver-coated Glass Particle/MWNT Composite

A Scanning electron microscope (SEM) was used to study the morphology of hybrid composite. Figure 7 shows the cross-sectional view of composite containing 20 wt. % silver-coated glass particles and 0.7 wt. % MWNT hybrid fillers at different magnification. Figure 7, at 260x magnification, shows the dispersion of silver-coated glass particles within polymer matrix. It is also noticeable from figure 7 and 8 (at 1200x magnification) that during fracture in liquid nitrogen atmosphere, silver-coated glass particles were totally separated from polymer matrix which denotes a weak surface interaction between matrix and silver-coated glass filler particles. Moreover, from figure 9, at 80000x magnification, it can be observed that isolated MWNT (shown as small bright contrast) are also dispersed in HDPE.

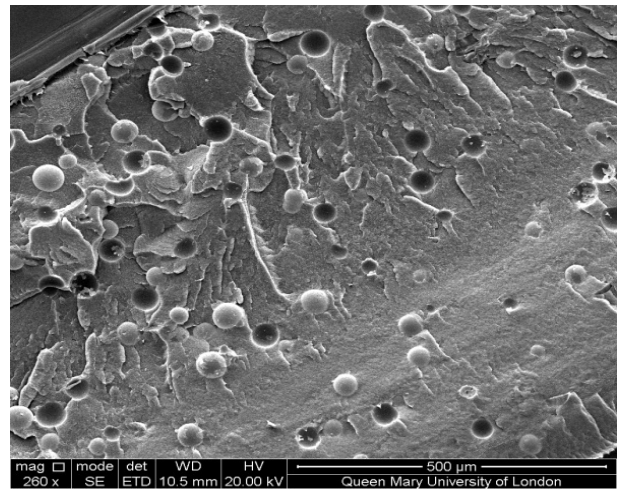


Fig. 7. SEM image of HDPE/20 wt. % silver-coated glass particle/0.7 wt. % MWNT composite at 260x magnification.

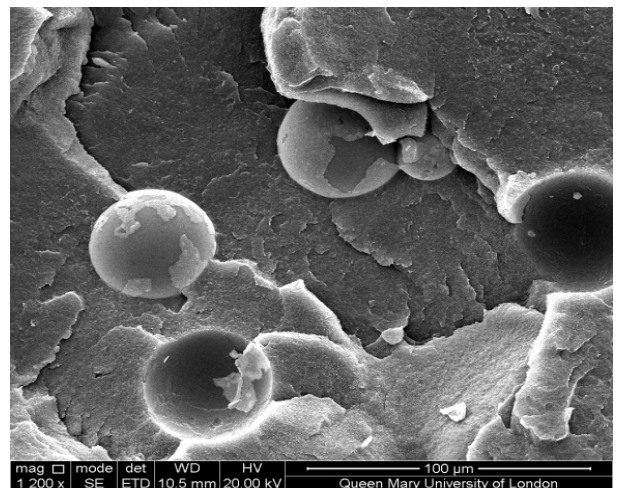


Fig. 8. SEM images of HDPE/20 wt. % silver-coated glass particle/0.7 wt. % MWNT composite at 1200x magnification.

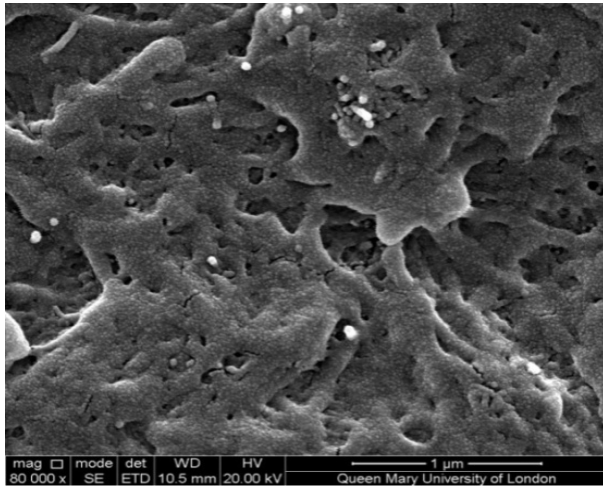


Fig. 9. SEM image of HDPE/20 wt. % silver-coated glass particle/0.7 wt. % MWNT composite at 80000x magnification.

IV. CONCLUSION

HDPE/20 wt. % silver-coated glass particle/MWNT and HDPE/25 wt. % silver-coated glass particle/MWNT hybrid conductive polymer composites fabricated by conventional melt mixing method displayed percolated behavior. Increase of MWNT content decreased the resistivity of hybrid composites with silver-coated glass particle content of 20 wt. % and 25 wt. %. It denotes that higher loading of hybrid fillers make more conductive paths within polymer matrix and therefore causing a reduction in resistivity. Moreover, displaying of lower resistivity by HDPE composites containing 25 wt. % silver-coated glass particles with different MWNT contents (0.4 wt. %, 0.5 wt. % and 0.6 wt. %) in comparison to HDPE composites containing 20 wt. % silver-coated glass particles with different MWNT contents (0.5 wt. %, 0.55 wt. % and 0.6 wt. %) clearly showed that resistivity decrease with higher loading of fillers and reason behind this is the increase of conductive paths within polymer matrix. Furthermore, PTC effect is the most important property of conductive polymer composite and hybrid composites prepared by silver-coated glass particle and MWNT fillers displayed PTC effect around the melting temperature (131°C) of composites because of expansion of polymer matrix.

REFERENCES

- [1] Bai, B. C., Kang, S. C., Im, J. S., Lee, S. H. and Lee, Young-Seak, "Effect of oxyfluorinated multi-walled carbon nanotube additives on positive temperature coefficient/negative temperature coefficient behavior in high-density polyethylene polymer switches.", *Materials Research Bulletin* vol.46,no.9, pp.1391-1397,2011.
- [2] Li, Q., Basavarajaiah, S., Kim, N. M., Heo, Seok-Bong and Lee, J. H. 2010. *Synergy effect of hybrid fillers on the positive temperature coefficient behavior of polypropylene/ultra-high molecular weight polyethylene composites*. [Online] Available: <http://dx.doi.org/10.1002/app.31478> [Accessed: 26 July 2013].
- [3] Mamunya, Y. P., Davydenko, V. V., Pissis, P. and Lebedev, E. V., "Electrical and thermal conductivity of polymers filled with metal powders." *European Polymer Journal*, vol. 38, no. 9, pp.1887-1897,2002.
- [4] Lee Joong-Hee, Kim, S. K. and Kim, N. H., "Effects of the addition of multi-walled carbon nanotubes on the positive temperature co-efficient

- characteristics of carbon-black-filled high-density polyethylene nanocomposites.", *Scripta Materialia*, vol. 55, no.12, pp. 1119-1122, 2006.
- [5] Bauhofer, W. and Kovacs, J. Z., "A review and analysis of electrical percolation in carbon nanotube polymer composites", *Composites Science and Technology*, vol. 69, no. 10, pp. 1486-1498, 2009.
- [6] Lisunova, M. O., Mamunya, Y. P., Lebovka, N. I. and Melezhyk, A. V. "Percolation behaviour of ultrahigh molecular weight polyethylene/multi-walled carbon nanotubes composites", *European Polymer Journal*, vol. 43, no.3, pp. 949-958,2007.
- [7] Shen, L., Wang, F. Q., Yang, H. and Meng, Q. R., "The combined effects of carbon black and carbon fiber on the electrical properties of composites based on polyethylene or polyethylene/polypropylene blend." *Polymer testing*, vol. 30, no.4, pp.442-448,2011.
- [8] Dang Zhi-Min, Li Wei-Kang and Xu Hai-Ping. (2009). "Origin of remarkable positive temperature coefficient effect in the modified carbon black and carbon fiber co-filled polymer composites." *Journal of Applied Physics*. [Online] Available: <http://link.aip.org/link/?JAP/106/024913/1> [Accessed: 30th November 2012].
- [9] Liu, F., Zhang, X., Li, W., Cheng, J., Tao, X., Li, Y. and Sheng, L., "Investigation of the electrical conductivity of HDPE composites filled with bundle-like MWNTs", *Composites Part A: Applied Science and Manufacturing*, vol. 40, no.11, pp. 1717-1721,2009.
- [10] Rybak, A., Boiteux, G., Melis, F. and Seytre, G., "Conductive polymer composites based on metallic nanofiller as smart materials for current limiting devices." *Composites Science and Technology*, vol.70, no. 2, pp. 410-416, 2010.
- [11] Stru"mpler, R. and Glatz-Reichenbach, J., "FEATURE ARTICLE Conducting Polymer Composites", *Journal of Electroceramics*, vol.3, no. 4, pp. 329-346, 1999.
- [12] Basir, A., "Experimental investigation of the electrical conductivity of MWNT filled HDPE composites." *11th International Conference on Mechanical Engineering (ICME)*, Dhaka, Bangladesh,2015.
- [13] He, X. J., Du, J. H., Ying, Z. and Cheng, H. M. (2005). "Positive temperature coefficient effect in multiwalled carbon nanotube/high-density polyethylene composites." *Applied Physics Letters*. [Online] Available: http://apl.aip.org/resource/1/apl/v86/i6/p062112_s1?bypassSSO=1 [Accessed: 17 July 2012].
- [14] Lee Joong-Hee, Kim, S. K. and Kim, N. H., "Effects of the addition of multi-walled carbon nanotubes on the positive temperature co-efficient characteristics of carbon-black-filled high-density polyethylene nanocomposites." *Scripta Materialia*, vol.55, no. 12, pp. 1119-1122, 2006.

Performance of Grid-voltage Synchronization Algorithms Build on Frequency-and Phase-locked Loop During Severe Grid Fault Conditions

Khaled Syfullah Fuad

Department of Computer Science & Engineering
Britannia University, Bangladesh
Email: khaled.fuad@britannia.ac

Eklas Hossain

Department of Electrical Engineering & Renewable Energy
Oregon Institute of Technology, USA
Email: eklas.hossain@oit.edu

Abstract—Since the beginning of twenty first century, Renewable energy resources have been developing into the fuel of choice to handle the upcoming electricity demand around the world. To implement the renewable energy resources, Microgrid—a localized group of electricity sources and loads—is considered as the most suitable option to shift over the conventional utility grid due to a number of advantages. In case of grid-tied microgrid application, the voltage synchronization is required to experience desirable performance. This paper evaluates the performance of two leading comprehensive synchronization algorithm, Decoupled Double Synchronous Reference Frame PLL (DDSRF-PLL) and Double Second-order Generalized Integrator FLL (DSOGI-FLL) under adverse grid fault conditions. Results show that their performance does not obey modern grid codes during the harmonic-polluted situation. An advanced algorithm, decoupled multiple Synchronous Reference Frame PLL (DMSRF-PLL) build on SRF-PLL, might exhibit accurate performance when the utility voltage contains harmonic distortion.

Index Terms—unbalanced voltage, grid-connected power converter, harmonics, phase-locked loop, second order generalized integrator, voltage sags.

I. INTRODUCTION

Due to profusion in nature and environment-friendly power generation approaches, power sector specialists and professionals prefer to adopt renewable energy resources to face the next generation energy crisis. To implement the renewable energy resources, microgrid system has been already constructed and available in service to reach electricity in the grid-isolated areas or to provide any mediocre demand in industrial setup. However, microgrid deals with utility grid connection, hence voltage synchronization is required between these two electricity sources for the desired operation. Without proper synchronization, efficiency may be hampered or any unwanted situation such as black-out or brown-out can be occurred. Hence, voltage synchronization is a prime concern for the power engineers and researchers around the world. Therefore, a number of research works have been accomplished on voltage synchronization techniques by the prominent researchers and professionals [1]-[5].

Phase-locked loop (PLL) conventionally a well-known method for Grid-voltage synchronization. A different algorithm formed by using frequency-locked loop (FLL), can be

a useful option as grid synchronization method [12]. PLL estimates the grid-phase angle, whereas FLL estimates the grid frequency as a fundamental component. Unbalanced voltages are the primary result of the maximum of grid faults. Due to this fact, exact plus quick extraction of the positive- and negative-sequence values of the voltage at the point of common coupling (PCC) are important to obey modern grid code requirements (GCRs) [23].

In forthcoming sections in this paper: 1) three-phase voltages are characterized into balanced and unbalanced space vector form and voltage sags are classified; 2) Decoupled Double Synchronous Reference Frame PLL (DDSRF-PLL) and Double Second-Order Generalized Integrator FLL (DSOGI-FLL) are studied and modeled accordingly 3) In the end, the performance between DDSRF-PLL and DSOGI-FLL under different fault conditions is tested and evaluated.

II. BALANCED AND UNBALANCED THREE-PHASE VOLTAGE

Balanced three-phase voltage abc is expressed as [19],

$$\mathbf{u}_{abc}(t) = \begin{bmatrix} u_a(t) \\ u_b(t) \\ u_c(t) \end{bmatrix} = \begin{bmatrix} U_a \cos(\omega t + \phi_a) \\ U_b \cos(\omega t - \frac{2\pi}{3} + \phi_b) \\ U_c \cos(\omega t + \frac{2\pi}{3} + \phi_c) \end{bmatrix} \quad (1)$$

where U_a , U_b and U_c denotes the maximum value of $u_a(t)$, $u_b(t)$ and $u_c(t)$ phase voltages, respectively. The unsymmetric or unbalanced three-phase system in (1) is decomposed with zero-, positive- and negative-sequence and given below:

$$\mathbf{u}_{abc}(t) = \underbrace{\begin{bmatrix} u_a(t) \\ u_b(t) \\ u_c(t) \end{bmatrix}}_{\mathbf{u}_{abc}^+} = \underbrace{\begin{bmatrix} U_+ \cos(\omega t + \phi^{+1}) \\ U_+ \cos(\omega t - \frac{2\pi}{3} + \phi^{+1}) \\ U_+ \cos(\omega t + \frac{2\pi}{3} + \phi^{+1}) \end{bmatrix}}_{\mathbf{u}_{abc}^+} + \underbrace{\begin{bmatrix} U_- \cos(\omega t + \phi^{-1}) \\ U_- \cos(\omega t + \frac{2\pi}{3} + \phi^{-1}) \\ U_- \cos(\omega t - \frac{2\pi}{3} + \phi^{-1}) \end{bmatrix}}_{\mathbf{u}_{abc}^-} + \underbrace{\begin{bmatrix} u_0(t) \\ u_0(t) \\ u_0(t) \end{bmatrix}}_{\mathbf{u}_{abc}^0} \quad (2)$$

where U_+ and U_- are the peak value of the positive and negative sequence components and $\mathbf{u}_{abc}^+$, $\mathbf{u}_{abc}^-$, $\mathbf{u}_{abc}^0$ represent positive-, negative- and zero-sequence voltages of three-phase voltage.

A. Space Vectors

$\mathbf{u}$ is the space vector in synchronous coordinates, symbolized by boldface and its components are expressed with subscripts d and q . The space vector in stationary reference frame $\alpha\beta$, $\mathbf{u}^s$ is symbolized with superscript s and subscripts α and β are used to denote the components of vector [12]. $\mathbf{u}^s$ is expressed for positive- and negative-sequence components as,

$$\mathbf{u}^s = \underbrace{U_+ \begin{bmatrix} \cos(\omega t) \\ \sin(\omega t) \end{bmatrix}}_{\mathbf{u}_+^s} + \underbrace{U_- \begin{bmatrix} \cos(-\omega t + \phi^{-1}) \\ \sin(-\omega t + \phi^{-1}) \end{bmatrix}}_{\mathbf{u}_-^s} \quad (3)$$

where $\mathbf{u}^s$ consist of two sub-vectors; $\mathbf{u}_+^s$, having angular frequency ω and $\mathbf{u}_-^s$, having negative $-\omega$.

B. Voltage Dip or Sag

Different type of voltage dip are categorized as 1. Three-phase and three-phase-to-ground fault (type A); 2. Single-phase-to-ground fault (type B); 3. Phase-to-phase fault (type C) and 4. Two-phase-to-ground fault (type E) [21]-[22].

Dispersion of these four type of voltage dips rise another three new voltage dips. Which are: 1. dispersion of type C (type D); 2. dispersion of type E (type F) and 3. dispersion of type F (type G). For a demonstration, type C can be expressed by the following equation:

$$\begin{bmatrix} U_+ \\ U_- \\ U_0 \end{bmatrix} = \begin{bmatrix} \frac{1}{2}(1+D) \\ \frac{1}{2}(1-D) \\ 0 \end{bmatrix} U_{pf} \quad (4)$$

here pre-fault voltage is symbolized by U_{pf} and phase angle jump is expressed by D .

III. DECOUPLED DOUBLE SYNCHRONOUS REFERENCE FRAME PLL

Three-phase synchronous reference frame (SRF) formed by using a traditional control method phase locked loop (PLL), mutually known as SRF-PLL is a traditional skill used for positive-sequence detection of three-phase voltage. Fig. 1 depicts a conventional diagram of SRF-PLL. PLL phase angle $\hat{\theta}$ and frequency $\hat{\omega}$ have the ability to trace the precise value of utility θ and ω , respectively during balanced voltage situation. But during unbalanced voltage situation, SRF-PLL cannot correctly determine the values [9]. So for the control of

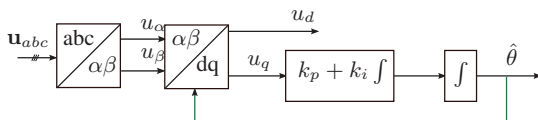


Fig. 1. Traditional three-phase SRF-PLL control diagram.

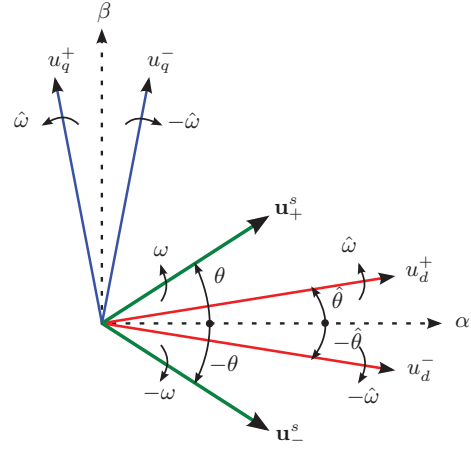


Fig. 2. Illustration of SDRF.

power converters, the conventional SRF-PLL technique cannot be used.

Now $\mathbf{u}^s$ is thrown on two distinct SRF's in Double SRF-PLL technique [10]. dq^+ reference frame having $\hat{\omega}$ is integrated with $\mathbf{u}_+^s$ and another dq^- reference frame having $-\hat{\omega}$ is integrated with $\mathbf{u}_-^s$. Illustration of synchronous double reference frame (SDRF) is shown in Fig. 2. Then $\mathbf{u}^s$ can be written for these reference frames as [9],

$$\mathbf{u}^+ = \begin{bmatrix} u_d^+ \\ u_q^+ \end{bmatrix} = \underbrace{U_+ \begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{dc \text{ part}} + \underbrace{U_- \begin{bmatrix} \cos(-2\omega t + \phi^{-1}) \\ \sin(-2\omega t + \phi^{-1}) \end{bmatrix}}_{ac \text{ part}} \quad (5a)$$

$$\mathbf{u}^- = \begin{bmatrix} u_d^- \\ u_q^- \end{bmatrix} = \underbrace{U_+ \begin{bmatrix} \cos(2\omega t) \\ \sin(2\omega t) \end{bmatrix}}_{ac \text{ part}} + \underbrace{U_- \begin{bmatrix} \cos(\phi^{-1}) \\ \sin(\phi^{-1}) \end{bmatrix}}_{dc \text{ part}} \quad (5b)$$

here superscript '+' indicates the vectors with positive ω and '-' indicates the vectors with negative $-\omega$, respectively and $\hat{\theta} = \int \hat{\omega} dt$ holds during transients. The dc part for $\mathbf{u}^+$ and $\mathbf{u}^-$ relies on U_+ and U_- , respectively in (5a) and (5b). The vectors $\mathbf{u}^+$ and $\mathbf{u}^-$ are rotating in opposite direction and 2ω denotes the oscillation due to the coupling of these two axes. As a result, 2ω creates the disturbance during fast and exact estimation of U_+ and U_- . Thus for the control of power converters, the traditional SRF-PLL method cannot be employed.

A cross-feedback decoupling network given in Fig. 3, simply cancels the oscillating term of both axes and can estimate the dc values of both positive and negative reference frames. As a consequence, $\mathbf{u}_{abc}^+$ and $\mathbf{u}_{abc}^-$ of unbalanced three-phase system can be track out easily [6]. Symbols with 'hat' express the estimated value from the system. Now decoupling cells

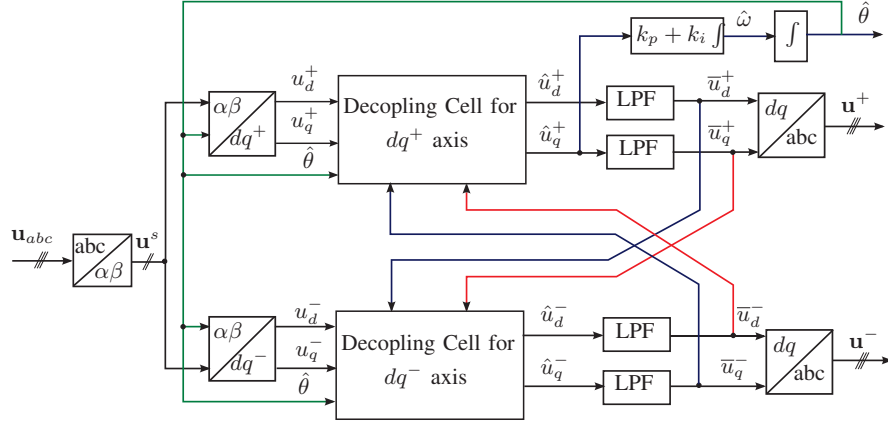


Fig. 3. Decoupled double synchronous reference frame PLL (DDSRF-PLL).

equations are,

$$\bar{u}_d^+ = \frac{\omega_f}{s + \omega_f} (u_d^+ - \bar{u}_d^- \cos 2\omega t - \bar{u}_q^- \sin 2\omega t) \quad (6a)$$

$$\bar{u}_q^+ = \frac{\omega_f}{s + \omega_f} (u_q^+ + \bar{u}_d^- \sin 2\omega t - \bar{u}_q^- \cos 2\omega t) \quad (6b)$$

$$\bar{u}_d^- = \frac{\omega_f}{s + \omega_f} (u_d^- - \bar{u}_d^+ \cos 2\omega t + \bar{u}_q^+ \sin 2\omega t) \quad (6c)$$

$$\bar{u}_q^- = \frac{\omega_f}{s + \omega_f} (u_q^- - \bar{u}_d^+ \sin 2\omega t - \bar{u}_q^+ \cos 2\omega t) \quad (6d)$$

In Fig. 3, low pass filter (LPF) block perform filtering and can be written as [11],

$$LPF(s) = \frac{\omega_f}{s + \omega_f} \quad (7)$$

where, LPF cut-off frequency is given by $\omega_f = \frac{\omega}{\sqrt{2}}$ [11]. Here a proportional-integral (PI) controller is employed as a loop filter, whose parameters k_i and k_p stand for integral and proportional gain, respectively. The value of k_p and k_i can be selected from $\omega_{PLL} = \sqrt{U_+ k_i}$ and $\xi_{PLL} = \frac{k_p}{2} \sqrt{\frac{U_+}{k_i}}$, where ω_{PLL} is control loop frequency and ξ_{PLL} is damping constant [8],[10], [20].

IV. DOUBLE SECOND-ORDER GENERALIZED INTEGRATOR FREQUENCY LOCKED LOOP (DSOGI-FLL)

In SOGI-FLL system, SOGI is tuned with the help of FLL and simultaneously perform as a grid synchronization algorithm [12]. Generalized integrator (GI), recognized as

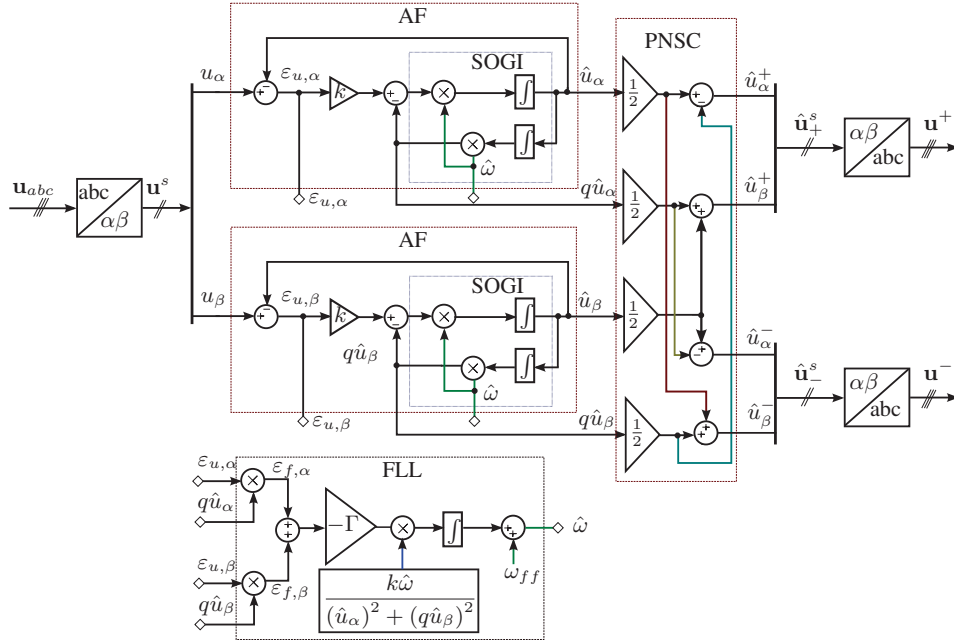


Fig. 4. Structure of DSOGI-FLL.

quadrature signals generator (QSG) can produce two signal perpendicular to each other [13][14]. These signals are utilized to measure the positive- and negative-sequence values of three phase voltage. DSOGI-FLL block diagram is depicted in Fig. 4. Lyon's transformation is utilized to form this novel three-phase grid synchronization method in synchronous coordinates ($\alpha\beta$) [18]. Here in Fig. 4, for the calculation of symmetrical component α and β , two SOGIs are used. Lyon's transformation is executed by Positive/Negative Sequence Calculation (PNSC) block [12]. The choice of the control parameters k and γ determine the precise response by SOGI-FLL. $k = \sqrt{2}$ is chosen for this paper and a justification for this value is given in [12]. γ is calculated as shown in (8) by means of feedback variables [17].

$$\gamma = -\frac{k\hat{\omega}}{U^2}\Gamma \quad (8)$$

The settling time $t_{\text{FLL}} = \frac{5}{\Gamma}$ of FLL solely rely on Γ [15]. $\Gamma = 50$ is chosen for this paper and a justification for this value is given in [12].

V. JUSTIFICATION OF DDSRF-PLL AND DSOGI-FLL METHOD UNDER ADVERSE GRID FAULT CONDITION

Above mentioned two methods will be examined in this section in various adverse grid fault situations. Both methods are modeled in Matlab Simulink environment and simulated under identical input voltage for the justification. The values for the PI controller parameters for the simulation below are $k_p = 2.22$ and $k_i = 246.7$, justification given in [12]; gives $\xi_{\text{PLL}} \approx 1/\sqrt{2}$ and reduced control loop bandwidth $\omega_{\text{PLL}} \approx 2\pi 25$ rad/s. The parameters $k = \sqrt{2}$ and $\Gamma = 50$ is chosen for DSOGI-FLL and identical for all the simulation. During unbalanced voltage conditions, both DDSRF-PLL and

DSOGI-FLL cancels the 2ω oscillation and perfectly detect the value of $\mathbf{u}_{abc}^+$ and $\mathbf{u}_{abc}^-$ of the input three-phase voltage [9].

A. Performance During Harmonic-polluted Situation

The input voltage is characterized with an unbalanced fifth harmonic component as: $U_+ = 100$ V, $U_- = 30$ V, $U_{+5} = 10$ V, $U_{-5} = 10$ V, $\omega = 50$ Hz and the prefault grid voltage is $U_{\text{pf}} = 100$ V. Fig. 5(a) and Fig. 5(b) illustrate DDSRF-PLL and DSOGI-FLL output graph, respectively. DDSRF-PLL can correctly track out the phase angle $\hat{\theta}$. It is reasonable because of the low-pass nature of (7) and the values are given for the simulation are $\omega_{\text{PLL}} \approx 2\pi 25$ rad/s and $\xi_{\text{PLL}} \approx 1/\sqrt{2}$. But for both methods, $\hat{\omega}$ and $\mathbf{u}_{abc}^-$ detection are disrupted due to the harmonic pollution. Meanwhile, adaptive filtering properties of DSOGI-FLL helped to detect $\hat{\omega}$ without serious oscillation. It can be decided that DSOGI-FLL exhibits reliable achievement for $\hat{\omega}$ detection and DDSRF-PLL performance is not acceptable during the harmonic-polluted situation.

B. Justification during Phase Angle jump and Voltage Dips

Both methods are examined for all above-mentioned category of voltage dips (A-G) with $D = 0.5 \angle 0^\circ$. Both methods output are nearly identical for all category voltage dips and both can correctly identify the voltage components of the unbalanced systems except type C [9]. For both algorithms, voltage dip type C is applied as input voltage having $D = 0.5 \angle -30^\circ$, $U_+ = 73 \angle -9.9^\circ$ V, $U_- = 25 \angle 23.8^\circ$ V, $\omega = 50$ Hz and $U_{\text{pf}} = 100$ V. Output of DDSRF-PLL and DSOGI-FLL is shown in Fig. 6(a) and 6(b), respectively. From figures, we can say that both algorithms depict sluggish dynamics and oscillation during the estimation of the positive- and negative-sequence voltage. Meanwhile, DDSRF-PLL produces high

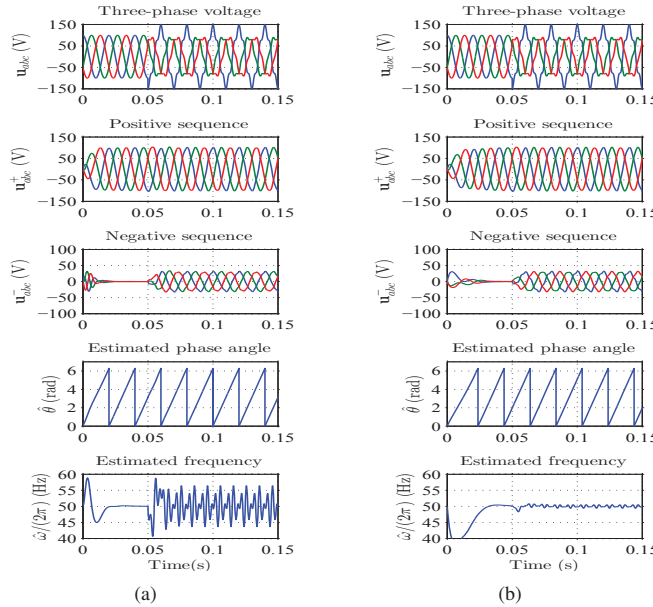


Fig. 5. (a) Output of DDSRF-PLL and (b) DSOGI-FLL during harmonic-polluted situation.

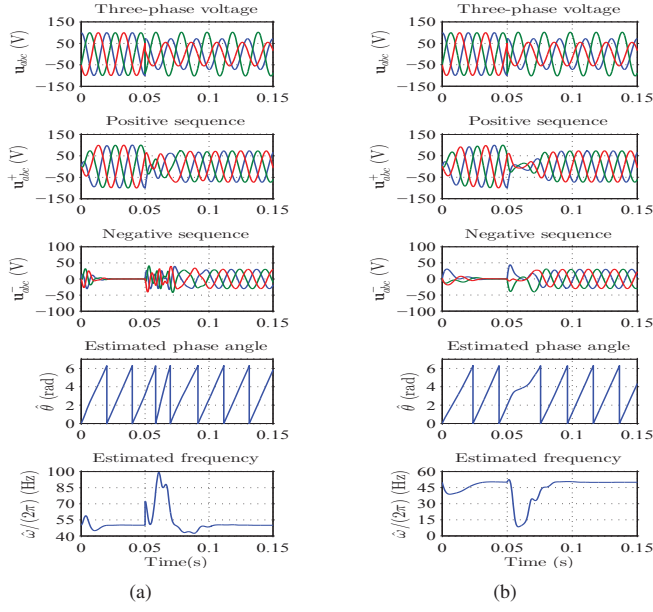


Fig. 6. Output of (a) DDSRF-PLL and (b) DSOGI-FLL during dip type C.

overshoot for the detection of $\hat{\omega}$ and as a consequence, it also unable to correctly detect $\hat{\theta}$. Therefore DSOGI-FLL displays more sharpened performance than DDSRF-PLL during sudden phase angle jump.

C. DDSRF-PLL having Control Loop Frequency 50 Hz

In preceding assessments, $\omega_{\text{PLL}} \approx 2\pi 25$ rad/s is chosen for DDSRF-PLL for experiments. But modern grid codes says that ω must be kept between 47.5 – 52 Hz and at most

can reach 46.5 Hz less and 53.5 Hz highest [23]. Therefore, DDSRF-PLL frequency must be also in between these limits. So $k_p = 4.44$ and $k_i = 986.96$ of PLL are changed to give $\omega_{\text{PLL}} \approx 2\pi 50$ rad/s. Now, to examine the performance of DDSRF-PLL with $\omega_{\text{PLL}} \approx 2\pi 50$ rad/s, which is same as grid frequency $\omega = 2\pi 50$ rad/s, above-mentioned experiments is simulated with these new parameters. DDSRF-PLL exhibits nearly identical results for the detection of $\mathbf{u}_{abc}^+$, $\mathbf{u}_{abc}^-$ and $\hat{\theta}$. However for the detection of $\hat{\omega}$, it gives high overshoot

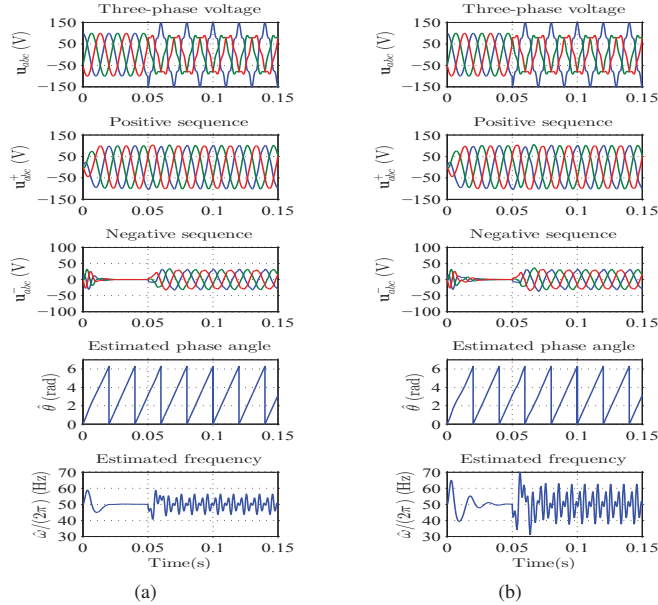


Fig. 7. Output of DDSRF-PLL during harmonic-polluted situation having (a) $\omega_{\text{PLL}} \approx 2\pi 25$ and (b) $\omega_{\text{PLL}} \approx 2\pi 50$.

when $\omega_{PLL} \approx 2\pi 50$ rad/s than $\omega_{PLL} \approx 2\pi 25$ rad/s. An unbalanced harmonic polluted voltage is selected for justification as earlier for illustration purpose. DDSRF-PLL output having $\omega_{PLL} \approx 2\pi 25$ rad/s and $\omega_{PLL} \approx 2\pi 50$ rad/s is given in Fig. 7(a) and 7(b), respectively. Since $\omega_{PLL} \approx 2\pi 50$ in Fig. 7(b), higher overshoot is clear enough than in Fig. 7(a).

VI. CONCLUSION

In this particular occasion for both algorithms, numerous points can be ameliorated, specifically events like phase angle jump and harmonic-polluted situation. Contrariwise, both algorithm satisfies the modern grid codes by exhibiting accurate FRT function and sound dynamic response during various faults. A new algorithm decoupled multiple SRF-PLL (DMSRF-PLL) based on SRF-PLL, might exhibit accurate performance throughout any harmonic polluted situation. DMSRF-PLL concept is identical as DDSRF-PLL. Decoupling of harmonics produce oscillations greater than 2ω , but DDSRF-PLL doesn't have the ability to remove it. Therefore, many decoupling cells will be used in DMSRF-PLL [16], which will remove oscillations greater than 2ω . As a consequence, $\omega_{PLL} \approx 2\pi 50$ rad/s can be used. All these criteria will fulfill the modern grid codes.

REFERENCES

- [1] Hossain, E., Kabalci, E., Bayindir, R. and Perez, R. *Microgrid testbeds around the world: State of art*, Energy Conversion and Management 86 (2014) 132-153.
- [2] Hossain, E., Kabalci, E., Bayindir, R. and Perez, R. *A Comprehensive Study on Microgrid Technology*, International Journal of Renewable Energy Research, Vol.4, No.4, 2014 1094-1107.
- [3] Kabalci, E., Hossain, E. and Bayindir, R. *Microgrid test-bed design with renewable energy sources*, International Conference on Power Electronics and Motion Control Conference and Exposition (PEMC), 2014, Pg. No. 907-911.
- [4] Kabalci, E., Bayindir, R. and Hossain, E. *Hybrid microgrid testbed involving wind/solar/fuel cell plants: A desing and analysis testbed*, Renewable Energy Research and Application (ICRERA), 2014 International Conference on, Milwaukee, WI, 2014, pp. 880-885.doi: 10.1109/ICRERA.2014.7016511.
- [5] Bayindir, R. and Hossain, E., Kabalci, E. and Billah, K. *Investigation on North American Microgrid Facility*, International Journal of Renewable Energy Research, Vol.5, No.2, 2015 558-574.
- [6] Teodorescu, R., Liserre, M. and Rodriguez, P. *Grid Converters for Photovoltaic and Wind Power Systems*, John Wiley and Sons, Inc., UK, 2011. ISBN 9780470057513.
- [7] Blaabjerg, F., Teodorescu, R., Liserre, M., Timbus, A. *Overview of control and grid synchronization for distributed power generation systems*, IEEE Transactions on Industrial Electronics, vol. 53, no. 5, pp 1398-1409, October 2006.
- [8] Chung, S. *A phase Tracking System for Three phase Utility Interface Inverters*, IEEE Transactions on Power Electronics, vol. 53, no. 3, May 2000.
- [9] Fuad, K. *Grid-voltage Synchronization Algorithms Based on Phase-locked Loop and Frequency-locked Loop for Power Converters*, Aalto University publication, Finland, June 2015.
- [10] Rodriguez, P., Sainz, L., Bergas, J. *Synchronous Double Reference Frame PLL applied to a Unified Power Quality Conditioner*, 10th International Conference on Harmonics and Quality of Power, vol.02, pp 614-619, October 2002.
- [11] Rodriguez, P., Pou, J., Bergas, J., Candela, J., Burgos, R. and Boroyevich, D. *Decoupled double synchronous reference frame PLL for power converters control*, IEEE Transactions on Power Electronics, vol. 22, no. 2, pp. 584-592, 2007.
- [12] Rodriguez, P., Luna, A., Munoz-Aguilar, A., Etxeberria-Otadui, I., Teodorescu, R. and Blaabjerg, F. *A Stationary Reference Frame Grid Synchronization System for Three-phase Grid-connected Power Converters Under Adverse Grid Conditions*, IEEE Transactions on Power Electronics, vol. 27, no. 1, January 2012.
- [13] Yuan, X., Merk, W., Stemmler, H., Allmeling, J. *Stationary-Frame Generalized Integrators for Current Control of Active Power Filters With Zero Steady-State Error for Current Harmonics of Concern Under Unbalanced and Distorted Operating Conditions*, IEEE Industry Applications Conference, 2000 , vol. 4, pp. 2143-2150, October 2000.
- [14] Teodorescu, R., Blaabjerg, F., Liserre, M., Loh, P.C. *Proportional-resonant controllers and filters for grid-connected voltage-source converters*, IEE Proceedings on Electric Power Applications, vol. 153, pp. 750-762, September 2006.
- [15] Franklin, G., Powell, J., Emami-Naeini, A. *Feedback control of Dynamic Systems*, Englewood Cliffs, NJ: Prentice-Hall, 2009.
- [16] Hadjidemetriou, L., Kyriakides, E. and Blaabjerg, F. *A New hybrid PLL for Interconnecting Renewable energy Systems to the Grid*, IEEE Transactions on Industry Applications, vol. 49, no. 6, November/December 2013.
- [17] Rodriguez, P., Luna, A., Candela, I., Teodorescu, R. and Blaabjerg, F. *Grid Synchronization of Power Converters using Multiple Second Order Generalized Integrators*, in Proc. 34th Annu. Conf. IEEE Ind. Electron., pp. 755-760, Nov. 10-13, 2008.
- [18] Lyon, W. *Application of the Method of Symmetrical Components*, New York: McGraw-Hill, 1937.
- [19] Harnefors, L. *Control of Variable-speed Drives*, Applied Signal Processing and Control, Department of Electronics, Mälardalen University, Sweden, 2003.
- [20] Gardner, F. *Phaselock Techniques*, John Wiley and Sons, Inc., USA, 1979. ISBN 0-471-04294-3.
- [21] Bollen, M. and Gu, I. *Signal Processing for Power Quality Disturbance*, Hoboken, NJ, USA: Wiley-IEEE Press, 2006, ch. 8.
- [22] Bollen, M. *Understanding Power Quality Problems: Voltage Sags and Interruptions (Power Engineering Series)*, Piscataway, NJ: IEEE Press, 2000.
- [23] Altin, M., Gksu, O., Teodorescu, R., Rodriguez, P., Jensen, B.-B. and Helle, L. *Overview of Recent Grid Codes for Wind Power Integration*, 12th International Conference on Optimization of Electrical and Electronic Equipment (OPTIM), 2010 , pp. 1152-1160, May. 20-22, 2010.

Class-Based Service Connectivity using Adaptive Guard Channel Reservation in Wireless Networks

Khan Tanjama Mustarin and Mostafa Zaman Chowdhury

Dept. of Electrical and Electronic Engineering, Khulna University of Engineering & Technology, Bangladesh.

E-mail: tanjama.mustarin@yahoo.com, mzceee@yahoo.com

Abstract: To reach the satisfied level of quality of service (QoS) for the maximum number of users, maximum bandwidth utilization is not the only key issue to be considered, some other factors like call dropping probability, call blocking probability, and overall connection blocking rate etc. are also needed to keep under deliberation. To accommodate large number of calls and to increase the utilization of bandwidth the proposed (*Adaptive-Guard-Band model*) is introduced. This proposed model is able to provide guard band for different classes of users as per as requirement with handover calls as highest priority. In this paper, we propose an adaptive guard band scheme which shows that the guard band allocations are dynamic in nature and it becomes changing as priority basis giving first priority to the handover calls. The proposed scheme considers call type, existing occupied channels, and average call life time to calculate the guard band reservation for different type of calls. The numerical and analytical results show that the proposed scheme provides negligible handover call dropping probability (HCDP) and satisfactory range of new call blocking probability (NCBP) as well as attain a significant increase in bandwidth utilization (BWU) while overall connection blocking rate (OCB) stays within reasonable range.

Keywords: Adaptive guard band, traffic class, QoS, call dropping probability, call blocking probability, and bandwidth utilization.

I. INTRODUCTION

The requirement of quality of service (QoS) for all traffic types not very firm. Hence more calls of higher priority can be accommodated by allocating few more reserved bandwidths from bandwidth adaptive calls. Therefore, the adaptive guard band allocation improves the overall forced call termination probability as well as provide priority of the traffic classes in terms of handover call dropping probability (HCDP) and new call blocking probability (NCBP) of higher priority classes without making significant reduction in bandwidth utilization.

A number of researches have already been done to grant higher priority for handover calls over new calls [1], [2]. By making a reservation for higher priority calls guard band based schemes can easily give priority to these calls. These guard band schemes not only can reduce the NCBP for the higher priority calls but also result in huge reduction of bandwidth utilization. To reduce the call blocking probability the flexible QoS measurement has been proposed by several researchers [1], [3]. To prioritize certain traffic classes over others an adaptive

resource allocation scheme is proposed by D. D. Vergados et al. [1]. More feasible and flexible guaranteed QoS proved through the QoS adaption techniques [4], [5] than the guard channel schemes [2]. QoS degradation of only low priority traffic classes to accept comparatively higher priority traffic call requests is mainly appeared on their proposed schemes. An adaptive QoS scheme which reduces the QoS levels of calls that carry more flexible and efficient in guaranteeing adaptive traffic to accept the handover call requests only is proposed by W. Zhuang et al. [2]. This is mainly known as handover-queuing scheme to reduce HCDP. The fixed or dynamically adjustable guard bands can improve the probability of successful handovers by reserving a number of channels exclusively for handovers [6]. Queuing is a way of delaying handover [7]. Because of complex nature, queuing is not used for real time services as well as not so popular for non-real time services.

Previously, in most of the research publications related to adaptive guard band consider a fixed amount of guard band for the ease of calculation and assumptions. Few research outcomes relating variable guard bands but as they consider call arrival rate as main parameter to calculate the amount of reserve bandwidth that make those methods complex. Here in this paper, mainly the existing number of call, average call life time, and call type/class are considered to calculate the variable amount of reserve bandwidth for different type/class of calls. It is very simple to measure all these parameters that make the simplicity of the proposed scheme. Additionally a multiplication factor is introduced to make trade-off between the bandwidth utilization and call blocking/dropping probabilities. Therefore, the scheme can reduce overall connection blocking rate significantly as well as NCBP for the higher priority classes. Also handover priority provides negligible handover call dropping probability, and adaptive guard band increases the number of call admission in the system. In this paper, our scheme is compared with different other schemes, mainly with non-priority and fixed priority schemes.

The rest of the paper is organized as follows. Section II describes the system model of the proposed adaptive guard channel allocation, performance analysis of the proposed scheme is shown in Section III. Finally, we concluded the work in Section IV.

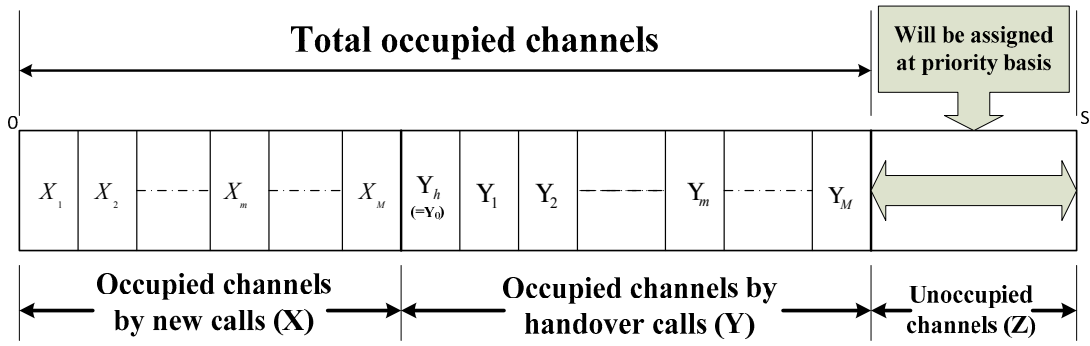


Fig 1. Channel allocation for proposed model

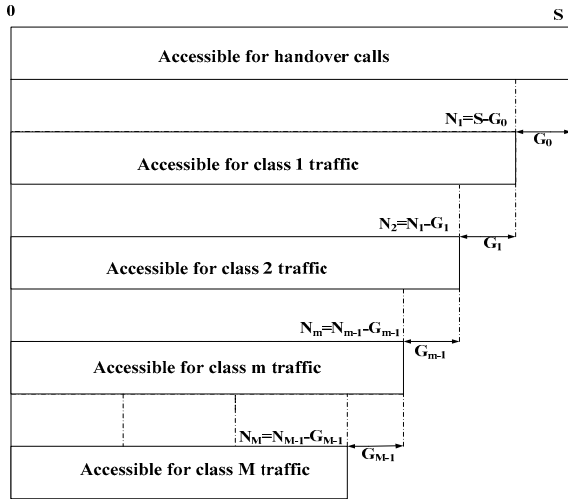


Fig 2. Proposed multi-level adaptive guard band model

II. THE PROPOSED ADAPTIVE GUARD CHANNEL RESERVATION MODEL

A new scheme of adaptive guard channel is proposed here allowing a calculated amount of guard band by considering the average call life time and existing no of calls in every class of traffic which provides highest priority to the handover calls. As fixed guard channel allocation causes waste of bandwidth utilization, so the proposed scheme derives the dynamic guard channel allocation to increase bandwidth utilization with a considering range of HCDP and NCBP of higher traffic classes. In the proposed scheme, we provide a new idea of bandwidth allocation considering negligible handover call dropping probability.

Fig. 1 shows the scheme for maximum S channels. Where X is the number of occupied channels for new calls, Y is the number of channels already occupied by handover calls, also an area of unoccupied channel Z shown in the figure which is used to assign for different traffic considering the existing calls of the traffic classes and with the introduced multiplication factor. This is done for avoiding any kind of handover call dropping

probability. That means handover stays at first priority, than class 1 traffic calls get the 2nd priority and serially the last priority is given to M^{th} class traffic calls.

Fig. 2 shows that, in every class there will be some allocated bandwidth for handover calls. Which is known as guard band for handover calls. But instead of making the guard fixed it will be dynamic for better bandwidth utilization. The accessible number of channels by a call of m^{th} class is denoted by N_m . The guard bands for handover (G_h) and new calls of different traffic classes are calculated using the following equations. Here, K denotes the introduced multiplication factor, X is the occupied channel and $1/\mu$ is the channel holding time. Channel holding time has direct relation with cell dwell time and call life time.

$$G_0 = KY_h \mu_h \quad (1)$$

$$G_m = KX_m \mu_m \quad (2)$$

$$G_M = KX_M \mu_M \quad (3)$$

$$X + Y + Z = S \quad (4)$$

where $m=1, 2, \dots, M$. μ_m is the average call release rate for new calls of traffic class m . μ_h is the average call release rate for existing handover calls. The multiplication factor K is used to make trade-off between the bandwidth utilization and call blocking/dropping probabilities.

Considering the existing number of calls and average call life time of different classes of traffic calls in the proposed scheme, it becomes simple to determine the reserve bandwidth for a particular type of call.

The HCDP (P_D) is calculated using (6). The new NCBP for the traffic class m is calculated using (7). In equation, λ_T denotes total call arrival rate. The average call release rate is denoted by μ_i for the state i of the system. N_M is the maximum number of available states for m^{th} priority class.

In (3), $P_{B,0} = P_D$

$$N_M = S - \sum_{i=0}^{M-1} G_i \quad (5)$$

$$P_D = P_S = P_0 \frac{\lambda_T^{N_M}}{(N_0)! \mu^N \sum_{i=N+1}^{N_0} \mu_i} \sum_{K=1}^M (\lambda_0 + \lambda_1 + \lambda_2 + \dots + \lambda_{M-K})^{N_{M-K} - N_{M-K+1}} \quad (6)$$

$$P_{B,m} = \sum_{i=N_m}^{N_0} P_i$$

$$= P_{B,m-1} + P_0 \sum_{i=N_m}^{N_{m-1}-1} \frac{\lambda_T^{N_M}}{i! (\mu_i)^i} (\lambda_0 + \lambda_1 + \lambda_2 + \dots + \lambda_{m-1})^{i-N_m} \sum_{k=m}^M (\lambda_0 + \lambda_1 + \lambda_2 + \dots + \lambda_k)^{N_k - N_{k+1}} \quad (7)$$

$$P_0 = \left[1 + \sum_{i=1}^{N_M} \frac{\lambda_T^i}{\mu^i i!} + \sum_{j=1}^M \sum_{i=N_j+1}^{N_{j-1}} \left\{ \frac{\lambda_T^{N_M} (\lambda_0 + \lambda_1 + \lambda_2 + \dots + \lambda_{j-1})^{i-N_j}}{i! (\mu_i)^i} \sum_{k=j}^M (\lambda_0 + \lambda_1 + \lambda_2 + \dots + \lambda_k)^{N_k - N_{k+1}} \right\} \right]^{-1} \quad (8)$$

III. PERFORMANCE ANALYSIS

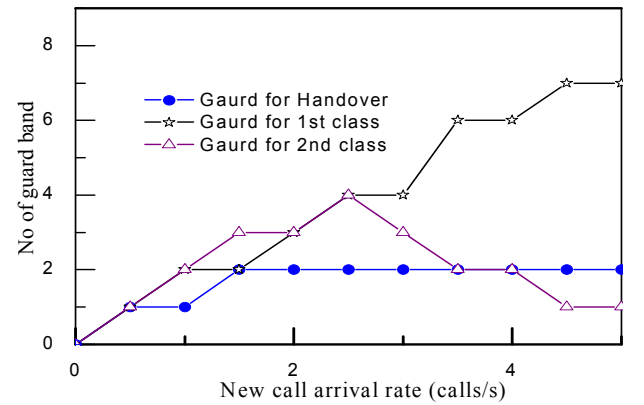
In this section, we present the numerical results of the analysis for the proposed scheme. We compared the performance of our proposed scheme with several other schemes (Fixed priority scheme and non-priority scheme). Here three classes of traffic are considered. The requested call sequence is assumed to be random here. Basic assumptions considered for the analysis are shown in Table 2. Here the guard bands are calculated considering the multiplication factor; $K=6$ and $K=3$. According to our proposed scheme in heavy traffic conditions, when there have no sufficient resources to accommodate more calls, calculated reserved guard bands used to admit the traffic in priority basis. Handover calls are given highest priority over any new class of traffic.

Table 2. Basic assumptions of parameters

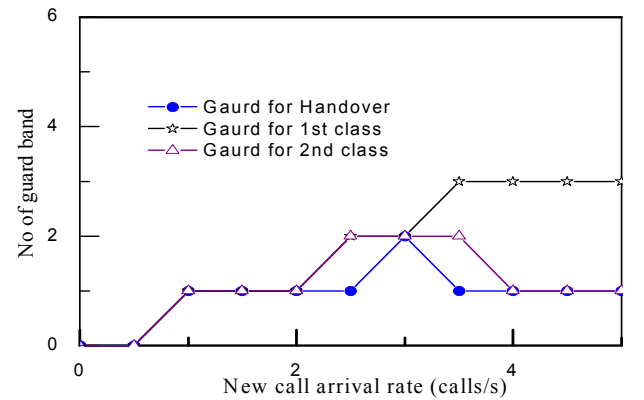
Basic assumptions for the traffic classes	
Traffic class (m)	Average call life time (sec)
Class 1	100
Class 2	80
Class 3	120
Basic assumptions for the traffic environment	
Parameter	Value
Total no of channels	120
Channel dwell time (1/η)	160s
Traffic of class 1: class 2: class : 3	0.25:0.25:0.5
Fixed guard band for each priority class	5%

Fig. 3 shows a comparison of guard band allocation for different classes of traffic calls. It mainly presents that for handover it keeps quite a constant guard band which proves that proposed scheme gives highest priority to handover calls. Next

the fluctuation in guard band allocations for 1st and 2nd classes represent that our proposed scheme works in reduction of wasting bandwidth and allocate it as per as requirement.

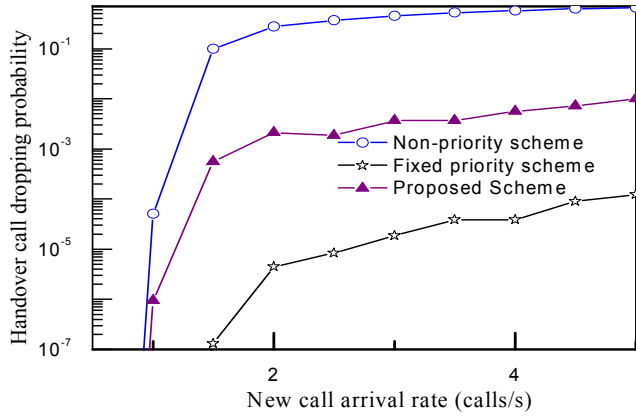


(a) for $K=6$

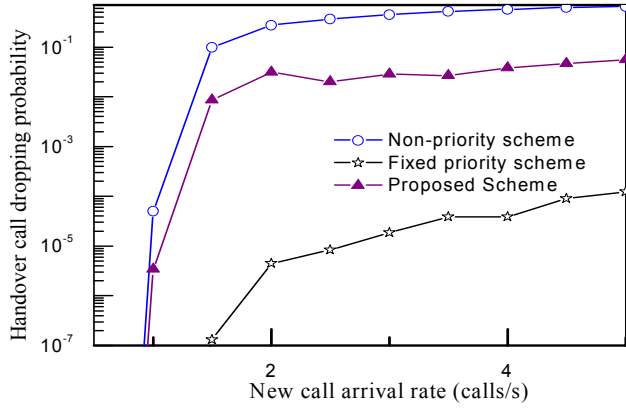


(b) for $K=3$

Fig 3. Variation in guard band allocation



(a) for $K=6$

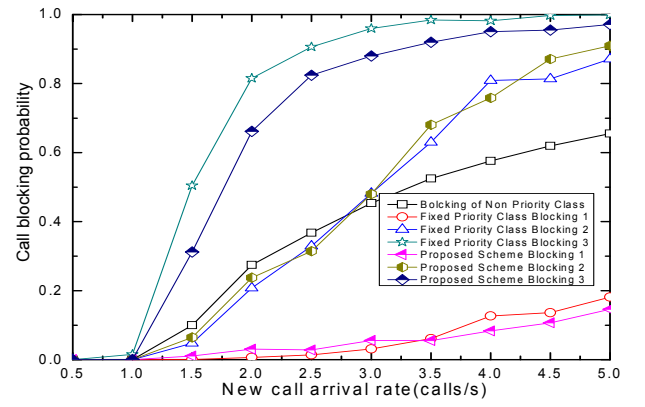


(b) for $K=3$

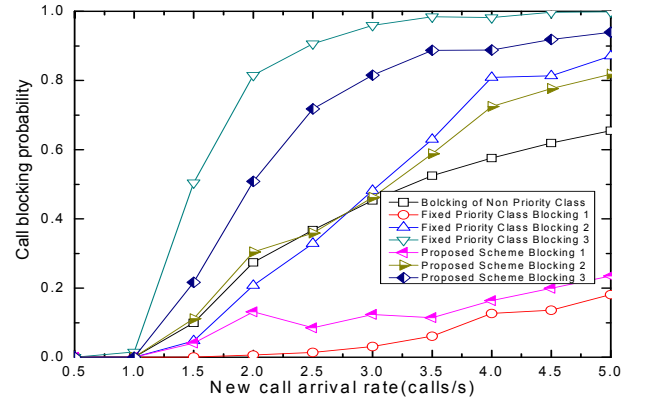
Fig 4. A comparison of handover call dropping probability

Fig. 4 shows that the proposed handover prioritization scheme offers negligible HCDP even for highly congested traffic condition. Generally dropping probability should stay within a range of 0.001 which is satisfied by the proposed scheme for both of the multiplication factors ($K=6$ and $K=3$). In the proposed scheme there occurs variation of performance between minimum handover dropping, NCBP, and higher bandwidth utilization. This trade-off condition can be minimized by varying the introduced multiplication factor K .

Fig. 5 shows that the proposed scheme provides significantly reduced NCBP for the higher priority traffic calls during the congested traffic condition. The non-priority scheme creates very high HCDP and NCBP for the higher priority traffic calls. The provision of maximum level of adaptive guard band without the priority of calls cannot offer a satisfactory level of HCDP and NCBP of the higher priority traffic calls.



(a) for $K=6$

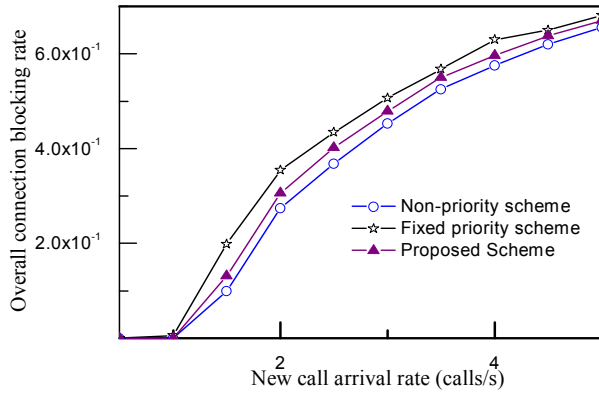


(b) for $K=3$

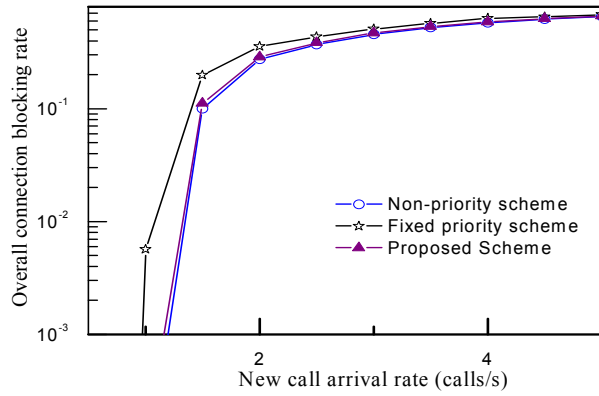
Fig 5. A comparison of new call blocking probability

Fig. 6 shows that the proposed scheme provides reduced forced call termination probability compared to the fixed guard band scheme for both of the multiplication factors ($K=6$ and $K=3$) and also provides almost equal forced call termination probability compared to the non-priority scheme. Therefore, the priorities for handover and other calls do not increase the overall forced call termination probability significantly. Whereas the fixed priority scheme results in very high overall forced call termination probability with respect to non-priority scheme.

Fig. 7 shows that the proposed scheme gives much higher bandwidth utilization with respect to fixed prioritization scheme. It gives nearer values to non-prioritized scheme. For multiplication factor ($K=3$) gives higher bandwidth utilization than ($K=6$). Therefore, our proposed scheme gives much better bandwidth utilization than other schemes considering other factors. Maximum bandwidth utilization is mandatory for fruitful communication. As non-priority scheme always gives highest bandwidth utilization so it is desired to get closer to it considering prioritization.

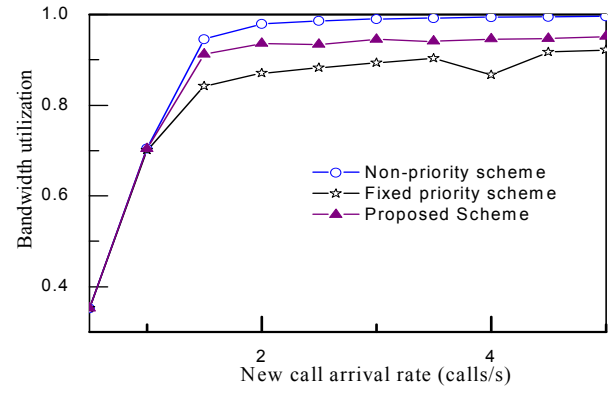


(a) for K=6

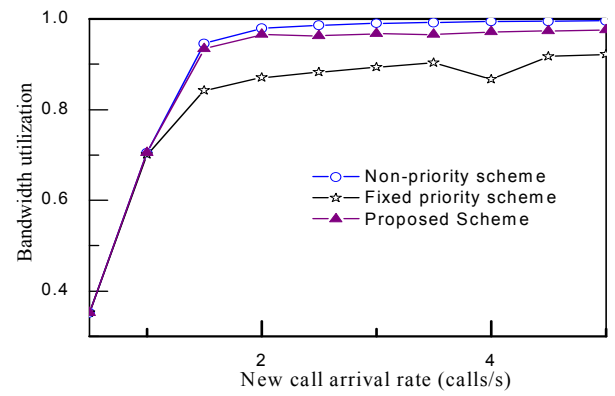


(b) for K=3

Fig 6. A comparison of overall forced call termination probability



(a) for K=6



(b) for K=3

Fig 7. A comparison of bandwidth allocation

IV. CONCLUSIONS

Here, in this paper we proposed a new adaptive guard band scheme for priority based call admission in wireless networks. To give highest priority to handover calls the proposed scheme does not stand behind to control other factors. Which implies that, the overall forced call termination probability is not increased significantly. Also using the scheme, the HCDP and NCBP reduces in a good manner.

As the proposed scheme is based on considering existing calls to allocate adaptive guard band for traffic classes, it becomes easy and simpler. Also by using different multiplication factors trade-off situation between maximum BWU, minimum HCDP, and OCB can easily be obtained.

REFERENCES

- [1] D. D. Vergados, "Simulation and Modeling Bandwidth Control in Wireless Healthcare Information Systems," *Simulation*, vol. 83, no. 4, pp. 347–364, November 2007.
- [2] W. Zhuang, B. Bensaou, and K. C. Chuas, "Adaptive Quality of Service Handover Priority Scheme for Mobile Multimedia Networks," *IEEE Transaction on Vehicular Technology*, vol. 49, no. 2, pp. 494–505, March 2000.
- [3] F. A. Cruz-Perez and L. Ortigoza-Guerrero, "Flexible Resource Allocation Strategies for Class-Based QoS Provisioning in Mobile Networks," *IEEE Transaction on Vehicular Technology*, vol. 53, no. 3, pp. 805–819, May 2004.
- [4] I. Habib, M. Sherif, M. Naghshineh, and P. Kermani, "An Adaptive Quality of Service Channel Borrowing Algorithm for Cellular Networks," *International Journal of Communication System*, vol. 16, no. 8, pp. 759–777, October 2003.
- [5] N. D. Tripathi, J. H. Reed, and H. F. Vanlandingham, "Handover in Cellular Systems," *IEEE Personal Communication*, vol. 5, no. 6, pp. 26–37, December 1998.
- [6] H. Chen, C. C. Cheng and H. H. Yeh, "Guard-Channel-Based Incremental and Dynamic Optimization on Call Admission Control for Next-Generation QoS-Aware Heterogeneous Systems," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 5, pp. 3064–3082, September 2008.
- [7] R. Sindal and S. Tokekar, "Analysis of Queuing Based Call Admission Control Scheme in CDMA Cellular Network for Variation in Mobility and Soft Handoff Threshold," In *Proceeding of International Conference on Industrial and Information Systems*, 2008, December 2008, pp. 1–6.
- [8] M. Z. Chowdhury and Y. M. Jang, "Class-based Service Connectivity using Multi-level Bandwidth Adaptation in Multimedia Wireless Networks," *Wireless Personal Communications*, vol. 77, no.4, pp. 2735–2745, February 2014.
- [9] M. Z. Chowdhury, Y. M. Jang and Z. J. Haas, "Call Admission Control Based on Adaptive Bandwidth Allocation for Wireless Networks," *Journal of Communications and Networks*, vol. 15, no.1, pp.15–24, May 2015.
- [10] H. Halabian, P. Rengaraju, C. H. Lung and I. Lambadaris, "A Reservation-Based Call Admission Control Scheme and System Modelling in 4G Vehicular Networks," *EURASIP Journal on Wireless Communication and Networking*, May 2015.

Zebra-crossing Detection Based on Geometric Feature and Vertical Vanishing Point

Md. Khaliluzzaman and Kaushik Deb*
Dept. of Computer Science and Engineering
Chittagong University of Engineering & Technology
Chittagong-4349, Bangladesh
khalil_021@yahoo.co.in, *debkaushik99@cuet.ac.bd

Abstract – Zebra-crossing region detection from a zebra-crossing image is a challenging activity to support visually impaired navigating safely in outdoor environments. This paper proposes a zebra-crossing detection framework based on unique geometrical features of zebra-crossing. The unique geometrical feature is zebra-crossing stripe's edges are arranged in sorted order. Initially, pre-process the zebra-crossing images by applying the directional Gabor filter to eliminate the noisy and shadow's effect and improve the resolution of the zebra-crossing horizontal edges. Then Sobel's operator is used to detect the horizontal edges. Potential horizontal edges are extracted by eliminating small and non-candidate edges. After that, the edge linking process is utilized to extract the longest horizontal edges. Extracted zebra-crossing edges are justified by arranging the horizontal edge segments in increasing parallel order. Finally, these parallel horizontal edges are used to measure the vanishing point to verify zebra-crossing edges from other analogous visual pattern and confirm the detection of zebra-crossing candidate region. Various zebra-crossing images are utilized to evaluate the proposed framework and presented outcomes demonstrate the adequacy.

Keywords - Visually impaired; zebra-crossing; geometrical property; Gabor filter; vanishing point

I. INTRODUCTION

Accurate object detection from a particular image is the challenging task in the area of the computer vision. Zebra-crossing region detection from a zebra-crossing image is such a task. This activity is used for visually impaired and autonomous systems in an unknown environment. In the real time system, the zebra-crossing region detection is harder for zebra-crossing shape, viewpoints, and different outdoor illumination conditions. To mitigate these problems and satisfy the requirement of intelligent autonomous systems the detection procedure should be performed as quickly as possible and extract every interest candidate object from the zebra-crossing image.

This section provides the descriptive summary of some procedures that are implemented as well as tested for zebra-crossing detection from zebra-crossing images. This topic has gained importance for the purpose of automatic navigation of autonomous systems. All the research works that have been done previously focused on the feature of the zebra-crossing is that a zebra-crossing edge image contains some long horizontal edges. As far as detection of the zebra-crossing region is concerned, researchers have found many methods to detect zebra-crossing region from zebra-crossing image. For

example, in [1], a method is described for detecting and tracing of pedestrians automatically with the camera setups, calibration, and image processing algorithms. An RGB-D based method is proposed in [2]. This method detects and recognizes the pedestrian crosswalks. For that, parallel concurrent lines are extracted by applying the Hough transforms. Finally, pedestrian crosswalks are recognized by using the depth information.

Pedestrian crossing detection approach for urban pedestrian safety system is introduced in [3]. In this system, authors first use the Fourier transformation then applied the Hough transform to detect lines and angles. The augmented bipolarity is utilized to detect the boundary between 'zero' and 'one' that has been extended to a histogram examination. Inverse perspective mapping template matching method is used for feature extraction. Finally, SVM method is used to classify the features and detect the crosswalk. An image processing based algorithm is used for pedestrian crossing detection in [4]. The algorithm starts with detecting bipolar region then perform the morphological operation. After that, features are extracted according to the corresponding relationships between an image and its Radon Transform result. Finally, the crossing is reconstructed.

In [5], an online pedestrian crossing detection system is introduced. This system uses pre-existing traffic-oriented video-sensors at a regular interval. And provide coarse spatial measurement on areas alone a crosswalk. This system also introduces the non-dedicated sensors for analyzing and detecting the pedestrians. Finally, this system presents an evidential oriented data fusion process for exploiting redundant information based on intra and inter-sensor fusion. A computer vision based CROSSWATCH system is introduced in [6]. This system is used with a camera cell phone to provide information about the location and orientation of crosswalks for the visually impaired. A few numbers of images are taken per second by the cell phone automatically. After that, each captured image is analyzed by the system in a fraction of a second. When the system detects the crosswalk, it sounds an audio tune. In [7], an approach is proposed for finding crosswalks for the blind based on grouping geometric features into a coherent structure. A safety based pedestrian detection is introduced in [8]. In [9], a method is proposed with a technique of grouping lines and checking for concurrency using the vanishing point constraint.

This paper has been proposed a framework based on geometrical features of a zebra-crossing. The basic property of every zebra-crossing is, it consists of an alternating pattern of black and white stripes. These stripes are considered as a group of consecutive edges in an edge image. These consecutive horizontal parallel stripes of a zebra-crossing are arranged in sorted order as shown in Fig. 1(a). These edges are parallel in the 3D space also and intersect at a vanishing point when those edges are projected onto the image. This is the key part of this zebra-crossing detection framework. Observing the zebra-crossing from a little distance of the zebra-crossing front side then every zebra-crossing has shown this property according to the human vision.



Fig. 1. Natural and unique geometrical property of zebra-crossing: horizontal edges appear in longest increasing subsequence order.

Since the zebra-crossing images are located in the outdoor environments with different lighting conditions and noisy environments. In order to deal with these situations, directional Gabor filter [11] is employed to preprocess the zebra-crossing image and eliminate the influence of the illuminations. After that, Sobel edge detector method is performed for HE detection. Furthermore, a method is applied to extract the potential horizontal edges by eliminating small and non-candidate edges. Then, an edge linking method is used to extract the longest horizontal parallel zebra-crossing edges. Those horizontal parallel edges are required to arrange them in longest increasing subsequence order. Finally, these parallel horizontal edge segments are used to measure the vanishing point (VP) [11] to confirm the detection of zebra-crossing candidate region.

The paper is organized as follows. Proposed zebra-crossing detection framework is described step by step in Section II. In Section III, the experimental results are explained. Finally, the conclusion is given in Section IV.

II. PROPOSED FRAMEWORK

This section describes the proposed zebra-crossing detection framework in details. The zebra-crossing detection framework is divided into six primary stages, i.e., (1) pre-processing zebra-crossing image with Gabor filter, (2) edge detection, (3) eliminating non-candidate edges, (4) edge linking and tracking, (5) arranging horizontal edges in longest increasing order, and (6) calculating vertical VP and finally, confirming and detecting the zebra-crossing candidate region. Fig. 2 shows the proposed zebra-crossing detection framework.

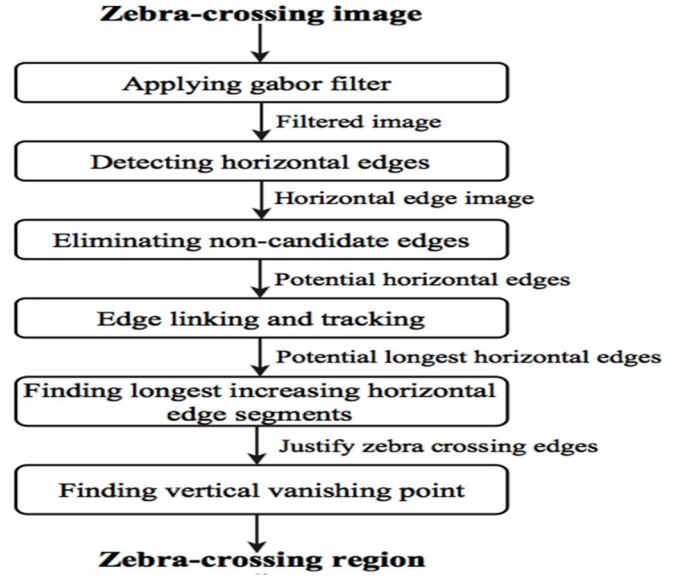


Fig. 2. The proposed zebra-crossing detection framework.

A. Applying Gabor Filter

In order to extract the edges accurately from the zebra-crossing image the noises should be removed from the zebra-crossing image. For that, the Gabor filter is employed on the zebra-crossing image with the convolution method. Before that, the zebra-crossing image is converted into the gray scale image to reduce the computational cost. A gray scale image has only one channel which contains intensity information for each pixel of the image. Then, the Gabor is applied on the zebra-crossing gray scale image to eliminate the influence of illumination and extract the zebra-crossing potential edges

Basically the zebra-crossing image is located in the outdoor environment with different lighting conditions and noisy environments. Gabor filters parameters are appropriately selected to eliminate the illumination varieties from the zebra-crossing image and reserve the multidirectional edge information. The basic form of the Gabor filter is defined in Eq. (1), Eq. (2), and Eq. (3).

$$G_{\lambda,\theta,\phi,\sigma,\gamma}(x,y) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos(2\pi \frac{x'}{\lambda} + \phi) \quad (1)$$

$$x' = x \cos\theta + y \sin\theta \quad (2)$$

$$y' = -x \sin\theta + y \cos\theta \quad (3)$$

here, (x, y) is the coordinate of the image pixel and the cosine factor of the wavelength is presented by λ , the angular orientation is defined by θ , phase offset of the filter is ϕ , the standard deviation and spatial aspect ratio are denoted by σ and γ respectively.

The Gabor filter effect on the zebra-crossing image is shown in Fig. 3(b).

B. Detection Horizontal Edges

Since the horizontal edges are needed to satisfy the geometrical features of the zebra-crossing, the Sobel's operator is used to detect the zebra-crossing edges from the filtered gray scale zebra-crossing image. Fig. 3(c) shows the result of the Sobel's edge detection [12].

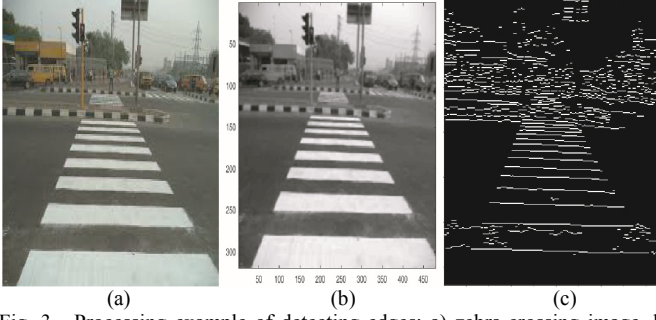


Fig. 3. Processing example of detecting edges: a) zebra-crossing image, b) Gabor filtered image, and c) horizontal edge image.

C. Eliminating Non-candidate Edges

This section extracting the potential HEs that are convey the information of the zebra-crossing by eliminating the non-interest edges. The Sobel edge image mainly contains the horizontal edges. However, many small edges still exist in the edge image. Those small edges are removed from the HE image by employing a filtering procedure. According to this procedure, if any edges are smaller than an *Edge_threshold* (a fixed size of edge) than the edges will be eliminated. Otherwise, the edges will be preserved. The value of the *Edge_threshold* is defined by $\max(\text{length}(i)) / \text{threshold}$, where the threshold is 6 pixels.

Fig. 4(a) shows the result of removing small edges from edge image. This operation removes all discontinuous and noisy edges, and can reduce the processing time for edge linking and tracking. In the HE image, there exist many edges that are not segment of the concurrent horizontal edges. Those edges are marked as non-candidate edges and removed from the edge image. Fig. 4(b) shows the result of removing the non-candidate edges.

D. Linking and tracking

This section mainly contains the potential horizontal stripe edges. However, those potential horizontal edges may have some gaps or breaks in different regions. This gaps or breaks are needed to be filled out to represent the all extracted edges are in the longest form. This is also important to represent the extracted edges are in the increasing sorted form. This paper introduces a method to fill the gaps or breaks in the horizontal edges. According to this method, the gaps of the horizontal edges with less than 4 pixels are filled automatically. Otherwise, the edge linking and tracking method will be employed.

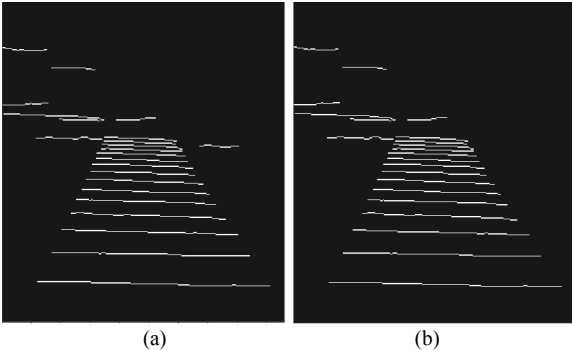


Fig. 4. Processing example of small and non-candidate edge elimination: a) eliminate small edges, and b) eliminate non-candidate edges.

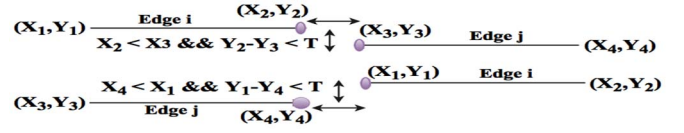


Fig. 5. Edge linking and tracking.

The edge linking and tracking method is shown in Fig. 5. The gap of the two edges is filled if the edges are followed the following conditions. For *edge_i* and *edge_j*, If $ey_r(\text{edge}_i) - ey_l(\text{edge}_j) < T_height$ and $ex_r(\text{edge}_i) < ex_l(\text{edge}_j)$ then merge *edge_i*th and *edge_j*th by combining the edge point $(ex_r(\text{edge}_i), ey_r(\text{edge}_i))$ and $(ex_l(\text{edge}_j), ey_l(\text{edge}_j))$. If $ey_l(\text{edge}_i) - ey_r(\text{edge}_j) < T_height$ and $ex_r(\text{edge}_j) < ex_l(\text{edge}_i)$ then merge the *edge_i*th and *edge_j*th by combining the edge point $(ex_r(\text{edge}_j), ey_r(\text{edge}_j))$ and $(ex_l(\text{edge}_i), ey_l(\text{edge}_i))$. Where, ex_r , ex_l , ey_r and ey_l are *x* and *y* coordinate of right and left end points of *edge_i* and *edge_j* and *T_height* is the threshold. Here, consider the value of the threshold is 3. Fig. 6(a) shows the result of final edge merging through the edge linking procedure. After completing the process of the edge linking the edge image mainly hold the longest parallel horizontal edges only.

E. Arranging horizontal edges in longest increasing order

This section considers the longest zebra-crossing stripe's horizontal edges that found in the previous section. In this stage, an important and unique property of zebra-crossing is utilized; that is, zebra-crossing stripe's edges are arranged in increasing parallel order. If any object edges are found in an increasing sequential order in the edge image, then considered these edges are from zebra-crossing looking object. Any object containing some long horizontal texture may produce long horizontal edges in its edge image. However, these edges will not appear in an increasingly sorted order. After arranging *N* horizontal edges in the current image, there are *L* edges in increasing order where, $N \geq L$. The arranging is done by comparing two edges from bottom to top by their magnitude and some conditions of longest increasing subsequent algorithm. The beginning edge is selected from first two edges comparing with their magnitude. The long magnitude horizontal edge is selected as a first arranging edge from zebra-crossing edge image. The rests of the arranging edges are selected using magnitude and additional conditions. This operation can be solved in $O(n \log n)$ time, where *n* is the number horizontal parallel edge of zebra-crossing. These horizontal edges are not large (< 50).

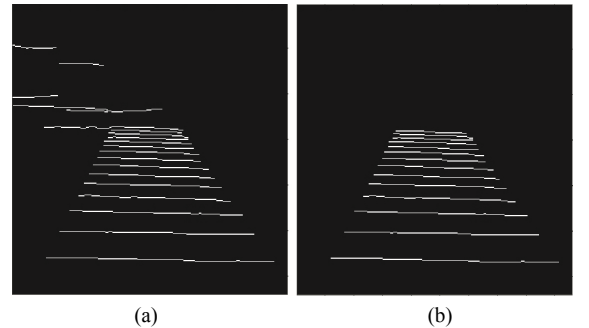


Fig. 6. Processing example of longest increasing subsequence: a) edge linking, and b) HE are arranged in increasing parallel order.

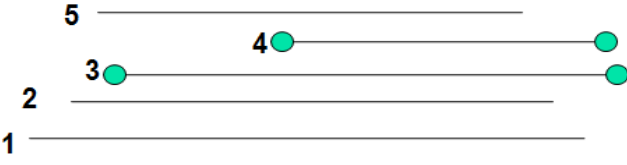


Fig. 7. Arranging horizontal edges in increasing parallel order.

The horizontal edges are arranged in increasing parallel order according to the following conditions. For two edges i and j , $i > j$ if magnitude $i > \text{magnitude } j$ and satisfy the following conditions for edge i and edge j , $L_x(i) < L_x(j)$ and $R_x(i) > R_x(j)$, and $L_y(i) > L_y(j)$ and $R_y(i) > R_y(j)$. Where, $L_x(i)$, $L_x(j)$, $R_x(i)$ and $R_x(j)$ are x coordinate of left and right endpoint of edge i and j . $L_y(i)$, $L_y(j)$, $R_y(i)$ and $R_y(j)$ are y coordinate of left and right endpoint of edge i and j .

In Fig. 7, edge 2 is decreasing edge with respect to edge 1 because it satisfies the above conditions. Edge 3 is increasing edge with respect to edge 1 and 2 because it fails to satisfy the condition 3 and 4. Accordingly edge 4 is decreasing with respect to edge 3 but not decreasing with respect to edge 1 and 2 because it fails at condition 3 and 4. Hence, there are increasing edges are edge 1, 2 and 5. If the edge 4 decreasing with respect to the second last edge that satisfied the above conditions to be arranged edge, i.e. edge 1 then the edge 4 will be considered as an decreasing edge. Fig. 6(b) shows the result of arranging horizontal edges in increasing parallel order.

G. Calculating vertical vanishing point

This section calculates the vertical vanishing point (VP) from the extracted increasing horizontal edges to verify whether the detected horizontal edges are extracted from zebra-crossing or any other similar visual pattern such as stair. The VP of a zebra-crossing can be stated as an imaginary point in a linear perspective drawing in which a group of parallel line are converged. In zebra-crossing, the VP can be defined as a point where two straight lines of the zebra-crossing intersect. These straight lines are passing through the right and left horizontal edges ending points. Basically, the intersection point (x, y) is located inside the zebra-crossing image ($y > 0$). However, zebra-crossing does not have the end side straight line. Hence, the VP of a zebra-crossing can be defined as an intersection point with two straight lines passing through the right and left side end points of the parallel horizontal edges respectively. The procedure of calculating the vanishing point is shown in Fig. 8.

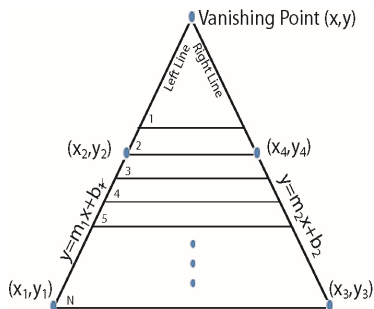


Fig. 8. Calculating vertical VP.

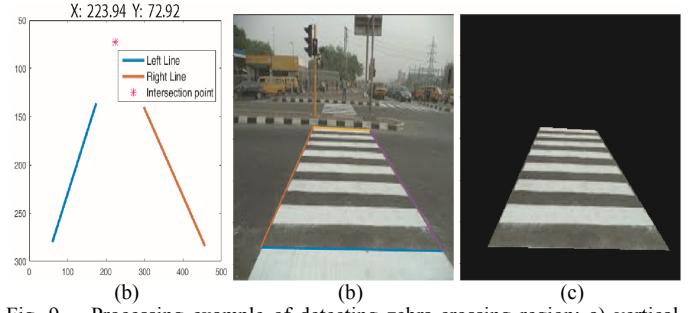


Fig. 9. Processing example of detecting zebra-crossing region: a) vertical vanishing point, b) detected ROI, and c) zebra-crossing candidate region.

According to the Fig. 8, points (x_1, y_1) and (x_2, y_2) are used to drawing the left straight line which is $y = m_1x + b_1$. The value of m_1 is $(y_2 - y_1) / (x_2 - x_1)$. Accordingly, points (x_3, y_3) and (x_4, y_4) are used to draw the right straight line which is $y = m_2x + b_2$. Where, the value of m_2 is $(y_4 - y_3) / (x_4 - x_3)$. The VP is measured from the intersection of these two straight lines that is (x, y) . The VP of the zebra-crossing will be used to confirm the detection of the zebra-crossing region. For that, y coordinate value of the VP is checked because this value for any zebra-crossing will be positive i.e. $y > 0$. And this y coordinate value will be resided to the range of $ZIH \geq YCVP > 0$, here, ZIH is image height, $YCVP$ is zebra-crossing y coordinate of the VP. If the VP of the zebra-crossing resides in the range of $ZIH \geq YCVP > 0$, then the extracted parallel horizontal edges are indicated as a zebra-crossing. Otherwise, those edges may be indicated as a zebra-crossing similar looking object such as stair. That is because the stair regions have the similar feature as like zebra-crossing. However, the VP of the stair is negative.

In Fig. 10, extracted stair step horizontal edges that are satisfied the property of arranging edges in longest increasing subsequence as like zebra-crossing edges. However, the calculated vanishing point is $(277.24, -87.20)$ as shown in Fig. 10(d). The y coordinate of the VP is -87.20 . The negative y value distinguishes this image from a zebra-crossing image. The variation of the VP in the stair and zebra-crossing image is due to the physical structure of the stair and zebra-crossing. The zebra-crossing does not have any slop with respect to the ground, where, stair is the opposite of the zebra-crossing.

The zebra-crossing region detected from zebra-crossing image is shown in Fig. 9, where y coordinate of the vanishing point is 72.92. Fig. 9(b) shows the zebra-crossing candidate region in original zebra-crossing image. The detected zebra-crossing candidate region is as shown in Fig. 9(c).

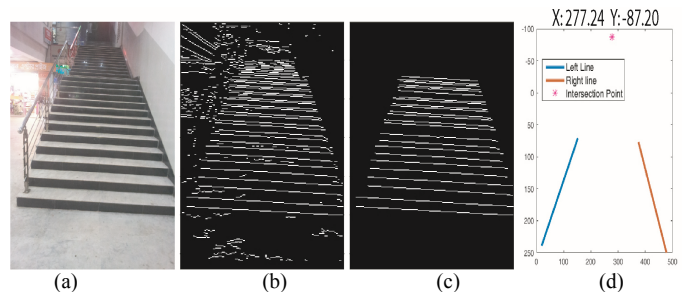


Fig. 10. Processing example of zebra-crossing like object stair: a) image, b) extracted edges, c) arranging edges longest increasing subsequences, and d) vanishing point.

III. EXPERIMENTAL RESULTS

The experimental results of some zebra-crossing images are explained in this section. All the experiments were performed on the Intel(R) Xeon(R) CPU E31220@3.10GHZ processor with 4GB RAM. In this paper, the experiments were performed on 480x320 zebra-crossing images under MATLAB simulator. 67 zebra-crossing images were used in the experiments. These zebra-crossing images are taken from outdoor environments with different illumination conditions, i.e., noisy, normal, and uneven to consider various types. Different illumination type's zebra-crossing images are presented in Fig. 11. All zebra-crossing images are captured from the front side of the zebra-crossing. The camera height from the ground should be a certain value that is not more than two meters. The angular displacement less than 30 degrees with respect to the zebra-crossing will work fine.

Fig. 12 shows the results of processing sample images. Input sample image, longest increasing HE segments, vertical VP, and the zebra-crossing detection region are shown in Fig. 12(a), Fig. 12(b), 12(c), and 12(d) respectively.

In Fig. 12, sample image 1, 2 and 3 are zebra-crossing images captured at day time under normal and noisy background. Where, sample 1 and 2 are straightforward zebra-crossing images captured from different heights. Sample 3 is the zebra-crossing image captured from a short distance. Zebra-crossing Sample 4 shows the experiment on the zebra-crossing at day time with the uneven illumination condition. Zebra-crossing Sample 5 shows the result on the colored zebra-crossing at uneven illumination condition. Each of these zebra-crossing samples has satisfied the condition of horizontal edges are arranged in longest increasing edge segments and has a vanishing point with positive y coordinate value. All these varieties of zebra-crossing are detected successfully by the proposed framework.

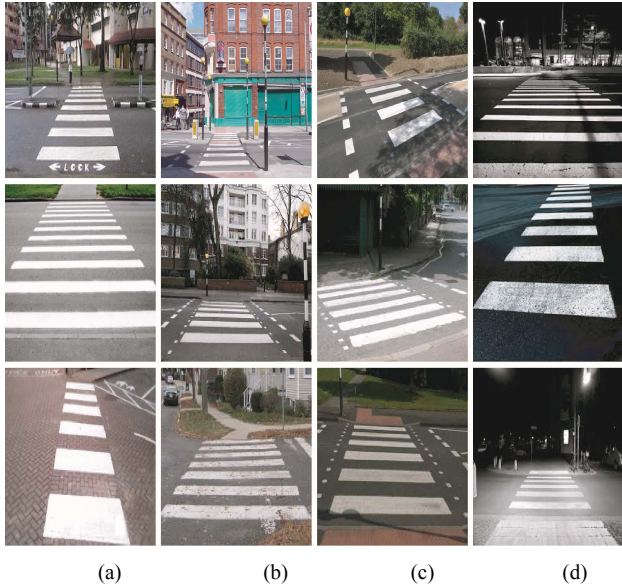


Fig. 11. Various types of zebra-crossing image: a) zebra-crossing with normal illumination, b) zebra-crossing image with noisy background, c) zebra-crossing with uneven illumination, and d) zebra-crossing image at night time with uneven illumination.

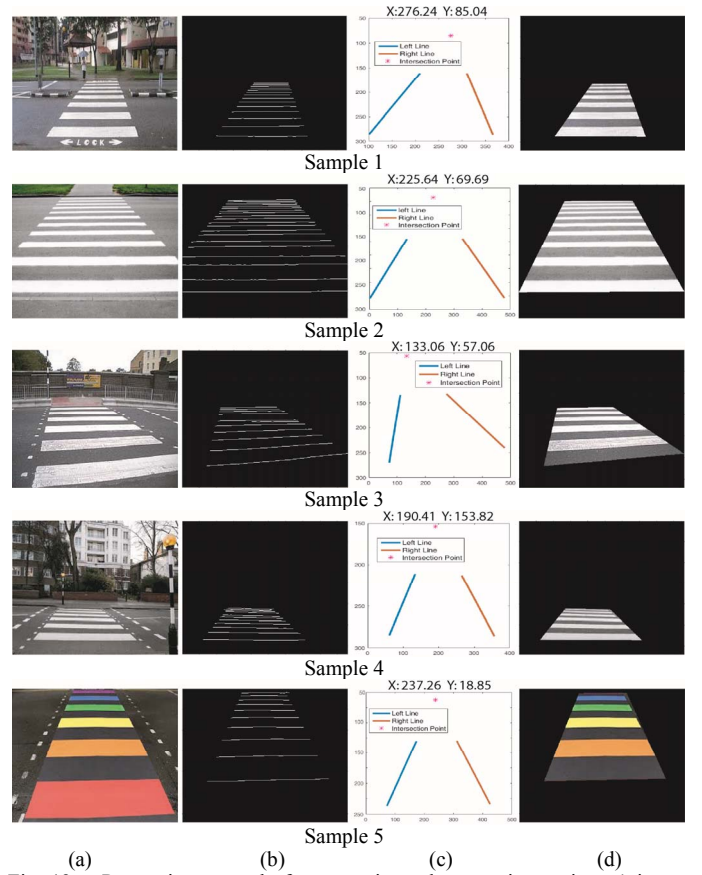


Fig. 12. Processing example for extracting zebra-crossing region: a) input image, b) arranging horizontal edges in longest increasing edge segments, c) vertical vanishing point, and d) zebra-crossing candidate region.

Fig. 13 shows some processing example of zebra-crossing at different environment such as rainy i.e. sample 6, under strip light i.e. sample 7, and human on the zebra-crossing line i.e. sample 8. If any objects are on the zebra-crossing line, the horizontal edges are cropped in the object places. However, the cropped edges are recovered by using the edge linking and tracking process.

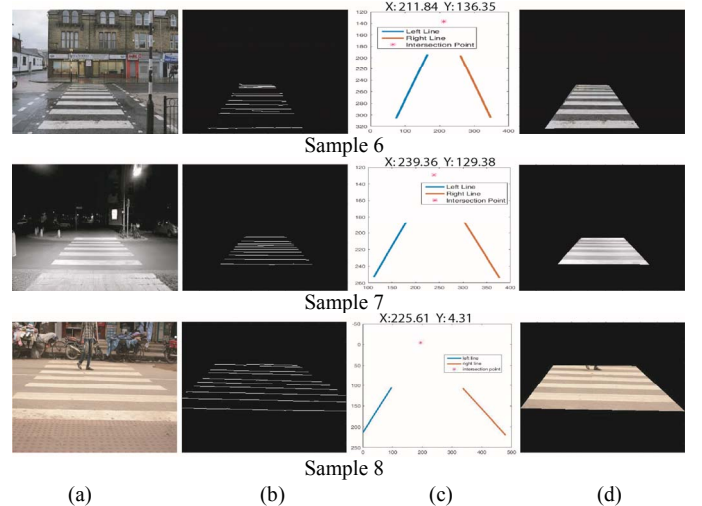


Fig. 13. Processing example for extracting zebra-crossing region in rainy and strip light environment: a) input image, b) arranging horizontal edges in longest increasing edge segments c) vertical vanishing point, and d) zebra-crossing candidate region.

From experimental results, it is shown that all the processing zebra-crossing sample images are detected the zebra-crossing region efficiently if the zebra-crossing framework can extract the zebra-crossing stripe edges effectively.

In Table I shows the detection accuracy of the zebra-crossing (ZC) that are measured from different types of zebra-crossing images. In Table II shows the average computational time at different stages of the proposed framework. The experimental processing results show that zebra-crossing with normal conditions provided highest response than others. It also gives the satisfactory response in other environmental conditions.

Table III shows the comparison of the proposed framework with [2] and [4] for zebra-crossing (ZC) detection. In [2], authors used the depth image information to detect the zebra-crossing region. However, it is hard to capture good quality depth images of pedestrian crosswalks compared to capture images such as stairs. The bipolarity theory concept is utilized in [2]. This concept is not works efficiently on the dark and noisy white stripes of zebra-crossing. On the other hand, in this paper, the zebra-crossing region is detected with the key feature, i.e., zebra-crossing horizontal stripe edges are arranged in increasing parallel order. This feature is measured from the zebra-crossing edge image within $O(n\log n)$ time, where, n is the number of horizontal parallel edges of a zebra-crossing. The average detection accuracy is computed in [2], in [4], and the proposed method over 52, 54, and 67 images respectively.

TABLE I. DETECTION ACCURACY AT DIFFERENT ENVIRONMENTAL CONDITIONS

Environmental conditions	Total image	Correctly detected	Detection accuracy (%)
Normal illumination	33	33	100
Uneven Illumination	21	20	95.23
Noisy background	13	12	92.31
Average	67	65	97.02

TABLE II. AVERAGE COMPUTATION TIME AT DIFFERENT STAGE OF THE PROPOSED FRAMEWORK

Different stage	Avg. time(s)
Filtering and detecting edges	0.020
Linking and tracking	0.018
Arranging and calculating vertical vanishing point	0.035
Total	0.073

TABLE III. COMPARISON WITH RESPECT TO ZEBRA-CROSSING DETECTION ACCURACY

Method	Detection accuracy (%)
[2]	78.9%
[4]	96.2%
The proposed method	97.02%

V. CONCLUSIONS

In this paper, a geometrical feature based framework has been proposed for extracting zebra-crossing region from a zebra-crossing image automatically. This framework has performed without any prior information about the position of the zebra-crossing in the zebra-crossing image to improve the

navigation process of autonomous systems. The utilized unique geometrical feature of zebra-crossing is zebra-crossing stripe's edges are arranged in the longest increasing order. The extracted increasing horizontal zebra-crossing edges are verified by the vertical vanishing point. In this paper, all zebra-crossing images are captured from the front side of zebra-crossing and with a height of less than two meters from the ground. The angular displacement less than 30 degrees with respect to the zebra-crossing will work fine. This framework was tested successfully on a group of zebra-crossing images with a varying style and different illumination conditions. The proposed framework detected zebra-crossing regions from these zebra-crossing images with acceptable running time and accuracy. The experiments demonstrate the effectiveness of the proposed framework. The proposed framework can detect the traditional zebra-crossing only. It is not provided satisfactory result for complex shapes zebra-crossing such as rounding zebra-crossing. Later, this framework will be updated for detecting complex shapes zebra-crossing.

References

- [1] W. Czajewski, P. Mrówka and P. Olszewski, "Video Processing for Detection and Tracking of Pedestrians and Vehicles at Zebra-crossings," In Tools of Transport Telematics, pp. 34-44, Springer International Publishing, 2015.
- [2] S. Wang, H. Pan, C. Zhang, C. and Y. Tian, "RGB-D image-based detection of stairs, pedestrian crosswalks and traffic signs," in Journal of Visual Communication and Image Representation, Vol. 25, No. 2, pp. 263-272, 2014.
- [3] S. Sichelschmidt, A. Haselhoff, A. Kummert, M. Roehder, B. Elias, and K. Berns, "Pedestrian crossing detecting as a part of an urban pedestrian safety system," In Intelligent Vehicles Symposium (IV), pp. 840-844. IEEE, 2010.
- [4] C. Yuzhen, C. Lushi, and J. Shuo, "An Image Based Detection of Pedestrian Crossing," In 2nd International Congress on Image and Signal Processing (CISP'09), pp. 1-5, IEEE, October, 2009.
- [5] L. Boudet, and S. Midenet, "Pedestrian crossing detection based on evidential fusion of video-sensors," Transportation research part C: emerging technologies, Vol.17, No. 5, pp. 484-497, 2009.
- [6] V. Ivanchenko, J. Coughlan, and H. Shen, "Detecting and locating crosswalks using a camera phone," In Computer Vision and Pattern Recognition Workshops (CVPRW'08), IEEE Computer Society Conference, pp. 1-8. IEEE, 2008.
- [7] J. Coughlan and H. Shen, "A fast algorithm for finding crosswalks using Fig-ground segmentation," In 2nd Workshop on Applications of Computer Vision, in conjunction with ECCV, Vol. 5. 2006.
- [8] W. Czajewski, P. Dąbkowski and P. Olszewski, "Innovate solutions for improving safety at pedestrian crossings," Arch. Transp. Syst. Telemat, Vol. 6, No.2, pp.16-22, 2013.
- [9] S. Se, "Zebra-crossing detection for the partially sighted." Proceedings. IEEE Conference on Computer Vision and Pattern Recognition, Vol. 2. IEEE, 2000.
- [10] F. Schaffalitzky and A. Zisserman, "Planar grouping for automatic detection of vanishing lines and points", in Image and Vision Computing, Vol. 18 Issue 9, pp. 647-658, 2000.
- [11] T. S. Lee, "Image representation using 2D Gabor wavelets", in IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 18, No. 10, 959-971, 1996.
- [12] L. Chen, "Laplacian Embedded Regression for Scalable Manifold Regularization", Neural Networks and Learning Systems, IEEE Transactions, Vol. 23, pp. 902 – 915, June, 2012.

A Novel Modified SFTA Approach for Feature Extraction

Md. Junayed Hasan

Department of Computer Science and Engineering
BRAC University
Dhaka 1212, Bangladesh.
junhasan@gmail.com

Jia Uddin

Department of Computer Science and Engineering
BRAC University
Dhaka 1212, Bangladesh
jia.uddin@bracu.ac.bd

Subroto Nag Pinku

Department of Computer Science and Engineering
BRAC University
Dhaka 1212, Bangladesh.
subroto.npi@gmail.com

Abstract—To increase the efficiency of conventional Segmentation Based Fractal Texture Analysis (SFTA), we propose a new approach on SFTA algorithm. We use an optimum multilevel thresholding hybrid method of Genetic Algorithm (GA) and Particle Swarm Optimization (PSO), called HGAPSO with the optimization technique for classification based on grey level range to get more accurate output. Experimental results show that proposed approach exhibits average 2% higher classification accuracy than conventional SFTA for our tested dataset.

Keywords—SFTA (Segmentation Based Fractal Texture Analysis), multilevel thresholding, HGAPSO, Otsu function.

I. INTRODUCTION

Preprocessing step, feature extraction, making the database of feature and matching the assigned image with the created database; these four steps can draw a solution for retrieval problem of images. Here, feature extraction plays a very important role as it defines the unique features of images. A number of researchers work on development of efficient feature extraction methods [1-4]. In [2], Costa *et al* proposed a new feature extraction method, called SFTA (Segmentation based Fractal Texture Analysis). On this conventional SFTA technique, the total process has two steps. One is two threshold Binary decomposition and another one is feature extraction.

On binary decomposition step, conventional SFTA uses multilevel Otsu algorithm for thresholding. It takes exponential time [6] to perform the trial and test for finding out the best number of threshold values to do the feature extraction technique, which is not efficient. Moreover the threshold values are not that much efficient to perform more accuracy. Again SFTA has a problem on finding out the number of binary images from binary decomposition portion. Conventional method is not actually well formed to find out the exact number of decomposed images for the creation of feature vector. In addition, conventional method creates dimensional mismatch.

To overcome the limitations; we propose our new model. We have also compared the classification accuracy of our novel modified SFTA approach with the conventional SFTA for our tested dataset.

The remainder of the paper is organized as follows - Section II displays how the conventional SFTA algorithm works. Section III describes the overview in depth of our proposed novel SFTA approach. The results of the experiments carried out and their analysis are included in Section IV and finally Section V concludes the paper.

II. LITERATURE REVIEW

To avoid the limitations of the conventional SFTA, in our novel approach, we use hybrid multilevel threshold method; defined as HGAPSO; for finding out significant threshold values with less time based on trial and test. After that; we have used optimization technique for classification based on grey level range; defined as OCGR; to attain more accurate threshold values from the collected ones from first step.

Ghamisi *et al.* proposed the HGAPSO method[3,4]; consisting of two main portions: one is for combining the standard velocity and another is for updating the rules of PSOs (Particle Swarm Optimization) with the ideas of collection, crossover and transfiguration from GA (Genetic Algorithm) [5]. We have used HGAPSO instead of Otsu because for an obvious reason. When Otsu method is comprehensive to solve multilevel thresholding, it is not efficient to define the ideal threshold values as for multilevel Otsu; because performing the trial and test to get the attained number of threshold will increase exponentially [6] along with increasing the number of thresholds. So we propose our new approach based on HGAPSO and OCGR in section III to get better accuracy on attained threshold values and avoid the exponential time necessity to obtain those values.

Conventional SFTA is a proficient texture feature extraction method; proposed by Costa *et al.* [2]; consisting

of two parts. The first part performs the decomposition technique for the input grey scale images to convert it into set of binary images by using Two-Threshold Binary Decomposition (TTBD). Then the second part is to compute the fractal dimension from its region margins for each subsequent binary image.

For the first step of the whole procedure, the TTBD considers a grayscale image as an input and we get a set of binary images as result. By using multilevel Otsu threshold algorithm, TTBD gives the resulting set T of threshold values; as the first step. Recursively; this Otsu algorithm is functional to each region of image till the preferred digit of thresholds is not attained.

The following step of the TTBD algorithm involves in disintegrating the grayscale image $I(x,y)$ which are measured as input; into a set of binary images; and it is accomplished by choosing couples of thresholds from that attained set of threshold values and relating a two-threshold segmentation [2]. Figure 1 demonstrate SFTA extraction process overview.

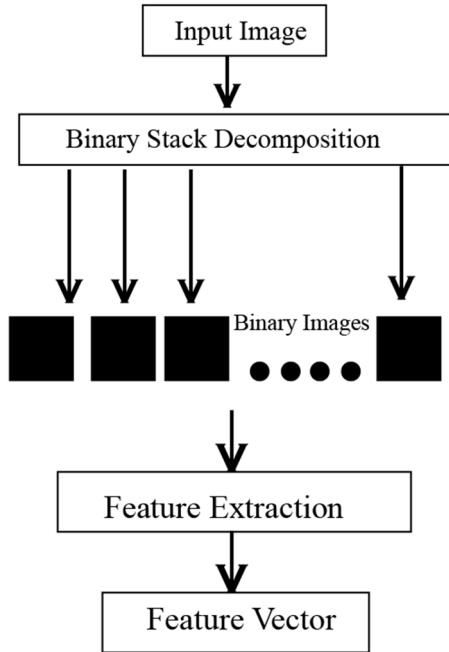


Fig. 1. SFTA Extraction Process Overview

According to the conventional SFTA; the number of causing binary images is $2n_t$ [2]; where n_t means number of thresholds; but here is a problem regarding the number of the subsequent binary images. On practical work, the number is not $2n_t$, it is $(2n_t-1)$. With relating the two threshold segmentation to the input image; the set of binary image is obtained; and this process use all pairs of adjoining thresholds from $TU\{n_i\}$ and all pairs of thresholds $\{t, n_i\}$, $t \in T$, where n_i corresponds to the maximum possible gray level in $I(x, y)$. Here n_i stands for gray level range; T for threshold. This way, finally, the number of resulting binary images is $2n_t - 1$. For example; if the number of threshold value is 4, on the conventional SFTA; the causing binary image would be 8; but according

to the correct and practical approach it would be 7. Figure 2 shows the detail process of this [23].

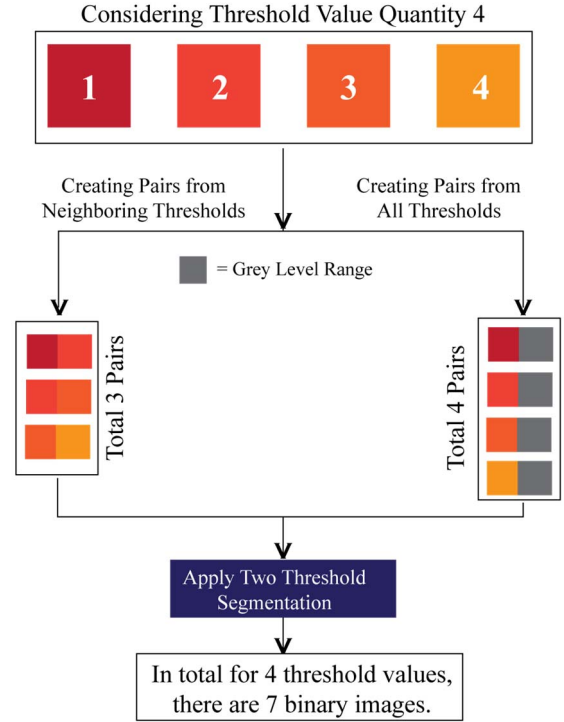


Fig. 2. Exact Binary Image Formation Process from Threshold.

III. PROPOSED MODEL

Figure 3 reveals a detailed implementation of our proposed model. This section demonstrates how the algorithm is set up.

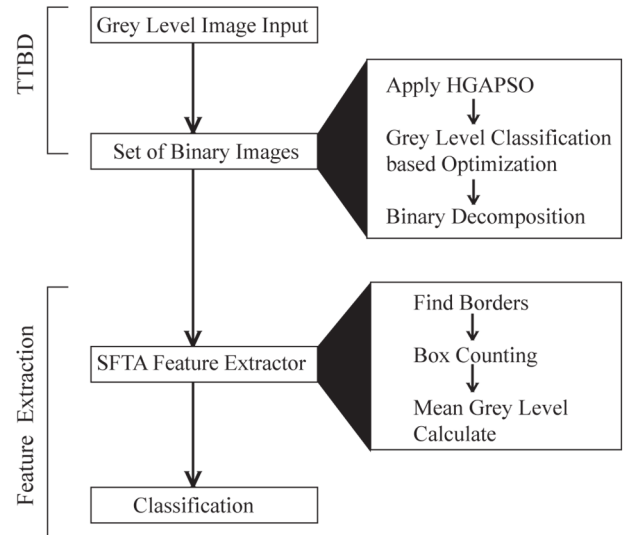


Fig. 3. The Proposed Model of Novel SFTA Method.

A. Apply HGAPSO

GA (Genetic Algorithm) and PSO (Particle Swarm Optimization) theorem [15-17] can perform together to

obtain a better optimization results [3,4]. In [3,4], Ghamisi *et al.* proposed the HGAPSO method; which has two main parts. One is to the association of the standard velocity and with the planning of selection, crossover and mutation from GA [5]; another one is to update the procedure of Particle Swarm Optimization.

Swarm or population is the set of solutions from PSO; in which each solution consists of a set of parameters and represents a point in the multidimensional space, denoted as a particle [21]. Having a predisposition of winged to a better search area over the pursuit crossways [17]; the rapidity of the particles are familiar based on the chronological behavior of each one and its neighbors. The velocity of the i -th particle in the k -th iteration is resolute as follows:

$$V_{id}^{k+1} = WV_{id}^k + c_1 r_1 (pb_{id}^k - x_{id}^k) + c_2 r_2 (gb_{id}^k - x_{id}^k) \quad (1)$$

Where c_1 and c_2 are defined as acceleration constants and r_1 and r_2 have random values ranging from 0 to 1. Here W stands for inertia weight [5]; which displays the location of each particle in the d -dimensional search space by the parameter x_{id}^k [5, 23]. The best previous position of each particle is represented by pb_{id}^k and considered as particle best position. gb_{id}^k is the best position of all particles, being denoted as global best particle [5,23]. The i -th particle position is updated by:

$$x_{id}^{k+1} = x_{id}^k + V_{id}^k \quad (2)$$

Then again, GA; based on the genetic process [5, 20]; is able to find an ideal solution by not exploring the whole search space [5]. Chromosome or individual is the optimization technique of GA which has several solutions. These chromosomes are consisting of several different valued genes [5]. The fitness value of the chromosome represents by the attributes of each individual; supported by these genes.

In our proposed method, to get the attained threshold value; by reducing the steps of exponential calculation of multilevel Otsu [6], we are using HGPASO and then again from those values; to select more significant threshold values we have used our own OCGR technique.

B. Optimization Technique for Classification based on Grey Level Range

From the previous step A, we get a number of threshold values. After that, we use those values to get more accurate number of threshold values. At the very beginning, we correspondingly categorize the grey level density 0 to 255 into equivalent classes. Then from all attained threshold values from previous step A, we calculate their incidence number on those classes. Then; from each class range, the mean values of thresholds are considered. If no incidence on any class, then we simply add the class range start point and class range end point and divide the attained summation by two, to consider the threshold value [23]. Thus our proposed OCGR process supports to attain the preferred number of improved threshold values. By using pseudo code, the whole algorithm of our proposed model is given below.

Algorithm 1 HGAPSO & OCGR Based SFTA extraction algorithm.

Requirement: I; the greyscale image and n_t the number of thresholds

01.	Threshold, T	$\leftarrow$ Hybrid GAPSO (I, n_t)
02.	Set of Improved Threshold, T_{op}	$\leftarrow$ OCGR (T , number of threshold obtained occurrence on each Grey Level Class)
03.	Threshold Set 1 from binary images, A	$\leftarrow \{\{t_j, t_{j+1}\} : t_j, t_{j+1} \in T_{op}, j \in [1.. T_{op} - 1]\}$
04.	Threshold Set 2 from binary images, B	$\leftarrow \{\{t_j, n_j\} : t_j \in T_{op}, j \in [1.. T_{op}]\}$
05.	j	$\leftarrow 0$
06.	for $\{\{lower\ threshold, upper\ threshold\} : \{lower\ threshold, upper\ threshold\} \in A \cup B\}$	do
07.	Binary Image	$\leftarrow$ TwoThresholdSegmentation (I , lower threshold, upper threshold)
08.	Border Image(x, y)	$\leftarrow$ Border Find (Binary Image)
09.	SFTA Feature Vector[j]	$\leftarrow$ Count Box (Border Image)
10.	SFTA Feature Vector [j + 1]	$\leftarrow$ GrayLevelMean (I , Binary Image)
11.	SFTA Feature Vector [j + 2]	$\leftarrow$ Count Pixel (Binary Image)
12.	j	$\leftarrow j+3$
13.		end for
14.	return	SFTA Feature Vector

IV. EXPERIMENTAL RESULT ANALYSIS

Figure 3 demonstrates how the proposed algorithm is set up. We applied our proposed algorithm on two publicly available dataset: KTH-TIPS [8] and Texture Surface [9]. On these two different dataset, we apply the original SFTA algorithm and our proposed novel SFTA algorithm. Like the conventional SFTA approach [2], we did not notice any major changes on classification accuracy increasing phase after the threshold value reached to 14. Till 14, the accuracy was significantly in increasing order. The interesting fact is our proposed novel SFTA algorithm gave more accuracy on different number of threshold values till 14 than the original SFTA algorithm. Mainly after 14, no additional texture patterns has been identified by our proposed TTBD phase. For individual training on each dataset for each image, we use SVM (Support Vector Machine).

The KTH-TIPS (Textures under varying Illumination, Pose and Scale) consist of 810 grayscale images [8] with 10 texture classes [2]. The Textured Surfaces Dataset [9] has numerous surfaces such as wood, marble and fur; these materials; under varying viewpoints, scales and illumination conditions [2].

A. Accuracy Measurement

Figure 4 shows the accuracy percentage for dataset KTH-TIPS [8] for number of threshold values. From here, it is clearly visible that our proposed algorithm can give more accuracy than SFTA [2] algorithm. Here from the figure 4, it is visible to us that our experimental result

shows the detailed comparison between the performance of SFTA and our proposed model. When the number of threshold value is 2,4,6,8,10,12 and 14, we get a much better accuracy than SFTA. For 16, 18, 20 numbers of threshold values also giving us the higher accuracy than conventional SFTA, but not that much in percent like the previous mentioned threshold number does. So till the number of threshold values 14, our proposed model can bring a significant improvement on classification accuracy.

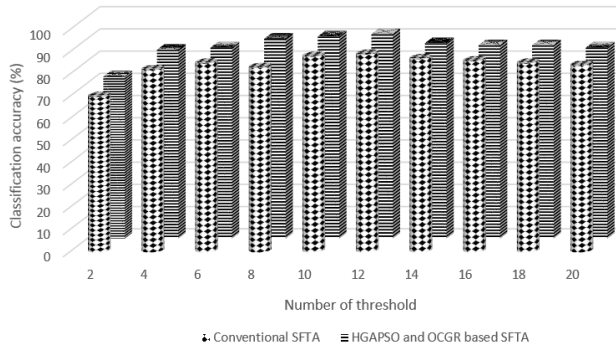


Figure 4: For **KTH-TIPS Dataset**; Classification Accuracy Comparison in Percentage between Conventional SFTA and Proposed Model.

Similarly Figure 5 shows the accuracy percentage for Texture Surfaces Dataset [9] for number of threshold values. From figure 5, along with better accuracy for our proposed model; we can also see that when the number of threshold value is 2, 4,6,8,10,12 and 14, we get a far better accuracy than SFTA. For 16, 18, 20 numbers of threshold values also giving us the higher accuracy than conventional SFTA, but not that much in percent like the previous mentioned threshold number does. So till the number of threshold values 14, for both experimented case, our proposed model can bring a significant improvement on classification accuracy.

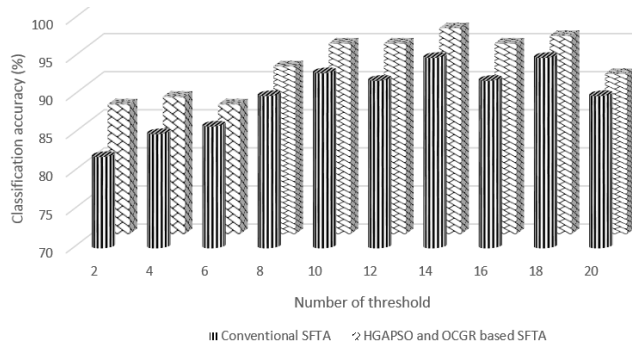


Figure 5: For **Texture Surface Dataset**; Classification Accuracy Comparison in Percentage between Conventional SFTA and Proposed Model.

V. CONCLUSION

This paper presents a HGAPSO and OCGR based SFTA texture feature extraction method. In this method, we used HGAPSO and OCGR to select the best number of threshold values and reduce steps of trail and test for the best number of feature selection. Which significantly impact on

performance of conventional SFTA algorithm. Experimental results show average 96% classification accuracy for our tested dataset, where conventional Otsu based thresholding SFTA displays 94% classification accuracy. On average for different number of testing and training images our proposed method gives 2% more accuracy than the conventional one.

REFERENCES

- [1] Ryszard S. Choras, "Image Feature Extraction Techniques and their applications for CBIR and Biometrics Systems," International Journal of Biology and Biomedical Engineering, Issue 1 vol. 1, pp. 6-16, 2007.
- [2] A. F. Costa, G. E. Humpire-Mamani, A. J. M. Traina, "An Efficient Algorithm for Fractal Analysis of Textures," Conference on Graphics, vol. 1, pp. 39-46, 2012.
- [3] P. Ghamisi, J. A. Benediktsson, "Feature Selection Based on Hybridization of Genetic Algorithm and Particle Swarm Optimization," IEEE Geoscience and Remote Sensing Letters, vol. 12, no. 2, pp. 309-313, Feb. 2015.
- [4] P. Ghamisi, "A Novel Method for Segmentation of Remote Sensing Images based on Hybrid GA-PSO," International Journal of Computer Applications, vol. 29, no. 2, pp. 7-14, September 2011.
- [5] S. K. Nahavandi, P. Ghamisi, L. Kumar, M. S. Couceiro, "A novel adaptive compression technique for dealing with corrupt bands and high levels of band correlations in hyperspectral images based on binary hybrid GA-PSO for big data compression," International Journal of Computer Applications, vol. 109, no. 8, pp. 18-25, 2015.
- [6] F. M. Ali, "Optimum Multilevel Thresholding Hybrid GA-PSO by Algorithm," Jurnal Ilmu Komputer dan Informasi, vol. 6, no. 1, p. 1-5, 2013.
- [7] Cheng, John, Max Grossman, and Ty KcKercher. Professional CUDAC Programming. Indianapolis: Wrox, 2014. Print.
- [8] B. Caputo and E. Hayman, "Class-specific material categorization," Computer Vision, vol. 2, pp. 1597-1604, 2005.
- [9] S. Lazebnik, C. Schmid, J. Ponce, "A sparse texture representation using local affine regions," Pattern Analysis and Machine Intelligence, vol. 27, no. 8, pp. 1265-1278, Aug. 2005.
- [10] M. Varma, A. Zisserman, "A statistical approach to material classification using image patch exemplars," IEEE Transactions on pattern analysis and machine intelligence, vol. 31, no. 11, pp.2032-2047, Nov. 2009.
- [11] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikäinen, X. Chen, W. Gao, "WLD: a robust local image descriptor," IEEE Transactions on pattern analysis and machine intelligence, vol. 32, no. 9, pp. 1705-20, Sep. 2010.
- [12] T. Ahonen, J. Matas, C. He, "Rotation invariant image description with local binary pattern histogram fourier features," Image Analysis, pp. 61-70, 2009.
- [13] Otsu N, A threshold selection method from gray-level histograms, IEEE., 1979.
- [14] J. N. Kapur, P. K. Sahoo, A. K. C. Wong, "A new method for gray-level picture thresholding using the entropy of the histogram," Computer Vision, Graphics and Image Processing, vol. 29, pp. 273-285, 1985.
- [15] Y. P. Yeng, "A fast scheme for multilevel thresholding using genetic algorithms," Signal Processing, vol. 72, pp. 85-95, 1999.
- [16] D. B. Fogel, "Evolutionary computation: Toward a new philosophy of machine intelligence," 2nd ed., Piscataway, NJ: IEEE Press, 2000.
- [17] R.C. Eberhart, J. Kennedy, "A new optimizer using particle swarm theory," Sixth International Symposium on Micro Machine and Human Science, pp. 39-43, 1995.
- [18] P. Ghamisi, Micael S. Couceiro, Jon Atli Benediktsson, Nuno M. F. Ferreira, "An Efficient Method for Segmentation of Images Based on Fractional Calculus and Natural Selection," Expert Systems with Application Publisher, vol. 39, pp. 12407-12417, 2012.
- [19] J. Kennedy and R.C. Eberhart, Swarm Intelligence, Morgan Kaufmann Publishers, San Francisco, 2001.

- [20] Goldberg, D. E. (1989). Genetic algorithms in search, optimization, and machine learning. Reading, Massachusetts, USA: Addison-Wesley Longman.
- [21] P. Ghamisi, M. S. Couceiro, J. A. Benediktsson, "A Novel Feature Selection Approach Based on FODPSO and SVM," IEEE Transaction on Geoscience and Remote Sensing, [In press].
- [22] P. J. Angeline. Evolutionary optimization versus particle swarm optimization: Philosophy and performance differences. In V. William Porto and et al., editors, Evolutionary Programming, volume 1447 of Lecture Notes in Computer Science, pages 601–610. Springer, 1998.
- [23] M. J. Hasan, N. K. Khan, N. A. Hridi, F. T. Ontora, "A novel parallel feature extraction method using HGAPSO and GLCO based SFTA," Undergraduate thesis, Spring 2016, BRAC University.

A Comparative Analytical Approach for Gate Leakage Current Optimization in Silicon MOSFET: A Step to More Reliable Electronic Device

Asif Abdullah Khan, Arun Audhikary, Md. Fahim-Al-Fattah¹, Md. Ashiquel Amin, Rajesh Nandi

Department of Electrical and Electronic Engineering

Khulna University of Engineering & Technology, Khulna, Bangladesh

¹Bangladesh Army University of Science & Technology, Saidpur

Email: abdullah1003065@gmail.com, arunkuet10@gmail.com, ¹fahim057@gmail.com

Abstract— A comparative analytical approach for performance evaluation of direct tunneling gate leakage current of ultrathin silicon MOS device has been introduced. Direct tunnel characteristics in NMOSFET with different gate dielectric material (HfO_2 , SiO_2 and Al_2O_3) at nanoscale regime has been analyzed. The variations of direct tunneling current with gate length, gate width, oxide thickness and various dielectric materials has been investigated in depth. Thus, scaling limit of gate dimension and choice of alternative gate dielectric for MOSFET is explored based on the direct tunnel characteristics for gate leakage current optimization.

Keywords—MOSFET, gate leakage current, direct tunneling, gate length, gate dielectrics, leakage current optimization.

I. INTRODUCTION

The scaling of semiconductor device has been the driving force for the technology advancements in the industries over the last few decades. Scaling indicates the reduction of the lateral geometric dimensions of MOSFET devices. Scaling improves cost, speed and power per function with every newer technology generation [1]. So, Size reduction has become the vital point of research for modern electronic device upgradation. However, to comply with the scaling trends some significant issues have come forward like short channel effect, off-state current or the leakage current of the MOSFETs. Leakage current optimization is the challenging task for the small scale electronic device fabrication, because power consumption, device lifetime and performance is related to this unwanted leakage current. To keep pace with Moore's law while device size minimization, reduction of gate length increases drive current that also necessitates a decrease of the gate oxide thickness to maintain electrostatic control of the charges induced in the channel. Therefore, while lateral dimension of the MOS device is being scaled, vertical dimensions e.g. oxide thickness, gate width must be reduced in order to keep pace with the reduction of gate length. So the oxide layer thickness, gate width, gate length's has individual effects on gate leakage current. Except the defects in material during fabrication, gate leakage current can be optimized by using different dielectric and proper scaling of device dimension. Due to the limitation of fabrication, all kinds of dielectric materials can't be used in MOS devices. SiO_2 is the

most widely used dielectric material but it is not possible to abate leakage current significantly with this. As the leakage current heated up the device, it has negative impact on device performance because semiconductor device is very sensitive to temperature change. So the researchers and device engineers are persistently working on reducing the leakage current to improve the device performance. However the previous research to analyze gate leakage current was only confined on a single model. For better realization of gate leakage current, sufficient understanding of the theoretical performance with distinct dielectric materials is of immense importance [2]. In this article, a comparative analytical approach is introduced with different dielectric materials considering device dimension to reduce leakage current. Here a significant reduction of leakage current has been noticed with HfO_2 rather than Al_2O_3 and SiO_2 . Also it is observed that there is a reduction of leakage current with comparatively small gate length as well as gate width. But the things are found reverse for small oxide thickness.

II. THEORETICAL ANALYSIS OF GATE LEAKAGE CURRENT

Gate oxide, which serves as insulator between the gate and channel, should be made as thin as possible to increase the channel conductivity and performance when the transistor is on and to reduce the subthreshold leakage when the transistor is in off condition. However, with current gate oxides with the proposed thickness of around 1.2nm the quantum mechanical phenomenon of electron tunneling occurs between the gate and channel, leads to increased power consumption.

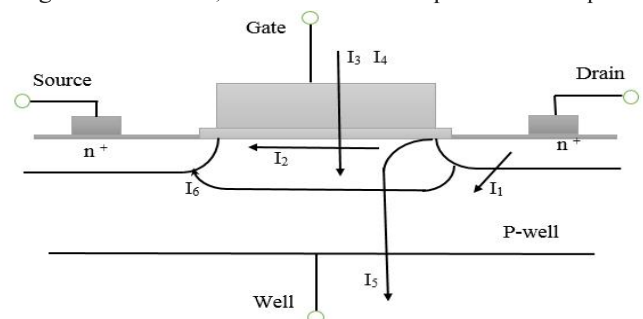


Figure 1. Leakage current mechanism in nanoscale MOSFET.

Different types of leakage mechanism has been illustrated in fig. 1. Among these components of leakage current, I_3 is the oxide tunneling current which occurs in both ON and OFF states of device.

Quantum tunneling refers to the quantum mechanical phenomenon where a particle tunnels through a barrier [3]. Tunneling is often explained by using the Heisenberg uncertainty principle and the wave-particle duality of matter [4]. The reduction of gate oxide thickness results in an increase in the field across the oxide. The high electric field across the oxide results in tunneling of electrons from substrate to gate and also from gate to substrate through the oxide, resulting in the gate oxide tunneling current[5]. The mechanism of tunneling between the substrate and gate can be divided into two types, namely:

(a) Fowler–Nordheim tunneling (b) Direct tunneling

The tunneling probability of an electron through barrier depends on the thickness of the barrier and its height, and the structure. Therefore, the tunneling probabilities of a single electron in FN tunneling and direct tunneling are not same, causing different tunneling currents. In case of Fowler Nordheim tunneling, electrons tunnel through a triangular potential well, whereas in case of direct tunneling, electrons tunnel through a trapezoidal well. In very thin oxide layers (less than 3–4 nm), electrons instead of tunneling into the conduction band of SiO_2 , directly tunnel to the gate through the band gap of the SiO_2 . In general, the direct tunneling current through band gap of a semiconductor is given by

$$J_T = \frac{2q}{(2\pi)^{3h}} \int_0^\infty (f_1 - f_2) \{ \iint P dk_y dk_z \} dE \quad (1)$$

where, P is the tunneling probability; E is total electron energy; K_y and K_z are the wave vectors in the plane of the barrier; f_1 and f_2 are the probabilities of occupation of the states, given by the Fermi–Dirac distribution functions[6]. The integration limits for the wave vectors are determined from momentum and energy conservation. Therefore, the expression for tunneling current density becomes[7]:

$$J_T = A E_{ox}^2 \exp \left\{ - \frac{B \left[1 - \left(1 - \frac{qV_{ox}}{\phi_{ox}} \right)^{3/2} \right]}{E_{ox}} \right\} \quad (2)$$

where, $A = (q^3/16\pi^2\hbar\phi_{ox})$ and $B = (4\sqrt{2m^*}\phi_{ox}^{3/2}/3\hbar q)$. Direct tunneling current is significant for thin oxide thickness. However, in order to predict the gate tunneling current, an accurate knowledge of gate oxide potential (V_{ox}) must be needed. V_{ox} can be calculated as[8]:

$$V_{ox} = V_{gs} - V_{FB} - \phi_s - V_{poly} \quad (3)$$

By solving the Poisson equation in the poly gate, it is found that,

$$V_{ge} = V_{FB} + \phi_{so} + \frac{q\epsilon_{si}N_{poly}T_{ox}^2}{\epsilon_{ox}^2} \cdot \left(\sqrt{1 + \frac{2\epsilon_{ox}^2(V_{gs} - V_{FB} - \phi_{so})}{q\epsilon_{si}N_{poly}T_{ox}^2}} - 1 \right) \quad (4)$$

where, V_{ge} gate voltage minus the gate-depletion voltage ($V_{gs} - V_{poly}$); V_{FB} is flat band voltage; ϵ_{si} and ϵ_{ox} dielectric constants of Si and SiO_2 ; N_{poly} is polysilicon gate doping concentration;

ϕ_{so} indicates surface potential equaling twice the Fermi potential, ϕ_f . The flat band voltage is the gate voltage considering there is no band bending in the semiconductor expressed as[9]

$$V_{FB} = \phi_{ms} - \frac{Q_{ox}}{C_{ox}} \quad (5)$$

where, ϕ_{ms} is the metal-semiconductor work function, Q_{ox} is the oxide charge per unit area and C_{ox} is the gate oxide capacitance[10]. Starting from the flat band position, there are three main regions of biasing of MOSFET structure and they are — accumulation, depletion and inversion[11]. When applied gate voltage is negative, positive and highly positive, these three modes appear consecutively. Therefore, equation (3) can be rearranged as

$$V_{ox} = V_{ge} - V_{FB} - \phi_s \quad (6)$$

This expression indicates that the gate voltage is distributed among oxide potential, flat band voltage and also surface potential. The surface potential, ϕ_s can be written as

$$\phi_s = \left[\frac{\gamma}{2} \left(-1 + \sqrt{1 + \frac{4(V_{gs} - V_{ge} - V_{FB})}{\gamma^2}} \right) \right]^2 \quad (7)$$

$$\gamma = \frac{\sqrt{2\epsilon_{si}qN_{sub}}}{C_{ox}} \quad (8)$$

where, ϕ_s is surface band bending of the substrate material and N_{sub} is channel doping concentration; γ is body-effect parameter. Using the above expression, it is easily possible to predict the tunneling gate leakage current. The electric field across the oxide can be defined as follows:

$$E_{ox} = \frac{V_{ox}}{t_{ox}} \quad (9)$$

where, t_{ox} is the thickness of the gate oxide layer. Therefore, it reveals that if the oxide thickness is decreased, the electric field increases which in turn causes the increase of gate leakage current. Direct tunneling gate leakage current, I_g can be expressed as follows:

$$I_g = J_T \cdot W \cdot L \quad (10)$$

where, W and L are the gate length and gate width respectively. This expression indicates that gate leakage optimization needs proper scaling of gate width and length.

TABLE I. SIMULATION PARAMETERS FOR TUNNELING PROCESS [6],[7],[9]

Parameters	Value used for simulation
m^*	$0.4m_0$
V_{FB}	-1 V
ϕ_{ox}	3.1 eV
ϵ_{ox}	3.45×10^{-13} F/cm
ϵ_{si}	1.04×10^{-12} F/cm
ϕ_f	0.46 eV
N_{sub}	$1 \times 10^{17} \text{ cm}^{-3}$
N_{poly}	$1 \times 10^{20} \text{ cm}^{-3}$

I. RESULTS AND DISCUSSION

The device used in this work is a Silicon NMOSFET with N^+ polysilicon gate. The gate oxide thickness of the device is taken 1.4 nm while considering constant gate oxide thickness. The effective mass has been assumed same for three biasing regimes of operation. Simulation is conducted with three dielectrics separately and also considering device scaling effect. MATLAB software has been used to conduct the simulation test of the gate leakage characteristics.

A. Gate Length Dependent Characteristics

The influence of different gate length on gate leakage current is simulated. For this purpose the gate voltage is varied and voltages at other terminals are kept at zero. The width and the oxide thickness of the tested device are $100\mu\text{m}$ and 1.4 nm respectively. The tunneling gate leakage current increases almost exponentially with the increase in gate length as a function of gate voltage as shown in fig.2.

For negative voltage gate leakage current, I_g is greater than that of the same positive voltage. This is because that the substrate used here p-type whose majority carrier is hole. In case of positive voltage the leakage current is due to the flow of electron. For negative voltage, it is due to the flow of hole and more hole generates from the p- type substrate that results in more gate leakage current.

B. Gate Width Dependent Characteristics

The influence of scaling of gate width on the gate leakage current is studied with similar consideration for a wide range of gate width. Results show that leakage current increases exponentially with increase in gate width as a function of gate voltage. Actually gate to source voltage, V_{gs} controls the potential barrier. When V_{gs} is high, barrier is sufficiently low for the carriers to flow. Therefore, gate leakage current increases with increase in gate bias as illustrated in fig. 3. Leakage current increases with gate width and vice-versa.

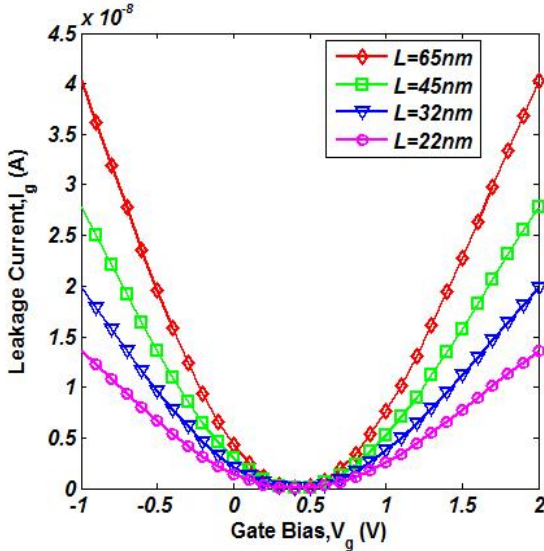


Figure 2. Direct tunneling gate leakage current vs. gate voltage for different gate lengths.

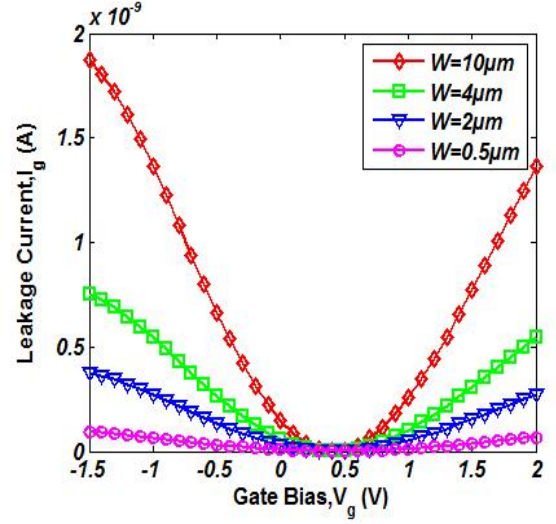


Figure 3. Direct tunneling gate leakage current vs. gate voltage for different gate widths.

Gate leakage current in constant gate length case is comparatively less than that of constant width case. As vertical dimensions in a MOS device are to be optimized in order to reduce the gate length, the effect of oxide thickness on gate leakage characteristics becomes very much important.

C. Oxide Thickness Dependent Characteristics

Different dielectric oxide thickness variation has effect on gate leakage current. In our article, 1.3nm, 1.4nm, 1.5nm thick oxide is used. As thickness decreases, gate leakage current increases exponentially. The rate of rise in leakage current is higher for negative gate voltage than that of positive voltage. If the oxide thickness is reduced, electric field across oxide is increased and, as a result, gate dielectrics are near to its break down point. Hence, oxide integrity has become very difficult to maintain and also the reliability of the device. In fig. 4, the relationship between tunneling leakage current and oxide thickness is illustrated considering gate length and width of 22nm and $100\mu\text{m}$ respectively.

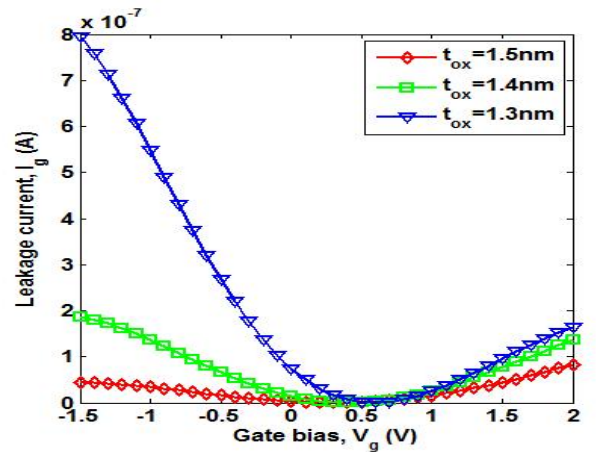


Figure 4. Direct tunneling gate leakage current vs. gate voltage for different oxide thicknesses.

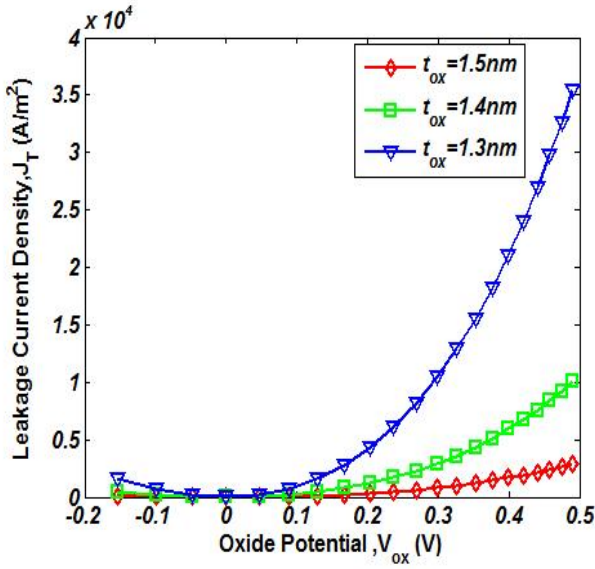


Figure 5. Direct tunneling gate leakage current density vs. oxide potential as a function of different oxide thicknesses.

The relationship between oxide potential and gate tunneling leakage current density is also studied as shown in fig. 5. Tunneling current density increases with the increase in oxide potential as a function of gate voltage and it is independent of gate dimension.

D. Impact of Different Dielectric Materials on Gate Leakage Current

Introduction of alternative gate oxide materials with high dielectric constants can be a solution to reduce gate tunneling leakage current and limit of scalability may be enhanced without increasing leakage current. Replacement of silicon dioxide with an alternative high-k dielectric material offers increased gate capacitance without the associated leakage effects. Fig. 6 shows that the gate leakage current density varies for different oxide material and Hafnium oxide allows the least gate leakage current.

HfO_2 has a relative dielectric constant (k) of ~ 25 , that is six times larger than that of SiO_2 . A 6 nm thick HfO_2 film is similar to 1 nm thick SiO_2 and both films can produce the same C_{ox} . So, this HfO_2 film has an equivalent oxide thickness 1 nm. Therefore, the HfO_2 film offers a thicker tunneling barrier to the electrons and holes. So leakage current is several orders of magnitude smaller than the leakage current with SiO_2 as gate dielectric material.

In this article, the impact of different gate dielectrics on tunneling leakage current is analyzed with gate voltage, V_g fixed at 1V. Fig.7 shows that gate leakage current decreases with increase in oxide thickness and Hafnium oxide offers better gate leakage control and may be a most reliable dielectric for advance MOSFET.

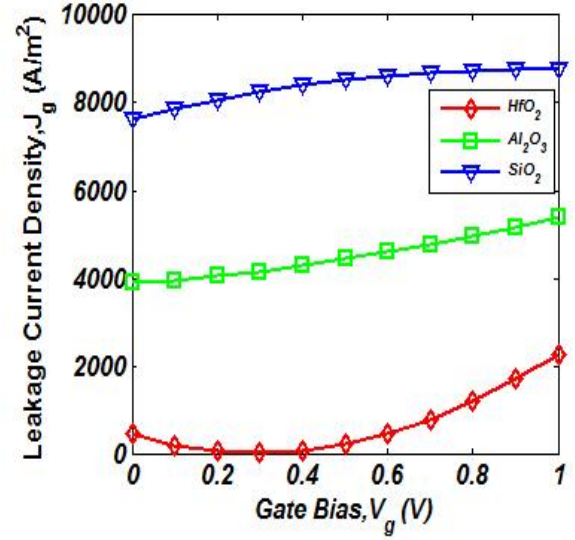


Figure 6. Direct tunneling gate leakage current density as a function of gate voltage for different dielectric materials.

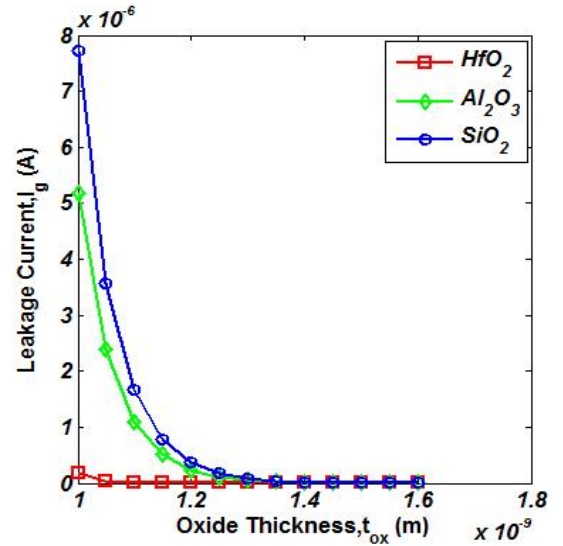


Figure 7. Direct tunneling gate leakage current as a function of oxide thickness for different dielectric materials.

II. CONCLUSION

In summary, a details analytical model for analyzing direct tunneling gate leakage current is studied. A details study on gate leakage optimization is done with comparative approach and with details illustration. The relationships of gate leakage current with device geometry and other relevant parameters are investigated in details based on direct tunnel characteristics. Results show that when silicon NMOS device has the same gate width and different gate lengths, the gate leakage current increases exponentially with increase in gate length and when the device has same gate length and different

gate width, the gate leakage current increases exponentially with increase in gate width. In this case, gate leakage current is found to be less than that of constant width case. With decreasing oxide thickness, the tunneling leakage current starts to increase. For different gate dielectrics similar analysis is performed. HfO_2 offers the least gate leakage current among those dielectrics. However, Al_2O_3 shows better leakage characteristics than SiO_2 although SiO_2 has more reliability than others. In terms of oxide thickness comparison, simulation also shows better leakage performance of HfO_2 . Thus, the results show additional insight into the analysis of gate leakage current of MOSFET that will help to increase reliability of electronic device.

REFERENCES

- [1] G. E. Moore, "Cramming more components onto integrated circuits," *Electronics*, vol. 38, pp. 114 -117, 1965.
- [2] Fahim-Al-Fattah, M., Tawabur Rahman, M., Sherajul Islam, M. and Bhuiyan, A.G., " A Study on Theoretical Performance of Graphene FET using Analytical Approach with Reference to High Cutoff Frequency" *International Journal of Nanoscience*, vol. 15 No. 2 p.1640001, 2015.
- [3] *MOSFET Technology Scaling, Leakage Current and Other topics*. Available: <http://www-inst.eecs.berkeley.edu/~ee130/sp06/chp7full.pdf>
- [4] D. A. Neamen, *Semiconductor Physics and Devices*. New York: The McGraw-Hill companies, 1992, ch.11 and ch.12.
- [5] K. Roy, S. Mukhopadhyay, and H. Mahmoodi-Meimand, "Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits," in *Proc. IEEE*, vol. 91, no. 2, pp. 305-327, 2003.
- [6] J. C. Ranuarez, M. J. Deen, and C. Hung Chen, "A review of gate tunneling current in MOS devices," *Elsevier, Microelectronics Reliability*, 46, pp. 1939-1956, 2006.
- [7] S. Hu, X. F. Wu, and Z. F. Xi, *Gate leakage current of NMOSFET with ultra-thin gate oxide*. Berlin, Heidelberg, Germany: Springer-Verlag, vol.19, pp. 3105-3109, 2012.
- [8] W. C. Lee and C. Hu, "Modeling CMOS tunneling currents through ultrathin gate oxide due to conduction and valence band electron and hole tunneling," *IEEE Trans. Electron Devices*, vol. 48, no. 7, pp. 1366-1373, 2001.
- [9] M. Marin, M. J. Deen, M. de Murcia, P. Llinares, and J. C. Vildeuil, "A new method for the channel length extraction in MOSFETs with sub 2 nm gate oxide," *IEEE Electron Device Lett.*, vol. 25, no. 4, pp. 202-204, 2004.
- [10] J. Singh, *Semiconductor Devices: Basic Principles*. New York: John Wiley & Sons Inc., 2001, ch.9.
- [11] C. H. Choi, K. Y. Nam, Z. Yu, and R. W. Dutton, "Impact of gate direct tunneling current on circuit performance: A simulation study," *IEEE Trans. Electron Devices*, vol. 48, no. 12, pp. 2823-2829, 2001.

An Innovative Implementation of Indoor Positioning System Using GPS

Sunjida Sultana¹, Mahdia Tahsin², Tasmia Reza³ and Md. Hossam-E-Haider⁴

Department of Electrical, Electronics and Communication Engineering

Military Institute of Science and Technology

Dhaka, Bangladesh

s_sunjida@yahoo.com, mahdia0369@gmail.com, tsm_rz@yahoo.com, haider8400@yahoo.com

Abstract— Global Positioning System is used by GPS tracking unit to determine the precise location of vehicles, persons or other assets and to record their position at regular intervals. The recorded data can be stored within the tracking unit or it may be transmitted to a data base or internet-connected computer device using a cellular, radio or satellite modem embedded in the unit. This will allow the location to be demonstrated as a map backdrop either considered in real time or later analyzing the track depending on different GPS Tracking System. This archetype is a combination of cellular device, Information and Communications Technologies (ICT), indoor and outdoor positioning systems and Server Software giving people information to guide them to their destination. The main motivation of this work is to introduce with a new tracking system which is an innovative additional device that can enable the individual to enjoy the flexibility in motion and develops an advanced sense of confidence and self-reliance.

Keywords—GPS;LBS;A-GPS;Wi-Fi;GSM.

I. INTRODUCTION

In the developed world the miniaturization and explosion of Information and Communications Technologies (ICT) such as cell phones, laptops, computers and personal wireless systems are rapidly and continuously evolving [1]. So a new approach to GPS tracking system is proposed to aid in accessing all forms of positioning information. This system can be used in navigating a destination and obtaining the information of its location. The offered system allows a very specific application of location-based services (LBS) which utilize the information of the location of the user or where the demand for information is indicated in a geographic sense. It requires an acceptable accuracy level in determining position. In general, sometimes GPS (or more generally the Global Navigation Satellite System - GNSS) fails to operate as it is hardly possible to receive the satellite signals inside most of the high-rise buildings and underground for indoor applications. So, other technologies must be incorporated.

In this paper we have proposed a theoretical design of GPS based tracking system and showed its implementation in an institution.

II. SATELLITE TRACKING

Satellite navigation is basically the means of several methods of determining or planning an object's position and course by geometry. Global navigation satellite system provides the user with a three-dimensional positioning solution by passive ranging using radio signals transmitted by orbiting satellites. This determination of positions mainly depends on the number of satellites that are available at a time. Availability depends on the constellation or geometry of the satellites. Good satellite geometry is obtained when the satellites are properly spread out in the sky [2]. Generally, the more the satellites are spread out in the sky, the better the satellite geometry, and vice versa. The satellite track is the path which is followed by satellite vehicle periodically. It provides the geometry of the satellite placement and has very important role in accuracy and reliability analysis. The satellites of GPS are unevenly distributed over six planes. GPS satellite constellation consists of total 24 satellites; each plane may contain 4 satellites, more or less.

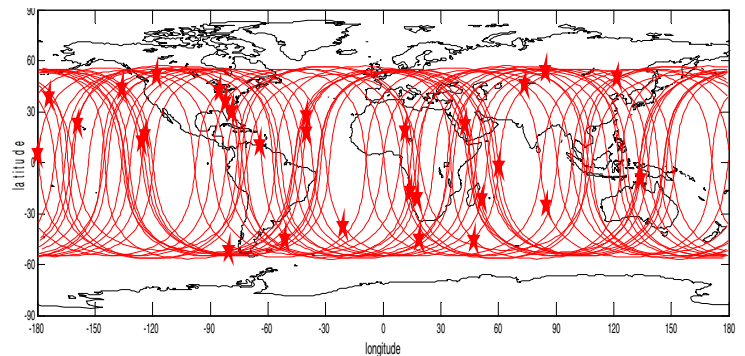


Fig. 1 (a).Satellite tracking of GPS.

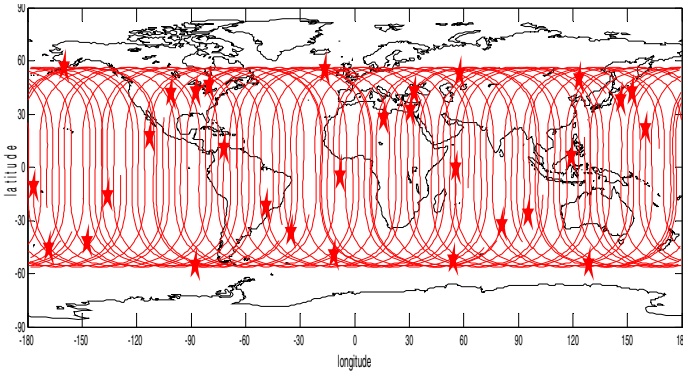


Fig. 1 (b).Satellite tracking of Galileo.

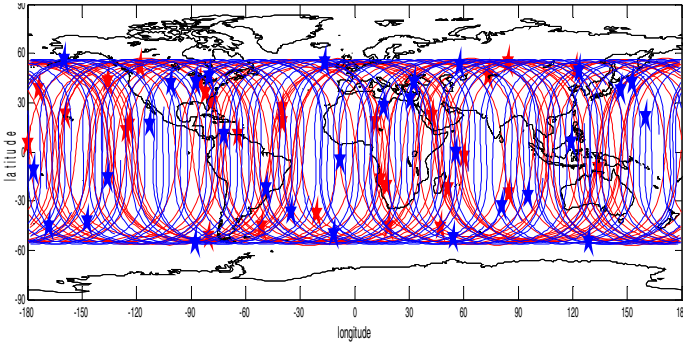


Fig. 1(c).Satellite tracking of GPS-Galileo.

In the above analysis at the same latitude and longitude the satellite tracking of GPS, Galileo and combined GPS-Galileo system is shown. From this we can come to an end that the combined system provides larger number of satellites then the individual ones and thus give better geometry of position. As we know, the number of satellites operating in the GLONASS system is quite low and Galileo system is still under development. So, in our proposed system the most well known GPS satellite system is used as we are focusing on the development of positioning technology, for establishing an efficient tracking device [3].

III. APPROACHES TO THE GPS TRACKING SYSTEM

Currently, Assisted-Global Navigation Satellite Systems (A-GNSS), or Assisted-Global Positioning Systems (A-GPS) are established as an effective technique for navigation to any destination. Now a day's A-GPS system strengthens location-based service (LBS) devices such as GPS-enabled mobile phones and personal digital assistants (PDA). Many urban environment assessments have verified that A-GPS can significantly reduce the undesired extra time for a positioning solution and tracks more satellites. Alternative positioning technologies can be used where GPS fails to direct. Therefore, wireless communication can be used as a more attractive and cheaper technique for positioning that makes it faster. Due to the widely arrayed Wi-Fi access points (Wi-Fi APs), Wireless Wide Area Network has become a well known alternative

positioning technology. Wi-Fi-based indoor positioning systems are very popular. Another Global positioning system (GPS) / Global System Mobile Communication (GSM) based System is one of the most important systems, which integrate both GSM and GPS technologies. It is necessary for many applications where both GSM and GPS systems are required. This system may be useful for communication process among any two points. Especially it is designed for users in land construction and transport business providing real-time information such as location, speed and expected arrival time of the user in a concise and easy-to-read format. We need a device that can provide location based information under indoor-outdoor environments. If we can integrate Wi-Fi & GSM as alternative location determination technologies with GPS, the system will provide higher accuracy in positioning than standalone GPS.

IV. SYSTEM DESIGN

In this design the user may be either an A-GPS or GNSS handset or an application that delivers a network of A-GPS or GNSS handsets. This system can be constituted to connect to one or more sources of GNSS data sources in order to cache it and serve it up to client's demand. The data is provided in a format that is useful for A-GNSS satellite acquisition and calculating the location of handsets [4].The proposed system can be defined into three parts:

- positioning technology,
- user interface and
- dynamic information provision (including information database structure).

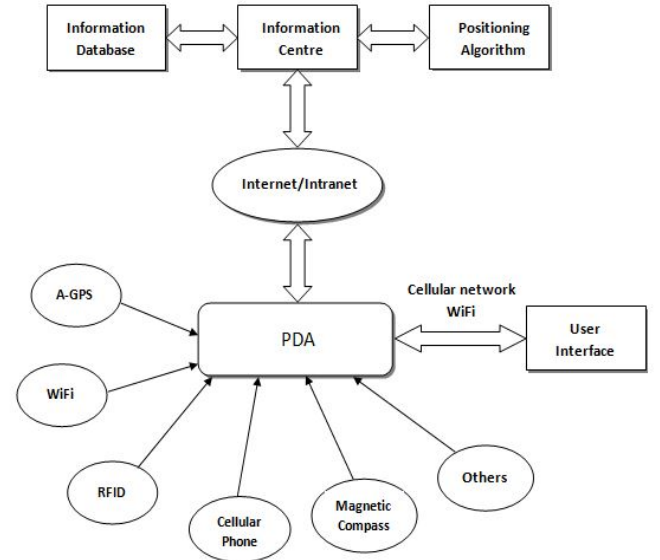


Fig. 2.The structure of the proposed system.

A. Positioning Technology

It is a major goal for most of the researchers to develop the positioning technologies for indoors and outdoors altogether. As an example, the software, Microsoft's Place Lab, permits devices such as PDAs, handheld cell phones and notebooks, to discover themselves using GSM mobile phone towers or Wi-Fi APs. The integration of the positioning technology is done by integrating several technologies such as A-GPS, Wi-Fi APs, Radio Frequency Identification (RFID), Inertial Navigation System (INS), etc. It can reduce the positioning problem to some extent. Both the A-GPS and ICT technologies are used to deliver the desired location information. A-GPS will be used in the outdoor environment, while in an indoor environment other technologies such as Wi-Fi APs or RFID will be used. A combination of all the methods may be used during the transition from one of these environments to the others. The proposed positioning system combines GPS, Wi-Fi and mobile phone positioning. Signals from Wi-Fi APs and mobile phone towers are need to trilaterate to find accurate position and on the other hand GPS is used to diagram the Wi-Fi and cellular scenery. However, the combination of this three- GPS, Wi-Fi and mobile phones is still a challenge for getting an accurate positioning and navigation.



Fig. 3. Integrate several positioning technologies to provide location.

B. User Interface

Mostly people always look for an easy-to-use Human User Interface (HUI) for submitting requests and establish a connection to a server via the cellular network or Wi-Fi. In most of the cases, human visualization is the best way to display where facilities located, how to reach them and other information linked to location. Keyboard entry, vibro-tactile, etc. are all possible interfaces.

C. Dynamic Information Provision

In this proposed system the dynamic information provision will consists of information database structure, information center and positioning algorithm. Once the location-related datas are classified, the server can be accessed via the cellular network or WiFi.

The output of this system is the data obtained from the GPS receiver and can be transmitted by using GSM module. The data contains information which is position of the user. The data from the GPS receiver will be sent using GSM module to the receiving end [5]. On the receiving end, the GSM modem will receive the data and displayed it on the screen of the mobile phone. The mobile phone also can be used to send command to the GPS receiver through the GSM module.

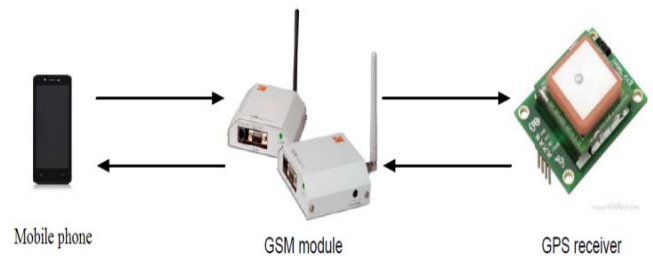


Fig.4. The proposed system using GPS/GSM based System.

Our proposed idea is to implement this system at any academic institution. Any outsider visits the institution can easily find his destination by using this tracking system. The mobile phone will act like a GPS receiver and it will guide him towards his destination showing his position in the campus. All required information will be available in the information database. The cellular network or Wi-Fi can access the server. Positioning technologies will help the individual to reach his exact location by providing precise position data and suitable user interface. We have prepared a hypothetical view on the mobile screen of this system which is implemented for the 4th floor of a building of that institution, in which the user position is indicated by a red dot. This dot will move with respect to the user position showing on the mobile screen in the form of map.

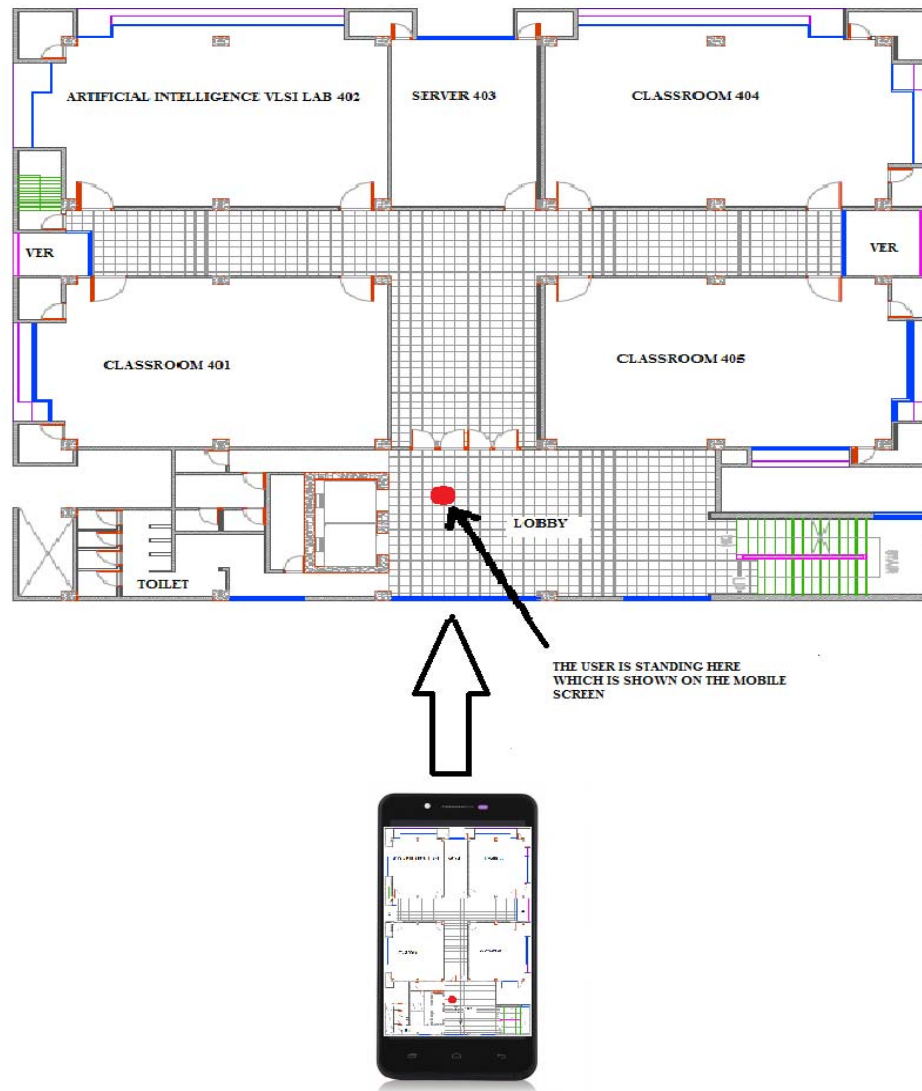


Fig.5.The hypothetical output on mobile screen.

V. CONCLUSION

As we know, the satellite tracking system is quite sophisticated in the positioning information point of view. The GPS tracking system will be very useful to the user and it will also contribute for the improvement of the accurate positioning technologies. The proposed design for navigation and information improves the freedom of movement and independency of the individual. Furthermore, an innovative design for an indoor/outdoor positioning and information purpose can be generated through this proposal. The goal of this paper is to develop navigation services for mobile networks. Combination of GPS-Galileo system can also be used in future as it is now only a theoretical idea. Further development in this tracking field can upgrade this proposed idea for the betterment of positional accuracy.

REFERENCES

- [1] E. D. Kaplan and C. J. Hegarty, "Understanding GPS: principles and applications," 2nd ed., Artech House Press, Boston, 2006.
- [2] Ahmed El-Rabbany, "Introduction to GPS-The Global Positioning System," Artech House Boston . London.
- [3] Pasi Kamppi, Jyri Rajamaki, Robert Guinness, "Information security risks for satellite tracking," International journal of computers and communications, issue 1, vol. 3, 2009.
- [4] Binghao Li, Euan Ramsey-Stewart, Kenny Johar, "More freedom for the Blind and Vision Impaired – A Proposed Navigation and Information System," International Global Navigation Satellite Systems Society IGNSS Symposium 2009.
- [5] J.C. Juang C.W. Jang, "Failure detection approach applying to GPS autonomous integrity monitoring," IEE Proc.- Radar.Sonar.Navig.vol. 145, no. 6. December 1998.

Modeling of an Efficient Microstrip Patch Antenna for Microwave Brain Imaging System

Reefat Inum, *Student Member, IEEE*, Md. Masud Rana, *Member, IEEE* and Md. Anwarul Quader

Department of Electrical and Electronic Engineering

Rajshahi University of Engineering and Technology

E-mail: romel.eee09@gmail.com, md.masud.rana@ruet.ac.bd, akshuvo10@gmail.com

Abstract—In this paper, we consider a system model for microwave brain imaging. As an important part of the considered system model, the radiating element is designed in the form of a microstrip patch antenna. To enhance the antenna performance, we propose two electromagnetic band gap (EBG) structures namely circular and rectangular, on the antenna ground plane. A comparison of the antenna performance parameters for the rectangular and circular EBG structures is presented. A six layer human head phantom model is designed which is located at 25 mm distance from the antenna. The specific absorption rate (SAR) of the modeled head tissue for the proposed microstrip patch antenna with circular EBG is determined using CST MWS software. The different SAR values are compared with the established standard SAR limit to provide a safety regulation of the considered microwave brain imaging system.

Keywords—microwave brain imaging, patch antenna, EBG, human head phantom model, SAR

I. INTRODUCTION

As a non-invasive or complementary non-ionizing diagnostic tool, microwave based head investigation has recently attracted a huge attention for various biomedical applications [1]-[5]. The conventional brain imaging tools like MRI, CT scan etc. are reported as expensive, time consuming, having static or bulky structures and employing ionization radiation [6]. To overcome these difficulties, researchers find motivation to explore new diagnosis technology employing microwave frequency spectrum.

However, two main aspects of any microwave brain imaging system are the efficient design of the transmitting/ receiving antenna and biocompatibility analysis. So far in the literature, different types of antenna structure [7]-[9] have been investigated as the radiating element in the microwave based imaging system. The microstrip patch antenna claims extra attention to be used in the microwave brain imaging system as it acquires some unique characteristics [10]. But the conventional patch antenna needs some modifications to be more efficient in the application. Incorporation of electromagnetic band gap (EBG) structure on the antenna ground plane is found in the literature to be an effective way

to enhance the performance of conventional microstrip patch antenna. The EBG structures can be categorized as a special type of metamaterial due to their unique band gap features. This unique property has been applied to design antenna systems with better gain and efficiency, lower side-lobes and back-lobe levels by suppressing surface wave modes [11], [12]. The effect of using EBG structure is found to be the reduction of the patch resonant length and the bandwidth as a result of stronger capacitive coupling.

The second aspect of a microwave brain imaging system is the assessment of specific absorption rate (SAR) of a suitable human head phantom model. This part of the imaging system must be considered carefully to ensure the safety of the patient under test. As the accurate modeling of human head consisting different types of complex tissue is cumbersome, various head phantom models have been investigated to mimic the practical scenario [13]-[15].

In this paper, we present an efficient design of a rectangular microstrip patch antenna by incorporating rectangular and circular EBG structures on the antenna ground plane. The performance of the proposed antenna with and without EBG structures is investigated for the considered brain imaging application. To determine the SAR value of the head tissue caused by the microwave EM radiation from the designed antenna, a six layer spherical human head phantom model is designed. The six different layers of the head model are the brain, CSF, dura, bone, fat, and skin. A feasible dimension of the head model is chosen that reflects the actual one. The SAR values for three different measuring techniques are compared with each other and with the established standard SAR limit.

II. SYSTEM MODEL FOR MICROWAVE BRAIN IMAGING

Microwave brain imaging system is based on the concept that a microwave signal is emitted from a transmitting antenna, which after being reflected from the target specimen is received by the receiving antenna. The received signal is stored and a suitable signal processing algorithm makes the image possible. The following Fig. 1 depicts the concept of our considered brain imaging system. In this system model, we specifically focus on the effective antenna modeling and

determination of SAR values for six layers human head model.

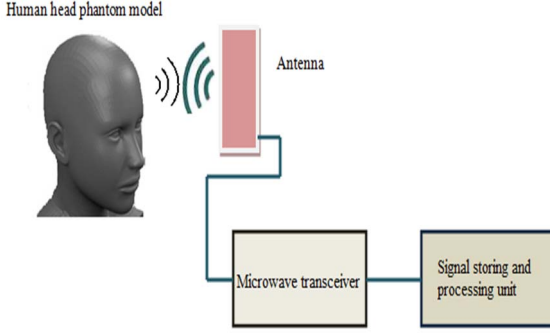


Fig. 1. Microwave brain imaging system model

A. Patch Antenna Design

Herein we design a microstrip line feed rectangular patch antenna. The antenna resonant frequency is taken as 7.3 GHz.

The following equations are used to design the rectangular patch [16].

Step-1: Patch width

$$w = \frac{v_0}{2f_r} \sqrt{\frac{2}{\epsilon_r + 1}} \quad (1)$$

Where, v_0 = The velocity of light.

ϵ_r = Dielectric constant of substrate.

Step-2: Effective Dielectric constant of the rectangular microstrip patch antenna.

$$\epsilon_{eff} = \frac{\epsilon_r + 1}{2} + \frac{\epsilon_r - 1}{2} \left(\frac{1}{\sqrt{1 + \frac{12h}{w}}} \right) \quad (2)$$

Step-3: Patch Length

$$L = L_{eff} - 2\Delta L \quad (3)$$

$$L_{eff} = \frac{c}{2f_r \sqrt{\epsilon_{eff}}} \quad (4)$$

Calculation of Length Extension:

$$\Delta L = 0.412h \frac{(\epsilon_{eff} + 0.3) \left(\frac{w}{h} + 0.264 \right)}{(\epsilon_{eff} - 0.258) \left(\frac{w}{h} + 0.8 \right)} \quad (5)$$

Step-4: Inset Feed

$$50 = R_{in} \left(\cos\left(\frac{\pi}{L} y_0\right) \right)^2 \quad (6)$$

Where, R_{in} is the input impedance at the leading radiating edge of the patch and 50 (Ω) is the desired impedance.

The complete design of the rectangular patch antenna with all required dimensions is shown in Fig. 2. Rogers R03003 is chosen as antenna substrate due to its flexibility for high frequency application compared to FR4 substrate. The

choice of appropriate substrate thickness plays an important role in the patch antenna design. The antenna efficiency and bandwidth can be increased by increasing the substrate height, but at the cost of introducing surface waves which extract power from the total available for direct radiation [16].

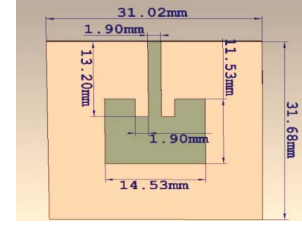


Fig. 2. Patch antenna without EBG.

B. EBG Structure Design

The wavelength of the microwave signal radiated by the proposed antenna is used to design the EBG structures. The effect of EBG structure is found to be the reduction of patch length, substrate length, inset gap and an increase in substrate height. Following are the equations required to design the proposed EBG structures.

$$L = 0.10\lambda \quad (7)$$

$$D = 0.02\lambda \quad (8)$$

$$r = 0.10\lambda \quad (9)$$

Where, L is the length and r is the radius of the rectangular and circular EBG structures respectively and D is the distance between two shapes in any case as evident from Fig. 3. Nicholson-Ross-Weir (NRW) technique is used as a conversion approach in designing the proposed EBGs. Both permeability and permittivity is obtained from S-parameter using this approach [17].

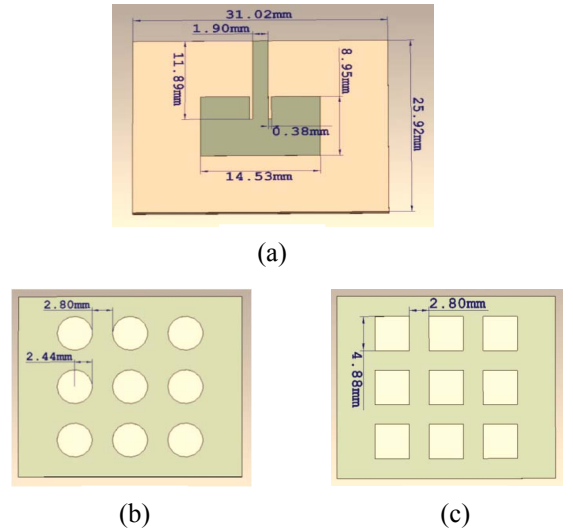


Fig. 3. (a) Patch antenna with EBG, (b) Geometry of circular EBG, (c) Geometry of rectangular EBG.

The initial dimension parameters of the patch antenna are obtained from the established equations. However, after incorporating the EBG structure on the antenna ground plane, the dimension parameters are slightly optimized to obtain enhanced performance. The effect of optimization is found to be the reduction in patch length, substrate length, and inset gap. Hence incorporation of the EBG structure makes the antenna more compact. Also the surface wave introduced in the antenna is sufficiently reduced due to the incorporation of EBG structure on the antenna ground plane.

C. Human Head Modeling

The dielectric properties, i.e. permittivity and conductivity of human body tissues vary with frequency, geometry, size of tissue, water contents and age. Since the SAR measurements and EM field evaluation inside the human body are quite difficult, phantom models of different shapes have been designed in the literature to conduct the desired EM measurement for any imaging system. A schematic representation of the designed spherical head phantom is shown in the following Fig. 4.

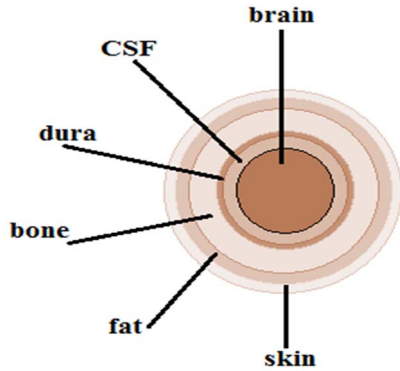


Fig. 4. Six layers of human head phantom model.

The dimension of each layer and their dielectric properties at 7.3 GHz is tabulated below.

Table I
Dimensions and electrical properties of head tissues

Tissue	Radius (mm)	Permittivity (ϵ_r)	Conductivity, σ (S/m)
Brain	81	43.22	1.29
CSF	83	70.1	2.3
Dura	83.5	46	0.9
Bone	87.6	5.6	0.03
Fat	89	5.54	0.04
Skin	90	45	0.73

To simplify the simulation study, a six layer specific anthropomorphic mennequin (SAM) phantom head is modeled considering the above realistic dielectric tissue properties in the frequency band of interest. Though anatomically non-realistic, as is often the case found in

literature, a spherical head phantom can be considered as a feasible one for SAR evaluation in the simulation environment. Patch antenna with circular EBG structure located at 25 mm distance from the head model is illustrated in the following Fig. 5.

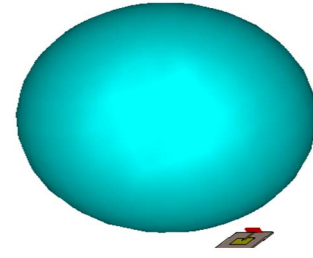


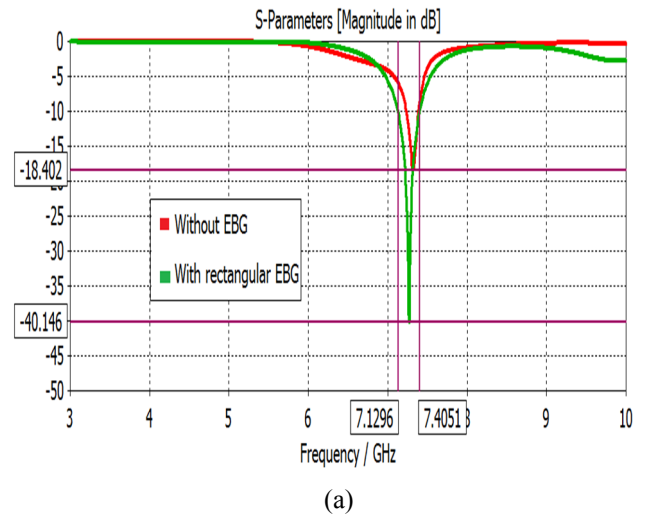
Fig. 5. Patch antenna and human head modeling in the CST MWS software.

III. SIMULATION RESULTS

Finite element method (FEM) based CST MWS software is used to simulate the proposed antenna and to evaluate the SAR values of the head tissues. The antenna performance parameters to be considered are the return loss, gain, and bandwidth. Biocompatibility analysis of the phantom head model is carried out by measuring the 1g average, 10g average, and point SAR. The averaging volume for maximum SAR is also plotted to clearly locate the maximum SAR position.

A. Antenna Return Loss, Bandwidth and Gain

The s-parameters of any antenna must be at least -10 dB to ensure the minimum reflection from antenna to the source. Particularly in a brain imaging application, the return loss must be as lowest as possible so that sufficient amount of power can be radiated to the target specimen.



In microwave imaging, the resolution of the reconstructed image mostly depends on the magnitude of the

scattered signal from the target which in turn depends on the amount of total radiated power from the transmitting antenna.

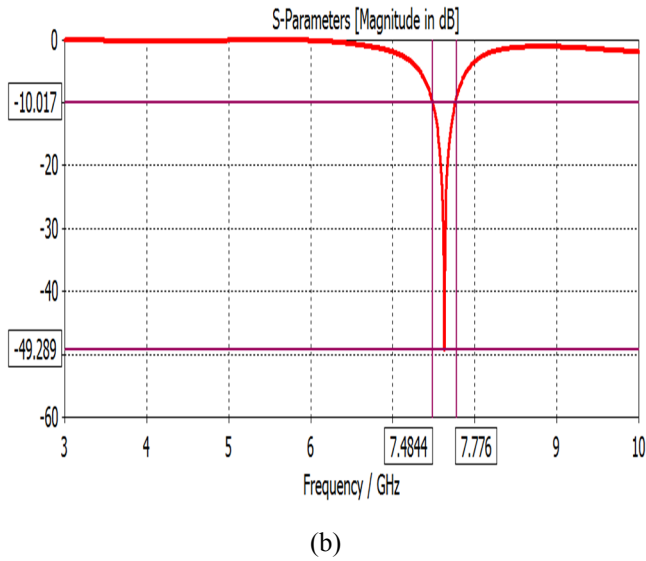


Fig. 6 (a) return loss of the patch antenna without EBG and with rectangular EBG. (b) antenna return loss with circular EBG.

Hence to ensure the good image quality, the return loss of the proposed antenna should be minimized. Fig. 6 (a) represents the return loss of the designed patch antenna without and with rectangular EBG structure. It's evident from the figure that rectangular EBG structure on the antenna ground plane improves the return loss from -18.402 dB to -40.146 dB. Fig. 6 (b) reveals a further enhancement in the reflection coefficient to -49.289 dB which is made possible by the incorporation of circular EBG structure on antenna ground plane. Bandwidth is another antenna performance parameter which needs to be taken into account for an appropriate imaging system. Ultra wideband antenna is always preferable in microwave brain imaging for a reliable system operation. In case of bandwidth, patch antenna without the EBG structure provides 156.20 MHz, with rectangular EBG structure provides 275.50 MHz, and with circular EBG structure provides 291.60 MHz. Hence the introduction of circular EBG structure perfectly improves the antenna performance in terms of return loss and bandwidth.

The polar plot of directivity and 3-D plot of gain of the patch antenna with circular EBG structure are shown in Fig. 7. Information obtained from Fig. 7 (a) indicates that the proposed patch antenna with circular EBG structure provides a directivity of 6.77 dBi in the direction of $\phi = 31^\circ$. The values of 3 dB beamwidth and side lobe level found from the same figure are 92.6° and -8.9 dB respectively. The increased beamwidth will ensure the resolution capabilities of the antenna while reduced side lobe level is an indication of greater radiation intensity in the main lobe.

Table II
Performance comparison of the proposed antenna with other existing antennas

Parameter	Proposed antenna	[20]	[11]	[21]
Return Loss (dB)	-49.289	-18	-22	-17.50
Gain (dB)	6.59	-	-	5.01
Bandwidth (MHz)	291.60	190	-	-
Directivity (dBi)	6.77	2.5	2.2	-

Gain is another useful measure to describe the performance of an antenna as it takes into account the efficiency as well as directional capabilities of the antenna. The simulated gain of the proposed antenna as shown in Fig. 7 (b) is 6.59 dB which is in good agreement to be used in brain imaging application.

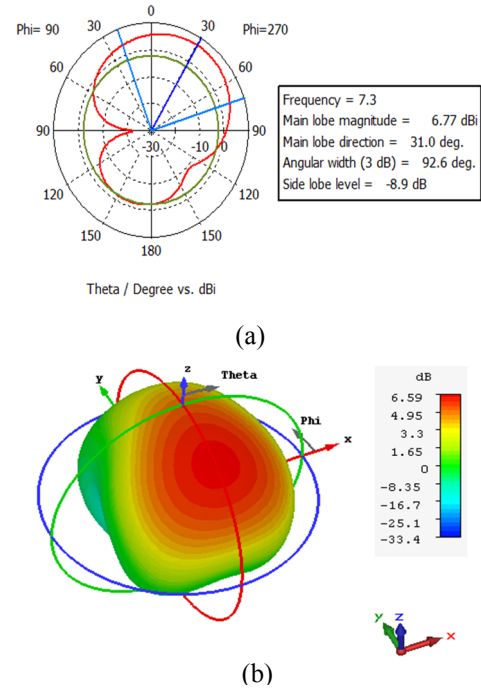


Fig. 7. (a) polar plot of directivity and (b) 3-D plot of gain of the patch antenna with circular EBG.

Finally, a comparison between the performance parameters of the proposed antenna and similar other existing antennas is given in Table II. Similar other antennas mean those designed incorporating any metamaterial structure. It's obvious from the comparison that the proposed microstrip patch antenna with circular EBG structure provides enhanced performances which can be a reliable candidate for microwave brain imaging.

B. Specific Absorption Rate (SAR)

The safety regulation of microwave based brain imaging

system must be maintained by proper assessment of SAR as a brief exposure to radiation may cause severe health hazard. SAR is a measure of the deposition of electromagnetic energy over time into human body tissue. The units are generally watts per kilogram of body mass. In mathematical form, it is defined as [18]:

$$\text{SAR} = \frac{\sigma}{\rho} |E|^2 = \frac{J^2}{\rho \sigma} \quad [\text{W/kg}] \quad (10)$$

Where, E is the rms value of the electric field strength in the tissue [V/m], J is the current density [A/m], σ is the conductivity of the head tissue [S/m], and ρ is the density of head tissues [kg/m³].

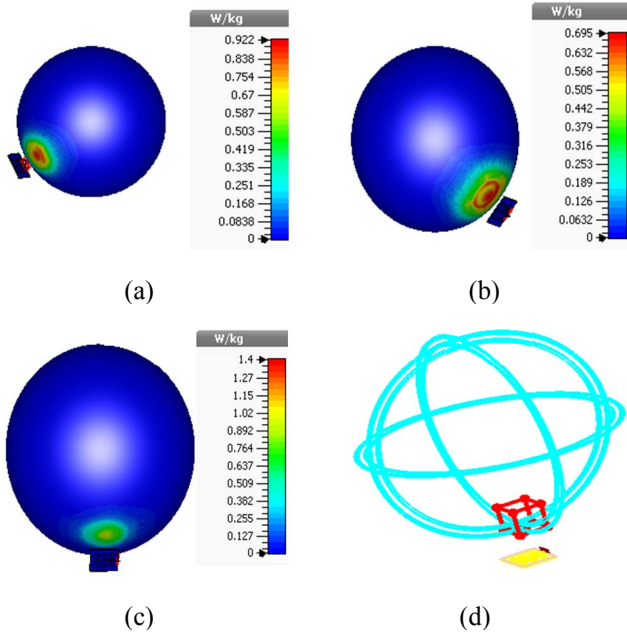


Fig. 8. (a) 1g average SAR, (b) 10g average SAR, (c) point SAR, and (d) 10g averaging volume for maximum SAR value for patch antenna with circular EBG structure.

Different techniques are available to assess the SAR value of human body tissue when exposed to EM radiation. The antenna type, operating frequency, and distance between the antenna and body tissue are the key parameters that affect the determination of SAR. Depending on national reporting, testing requirement, and the network band, the international SAR limit may vary. Averaged on 1g of tissue, the SAR limit set by US is 1.6 W/kg while in EU, the limit is 2.0 W/kg, averaged on 10g of tissue [19]. The 3-D plot obtained by numerical method based EM simulation of SAR distribution in the six layered head phantom model is depicted in Fig. 8. For mass averaged SAR, the power loss density is integrated over a cube with a defined mass, typically 1g or 10g. The integral power loss is then divided by the cube's mass. Thus the 1g average and 10g average SAR values are evaluated which are evident from Fig. 8(a) and 8(b) respectively. It's also obvious from Fig. 8(a) and 8(b) that the 3-D maximum SAR values are 0.922 W/kg and 0.695 W/kg, which are less than maximum standard limit as

1.6 W/kg and 2.0 W/kg, averaged on 1g and 10g of tissue respectively. Fig. 8(c) represents the point SAR which is local SAR without mass or volume averaging. In all cases, the reference stimulated power is taken as 0.5 W. In Fig. 8(d), 10g averaging maximum SAR cube is visualized. It's clear from Fig. 8 that the SAR value is higher near the antenna and then rapidly decreases to zero after a short distance. Hence, this microwave brain imaging system employing circular EBG based microstrip patch antenna maintains the safety regulation.

IV. CONCLUSION

A brain imaging system working in microwave frequency range is envisaged. As two important parts of the brain imaging system, modeling of human head phantom and efficient microstrip patch antenna is focused. Appropriate tissue properties including permittivity and conductivity at the antenna resonant frequency is chosen to model the head phantom. Also, a spherical shape having six layers namely brain, CSF, dura, bone, fat, and skin is considered to replicate the realistic human head model. In designing the antenna as the transmitting and receiving element of the system model, at first a simple microstrip patch antenna is devised whose performance is later improved by slotting EBG structure in the antenna ground plane. Patch antenna with circular EBG structure shows decent improvement in the antenna performance parameters in comparison with the rectangular EBG. The simulation results depict that incorporation of circular EBG on the antenna ground plane provides an improvement of 22.77% in return loss, 5.84% in bandwidth, and 16.53% in antenna gain with respect to the patch antenna with rectangular EBG. However, the limitation of this design is that the patch antenna without EBG structure presents higher gain than that with either EBG structure, yet 6.59 dB for circular EBG structure is taken as enough for the considered brain imaging system. Thus the patch antenna with circular EBG structure is accepted as a perfect candidate for the imaging system. The total SAR value found in the six layered head tissue stimulated by the circular EBG based patch antenna is 0.001765 W/kg. The 1g and 10g mass averaged SAR and local SAR values are also evaluated which are much below the internationally standard SAR limits. Hence the envisaged microwave brain imaging system properly maintains the safety regulation and can be safely used to diagnose human brain diseases.

REFERENCES

- [1] M. Persson *et al.*, "Microwave-based stroke diagnosis making global prehospital thrombolytic treatment possible," *IEEE Trans. Biomed. Eng.*, vol. 61, pp. 2806-2817, Nov. 2014.
- [2] S. Semenov, B. Seiser, E. Stoeckmann and E. Auff, "Electromagnetic tomography for brain imaging: From virtual to human brain," *IEEE Conf. Antenna Measurements & Applications*, pp. 1-4, Nov. 2014.
- [3] X. P. Li *et al.*, "The dynamic dielectric at a brain functional site and an EM wave approach to functional brain imaging," *Sci. Rep.* 4, article number. 6893, Nov. 2014.

- [4] J. J. Korfhagen, M. A. Kandadi, J. F. Clark, O. Adeoye, and G. J. Shaw, "A prototype device for non-invasive continuous monitoring of intracerebral hemorrhage," *J. Neurosci. Methods*, vol. 213, pp. 132-137, Feb. 2013.
- [5] B. J. Mohammed, A. M. Abbosh, S. Mustafa, and D. Ireland, "Microwave system for head imaging," *IEEE Trans. Instrum. Meas.*, vol. 63, pp. 117-123, Jan. 2014.
- [6] R. Scapaticci *et al.*, "A feasibility study on microwave imaging for brain stroke monitoring," *Prog. Electromag. Research B*, vol. 40, pp. 305-324, 2012.
- [7] M. T. Islam *et al.*, "A negative index metamaterial-inspired UWB antenna with an integration of complementary SRR and CLS unit cells for microwave imaging sensor applications," *Sensors*, vol. 15, pp. 11601-11627, May. 2015.
- [8] A. T. Mobashsher and A. Abbosh, "CPW-fed low profile directional antenna operating in low microwave band for wideband medical diagnostic system," *Electron. Lett.*, vol. 50, pp. 246-248, Feb. 2014.
- [9] M. Klemm, I. J. Craddock, A. Preece, J. Leendertz and R. Benjamin, "Evaluation of a hemi-spherical wideband antenna array for breast cancer imaging," *Radio Sci.*, vol. 43, pp. 1-15, 2008.
- [10] I. Singh *et al.*, "Microstrip patch antenna and its applications: a survey," *Int. J. Comput. Applicat. Tech.*, vol. 2 (5), pp. 1595-1599, 2011.
- [11] J. Liang, and H. Y. D. Yang, "Microstrip patch antennas on tunable electromagnetic band-gap substrates," *IEEE Trans. Antennas Propag.*, vol. 57, pp. 1612-1617, Jun. 2009.
- [12] M. F. Abedin, M. Z. Azad, and M. Ali, "Wideband smaller unit-cell planar EBG structures and their application," *IEEE Trans. Antennas Propag.*, vol. 56, pp. 903-908, Mar. 2008.
- [13] M. R. I. Faruque *et al.*, "Effect of human head shapes for mobile phone exposure on electromagnetic absorption," *InformacijeMIDEM*, vol. 40, pp. 232-237, 2010.
- [14] M. S. M. Said, N. Seman, and H. Jaafar, "Characterization of human head phantom based on its dielectric properties for wideband microwave imaging application," *J. Teknologi (Sci. Eng.)*, vol. 73, no. 6, pp. 43-49, Apr. 2015.
- [15] S. Mustafa, A. M. Abbosh and P. T. Nguyen, "Modeling human head tissues using fourth-order Debye model in convolution-based three dimensional finite-difference time-domain," *IEEE Trans. Antennas Propag.*, vol. 62, pp. 1354-1361, Mar. 2014.
- [16] C. A. Balanis, "Antenna theory analysis and design," *A John Wiley & Sons, Inc.*, 3rd edition, 2005.
- [17] C. Y. Kuek "Measurement of dielectric material properties," *Rohde & Schwarz Application Note, RAC0607-0019*, 2006.
- [18] M. K. Hosain *et al.*, "Development of a compact rectenna for wireless powering of a head-mountable deep brain stimulation device," *IEEE J. Transl. Eng. Heal. Medic.*, vol. 2, Jan. 2014.
- [19] Antenna-Theory.com, "Specific Absorption Rate (SAR)," 2014. [Online]. Available: <http://www.antennatheory.com/definitions/sar.php>
- [20] A. S. Bhadouria, and M. Kumar, "Microstrip patch antenna for radiolocation using DGS with improved gain and bandwidth," *IEEE Int. Conf. Advances Eng. Tech. Research*, pp. 1-5, Aug. 2014.
- [21] S. E. Mendhe, and Y. P. Kosta, "Gain enhancement and broadband using helical resonating metamaterial superstrate in stacked microstrip patch antenna," *Microw. Optic. Tech. Lett.*, vol. 56, no. 9, pp. 1978-1982, Sep. 2014.

Design and Performance Study of Optical Logic Gates Using Ti:KNbO₃ Based Waveguide at 1.55 μm

Rabeya Khatun, A. Zahed Chowdhury

Department of Applied Physics, Electronics & Communication Engineering

University of Chittagong

Chittagong-4331, Bangladesh

shompa.apece@gmail.com, zahed_apece@cu.ac.bd

Abstract—Optical logic gates have become the most essential means of performing optical signal processing functions. These gates have potential for the development of large scale integrated optical processor. In this paper we tried to bring out the best performance of AND, NOT, NAND, NOR, XOR, XNOR logic gates based on electro-optic effect by implementing these devices with the use of Ti:KNbO₃ waveguide. To implement these gates we used an optimal performing 2x2 Mach-Zehnder Interferometer electro-optic switch based on Ti:KNbO₃ waveguide as basic building block. Two or more switches were needed for designing the gates. Signals in second electrodes which are also referred as control signals are the inputs of the gates. The novelty of this work is to introduce AND, NOT, NAND, NOR, XOR, XNOR logic gates using Ti:KNbO₃ based waveguides at 1.55 μm. Finite difference beam propagation method (FDBPM) was used as the performance analyzing tool of these gates with the help of Opti-BPM 12.2 simulation software.

Keywords—Optical logic gates; Mach-zehnder Interferometer; Opti-BPM; Electro-Optic; FDBPM.

I. INTRODUCTION

Larger communication bandwidth, higher transmission speed and lower cost are the most demanding performance factor in the world of communication. Due to incapability of handling larger communication traffic with limited bandwidth availability, optical communication is taking over electronic media. To handle such large number of user, there is a vast need of building larger communication network. Logic gates, signal router, multiplexer, modulator, all-optical switch, filter, converter etc. are basic building blocks in optical communication. For the development of the most demanding optical communication, these building blocks have to be updated, smaller sized and lower loss oriented to cope up with upcoming technologies. Photonic computing uses photons which are produced by lasers or diodes for the purpose of computation. For several decades, photons have promised to allocate higher bandwidth than the electrons which are used in conventional computers [1-3]. Although in the past various researchers have shown their intense interest in designing a number of optical logic gates (for example XNOR, XOR, AND, OR, NAND, NOR and NOT) using different optical processing techniques [4-9], but device with smaller length, lower power requirement and higher speed are in interest for further investigation. MZI can play a major role in the field of

ultra-fast optical signal processing [7, 9-10] up to 10 Gbit/s [11-14]. We can find the XOR using semiconductor optical amplifier based MZI gate being analyzed by the use of numerical simulations, in order to specify the optimal parameters along with the practical conditions for operation in [15]. Mach-zehnder interferometer has also been used broadly to implement several logical devices (XOR, AND, XNOR), adder/subtractor, multiplexer, flip-flop and encoder using SOA, when using SOA there is limitation in speed (10 to 20 Gb/s) due to carrier lifetime [15–19]. Quantum dot has also been used to realize XOR, XNOR and NOT gate up to 160 Gb/s, but these require complicated loss compensation method reducing designer flexibility [6-7, 9]. In 2014-2015 A. Kumar proposed novel way of implementing AND, NAND, NOR, XOR and XNOR gate using Ti:LiNbO₃ based waveguide at cost of higher switching voltage of 8 V [20-21]. In this paper we are about to show implementation of AND, NOT, NAND, NOR, XOR and XNOR gate using Ti:KNbO₃ based waveguide as it shows better performance along with low power requirement (4 V), low insertion loss (<0.04 dB), high electro-optic coefficient, high speed of transmission (>40 Gbit/s), easy availability in nature and can be implemented using available fabrication method [22]. Section II presents mathematical explanation of these logic gates, where in section III design and results of simulations along with comparative study with respect to previous researches are shown. At last we tried to draw a conclusive study.

II. MATHEMATICAL IDEA OF OPTICAL LOGIC GATE USING TWO OR MORE CASCADED SWITCH

XOR, XNOR, AND, NOT, NOR and NAND gates can be implemented using two or more switch simply by cascading them. Let us consider the case of two cascaded MZIs as in Fig. 1. Mathematical formulation of normalized power at different output ports for two and three cascade switches can be found by means of mathematical idea of single MZI configuration in [23], for example output at port 1 of MZI2 is given by [21],

$$OUT1_{MZI2} = \left\{ j e^{-j(\phi_0 MZI1)} \cos\left(\frac{\Delta\phi_{MZI1}}{2}\right) \right\} \times \left\{ j e^{-j(\phi_0 MZI2)} \cos\left(\frac{\Delta\phi_{MZI2}}{2}\right) \right\} E_{in} \quad (1)$$

$$OUT1_{MZI2} / E_{in} = \left\{ j e^{-j(\phi_0 MZI1)} \cos\left(\frac{\Delta\phi_{MZI1}}{2}\right) \right\} \times \left\{ j e^{-j(\phi_0 MZI2)} \cos\left(\frac{\Delta\phi_{MZI2}}{2}\right) \right\} \quad (2)$$

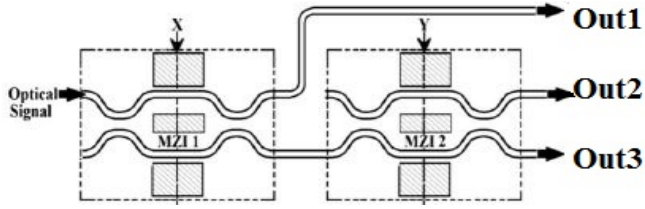


Fig. 1. Structure of two cascaded MZIs [21].

$$m_0 = |OUT1_{MZI2} / E_{in}|^2 = \cos^2\left(\frac{\Delta\phi_{MZI1}}{2}\right) \cos^2\left(\frac{\Delta\phi_{MZI2}}{2}\right) \quad (3)$$

For the calculation of equation 3, it has been assumed that,

$$\left. \begin{aligned} \phi_{0MZI1} &= \frac{\phi_{1MZI1} + \phi_{2MZI1}}{2} \left\{ \begin{aligned} \Delta\phi_{MZI1} &= \phi_{1MZI1} - \phi_{2MZI1} = \frac{\pi}{V_\pi} X \\ \phi_{0MZI2} &= \frac{\phi_{1MZI2} + \phi_{2MZI2}}{2} \left\{ \begin{aligned} \Delta\phi_{MZI2} &= \phi_{1MZI2} - \phi_{2MZI2} = \frac{\pi}{V_\pi} Y \end{aligned} \right. \end{aligned} \right\} \quad (4)$$

Similarly we can find outputs of different ports for the MZIs used to build the logic gates. ϕ_{1MZI1} and ϕ_{1MZI2} represent the phase angles produced on the upper arm of MZI1 & MZI2 respectively. Again ϕ_{2MZI1} and ϕ_{2MZI2} present the phase angles produced on the lower arm of MZI1 & MZI2 respectively. MZI1 can be controlled by signal X which is the voltage applied at the second electrode while keeping the other two electrodes at the zero potential. In the same way, MZI2 can be controlled by control signal Y. Basically, the values of the control signals are assumed as 0 (0 V) & 1 (4 V) and these are supplied at the control (2nd) electrodes of each MZI as we used Ti:KNbO₃ based MZI switch. Therefore, we can present the normalized power at the 1st output port of MZI2 by equation 3. Similarly we can derive equation for OUT2 and vice versa. The proper combination of min-terms provides the expression of the XOR, XNOR, AND, NOT, NOR and NAND logics.

III. DESIGN AND SIMULATION OF PHOTONIC LOGIC

Before implementing our proposed logic gates, we designed a single 2x2 MZI electro-optic switch using Ti:KNbO₃ waveguide by following the same procedure as it was described in ref [22] with the help of Opti-BPM 12.2 software. Then we evaluated the best possible performance of 2x2 MZI switch based on electro-optic effect. The optimization was done by following the same procedure as in our previous paper ref [24], i.e. we calculated lowest insertion loss along with highest extinction ratio for six parameters of the switch which are strip thickness of titanium, length of interferometer arm and arm gap, width of waveguide, separation in directional coupler (3 dB), then finally location of second electrode above original axis of the switch. The ultimate values of insertion loss (I.L.) and extinction ratio (Ex.R.) are 0.00925 dB and 32.157 dB respectively for both cross state (0 V) and bar state (4 V) at 1.55 μ m [25], where in [22] maximum extinction ratios are 29.4 dB and 26.6 dB & minimum insertion losses were 0.032 dB and 0.0261 dB for cross state (0 V) and bar state (4 V) respectively found with a

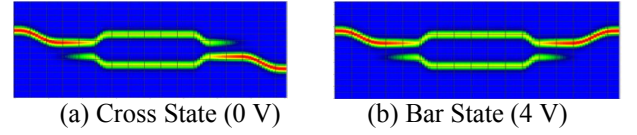


Fig. 2. Optical field propagation of 2x2 MZI electro-optic switch.

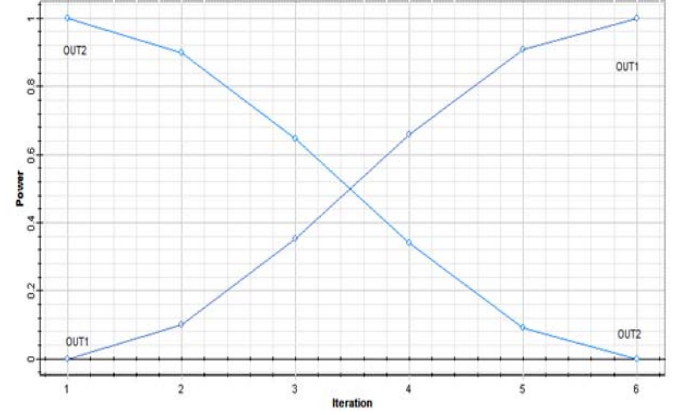


Fig. 3. Power in output waveguides.

thickness of 0.0525 μ m at 1.3 μ m operating wavelength. So, our proposed switch shows better performance than that of [22], as I.L. reduces 71.09% in 0 V & 64.56% at 4 V and Ex.R. also increases 9.38% at 0 V and 20.89% at 4 V. Again we can say that performance is better at 1.55 μ m. Fig. 2 shows optical field propagation of this optimized switch and Fig. 3 presents optical power distribution in output waveguides. Finally the optimized structure of the Ti:KNbO₃ based 2x2 MZI electro-optic switch includes,

- Strip thickness of Titanium: 0.0825 μ m
- Waveguide width: 8 μ m
- Arm length: 9900 μ m, Arm gap: 32 μ m
- Gap in coupling region: 14.5 μ m
- Center position of second electrode: 5.5 μ m
- Length & width of switch: L=32900 μ m, W=100 μ m

In the next subsections we are going to present the structure and operation of AND, NOT, XOR, XNOR, NAND and NOR gates.

A. AND & NOT Logic Gates

To construct AND gate the number of MZIs needed is 2 and its layout is shown in Fig. 4. This structure can also be used as NOT gate at the same time. The optical input signal is provided at input port 1 of 1st MZI. There is a joint between output port 1 of first Mach-Zehnder Interferometer and input port 1 from the 2nd MZI. Signal in second electrodes of MZI1 and MZI2 are defined as X and Y. So, if control signal X and Y is in the high logic, we will find optical signal appearing in the 1st port of 2nd Mach-Zehnder Interferometer. For each and every other combination of X and Y, the signal is not switched to output port 1 from the 2nd Mach-Zehnder Interferometer. Therefore, output port 1 from the 2nd Mach-Zehnder Interferometer operates like AND gate.

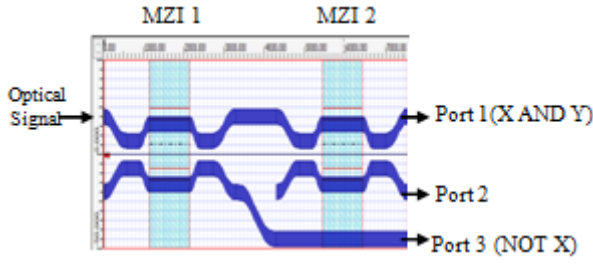


Fig. 4. Layout of AND & NOT gates.

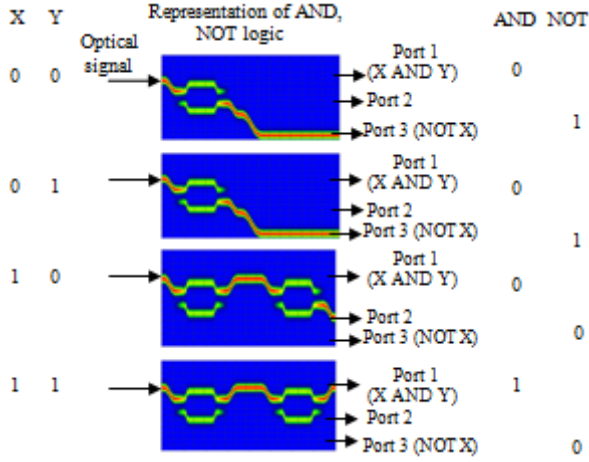


Fig. 5. Representation of AND and NOT logic observed from Opti-BPM.

TABLE I. TRUTH TABLE FOR PROPOSED AND & NOT GATE

Control Signal X	Control Signal Y	Port 1 (X AND Y)	Port 2	Port 3 (NOT X)
0	0	0	0	1
0	1	0	0	1
1	0	0	1	0
1	1	1	0	0

Results for the combinations of X and Y are listed in table I, e.g. with X high optical power goes to input port 1 of MZI2, then with Y high optical power goes to out1 of MZI2. From table I we can see that output at port 3 acts as NOT gate of signal X. Device dimension are ; Length $L=75800 \mu\text{m}$, Width $W=100 \mu\text{m}$, Center of second electrodes of both MZI is $5.5 \mu\text{m}$. In table I, 0 represents 0 V and 1 represents 4 V. Fig. 5 represents the result found from layout of Fig. 4. It also represents that port 1 of our proposed device operates as an AND logic gate for the various combination of the control signals X and Y. Now port 1 will be high, if both control signals X and Y acquire 4 V at second electrodes of both MZI, which is supposed as logic '1' for this gate. Output powers in every case of high state in table I are 0.992529 mW (insertion loss - 0.033 dB). If we compare our proposed switch to the previous one in ref [20], length of the device are not same as length in ref [20] is $76000 \mu\text{m}$ (because MZIs with $33000 \mu\text{m}$ is used to build the gate), whereas our proposed one is $75800 \mu\text{m}$ in length. Again operating wavelength are different and by comparing optical field propagation we can see that in figure 10 of ref [20] (shown in table II) there is some power leakage which can be observed by the disconnected dark shadowed line in optical field, where in our proposed switch it is almost absent. Table II shows a comparative study discussed above.

TABLE II. COMPARATIVE STUDY ON AND GATE

Specification	Proposed AND gate	Existing [20]
Length	$75800 \mu\text{m}$	$76000 \mu\text{m}$
Operating Wavelength	$1.55 \mu\text{m}$	$1.3 \mu\text{m}$
Switching Voltage	4 V for Ti:KNbO ₃ based switch	6.75 V with Ti:LiNbO ₃ based switch
Applications	NOT gate also at port 3	NOT gate not shown
Power Loss	Lower- no shadowed line found at optical field propagation	Higher- dark shadowed line present at optical filed propagation indicated by the white arrowed line
Optical Field Propagation		

B. XOR & XNOR Logic Gates

For constructing XOR & XNOR gates 2 MZIs were needed as depicted in Fig. 6. The optical input signal is provided at 1st input port from MZI1. From Fig. 6, we can see that port 1 & port 2 in the output section of the MZI1 are the 1st and 2nd input for the MZI2. X and Y are the control signal which are provided respectively at the 2nd electrode of MZI1 & MZI2. In this layout shown in Fig. 6, the output port 1 act as XNOR logic and output port 2 as XOR logic. The operation of this structure includes 4 cases as listed in table III (0 V is presented as bit 0 and 4 V is presented as bit 1) e.g. with X zero optical power is switched to input port 2 of MZI2, then with Y zero optical power is at out1 of MZI2. The structure dimensions of the device are : Length $L= 65800 \mu\text{m}$, Width $W= 100 \mu\text{m}$, Center position of second electrodes of both MZI is $5.5 \mu\text{m}$. Optical field propagation corresponding to each of the logics in table III is shown in Fig. 7. If we compare this figure with the previous one implemented using single MZI switch in ref [24], we can see that power at desired port is higher in case of cascaded one. But in terms of size, XOR/XNOR gate using single switch is better as it cost only $32900 \mu\text{m}$ of length, whereas gate using two cascaded switch is $65800 \mu\text{m}$ in length. Output power in every case of high state in table III is 0.992529 mW (insertion loss - 0.033 dB).

TABLE III. TRUTH TABLE FOR XNOR & XOR LOGIC GATE

Control Signal X	Control Signal Y	Port 1 (X XNOR Y)	Port 2 (X XOR Y)
0	0	1	0
0	1	0	1
1	0	0	1
1	1	1	0

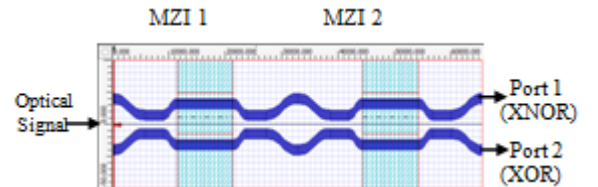


Fig. 6. XOR and XNOR logic layout structure.

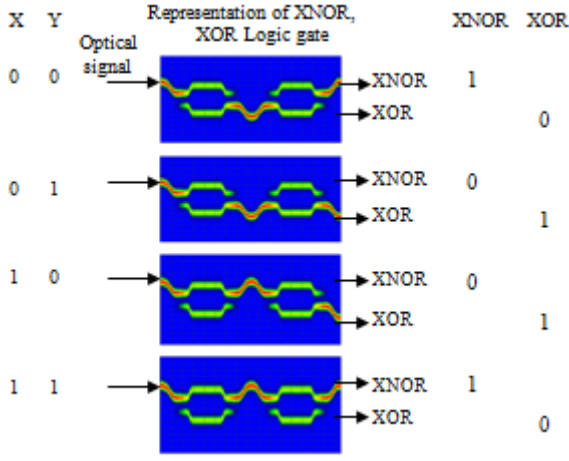


Fig. 7. Representation of XNOR and XOR logic observed from Opti-BPM.

We found the same XNOR/XOR gate using two cascaded MZI switches in ref [20]. But when compared to our designed XNOR/XOR logic gate, ours one is better in the sense of having smaller length, which is 65800 μm , whereas the gate in ref [20] is 76000 μm in length (as outputs of 2 cascaded MZIs of 33000 μm each are connected with 10000 μm waveguides). Again power to the desired port is higher in ours than in ref [20] as we can see that the optical field propagation in figure 8 of ref [20] (shown in table IV) is not uniformly distributed, it seems like there are some leakage of power which can be seen if we look at the disconnected dark shadowed lines indicated by arrow. That means power of some ports are high enough than zero in ref [20]. These are summarized in table IV.

TABLE IV. COMPARATIVE STUDY ON XNOR/XOR GATE

Specification	Proposed XOR/XNOR gate	Existing [20]
Length	65800 μm	76000 μm
Operating Wavelength	1.55 μm	1.3 μm
Switching Voltage	4 V for Ti:KNbO ₃ based switch	6.75 V with Ti:LiNbO ₃ based switch
Power Loss	Lower- no shadowed line found at optical field propagation	Higher- dark shadowed line present at optical filed propagation
Optical Field Propagation		

C. NAND Logic Gate

To implement the NAND gate 3 MZIs are required, as shown in Fig. 8. The first optical signal is applied at the 1st input port of MZI1. Then, we took the first output port of MZI1 as the 1st output & the 2nd output port of MZI1 is connected with the 1st input port of MZI3. Similarly, the 2nd optical signal is applied at 1st input port of MZI2 & the 2nd output port of MZI2 is in a joint with the 2nd input port of

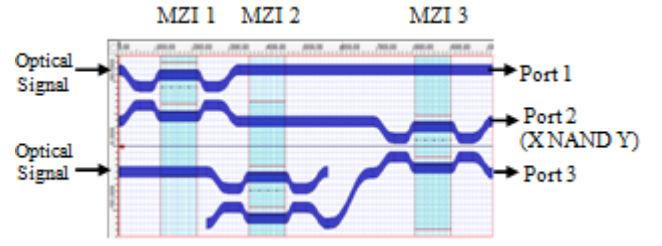


Fig. 8. Layout of proposed NAND gate.

TABLE V. TRUTH TABLE FOR PROPOSED NAND GATE

Control Signal X	Control Signal Y	Port 1	Port 2 (X NAND Y)	Port 3 (X NOR Y)
0	0	0	1	1
0	1	0	1	0
1	0	1	1	0
1	1	1	0	0

MZI3. We found output port 2 as the NAND logic gate. We can find equation 5 for port 2 in Fig. 8 using equation 3 at section II. Where m_0 , m_1 & m_2 are normalized output power at various output ports and found by the similar way as in equation 3 and their corresponding equations are also given in equation 5 [21].

Output port 2 (NAND logic gate) =

$$\begin{aligned}
 m_0 + m_1 + m_2 &= \frac{OUT_{1_{MZI3}}}{E_{in}} \\
 &= \cos^2\left(\frac{\Delta\phi_{MZI1}}{2}\right) \cos^2\left(\frac{\Delta\phi_{MZI2}}{2}\right) \cos^2\left(\frac{\Delta\phi_{MZI3}}{2}\right) \\
 &+ \cos^2\left(\frac{\Delta\phi_{MZI1}}{2}\right) \sin^2\left(\frac{\Delta\phi_{MZI2}}{2}\right) \sin^2\left(\frac{\Delta\phi_{MZI3}}{2}\right) \\
 &+ \sin^2\left(\frac{\Delta\phi_{MZI1}}{2}\right) \cos^2\left(\frac{\Delta\phi_{MZI2}}{2}\right) \cos^2\left(\frac{\Delta\phi_{MZI3}}{2}\right)
 \end{aligned} \quad (5)$$

Fig 9 shows the simulation result of the NAND gate which can be verified by table V. Control signal X is applied at second electrode of MZI1 and control signal Y is at the second

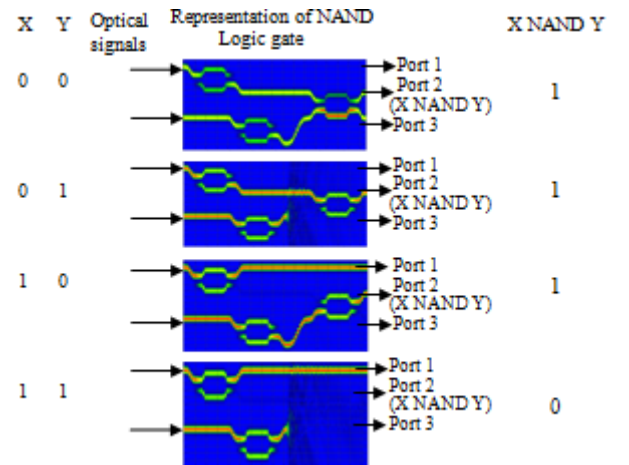
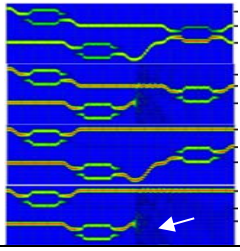
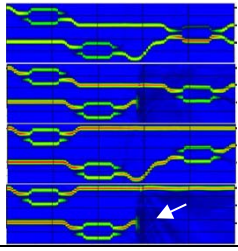


Fig. 9. Representation of NAND logic obtained from Opti-BPM.

electrodes of MZI2 and MZI3. Device dimensions are, Length $L=101800$, Width $W=140$, Center positions of 3 MZI are, MZI 1= $45.5 \mu\text{m}$, MZI 2= -34.5 , MZI 3= $4.5 \mu\text{m}$ above their original axis. Power in high states is 0.992529 mW which gives 0.033 dB of I.L. If we compare our proposed switch to the previous one in ref [21], in case of length our proposed NAND structure is smaller and operating wavelength are also different. By comparing optical field propagation we can see that in figure 9 of ref [21] (shown in table VI) there is some power leakage which can be seen if we look at the disconnected dark shadowed lines indicated by the white arrow in the table, where in our proposed switch is almost absent. Table VI shows a comparative study discussed above. If we take a close look on table V, we can see that port 1 exactly repeats the logic as in control signal X, so along with being NAND gate, this proposed structure can be used as signal repeater too. And again port 3 acts as NOR gate.

TABLE VI. COMPARATIVE STUDY ON NAND GATE

Specification	Proposed NAND gate	Existing [21]
Length	$101800 \mu\text{m}$	$102000 \mu\text{m}$
Operating Wavelength	$1.55 \mu\text{m}$	$1.3 \mu\text{m}$
Switching Voltage	4 V for Ti:KNbO ₃ based switch	6.75 V with Ti:LiNbO ₃ based switch
Applications	Repeater at port 1 and NOR gate at port 3	NOR gate not shown
Power Loss	Lighter shadowed line found at optical field	Darker shadowed line present at optical field
Optical Field Propagation		

D. NOR Logic Gate

The application of universal logic gate is very essential for improving the flexibility in complex digital circuits. A layout diagram of the NOR gate using two MZIs are shown in Fig. 10. Optical signal is given to the 1st input port of MZI1. The 1st output of MZI1 is directly taken as output port 1. The 2nd output port of MZI1 is in a joint with the 2nd input port of MZI2. The 1st output port of MZI2 acts as the NOR gate. The expression for the NOR gate can be obtained by the following equation [21]:

Output port 2 (NOR logic gate) =

$$|OUT_{1_{MZI2}} / E_{in}|^2 = \cos^2\left(\frac{\Delta\phi_{MZI1}}{2}\right) \cos^2\left(\frac{\Delta\phi_{MZI2}}{2}\right) \quad (6)$$

In Opti-BPM layout designer the device was drawn and the device dimensions i.e. length, width of the device and position of center electrodes are, Length $L= 71800 \mu\text{m}$, Width $W= 100 \mu\text{m}$, both the center electrodes are at $5.5 \mu\text{m}$ above the original axis. For different combinations of control signals (X and Y) result of NOR operation is obtained by the beam propagation

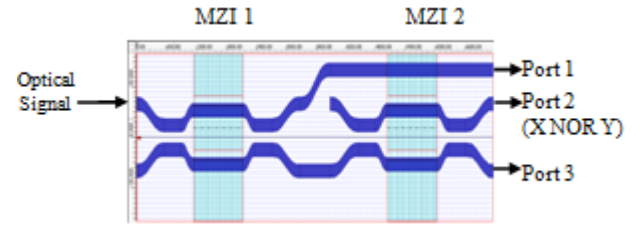
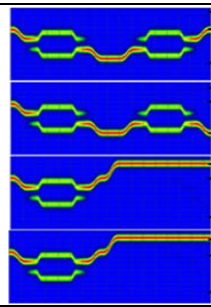
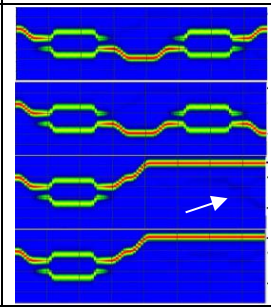


Fig. 10. Layout of NOR gate.

TABLE VII. TRUTH TABLE FOR PROPOSED NOR GATE

Control Signal X	Control Signal Y	Port 1 (Repeater of X)	Port 2 (X NOR Y)	Port 3
0	0	0	1	0
0	1	0	0	1
1	0	1	0	0
1	1	1	0	0

TABLE VIII. COMPARATIVE STUDY ON NOR GATE

Specification	Proposed NAND gate	Existing [21]
Length	$71800 \mu\text{m}$	$72000 \mu\text{m}$
Operating Wavelength	$1.55 \mu\text{m}$	$1.3 \mu\text{m}$
Switching Voltage	4 V for Ti:KNbO ₃ based switch	6.75 V with Ti:LiNbO ₃ based switch
Applications	Repeater at port 1	Not mentioned
Power Loss	Lower-no shadowed line found at optical field propagation	Lighter shadowed line found at optical field propagation indicated by the arrow in the next row.
Optical Field Propagation		

method. Fig. 11 shows the different combinations of the control signals X and Y (applied at second electrodes of MZI1 and MZI2 respectively) & their corresponding outputs at port 2. These results obtained from our proposed structure can be confirmed with the truth table in table VII. Again output power in every case of high state in table VII is 0.992529 mW (insertion loss - 0.033 dB). If we compare our proposed switch to the previous one in ref [21], length of the devices are different. And operating wavelength are also different, when we compare optical field propagation we can see that in figure 7 (shown in table VIII) of ref [21] there is some power leakage, which in our proposed switch is almost absent. Table VIII shows a comparative study discussed above. If we take a close look on table VII, we can see that port 1 exactly repeats the logic as in control signal X, so along with being NOR gate, this proposed structure can be used as signal repeater too. Section C also shown another representation of NOR gate, but if we observe carefully we can see that NOR gate at section D is higher in power transmission, where at port 3 of Fig. 9 power loss is larger.

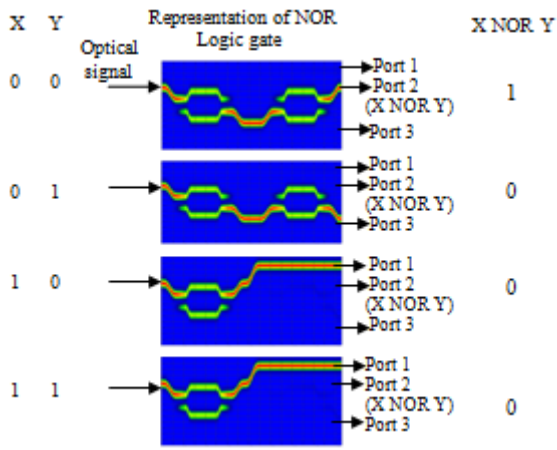


Fig. 11. Representation of NOR logic obtained from Opti-BPM.

IV. CONCLUSION

This paper focuses on design aspects of logic gates such as AND, NOT, XOR, XNOR, NAND and NOR gates using the electro-optic effect in Ti:KNbO₃ based MZI. In our study the proposed optical logic devices are designed by Opti-BPM 12.2 software. By optimizing various device parameters such as length of MZI arm, arm gap, 3 dB coupler region, waveguide width, strip thickness of titanium and center position of electrode; we got better performances with respect to previous researches especially in case of smaller length and lower insertion loss. We found 0.992529 mW power in each high state of the gates which gives very small insertion loss of 0.033 dB. Again all the devices are $\leq 101800 \mu\text{m}$ in length.

ACKNOWLEDGMENT

We want to acknowledge gratefully the support from Information and Communication Technology Division, Government of the People's Republic of Bangladesh provided by the ICT Fellowship 2015-2016 for this work.

REFERENCES

- [1] D.D. Nolte, "Mind at Light Speed: A New Kind of Intelligence," Simon and Schuster, ISBN 978-0-7432-0501-6, pp. 34, 2001.
- [2] Feitelson, Dror G., "Optical Image and Signal Processing" Optical Computing: A Survey for Computer Scientists. Cambridge, MA: MIT Press. ISBN 0-262-06112-0, 1988.
- [3] S. K. Kim, K. Goda, A. M. Fard, B. Jalali, "Optical time-domain analog pattern correlator for high-speed real-time image recognition". Optics Letters, vol. 36, no. 2, pp. 220, 2011.
- [4] Y. Ishizaka, Y. Kawaguchi, K. Saitoh, M. Koshiba, "Design of ultra compact all-optical XOR and AND logic gates with low power consumption," Opt. Commun., vol. 284, 3528–3533, 2011.
- [5] H. Sun, Q. Wang, H. Dong, N. K. Dutta, "XOR performance of a quantum-dot semiconductor optical amplifier-based Mach-Zehnder interferometer," Opt. Express, vol. 13, pp. 1892–1899, 2005.
- [6] H. Han, M. Zhang, P. Ye, F. Zhang, "Parameter design and performance analysis of an ultrafast all-optical XOR gate based on quantum-dot semiconductor optical amplifiers in nonlinear Mach-Zehnder interferometer," Opt. Commun., vol. 281, pp. 5140–5145, 2008.
- [7] E. Dimitriadou, K. E. Zoiros, "On the design of ultrafast all-optical NOT gate using quantum-dot semiconductor optical amplifier-based Mach-Zehnder interferometer," Opt. Laser Technol., vol. 44, pp. 600–607, 2012.
- [8] A. Kotb, K. E. Zoiros, "Simulation of all-optical logic XNOR gate based on quantum-dot semiconductor optical amplifiers with amplified spontaneous emission," Opt. Quantum Electron., vol. 45, pp. 1213–1221, 2013.
- [9] E. Dimitriadou, K. E. Zoiros, "Proposal for ultrafast all-optical XNOR gate using single quantum-dot semiconductor optical amplifier-based Mach-Zehnder interferometer," Opt. Laser Technol., vol. 45, pp. 79–88, 2013.
- [10] J. Y. Kim, J. M. Kang, T. Y. Kim, S. K. Han, "All-optical multiple logic gates with XOR, NOR, OR, and NAND functions using parallel SOA-MZI structures: theory and experiment," J. Lightwave Technol., vol. 24, issue 9, pp. 3392–3399, 2006.
- [11] D. K. Gayen, A. Bhattacharyya, T. Chattopadhyay, et al., "Ultrafast All-Optical Half Adder Using Quantum-Dot Semiconductor Optical Amplifier Assisted Mach-Zehnder Interferometer," J. Lightwave Technol., vol. 30, pp. 3387–3393, 2012.
- [12] D. K. Gayen, T. Chattopadhyay, "Designing of Optimized All-Optical Half Adder Circuit Using Single Quantum-Dot Semiconductor Optical Amplifier Assisted Mach-Zehnder Interferometer," J. Lightwave Technol., vol. 31, pp. 2029–2035, 2013.
- [13] Y. D. Wu, M. L. Huang, M. H. Chen, et al., "All-Optical Switch Based on the Local Nonlinear Mach-Zehnder Interferometer," Opt. Exp., vol. 15, pp. 9883–9892, 2007.
- [14] C. Taraphdar, T. Chattopadhyay, J. N. Roy, "Mach-Zehnder Interferometer-Based All-Optical Reversible Logic Gate," Opt. Laser Technol., vol. 42, pp. 249–259, 2010.
- [15] T. Houbavlis, K.E. Zoiros, G. Kanellos, C. Tsekrekos, "Performance analysis of ultrafast all-optical Boolean XOR gate using semiconductor optical amplifierbased Mach-Zehnder Interferometer," Opt. Commun., vol. 232 (1–6), pp. 179–199, March 2004.
- [16] Q. Wang, G. Zhu, H. Chen, J. Jaques, J. Leuthold, A.B. Piccirilli, N.K. Dutta, "Study of all-optical XOR using Mach-Zehnder interferometer and differential scheme," IEEE J. Quantum Electron., vol. 40, no. 6, pp. 703–710, June 2004.
- [17] Min Zhang, Yongpeng Zhao, Ling Wang, Jian Wang, Peida Ye, "Design and analysis of all-optical XOR gate using SOA-based Mach-Zehnder interferometer," Opt. Commun., vol. 223, pp. 301–308, 2003.
- [18] T. Houbavlis, K.E. Zoiros, G. Kanellos, C. Tsekrekos, "Performance analysis of ultrafast all-optical Boolean XOR gate using semiconductor optical amplifier based Mach-Zehnder Interferometer," Opt. Commun., vol. 232, pp. 179–199, 2004.
- [19] Huining Han, Min Zhang, Peida Ye, Fangdi Zhang, "Parameter design and performance analysis of a ultrafast all-optical XOR gate based on quantum dot semiconductor optical amplifiers in nonlinear Mach-Zehnder interferometer," Opt. Commun., vol. 281, pp. 5140–5145, 2008.
- [20] A. Kumar, S. Kumar, S. K. Raghuvanshi, "Implementation of XOR/XNOR and AND logic gates using Mach-Zehnder interferometers," Optik 125, pp. 5764–5767, 2014.
- [21] Santosh Kumar, Gurdeep Singh, Ashish Bisht, Sandeep Sharma, Angela Amphawan, "Proposed new approach to the design of universal logic gates using the electro-optic effect in Mach-Zehnder interferometers," Applied Optics, vol. 54, no. 28, pp. 8479–8484, 2015.
- [22] Nazmi A. Mohammed, Huda S. Abo Elnasr, Moustafa H. Aly, "Analysis and design of an electro-optic 2×2 switch using Ti: KNbO₃ as a waveguide based on MZI at $1.3 \mu\text{m}$," Opt Quant Electron (springer journal), vol. 46, pp. 295–304, 2014.
- [23] S. K. Raghuvanshi, A. Kumar, S. Kumar, "1×4 signal router using three Mach-Zehnder interferometers," Opt. Eng., vol. 52, pp. 035002, 2013.
- [24] Rabeya Khatun, Kazi Tanvir Ahmmed, A. Zahed Chowdhury, Raqibul Hossen, "Optimization of 2x2 MZI Electro-Optic Switch and It's Application as Logic Gate," 18th International Conference on Computer and Information Technology, pp. 294–299, 2015.
- [25] Rabeya Khatun, A. Zahed Chowdhury, "Performance Evolution of a 2x2 MZI Electro-Optic Switch Using Ti:KNbO₃ as Waveguide at $1.55 \mu\text{m}$," 3rd ICEEICT, 2016, in press.

Design, Analysis and Optimization of Grid-Connected Solar Photovoltaic System Using RETScreen in Eight Divisions of Bangladesh

Shagorika Mukherjee* and M. Abdur Razzak**

Department of Electrical and Electronic Engineering

Independent University, Bangladesh

Plot-16, Block-B, Bashundhara, Dhaka-1229, Bangladesh

E-mail: shagorikam@gmail.com*, razzak@iub.edu.bd**

Abstract—Design, technical and financial analysis, and optimization of 100 kW grid connected solar photovoltaic system at each division of Bangladesh have been carried out using RETScreen along with NASA's data of location and solar radiation. Slope and azimuth of placing solar panels affect absorption of solar irradiation by the panels and this has considerable impact on the amount of electricity being generated by a solar grid, the technical and financial aspects of the solar grid. Bangladesh is an Asian country having eight divisions. Due to varying geographical locations, the amount of solar radiation being incident on the divisions vary significantly. Thus, placing solar panels at the same slope and azimuth in all the divisions does not allow the panels to absorb maximum amount of solar radiation. This paper discusses about the design of 100kW grid connected solar photovoltaic system, performs technical analysis and demonstrates the optimum slope and azimuth of placing solar panels at different divisions of Bangladesh. This paper also represents financial analysis of the solar grid for all the divisions. The optimized solar grid is capable of reducing approximately 166 tons of carbon dioxide emission annually which is equivalent to eliminating the use of 30.4 cars and light trucks annually.

Keywords—photovoltaic system; RETScreen; technical analysis; financial analysis; solar grid.

I. INTRODUCTION

Based on the latest United Nations Population Division Estimates, Bangladesh has a population of 162,910,864 among which 76% has access to electricity [1-2]. Each year, the population of Bangladesh is increasing significantly. The current estimated population growth rate of Bangladesh is 1.19% [1] which will keep on increasing the population with time and hence increasing the demand of electricity. According to the installed capacity as on April, 2016, natural gas is the main fuel for generating electricity in Bangladesh and it generates 61.82% of electricity while 2.03% and 29.43% of electricity are generated by coal and oil, respectively. 1.86% of the electricity is obtained from hydro and 4.86% is obtained by importing it [3]. Thus, Bangladesh is highly dependent on fossil fuel for generating electricity while fossil fuel has tremendous impact on global warming, it is non-renewable and almost on the verge of extinction, it is not clean and its price is increasing day by day which in turn is increasing the tariff rate of electricity.

Therefore, the use of fossil fuels in generating electricity is neither a solution to the power crisis of Bangladesh nor a solution to meet the future demand of electricity, while the government aims to make the entire country electrified with demand and supply of electricity being equal. From the point of worldwide view, the use of solar energy is one of the finest options since it is clean, renewable, and sustainable as well as environment friendly [4]. Furthermore, the geographic position of Bangladesh is such that it gets a large quantity of solar radiation and thus implementing cost effective and effectual solar photovoltaic systems can lessen power crisis by generating power at very low price [4-5]. Therefore, solar energy is considered to be one of the top growing energy sources in Bangladesh amongst all the renewable energy sources [6].

Bangladesh Power Development Board installed a micro photovoltaic grid system of 10kW in Barkal, Rangamati Hill District in March 2005 [7]. The design of the grid has been verified using RETScreen and its analysis has been performed too [7]. The feasibility analysis of technical as well as financial aspects of a 500kW grid connected solar photovoltaic system at Rajshahi, Bangladesh has been carried out [8].

On the other hand, this paper considers all the eight divisions of Bangladesh separately since the amount of solar radiation being incident on each division vary significantly due to their varying geographical locations, which in turn has considerable impact on the amount of electricity being generated by a solar grid, the technical aspects as well as the financial aspects of the solar grid. The eight divisions of Bangladesh are - Dhaka, Chittagong, Rangpur, Khulna, Sylhet, Barisal, Maimansingh and Rajshahi, and in this paper the divisions are represented by the short names DHK, CTG, RNG, KHL, SYL, BAR, MMS and RAJ respectively.

The objectives of this study are as the following: (1) to design a 100kW grid connected solar photovoltaic system, (2) to perform technical analysis to figure out various technical parameters and to optimize the solar grid by finding out the optimum angles of slope and azimuth for placing the solar panels in each division of Bangladesh in order to allow the solar panels to absorb the maximum amount of solar radiation, (3) to perform financial analysis of the solar grid based on each division.

II. LOCATIONS AND ANNUAL SOLAR RADIATION

This paper uses the RETScreen software, which is a Clean Energy Project Analysis Software developed by the Ministry of Natural Resources, Canada. As a part of its analysis tools, RETScreen uses NASA's global climate data [9]. In this paper, NASA Surface meteorology as well as Solar Energy (SSE) data sets from RETScreen have been used to obtain the latitude, longitude and annual solar radiation-horizontal of each division of Bangladesh as listed in Table I.

TABLE I. LOCATION AND ANNUAL SOLAR RADIATION OF EACH DIVISION

Division	Latitude (North)	Longitude (East)	Annual Solar Radiation-Horizontal (kWh/m ² /d)
DHK	23.7°	90.4°	4.65
CTG	22.3°	91.8°	4.55
RNG	25.7°	89.3°	4.86
KHL	22.8°	89.6°	4.55
SYL	24.9°	91.9°	4.57
BAR	22.7°	90.4°	4.51
MMS	24.8°	90.4°	4.64
RAJ	24.4°	88.6°	4.87

III. SYSTEM DESIGN, TECHNICAL ANALYSIS AND OPTIMIZATION

This study is based on a 100kW solar grid. Therefore, the target is to design a grid which will be capable of generating a total daily power of 100kW. The unit of electricity required to be produced annually to get a daily power of 100kW is calculated using the following equation.

$$\text{Daily Power} = \frac{\text{Unit of Electricity}}{\text{Peak Sunshine Hours} \times 365} \quad (1)$$

The peak sunshine hours are considered to be from 9am to 4pm every day for entire year. According to this calculation, it has been found that 255500kWh of electricity is required annually which is approximately equal to 266MWh. Table II represents the summary of PV module and inverter being considered in this study.

TABLE II. SUMMARY OF PV MODULE AND INVERTER

Component	PV Module	Inverter
Type	Mono-Crystalline	Pure Sine Wave
Manufacturer	SolarWorld	Shenzhen JingFuYuan Tech Co., Ltd (JFY)
Model	SW275 mono black	According to required power capacity
Efficiency (%)	16.4	98.0
Capacity(kW)	0.275	As per division

For the design, additional technical parameters are listed in Table III which provides the information of the required units of solar panel to generate 100kW electricity in each division

along with the output of the panels, area required by the panels and the minimum required power capacity of the inverters.

TABLE III. FURTHER TECHNICAL ANALYSIS

Division	Required Units of Solar Panel	Output of Solar Panels (kW)	Solar Collector Area (m ²)	Required Power Capacity of Inverter (kW)
DHK	564	155.10	831.50	156
CTG	579	159.23	853.60	160
RNG	527	144.93	777.00	146
KHL	580	159.50	855.00	160
SYL	562	154.55	829.00	155
BAR	583	160.33	859.50	161
MMS	557	153.18	821.50	154
RAJ	537	147.68	792.00	148

In this study, a 100kW grid is considered for technical analysis in order to find out the optimum slope and azimuth angles for each division of Bangladesh to place the solar panels, so that maximum amount of solar radiation can be absorbed by the panels for generating electricity. The data obtained for eight divisions are listed in Table IV. From the table, it can be seen that in Dhaka division, placing the PV panels at an azimuth of 0° with a slope of 25°/26° or an azimuth of 1° with a slope of 26° generates maximum power annually which is 255.90MWh. In case of Chittagong division, the azimuth has to be 0° while the slope has to be 24° to get the maximum power annually. For Rangpur division, the best azimuth is 0° while the best slope is 29°. An azimuth of 0° with a slope of 24°/25° or an azimuth of 1° with a slope of 25° is the best selection for Khulna division. A slope of 28° with an azimuth of either 0° or 1° is the best match for Sylhet while for Barisal the optimum slope is 25° for the same azimuths as Sylhet. An azimuth of 0° along with a slope of 27°/28° is optimum for Maimansingh whereas for Rajshahi the optimum slope is 26° for the same azimuth as Maimansingh.

In this study, the inverters being considered are wall mounted or can be mounted below the panels in the structure itself while the entire system is considered as roof top, and thus the inverters do not require any additional area on the roof.

IV. FINANCIAL ANALYSIS

Financial analysis is divided into three parts and they are the initial project cost at each division, annual cost for operation and maintenance and further financial parameters as shown in tables V, VI and VII respectively. The installation is considered in such a way that four pieces of solar panels are mounted on six pieces of solar mounting system. The mounting system considered in this study is fixed. The transportation cost is not calculated separately but it is considered within the total initial cost of the system in each division.

TABLE IV. TECHNICAL ANALYSIS TO OPTIMIZE THE GRID

Azimuth in degrees	Slope in degrees	Annual Electricity Exported to Grid (MWh)							
		Division							
		DHK	CTG	RNG	KHL	SYL	BAR	MMS	RAJ
0	22	255.56	255.61	254.70	255.67	254.80	255.37	255.13	255.40
	23	255.72	255.70	255.04	255.78	255.08	255.49	255.39	255.59
	24	255.84	255.75	255.33	255.84	255.31	255.56	255.60	255.73
	25	255.90	255.74	255.66	255.84	255.49	255.58	255.76	255.82
	26	255.90	255.68	255.74	255.79	255.61	255.55	255.87	255.86
	27	255.86	255.57	255.86	255.69	255.68	255.47	255.92	255.84
	28	255.76	255.41	255.93	255.55	255.70	255.33	255.92	255.77
	29	255.61	255.20	255.95	255.34	255.67	255.15	255.87	255.65
	30	255.41	254.94	255.91	255.09	255.58	254.91	255.76	255.47
1	22	255.56	255.60	254.69	255.67	254.79	255.37	255.13	255.39
	23	255.72	255.70	255.03	255.77	255.08	255.49	255.39	255.59
	24	255.83	255.74	255.32	255.83	255.30	255.56	255.60	255.73
	25	255.89	255.74	255.55	255.84	255.48	255.58	255.76	255.82
	26	255.90	255.68	255.73	255.79	255.60	255.54	255.86	255.85
	27	255.85	255.57	255.85	255.69	255.68	255.46	255.91	255.83
	28	255.76	255.41	255.93	255.54	255.70	255.33	255.91	255.76
	29	255.61	255.20	255.94	255.34	255.66	255.14	255.86	255.64
	30	255.41	254.93	255.91	255.09	255.58	254.90	255.75	255.46
2	22	255.54	255.59	254.68	255.65	254.78	255.35	255.11	255.38
	23	255.70	255.69	255.02	255.76	255.06	255.47	255.37	255.57
	24	255.82	255.73	255.30	255.82	255.29	255.54	255.58	255.71
	25	255.88	255.72	255.53	255.82	255.46	255.56	255.74	255.80
	26	255.88	255.66	255.71	255.77	255.59	255.53	255.84	255.84
	27	255.84	255.55	255.84	255.67	255.66	255.44	255.89	255.82
	28	255.74	255.39	255.91	255.52	255.68	255.31	255.89	255.75
	29	255.59	255.18	255.92	255.32	255.64	255.12	255.84	255.62
	30	255.39	254.91	255.89	255.07	255.56	254.88	255.73	255.45

TABLE V. INITIAL COST

Component	Division							
	DHK	CTG	RNG	KHL	SYL	BAR	MMS	RAJ
Solar Modules 68,000 BDT/kW	10,546,800	10,827,640	9,855,240	10,846,000	10,509,400	10,902,440	10,416,240	10,042,240
Inverter 8,500/kW	1,326,000	1,360,000	1,241,000	1,360,000	1,317,500	1,368,500	1,309,000	1,258,000
Mounting System 200BDT/piece	169,200	173,800	158,200	174,000	168,600	175,000	167,200	161,200
Cables, wires, nuts& bolts	200,000	200,000	200,000	200,000	200,000	200,000	200,000	200,000
Development Cost	120,000	120,000	120,000	120,000	120,000	120,000	120,000	120,000
Subtotal	123,62000	126,81500	115,74500	127,00000	123,15500	127,66000	122,12500	117,81500
Contingency (1%)	123,620	126,815	115,745	127,000	123,155	127,660	122,125	117,815
Total Initial Cost	12,485,620	12,808,255	11,690,185	12,827,000	12,438,655	12,893,600	12,334,565	11,899,255

The Annual Cost of System Operation and Maintenance and the Contingences for each division are considered to be the same since these parameters do not vary much from division to division. Table VI shows the annual cost while the data of Table VII represents additional financial parameters which are also considered to be the same for all divisions. An Inflation Rate of 5.6% is considered as on April, 2016 and it is assumed that 70% of project expense is supported by a loan.

TABLE VI. ANNUAL COST

Parameters	Unit	Price
Operation & Maintenance	BDT	125,000
Contingencies (%)	20%	25,000
Total Annual Cost	BDT	150,000

TABLE VII. OTHER FINANCIAL PARAMETERS

Project Life	25 years
Debt Ratio	70%
Debt Interest Rate	2%
Debt Term	12 years
Fuel Cost Escalation Rate	3%
Inflation Rate	5.6%
Discount Rate	5%
Electricity Export Escalation Rate	3%

TABLE VIII. EFFECT OF ELECTRICITY SELLING RATE ON VARIOUS FINANCIAL PARAMETERS

Electricity Selling Rate BDT/kWh	Division	Annual Savings/Income (BDT)	IRR(%) Equity	IRR(%) Assets	Simple Payback (Years)	Equity Payback (Years)	Benefit-Cost Ratio	NPV (BDT)
3	DHK	767,710	negative	negative	20.2	17.8	-1.31	-8,659,321
	CTG	767,244	negative	negative	20.7	18.2	-1.38	-9,115,680
	RNG	767,851	negative	negative	19.1	17.0	-1.14	-7,553,449
	KHL	767,520	negative	negative	20.7	18.2	-1.38	-9,140,071
	SYL	767,106	negative	negative	20.2	17.8	-1.31	-8,611,558
	BAR	766,744	negative	negative	20.8	18.3	-1.40	-9,244,763
	MMS	767,756	negative	negative	20.0	17.7	-1.28	-8,449,715
	RAJ	767,573	negative	negative	19.4	17.2	-1.19	-7,857,035
4	DHK	1,023,613	negative	negative	14.3	13.1	0.03	-3,628,868
	CTG	1,022,992	negative	negative	14.6	13.4	-0.07	-4,088,278
	RNG	1,023,801	negative	negative	13.5	12.5	0.29	-2,522,075
	KHL	1,023,360	negative	negative	14.6	13.4	-0.07	-4,110,861
	SYL	1,022,808	negative	negative	14.3	13.1	0.04	-3,585,061
	BAR	1,022,326	negative	negative	14.7	13.4	-0.10	-4,220,638
	MMS	1,023,675	negative	negative	14.1	13.0	0.08	-3,418,959
	RAJ	1,023,431	negative	negative	13.7	12.6	0.21	-2,827,478
5	DHK	1,279,517	2.0	positive	11.1	8.3	1.37	1,401,585
	CTG	1,278,741	2.8	positive	11.3	8.7	1.25	939,124
	RNG	1,279,751	0.4	positive	10.4	7.3	1.71	2,509,299
	KHL	1,279,200	2.8	positive	11.3	8.8	1.24	918,348
	SYL	1,278,511	1.9	positive	11.0	8.3	1.39	1,441,437
	BAR	1,277,907	3.0	positive	11.4	8.9	1.21	803,488
	MMS	1,279,594	1.7	positive	10.9	8.1	1.43	1,611,797
	RAJ	1,279,289	0.8	positive	10.6	7.6	1.61	2,202,080
6	DHK	1,535,420	-2.9	positive	9.0	5.5	2.72	6,432,038
	CTG	1,534,489	-2.5	positive	9.2	5.7	2.56	5,966,526
	RNG	1,535,701	-4.0	-1.0	8.5	4.9	3.13	7,540,673
	KHU	1,535,040	-2.5	positive	9.2	5.7	2.55	5,947,558
	SYL	1,534,213	-3.0	positive	9.0	5.4	2.73	6,467,935
	BAR	1,533,489	-2.4	positive	9.3	5.8	2.51	5,827,614
	MMS	1,535,512	-3.1	positive	8.9	5.4	2.79	6,642,553
	RAJ	1,535,147	-3.7	0.2	8.6	5.1	3.01	7,231,638

Electricity Selling Rate BDT/kWh	Division	Annual Savings/Income (BDT)	IRR(%) Equity	IRR(%) Assets	Simple Payback (Years)	Equity Payback (Years)	Benefit- Cost Ratio	NPV (BDT)
7	DHK	1,791,323	-6.1	-4.9	7.6	4.0	4.06	11,462,491
	CTG	1,790,237	-5.7	-4.3	7.8	4.2	3.87	10,993,928
	RNG	1,791,651	-7.1	-6.2	7.2	3.6	4.56	12,572,047
	KHU	1,790,880	-5.7	-4.3	7.8	4.2	3.86	10,976,767
	SYL	1,789,915	-6.2	-5.0	7.6	4.0	4.08	11,494,433
	BAR	1,789,070	-5.6	-4.1	7.8	4.2	3.82	10,851,739
	MMS	1,791,431	-6.3	-5.2	7.5	3.9	4.15	11,673,309
	RAJ	1,791,004	-6.8	-5.9	7.3	3.7	4.41	12,261,196
8	DHK	2,047,226	-8.8	-8.2	6.6	3.2	5.40	16,492,944
	CTG	2,045,985	-8.4	-7.7	6.7	3.3	5.18	16,021,329
	RNG	2,047,601	-9.8	-9.4	6.2	2.9	5.98	17,603,421
	KHU	2,046,720	-8.3	-7.7	6.7	3.3	5.17	16,005,976
	SYL	2,045,617	-8.8	-8.2	6.6	3.2	5.42	16,520,931
	BAR	2,044,652	-8.2	-7.6	6.8	3.3	5.12	15,875,865
	MMS	2,047,350	-8.9	-8.4	6.5	3.1	5.51	16,704,065
	RAJ	2,046,862	-9.5	-9.0	6.3	3.3	5.81	17,290,754
9	DHK	2,303,130	-11.1	-10.8	5.8	2.6	6.75	21,523,397
	CTG	2,301,733	-10.7	-10.4	5.9	2.7	6.49	21,048,731
	RNG	2,303,552	-12.2	-12.0	5.5	2.4	7.40	22,634,795
	KHU	2,302,560	-10.7	-10.3	5.9	2.7	6.48	21,035,186
	SYL	2,301,319	-11.1	-10.9	5.8	2.6	6.77	21,547,429
	BAR	2,300,233	-10.6	-10.2	6.0	2.7	6.42	20,899,990
	MMS	2,303,269	-11.3	-11.0	5.7	2.6	6.86	21,734,821
	RAJ	2,302,720	-11.9	-11.6	5.6	2.4	7.21	22,320,311
10	DHK	2,559,033	-13.3	-13.1	5.2	2.2	8.09	26,553,849
	CTG	2,557,481	-12.9	-12.7	5.3	2.3	7.81	26,076,133
	RNG	2,559,502	52.2	17.7	4.9	2.0	8.82	27,666,169
	KHU	2,558,401	-12.8	-12.7	5.3	2.3	7.79	26,064,395
	SYL	2,557,021	-13.3	-13.2	5.2	2.2	8.12	26,573,926
	BAR	2,555,814	-12.7	-12.5	5.3	2.3	7.73	25,924,116
	MMS	2,559,187	-13.5	-13.4	5.1	2.2	8.22	26,765,576
	RAJ	2,558,578	-14.1	-14.0	5.0	2.1	8.61	27,349,869
11	DHK	2,814,936	55.2	18.8	4.7	1.9	9.43	31,584,302
	CTG	2,813,229	53.6	18.2	4.8	2.0	9.12	31,103,535
	RNG	2,815,452	59.4	20.1	4.4	1.8	10.25	32,697,543
	KHU	2,814,241	53.5	18.2	4.8	2.0	9.11	31,093,605
	SYL	2,812,723	55.3	18.8	4.7	1.9	9.46	31,600,424
	BAR	2,811,396	53.1	18.1	4.8	2.0	9.03	30,948,242
	MMS	2,815,106	56.0	19.0	4.6	1.9	9.58	31,796,332
	RAJ	2,814,436	58.2	19.7	4.5	1.8	10.01	32,379,427
12	DHK	3,070,840	62.1	21.0	4.3	1.7	10.78	36,614,775
	CTG	3,068,977	60.3	20.4	4.4	1.7	10.43	36,130,937
	RNG	3,071,402	66.8	22.5	4.0	1.6	11.67	37,728,918
	KHU	3,070,081	60.2	20.4	4.4	1.7	10.42	36,122,814
	SYL	3,068,425	62.3	21.0	4.3	1.7	10.81	36,626,922
	BAR	3,066,977	59.8	20.3	4.4	1.8	10.34	35,972,367
	MMS	3,071,025	62.9	21.3	4.2	1.7	10.94	36,827,088
	RAJ	3,070,293	65.4	22.0	4.1	1.6	11.41	37,408,985

V. EFFECT OF ELECTRICITY SELLING RATE

Table VIII signifies the effect of electricity selling rate on various financial parameters based on each division. It can be seen that the project is not feasible in any of the divisions when the selling rate of electricity is BDT 3 to 4 since the Net Present Values (NPV), Equity and Asset IRRs are all negative with some of the Benefit-Cost Ratios being positive while the Simple and Equity Paybacks are high. For the selling rate of electricity being in the range of BDT 5 to 12, the projects are feasible in all the divisions with the NPV and B-C ratio being positive while IRRs are either positive or negative. Whereas, when the electricity selling rate is BDT 11 or higher, the project is highly feasible in all the divisions with all the financial parameters being positive. Annual savings/income increases gradually with the increase in selling rate of electricity, the B-C ratio increases and the paybacks decrease.

VI. CONCLUSION

The design, technical & financial analysis and optimization of 100kW grid connected solar photovoltaic system at each division of Bangladesh are presented using RETScreen software along with NASA's data of location and solar radiation. From the analysis it is revealed that an azimuth of 0° is optimum for all of the divisions with varying angles of slopes at each division in case of placing the solar panels to let the panels to absorb maximum amount of solar radiation. On the other hand, in terms of financial aspect, when the electricity selling rate is BDT 11 or higher, the project is highly feasible with all the financial parameters being positive in all the divisions, whereas the project is not feasible in any of the divisions when the selling rate of electricity is less than BDT 5. It is also observed that the annual savings/income increases gradually with the increase in selling rate of electricity, the B-C ratio increases and the paybacks decrease.

The project also has huge environmental impact since it is capable of reducing approximately 166 tons of carbon dioxide emission annually.

ACKNOWLEDGMENT

The authors acknowledge the information of components of the solar grid provided by the Green Energy Research Center of Independent University, Bangladesh and RK Tech (India) Pvt. Ltd.

REFERENCES

- [1] www.worldometers.info(Accessed on May, 2016)
- [2] Power Cell. Power Division, Ministry of Power, Energy & Mineral Resources, and Government of the People's Republic of Bangladesh, March, 2016
- [3] Bangladesh Power Development Board, April, 2016
- [4] Statistical pocketbook of Bangladesh, 2010.
- [5] S. Nahar, A. T. M. G. Sarwar and A.S. Chowdhury, "A theoretical Analysis of optimizing Solar Irradiance: Bangladesh Perspective," 1st International Conference on the Developments in Renewable Energy Technology, 17-19 December, 2009, pp.161-164.
- [6] Sajib Chakraborty, M. A. Razzak, "Design of a Transformer-less Grid-Tie Inverter using Dual-Stage Buck and Boost Converters". International Journal of Renewable Energy Research, Vol.4, No.1, 2014
- [7] Shahidul I.Khan, Md. Abul Kashem, Md. Ariful Hoque, "Design and Analysis of a mini Solar Grid in Remote Area of Bangladesh," IEEE Transl. North American Power Symposium, October 2009.
- [8] Md. Alam Hossain Mondal, A.K.M. Sadrul Islam, "Techno-economic Feasibility of Grid connected Solar PV System in Bangladesh," IEEE Transl. Developments in Renewable Energy Technology (ICDRET), Decemnberr 2009.
- [9] The Researcher News, NASA, http://www.nasa.gov/centers/langley/news/researchernews/rn_RETScreen.html (Accessed on May, 2016)
- [10] RETScreen International, Renewable Energy Project Analysis Software, Natural Resources Canada, <http://www.nrcan.gc.ca/>(Accessed on February,2016)

Developing Load Profile for Domestic Customers of Dhaka City through Statistical Prediction

Abidur Rahman

Lecturer, EEE Department,
Northern University Bangladesh,
Dhaka, Bangladesh.
abidr92@gmail.com

Syed Ishraq Chowdhury Arnob

Jr. Electrical Engineer, EME Division,
WASO Engineers & Consultants (BD) Ltd.
Dhaka, Bangladesh.
c_ishraq@yahoo.com

Abstract—Due to ever growing population, the electricity demand is continuously increasing in Dhaka city. Electricity demand is influenced by a number of factors such as number of customers, customer behavior, working hours, weather, weekday-weekend, holidays etc. In order to reduce future energy consumption by city-dwellers, understanding the patterns of present electricity consumption is necessary. By taking the reading of individual load consumption in real time for 24 hours, the load curve can be generated. However, due to wide diversity of customers, a statistical methodology needs to be followed for the preparation of aggregated load curves. By doing a survey on number of domestic users filling out a questionnaire, a prediction of domestic load curve has been obtained. Analysis of these load curves provides an understanding about the behavior of residential users in different time and day. Results have been verified with feeder data of corresponding substation.

Keywords: load curve; residential customer; survey; feeder data.

I. INTRODUCTION

The concept of developing a load profile has been used as an efficient, organized and systematic tool for energy tariff scheme design, planning and load management in many countries [1]. Proper analysis and subsequent implications of accurate load profiles can be used in balancing the load demand, which eventually helps to enhance the stability and reliability of the system.

Bottom-up & top-down approach, time resolution and spatial resolution are some of the most common methods used worldwide along with other accepted techniques [2]. The choice of the most appropriate method to implement before analysis is usually determined by the available data sources and the forecasting intentions. Most of these methods do not study human behavior. Since different consumers have different behavior, there is a need for a tool that can link the behavior of consumers to generate load profiles in local communities. Domestic energy consumption has always been a large share in total electricity consumption in any network. Therefore it has to be considered as a vital part in energy solutions [3]. The pie chart in Figure 1 illustrates the total electricity consumption of different types of users in Dhaka city [4].

The objective of this paper is to develop a methodology that enables the prediction of realistic electricity profiles for local

resident communities of Dhaka city. The idea is to generate load profiles for the domestic customers of Dhaka city through statistical prediction and to increase knowledge and understanding of load usage of urban communities. Furthermore, this paper will try to compare and correlate the data and analyze the results in order to determine different factors that influence the load profiles.

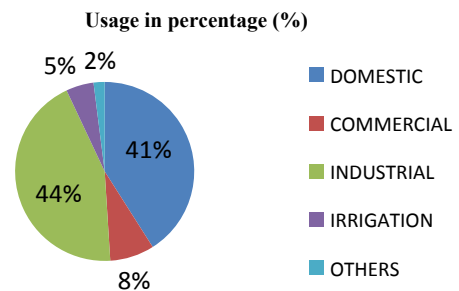


Figure 1. Consumption categories (as of July 2006-June 2007) in Dhaka [4]

This paper is comprised of five sections. Section I introduces the importance of generating load profiles and identifies the objective the work. Section II consists of brief descriptions of factors influencing load profile generation and the methodology of the project is in section III. The results and analysis are explained in Section IV. The findings are summarized and the future scopes identified in Section V.

II. INFLUENCING FACTORS OF LOAD CURVES

Daily load profile is a tool through which the information on the consumers' consumption patterns can be gathered. To analyze consumers' information, the demand characteristic has to be considered as a very important factor.

Nowadays, developing a load profile has become one of the most appropriate methods to approach and cope effectively with the consumers' load shape problems. Choosing the most appropriate load profiling method for any situation depends on factors including data availability, equipment availability, cost, regulatory requirements, accuracy requirements, and the needs of the utility distribution company. The target is to distribute consumers with similar load profiles into consistent groups.

Load profiling is commonly used for billing, demand predicting, and tariff designs.

Load profiles can significantly vary due to some important factors. Customer behavior, working hours, weekday-weekend, change in weather, national holiday are some of the major factors. These factors influence the load curves which should be taken into consideration while generating load profiles for a certain community. For example, in residential community the load curves of some of the families under observation will change over time because of these factors, even if major calamity change does not occur. Also there is a noticeable difference in the load profiles of urban and rural residential communities.

Understanding the main determinants affecting a households' energy use is important for the planning and implementation of organized strategies to reduce energy use in the residential sector. Traditional methods do not study the behavior of people. The issue is that different households have different behavior. Therefore, traditional forecasting methods are not suitable for an intervention approach at local community level because there is no clarity on how the intervention could affect behavior. So, there is a need for a tool that can link the behavior of households for local planners.

III. METHODOLOGY

Figure 2 illustrates the general algorithm for developing load profiles from survey conducted on residential consumers. Modeling of residential energy consumption depends on input data. Comprehensive and detailed electricity load profiles for residential buildings are an important requirement and an essential element for the accurate analysis of demand side management. The first step for building a load profile model is to have sufficient amount of data available. Demographic information, daily occupancy information and annual electricity consumption patterns is included in the input data which is required to develop the residential energy model. Normally the number of households in the community is considered as the initial input. There is a difference in lifestyles amongst the households, which mean different types of consumers with diverse characteristics have differing daily load profiles. A survey questionnaire is prepared with the support of [5]. Screenshot of the questionnaire is given in Appendix.

The questionnaire is divided into three main sections. The first section requests to give general information about the household in order to obtain a generic outline of the household. This helps to discriminate and categorize the household in a particular type. The second section is the primary and the most important part of this questionnaire. This section requests to give general information about the use of electric appliances in order to accumulate specific data on when and how often occupants use their electrical appliances at different times of the day. The time intervals for both the questionnaire & feeder data are set to 1 hour so that the load curves can be compared. The third section aims to gather information about lighting use [6].

Cumulative distribution functions (CDF) are used to describe the appliances usage time for different groups of consumers and it plays a vital role in the methodology that is

presented here. The groups were categorized by demographic distribution. This methodology also goes through the process of briefly describing the probability distribution of their random operations. The CDFs are based on actual values from the surveyed data sample which gives information about households' occupancy. This represents the probability of performing a particular activity for a household during any hour of the day. This method can be used to predict the daily load profiles starting from the individual dwelling to community. A further modification was done by normalizing the CDF which assisted in analyzing the profiles [7]. This was used to make the comparison visually comprehensible.

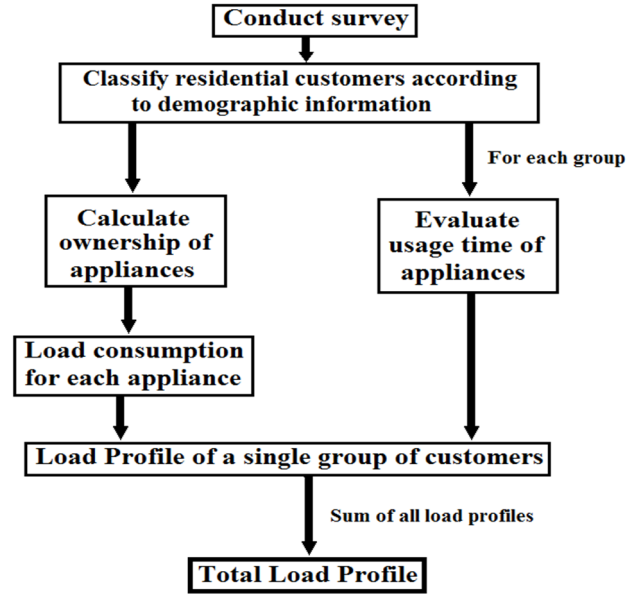


Figure 2. Algorithm of Load Profile Generation

IV. RESULTS AND ANALYSIS

The proposed methodology is implemented through a survey questionnaire among the inhabitants of Japan Garden City located at Mohammadpur, Dhaka (GPS location Latitude: 23.76473, Longitude: 90.358066, Altitude: 9 meters). This survey was carried out to generate data for 400 households and form an acceptable pattern of their household energy consumption.

A. Categorizing and allocation of household types

It goes without saying and is almost certain that there is a difference in knowledge about electricity consumption amongst the consumers in different households, and these households have different attitudes and different energy-using or saving practices because of the range of variety and diversity of people in the Dhaka (Bangladesh) population. Among the family members, children, housewives, senior citizens and unemployed members occupy the households almost all the time. Different households use various kinds of appliances. After considering all these factors, the consumers were categorized into 5 groups.

Allocating households to different types:

TABLE I. HOUSEHOLD CATEGORIES

Household Types	Percentage Share (%)
Three adults	14
Four adults	32
Five adults with children	14
Five adults	25
More than Five adults	15
Total	100

The information in Table I for 400 households in test community was analyzed further and also the physical location of the households was allocated using a rand function in Excel.

B. Electric appliance consumption and ownership pattern

Findings from survey as illustrated in Table II shows that all of the households own a television, fridge and computer. A large number of households own microwave oven and iron. Air conditioner, washing machine and even kettle are owned by good number if not majority of the households. Dish washer, vacuum cleaner, video recorder is also owned in a small number of households. Table II shows the ownership level of appliances from the questionnaire survey.

TABLE II. CATEGORIES OF APPLIANCES

Appliance Type	Ownership Level (%)
Computer	100
Kettle	21.42
AC	46.42
Fridge	100
Microwave Oven	75
Video Recorder	3.57
TV	100
Washing machine	32.14
Dish Washer	10.71
Iron	75
Vacuum Cleaner	10.71

C. Time of Use Probability Profiles

The time of use probability profiles attempts to measure the probability of the particular activity being commenced or undertaken as a function of time of day. The probability profiles represent the probability of a household carrying out a specified activity during a 24 hour period.

Generally, relative frequency is used for approximation of the relative probability of an event. Relative frequency can be said as the fraction of times that the events occur. To be more precise, it is the number of times an experimental event occurs, divided by the total number of trials [8]. Relative probability is given by the following equation:

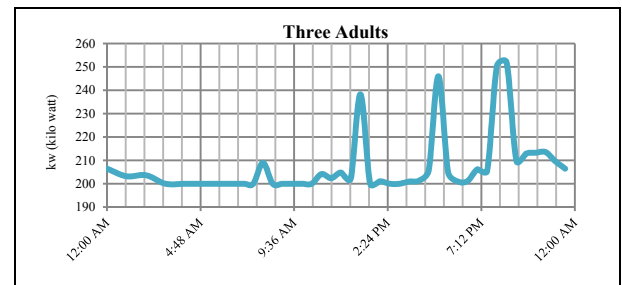
$$P_R \approx \frac{N_x}{N_t} \dots \dots \dots (1)$$

Where P_R is the relative probability, N_x is the number of times an event x occurred, and N_t is the total number of trials. These data are obtained from the questionnaire survey and a behavior pattern was formed which assisted in analyzing the generated load profiles and had an effect on the overall survey system.

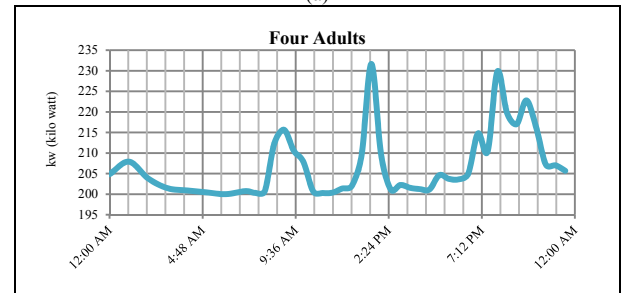
D. Load profile generation

The load profiles are created and implemented in Excel for the creation of load profiles on an hourly basis after midnight to 7 AM and on a half hourly basis for the rest of the day up to midnight with a view to generating daily electrical load profiles for the household types. Microsoft Excel has been used to plot the profile due to its acceptance as powerful and widely used Spreadsheet software developed by Microsoft [9].

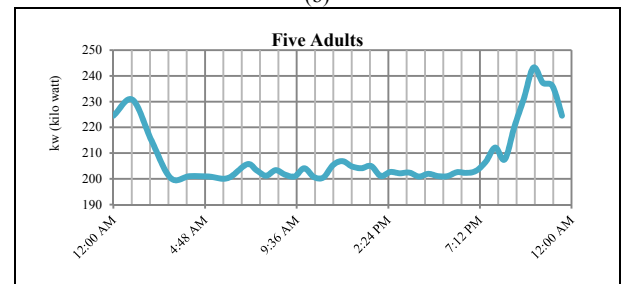
The effect of electrical hydraulic pumps lifts an additional deep freezer and many more heavy loads create peaks in the load curve. Data collected from the substation are not purely residential. A substation gives power to schools, colleges, small hospitals, mosques, madrasas etc. The substation gives no output when it is under maintenance or faults. The power factor was assumed to be 0.9 all the time.



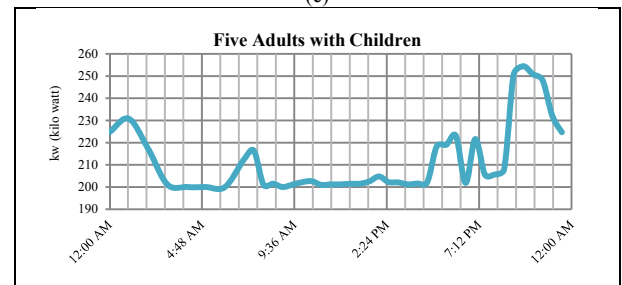
(a)



(b)



(c)



(d)

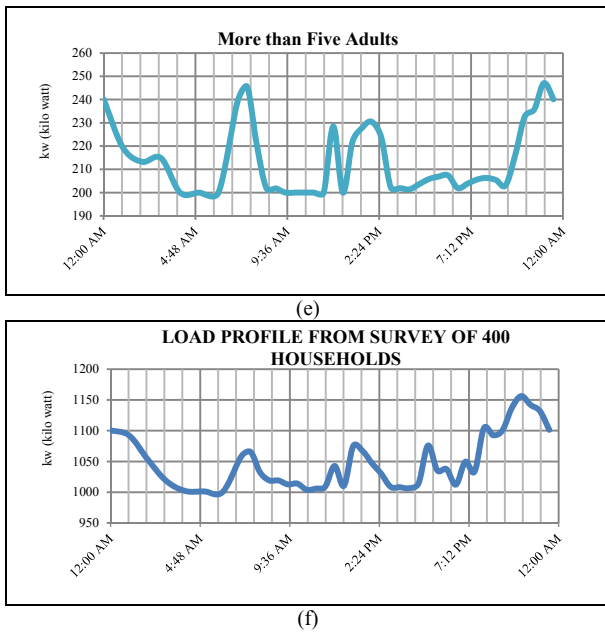


Figure 3. Load Curves of Different Types of Households. (a) Load curve for three adults, (b) Load curve for four adults, (c) Load curve for five adults, (d) Load curve for five adults with children, (e) Load curve for more than five adults, (f) Load curve for 400 households.

Figure 3 shows the load profile of different types of families in Dhaka city. All of the curves show a significant peak starting around 7:30 PM. Since all the members of the household are present and are occupied with several electrical appliances during that period, the peak is relatively higher and its duration is much wider than any other peaks at different times. The peak starts to fall during midnight when people of the city are prone to sleep.

Other than that some small peaks are found early in the morning around 9:00 AM when people get prepared to go to offices or schools & colleges etc. And the rest of the time all the full time loads are running which consume constant power e.g. fridge.

Another noticeable behavior among these curves is that few households create a peak at around 1:00 PM. The specific reason of this peak is still unknown. However a hypothetical guess is that people might use washing machines at that time creating the small peak. Also school going kids return home and turn on the television, microwave oven, air cooler etc.

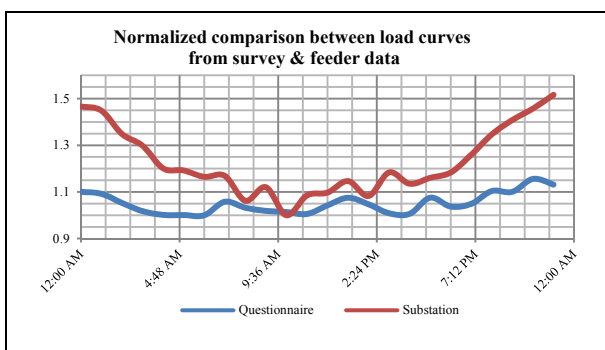


Figure 4. Normalized Comparison Among Load Curves of Different Methods

Figure 4 represents a comparison between the two load-curves found by using two different methods. The curves have been normalized in order to make the comparison visually easier. Since the feeder data from the substation is not purely residential, the curves have some deviations between them. However the curve found from the survey does not decrease that much like the other one. The reason for this deviation could be due to some illegal steel & welding factories which start their work during night time up to sunrise. The usage of water pump which can create a noticeable change in a load curve was not included in the questionnaire since the customers do not have the access to it. The water pump is controlled centrally by the authorities of that area. These pumps are usually operated at midnight for about 2 hours and 30 minutes which create a large amount of energy consumption. From the figure it is clear that the rising and falling of the peak of the two curves occur at the same time and the timing of the peak in day time is somewhat similar to that of the other curves.

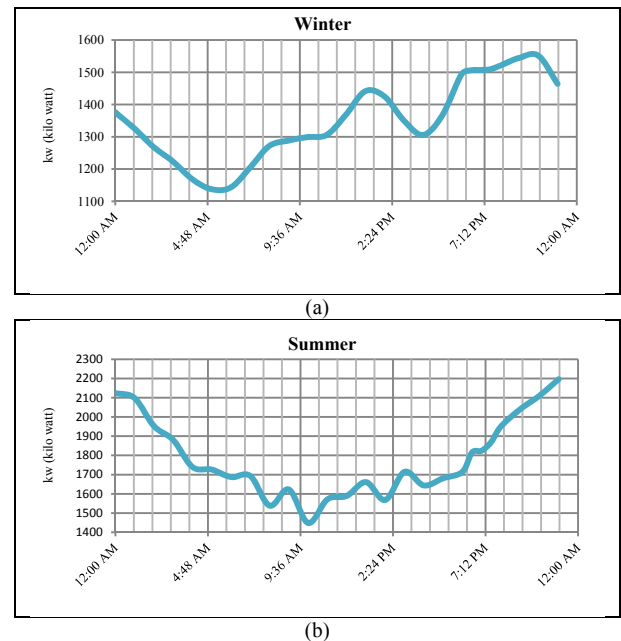


Figure 5. Load Curves for Different Weather Factors of Residential Users of Dhaka City. (a) Load curve during winter season, (b) Load curve during summer season.

Figure 5 illustrates the load curve of the people of Dhaka city during summer & winter season. The curve has been generated by using the data collected from the substation. One year data of electricity consumption from a substation feeder of a residential area were taken. The load curve generated was based on feeder data collected from substation located at Japan Garden City, Mohammadpur, Dhaka. The feeder data represented consumption of around 1500 customers connected to it. The curve of summer shows an increase in consumption starting from around 5:00 PM which starts to decrease at around 1:00 AM. Fulfilling this huge amount of demand is a great challenge for Bangladesh Government. The lowest consumption which is around 1400 kW during summer season is the highest consumption during winter. The reason behind it is the usage of air cooler and ceiling fans. During summer the usage of these appliances increase at a great number but during

winter the usage decreases drastically. The increase of consumption after evening during winter is because of the lighting load, TV, computers etc.

V. CONCLUSIONS

This paper presents an efficient process for the classification as well as load profiling of the residential consumers residing in Dhaka city. The main objective was to generate load profiles for residential users in Dhaka city in a particular method where the behavior of the occupants and appliance end-users are correlated. Two methods were utilized in order to make a comparison between real data which was collected from a feeder and data found by conducting a survey on residential customers. This paper gives a better approach to reveal the reasons behind the variations in the load curves. The load profiles varied significantly between communities due to different occupancy types, different housing data, and different consumer behavior. The results can be used to forecast the load demand in that region. Findings of this work would also be helpful in adopting correct measure of demand side management for residential customers. Future work would focus on improving the level of prediction accuracy of proposed method in comparison to other statistical prediction techniques of load profiling.

ACKNOWLEDGMENT

The authors would like to thank Dr. Tareq Aziz, Ahsanullah University of Science & Technology for giving his support and guidelines to conduct this research. The authors would also like to thank DPDC for permitting and the authorities of Japan Garden City substation for providing feeder data.

REFERENCES

- [1] Union of the Electricity Industry, "Metering Load, Metering Load Profiles and Settlement in Deregulated Markets", March 2000.
- [2] Weblink: <http://publications.aeic.org/lrc/ArnaudGrandjean2.pdf> (last accessed on June 19, 2016).
- [3] V. Figueirido et al., "Electric Energy Customer Characterization by Clustering", IEEE Intelligent Systems Applications to Power Systems, Lemnos, Greece, January 2003.
- [4] Available at, buet official website: "<http://www.buet.ac.bd/ces/conventional-energy.htm>" (last accessed on June 15, 2016).
- [5] Abdel-Baset Mostofa Imbarek Ihbal, "Investigation of Energy Demand Modeling and Management for Local Communities", Ph.D. Thesis, School of Engineering, Design and Technology, University of Bradford, 2012.
- [6] Syed Ishraq Chowdhury, Abidur Rahman, Raivy Shahriar, Ishar Ibtisam, Mirza Rafsan Rahat, "Developing Load Profile for Domestic Customers of Dhaka City through Statistical Prediction", B.Sc. Thesis, Department of Electrical & Electronic Engineering, Ahsanullah University of Science & Technology, 2015.
- [7] Weblink: <http://www.mathworks.com/help/stats/normcdf.html> (Last accessed on June 19, 2016).
- [8] Ken Black, Business Statistics: Contemporary Decision Making, 6th edition, John Wiley & Sons Inc., 2010.
- [9] Weblink: <http://office.microsoft.com/en-us/excel/> (last accessed on June 15, 2016).

THE QUESTIONNAIRE

Enter a number

Put $\checkmark$ marks to fill up the table.

[illegible]

C1: How many light bulbs in each room are usually used? Please enter a number.

Bedrooms	
Kitchen	
Bathrooms	
Living Room	
Others	

C2: Put $\checkmark$ marks to fill up the table.

Time	Bathroom	Kitchen	Living Room	Bedrooms	Others
1:00 AM					
2:00 AM					
3:00 AM					
10:00 PM					
11:00 PM					
12:00 AM					

A Feasibility Study & Simulative Development of Hybrid Power Plant Using Renewable Energy at Isolated Island Kutubdia in Bangladesh

Sree Bash Chandra Debnath

Department of Engineering and Computer Science
University of Trento
Trento, Italy
sreebashcuet@gmail.com

Md. Rashedul Islam

Dept. of EEE
Premier University
Chittagong, Bangladesh
rashed.eee38@gmail.com

Farhana Afroz

Dept. of EEE
Chittagong University of Science & Technology
Chittagong, Bangladesh
faru_042@yahoo.com

Biki Barua

School of Engineering (Telecommunication)
Université du Québec en Abitibi-Témiscamingue
Quebec, Canada
biki.barua@uqat.ca

Samina Alam

Dept. of EEE
Chittagong University of Science & Technology
Chittagong, Bangladesh
samina0602107@yahoo.com

Abstract—To satisfy the demand of electricity with efficiently is a challenging work to the developing country like Bangladesh. Moreover, the radical utilization of conventional energy forces the authority to think about the alternative source. Also, the energy system faces major problem being distributed the power in remote area from a centralized system. Seeing these complications, some solutions have been proposed but still left a good number of problems. In this paper, the study of power demand and feasibility analysis of renewable energy has been done in a remote area such as Kutubdia island, Bangladesh. Alongside we proposed an integrated model of the power plant for meeting the demand of power or electricity in the remote area.

Keywords—feasibility study; renewable energy; weibull's distribution; cost; integrated model; homer.

I. INTRODUCTION

The demand of the electricity in the global world is a major issue in the era of 21st century. But the global warming effect via the conventional energy put the demand of electricity production in the challenging aspects. Moreover, the inadequate production of electricity to aside the economy & life of remote area in danger as well as out of progress. In Bangladesh most of the power plant is based on conventional energy mostly depends on natural gas. According to the recent

survey of BPDB, it has shown that the reservation of natural gas is going to be decreased & it's won't be enough to sustain the production. So it's a big concern to meet the demand & to sustain the progress of production of power involving the alternative energy along with conventional energy. Following the policy on 18 December 2008, Bangladesh government has been targeted to produce the 5% of the total electricity by 2015 & 10% by 2020 from renewable energy [1].

Long distance of sea from the national grid to Island, it is almost impossible to maintain grid-connection to the Island. That's why the immersion of renewable energy for conveying power in island or not interconnected areas are necessities to get expand in light up. In general bio energy-bio fuels, solar, wind, hydropower energy is the main forms of the renewable energy. Among all of these sun and wind are the main sources of renewable energy. Following the geological condition, Bangladesh considers as a perfect placement for renewable energy. The statement of global analysis indicates that Bangladesh experiences average 7.6 hours sunshine in a day. [2]. Unfortunately, the sun & wind depend on the weather, as a result, both are uncertain & uncontrollable as well as can't be depended on as independent power generation units [3]. So, to meet the demand & make the electricity production sustainable, the concept of combining several sources with the storage device is coming out, that's considered as a way to

overcome the pervious drawback along with to build the entrance level of renewable vitality of the framework [4]. This types of power generation systems are called Hybrid Power Systems (HPS).

In our research work, we did the feasibility test of renewable energy for Kutubdia Island. There is a 250KW diesel power station in this island. It is latent since quite a while, due to low proficiency, costly, troublesome and very unsafe and tedious to transport diesel to this disconnected island. Another alternative source of power of the island is 1MW capacity Wind Battery Hybrid Power Plant (WBHPP). BPDB have designed 11KV isolated grid between two turbines and weak control system. But the problem is 1000pcs 12 V DC rechargeable batteries as it's unable to supply power without charge of its battery though the availability of wind velocity. In our work, cost estimation has been shown for using renewable energy in Kutubdia & we proposed a new integrated concept & simulative approach of power plant using Homer software.

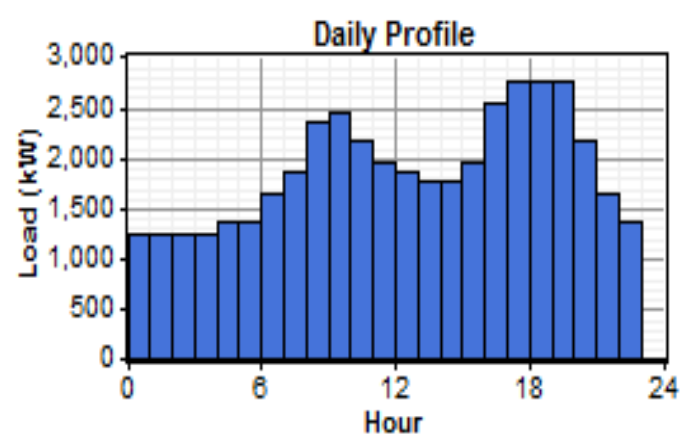


Fig. 1. Daily load courve of Kutubdia island.

II. FESABILITY ANALYSIS OF SOLAR ENERGY IN KUTUBDIA

Geographically Bangladesh is a good side of solar radiation & Kutubdia is situated between 21°54.71 North Latitude & 91°52.43 East Longitude which is an ideal location for sun-powered vitality usage. Regularly the approximate variation of solar radiation in between 4 to 6.5 kWh per square meter. The framework of the radiation is maximum in the month of March-April and minimum in December-January [12].

By analyzing the data from [8], [11] and [12], we have a demonstrative idea about annual rain, humidity, temperature, sunshine and radiation for Kutubdia island in table I.

Month	Rain Fall	Humidity	Temp.	Sun Shine	Radiation
Jan.	0.00	2.40	2.70	8.00	2.70
Feb.	0.00	2.30	4.00	6.70	4.20
Mar.	0.00	2.20	4.70	8.40	4.50
Apr.	0.40	2.60	4.50	8.30	4.90

May	4.70	2.60	4.40	6.00	4.80
Jun.	4.30	2.70	4.30	2.70	4.70
July	6.70	2.80	4.30	4.30	4.20
Aug.	4.80	2.90	4.50	4.60	4.50
Sept.	4.30	2.70	4.40	4.50	4.40
Oct.	4.20	2.70	4.30	4.80	3.20
Nov.	0.20	2.50	4.00	6.20	3.10
Dec.	0.00	2.4	2.9	6.7	2.5

From the given data's, it can be said that Kutubdi island is the ideal spot for utilizing sun oriented vitality and also to meet the power demand.

III. FESABILITY ANALYSIS OF WIND ENERGY IN KUTUBDIA

In Bangladesh, satisfactory data of wind velocity over the nation and especially on wind speed at center point statures of wind machines is not accessible. In Kutubdia, the wind monitoring stations (91.85°E, 21.82°N) of BMD (Bangladesh meteorological department) reports that the annual average wind speed over Kutubdia, in last 7 years (2008–2014) to be 1.8 m/s at a height of 13m. But after that, some measurements has done by BCAS (Bangladesh Center of Advance Studies) & finds the different data that's near to seashore, (91.84°E, 21.86°N) prompts a yearly estimation of 4.4 m/s at a tallness of 25m. Alternatively, the average data from NASA over 1° latitude × 1° longitude is 3.6 m/s in 50m height around Kutubdia [10]. Based on this data the variation of the wind at Kutubdia in the period of 2002 to 2012 [11] is shown in figure 2 and 3. Where it's also showing the mean wind speeds in Kutubdia were overhead 5.5 m/s in every year from 2008 to 2014 [11].

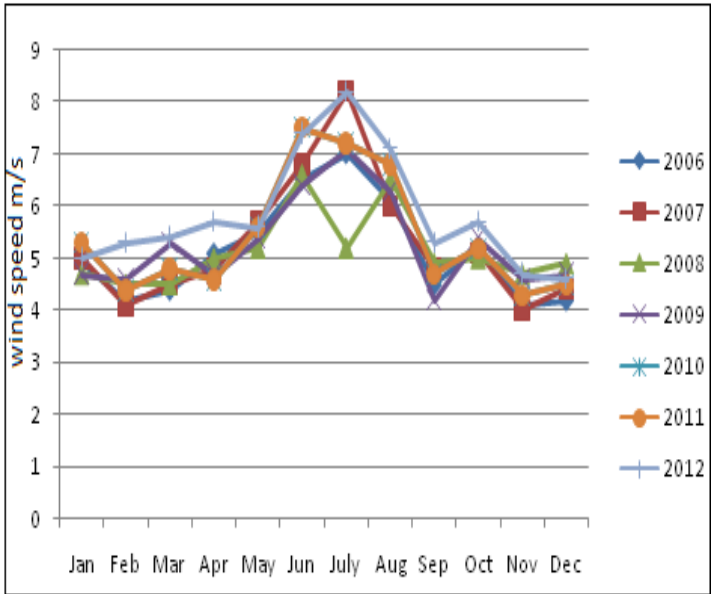


Fig. 2. Wind Speed Variation (Monthly) at Kutubdia Island (2006-2012).

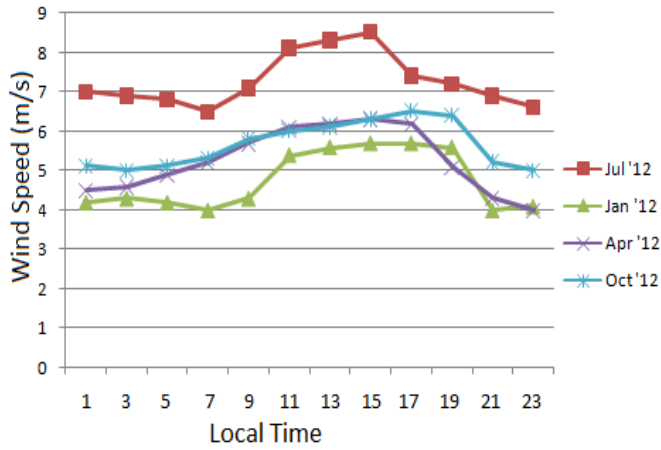


Fig. 3. Wind Speed Variation (Hourly) at 2012 in Kutubdia Island.

A. Analysis of Weibull's Distribution

We used Weibull distribution algorithm, which is basically a statistic algorithm & popular because of it's some special character like to aptitude & ranging from various operation field as well as for observation in the domain of economics, health, physical, biology & in the engineering sciences. There are three common methods for calculating weibull parameters like Weibull scale factor (C) and Weibull shape factor (K) and these are: (a) Weibull paper method, (b) Energy pattern factor method & (c) Standard deviation method. In our work, by using existing acknowledge data we make the solution to find the value of Weibull shape factor (K) & Weibull scale factor (C). The simulation result is shown below:

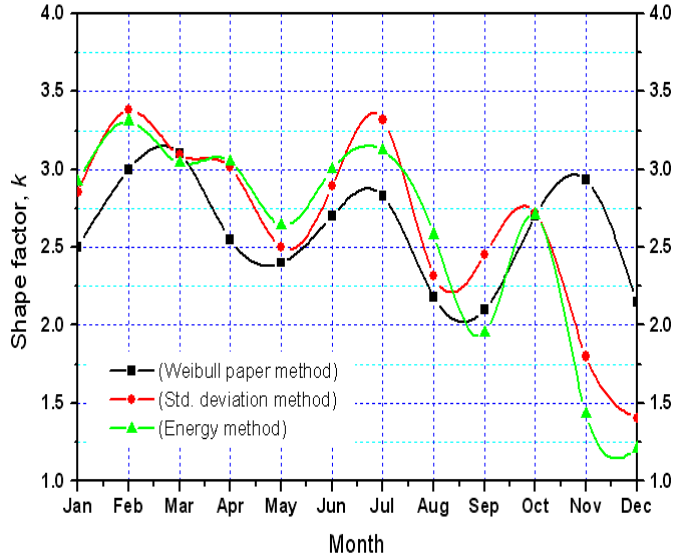


Fig. 4. Variation of Weibull's Shape factor (k) at Kutubdia Island.

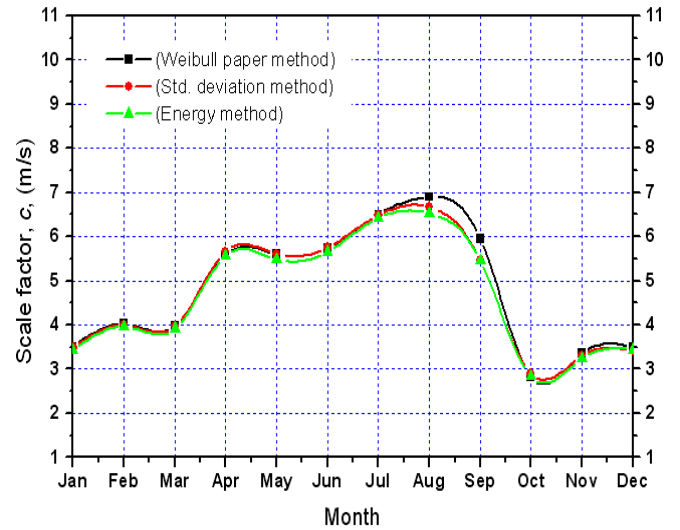


Fig. 5. Variation of Weibull's Scale factor (c) at Kutubdia Island

The solution of shape factor (K) and scale factor (C) were resolute having monthly consideration. Where the value of K rests in between 1.4 to 3.38 and the value of C rests between 2.83 to 6.90. In general the functions of the Weibull's distribution follow the Raleigh function ($K=2$) for the selected site. So, it's showing that the wind waves have enough potential for power generation.

IV. PROPOSED MODEL

Our proposed integrated model is consisting up with the combination solar & wind energy system, that's projected to continue the operation without any interruption. Here design & analysis has been done using Homer software.

A. Plant Description

The projected framework of the power plant is shown in Figure 6. The system is composed up with 20 sections solar array system & 4 sections wind turbine System. To form the PV array the PV Modules can be wired together. Where the series wiring increases the available voltage & parallel wiring increase the available current, but in either way the production of power is same. A typical PV module has a size of about 2.6 feet by 5.2 feet and with the full sun radiation; the rating of DC electricity is 220 watts. In case of wind turbine three-bladed wind turbines are operated as "upwind," and as soon as wind turns the blades, the shaft starts to spin, that's associated with generator and produces electricity & the approximate up-to 20 kW. In the following projection, each section will be referred to as a "string." & each string consists of a 50kW solar array, two 25 kW charge controller connected in parallel, and a pair of parallel connected inverters. Here charge controller is using maximum power point tracking (MPPT) system for enhanced power usage. The utility transformer inverts the voltage level from 440 V level to 11 kV.

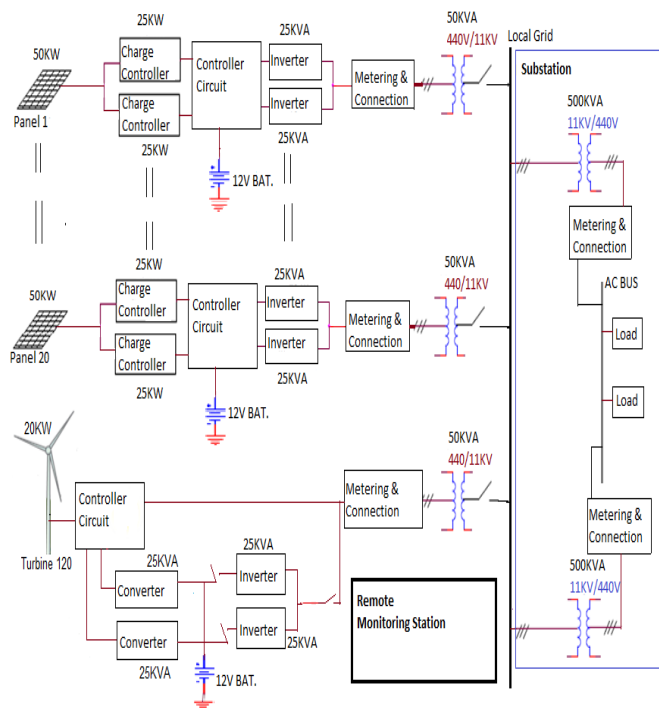


Fig. 6. Proposed Design for the Power Plant.

B. Control System

The controlling development method of this integrated system in lead up the PLC shown in figure 7. The control method depends on the analogue sensors (wind anemometer, solar sensor, voltage sensor), and output contactors. And these sensors must need to be integrated with analogue system or input card. The reading throughout the analog system gets convert into the digital system or output card and shows up in the display.

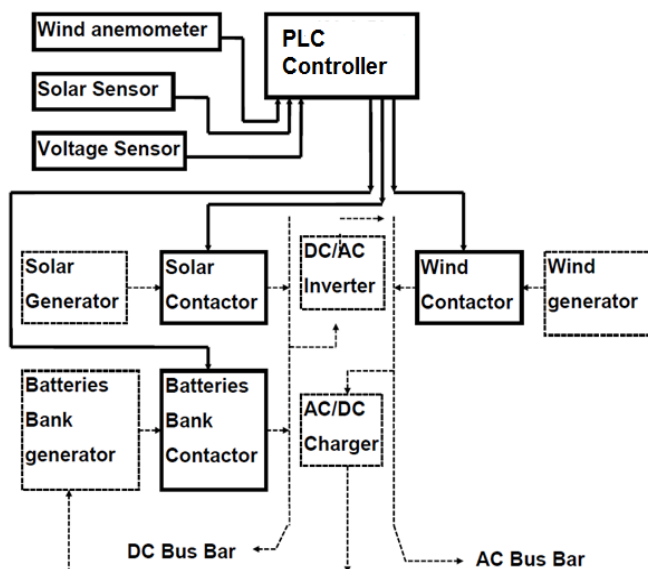


Fig. 7. Block diagram of a control developed method.

C. Load Analysis

For the load analysis, we consider typical loads of a single home. And for estimation, we consider two 65 Watt fans, three 25 Watt CFL bulb, one 15 Watt CFL bulb, one 100 Watt TV (colour) and other small loads. In HOMER, we made a 24 hours load profile for a single home. Among 74,530 families in the island, if we consider an average load demand of all the houses is same & determinate demand for the island is 43.6MW/day. And projected simulative result is shown in figure 8.

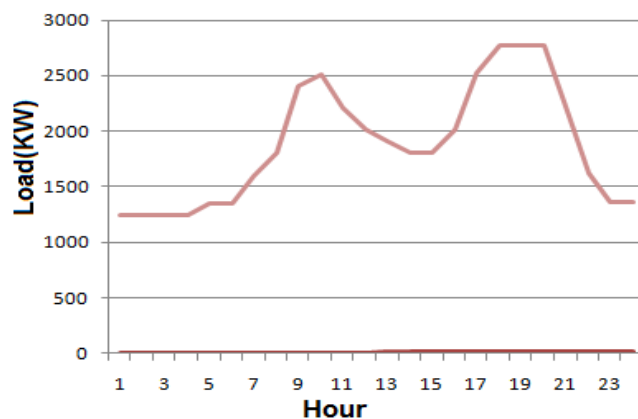


Fig. 8. Average load for total houses of Kutubdia Island. (Hourly).

D. Cost Analysis

For the analysis of cost, we have utilized the net present worth (NPV) strategy to ascertain the monetary possibility of the hybrid energy system. We try to figure out whether the proposed crossover framework pays for itself in a time of 20 years. Our cost analysis has been composed up with the initial project cost that includes equipment and installation cost as well as calculates net present value, using a one-year interval that estimates operation and maintenance cost. So a cost base demonstration is given in table II:

TABLE II. TOTAL COST OF HYBRID SYSTEM

System Name	Power(MW)	Cost(Tk)
Wind System	2.4MW	2400000
Solar System	1.3MW	1879750
Total Power =3.7MW		Total cost of Svsystem=4279750

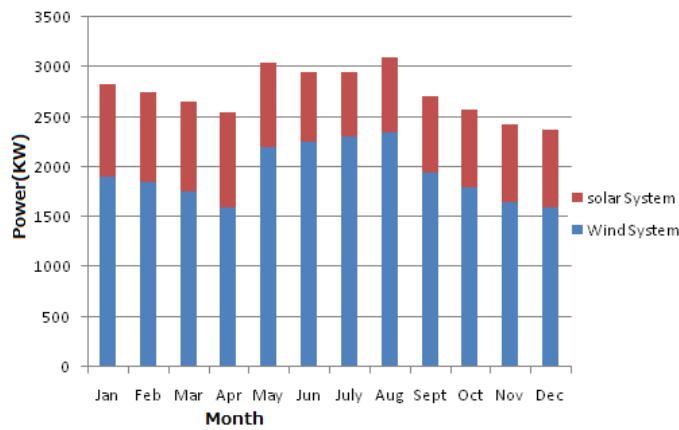


Fig. 9. Monthly Average Electric Production from Solar Module and Wind Turbine.

E. Comparison

Table III is showing the cost analysis of different sources for power plant. And figures here described the cost variation with our model to the other's sources.

TABLE III. DIFFERENT POWER PLANT PROJECTED COST

Energy Technology	Fuel Cost KWh(TK.)	Initial Investment KWh(Tk.)	Transmission and Distribution Cost KWh(TK.)	Total Cost per KWh(TK.)
Coal	7	2.25	2.15	11.25
Oil	30.10	2.20	2.15	34.45
Gas	1.12	1.10	1.00	3.22
Integrated System(PV+ Wind)	0.00	5.50	1.50	7.00

V. CONCLUSION

In this paper, we proposed a simulative hybrid power plant, that works well for the remote area like Kutubdia island. Where, we used the renewable resources for electricity production in the island to fulfill the demand of power. The key advantages of the proposed system are different suppliers can be selected for every string without modifying the performance of the overall system. Individual string power output can be monitored and efficient processing units can be singled out. In

future studies, we are planning to involve the MPPT system as well as intelligent control system.

REFERENCES

- [1] Islam Sharif "Renewable energy development in Bangladesh", Executive Exchange on the use and integration of Renewable Energy in the Power Sector, Madrid, Spain, October 19-23, 2003.
- [2] Shakir-ulhaqueKhan, Towfiq-ur-Rahman & Shadat Hossain, "A brief study of the prospect of solar energyin generation of electricity in bangladesh " Multidisciplinary Journals in Science and Technology, Journal of Selected Areas in Renewable and Sustainable Energy (JRSE), June Edition, 2012.
- [3] Zakaria Mahbub, Husnain-Al-Bustam, Suvro Shahriar, T.M. Iftakhar Uddin, Abrar Saad, "Comparison Between Conventional Power Generation And Renewable Energy Power Generation In Bangladesh," International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com, Vol. 2, Issue 2,Mar-Apr 2012, pp.896-902
- [4] Sarkar M and Hussain M, The potential of wind electricity generation in Bangladesh, Renewable Energy 1991; 1; 855-7.
- [5] Eusuf M, Bangladesh Centre for Advanced Studies, Wind Energy Study (WEST) Project, Bangladesh, Final Report, 1998.
- [6] GTZ, TERNA Wind Energy Programme, Final Report, Bangladesh, 1997.
- [7] Renewable Energy Information Network, Wind Energy Resource Mapping, LGED, <http://www.lged-rein.org/wind/werm.htm>
- [8] Renewable Energy Development & Energy Efficiency Measures, Bangladesh power development board website, <http://www.bpdb.gov.bd/renewables.htm>
- [9] Burton T., Sharpe D., Jenkins N., Bossanyi E., "Wind Energy Handbook", John Wiley & Sons Ltd., England, 2001, pp. 41-170.[wind powe-7]
- [10] Wind Energy Study (WEST) Project, Bangladesh Centre for Advanced Studies, Final Report, Bangladesh, 1998. Wind Energy Study (WEST) Project, Bangladesh Centre for Advanced Studies, Final Report, Bangladesh, 1998.
- [11] Bangladesh Metrological Department(BMD) <http://www.bmd.gov.bd/bmd/>
- [12] Bangladesh Renewable Energy Report, APCTT-UN.

Forecasting the Long Term Energy Demand of Bangladesh Using SPSS from 2011-2040

MD Zakaria Rahman, MD. Nahiduzjaman Sajib, M.M.Safayet Hasan Rifat, Md Hossam-E-Haider,
Md. Ali Azam Khan
Electrical, Electronic & Communication Engineering
Military Institute of Science and Technology (MIST)
Dhaka, Bangladesh
zakariarahman.mist@gmail.com

Abstract—Energy consumption is promptly increasing in most developing countries since they foster their economic growth. Modern technological electrical devices in day to day activities require high consumption of energy. People's requirement for higher living standard and rapid population growth are among the determinants in the growth of energy consumption. For optimal use of facility systems and power sources forecasting energy consumption is very important. This paper portrays the forecast of long term energy demand of Bangladesh as dependent variable and independent variables are GDP per capita, price of oil, population growth and lagged electricity consumption. The findings are similar and correspond with the assumptions in the relationship between variables.

Keywords—Energy demand forecasting, electricity demand, natural gas consumption, oil consumption, SPSS, long term demand, energy scenario of Bangladesh.

I. INTRODUCTION

Developing and emerging world economies are facing the challenge of meeting the demands of billions of people with no electricity and concurrently transition to clean, low carbon energy sources. A report shows that there were about a billion people living on our planet just few centuries ago. A hundred years later, in the year 1900, the world population grew a moderate amount to about 1.6 billion people [1]. But, over the next century, the world's population exploded, and today, there is about seven billion inhabitant on planet Earth. Around 1.6 billion of people in the world are living without electricity as of 2009 [2]. Bangladesh, which is a country in Southeast Asia, is of great significance to world energy markets because of its enormous potential natural gas reserves. Bangladesh's position near India, moreover in Southeast Asia, further makes it a potentially important regional energy trading hub. The population of Bangladesh is above 130 million and the population density is 750 per sq. km which is among the highest in the world [3]. Roughly 62% of the population has access to electricity [6]. The per capita energy consumption in Bangladesh is one of the lowest

(321 kwh) in the world. There is sufficient room to unfold if the country is willing to comply with the changes. Every year, a lot of subsidy is provided for the power sector. Energy professionals has pointed out that the fundamental reason behind this is the country is highly dependent on fuel oil to produce its electricity, adding that there would not be any viable development in the energy sector unless such dependency is reduced. Like prior budgets, the imminent one is also reportedly seeing a surge for energy sector subsidies. Looking at the data of last six years, it can be seen that subsidies for the energy sector have seen a growth of over 100% during this period [4]. The energy demand from 2010 till 2040 is expected to grow rapidly. The present energy supply cannot be treated as a sustainable source of energy since the costs of energy are exponentially increasing. Forecasting energy consumption is very important in order to optimally use facility systems and power sources. With a good energy demand forecasting, the quantity and quality of energy supply can accomplish the urgency of consumers with minimum operational cost. Daily forecasts are required to schedule generating units for optimum cost-cutting measure. Weekly or monthly forecasts are required for maintenance scheduling. The purpose of this research is to forecast the long-term energy demand of Bangladesh for 2010-2040 and to provide some mathematical data which can help the policy makers to formulate a particular policy in the field of energy supply in Bangladesh.

II. METHODOLOGY

It is very important to forecast energy demand precisely. An underestimation can cause insufficiency, which may further result in poor service. An overestimation can allow the approval of some facilities that may not be needed for quite a few years. A developing country like Bangladesh hardly has enough resources to cover such a cost without offsetting revenues. Therefore, it is very important to

forecast the future energy demand. Based on the forecast the government should make plans to construct some facilities and begin development of these facilities so that they can meet up the future demand. Energy demand forecasts can also help determine the impact that a new technology will have on the energy consumption. This will help the planners understand and calculate the benefits of investing in the new technology. Many forecasting techniques have been developed to deal with all of the above, ranging from very simple extrapolation methods to more complex time series techniques, optimization methods, or even hybrid models that use a combination of these for purposes of projection.

A. Time Series Methods

A time series is a set of data values of a particular variable. Time series models are basically econometric models. It uses the past values of some variable in order to forecast the future. The perception behind time-series processes is that the prospective behavior of variables is related to its prior values, both authentic and predicted, with some adjustment built-in to take care of how past understandings averted from those expected [10]. Thus, the data of last 20 to 30 time periods are vital prerequisite for a time series forecasting technique. The variation between econometric models, which is based on time series data and time series models lies in the analytical variables used. Nearly, the most important purpose of time series comes from the development of short-term forecasts. Moreover time series models are structurally simple. They do not require accumulation of data on numerous variables. A disadvantage of these models is that they do not explain a cause-and-effect conjunction. Thus, a time series does not provide intuitions into why changes occurred in the variable. Usually, while analyzing time series data by using econometric methods or time series models, there exists some technical difficulties where more than one variable is highly correlated with another. This is called multi-co linearity.

B. Modeling Tool

IBM SPSS Forecasting is used for time series module. A time series is a set of data obtained by measuring a single variable frequently over time [9]. Time series forecasting predicts the future events based on known prior events. As opposed to regression techniques, in time series each of the cases are correlated to each other, as they exhibit the same phenomena over time. Because of this, the time factor is in itself, a forecaster of the dependent variable. In other words, it can be said that in time series, the past provides a model for the future. In order to increase the accuracy of the forecast time series analysis uses

some independent variables. For instance in this forecast GDP, population growth as independent variables have been used to help predict the future energy demand. GDP per capita, income per capita and population growth of Bangladesh is given in Table I below.

TABLE I. GDP per capita, income per capita and population growth of Bangladesh [2] & [6].

GDP per capita	957.82 USD
Income per Capita	1314 USD
Population Growth	1.20%

Instead of manually defining the parameters and other settings of time series models here we have used the Expert Modeller. It is an important feature of the IBM SPSS Forecasting module. It helped to automatically determine and to evaluate the best-fitting ARIMA or exponential smoothing model. Thus it helped to eliminate a good deal of the trial and error which usually occurs if the parameters are manually defined.

III. RESULT AND DISCUSSION

Energy security is the most important issue for all over the world. Uninterrupted energy supply is the main prerequisite for the development of any country. In Bangladesh the per capita commercial energy consumption is among the lowest within South Asia. Therefore it is essential to know the energy requirement in the future to take proper steps. In the section below, the energy projection by using IBM SPSS modeling software for long term energy forecasting for Bangladesh has been calculated.

In this section three major energy sectors (natural gas, electricity, oil) of Bangladesh from 1990 to 2010 has been analyzed and the future energy demand up to 2040 has been forecasted.

The Consumption of Natural Gas, Oil and the total demand of Electricity from 1990 to 2010 are provided to SPSS to forecast the future. The provided data is shown in the Table II with one year interval.

TABLE II. Consumption of Natural Gas, Oil and the demand of Electricity From 1990-2010 [4], [5], [11] & [12].

Fiscal year	Natural Gas Consumption (Billion Cubic Feet)	Oil Consumption (Thousand barrels per year)	Electricity Demand (MW)
1990-91	167.83	13140	1731
1992-93	188.48	13618.15	1826
1994-95	223.76	16768.1	2038
1996-97	365.51	19027.45	2419
1998-99	282.02	21330.6	2881
2000-01	332.35	25112	3394
2002-03	391.53	30211.05	3947
2004-05	454.59	30893.6	4597
2006-07	526.72	33131.05	5112
2008-09	600.86	34335.55	6066
2010-11	703.6	39379.85	6765

A. Natural Gas

In 1990 the total natural gas consumption was 167.83 billion Cubic Feet while in 2010 it was 703.6 Billion Cubic Feet. It has already become greater than four times within this short duration. The data of the Natural Gas Consumption from 1990-2010 are shown in Table II. Analyzing the consumption of 1990 to 2010 we can forecast the data using SPSS modeling software. In Fig. 1 projected values are shown graphically.

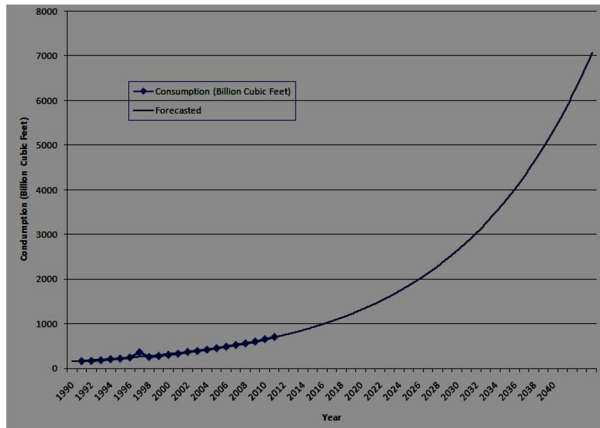


Fig. 1. Consumption of natural gas from 1990-2010 and forecast of the consumption from 2011-2040.

The forecasted data of the consumption of gas from the year 2011-2040 are shown in Table III. From Fig-1 and Table III it is seen that at 2040 the consumption of natural gas will be 5267.11 Billion Cubic Feet.

B. Petroleum Oil

In 1990 the consumption of petroleum oil was 13140 thousand barrels per year on the other hand, in the year 2010, consumption became 39379.85 thousand barrels per year which is almost 3 times in 20 years.

Analyzing the consumption from 1990-2010 of Table II we can forecast the data using SPSS modeling software. From figure-2 it is assumed that the consumption of petroleum oil will reach 210284.4997 thousand barrels per year by 2040.

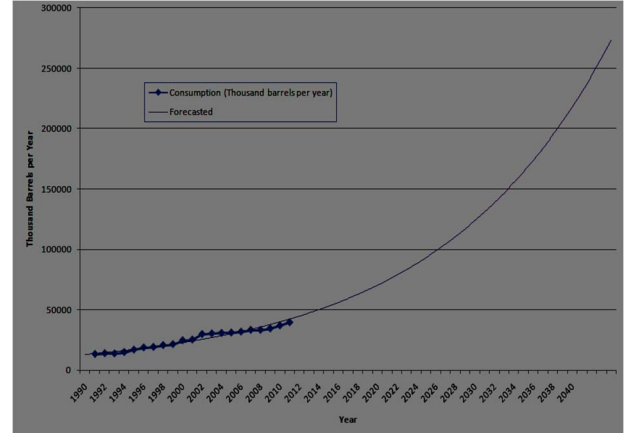


Fig. 2. Consumption of Petroleum oil from 1990-2010 and the forecasting of the consumption from 2011-2040.

The year wise forecasted data from the Fig. 2 are given at Table III with one year interval.

C. Electricity

The demand of electricity consumption became 6765 MW in the year 2011 where as it was 1731 MW in the year 1990. The electricity consumption rate increases rapidly as Bangladesh is a developing country. People of this country are becoming so much dependent on electricity as its rural areas are also developing rapidly. It has become tougher than before to meet up this huge demand for the government.

The graph of electricity consumption Fig-3 has been plotted based on the data of Table II above, the future consumption shown by SPSS modeling software. We found that, in the year 2040, the consumption will become 57447.06 MW.

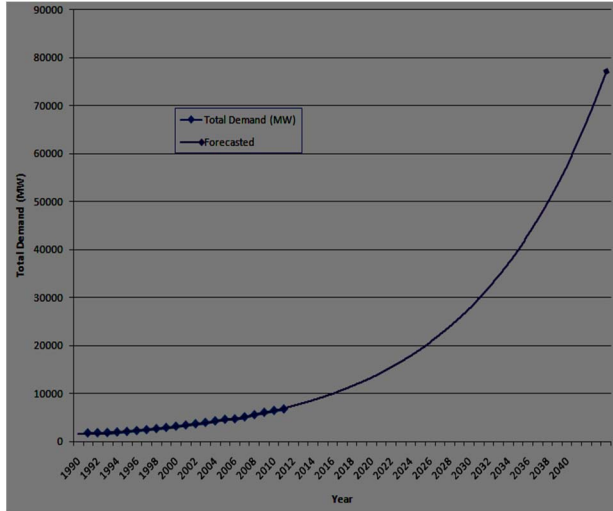


Fig. 3. Total demand of Electricity from 1990-2010 and the of the forecasting demand from 2011-2040.

Year wise projected data from the year 2011-2040 from the Fig.1, 2 and 3 are given in Table III with one year interval.

TABLE III. Forecasted data of consumption of Natural Gas, Oil and demand of Electricity from 2011-2040.

Fiscal year	Natural Gas Consumption (Billion Cubic Feet)	Oil Consumption (Thousand barrels per year)	Electricity Demand (MW)
2011-12	691.75	41448.92	6916.18
2013-14	795.7	46361.12	8003.37
2015-16	915.29	51855.52	9261.48
2017-18	1052.83	58001.06	10717.34
2019-20	1211.04	64874.93	12402.07
2021-22	1393.03	72563.45	14351.63
2023-24	1602.36	81163.153	16607.65
2025-26	1843.16	90782.03	19218.31
2027-28	2120.14	101540.86	22239.36
2029-30	2438.74	113574.77	25735.31
2031-32	2805.22	127034.83	29780.8
2033-34	3226.77	142090.11	34462.22
2035-36	3711.67	158929.61	39879.55
2037-38	4269.48	177764.81	46148.47
2039-40	4911.02	198832.22	43402.836
2040-41	5267.11	210284.5	57447.06

Here, the actual data of the consumption of Natural Gas, Petroleum oil and the demand of Electricity for the year 2011 and 2012 have been shown and those data have been compared with forecasted data of 2011 and 2012.

TABLE IV. Comparison between actual data and forecasted data [4], [5], [7] & [11].

Year	Natural Gas Consumption (Billion Cubic Feet)			Petroleum Oil Consumption (Thousand Barrel per Year)			Electricity Demand (MW)		
	Actual Value	Forecasted Value	Error	Actual Value	Forecasted value	Error	Actual value	Forecasted Value	Error
2011	708.9	691.76	2.42%	40150	41448.90	3.6%	7175	6916.18	3.24 %
2012	743.57	741.92	0.22%	41478.60	43836.26	2.6%	7639.95	7439.95	5.68 %

The preceding data table shows that the actual consumption of natural gas was 708.9 billion cubic feet whereas the forecasted data is 691.76 billion cubic feet giving an error of 2.42%. Similarly comparison of the data of 2012 for natural gas consumptions gives an error of 0.22%. Again for the consumption of petroleum oil the projection error was 3.60% and 2.61% for the year 2011 and 2012 respectively. The actual electricity demand for the year 2011 was 7175 MW which is 3.24% more than our projected demand. Similarly for the year 2012 it is 5.68%. The average error for natural gas consumption petroleum oil consumption and energy demand is 1.32%, 3.01% and 4.46% consecutively.

IV. CONCLUSION

Long-term energy demand forecasting is a sophisticated task because it is directly or indirectly influenced by various factors primarily affiliated with the economy and the population. In this research Exponential smoothing method was used. The method provided highly accurate forecasts for the year 2010-2040. The forecasts presented in this project shows that there will be a rapid increase in the energy demand in Bangladesh. The forecast indicate expressive increase in energy consumption and is persistent with the anticipated, comparatively moderate rate of economic growth in Bangladesh in the forthcoming future. Moreover, the faster development in energy consumption also reflects the fact that there will be structural changes in the economy of Bangladesh and consecutively that some energy-intensive sectors in the economy are expected to grow. Committing with the faster growth in energy demand oil, natural gas and hydroelectricity are expected to become increasingly vital energy sources at the expense of coal, reflecting government interests towards the use of cleaner renewable energy in Bangladesh.

REFERENCES

- [1] "World population data." [Online]. Available: https://en.wikipedia.org/wiki/World_population. [Jan. 29, 2015].
- [2] "World Bank Data." [Online]. Available: <http://www.worldbank.org/>. [Jan. 30, 2015].
- [3] "Bangladesh: Economy, Asian Development Bank." [Online]. Available: <http://www.adb.org/countries/bangladesh/economy>. [Feb. 13, 2015].
- [4] "Annual Report, Bangladesh petroleum Corporation." [Online]. Available: <http://www.bpc.gov.bd/>. [Feb. 13, 2015].
- [5] "Energy Statistics, Bangladesh Energy Regulatory Commission." [Online]. Available: <http://www.berc.org.bd/>. [July 17, 2015].
- [6] Annual report of Bangladesh Bureau of Statistics (BBS), 2011-12.
- [7] "Bangladesh Power Development Board (BPDB)." [Online]. Available: <http://www.bpdb.gov.bd/>. [July 18, 2015].
- [8] Balocco, C., Grazzini, G., 1997, "A statistical method to evaluate urban energy needs", Int.J.Energy Res.21 (14), 1321-1330.
- [9] Choudhary, M.A., Khan, N., Ali, A., and Abbas, A, "Achievability of Pakistan's 2030 Electricity Generation Goals Established under Medium Term Development Framework (Mtdf): Validation Using Time Series Models and Error Decomposition Technique", [Energy 2030 Conference, ENERGY 2008. IEEE. 2008].
- [10] Matei Demetrescu. 2012. Time series and stochastic processes. University of Bonn.
- [11] "Production & Marketing Division, Petrobangla." [Online]. Available: <http://www.petrobangla.org.bd/>. [Aug. 15, 2015].

BER Performance Analysis of a FSO OCDMA Communication System with Receive Diversity using SIK Dual Photodetector Receiver

A. K. M. Nazrul Islam

Faculty member, Department of EECE
Military Institute of Science and Technology (MIST)
Dhaka-1216, Bangladesh
e-mail: nazrul_4620@yahoo.com

S.P. Majumder

Department of EEE, BUET
Dhaka-1205, Bangladesh
e-mail: spmajumder@eee.buet.ac.bd

Abstract— An analytical approach is represented to evaluate the bit error rate (BER) performance results of an optical CDMA (OCDMA) free space optical (FSO) communication system with receive diversity using multiple photodiode dual-detect Sequence Inverse Keying (SIK) receiver. The analysis is presented to find the expression for the signal current and multi access interference (MAI) current at the output of an equal gain combiner (EGC) in presence of atmospheric channel effect, atmospheric turbulence and pointing error. The performance results in terms of BER are evaluated for different values of code lengths, number of receiver antennas, number of simultaneous user and the system parameters. It is found that, significant degradation in the receiver BER performance due to the atmospheric channel effect and in presence of pointing error can be substantially reduced by increasing the number of photodetectors. Further improvement in the receiver sensitivity can be obtained by increasing the optical code length for a given number of user at a given BER of 10^{-9} . For example, at a BER of 10^{-9} and number of user $K=16$ sensitivity improvement due to receive diversity is found to be 2.15 dB, 2.35 dB, 2.6 dB corresponding to processing gain of 64, 256 and 512 respectively for number of photodetector receiver of 4.

Keywords—BER, diversity scheme, Multiple Access Interference (MAI), OCDMA, Sequence Inverse Keying (SIK), Single Input Single Output (SIMO).

I. INTRODUCTION

OCDMA with SIK receiver are very much attractive for optical networking with wavelength division multiplexing (WDM) [1, 2]. Future optical networking is going to use FSO links to avoid complexity in fiber deployment in metropolitan city area [2]. Use of OCDMA system in a FSO communication system is recently reported which has a great potentiality in deployment of multiuser multi-wavelength FSO networks. However, the performance of FSO system are very much limited due to its channel effects like fog, rain, smoke and specially the effect of atmospheric turbulence and pointing error [3-8]. Numerous research works are reported for modeling the FSO channel effects due to atmospheric turbulence and pointing error using receive and transmit diversity on FSO communication [5-10]. Mathematical models

are developed for weak, moderate and strong atmospheric turbulence and pointing error [4-7].

Several research works have been carried out on the performance of FSO communication system with multiuser diversity and multiple input multiple output (MIMO) configuration in presence of atmospheric turbulence and pointing error using different modulation formats [8-13]. Performance of single input multiple outputs (SIMO) and multiple input single outputs (MISO) FSO links through the weak and strong turbulence has been studied in [4]. A Novel closed form asymptotic BER expressions are analytically developed in [11] taking into account the cooperative FSO communication with decode-and-forward relaying with EGC. Further, investigation on BER performance results is evaluated in [12] using multi-hop FSO communication system. More recently, BER performance of FSO communication system with receive diversity in presence of pointing error is presented in [13].

Time-diversity scheme for OCDMA FSO communication system to minimize the influence of atmospheric turbulence is presented in [15] and novel transceiver architecture for atmospheric OCDMA FSO systems based on polarization modulation with direct detection is also presented in [16]. Performance of an OCDMA FSO system over FSO channels in presence of weak atmospheric turbulence using SIK correlator receiver are also studied analytically in [17]. Most of the above research works presented based on the effect of atmospheric turbulence on OCDMA FSO wireless communication system. An analytical approach to investigate the OCDMA FSO communication system in presence of pointing error using SIK dual photodetectors receivers and SIMO configuration is yet to be reported.

In this paper, following an OCDMA encoder and SIK dual photodetector receiver [14, 17] using SIMO configuration we investigated a novel analytical approach to evaluate the performance of OCDMA FSO communication system in presence of atmospheric turbulence and pointing error using EGC at the output. The BER performance results and receiver sensitivity improvements are numerically evaluated for different OCDMA encoder code lengths and number of

simultaneous user in presence of pointing error taking the system parameters into account.

This paper is organized as follows: Section II represents SIMO OCDMA FSO system model using multiple SIK dual photodetectors receivers. Analysis on BER performance results in presence of atmospheric turbulence and pointing error with SIMO SIK dual photodetectors receivers are represented in section III. The section IV presents the numerical evaluation and finally, section V represents conclusion.

II. SIMO OCDMA FSO SYSTEM MODEL

The block diagram of a SIMO OCDMA FSO communication system with single transmitter and multiple SIK dual photodetector receivers is shown in Fig. 1. In the transmitter we consider intensity modulation (IM) or on-off keying (OOK) of the transmitted laser of the incoming data bits. In the receiver, the OOK optical signal is detected by multiple numbers of dual photodetectors followed by N_r number of optical CDMA SIK receivers. In case of multiple photodetector receivers we consider equal gain combining (EGC) technique at the output of the OCDMA receivers.

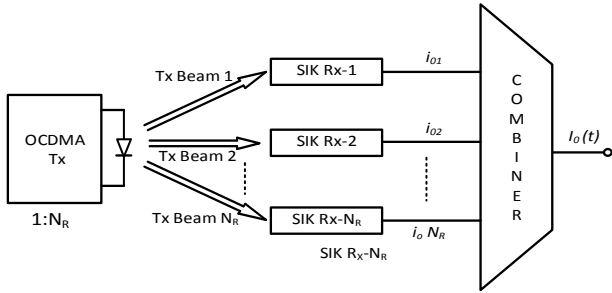


Fig. 1. Block diagram of an OCDMA FSO communication system with single transmitter and multiple SIK dual photodetector receiver (SIMO).

III. SYSTEM ANALYSIS

OCDMA signal corresponding to the m^{th} user transmitter is represented by as [2]:

$$S^m(t) = 2P_T \cdot \sum_{k=0}^{\infty} a_k^m \cdot g(t - kT_b) \cdot C^m(t) e^{j\omega_c t} \quad (1)$$

where P_T represents the average optical power transmitted by the laser diode; $C^m(t)$ is the code of the m^{th} user; a_k^m is the k^{th} bit of the m^{th} user; $g(t)$ represents the pulse shape of the information bit and T_b represents the bit period. The signature sequence $C^m(t)$ of the m^{th} user can be represented by [2] as:

$$C^m(t) = \sum_{l=0}^{L_c-1} C_l^m p(t - lT_c) \quad (2)$$

where C_l^m is the l^{th} chip of the m^{th} codeword; L_c is the code length or processing gain, Gp ; T_c represent the chip period as $T_c = T_b/L_c$ and the pulse shape of the individual chip is represents by $p(t)$.

The optical signal is received by N_r number of photodetectors followed by N_r number of optical CDMA SIK receivers. The optical signal at the input of j^{th} photodiode due to m^{th} user signal and is given by:

$$r_j^m(t) = S^m(t) \cdot e^{-\alpha d_m} \cdot h_m \quad (3)$$

where $h_m = h_{am} \cdot h_{pm}$ represent the effect of atmospheric channel corresponding to the m^{th} user communication system; h_{am} is the effect of atmospheric turbulence and h_{pm} is the attenuation due to geometric spread and pointing error; α is the atmospheric loss coefficient, d_m is the distance of the j^{th} photodiode from the m^{th} user transmitter.

The optical signal at the input to j^{th} photodiode for K number of simultaneous user can be represented by:

$$r_j(t) = \sum_{m=1}^K r_j^m(t) + n_o(t) \quad (4)$$

where $n_o(t)$ represents optical channel background noise.

The optical output current of j^{th} photodiode can be represented by:

$$i_{oj}(t) = R_d \cdot \sum_{m=1}^K r_j^m(t) + i_n(t) \quad (5)$$

The output of the j^{th} optical CDMA decoder can be represented as:

$$i_{out,j}(t) = i_{oj}(t) \times \sum_{l=0}^{L_c-1} C_l^m p(t - lT_c) + i_n(t) \quad (6)$$

Finally, the output of the combiner for m^{th} user can be represented as:

$$i_{out}(t) = \sum_{j=1}^{N_r} G_j \cdot \sum_{m=1}^K i_{out,j}(t) + i_{on}(t) \quad (7)$$

where G_j is the gain of the combiner; $i_{on}(t)$ represent the output noise current with zero mean Gaussian with variance σ_n^2 . Inserting (1) to (6) into (7) we get the optical signal received by the m^{th} user receiver in SIMO configuration as:

$$i_o(t) = 2R_d P_r \cdot \sum_{m=1}^K \sum_{j=1}^{N_r} G_j \cdot \sum_{k=0}^{\infty} a_k^m \cdot g(t - kT_b) \times \sum_{l=1}^{L_c} C_l^m p(t - lT_c) h_{am} h_{pm} + i_{on}(t) \quad (8)$$

where, $P_r = P_t \cdot e^{-\alpha d_m}$

Considering SIK correlator receiver, corresponding to j^{th} user, output current of the receiver can be expressed as [2, 14, and 17]:

$$Z_i^j(t) = R_d P_r \cdot \int_0^{T_b} \sum_{j=1}^{N_r} G_j \cdot a_k^m g(t - kT_b) \cdot \sum_{l=1}^{L_c} C_l^m p(t - lT_c) \times \left\{ C^j(t - lT_c) - \overline{C^j(t - lT_c)} \right\} h_{am} h_{pm} dt + \int_0^{T_b} i_{on}(t) dt \quad (9)$$

Let, $c_i(.) = \{C^i(.) - \overline{C^i(.)}\}$ which represents the bipolar form of $C^i(.)$ and $B^i(.) \circ A^i(.) = \left\{1 + b^i(.)a^i(.)\right\} / 2$,

where, $c^i(.)$, $b^i(.)$ and $a^i(.)$ are the bipolar form of $C^i(.)$, $B^i(.)$ and $A^i(.)$ respectively [14, 17]. The output photocurrent $Z_i^j(t)$ can now be represented by:

$$Z_i^j(t) = \frac{R_d P_r N_r}{4} \sum_{j=1}^{N_r} G_j \int_0^{T_b} \sum_{l=0}^{L_c-1} \left\{ p^m(t-lT_c) c^i(t-lT_c) \right\} h_{am} h_{pm} \\ + \frac{R_d P_r N_r}{4} \sum_{j=1}^{N_r} G_j \int_0^{T_b} \sum_{l=0}^{L_c-1} \left\{ p^i(t-lT_c) b^i(t-lT_c) |c^i(t-lT_c)|^2 \right\} h_{am} h_{pm} \\ + \frac{R_d P_r N_r}{2} \sum_{j=1}^{N_r} G_j \sum_{l=0}^{L_c-1} p^m(t-lT_c) \left\{ b^m(t-lT_c) c^m(t-lT_c) c^i(t-lT_c) \right\} h_{am} h_{pm} + \int_0^{T_b} i_n(t) dt \quad (10)$$

Total output photocurrent for K number of user is given by:

$$Z_i(t) = \sum_{m=1}^K Z_i^j(t) \quad (11)$$

The mean value of photo current $Z_i(t)$ for a given value of h_{am} and h_{pm} can be expressed by:

$$U(h_o) = \sum_{j=1}^{N_r} \frac{G_j R_d P_r}{4 T_s} \int_0^{T_b} \sum_{l=0}^{L_c-1} I_s \cdot h_o \cdot p(t-lT_c) dt \times \sum_{m=1}^K h_m \quad (12)$$

$$\text{and, } \overline{U^2(h_o)} = \int_0^\infty U^2(h_o) p(h_o) d(h_o) \quad (13)$$

$$\text{where, } h_o = \sum_{m=1}^K h_{am} \cdot h_{pm} = \sum_{m=1}^K h_m \quad (14)$$

The variance of noise current $i_{on}(t)$ is given by:

$$\sigma_n^2 = \sigma_{th}^2 + \sigma_{shot}^2 = \frac{4K_b T B}{R_L} + 2eB(I_s + I_b) \quad (15)$$

The variance of MAI for a given value of h_o can be presented by [14] as:

$$\sigma_{MAI}^2(h_o) = \sum_{j=1}^{N_r} U^2(h_o) \cdot \frac{2(K-1)}{3L_c} \quad (16)$$

The SNIR conditioned on a given value of h_o is then expressed by:

$$\xi(h_o) \equiv \frac{N_r U^2(h_o)}{\sigma_{th}^2 + \sigma_{shot}^2 + N_r \cdot \sigma_{MAI}^2(h_o)} \quad (17)$$

The conditional BER conditioned on a given value of channel coefficient h_o can be represented as:

$$P_b(h_o) = \frac{1}{2} \operatorname{erfc} \left[\sqrt{\xi(h_o)} / 2\sqrt{2} \right] \quad (18)$$

The pdf of h_o can be derived by $(K-1)$ fold convolution as:

$$p(h_o) = p(h_1) \otimes p(h_2) \otimes p(h_3) \dots \otimes p(h_m) \quad (19)$$

The average BER for a given h_o can be expressed as:

$$BER = \int_{-\infty}^{\infty} \sum_{m=1}^K P_b(h_m) p(h_m) d(h_m) \quad (20)$$

IV. RESULTS AND DISCUSSIONS

Based on the analytical approaches presented in section III, we evaluate the BER performance result as well as the improvement of receiver sensitivity for SIMO OCDMA FSO communication system for different code length, number of simultaneous user in presence of pointing error utilizing the system parameters shown in Table I.

TABLE I: SYSTEM PARAMETERS

Parameter	Symbol	Value
Data rate	R_b	10 Gbps
Receiver bandwidth	B	10 GHz
Responsivity	R_d	0.80 A/W
Receiver temperature	T	300° K
Load resistance of receiver	R_L	50 Ω
Boltzmann's constant	K_b	1.38×10^{-23} W/K/Hz
Electron charge	e	1.6×10^{-19} C
Normalized beamwidth	ω_z/r	5 and 10
Detector aperture radius	r	20 cm
Number of transmitter	M_t	1
Number of photodetector receiver	N_r	1 to 4
Normalized pointing error standard deviation	σ_p/r	1.0
Operating wavelength	λ	1550 nm
Number of user	K	1 – 256
Code length (processing gain)	$L_c(Gp)$	32 – 1024
Background current	I_b	10 nA
Power penalty at BER	-	10^{-9}

Plots of BER as a function of transmitted optical power Pt(dBm) using code length Gp as a parameter are presented in Fig. 2 in presence of pointing error. It is noticed that, the BER performance result improves with the higher processing gain Gp and the BER performance results degrades significantly with the decrease of processing gain for the same system parameters. It is also observed that, BER floor starts to occur at a processing gain of 256 and continues to increase significantly till the processing gain decreases to 32 due to the effect of MAI in presence of optical channel effects and pointing error.

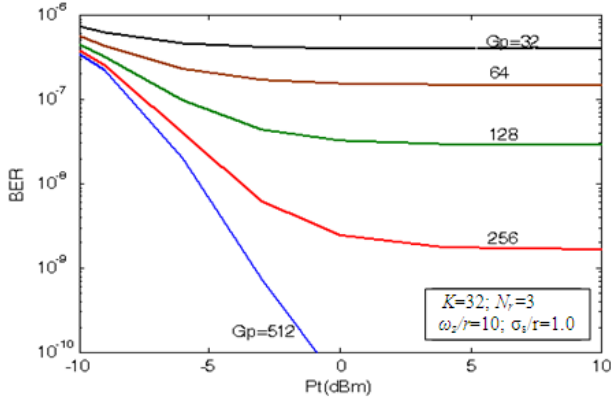


Fig. 2. BER vs. transmitted power for number of receiver antenna $N_r=3$, number of user $K=32$, normalized beamwidth $\omega_z/r=10$ and normalized jitter standard deviation $\sigma_s/r=1.0$ at a data rate $R_b=10$ Gbps using code length Gp as a parameter.

The plots of BER versus transmitted optical power P_t (dBm) are depicted in Fig. 3 for a code length $Gp=512$, number of user $K=32$ at a bit rate of 10 Gbps in the presence of pointing error using number of receiver antennas as a parameter. It is noticed that the BER performance result improves with the increase of number of receiver antennas $N_r=1$ to 4 (SISO to SIMO) with all other parameters as constant. The effect of MAI on BER flooring is also significant in presence of channel effects and pointing error.

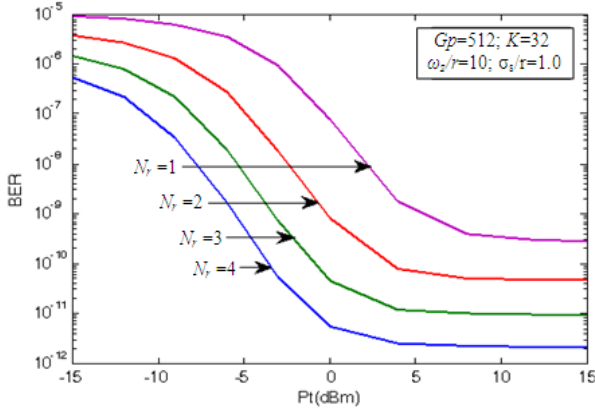


Fig. 3. BER vs. transmitted power for code length $Gp=512$, number of user $K=32$, normalized beamwidth $\omega_z/r=10$ and normalized jitter standard deviation $\sigma_s/r=1.0$ at a data rate $R_b=10$ Gbps using number of receiver antenna N_r as a parameter.

The plots of BER as a function of transmitted optical power P_t (dBm) are shown in Fig. 4 for a code length $Gp=1024$, at a bit rate of 10 Gbps using number of simultaneous users as a parameter. It is noticed that, the BER increases with the increase of number of simultaneous users with even all other parameters remain constant. It is also noticed from the Fig. 4 that, BER floor starts to occur for the numbers of simultaneous users 100 and higher for the processing gain of 1024 in the presence of channel effects and pointing error which is the significant improvement of OCDMA FSO communication system in comparison to SISO and SIMO FSO communication system.

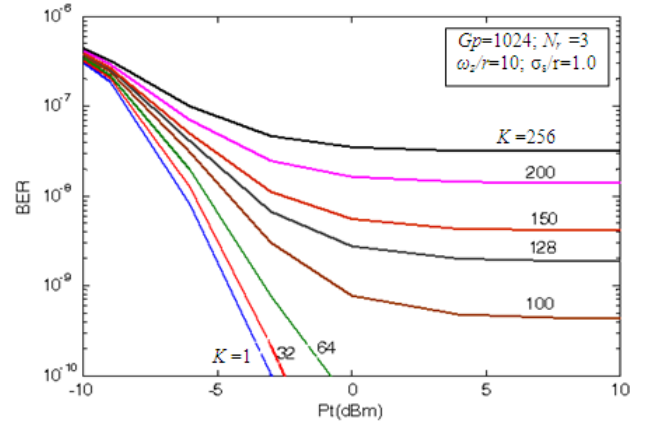


Fig. 4. BER vs. transmitted power for number of receiver antenna $N_r=3$, code length $Gp=1024$, normalized beamwidth $\omega_z/r=10$ and normalized jitter standard deviation $\sigma_s/r=1.0$ at a data rate $R_b=10$ Gbps using number of user K as a parameter.

Improvement of receiver sensitivity as a function of number of receivers are shown in Fig. 5 plotted at a BER= 10^{-9} using code length Gp as a parameter. It is observed from Fig. 5 that, the improvement of receiver sensitivity depends significantly on the code length in presence of MAI due to channel effect and pointing error. Receiver sensitivity improves significantly with the higher code length and as well as the higher number of receivers and vice versa. It is also noticed from the figure that, the receiver sensitivity increases to 2.75 dB with the increases of processing gain Gp from 64 to 512 with the increase of number of antennas from 1 to 4 for normalized beamwidth $\omega_z/r=5$, normalized pointing error $\sigma_s/r=1.0$ and number of users $K=16$ at a data rate $R_b=10$ Gbps.

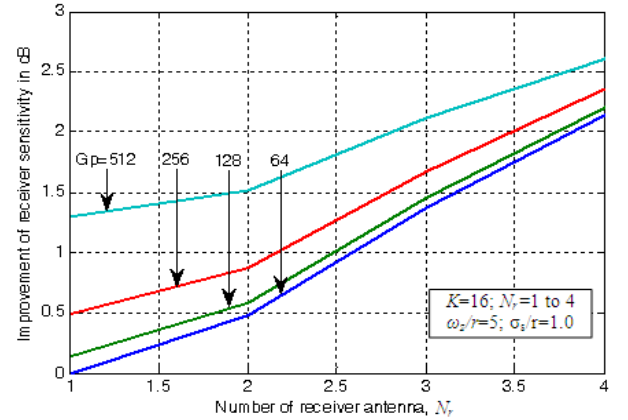


Fig. 5. Improvement of receiver sensitivity vs. number of receiver for normalized beamwidth $\omega_z/r=5$, normalized jitter standard deviation $\sigma_s/r=1.0$ and number of users $K=16$ at a data rate $R_b=10$ Gbps using code length as a parameter.

V. CONCLUSIONS

The BER performance results and improvement of receiver sensitivity are evaluated for a SIMO OCDMA FSO communication system with SIK receiver in the presence of pointing error with receive diversity by developing the pdf of the SNR at the output photodetector receiver combiner. The results show that, the BER and receiver sensitivity degradation are significant due to the increase number of simultaneous

users with lower code length in presence of channel effects and pointing error. The receiver sensitivity can be substantially improved by using longer code length using higher number of receiver antenna. The research work will find its application in designing an OCDMA FSO communication system in the presence of the combined effect of atmospheric turbulence and pointing error using transmit and receive diversity schemes.

Acknowledgment

The work is carried out as a part of Ph. D dissertation in the department of EEE of Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh.

References

- [1] P. C. Teh, M. Ibsen, J. H. Lee, P. Petropoulos, and D. J. Richardson "Demonstration of a four-channel WDM/OCDMA system using 255-chip 320 Gchip/s quaternary phase coding gratings," *IEEE Photonics Technology Letters*, vol. 14, no. 2, pp. 227-229, February 2002.
- [2] A. K. M. N. Islam and S. P. Majumder, "Effect of Pointing Error on BER Performance of an Optical CDMA FSO Link with SIK Receiver", 'in press' *Journal of Optical Communications*, 2016.
- [3] T. Tsiftsis, H. G. Sandalidis, G. K. Karagiannidis, and M. Uysal, "Optical wireless links with spatial diversity over strong atmospheric turbulence channels," *IEEE Transactions on Wireless Communications*, vol. 8, no. 2, pp. 951-957, 20 February 2009.
- [4] M. Mofidi and A. Chaman-Motlagh, "Error performance of SIMO and MISO FSO links over weak and strong turbulent channels," *8th International Symposium on Communication Systems, Networks & Digital Signal Processing (CSNDSP)*, pp. 1-4, 18-20 July 2012, Poznan, Poland.
- [5] X. Tang, Z. Xu, and Z. Ghassemlooy, "Coherent polarization modulated transmission through MIMO atmospheric optical turbulence channel," *Journal of Lightwave Technology*, vol. 31, no. 20, pp. 3221-3228, 09 September 2013.
- [6] P. Kaur, V. K. Jain, and S. Kar, "BER performance improvement of FSO links with aperture averaging and receiver diversity technique under various atmospheric conditions," *9th International Conference on Industrial and Information Systems (ICIIS)*, pp. 1-6, 15-17 December 2014, Gwalior, India.
- [7] T. H. Duyen, "Performance Analysis of MIMO/FSO Systems Using SC-QAM Signaling over Atmospheric Turbulence Channels," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 97, no. 1, pp. 49-56, January 2014.
- [8] L. Yang, X. Gao, and M. S. Alouini, "Performance analysis of free-space optical communication systems with multiuser diversity over atmospheric turbulence channels," *IEEE Photonics Journal*, vol. 6, no. 2, pp. 1-17, 12 March 2014.
- [9] D. Luong and A. T. Pham, "Average capacity of MIMO free-space optical Gamma-Gamma fading channel," *IEEE International Conference on Communications (ICC)*, pp. 3354-3358, 10-14 June 2014, Sydney, NSW.
- [10] M. A. Kashani, M. Uysal and M. Kavehrad, "On the Performance of MIMO FSO Communications over Double Generalized Gamma Fading Channels," *Cornell University Library, arXiv:1502.00365 [cs.IT]*, 02 February 2015.
- [11] A. García-Zambrana, C. Castillo-Vázquez, B. Castillo-Vázquez, and R. Boluda-Ruiz, "Bit detect and forward relaying for FSO links using equal gain combining over Gamma-Gamma atmospheric turbulence channels with pointing errors," *Optics express*, vol. 20, no. 15, pp. 16394-16409, 16 July 2012.
- [12] M. Sheng, P. Jiang, Q. Hu, Q. Su, and X.-x. Xie, "End-to-end average BER analysis for multihop free-space optical communications with pointing errors," *Journal of Optics*, vol. 15, no. 5, pp. 1-7, 17 April 2013.
- [13] A. K. M. N. Islam and S. P. Majumder, "Analytical Evaluation of Bit Error Rate Performance of a Free-Space Optical Communication System with Receive Diversity Impaired by Pointing Error," *Journal of Optical Communications*, vol. 36, no. 2, pp. 161-168, March 2015.
- [14] T. O'Farrell and S. Lochmann, "Switched correlator receiver architecture for optical CDMA networks with bipolar capacity," *Electronics Letters*, vol. 31, no. 11, pp. 905-906, 25 May 1995.
- [15] P. Liu, P. T. Dat, K. Wakamori, and M. Matsumoto, "A new scheme on time-diversity atmospheric OCDMA system over atmospheric turbulence channels," *IEEE GLOBECOM Workshops (GC Wkshps)*, pp. 1020-1025, 2010.
- [16] F. Bai, Y. Su, and T. Sato, "Performance Analysis of Polarization Modulated Direct Detection Optical CDMA Systems over Turbulent FSO Links Modeled by the Gamma-Gamma Distribution," *Photonics*, vol. 2, no. 1, pp. 139-155, 29 January 2015.
- [17] A. K. M. N. Islam and S. P. Majumder, "Effect of Weak Atmospheric Turbulence on the BER Performance of an Optical CDMA FSO Link with SIK Receiver" *IEEE International Conference on Telecommunications and Photonics (ICTP 2015)*, pp. 1-5, 26-28 December 2015, Dhaka, Bangladesh.

Performance Evolution of an 2x2 MZI Electro-Optic Switch Using Ti:KNbO₃ as Waveguide at 1.55 μm

Rabeya Khatun, A. Zahed Chowdhury

Department of Applied Physics, Electronics & Communication Engineering

University of Chittagong

Chittagong-4331, Bangladesh

shompa.apece@gmail.com, zahed_apece@cu.ac.bd

Abstract—All optical switches have become crucial components in Wavelength division multiplexing (WDM) technology to build large optical cross connect (OXC) for wide-area communication network and are well known mainly because of their high speed of operation. This paper shows the optimum performance of a 2x2 MZI electro-optic switch using a channel profile based on titanium diffused potassium niobate (Ti:KNbO₃) at 1.55 μm . Extinction ratio and insertion loss are two major performance factors to obtain the design parameters, which are evaluated by varying titanium strip thickness, waveguide width, length of interferometer arm, gap in 3 dB coupler region and gap between interferometer arm. Finally the position of center electrode regarding main axis of the device is also varied. Optimization leads to the value of 0.00925 dB insertion loss and 32.157 dB extinction ratio for both cross state (0 V) & bar state (4 V). The originality of this work is to investigate all the device parameters of a 2x2 Ti:KNbO₃ based switch at 1.55 μm .

Keywords—Electro-optic switch; potassium niobate; Insertion loss (IL); Mach-Zehnder interferometer; Titanium; Opti-BPM12.2; Extinction ratio (Ex.R).

I. INTRODUCTION

Broadening of always-on internet access network is pressurizing the global communication network industry in an implausible way. Incredibly ascending number of global subscriber is creating the problem of data congestion in wireless communication. Optical communication has become more reliable as light with its high frequency (10^{13} - 10^{14} Hz) allows larger data (large videos, complex files etc) transmission with higher speed. Wavelength division multiplexing (WDM) cross connect network (WXC) is most widely used optical systems. Optical devices required by this system are optical switch, modulator, multiplexer, optical sensor, signal router etc [1]. In optical devices for providing external control on the entrapped light for creating deflection, switching, focusing, and modulation alternative mechanism like acousto-optic, electro-optic, thermo-optic and magneto-optic are in need [2]. Acousto-optic devices however had found a lesser role, first and foremost in the area of beam reflection. Again magneto-optics devices that utilizes Faraday Effect are not broadly used as generating electric fields are lot easier than generating magnetic one [3]. Thermo-Optical devices especially switches (2-3 ms of speed) are slower than electro-optical switches (< microseconds of speed) due to their

inherently slow stabilization process and temperature flow [4-5]. So far, interest has centered on the utilization of electro-optic effect because of easy controlling of electric fields by the use of electrodes and superiorly performing electro-optic devices. Such devices show incredible speed comparative to other technologies such as bubble shift SW (about 10 ms), 2-D MEMS SW (0.5 ms), 3-D MEMS SW (< 10 ms), liquid crystal SW [4-9]. So we chose electro-optical switch. Multimode interference coupler, Y-shaped devices, directional coupler, s-bend devices etc. are useful as splitter and combiner to such switches [10-13]. For using directional coupler in optical switch based on electro-optic effect as power splitter/combiner, highly nonlinear materials for example Si, SOI, InP, LiNbO₃, KNbO₃, InGaAs, GaAs-GaAlAs etc have been used. Among these material, lowest switching voltage requirement is given by Polymer (± 0.925 V), GaAs material (1.35-1.48 V), then InP material (7.5 V) but loss are higher (> 2 dB) [14-21]. There are numerous drawbacks coupled with the formation of these heterostructures, for instance one must choose the alloy composition very carefully in order to match the lattice constant of the quaternary exactly with the lattice constant of the host substrate. Again, the apparatus & techniques required to form the alloys may include complex techniques such as molecular beam epitaxy and metal-organic chemical vapor deposition. On the other hand it is essential to grow successively alternating layers of different materials to form the multi-layer heterostructures [22].

Ti:LiNbO₃ based Mach-zehnder interferometer switches show the advantages of large capacity, low insertion loss (< 0.02 dB), available fabrication technology, availability of materials in nature, speed > 40 Gbit/s at large distance optical communication [10], large electro-optic coefficient, less complexity. These switches convey disadvantages like higher switching voltage, crosstalk, dependency in polarization [23]. On the other hand Ti:KNbO₃ overcomes these limitations incredibly along with the same advantages of LiNbO₃ [24]. Ti:KNbO₃ based switches require 50% less power than Ti:LiNbO₃ based one but cost slightly more power loss.

In our study we did not find any work at 1.55 μm on 2x2 electro-optic switch using Ti:KNbO₃ as channel profile. So here we have evaluated the values of the switch parameters at which the device performance is optimum at 1.55 μm to get very low dispersion and attenuation at the same time, whereas

1.3 μm gives minimized dispersion but creates larger attenuation [25-26]. The design parameters are titanium strip thickness, width of waveguide, length of interferometer arm, separation in 3 dB couplers, gap between interferometer arms and position of second electrode. The optimal value of the parameters are estimated by the use of FDBPM [27-29] in OptiBPM software. Then a comparative study between Ti:LiNbO₃ based switch and Ti:KNbO₃ is made in tabular form to ease the choice of electro-optic material for implementing large optical network. In section II there is basic mathematical idea of 2x2 Mach-Zehnder interferometer switch, design and simulation results are in section III & section IV represents conclusion on the paper.

II. BASIC MATHEMATICAL MODEL OF 2X2 ELECTRO-OPTIC MZI SWITCH

Ti:KNbO₃ waveguide is made by diffusing the titanium dopant into the potassium niobate host by heating which increases the extraordinary and ordinary refractive index. The Titanium ions collide with host substrate thereby forming graded index profile. We can characterize the graded index waveguide profile using the equations below [24],

$$n_i(\lambda, x, y) = n_i^{(0)}(\lambda) + \Delta n_i(\lambda, x, y) \quad i = o, e \quad (1)$$

Here $n_i^{(0)}$ represents bulk crystal index and Δn_i is change in diffusion induced index. From the classical diffusion theory we can derived the concentration profile [24].

$$c(x, y) = c_o \left\{ \left[\text{erf} \left[\frac{w}{2D_H} \left(1 + \frac{2x}{w} \right) \right] + \text{erf} \left[\frac{w}{2D_H} \left(1 - \frac{2x}{w} \right) \right] \right] \right\} \exp \left(-\frac{y^2}{D_V^2} \right) \quad (2)$$

In equation 2, c_o indicates profile constant, w is strip width of dopant before diffusion. D_H & D_V are the tangential and vertical diffusion length respectively which are given by the following equations [24].

$$D_H = 2 \sqrt{t D_{OH} \exp \left(-\frac{T_o}{T} \right)} \quad (3)$$

$$D_V = 2 \sqrt{t D_{OV} \exp \left(-\frac{T_o}{T} \right)} \quad (4)$$

D_H & D_V are dependent on diffusion temperature T , temperature coefficient T_o , diffusion time t and diffusion constant D_{OH} & D_{OV} . Concentration profile constant C_o depends on strip thickness before diffusion τ , vertical diffusion length and dopant constant C_m , as given in equation 5 and 6.

$$C_o = \frac{\tau C_m}{\sqrt{\pi D_V}} \quad (5)$$

$$C_m = \left(\frac{\rho}{M_{at}} \right) N_A \quad (6)$$

C_m depends on Avogadro's number N_A , dopant density ρ and atomic weight M_{at} . For titanium dopant, $\rho = 4.51 \text{ g cm}^{-3}$, $M_{at} = 47.9 \text{ g mol}^{-1}$, as a result $C_m = 5.67 \times 10^{22} \text{ cm}^{-3}$ [30]. Equations 1 to 6 give elaborate descriptions of titanium during the process of diffusion to establish the waveguide structure in

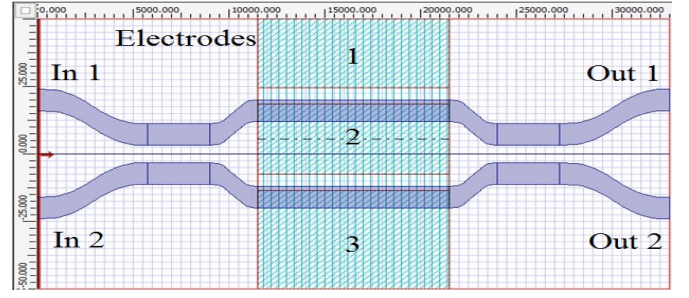


Fig. 1. 2x2 MZI Electro-Optic (MZI-EO) switch.

building 2x2 MZI-EO switch. Fig. 1 shows a 2x2 MZI-EO switch which consists of two directional coupler (3 dB) functioning as splitter/combiner respectively. The relative change in phase due variation in electrode voltage is defined by the following equation [31].

$$\Delta \phi = \frac{2\pi}{\lambda} \left[\left(\frac{n^3}{2} \right) \frac{V}{d} r \right] L. \quad (7)$$

L presents required length of interferometer arm & V is voltage difference (Fig. 1), r is electro-optic coefficient and d is the gap between the electrodes (Fig. 1). Now output powers at port 1 and port 2 are [31],

$$P_{out1} = \sin^2 \left(\frac{\Delta \phi}{2} \right). \quad (8)$$

$$P_{out2} = \cos^2 \left(\frac{\Delta \phi}{2} \right). \quad (9)$$

Insertion loss and extinction ratio of the device can be found by using the results from equation 10 and 11. Insertion loss can be measured by the lost power. Low insertion loss gives measure of good performance. Extinction ratio is output power in on state divided by the output power in off state which should be large enough [26].

$$I.L.(dB) = 10 \log_{10} \left(\frac{P_{out}}{P_{in}} \right). \quad (10)$$

$$Ex.R.(dB) = 10 \log_{10} \left(\frac{P_{on}}{P_{off}} \right). \quad (11)$$

The excess loss in switch is a quantity of the input power, which does not appear at the output ports defined by equation 12. Again, Crosstalk is defined as the difference between the power at the switched port and the power at the other port in dB as shown in equation 13 [32].

$$EL(dB) = 10 \log_{10} \left(\frac{P_{in}}{P_{Total}} \right). \quad (12)$$

$$CT(dB) = 10 \log_{10} \left(\frac{P_{switched}}{P_{other}} \right). \quad (13)$$

III. DESIGN PROCEDURE AND RESULTS OF SIMULATION

For designing MZI switch, we used titanium diffused potassium niobate based waveguide which is prepared on Z-cut wafer having air as surrounding. Table I shows the

waveguide properties. And strip thickness of titanium is varied from a range of 0.0725-0.09 μm as shown in table II. Table II also shows the values of D_V and D_H .

TABLE I. WAVEGUIDE PROPERTIES

Dielectric Material		Diffused Material		
Refractive Index	Material	Propagation Direction	Crystal Cut	Crystal Name
1	Air	Y	Z	KNbO ₃
Diffused Profile				
Ti:KNbO ₃				

TABLE II. TITANIUM STRIP SPECIFICATIONS

Strip Thickness Before Diffusion	Dopant constant	Diffusion length	
		D_H	D_V
0.0725-0.09 μm	5.67e+022 per cm ³	4.0 μm	3.5 μm

TABLE III. POTASSIUM NIOBATE PARAMETER

Crystal Name	Electro-Optic Coefficients 10 ⁻¹² (m/V) :			
	r_{33}	r_{51}	r_{13}	r_{22}
KNbO ₃	64	270	28	1.3
Refractive index				
Ordinary n_o		Extraordinary n_e		
2.25		2.169		

TABLE IV. WAFER PROPERTIES

Wafer Dimensions (μm)		Width (μm)	Wafer Profile
<i>breadth</i>	<i>Length</i>	8.0	Ti:KNbO ₃
100	33000		
2D Wafer Properties			
Wafer Refractive index: Potassium niobate			
3D Wafer Properties			
<i>Substrate</i>		<i>Cladding</i>	
Thickness (μm)	Material	Thickness (μm)	Material
10	KNbO ₃	2.0	air

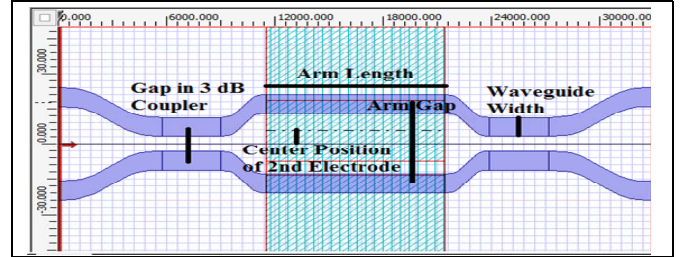
We used potassium niobate substrate whose electro-optic coefficient is very large shown in table III gives specification of potassium niobate as substrate. From table IV we can find the wafer properties including the dimensions and design specifications. The length of the wafer is 33mm where it's width is not more than 100 μm . The width of the waveguide is 8.0 μm along X-axis. Table IV also shows 3D wafer properties which include air as cladding with a thickness of 2.0 μm and substrate as potassium niobate with a thickness of 10 μm . Three electrodes with a thickness of 4.0 μm are used to create electro-optic effect and these electrodes are placed on the top of a buffer layer. The thickness of this buffer layer is 0.3 μm with a refractive index of 1.47. Horizontal permittivity and vertical permittivity of the buffer layer are set to 4. Width and voltage of the first and third electrodes are 50 μm and 0 V respectively. The second one is set to a width of 26 μm and voltage of 0 V. The separations between 1-2 and 2-3 are 6 μm . We can find the list of simulation parameters in table V. The ranges for 6 variable parameters are shown in table VI and also depicted in a figure which is included in the same table. I.L. and Ex.R. for each of the 6 variables are calculated to find the optimum parameters of the switch for which switch performance is optimal at 1.55 μm .

TABLE V. SIMULATION PARAMETERS

Parameters	Definitions
Wavelength	1.55 μm
Polarization	Transverse magnetic (TM)
Reference index	Modal
Mesh-number of points	500
BPM solver	Paraxial
Boundary condition	Transparent Boundary Condition
Engine	Finite Difference
Scheme parameter	0.5
Propagation step	1.55

TABLE VI. VARIABLE PARAMETERS

Parameters	Range
Dopant (Ti) stripe thickness (t_s)	0.0725-0.09 μm
Waveguide width (W)	7-10 μm
Interferometer arms length (L_{INF})	9850-10100 μm
Interferometer arms gap (G_{INF})	31-35 μm
Gap in 3-dB coupler (g)	13.5-15.5 μm
Center position of second electrode	5.4-5.6 μm



A. Variation in strip thickness of Titanium

To estimate the optimum value of titanium strip thickness in the host, we have done 2D isotropic simulation over a range of 0.0725 μm to 0.09 μm [33], while keeping other parameters constant. To compare the performances over this range we have collected the values of I.L. and Ex.R. The results are shown in table VII. The results are compared with the results of [24]. Table VII states that the performance is better at strip thickness of 0.0825 μm . Best figure of extinction ratio and insertion loss calculated are 31.43 dB & 0.0097 dB respectively with titanium trip thickness of 0.0825 μm in both crossbar and bar state at 1.55 μm operating wavelength, where in [24] best extinction ratios are 29.4 dB, 26.6 dB and minimum insertion losses are 0.032 dB, 0.0261 dB at crossbar & bar state respectively found with thickness of 0.0525 μm at 1.3 μm operating wavelength. These are summarized in table VIII.

So, we can say that performance of the switch is better at 1.55 μm . The device shows an optimum performance for 0.0825 μm strip thickness of Titanium. Power in both output waveguides for 0.0825 μm strip thickness is shown in Fig. 2. 6 iterations were needed to complete the operation of switching. Iteration and switching voltage are directly related. At iteration 1 voltage is 0 V and then for each step there is an increase of 0.8 V, which in turn results in 4 V at iteration 6. With second electrode of the switch at 0 V, we find out 1 and out 2 as 0.000717271mW & 0.997764mW. After 6 iterations when applied voltage is 4 V out 1 is 0.997764mW and out 2 is 0.000717271mW. This switch only needs 4 V for switching operation where in case of Ti:LiNbO₃ switch it was 8 [34].

TABLE VII. EFFECTS OF VARYING TITANIUM STRIP THICKNESS

Titanium thickness (μm)	Driving voltage (V)	In/Out Power (mW)		I.L. (dB)	Ex.R. (dB)
0.0725	0	In 1	1	2.291	Out 2 16.9
		Out 1	0.403513		
		Out 2	0.590072		
	4	In 1	1	0.0809	Out 1 3.816
		Out 1	0.981541		
		Out 2	0.012045		
0.0775	0	In 1	1	0.43	Out 2 26.20
		Out 1	0.0896064		
		Out 2	0.904624		
	4	In 1	1	0.0346	Out 1 9.814
		Out 1	0.992059		
		Out 2	0.00217151		
0.08	0	In 1	1	0.0846	Out 2 29.29
		Out 1	0.020058		
		Out 2	0.980702		
	4	In 1	1	0.0172	Out 1 16.96
		Out 1	0.99645		
		Out 2	0.001157889		
0.0825	0	In 1	1	0.0097	Out 2 31.43
		Out 1	0.000717271		
		Out 2	0.997764		
	4	In 1	1	0.0097	Out 1 31.43
		Out 1	0.997764		
		Out 2	0.000717271		
0.0875	0	In 1	1	0.3404	Out 2 20.35
		Out 1	0.0706372		
		Out 2	0.924618		
	4	In 1	1	0.0444	Out 1 20.65
		Out 1	0.989831		
		Out 2	0.00852946		
0.09	0	In 1	1	0.6667	Out 2 17.65
		Out 1	0.137359		
		Out 2	0.85768		
	4	In 1	1	0.0594	Out 1 8.562
		Out 1	0.986419		
		Out 2	0.0147507		

TABLE VIII. EVALUATION TABLE 1

Performances	Results of simulation at 1.55 μm		Results at 1.3 μm [24]	
	0 V	4 V	0 V	4 V
I.L. (dB)	0.0097	0.0097	0.032	0.0261
Ex.R. (dB)	31.43	31.43	29.4	26.6
Performance evolution				
I.L. reduced to 69.69% at 0V and 62.84% at 4 V	Ex.R increased 6.9% at 0 V and 18.16% at 4 V		Power requirement is same	

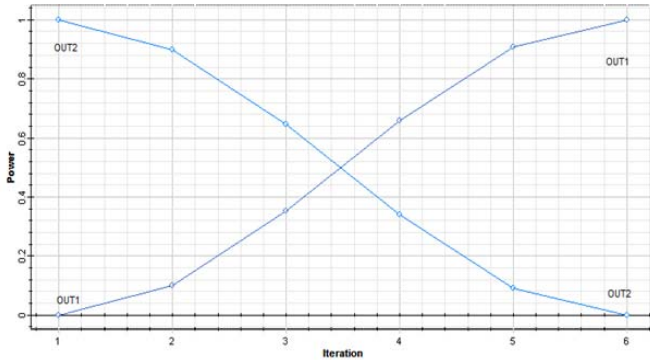


Fig. 2. Power variation in output waveguides.

B. Variation in arm length of Interferometer

After varying titanium strip thickness, we also varied arm length of the switch over a range between 9850 μm to 10100 μm with the strip thickness 0.0825 μm . The results of variation in arm length are listed in table IX. Better results are obtained with a length of 9900 μm . From Fig. 3 we can observe the differences in output power from 0 V to 4 V at out 1 and out 2. The outputs at 0 V in port 1 and port 2 are 0.000607266mW and 0.997872mW respectively and outputs at 4 V in port 1 and port 2 are 0.997872 mW and 0.000607266mW respectively. From these results the calculated insertion loss is 0.00925 dB and extinction ratio is 32.157 dB, which is slightly better than section A, so it can be said that performance of the switch is optimized at 9900 μm arm length.

TABLE IX. EFFECTS OF VARYING LENGTH OF INTERFEROMETER ARM

Length of Arm (μm)	Driving voltage (V)	In/Out Power (mW)		I.L. (dB)	Ex.R. (dB)
9850	0	In 1	1	0.00953	Out 2 31.599
		Out 1	0.000690513		
		Out 2	0.997808		
	4	In 1	1	0.00953	Out 1 31.599
		Out 1	0.997808		
		Out 2	0.000690513		
9900	0	In 1	1	0.00925	Out 2 32.157
		Out 1	0.000607266		
		Out 2	0.997872		
	4	In 1	1	0.00925	Out 1 32.157
		Out 1	0.997872		
		Out 2	0.000607266		
9950	0	In 1	1	0.00938	Out 2 31.859
		Out 1	0.000650388		
		Out 2	0.997842		
	4	In 1	1	0.00938	Out 1 31.859
		Out 1	0.997842		
		Out 2	0.000650388		
10000	0	In 1	1	0.00972	Out 2 31.43
		Out 1	0.000717271		
		Out 2	0.997764		
	4	In 1	1	0.00972	Out 1 31.43
		Out 1	0.997764		
		Out 2	0.000717271		
10050	0	In 1	1	0.00980	Out 2 31.434
		Out 1	0.000717122		
		Out 2	0.997746		
	4	In 1	1	0.00980	Out 1 31.434
		Out 1	0.997746		
		Out 2	0.000717122		
10100	0	In 1	1	0.00986	Out 2 31.436
		Out 1	0.000716873		
		Out 2	0.997733		
	4	In 1	1	0.00986	Out 1 31.436
		Out 1	0.997733		
		Out 2	0.000716873		

C. Miscellaneous

We have observed the results of variation in arm gap, width of waveguide, separation in coupling region of both the 3 dB couplers and the center position of control electrode (2nd) too. But overall performance of the switch is as same as section B for 32 μm arm separation, 8 μm waveguide width, 14.5 μm gap in directional coupler and 5.5 μm position of

second electrode. The ultimate values of I.L. and Ex.R. are 0.00925 dB and 32.157 dB respectively for cross state and bar state both, which shows better performance than that of ref [24], because I.L. decrease further as 71.09% in 0 V and 64.56% at 4 V. Ex.R. increases 9.38% at 0 V and 20.89% at 4 V. Optical field propagation of the optimized switch for both 0 V & 4 V are shown in Fig. 4 and Fig. 5 respectively. Fig. 6 and Fig. 7 show distribution electric field of this switch regarding effective index and switch width at 0 V and 4 V respectively. Here the red curve represents ideal switching behavior where the blue line shows the same for our proposed switch. Power overlap integral (POI) characteristics along the effective distance of the switch for out 1 are shown in Fig. 8 & Fig. 9. POI is known as integrated result between simulated DFT that results in a selected x-y plane and the input wave in the input plane. At the end of this curve it shows any undesirable value of power at out 1. At iteration 1, where applied electric voltage is 0 V, POI ends almost at zero and similarly at iteration 6 it shows some loss experienced by the waveguide at the end of out 1 which in turn prevents it to be unity.

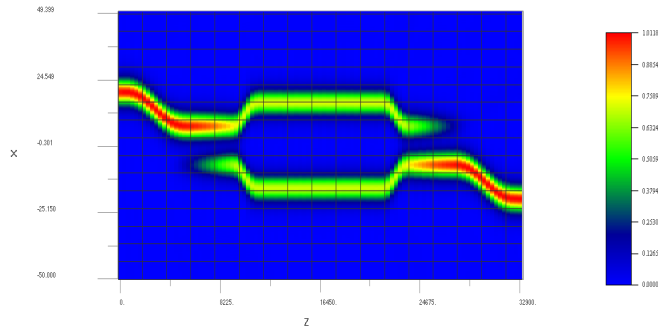


Fig. 3. Optical field propagation at 0 V.

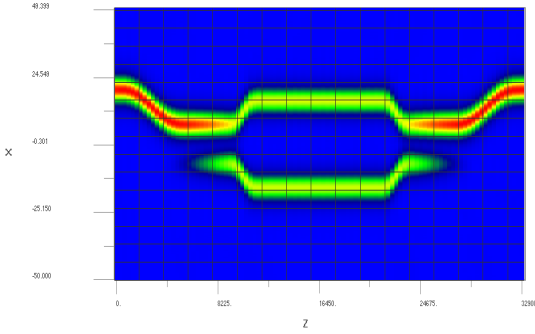


Fig. 4. Optical field propagation at 4 V.

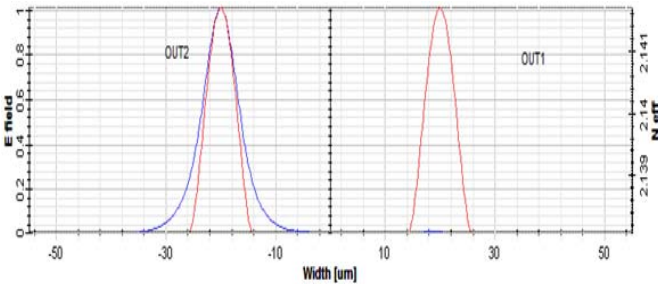


Fig. 5. Distribution of electric field at 0 V.

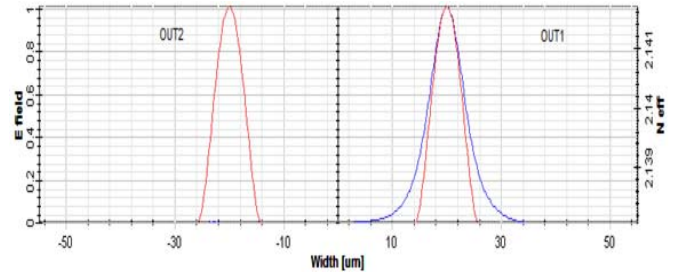


Fig. 6. Distribution of electric field at 4 V.

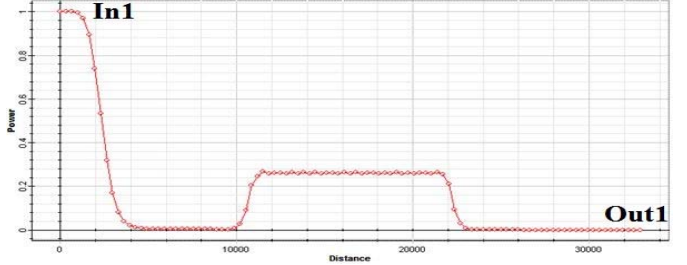


Fig. 7. Power overlap integral at 0 V.

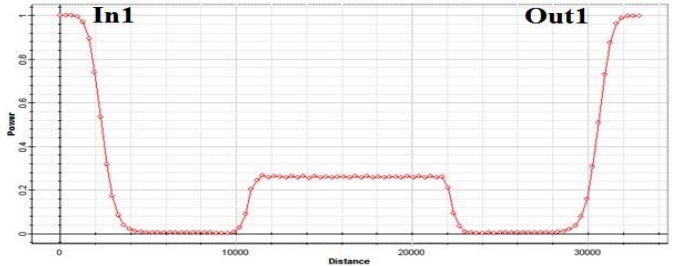


Fig. 8. Power overlap integral at 4 V.

The proposed structure shows enhanced performance than the previous one in ref [34] in the sense of power requirement as power requirement is reduced to 50% but in case of I.L. and Ex.R., Ti:LiNbO₃ is slightly better. As refractive index of KNbO₃ is larger than LiNbO₃, the phase change between the interferometer arms is faster in Ti:KNbO₃ based switch resulting in low switching voltage [24]. Switching performance of these two types of material is mainly dependent of the value of refractive index and electro-optic co-efficient. Considering a tradeoff between power requirement and loss factor performance of switch based on Ti:KNbO₃ is better than that of Ti:LiNbO₃. But if we need to implement a long interconnecting network we should emphasize to loss factor as with increasing distance power reduction at output port can be large enough to produce error in transmission. These are summarized in table X.

TABLE X. EVALUATION TABLE 2

Performances	Ti:LiNbO ₃ based switch [34]		Ti:KNbO ₃ based switch	
	0 V	8 V	0 V	4 V
I.L. (dB)	0.00881	0.00881	0.00925	0.00925
Ex.R. (dB)	33.581	33.581	32.157	32.157
EL (dB)	0.0069	0.0069	0.0066	0.0066
CT (dB)	33.58	33.58	32.16	32.16
Performance Evolution	4.994% less I.L. and 4.24% higher Ex.R.		50% less swicthing voltage & 4.55% less E.L.	

IV. CONCLUSION

The optimum parameters of 2x2 MZI switch based on Ti:KNbO₃ for instance titanium strip thickness, width of waveguide, arm length, arm separation, coupling region separation & position of control electrode for evolution of greater performance at 1.55 μm are presented in this paper. From results of simulation observed using OptiBPM12.2 software, it is found that the switch shows optimum performance with 0.0825 μm strip thickness titanium, arm length of 9900 μm , 8 μm waveguide width, 32 μm arm gap, 14.5 μm gap in 3dB coupler and position of control electrode at 5.5 μm just above original axis of the switch. We found minimum insertion loss of 0.00925 dB & maximum extinction ratio of 32.157 dB for both cross & bar state.

ACKNOWLEDGMENT

We gratefully acknowledge the support for this work provided by the Information and Communication Technology Division, Government of the People's Republic of Bangladesh through the ICT Fellowship 2015-2016.

REFERENCES

- [1] O'Mahony, J. Michael, "Future optical networks," IEEE J. Light. Tech., vol. 24, no. 12, pp. 4684-4696, 2006.
- [2] J. Wilson, J.F.B. Hawkes, "Optoelectronics: An Introduction," Prentice Hall, 2nd edition, chapter 3, 1989.
- [3] T. Tamir (Ed.), "Integrated Optics," Springer-Verlag, 2nd edition, 1979.
- [4] Javier Aracil, Franco Callegati, "Enabling Optical Internet with Advanced Network Technologies," Springer Science & Business Media, ISBN: 1848822782, 9781848822788, pp. 176, 2009.
- [5] Achyut K. Dutta, Niloy K. Dutta, Masahiko Fujiwara, "WDM Technologies: Optical Networks," Academic Press, pp. 104, 2004.
- [6] D. Chen, S. Close et al, "An optical cross-connect switch based on micro-bubbles," ASME 2000, MEMMS, vol. 2, pp. 35-37, 2000.
- [7] K. Noguchi, "Optical multichannel switch composed of liquid-crystal light modulator arrays and bi-refractive crystals" IEEE Electron. Lett., vol. 33, no. 19, pp. 1627-1628, 1997.
- [8] L. Y. Lin, E. L. Goldstein, R. W. Tkach, "Free-space micromachined optical switches for optical networking," IEEE J. Sel. Top. Quantum Electron., vol. 5, no. 1, pp. 4-9, 1999.
- [9] D.T. Nelson, V.A. Aksyuk, et al, "Fully provisioned 112x112 micro-mechanical optical cross connect with 35.8 Tb/s demonstrated capacity," OFC'00, PD12-1, 2000.
- [10] G. Singh, "Ti indiffused Lithium Niobate (Ti: LiNbO₃) Mach-Zehnder interferometer all optical switches: A review," New Advanced Technologies, ISBN 978-953-307-067-4, pp. 312-322, March 2010.
- [11] Raqibul Hossen, Kazi Tanvir Ahmmed, A. Zahed Chowdhury, Md. Shakhawat Hossain, R.M. Murad Hasan, "Optimal Design and Performance Study of a Low-Index-Contrast MMI Optical Coupler," Proceedings of The 9th International Forum on Strategic Technology (IFOST), pp. 230-233, 2014.
- [12] A. Zahed Chowdhury, "Performance Study of Silica-on-Silicon Based Multimode Interference (MMI) Optical Coupler," Photonic Sensors, vol. 4, no. 1, pp. 34-42, 2014.
- [13] George T. Palocz, Yanyi Huang, Amnon Yariv, Jingdong Luo, Alex K.-Y. Jen, "Replica-molded electro-optic polymer Mach-Zehnder modulator," Appl. Phys. Lett., vol. 85, issue 10, pp. 1662, 2004.
- [14] C. T. Zheng, C.S. MA, X. Yan, D. M. Zhang, "Design of a spectrum-expanded polymer MZI electro-optic switch using two-phase generating couplers," Applied Physics B: Lasers and Optics, vol. 102, no. 4, pp. 831-840, 2011.
- [15] Wol-Yon Hwang, Min-Cheol Oh, Hyang-Mok Lee, Heuk Park, Jang-Joo Kim, "Polymeric 2x2 electrooptic switch consisting of asymmetric Y junctions and Mach-Zehnder interferometer," IEEE Photonics Technology Letters, vol. 9, no. 6, pp. 761-763, 1997.
- [16] H. P. Li, J. K. Liao, X. G. Tang, R. G. Lu, Y. Z. Liu, "2x2 polymeric electro-optic MZI switch using multimode interference couplers," Proceedings of SPIE 7509, article 75090X, 2009.
- [17] Xu Xue-Jun, Chen Shao-Wu, Xu Hai-Hua, Sun Yang, Yu Yu-De, Yu Jin-Zhong, Wang Qi-Ming, "High-speed 2x2 silicon-based electro-optic switch with nanosecond switch time," Chinese Physics B, vol. 18, no. 9, pp. 3900-3904, 2009.
- [18] E. H. M. Yahya, "Mach-Zehnder Interferometer," M. Tech. Thesis, Faculty of Electrical Engineering, University of Technology Malaysia, April 2007.
- [19] M. S. A. Rahman, K. M. Shaktur, R. Mohammad, "Analytical and simulation of new electro-optic 3x3 switch using a Ti:LiNbO₃ waveguide medium," International Conference on Photonics (ICP), pp. 1-5, 2010.
- [20] Maat D.H.P., "InP-Based Integrated MZI Switches for Optical Communication," PhD Thesis, Department of Applied Physics, Delft University of Technology, Netherlands, April 2001.
- [21] Shaochun Cao, Liping Sun, Savoie M., "2x2 MMI-MZI GaAs-GaAlAs carrier-injection optical switch," 2010 IEEE Photonics Society Summer Topical Meeting Series, pp. 207-208, 2010.
- [22] Mikami et al, "Waveguided Optical Switch in InGaAs/InP Using Free-Carrier Plasma Dispersion," Electronic Letters, vol. 20, no. 6, pp. 228-229, March 1984.
- [23] G. Singh, "Design of 2x2 Optoelectronic Switch Based on MZI and Study the Effect of Electrode Switching Voltages," Journal of World Academy of Science, Engineering and Technology, vol. 39, pp. 401-407, 2008.
- [24] Nazmi A. Mohammed, Huda S. Abo Elnasr, Moustafa H. Aly, "Analysis and design of an electro-optic 2 x 2 switch using Ti: KNbO₃ as a waveguide based on MZI at 1.3 μm ," Opt Quant Electron (springer journal), vol. 46, pp. 295-304, 2014.
- [25] Senior John M., "Optical Fiber Communications: Principles and Practice," Pearson Education Limited, 3rd edition, ISBN: 0-13-635426-2, pp. 4, September 2009.
- [26] Nazmi A. Mohammed, Huda S. Abo Elnasr, Moustafa H. Aly, "Performance Evaluation and Enhancement of 2x2 Ti: LiNbO₃ MachZehnder Interferometer Switch at 1.3 μm and 1.55 μm ," The Open Electrical & Electronic Engineering Journal, vol. 6, pp. 36-49, 2012.
- [27] Y. Chung, N. Dagli, "An assessment of finite difference beam propagation method," J. Quantum Electron., vol. 26, no. 8, pp. 1335-1339, 1990.
- [28] Y. Chung, N. Dagli, "Analysis of Z-invariant and Z-variant semiconductor rib waveguides by explicit finite difference beam propagation method with nonuniform mesh configuration," J. Quantum Electron., vol. 27, no. 10, pp. 2296-2305, 1991.
- [29] G. R. Hadley, "Low-truncation-error finite difference equations for photonics simulation. I. Beam propagation," J. Lightwave Technol., vol. 16, no. 1, pp. 134-141, 1998.
- [30] A.H. Tehranchi, "Optical beam propagation in Ti:LiNbO₃ Mach-Zehndermodulator," Proceeding of Laser and Fiber Optical Networks Modeling, Kharkiv, pp. 613-625, 2002.
- [31] S. K. Raghuwanshi, A. Kumar, "1x4 Signal Router Using 3 Mach-Zehnder Interferometers," Opt. Eng., vol. 52, no. 3, pp. 035002, 2013.
- [32] A. K. Ghatak, B. Culshaw, V. Nagarajan and B. D. Khurana, "Trends in fiber optics and Optical Communications", Viva Books Pvt. Ltd. (New Delhi), Ch. 4, pp. 45, 1995.
- [33] G. Singh, "Modeling of a 2x2 electro-optic Mach-Zehnder Interferometer optical switch with s-bend arms," Photonics Letters of Poland, vol. 3, no. 3, pp. 119-121, 2011.
- [34] Rabeya Khatun, Kazi Tanvir Ahmmed, A. Zahed Chowdhury, Raqibul Hossen, "Optimization of 2x2 MZI Electro-Optic Switch and Its Applications as Logic Gate," 18 th International Conference on Computer & Information Technology, pp. 294-299, 2015.

Modelling and simulation of power system of battery, solar and fuel cell powered Hybrid Electric Vehicle

Muhammad Sifatul Alam Chowdhury, Khandakar
Abdulla Al Mamun
Department of Electrical and Electronic Engineering
International Islamic University Chittagong
Chittagong, Bangladesh
m.sifatul.alam@gmail.com, k.a.a.mamun@gmail.com

Al Mahmudur Rahman
Institute of Engineering Mechanics
Karlsruhe Institute of Technology
Karlsruhe, Germany
al.rahman@partner.kit.edu

Abstract— To meet the never ending fuel demand of the vehicle, Hybrid Vehicle was introduced in the research. Generally a hybrid vehicle is the combination of an internal combustion engine (ICE) and electrical drive system. The performance of the vehicle is largely dependent on the accuracy and efficiency of the electrical system of the vehicle powertrain. This paper analyzed an electrical system for the powertrain of Hybrid Electric vehicle which is powered by Fuel cell, Battery and PV panel. A Simulink model is prepared and simulated successfully. Permanent magnet synchronous motor drive system is used in the Simulink model. The simulation results like rotor speed, electromagnetic torque, current, DC-DC converter current, voltage, state of charging, grid charger performance, photovoltaic panel performance and mechanical torque are discussed and compared with the state of the art methods. Including this the research also compared the performance of energy storage system with the reference which is frequently used in the electrical system of the HEV.

Keywords— Hybrid electric vehicle; Power system of vehicle; Fuel cell vehicle; Solar vehicle.

I. INTRODUCTION

According to the center of solar energy and hydrogen research a total number of 550,000 new hybrid electric vehicle registered till 2015 [1]. Based on the use of energy storage system electric vehicle can be classified into battery electric vehicle, hybrid electric vehicle and plug-in hybrid electric vehicle [2]. The key factor behind the revolution of Hybrid Electric Vehicle is the improved performance of power electronics and performance oriented motor drives. The performance of the powertrain of HEV is solely dependent on the performance of motor drives. The demand of hybrid electric vehicle are increasing day by day due to immense improvement in the performance of various power electronics devices like converter, switches and motor drives [3]. Besides, the development in the hardware of the hybrid electric vehicle, a silent revolution is also happened in the field of software of HEV. A hybrid electric vehicle is generally consists of an ICE and motor for propulsion. Vehicle manufacturer are emphasizing on hybrid electric vehicle because safety of environment is a global manifesto. To keep the environment safe hydrogen was introduced as a fuel source of the vehicle.

Fuel cell technology specifically proton exchange membrane fuel cell (PEMC) used vastly to utilize hydrogen as major fuel source of the vehicle. As an environmentally safe and efficient fuel hydrogen is the vital choice for the manufacturer in case electric vehicle.

The main advantage of hybrid electric vehicle beside the improved fuel economy, is the reduction of carbon-di-oxide emission [4]-[5]. The carbon dioxide emission rate on May 2016 was near 407.70 ppm and it is simultaneously increasing [6]. Hybrid electric vehicle uses many state of the art environment friendly technologies like regenerative braking system, fuel cell etc. which will greatly reduce the emission of harmful elements whose are responsible for environment pollution. Along with this green technologies use of efficient power electronics and lighting system will greatly improve the overall performance of hybrid electric vehicles [7].

This paper introduces a Simulink model for the complete power system of powertrain of the hybrid electric vehicle powered by fuel cell and photovoltaic panel. Grid charger also added in this model to make the model complete. In this paper the detail analysis of what will happen after adding the photovoltaic panel to charge the auxiliary battery of the vehicle is discussed. Beside this the comparative discussion on the performance of various power electronic devices and motor drives are analyzed.

II. MATHEMATICAL MODELLING

The mathematical modelling of permanent magnet synchronous machine are briefly discussed in [8]-[10]. During the modelling of Permanent magnet synchronous machine the modelling is done in rotor d-q frame. Park transformation is used to transform the frame abc to frame d-q. From park transformation we can write for stator current,

$$\frac{d}{dt}(i_d) = \frac{1}{L_d}(V_d - R_s i_d + L_q i_q \omega_r) \quad (1)$$

$$\frac{d}{dt}(i_q) = \frac{1}{L_q}(V_q - R_s i_q - L_d i_d \omega_r - \lambda_{pm} \omega_r) \quad (2)$$

Here,

V_q , is the quadrature axis voltage

V_d , is the direct axis voltage

i_d , is the direct axis current

i_q , is the quadrature axis current

L_q , is the quadrature axis inductance

L_d , is the direct axis inductance

R_s , is the stator resistance

λ_{pm} , is the permanent magnet flux linkage

ω_r , is the rotor speed

For d-q frame electromagnetic torque will be,

$$T_e = 0.75P * (\lambda_{pm} i_q + L_d i_d i_q - L_q i_d i_q) \quad (3)$$

$$T_e = 0.75P * (\lambda_{pm} i_q + (L_d - L_q) i_d i_q) \quad (4)$$

Here P is the number of pole. Relation between electromagnetic torque and the load torque can be defined as below,

$$\frac{d}{dt} (\omega_r) = \frac{1}{J} (T_e - T_L + B \omega_r) \quad (5)$$

III. SIMULATION STUDIES AND RESULT

A. Permanent Magnet Synchronous Machine

A salient pole permanent magnet synchronous machine is used in the Simulink model where back emf waveform will be Sinusoidal. According to Figure 2 when the simulation is done electromagnetic torque of the permanent magnet synchronous motor is varied from -100 Nm to 300 Nm at $t=16$ sec. Rotor speed is simultaneously increasing and increase till the speed goes to 7000 rpm and tends to follow the expected. The mechanical power is varied according to the reference within the range 0 W to 15×10^4 W. There are some distortion found

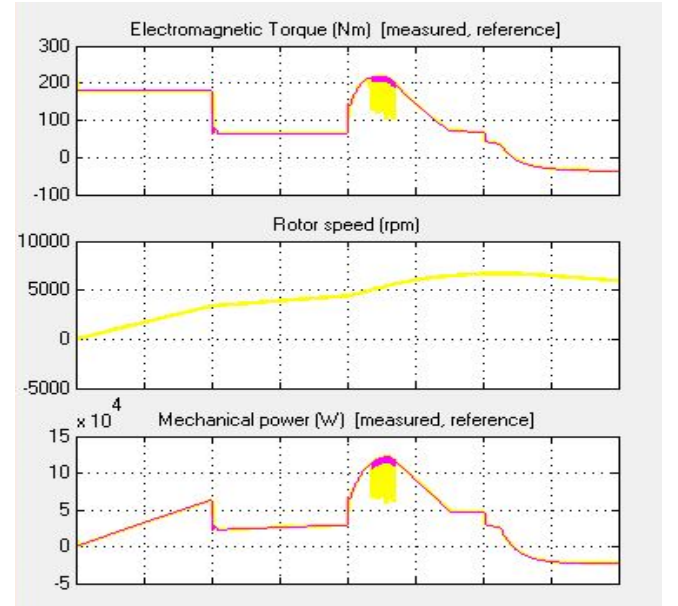


Figure 2. Performance Variation of PMSM

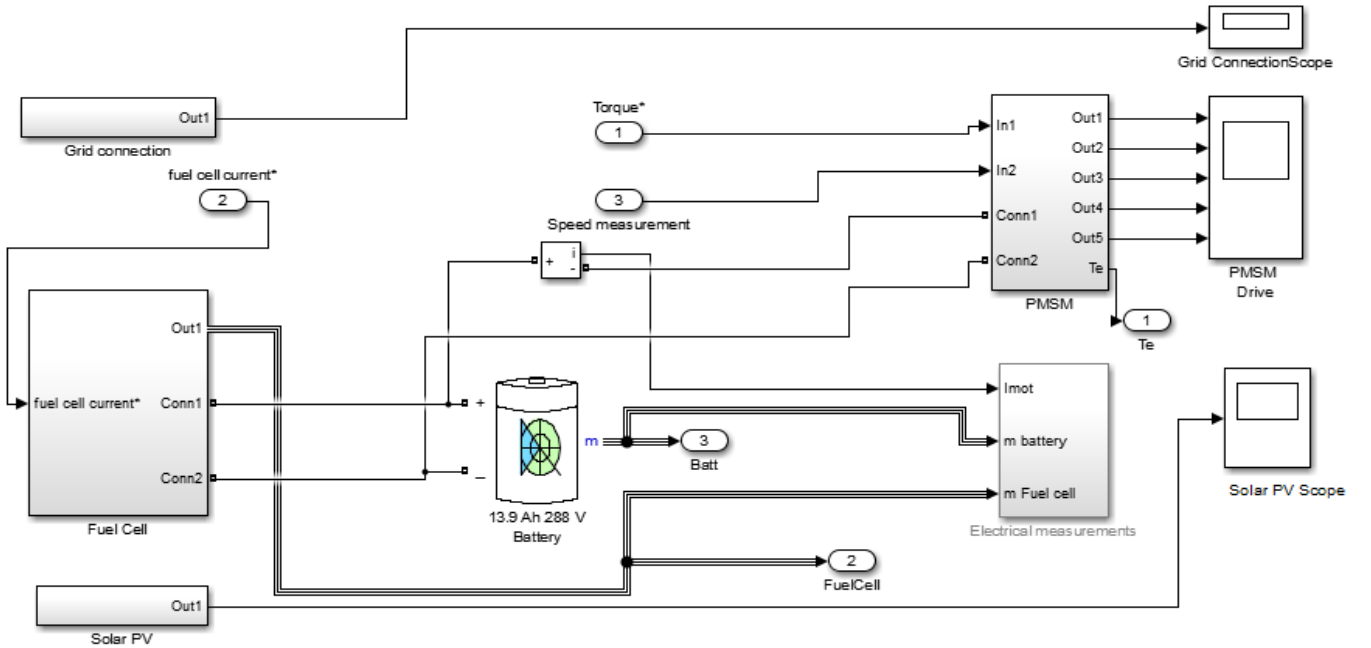


Figure 1. Simulink model of HEV power system

in case of electro-magnetic torque and mechanical power with the reference output. However this performance fluctuation is for a small period of time and can be neglected as overall rotor speed is continuously increasing. The stable output of rotor speed ensures that this two slightly fluctuated value has a very low impact on the overall stability of the entire system.

According to Fig. 3 quadrature current varies within the range 0 A to 500 A and the waveform also tends to follow the expected. Similarly the reference voltage, quadrature voltage varies within the range of 0 V to 200 V and direct voltage varies within the range of 0 V to -200 V at $t=16$ sec.

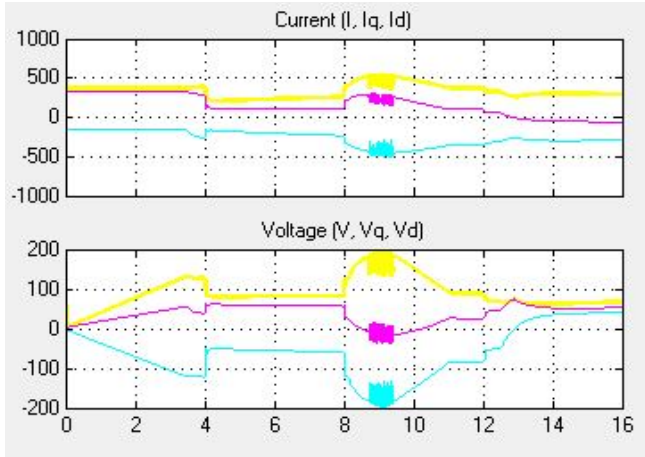


Figure 3. Current and Voltage response of PMSM

B. Grid Charger

A grid charger is an electronic device that can charge the battery pack of hybrid electric vehicle at a very low current. In near future vehicle to grid charge concept can be seen where electric vehicles can be charged directly from the grid [11]-[12]. Its name implying that it will use AC power for charging. During parking electric vehicle can be recharged and by vehicle to grid technology a vehicle can recharge its battery pack anywhere. This will be a great improvement as battery electric vehicles are struggling with their battery packs. Widespread implementation of vehicle to grid technology will remove the driving range limitations of electric vehicle. Detail specification of grid charger used in the Simulink model are given below.

TABLE I. GRID CHARGER SPECIFICATION

Parameter	Value
Peak amplitude	312 V
Frequency	50 Hz
Phase	0°
Transformer Primary	312 V
Transformer Secondary	280 V
Diode Parameters	
Snubber resistance	500 ohm
Forward voltage	0.8 V
Resistance Ron	.001 ohm

The Simulink model of grid charger which is prepared to combine with the Simulink model of HEV power system are visualized below. The model is consist of a transformer, four diodes and other RLC elements.

According to Figure 5 the output voltage of the grid charger varies within the 255 V to 280 V at $t=16$ sec which meets the desired system requirement of 280 V.

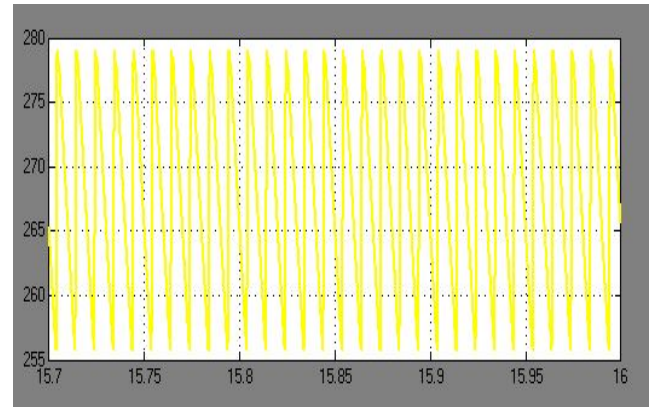


Figure 5. Response curve of Grid Charger

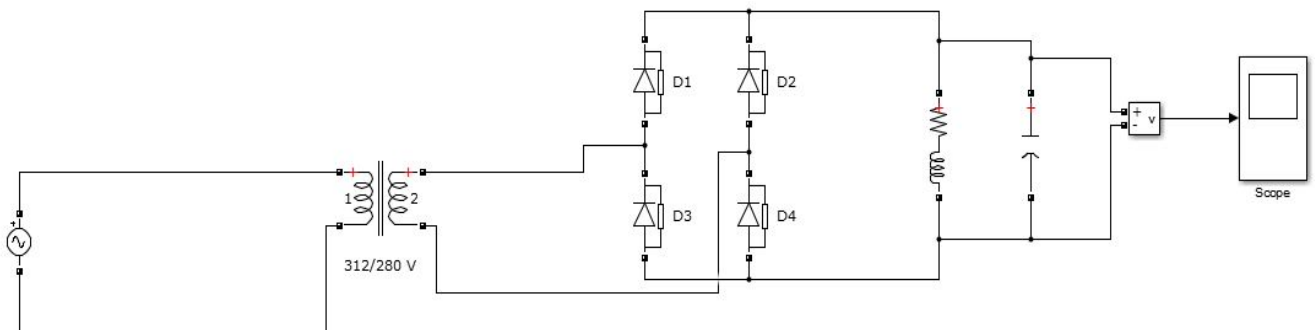


Figure 4. Simulink model of Grid Charger

C. Photovoltaic Panel

Commercial photovoltaic panels cannot provide equal performance due to average thermal conductivity of substrate material [13]. Here eighteen photovoltaic cell are used to complete the model of photovoltaic panel. According to Figure 6 the Simulink model of photovoltaic panel used in the HEV power system to support the battery is shown.

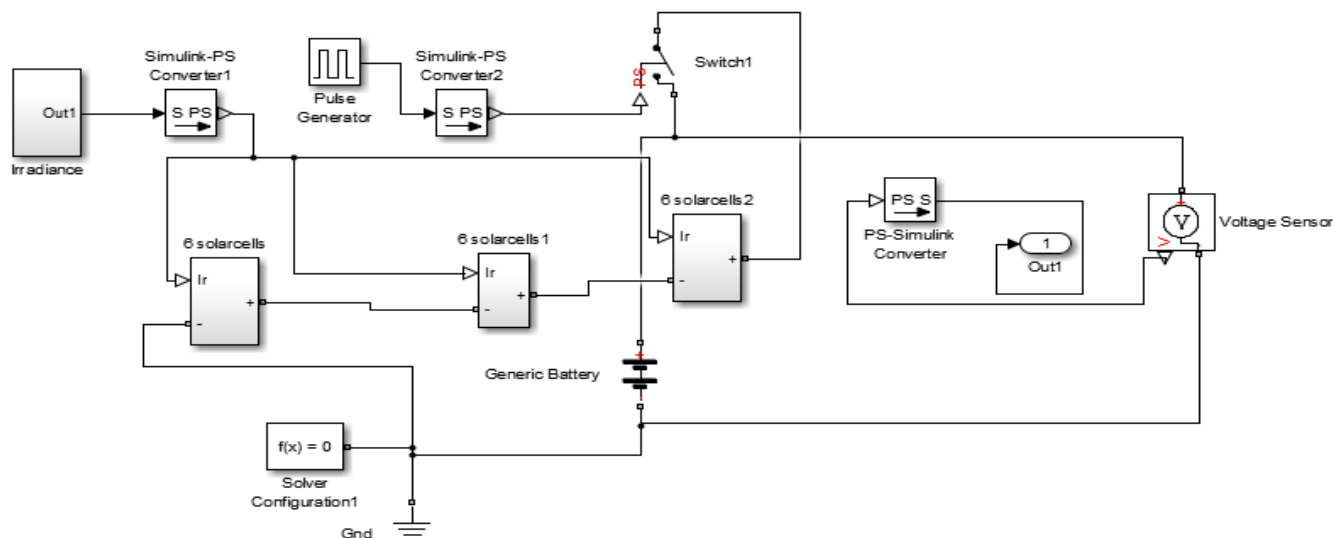


Figure 6. Simulink model of Photovoltaic Panel

According to the Fig. 7 the output response of the photovoltaic panel is shown. It is mentioned earlier that photovoltaic panel will support the auxiliary battery of the vehicle which is typically contains the rating of 12 V. From the simulation result it is seen that initially the panel generating 11.929 V however the generation is slightly reduced in course of time which remains till 15.95s. In 16th second and onwards a constant supply of 11.928 V is achieved which is very close to recharge a 12 V battery. That means photovoltaic panel successfully meets the system requirements.

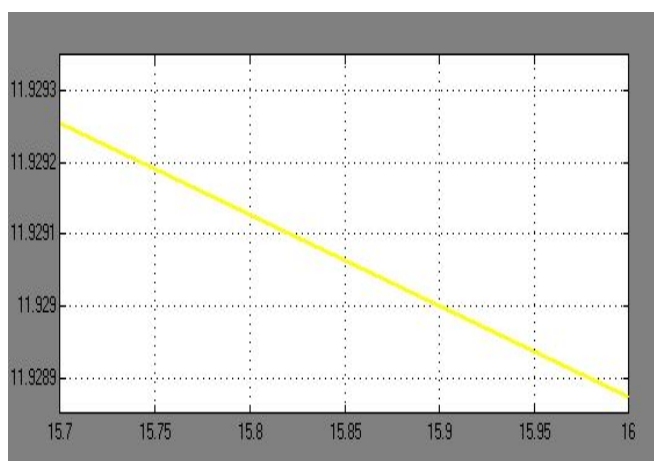


Figure 7. Solar panel response curve

D. Fuel Cell

Fuel cell is an integral part of the proposed hybrid electric vehicle power system. The use of fuel cell was initiated by NASA as supporting power source to meet the power demand of their space shuttle. Overall performance of the vehicle specifically the fuel economy of the vehicle depends on the vehicle energy management system where the power distribution between the fuel cell and battery is done [14].

That means to improve the fuel economy of the vehicle it is equally important to improve energy management system. To measure the performance of the fuel cell system generally the initial state of charging and final state of charging are considered. Specification of fuel cell used in the Simulink model are given below.

TABLE II. FUEL CELL SPECIFICATION

Parameter	Value
Number of cell	400
Stack efficiency	57%
Operating temperature	95°
Nominal air flow rate	1698 lpm

Total number of fuel cell stack used in this Simulink model is 400 where stack efficiency is 57%. Operating temperature plays important role in case of fuel cell. In this system the system temperature kept 95°C. Vehicle fuel cell system can be classified into four sub-system and the air flow sub-system plays important role with having greater impact on fuel cell overall performance [15]. Air flow system is generally consists of cooler, compressor, valve and pipes etc. To maintain the stable and reliable performance air flow control is an impressive method for fuel cell system. Here nominal air flow rate is 1698 lpm.

According to Figure 8 it is shown that 400 fuel cell stacks are sufficient to produce the expected voltage of 280 V and the power produced by the stacks is 85.5 kW when the current is 280 A and produced power is 100.02 kW when the current is 350A. As auxiliary loads are supported by the photovoltaic panel so power produced from the fuel cell will be used completely for the propulsion of the vehicle. The relation between stack voltage vs. current and stack power vs. stack current reflects the accurate power production from fuel cell.

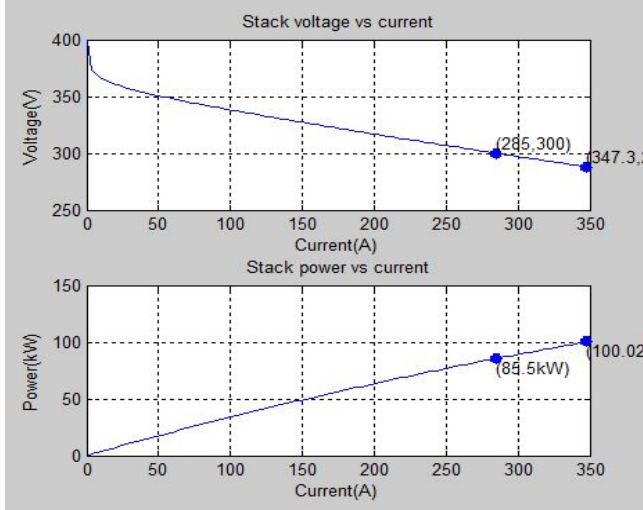


Figure 8. Fuel cell response curve

E. Battery

In any energy management system (EMS), energy storage system as well as battery play the key role. Lithium-ion battery is used in this Simulink model. Detail specification of the battery used in the Simulink model are visualized below.

TABLE III. BATTERY SPECIFICATION

Parameter	Value
Battery type	Lithium-ion
Nominal voltage	288 V
Rated capacity	13.9 Ah
Initial state of charge	40.32%
Fully charged voltage	335.2283 V
Battery response time	30 sec

IV. OVERALL PERFORMANCE MEASUREMENT

Figure 9 represents an overall electrical measurement of the HEV power system. Here the voltage through DC-DC converter varies within 0 V to 400 V. From output response it is seen that the converter output failed to follow the reference waveform accurately though conversion in different time period are done successfully. If the focus can be given on the battery and fuel cell voltage, some distortion in the battery performance can be found though fluctuation for a very small

period of time so this fluctuation can be neglected. Though optimum voltage for the propulsion is 288V but the voltage production from fuel cell varies within the range of 300-450 V which is quite satisfactory for the smooth operation of the system. Current production of fuel cell is varies within the range 0 A to 400 A which is within the desired range of the system. According to the figure the state of charge (SOC) lies between the range 39.8% to 40.8% where system requirement is 40.32% so the output response of state of charge (SOC) clearly satisfies the demand.

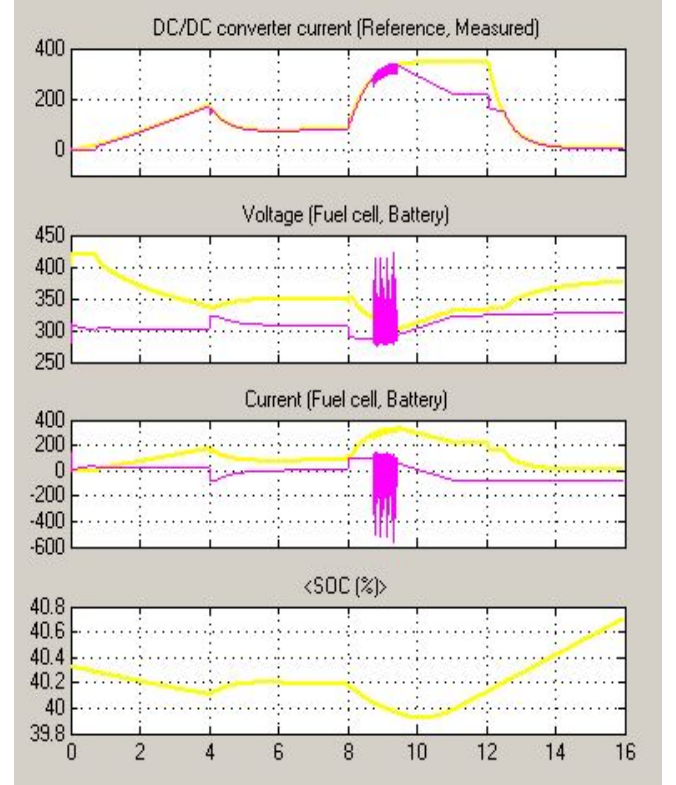


Figure 9. Performance overview of HEV power system

CONCLUSION

In this paper mainly discussed modelling and simulation of HEV power system. More specifically performance of various electrical components of HEV power system are closely observed and analyzed. Comparison of the performance of various components of HEV power system with the state of the art methods are done. The growth of market potential of hybrid electric vehicle can be kept up by keeping the uninterrupted supply of power. That means the long lasting success of HEV is largely dependent on the sustainability and flexibility of its power system. Automotive market is currently struggling with the high price of the fuel and the most convenient way to face this is to rely on hybrid electric vehicle. Beside this a major advantage of hybrid electric vehicle is the lower emission of CO₂ which ensures the safety of environment. This paper represented a complete Simulink model of hybrid electric vehicle power system which is powered by PEM fuel cell, Lithium-ion battery and photovoltaic panel.

The model represented in this paper ensures the detail and accurate electrical measurement. The simulation is done with MATLAB Simulink. The efficiency of the system met with the expectation which is ensured by comparing resultant output with the state of the art methods but this can be improved by selecting more efficient power electronic equipment's. Hybrid Permanent Magnet Synchronous Machine can play the key role to improve the efficiency of this model.

REFERENCES

- [1] Number of electric car worldwide can be accessed at: <https://evannex.com/blogs/news/77801925-number-of-electric-cars-worldwide-climbs-to-1-3-million-tesla-model-s-takes-top-spot-among-new-ev-registrations> [August 21,2016].
- [2] M. Anderman, "Status and trends in the HEV/PHEC/EV battery industry," Rocky Mountain Institute, 2008.
- [3] Subotic, Ivan, and Emil Levi. "A review of single-phase on-board integrated battery charging topologies for electric vehicles." *Electrical Machines Design, Control and Diagnosis (WEMDCD)*, 2015 IEEE Workshop on. IEEE, 2015.
- [4] Chowdhury, Muhammad Sifatul Alam, Al Mahmudur Rahman, and Nahidul Hoque Samrat. "A comprehensive study on green technologies used in the vehicle." *Green Energy and Technology (ICGET)*, 2015 3rd International Conference on. IEEE, 2015.
- [5] Cheng, K. W. E., X. D. Xue, and K. H. Chan. "Zero emission electric vessel development." 2015 6th International Conference on Power Electronics Systems and Applications (PESA). IEEE, 2015.
- [6] Detail statistics about CO₂ emission is available at: <http://co2now.org/> [August 21,2016].
- [7] Chowdhury, Muhammad Sifatul Alam, and Al Mahmudur Rahman. "Electric efficiency of the lighting technology of auto industry: recent development and future prospect." 2016 4th International Conference on the Development in the in Renewable Energy Technology (ICDRET). IEEE, 2016.
- [8] Simulink model exchance community on official Matlab website: <http://www.mathworks.com/> [August 21,2016].
- [9] R Krishnan, "Permanent Magnet Synchronus and brushless DC Motor drives", PHI,ch 9 pp 518-555 .
- [10] AE Fitzgerald, Stephen D Umans,"Electrical Machinery," TATA McGraw 6th edition,ch 5 pp 245-293, 578 .
- [11] Shetty P, Downee S, "Modeling and simulation of the complete electric power train of a Hybrid Electric Vehicle ," International conference on Magnetics, Machines and drives, July 24-26,2014.
- [12] N. Weise, K. Mohapatra, and N. Mohan, "Universal utility interface for plug-in hybrid electric vehicles with vehicle-to-grid functionality," in *Power and Energy Society General Meeting*, 2010 IEEE, july 2010, pp. 1 –8.
- [13] Zakharchenko, R., Licea-Jiménez, L., Pérez-García, S. A., Vorobiev, P., Dehesa-Carrasco, U., Pérez-Robles, J. F., & Vorobiev, Y. "Photovoltaic solar panel for a hybrid PV/thermal system. *Solar Energy Materials and Solar Cells*", 82(1), 253-261.
- [14] Zheng, C. H., Oh, C. E., Park, Y. I., & Cha, S. W. "Fuel economy evaluation of fuel cell hybrid vehicles based on equivalent fuel consumption".*International Journal of Hydrogen Energy*, 37(2), 1790-1796.
- [15] Guo, A., Chen, W., Li, Q., Liu, Z., & Que, H.. "Air flow control based on optimal oxygen excess ratio in fuel cells for vehicles". *Journal of Modern Transportation*, 21(2), 79-85.

Analyses of Different Types of Errors for Better Precision in GNSS

Sushmita Joardar¹, Tanveer Ahmed Siddique², Sakila Alam³, Md Hossam-E-Haider⁴

Electrical, Electronics and Communication Engineering Department

Military Institute of Science and Technology

Dhaka, Bangladesh

tratina9@yahoo.com¹, tanveer2908@gmail.com², sakila_alam@yahoo.com³, haider8400@yahoo.com⁴

Abstract—World is moving forward and getting brighter future day by day because of the technological advancement. Each creation of innovation has its own particular importance. Global Navigation Satellite System (GNSS) is such an innovation which is playing a vital role. The most important components are Global Positioning System (GPS) and Galileo but some other systems are also mentioned like GLONASS, BEIDOU and QZSS (quasi-zenith satellite system). While a satellite orbiting about 12,551 miles above the earth surface transmitting a signal must have some errors. For better accuracy, errors must be minimized to an optimum level. The most common errors that affect GNSS receivers have been studied in this paper. After knowing each system, the main sources of errors and how they affect the quality of precision of the receiver can be determined which will lead to attain better performance.

Keyword—Global Navigation Satellite System (GNSS), Dilution of Precision (DOP), User Equivalent Range Error (UERE), Ionospheric Error, Ephemeris and Clock Error, Multipath Error.

I. INTRODUCTION

The Global Positioning System (GPS) is a space based navigation system providing location and time information in all weather conditions, anywhere on or near the earth where there is an unobstructed line of sight to four or more GPS satellites. The error analysis is important for communicating more clearly in global positioning system. So here we discussed about different error sources and their solutions. Generally all visible satellites are not affected by scintillation. The positioning precision is affected by the number of satellites and their positions. The Dilution of Precision (DOP) is usually used to quantify this precision. The DOP is related to the geometrical distribution of the visible satellites. The User Equivalent Range Error (UERE) is related in both space and control segment error. From these two errors we can find the total positioning error (σ_p) [1]:

$$\sigma_p = \text{DOP} \times \text{UERE} \quad (1)$$

II. DILUTION OF PRECISION

Dilution of precision means error in positional measurement precision which occurs due to additional multiplicative effect of navigation satellite geometry.

The effectiveness of determining the position accurately is reduced by scintillation. Signals are affected by the satellite's location and properties of irregularities. To determine a position each receiver needs to be enabled which can be done by minimum of four satellites. Moreover, 9 to 11 satellites can be detected by each receiver at a time. DOP is mainly measured at each time depending on visible satellite. In general, more satellites will make the DOP value upgrade. And low DOP value will make the positioning more accurate.

$$GDOP = \frac{\Delta(\text{output location})}{\Delta(\text{measured data})}$$

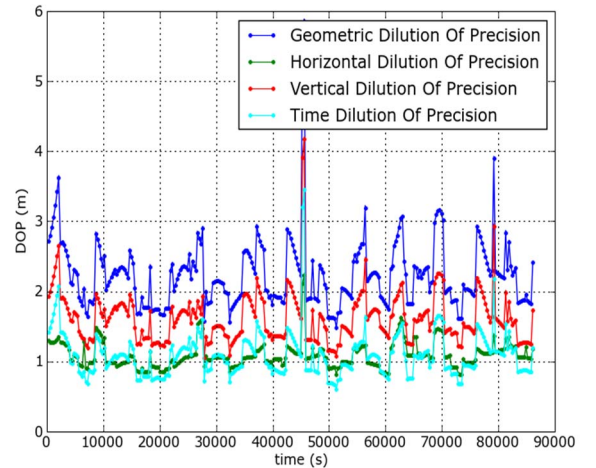


Fig-1 Plot of DOPs

From the above figure we can see that the value of geometric dilution of precision is much more than any other dilution of precision. To minimize the overall dilution of precision we have to give much effort on minimizing the geometric dilution of precision.

DOP can be measured in a number of ways. There are 5 measures of DOP:

- Geometrical Dilution of Precision (GDOP)
- Position Dilution of Precision (PDOP)
- Horizontal Dilution of Precision (HDOP)
- Vertical Dilution of Precision (VDOP)
- Time Dilution of Precision (TDOP)

PDOP consists of HDOP and VDOP. GDOP consists of PDOP and TDOP.

DOP calculation requires the satellites and receiver position in World Geodetic Survey 1984 (WGS84) Earth-Centered Earth-Fixed (ECEF) coordinate. Position of each of the satellites in longitude, latitude and altitude is given for each epoch by the Global Ionospheric Scintillation Model (GISM) program. The equation below is used to convert the position to ECEF co-ordinate [2].

$$x = \left[\frac{a}{\sqrt{1-e^2 \sin^2 \lambda}} + h \right] \cos \lambda \cos \phi \quad (2)$$

$$y = \left[\frac{a}{\sqrt{1-e^2 \sin^2 \lambda}} + h \right] \cos \lambda \sin \phi \quad (3)$$

$$z = \left[\frac{a(1-e^2)}{\sqrt{1-e^2 \sin^2 \lambda}} + h \right] \sin \lambda \quad (4)$$

Where,

- x = ECEF x-coordinate (m)
- y = ECEF y-coordinate (m)
- z = ECEF z-coordinate (m)
- a = Equatorial Earth's radius (6378137 m)
- e = First eccentricity (0.081819190842622 m)
- h = Height above WGS84 ellipsoid (20200000 m)
- λ = Geodetic latitude (radians)
- ϕ = Geodetic longitude (radians)

Then, pseudoranges (R_i) from receiver position to each of the satellites are calculated using equation

$$R_i = \sqrt{(SV_{x_i} - R_x)^2 + (SV_{y_i} - R_y)^2 + (SV_{z_i} - R_z)^2} \quad (5)$$

Where,

- SV_{x_i} = i^{th} satellite position in ECEF x-coordinate (m)
- SV_{y_i} = i^{th} satellite position in ECEF y-coordinate (m)
- SV_{z_i} = i^{th} satellite position in ECEF z-coordinate (m)
- R_x = Receiver position in ECEF x-coordinate (m)
- R_y = Receiver position in ECEF y-coordinate (m)
- R_z = Receiver position in ECEF z-coordinate (m)

Directional derivatives (D) for x , y , z (D_x , D_y , D_z) and time (D_t) was computed:

$$D_{x_i} = \frac{SV_{x_i} - R_x}{R_i} \quad (6)$$

$$D_{y_i} = \frac{SV_{y_i} - R_y}{R_i} \quad (7)$$

$$D_{z_i} = \frac{SV_{z_i} - R_z}{R_i} \quad (8)$$

$$D_{t_i} = \frac{t_i - t_r}{t} = -1 \quad (9)$$

Where,

t_i = i^{th} satellite time (s)

t_r = receiver time (s)

t = signal travelling time from i^{th} satellite to receiver (s)

From the directional derivatives, covariance matrix (A) for each of the satellites was produced as below:

$$A = \begin{bmatrix} Dx_1 & Dy_1 & Dz_1 & Dt_1 \\ \vdots & \vdots & \vdots & \vdots \\ Dx_i & Dy_i & Dz_i & Dt_i \end{bmatrix} \quad (10)$$

$$Q = (A^T \cdot A)^{-1} \quad (11)$$

Where,

A^T = transpose A

$$Q = \begin{bmatrix} Q_{11} & Q_{12} & Q_{13} & Q_{14} \\ Q_{21} & Q_{22} & Q_{23} & Q_{24} \\ Q_{31} & Q_{32} & Q_{33} & Q_{34} \\ Q_{41} & Q_{42} & Q_{43} & Q_{44} \end{bmatrix} \quad (12)$$

Then, DOP was calculated using this equation.

$$GDOP = \sqrt{Q_{11} + Q_{22} + Q_{33} + Q_{44}} \quad (13)$$

$$PDOP = \sqrt{Q_{11} + Q_{22} + Q_{33}} \quad (14)$$

$$HDOP = \sqrt{Q_{11} + Q_{22}} \quad (15)$$

$$VDOP = \sqrt{Q_{33}} \quad (16)$$

$$TDOP = \sqrt{Q_{44}} \quad (17)$$

III. USER EQUIVALENT RANGE ERROR (UERE)

The term user equivalent range error (UERE) refers to the error of a component in the distance from receiver to a satellite. It represents the combined effect of space and control segment errors (satellite vehicle position and clock errors), atmospheric (ionospheric and tropospheric), and user equipment errors (receiver measurement uncertainties).

User Range Accuracy (URA) indicates the errors of Space and control segment. URA value is transmitted by each of the satellites which consist of all errors. Errors such as Selective Availability and satellite clock error. The control segment on earth will correct all of these errors and will be uploading to each of the satellites. The URA index broadcast by each satellites will change over time and its value depends on the length of time since the last control segment upload to that satellite.

The uncertainties in the measurements cause the errors which are User Equipment Errors (UEE). The design of tracking loops, and the strength of satellite signals, receiver noise,

multipath, and antenna orientation are included in these uncertainties.

UERE cannot be broadcast by the satellites and can only be estimated at the receiver. By using the mean of range error we can calculate UERE:

$$UERE = \sum \frac{R_i}{i} \quad (18)$$

Where,

R_i = Range error

i = Number of satellites

Some sources of User Equivalent Range Error (UERE) have given below:

Table-1 Source of User Equivalent Range Error (UERE)

Source	Error(m)
Signal arrival C/A	$\sigma \approx \pm 3$
Signal arrival P(Y)	$\sigma \approx \pm 0.3$
Ionospheric effects	$\sigma \approx \pm 5$
Ephemeris errors	$\sigma \approx \pm 2.5$
Satellite clock errors	$\sigma \approx \pm 2$
Multipath distortion	$\sigma \approx \pm 1$
Tropospheric effects	$\sigma \approx \pm 0.5$
C/A	$\sigma \approx \pm 6.7$
P(Y)	$\sigma \approx \pm 6$

User equivalent range errors (UERE) are shown in the table-1. There is also a numerical error with an estimated value, σ_{num} , of about 1 meter. The standard deviations, σ_R , for the C/A and P(Y) codes are also shown in the table.

IV. ATMOSPHERIC ERROR AND MITIGATION

GPS signals are interrupted while passing through earth's atmosphere due to atmospheric inconsistency. It is caused by the ionosphere and troposphere affecting the satellite RF signal. GPS position accuracy can be improved by correcting these errors.

• Ionospheric Error

Ionosphere has a significant concentration of ions and free electrons in the upper strata of the atmosphere. It extends for approximately 50 km to 1000 km above the surface of the earth and has a dispersive effect on GPS signal frequencies. Free electrons and ions are produced by Ultra-violet light and x rays from the sun. During the night these ions and electrons are slowly recombined but at sunrise there are still available

ions and electrons. The error will be determined by the number of electrons that the signal will find between the satellite and the receiver. This can be expressed by the Total Electron Content (TEC). The ionosphere is a dispersive medium and so the code is delayed and the phase is advanced. Using the pseudo-distance concept, two frequency receivers will be able to easily mitigate this error.

Fig-2 shows various ionospheric error components in several times of the day. The blue color indicates the ionospheric error component when all pseudorandom noise code (PRN) is activated and the red color indicates the error component when only PRN-13 is activated. Ionospheric error component is present all the time. So it is a very important component to mitigate for precisely finding the position of the user.

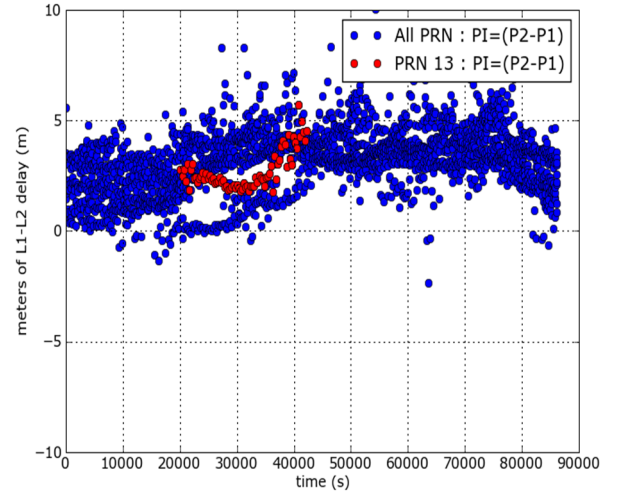


Fig-2 Ionospheric Combination

Klobuchar algorithm is used by GPS to mitigate the errors due to the ionosphere in single frequency receivers. This algorithm results from a compromise between the computational complexity, the knowledge of daily TEC variations, the number of coefficients in the satellite message to ionospheric corrections and the probability that a single frequency user is used in a certain region. Using this equation about fifty percent of errors can be corrected.

The initial values known are user latitude (Φ_u), user longitude (λ_u), satellite elevation angle (E) and azimuth (A). The steps are the following [3]:

1. Earth angle determination:

$$\psi = \frac{0.0137}{E+0.11} - 0.022 \quad (19)$$

2. Sub-ionospheric latitude determination:

$$\Phi_I = \Phi_u + \psi \times \cos A \quad (20)$$

$$\begin{aligned} \text{If } \Phi_I > +0.416 & \quad \text{then} \quad \Phi_I = +0.416 \\ \text{If } \Phi_I < -0.416 & \quad \text{then} \quad \Phi_I = -0.416 \end{aligned}$$

3. Sub-ionospheric longitude determination

$$\lambda_l = \lambda_u + \frac{\psi \times \sin A}{\cos \Phi_l} \quad (21)$$

4. Geomagnetic latitude determination

$$\Phi_m = \Phi_l + 0.064 \times \cos(\lambda_l - 1.617) \quad (22)$$

5. Local time determination

$$t = 4.32 \times 10^4 \lambda_l + GPStime(s) \quad (23)$$

6. Obliquity factor

$$F = 1.0 + 16.0 \times (.53 - E)^3 \quad (24)$$

7. Ionospheric delay determination

$$T_{IONO} = F \times \left[5 \times 10^{-9} + \sum_{n=0}^3 \alpha_n \Phi_m^n \times \left(1 - \frac{x^2}{2} + \frac{x^4}{24} \right) \right] \quad (25)$$

With:

$$x = \frac{2\pi(t - 50400)}{\sum_{n=0}^3 \beta_n \Phi_m^n}$$

There are some constraints regarding this Algorithm. Assumption is that TEC is concentrated in a very thin layer at 350 km high and during the night a constant of 5ns is considered.

• Tropospheric Error

A variable delay caused by Humidity results in errors in troposphere which is similar to ionospheric delay. This effect is more localized and changes more quickly than ionospheric effects, and is not frequency dependent. These traits make precise measurement and compensation of humidity errors more difficult than ionospheric effects.

We will discuss about ITU-R. The ITU-R P.618-9 (2007) model is based on indirect measurements, and the data were obtained from the International Telecommunication Union (ITU) data bank.

$$\partial_{pre} = \frac{\partial_{ref} f^{7/12} g(x)}{(\sin \theta)^{1.2}} \quad (26)$$

$$As(p) = a(p) \cdot \partial_{pre} (dB) \quad (27)$$

Where,

$$\partial_{ref} = 3.6 \times 10^{-3} + N_{wet} \times 10^{-4} \text{ dB} \quad (28)$$

$$N_{wet} = \frac{3732 H es}{(273+T)^2} \quad (29)$$

$$g(x) = \sqrt{3.86(x^2 + 1)^{\frac{11}{12}} \times \sin\left[\frac{11}{6} \arctan\left(\frac{1}{x}\right) - 7.08x^{5/6}\right]} \quad (30)$$

$$L = \frac{2h_L}{\sqrt{(\sin \theta)^2 + 2.35 \times 10^{-4} - 4 + \sin \theta}} \quad (31)$$

Where h_L is the height turbulent layer; the value to be used is $h_L = 1,000\text{m}$.

$$a(p) = -0.061 * (\log p)^3 + 0.072 * (\log p)^2 - 1.71 * (\log p) + 30.01\% < p < 50.0\% \quad (32)$$

V. EPHEMERIS AND CLOCK ERROR

Satellite clock error appears when true time (GPS time) cannot be synchronized with satellite oscillator. Quadratic function of time is used to determine the difference between a particular clock and GPS system. Before broadcasting, satellite is uploaded with the estimated values of the parameters of the above model. Satellite clock error can be computed using the below steps.

The transmission time of satellite is calculated as [5],

$$t_t = t_r - (P \div c) \quad (33)$$

Where,

t_t = Transmission time of satellite (sec)
 t_r = Reception time at receiver of satellite (sec)
 P = Pseudorange (meters)
 c = Speed of light (meters/sec)

Here, propagation time is represented by P/c . The satellite clock offset (dt) is estimated,

$$dt = t_t - t_o \quad (34)$$

Where,

t_t = Transmission time of satellite (sec)
 t_o = Time of clock (sec)

In the below equation, the second order polynomial has been used to represent the satellite clock error (ε^{sat}). Here, the parameters which represents the behavior of the satellite clock are clock bias (a_{f0}), clock drift (a_{f1}) and clock drift rate (a_{f2}).

$$\varepsilon^{sat} = a_{f0} + a_{f1}dt + a_{f2}dt^2 \quad (35)$$

Where,

a_{f0} = Clock bias (sec)
 a_{f1} = Clock drift (sec/sec)
 a_{f2} = Clock drift rate (sec/sec²)
 dt = Satellite clock offset (sec)

Finally transmission time of GPS is calculated as below,

$$T_{GPS} = t_t - \varepsilon^{sat} \quad (36)$$

Where,

T_{GPS} = Corrected transmission time of GPS (sec)
 t_t = Transmission time of satellite (sec)
 ε^{sat} = Satellite clock error (sec)

The final equation represents the corrected transmission time of GPS (T_{GPS}) which is used to determine satellite position.

Fig-3 shows each satellite's orbital solution vary from their actual location. It is computed from broadcast ephemerides.

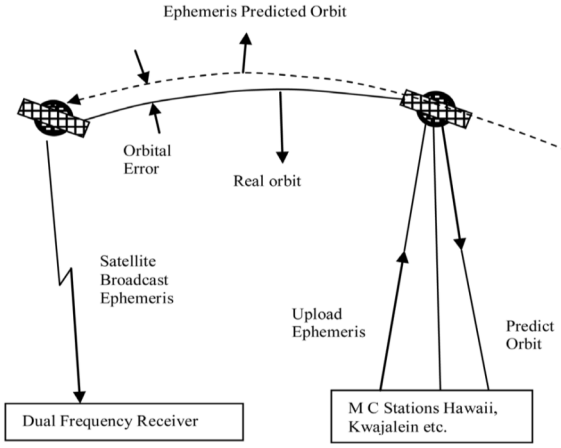


Fig-3 Satellite Ephemeris Errors

VI. MULTIPATH EFFECTS

When radio signals reflect off encompassing landscape; structures, gulch dividers, hard ground, and so on, GPS signals can be biased and thus occur multipath error. As different GPS signals have different wavelengths, multipath effect cause error in signal in different level.

If long delay multipath occurs, the receiver can determine the unwanted signal and diminish it. But when shorter delay multipath occurs, specialized antennas may reduce the signal power. Short delay multipath is harder to filter out.

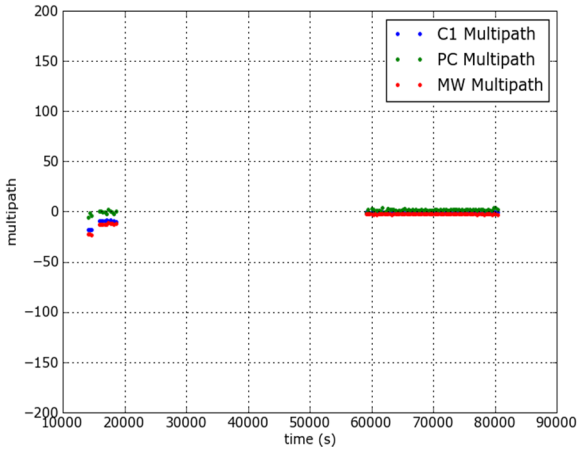


Fig-4 Multipath error component

The fig-4 shows various component of multipath. C1 code multipath is presented for a very short time, but PC multipath and MW multipaths are presented for comparatively long time. Multipath error does not exist for the whole period of time rather it vanishes within some period. So it is not as prominent as ionospheric error and thereby mitigation process is comparatively easy.

Septentrio has developed an experimental way by utilizing the measure of the amount of the distortion of the correlation peak for alleviating the multipath impact. The technique used here is known as A-Posteriori Multipath Estimation technique (APME). The assumption in this equation is that for multipath free signals, the slope of the correlation peak is same. As indicated by the following formula by the estimations of the multipath delay, it corrects the measurements [Sleeweagen and Boon, 2001]:

$$MP = -0.42(1 - \frac{I_{+2}}{I_0(1-d)}) \quad (37)$$

Here,

I_0 = punctual correlation value.

I_{+2} = correlation value which is measured when d chips delay from their original punctual code.

Thus, the multipath error affecting the code-phase measurement tracking by estimating a-posteriori by the APME is subtracted from the code-phase measurement. Also the standard deviation of the multipath estimates is provided by APME. Now consider the model:

$$E\{y\} = (B \ A) \begin{pmatrix} b \\ a \end{pmatrix} + M_m ; D\{y\} = Q_y \quad (38)$$

Here,

M is a vector related to one or more satellites with the multipath mistakes on the code estimations. M indicates which estimations are influenced.

Without considering this term M_m if the float solution would be calculated, we would obtain a biased solution, with

$$\nabla_y = M_m \quad (39)$$

Before computing the float solution, we don't consider other parameters while estimating min APME. And thus estimations of $\hat{m}$ and related variance matrix $Q_{\hat{m}}$ are obtained. This assessment is then used to correct the observations:

$$y^c = y - M_{\hat{m}} ; Q_{y^c} = Q_y + M Q_{\hat{m}} M^T \quad (40)$$

As we have to take into account the uncertainty in multipath estimates, the precision of y^c is lower than that of the uncorrected observations. Nevertheless we have to consider this because without the correction we would obtain a biased solution.

CONCLUSION

As the use of GPS system is increasing, the investigation of errors is becoming more important for better accuracy, integrity and availability. Several types of errors have been discussed in this paper, which are DOP, UERE, ionospheric error, tropospheric error, ephemeris and clock error and multipath error. Solution has been provided of these errors to some extent for more accuracy of position. Reducing dilution of precision (DOP) error, we may find out users position

precisely for any constellation of satellite. Then reducing user equivalent range error (UERE), signals at long ranges can be received with more accuracy. Moreover reducing ionospheric and tropospheric errors, convergence time will be minimized to an optimum level. Thus minimizing the errors, user will be able to receive the signals with more accuracy and precision.

REFERENCES

- [1] Beniguel, "Global Ionospheric Scintillation Model," Technical Manual Vers. 5.1, IEEE, ITU-R, Oct 2004.
- [2] Y. H. Ho, S. Abdullah, M. H. Mokhtar, "Global Positioning System (GPS) Positioning Errors During Ionospheric Scintillation Event," March 2014.
- [3] Teresa Sousa Silva, "Error Analysis in GNSS receivers," October 2007
- [4] James Bao-Yen Tsui, "Fundamentals of Global Positioning System Receivers: A Software Approach," Chapter- 2 & 5, 2000.
- [5] Bharati Bidikar, Gottapu Sasibhushana Rao, Laveti Ganesh, MNVS Santosh Kumar, "Satellite Clock Error and Orbital Solution Error Estimation for Precise Navigation Applications," December 2013.
- [6] Sandra Verhagen and Dennis Odijk, Frank Boon, Jose Maria Lopez, "Reliable Multi-Carrier Ambiguity Resolution in the Presence of Multipath," ION GNSS 20th International Technical meeting of the Satellite Division, 25-28 September 2007, Fort Worth, TX.

Theoretical Efficiency and Cell Parameters of AlAs/GaAs/Ge Based New Multijunction Solar Cell

¹Syeda Puspita Mouri, ¹Syed Nazmus Sakib, ²Shahidul Hoque, ³M. Shamim Kaiser

¹Department of Electrical, Electronic and Communication Engineering,
Military Institute of Science and Technology, Mirpur Cantonment, Dhaka, Bangladesh

²Department of Mechanical Engineering, Khulna University of Engineering and Technology, Khulna, Bangladesh

³Institute of Information technology, Jahangirnagar University, Savar, Dhaka, Bangladesh

Email: ¹mourimisteece10@yahoo.com

Abstract— This paper represents theoretical efficiency and cell parameters of AlAs/GaAs/Ge based new multijunction solar cell. A very new and unique model of multijunction solar is introduced in this paper. Using the modified version of spectral p-n junction solar cell model, the calculations of the solar cell parameters are done. All the calculations done for this work are completed by using MATLAB simulation using different appropriate equations given in various literature. The J-V characteristics are plotted for different solar radiation. In this paper, we have shown the calculations for different Air Mass (AM) condition such as AM 1.5D, AM 1.5G and AM 0 solar spectrum. The efficiency of the solar cell varies with the change in the Air Mass conditions. Solar cell parameters like short circuit current density, open circuit voltage, fill factor, maximum power are calculated. The efficiency for AM0 condition is found 50.7484% while for AM 1.5G condition the efficiency is 44.5156%. For AM 1.5D condition the efficiency is 48.2183%.

Keywords—renewable energy; solar energy; solar cell; multijunction solar cell; air mass; efficiency.

I. INTRODUCTION

Power crisis has been the most serious and important issue from the last few decades. Power being a crying need to every human, becomes crucial factor when it does not meet the full demand. Engineers and researchers across the whole world has been working relentlessly to find out an appropriate solution for the problem. Even lots of works have been undertaken to reduce the consumption of the electrical energy by introducing the latest smart technology [1]. With the limited amount of natural sources like biofuels, natural gas, coal etc. power generation has become very difficult. It is so because the cost of extraction of various minerals are increasing and the amount of it are decreasing. But to generate the power, these minerals are burning and creating a huge amount of pollution to the environment. The adverse effect of this pollution has made the world to think in a different way to generate the power for the people. Hence is the importance of the renewable energy.

Renewable energy is getting more popularity because of its unlimited availability and cost effectiveness. Where every other sources of energy is creating massive pollution in the environment, renewable energy offers power with nontoxicity and pollution free. The sources of renewable energy are solar, wind, rain, tides, hydro and geo thermal heat, biodiesel, biofuel. Among them solar energy is catching more attention towards it as a source of electrical energy. Not only in the power generation but also solar energy is now being used as a

source of energy for the vehicles [2][3]. Solar vehicles are performing very well in different countries. Solar cell or Photovoltaic cell is used to convert the solar energy to useable electrical energy. Its distribution and installation is very simple. The areas, where it is difficult to supply the generated power, solar cell are the best and only suitable solution.

Solar radiation is not same in all places in the world. It also varies with the Air mass. Different solar radiation for various air mass is shown in fig.1. As the solar spectrum varies, the condition for solar cell varies along with the efficiency. The work on increasing the efficiency of solar cell is still on. Every researcher is trying to improve the solar cell efficiency. This paper also represents a work on improved efficiency by using different materials for solar cell. For making high efficiency solar cell, material choice is a big issue. All the challenges and the outcomes of our work is described in the later sections with proper data analysis and reference.

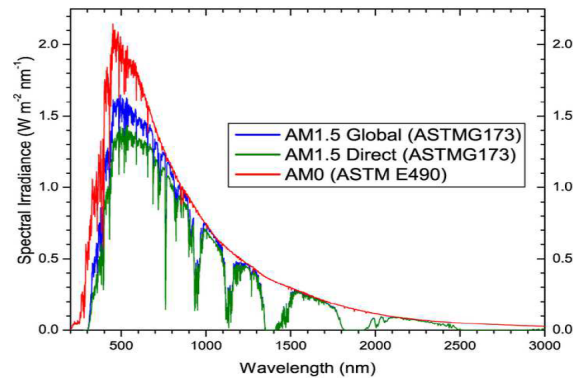


Fig. 1. Solar spectrum for different Air Mass condition [4].

In this paper section I introduces the idea about our work, section II depicts the literature review that we have used to do this research work, section III describes about the multijunction solar cell, section IV gives the details about our proposed solar cell model, section V shows the assumptions and simulation parameters, sections VI reveals the results obtained by MATLAB simulation and the work is concluded in the next section along with the references used mentioned in the later part.

II. LITERATURE REVIEW

Earlier photovoltaic cells or solar cells were used mostly in space applications because of its rare availability and high cost.

For the satellites and vehicles in space, the solar cells are the only long duration power supply device available to the researchers. In terrestrial applications, they were used in a very small scale. As time has changed and the technology advances, this concept has been changed by the researchers. Now people all over the world are keener on using solar panels for getting the power for their houses as well as for the industry. As the energy demand of the whole world is increasing day by day, conventional energy sources will be completely used up in the near future. Because of its capability to convert sunlight directly to electricity with high conversion efficiency, solar cell can provide pollution free and cost effective power in a permanent basis. Lots of works are going on for making the solar cell financially feasible with keeping the high efficiency for using on a large scale.

With an efficiency of 6%, first solar cell was made in bell laboratory using silicon [5] in 1954. Later that year Reynolds developed another one using cadmium sulfide [6]. The early days' solar cell were of single junction. But as technology advances, the introduction of multijunction solar cell made a tremendous change in the field of solar energy research. The idea of multijunction solar cell was introduced in 1995 [7] [8]. A single junction solar cell with an efficiency of 33.7% was made and developed by William Shockley and Hans Queisser [9]. Another model was developed for spectral p-n junction model including both multijunction and single junction solar cell [10]. A very new GaAs based mathematical model and electrical equivalent circuit of multi-junction solar cell is introduced in [11]. A new solar simulator was developed based on the idea of supplementary-light method for performance measurement of multi-junction solar cells [12]. In fig.2 the daily solar radiation data is shown for Dhaka city. From the figure it can be seen that the solar radiation is good for solar cell to work effectively. With the introduction of concentrator, the multijunction solar cell becomes more efficient. Different multi junction solar cell models are described in [13] [14] [15]. All these models described in various literature have their own way of representing the solar cell model. For our calculation and modelling process, we have considered the spectral p-n junction model which is widely used for research and experiments in practice.

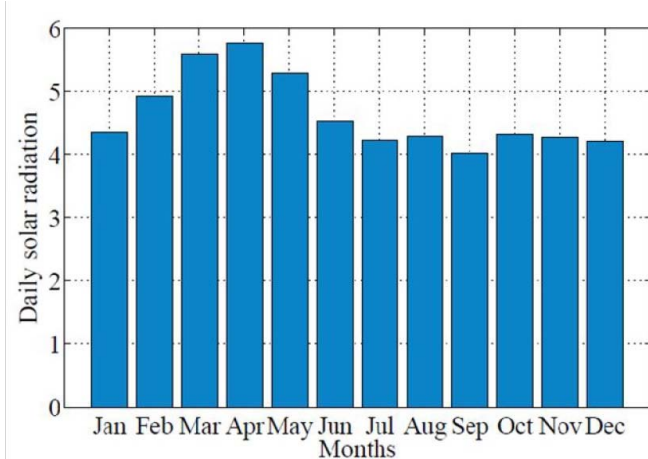


Fig. 2. Bar plot of daily solar radiation data [2].

III. MULTIJUNCTION SOLAR CELL

In a single junction solar cell there is only one layer. The band gap of the material defines the range of wavelength which will be absorbed. The material only absorbs incident photon having energy ($h\nu$) greater than the band gap energy (E_g) of the material. So the incident photons having energy less than the material's band gap energy will remain unabsorbed. This unabsorbed photons lessen the efficiency of a single junction solar cell.

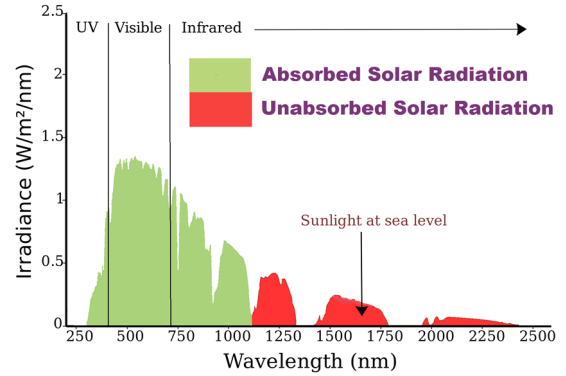


Fig. 3. Solar radiation absorption by single junction solar cell.

In a multijunction solar cell different layer consists of different bandgap material. The materials are ordered with decreasing bandgaps (E_g), allowing photon to transmit to the lower sub-cells. When the sun ray is incident on it, each layer will absorb the photon having energy greater than the material's bandgap. This states that the photon having wavelength lower than the corresponding wavelength of the material will be absorbed. This absorbed photon energy will be converted to electrical energy. The photon power conversion efficiency is maximum when the bandgap (E_g) of a material is equal to the incident photon energy ($h\nu$). Each layer is designed to have a specific bandgap in order to maximize the power conversion over the whole solar spectrum [4]. So for multijunction solar cell the absorption of solar radiation much higher than than the conventional single junction solar cell. Fig. 3 and fig. 4 shows the absorption of photon in single junction and multijunction solar cell.

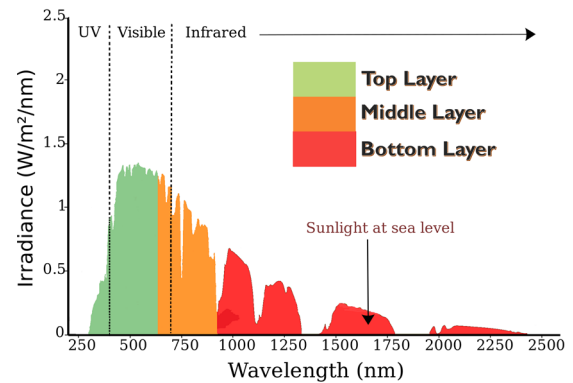


Fig. 4. Solar radiation absorption by multijunction solar cell.

IV. PROPOSED SOLAR CELL MODEL

The proposed model is a three-junction model. 3 junctions are connected in series. Fig. 5 shows the proposed model of 3 junction solar cell.

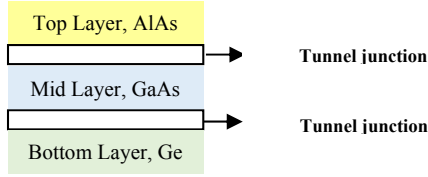


Fig. 5. Proposed Model

The material used here are lattice matched [16][17][18]. For lattice mismatch condition there will be phonon generation and that may lead to decrease of efficiency. In each layer there is a p-n junction and that is the reason why this model is called a 3-junction solar cell. The calculation of efficiency needs a number of parameters which are mentions below.

Nomenclature:

J_{sc} = Short circuit current

J_0 = Reverse saturation current or dark current

e = Charge of an electron

F = Photon flux

λ = Wavelength

I = Irradiance

h = Planck's constant

c = Speed of light

D_e = Diffusion current constant for electron

D_h = Diffusion current constant for hole

N_A = Acceptor concentration

N_D = Donor concentration

τ_e = Minority carrier lifetime for electron

τ_h = Minority carrier lifetime for hole

S_e = Surface recombination velocity for electron

S_h = Surface recombination velocity for hole

n_i = Intrinsic carrier concentration

x_p = Thickness of the p-layer

x_n = Thickness of n-layer

μ_e = Electron mobility

μ_h = Hole mobility

k = Boltzmann's constant

τ_{SRH} = Shockley-Read-Hall lifetime

B = Direct band-band recombination coefficient

M_c = Number of equivalent minima in the conduction band

M_v = Number of equivalent minima in the conduction band

N_c = Effective Density of States at Conduction Band edge

N_v = Effective Density of States at Valence Band edge

T = Temperature

E_g =Energy gap of the material

m_e^* = Effective mass of electrons

m_h^* = Effective mass of holes

P_{in} = Input power density

$F.F.$ = Fill Factor

The short circuit current density J_{sc} can be calculated using the charge of electron irradiance data [19].

$$J_{sc} = e \times F \quad (1)$$

Where,

$$F = \lambda I / hc$$

Rewriting (1) gives the following equation,

$$J_{sc} = \frac{e \lambda I}{hc} \quad (2)$$

The equation for the reverse saturation current density can be expressed as [19][20][21]:

$$J_0 = e \left(\frac{D_e}{\tau_e} \right)^{\frac{1}{2}} \frac{n_i^2}{N_A} \left(\frac{S_e \left(\frac{\tau_e}{D_e} \right)^{\frac{1}{2}} \cosh \left(\frac{x_p}{\sqrt{D_e \tau_e}} \right) + \sinh \left(\frac{x_p}{\sqrt{D_e \tau_e}} \right)}{S_e \left(\frac{\tau_e}{D_e} \right)^{\frac{1}{2}} \sinh \left(\frac{x_p}{\sqrt{D_e \tau_e}} \right) + \cosh \left(\frac{x_p}{\sqrt{D_e \tau_e}} \right)} \right) + e \left(\frac{D_h}{\tau_h} \right)^{\frac{1}{2}} \frac{n_i^2}{N_D} \left(\frac{S_h \left(\frac{\tau_h}{D_h} \right)^{\frac{1}{2}} \cosh \left(\frac{x_n}{\sqrt{D_h \tau_h}} \right) + \sinh \left(\frac{x_n}{\sqrt{D_h \tau_h}} \right)}{S_h \left(\frac{\tau_h}{D_h} \right)^{\frac{1}{2}} \sinh \left(\frac{x_n}{\sqrt{D_h \tau_h}} \right) + \cosh \left(\frac{x_n}{\sqrt{D_h \tau_h}} \right)} \right) \quad (3)$$

For the ease of calculation on the basis of theoretical assumptions the above equation can be simplified as,

$$J_0 = e S_e \frac{n_i^2}{N_A} + e S_h \frac{n_i^2}{N_D} \quad (4)$$

From the Einstein relationship the equation of diffusion current constant for electron and hole can be written as,

$$D_e = kT \mu_e / e$$

$$D_h = kT \mu_h / e$$

The minority carrier life time equation stands for:

$$\tau_e = \frac{1}{\frac{1}{\tau_{SRH}} + B N_A}$$

$$\tau_h = \frac{1}{\frac{1}{\tau_{SRH}} + B N_D}$$

The equations for surface recombination velocity are:

$$S_e = \left(\frac{D_e}{\tau_e}\right)^{\frac{1}{2}}$$

$$S_h = \left(\frac{D_h}{\tau_h}\right)^{\frac{1}{2}}$$

The intrinsic carrier concentration:

$$n_i^2 = 4 M_c M_v \left(\frac{2\pi kT}{h^2}\right)^3 (m_e^* m_h^*)^{\frac{3}{2}} \exp\left(\frac{-E_g}{kT}\right)$$

$$n_i^2 = N_c N_v \exp\left(\frac{-E_g}{kT}\right)$$

The open circuit voltage equation is expressed as:

$$V_{oc} = \left(\frac{kT}{e}\right) \ln \left[\left(\frac{J_{sc}}{J_0}\right) + 1 \right] \quad (5)$$

The maximum voltage is:

$$V_m = V_{oc} - \frac{1}{\beta} \ln(1 + \beta V_m)$$

Where, $\beta = e/KT$

Maximum current density is:

$$J_m = J_{sc} \left(1 - \frac{1}{\beta V_m}\right)$$

The final equation for the efficiency of the solar cell is:

$$\eta = \frac{V_{oc} J_{sc} FF}{P_{in}} \times 100\%$$

Where,

$$F.F. = \frac{V_m J_m}{V_{oc} J_{sc}}$$

So the equation for efficiency becomes,

$$\eta = \frac{V_m J_m}{P_{in}} \times 100\%$$

V. ASSUMPTIONS AND SIMULATION PARAMETERS

The parameters required for calculation are considered at the temperature of 298K. The modified version of spectral p-n junction model [22][23] is used for calculation. The materials are lattice matched and the lattice constant is 5.65 Å [16][17][18]. The input power and the irradiance value changes with different AM condition [19]. The value of irradiance also changes with the wavelength which entirely depends on the material. Table I depicts the irradiance value from ASTM [24] and input power densities for different air mass condition.

TABLE I. INPUT POWER DENSITY AND IRRADIANCE FOR DIFFERENT AM CONDITION [19][24][25][26][27]

AM Condition	Materials	Input Power Density (W/m ²)	Irradiance (Wm ⁻² nm ⁻¹)
AM 0	AlAs	1353	1881
	GaAs		935.6
	Ge		176

AM 1.5G	AlAs	1000	1519.8
	GaAs		926.9
	Ge		114.8
AM 1.5D	AlAs	900	1359.5
	GaAs		862.04
	Ge		112.13

The optoelectronics parameters are enlisted in Table III. Standard units are given in the table while conversion of unit is done while calculation. It has been considered that, there is no carrier generation or recombination in the space charge region. It is also assumed that neutral region operates at low injection condition and electric field in neutral region is negligible.

TABLE II. OPTOELECTRONIC PARAMETERS [28][29]

Parameter	AlAs	GaAs	Ge
E _g	2.17 eV	1.42 eV	0.67 eV
λ	0.5730e-06 m	0.875e-06 m	1.775e-06 m
μ _e	0.02 (m ² /Vs)	0.2322 (m ² /Vs)	0.39 (m ² /Vs)
μ _h	0.01 (m ² /Vs)	0.02 (m ² /Vs)	0.19 (m ² /Vs)
m _e [*] /m _e	0.146	0.067	1.64
m _h [*] /m _h	0.760	0.473	0.28
τ _{SRH}	10 ⁻⁵ (s)	10 ⁻⁵ (s)	10 ⁻⁵ (s)
B	7.5e-16 (s ¹ m ³)	7.5e-16 (s ¹ m ³)	7.5e-16 (s ¹ m ³)
N _A	1.7e24 (m ⁻³)	9e23 (m ⁻³)	1e23 (m ⁻³)
N _D	1.5e24 (m ⁻³)	7.8e23 (m ⁻³)	2e24 (m ⁻³)
X _p	1e-07 m	1e-07 m	1e-07 m
X _n	2.08e-07 m	3e-07 m	4e-07 m

VI. RESULTS

Using the different equations mentioned in section IV and parameters from section V the following results are found using MATLAB. For ease of calculation the losses due to reflection and series resistance is considered as negligible. The proposed multijunction solar cell is considered to be completely optically transparent and the current conductivity from top layer to the bottom layer is very high. In spite of the presence of tunnel junction, the loss causes because of it is negligible. Table III comprises the result for each layer individually for the three AM condition. Table IV includes the

comparison in total cell parameters for AM 0, Am 1.5G and AM 1.5D condition.

TABLE III. COMPARISON OF DIFFERENT LAYERS

Parameter	For AM 0	For AM 1.5G	For AM 1.5D
Top layer GaInP₂ ($E_g = 1.9$ eV)			
Short circuit current density (J_{sc1})	867.0190 (A/m ²)	700.53 (A/m ²)	626.6413 (A/m ²)
Open circuit voltage (V_{oc1})	1.7052 V	1.7 V	1.6968 V
Middle layer GaAs ($E_g = 1.42$ eV)			
Short circuit current density (J_{sc2})	658.5561 (A/m ²)	652.39 (A/m ²)	606.7652 (A/m ²)
Open circuit voltage (V_{oc2})	1.0418 V	1.0415 V	1.0396 V
Bottom layer Ge ($E_g = 0.67$ eV)			
Short circuit current density (J_{sc3})	251.3022 (A/m ²)	163.9175 (A/m ²)	160.1052 (A/m ²)
Open circuit voltage (V_{oc3})	0.1325 V	0.1216 V	0.1210 V

Table III depicts the variations of short circuit current density and the open circuit voltage of the three layers for different AM condition. For each condition short circuit current density in bottom layer is the lowest and this minimum current flows all through the cell.

TABLE IV. COMPARISON OF THE TOTAL CELL PARAMETERS

Parameters	For AM 0	For AM 1.5G	For AM 1.5D
Current Density, J_{sc}	251.3022 (A/m ²)	163.9175 (A/m ²)	160.1052 (A/m ²)
Total open circuit voltage, V_{oc}	2.8794V	2.8628V	2.8574V
Maximum Voltage (V_m)	2.7580V	2.7414 V	2.7362V
Maximum Current Density	248.9602 (A/m ²)	162.3807 (A/m ²)	158.6012 (A/m ²)
Fill factor	0.9489	0.9486	0.9486
Efficiency, η	50.7484 %	44.5156 %	48.2183 %

Table IV shows the various cell parameters in different AM condition. For AM 0 condition the maximum value of open circuit voltage is found and hence the highest efficiency is calculated. This is because of the higher irradiance value [24] and input power. As described in section IV equation (2) for a given wavelength or material the short circuit current density is directly proportional to the irradiance value. According to equation (5) in section IV, with the increase of short circuit current density in each layer the open circuit voltage for each

layer slightly increase. Again as the three layers or junctions are connected in series the total open circuit voltage is the sum of open circuit voltages in each layer. So this higher short circuit current and higher open circuit voltage give a boost to efficiency.

The total leakage current density VS the open circuit voltage of the multijunction solar cell for each layer under different air mass solar radiation spectrum is plotted in the graph.

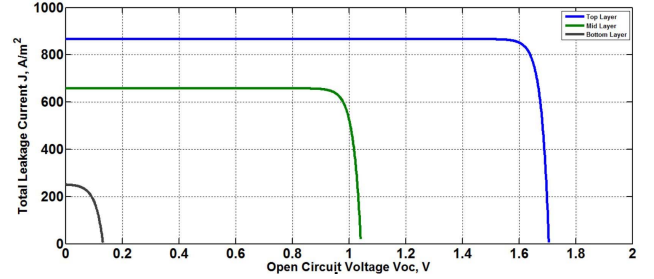


Fig. 6. J-V Characteristics for AM0

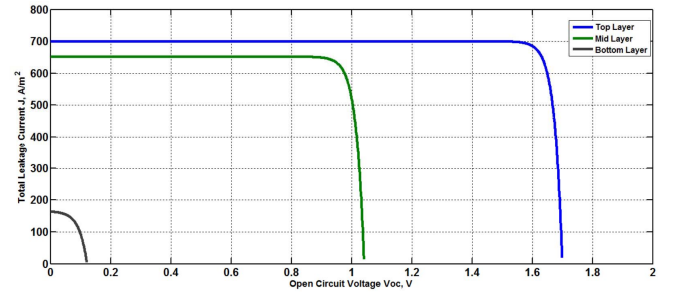


Fig. 7. J-V Characteristics for AM 1.5G

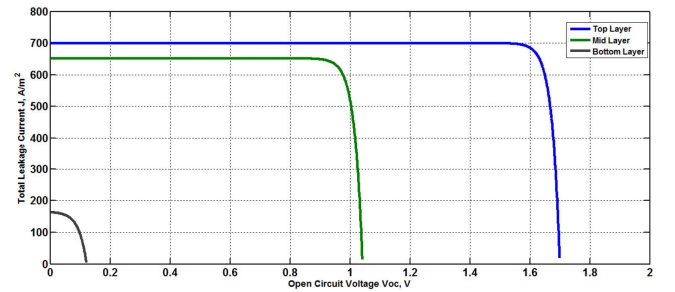


Fig. 8. J-V Characteristics for AM 1.5D

Fig. 6, fig. 7 and fig. 8 shows the J-V characteristics curve each layer of the cell in different AM condition. For certain increment of open circuit voltage the leakage current remains same.

This characteristic may change in the practical scenario because of different losses such as losses due to reflection, grid loss, losses due to series resistance etc. Adverse situations like carrier recombination in the space charge region, the amount of electric field in the neutral region, effects of photon recycling can lessen the leakage current density and hence the open circuit voltage. Decrement of open circuit voltage will cause reduction in overall efficiency.

VII. CONCLUSION

The paper represents the variations in efficiency of AlAs/GaAs/Ge based multijunction solar cell in different AM conditions such as AM0, AM1.5D & AM 1.5G. The highest efficiency is found in AM0 condition and that is 50.7484%. Second highest efficiency is 48.2183% and it is calculated for AM 1.5D condition which is slightly higher than that of AM 1.5G condition.

VIII. REFERENCES

- [1] S. P. Mouri, S. N. Sakib, Z. Ferdous and M. A. Taher, "Automatic Lighting and Security System Design Using PIR Motion Sensor," *JU Journal of Information Technology (JIT)*, Vol. 4, pp. 15-18, July 2015.
- [2] A. R. M. Siddique, S. N. Sakib, S. M. Tahsien and M. S. Kaiser, "Performance evaluation of a microcontroller-based solar powered auto-rickshaw," *International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT)*, Dhaka, Bangladesh, pp. 1-6, 10-12 April, 2014.
- [3] S. N. Sakib, S. P. Mouri, Z. Ferdous and M. S. Kaiser, "A study on low cost solar powered wheel chair for disabled people of Bangladesh," *18th International Conference on Computer and Information Technology (ICCIIT)*, Dhaka, Bangladesh, pp. 27-31, 21-23 Dec, 2015.
- [4] Standard Solar Spectra, ASTM G-173-03 (International standard ISO 9845-1, 1992), <https://www.pveducation.org/pvcdrom/appendices/standard-solar-spectra>.
- [5] D.M. Chapin, C. S. Fuller and G. L. Pearson, "A New Silicon p-n Junction Photocell for Converting Solar Radiation into Electrical Power," *Journal of Applied Physics* 25, 676.
- [6] D. C. Reynolds, G. Leies, L. L. Antes and R. E. Marburger, "Photovoltaic Effect in Cadmium Sulfide," *Phys. Rev.* 95, 533.
- [7] D. Trivich and P. A. Flinn, "Maximum efficiency of solar energy conversion by quantum processes," in *Solar Energy Research*, F. Daniels, J. A. Duffie, Eds. Madison, WI: Univ. of Wisconsin Press, 1955, p. 143.
- [8] E. D. Jackson, "Areas for improvement of the semiconductor solar energy converter," in *Trans. Intern. Conf. Use of Solar Energy The Scientific Basis*, vol. V, p. 122, 1955.
- [9] William Shockley and Hans J. Queisser, "Detailed Balance Limit of Efficiency of p-n Junction Solar," *Journal of Applied Physics*, 32, 1961, 510-519.
- [10] E. M. Nell and A. M. Barnett, "The spectral p-n junction model for tandem solar-cell design," *IEEE Transaction on Electron Devices*, Ed-34, No. 2, 1987, 257-266.
- [11] O. Shekoofa, J. Wang, "Multi-diode modeling of multi-junction solar cells," *23rd Iranian Conference on Electrical Engineering*, pp. 1164-1168, Tehran, Iran, 10-14 May, 2015.
- [12] F. Nagamine, R. Shimokawa, M. Suzuki, T. Abe, "New solar simulator for multi-junction solar cell measurements," *23rd IEEE Photovoltaic Specialists Conference*, pp. 686-690, Louisville, KY, 10-14 May, 1993.
- [13] H. Cotal, J. Frost, "Heat transfer modeling of concentrator multijunction solar cell assemblies using finite difference techniques," *35th IEEE Photovoltaic Specialists Conference (PVSC)*, pp. 000213 – 000218, Honolulu, HI, 20-25 June 2010.
- [14] J. R. Wilcox, A. W. Haas, J. L. Gray, R. J. Schwartz, "A solar cell model for use in optical modeling of concentrating multijunction photovoltaic systems," *35th IEEE Photovoltaic Specialists Conference (PVSC)*, pp. 003104 – 003108, Honolulu, HI, 20-25 June 2010.
- [15] John F. Geisz, Myles A. Steiner, Iván García, Ryan M. France, William E. McMahon, Carl R. Osterwald, Daniel J. Friedman, "Generalized Optoelectronic Model of Series-Connected Multijunction Solar Cells," *IEEE Journal of Photovoltaics*, Vol: 5, Issue: 6, pp. 1827 – 1839, 2015.
- [16] Physical properties of Gallium Arsenide, <http://www.ioffe.ru/SVA/NSM/Semicond/GaAs/basic.html>.
- [17] Physical and electrical properties of Ge, <http://www.ioffe.ru/SVA/NSM/Semicond/Ge/basic.html>.
- [18] PROPERTIES OF THE III-V COMPOUND SEMICONDUCTORS, <http://www.semiconductors.co.uk/propiii5653.html>.
- [19] S. N. Sakib, S. P. Mouri, Z. Ferdous, A. Kowsar, M. S. Kaiser, "Effect of Different Solar Radiation on the Efficiency of *GaInP2/GaAs/Ge* based Multi Junction Solar Cell," *EICT 2015*, 10-12 December 2015, Khulna, Bangladesh.
- [20] Andrew Bates, Sherif Michael, *The Design and Optimization of an Advanced Four Junction Solar Cell*, Naval Postgraduate School, Monterey, California, 93955.
- [21] R. Hulstrom, R. Bird and C. Riordan, *Spectral solar irradiance datasets for selected terrestrial condition*, solar cells 15(1985) 365 391.
- [22] A. Kowsar, M. S. Islam, K. R. Mehzabeen and Z. H. Mahmood, *Study on the Efficiency of the GaInP2/GaAs/Ge Multijunction Solar Cell*, *Proc. of International Conference on Environmental Aspects of Bangladesh (ICEAB10)*, Japan, Sept. 2010.
- [23] S. R. Kurtz, P. Faine, J. M. Olson, *Modeling of two-junction, series-connected tandem solar cells using top-cell thickness as an adjustable parameter*, *Journal of Applied Physics*, 68 (4), 1990, 1890 1895.
- [24] ASTM G173-03 Reference Spectra Derived from SMARTS v. 2.9.2.
- [25] ASTM Standard G173, *Standard Tables for Reference Solar Spectral Irradiance: Direct Normal and Hemispherical on 37° Tilted Surface*, Amer. Society for Testing Matls. West Conshohocken PA, USA 2007.
- [26] K. W. Mitchell, *High efficiency concentrator solar cells*, *Conf. of 15th IEEE Photovoltaic specialists con.*, pp. 142-146, 1981.
- [27] Abu Kowsar, Abdullah Yousuf Imam, Mashudur Rahaman, Muhammad Shahriar Bashar, Md. Saidul Islam, Sumona Islam, Nowrin Akter Surovi, Zahid Hasan Mahmood, *IOSR Journal of Applied Physics (IOSR-JAP) Issue 6 Ver. IV (Nov.-Dec. 2014)*, PP 13-17.
- [28] Hamza Bennacer, Smail Berrah, Abdelkader Boukourt, Mohamed Issam Ziane *Electronic and optical properties of GaInX2 (X= As, P) from first principles study*, *Indian Journal of pure & Applied Physics*, Vol. 53, March 2015, pp. 181-189.
- [29] Syeda Puspita Mouri, Syed Nazmus Sakib, Zannatul Ferdous, Mohammad Ziaur Rahman, M. Shamim Kaiser, "Comparison of Theoretical Efficiencies of Three and Four Layer Multijunction Solar Cell and the Effects of Varying the Bottom Layer Material", *ICIEV 2016*, 13-15 May 2016, Dhaka, Bangladesh.

Throughput Analysis of Narrowband Cognitive Radio Networks For Fading Channel

Sk. Shariful Alam, Juthika Chowdhury, M. R. Rashed, Sariha Anjuman, Rokhsana Zaman

Department of Electronics & Communication Engineering

Khulna University of Engineering & Technology

Khulna, Bangladesh

Emails: {ssa1996, rashed413}@gmail.com, {juthika_chowdhury, sariha.oishi, roksana_zaman}@yahoo.com

Abstract— At present, spectrum scarcity is a major concern in the field of mobile communication. The necessity of spectrum is increasing gradually because the usage of wireless communication is rising tremendously. Cognitive radio (CR) is capable of dealing with this problem. This is referred to as an intelligent and reconfigurable radio which enables efficient usage of unused spectrum while avoiding any kind of interference. There are two groups of users in cognitive radio network. Primary users (PU) are the authorized users of the frequency band, the secondary users (SU) make use of those bands while primary users don't use those bands. So the unlicensed users i.e secondary users must sense the spectrum holes which are vacate as quickly as possible for conserving this spectrum reuse consequence and if it is detected that the channel is occupied by primary users again, it is obvious that the secondary users should vacate the channel in between a definite amount of time. Hence spectrum sensing plays a vital role in cognitive radio technology. Again, enhancing throughput is one of the fundamental issue in this field. If the sensing time large then CR gets less time to transmit data. Thus, throughput will degrades and if sensing time is low throughput will be high. In our paper, we try to improve throughput for different fading channel in narrowband with the help of a proposed model.

Keywords—Cognitive radio, spectrum reuse, energy detection, sensing time, achievable throughput.

I. INTRODUCTION

Cognitive Radio is a latest method of wireless communication which is a solution to spectrum scarcity. To sense vacant band efficiently and by using those band to attain communication without creating any disturbance to the licensed users is the main focus of this technology. In this arrangement, unlicensed or secondary users can manipulate unused spectrum bands when licensed users of those band are not active. So the spectrum is used properly [1]. To ensure efficient usage of spectrum, spectrum management scheme is used. However each of the secondary user uses one frame which is consists of two part: one part is for sensing and one part for transmitting data. If the sensing duration is small, then the duration for transmitting data will be large. Though the longer sensing time lessens the time for data transmission but it improves sensing capability for a particular frame length. Therefore it is demandable to sort out an absolute sensing time that improves achievable rate while protecting primary users

from interference. When any empty spectrum is occupied by cognitive radio, primary users can reappear. At that time cognitive radio must be capable of recognizing the existence licensed users as soon as possible and move out of the band. Consequently, sensing procedure should be selected in such that the existence of primary users can be detected in between a definite amount of time. The capability of sensing algorithm is limited by the needs of this prerequisite. It creates a challenge also for cognitive radio implementation. A tradeoff between the speed (sensing time) and constancy of sensing is controlled by the preference of sensing specification. The secular individuals of primary users in the circumstances and capacity of secondary users effect the best selected value [2]. However the time of sensing effects the ability of secondary users which is analyzed in [3]. Still searching an accessible band and monitoring a used band are done by optimum sensing period. It is preferable to improve the throughput of secondary users and securing primary users at the same time. The valid radio resources can be used properly and effectively by reducing the duration and periodicity of spectrum sensing phase. The time required to sense is represented in such a manner that the throughput will maximum for a given frame duration without creating interference for PU which is done by maintaining a certain threshold lower than the detection probability in [4]. Such kind of situation has been investigated in [4-5] where for a particular sensing time frame length is optimized. In [5], the limitation was on the possibility of interference with the licensed users and the throughput of secondary users was developed in the meanwhile securing the primary users. The remainder of the paper deals with the following section. In section II spectrum sensing is discussed. Section III describes fading model, section IV deals with proposed model for enhancing throughput, sensing throughput is focused on the section V. Simulated result is shown in section VI and last of all conclusion is shown in section VII.

II. SPECTRUM SENSING

The most challenging issue in cognitive radio network is to detect that whether a primary user is active or not. If the secondary user finds a vacant spectrum it want to access the spectrum as early as possible to avoid interference with the primary user. This is referred to as spectrum sensing. There are three parameters which are involved with spectrum sensing: probability of detection, P_d , probability of false alarm, P_f and

probability of miss detection, P_m . When PU is available in the channel and SU detects the possibility of existence of PU it is called detection probability. On the contrary, the probability that SU detects the presence of PU but literally there is no PU is known as false alarm probability. When PU exists in the band but SU fails to identify the existence of PU it is called probability of miss detection. Hence there prevails a theoretical interconnection that links performance of sensing and throughput of cognitive radio network. If we specify sensing time then we can achieve this goal. If the sensing time is small then the throughput will be greater. By considering the restriction of detection probability we need to minimize the false alarm probability to develop the sensing-throughput tradeoff problem. If the primary user re-appear and detect that the band is busy due to the secondary user, then secondary user looks for an out-band to continue the transmission without interrupting the PU. Thus the average search time is minimizes while aiming huge feasibility that the SU can be able to discover at least a vacant out-band channel [6]. Since in CR network technology spectrum sensing is a vital task, there are various technique for signal detection as energy, cyclostationary, matched filter based detectors. Among these detection scheme energy detection (ED) is the simplest as it doesn't require any prior knowledge about primary users.

A. Energy Detector

Energy detection is the fundamental technique of spectrum sensing. It is one of the popular method for spectrum sensing considering its low computational complexity and lower sensing time. Energy detection is chosen here because of its simplicity. In this scheme of spectrum sensing, computation is done for energy of received signal. A threshold value is predefined. Then the calculated energy is compared with that threshold value. If the predefined threshold value is lower than the calculated energy it is assumed that primary user is present there and if the threshold value is higher it is considered as a vacant channel. A simple block diagram of energy detector is shown in Fig. 1.

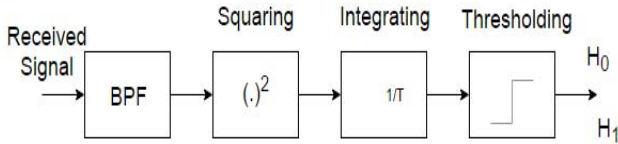


Fig. 1. Block diagram of a typical energy detector

In case of energy detection two hypothesis is considered: When the primary user is inactive then

$$y(n) = w(n) \quad (1)$$

This is named as hypothesis H_0 . When the primary user is active then it can be written as:

$$y(n) = x(n) + w(n) \quad (2)$$

Which is known as hypothesis H_1 .

If total number of sample is N energy can be calculated as follows:

$$S(y) = \frac{1}{N} \sum_{n=1}^N |y(n)|^2 \quad (3)$$

The test constant $S(y)$ is a random variable whose probability density function (PDF), $p_0(x)$ is a Chi-square distribution under hypothesis H_0 . Let assume detection threshold is ϵ then the probability of false alarm can be written as

$$P_f(\epsilon, \tau) = P_r(S(y) > \epsilon | H_0) = \int_{\epsilon}^{\infty} p_0(x) dx \quad (4)$$

Again we consider another hypothesis H_1 for test static $S(y)$ if PDF is $p_1(x)$ then probability of detection can be written as

$$P_d(\epsilon, \tau) = P_r(S(y) > \epsilon | H_1) = \int_{\epsilon}^{\infty} p_1(x) dx \quad (5)$$

III. FADING MODEL

In this paper we have done our simulation for four types of fading channel-Additive White Gaussian Noise (AWGN) channel, Rayleigh channel, Rician channel and Nakagami channel.

A. AWGN channel

This is the basic noise model which is not realistic. In this type of model, the channel noise is considered to be Gaussian, additive and white in nature. The AWGN channel provides the maximum bit error and the approach is figured out to give accuracy in AWGN channel is assumed to yield perfect outcome in other realistic channels when comparing with other equivalent channels. But the real performance may vary. The AWGN channel is considered as an acceptable model for different application as many satellite and deep space communication.

B. Rayleigh channel

Rayleigh fading model is another prominent model. It is introduced where signal is received from multipath. It is a satisfactory model for ionosphere and troposphere's signal propagation. It is also applicable in urban area. Rayleigh fading is most appropriate if line of sight (LoS) is absent between transmitter and receiver. If the signals arrive from multiple direction, the central limit theorem is used to imitate the time-variant impulse response of the channel as a complex-valued Gaussian random process. When the impulse response is modeled as a zero mean complex-valued Gaussian process, the channel is said to be a Rayleigh fading channel [7].

C. Rician channel

Rician fading is close to Rayleigh fading in many respect, but it possess a strong dominant component. Usually the dominant component is LoS component. This is modeled by using two Gaussian random variables – one with zero mean and another with non-zero mean [7]. This fading is reliable for satellite communication.

D. Nakagami channel

There are various types of fading model which are proposed to distinguish the fading envelope for multipath fading in the sphere of wireless communication. The Nakagami- m distribution is an important model among those analytical model. It is applicable for empirical fading data. It is used to model signal for excessive to temperate fading case by properly setting the value of Nakagami parameter, m .

IV. PROPOSED MODEL

In this section, we proposed a system model which is shown in Fig. 2 for improving throughput of CR networks. Let consider a narrowband signal, $x_c(t)$ which bandwidth is w Hz. One part of its bandwidth is reserved for primary user of a communication system and other portion is free for CR to access opportunistically. Let assume the secondary user has equipped with k number of BPFs which are identical and the output of BPFs are denoted by x_k . The bandwidth of each BPF will be $w_k = w/k$ Hz where $k=0, 1, 2, \dots, K$. Then the average energy E_k of each BPF is calculated and compare those average energies E_k ($k=1, 2, 3, 4, 5, \dots, K$) by means of energy estimate and compare section. In this proposed model, the block which is used for energy computation and comparison, has a great significance. An important fact is recommended here that no of sample which is considered has effect on the accuracy of energy comparison. The result of energy comparison would be more accurate if we increase the no of sample.

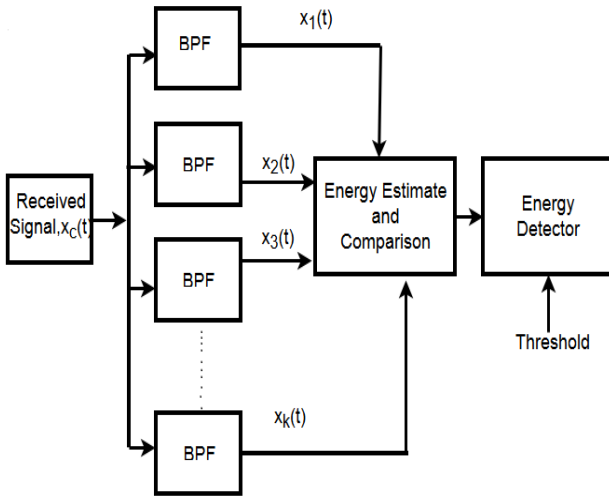


Fig. 2. Proposed model for improving throughput [8].

The comparator of the block also put back the samples related to that BPF which contains lower average energy, during the comparison of average energy of a group of samples corresponding to the bandwidth. We are interested to find out that BPF which contains minimum energy because minimum energy indicates that there is minimum number of PU in that BPF. Then, we applied ED scheme to find a vacant channel for SU. By using the outcome of energy detector secondary users determine whether the channel is vacant or not, the decision of spectrum sensing evaluating the sub-channel of consideration.

V. SENSING THROUGHPUT

Throughput is a computation of how many units of information a system can process in a given amount of time. To determine the achievable throughput for CR network let consider Fig. 3, a simple problem where PU is absent and there is no false alarm caused by CR. Let, the total frame length is T , the reserved time slot for sensing is τ and the data transmission slot is $(T-\tau)$.

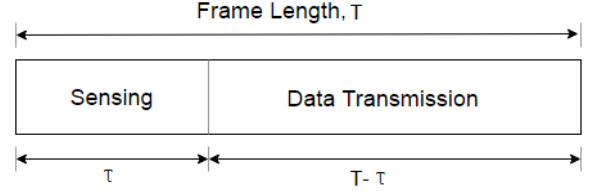


Fig. 3. Formation of a frame structure of cognitive radio.

Considering primary user is absent, the achievable rate of a secondary user is denoted by C_0 where $C_0 = \log_2(1 + SNR_s)$; SNR_s is the signal to noise ratio of the CR link. Let assume a specific band where data transmission by PU is off and denoted by $P(H_0)$. Then the optimal achievable rate can be found by following equation [8]:

$$R(\tau) = C_0 P(H_0) \left(1 - \frac{\tau}{T}\right) (1 - Q(\alpha + \sqrt{N}\gamma)) \quad (6)$$

Where, $\alpha = \sqrt{2\gamma + 1} Q^{-1}(P_d)$ and $Q(x)$ is a monotonically decreasing function. From (6) it is clear that with the variation of sensing slot and frame duration the throughput of a CR user also fluctuates. If the sensing time τ is short for any fixed given frame length T the throughput will large. But if a CR node spends much time to detect a vacant spectrum band then it will get comparatively less time for data transmission. Thus throughput will degrades. If the throughput of SU is low it affects the performance of SU for instance, extension of transmission time and increasing the energy dissipation. So it is necessary to find out a trade-off between the sensing period and frame duration.

VI. SIMULATION RESULT

In this segment, computer simulation results are presented to estimate the achievable rate by using energy detection technique for AWGN, Rayleigh, Rician and Nakagami channel. For our proposed method we have used four BPFs. We have done this simulation for three different SNR i.e 0dB, 5dB and 15dB. It is assumed that secondary user wants to transmit data while primary user is not active. We choose it as, $P(H_0)=0.9$ when a CR node wants to transmit data. To achieve desired throughput for a particular frame duration, (6) is applied where frame length is 100ms. Fig. 4 shows the simulated result of sensing time vs achievable throughput for AWGN channel. These throughput are calculated using real

value of those time which were required for sensing for the corresponding channel.

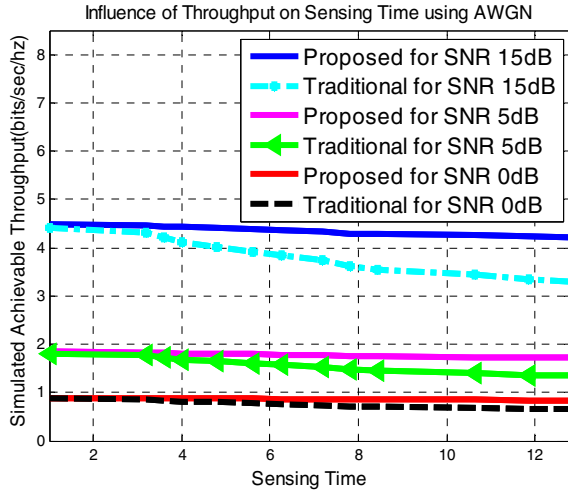


Fig. 4. Simulated achievable throughput result against sensing time for AWGN channel

From this figure it is seen that for 15dB at the initial moment the throughput was highest and it is 4.4bits/Sec/Hz. The throughput was almost same for sensing time 0ms to 3.5 ms for both proposed and traditional model. After 3.5ms the throughput is decreasing slowly. But we can see that throughput for traditional model was decreasing more rapidly than our proposed model because CR receiver is accommodated with three BPFs. But when we consider comparatively low SNR i.e for 5dB and 0dB the throughput is decreasing very slowly for proposed method. But for low SNR traditional throughput decrease more than proposed. Fig. 5 shows the simulation result which is done by using Rayleigh channel with addition to AWGN for three SNR value: 0dB, 5 dB, 15dB. The frame length is same also here.

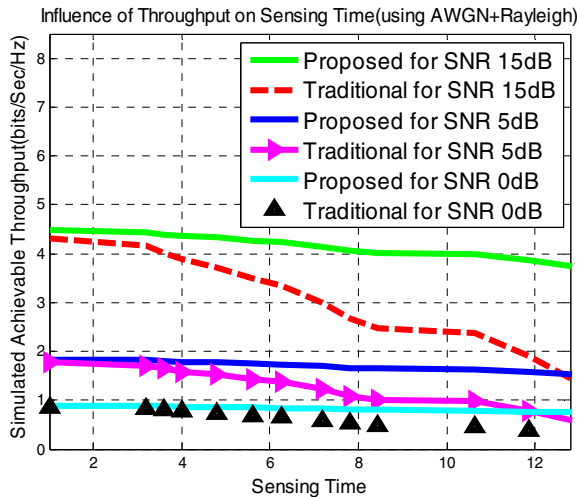


Fig. 5. Simulated achievable throughput result against sensing time for Rayleigh channel.

For Rayleigh fading channel the throughput is decreasing with the increase of sensing time rapidly for all SNR values.

Here also the performance of proposed model is better than traditional one. The throughput for traditional model is decreasing more harshly than proposed model. At initial the throughput were 4.4bits/Sec/Hz, 1.8bits/Sec/Hz and 0.9 bits/Sec/Hz for 15dB, 5dB and 0dB respectively. The throughput was highest and almost linear when the sensing time was below 3.1ms and was 3.7bits/Sec/Hz when sensing time exceeds 12ms. When SNR value is low then the throughput degradation is also low. By comparing Fig. 4 and Fig. 5 it is observed that the performance of Rayleigh channel is comparatively poor than AWGN. Fig. 6 and Fig. 7 shows the simulation result for the combination of AWGN and Rician and for the combination of AWGN and Nakagami channel respectively under the SNR value of 0dB, 5dB and 15dB for the fixed frame length 100ms.

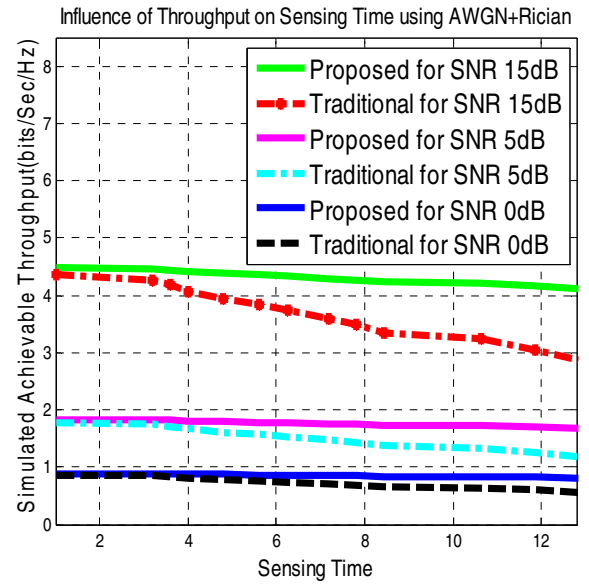


Fig. 6. Simulated achievable throughput result against sensing time for Rician channel

In Fig. 6 achievable throughput under Rician channel is observed for different values of SNR. For 15dB initially the throughput was 4.4bits/Sec/Hz and it was remain same till the sensing time was up to 3.2ms. After 3.2ms the throughput was decreasing gradually. The throughput was 4bits/Sec/Hz when sensing time exceeds 12ms. Initially throughput for both model was same. But it is seen from the Fig. 6 that throughput of proposed model decreases slowly than traditional. For traditional model throughput is decreasing more roughly. For 5dB throughput was 1.9bits/Sec/Hz and decreases slowly than 15dB. But the performance of traditional model is also low than our proposed model. For 0dB the throughput for both proposed and traditional model is almost same. As we have done simulation using energy detection so in case of low SNR simulation result is not so good. The performance of throughput under Rician channel is better than Rayleigh channel as seen from Fig. 5.

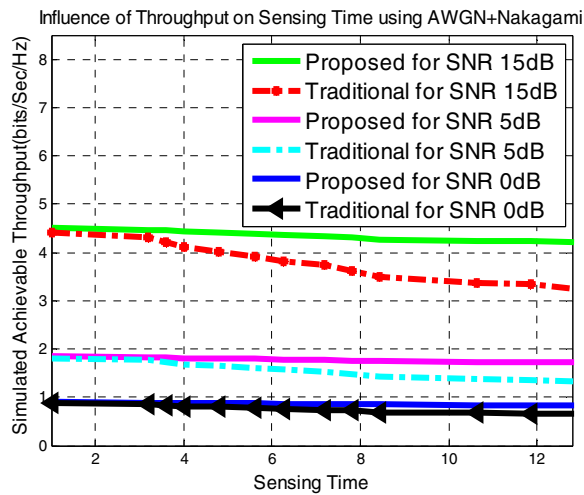


Fig. 7. Simulated achievable throughput result against sensing time for Nakagami channel

It is seen from the Fig. 7 that the achievable throughput is 4.5bits/Sec/Hz, 1.98bits/Sec/Hz and 0.99bits/Sec/Hz for SNR value 15dB, 5dB and 0dB respectively for proposed model when analyzing these simulation result it is observed that the proposed model has less computational complexity than traditional one as it accommodated with a number of BPFs. Thus it improves throughput of a CR node.

VII. CONCLUSION

In the field of wireless communication spectrum is very precious resource. But with the rapid growth of communication spectrum scarcity becomes a problem now a days. To provide efficient utilization of spectrum cognitive radio offers a way for tactfully utilizing the unused spectrum hole. The time required for spectrum sensing is an elementary feature of cognitive radio technology. Because the data rate of cognitive radio mainly depends on sensing time. By varying sensing time we can also vary transmission rate. The throughput of cognitive radio

differs with the sensing time additionally frame length. As a consequence when sensing time is small throughput of secondary user is large and vice versa. So we should try to find out the perfect sensing period to achieve desired throughput for cognitive radio. This paper presented a system model for maximizing throughput which is verified in our simulation result.

REFERENCES

- [1] Y.-C. Liang, Y. Zeng, E. Peh, A. T. Hoang, "Sensing-throughput tradeoff for cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 4, pp. 1326-1337, Apr. 2008.
- [2] S. D. Jones, E. Jung, X. Liu, N. Merheb, I.J. Wang "Characterization of spectrum activities in the U.S. public safety band for opportunistic spectrum access," *IEEE Int. Symposium on New Frontiers in Dynamic Spectrum Access Networks*, Dublin, Ireland, 2007, pp. 137-146.
- [3] A. Ghasemi, E. Sousa, "Optimization of spectrum sensing for posium on New Frontiers in Dynamic Spectrum Access Networks, opportunistic spectrum access in cognitive radio networks," *IEEE Consumer Commun. and Networking Conf.*, LasVegas, Nevada.
- [4] Yuhua Xu, Jinlong Wang and Qihui Wu, "Interference-throughput tradeoff in dynamic spectrum access: Analysis based on discrete-time queuing subjected to bursty preemption," *2009 4th International Conference on Cognitive Radio Oriented Wireless Networks and Communications*, Hannover, 2009, pp. 1-6.
- [5] Y. Pei, A. T. Hoang and Y.C. Liang, "Sensing-throughput tradeoff in cognitive radio networks: How frequently should spectrum sensing be carried out?," *2007 IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications*, Athens, 2007, pp. 1-5. doi: 10.1109/PIMRC.2007.4394632
- [6] Hyoil Kim and Kang G. Shin. 2008. In-band Spectrum Sensing in Cognitive Radio Networks: Energy Detection or Feature Detection? *Proceedings of the 14th ACM international conference on Mobile computing and networking (MobiCom '08)*. ACM, New York, NY, USA, 14-25.
- [7] G. A. Bhalerao, R.G. Zope, "BER Improvement of DSCDMA with Rake Receiver Using Multipath Fading Channel" *International Conference on Recent Trends in engineering & Technology – 2013 (ICRTET'2013)*.
- [8] S. S. Alam, L. Marcenaro and C. S. Regazzoni, "Improved Throughput Performance in Wideband Cognitive Radios via Compressive Sensing," *Modelling and Simulation (EUROSIM), 2013 8th EUROSIM Congress on*, Cardiff, 2013, pp.585-590. doi: 10.1109/EUROSIM.2013.102

Exploring the Opportunity of Using Graphene as the Transparent Conducting Layer in CZTS-based Thin Film Solar Cells

¹Kashfia Naz Nikita, ²Md. Abdul Gaffar, ³Md. Iqbal Bahar Chowdhury

Department of Electrical and Electronics Engineering (EEE)

United International University, Dhaka, Bangladesh

Email: ¹kashfia.nikki@gmail.com, ²m.a.gaffar012@gmail.com, ³ibchy@eee.uiu.ac.bd

Abstract— This work explores the opportunity of the use of graphene as the transparent conducting (TC) layer instead of conventional transparent conducting oxide (TCO) such as indium tin oxide (ITO) in the CZTS based solar cells which is recently evolved as one of the most promising thin film solar cells. It has been known that graphene has higher transmittance and conductance and is cheaper than the widely used ITO. Based on the MATLAB simulations, this work proves that graphene can be used as the better alternative than ITO as the transparent conducting layer. Moreover, simulations show that the use of graphene can reduce the amount of CZTS material and hence, reduce the cost of solar cell as optimum absorber width based on maximum possible conversion efficiency is found less for the graphene-TCL based solar cells.

Keywords—Graphene, Indium Tin Oxide, CZTS solar cell, Transparent Conducting Layer, Optimum Absorber Width.

I. INTRODUCTION

Although thin film solar cells usually made up of multiple thin layer of photovoltaic materials are much flexible, light and cheap, their efficiency is less than that of the conventional crystalline silicon solar cells. To improve the efficiency keeping the cost as low as possible many researches are ongoing. In this context, Copper Zinc Tin Sulfide (CZTS) based solar cells have been evolved as the most promising thin film solar cells. CZTS has a direct band gap of about 1.4-1.5eV [1] and hence, can absorb a higher quantity of incident light in the first few micrometres. Being made from abundant materials: copper, zinc, tin and sulphur, it is cheaper than the widely used Cadmium-Telluride (CdTe) and Copper-Indium-Gallium-Selenide (CIGS) absorbers. It has no toxicity problem as it is free from toxic cadmium and selenium. Recently, a team led by Dr Xiaojing Hao of the 'Australian Centre for Advanced Photovoltaics' at the UNSW School of Photovoltaic and Renewable Energy Engineering achieved an efficiency beyond 20% for a full-sized CZTS based solar cell [2]. For emitter layer Zinc Oxide (ZnO), Zinc Sulphide (ZnS) and Cadmium Sulphide (CdS) can be chosen in CZTS based solar cells. However, owing to the highest Bandgap ZnO has higher efficiency than the other two. Moreover, ZnO is more environment-friendly as ZnS contains environmentally harmful sulphur and CdS causes environmental hazards because of the presence of cadmium. Finally, ZnO is

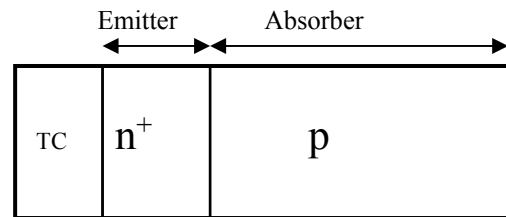


Fig.1: Solar cell with TC Layer

structurally compatible with CZTS and its high absorption coefficient and transmittance eases absorption of the photons [3].

As seen in Fig. (1), solar cells generally use a transparent conducting (TC) layer at the front end which functions as a metal contact as well as allows light to get through. In order to obtain higher conversion efficiency, this TC layer becomes an integral part of thin film fabrication and hence, the appropriate choice of TC layer is very important for thin film solar technology. A good TC layer should fulfil some requirements such as (i) highly conductive (ii) sheet resistance of less than 10 Ω /sq. (iii) transmittance of greater than 90% (iv) The refractive index should be high (typically more than 3) [4]. In the literature, a number of transparent conducting oxides are proposed as TC layer such as tin oxide (SnO_2), Indium Tin Oxide (ITO), Fluorine doped Tin Oxide (FTO), boron doped zinc oxide (ZnO:B), aluminium doped zinc oxide (ZnO:Al) etc. SnO_2 is not stable in saline plasma and cannot be used as a front contact in superstrate configuration. FTO devices cause current leakage due to its structural defects [5]. Since ITO fulfils almost all the requirements of TC, it is most widely used. However, the cost of ITO increases rapidly as the indium is not abundant on earth and the fabrication process becomes costlier for ITO [6]. Therefore, new materials need to be explored to achieve lower cost as well as higher efficiency. Some potential replacement materials include thin film metals [7], metal oxides [8], and metal grids [9]. Low sheet resistance and high optical efficiency are the fundamental requirement for these electrodes. ZnO and graphene are the other alternatives as TC material. But graphene has more optical transparency compared to ZnO. It satisfies almost all the requirements of TC layer such as high transmittance, a refractive index of 2.37 [10]

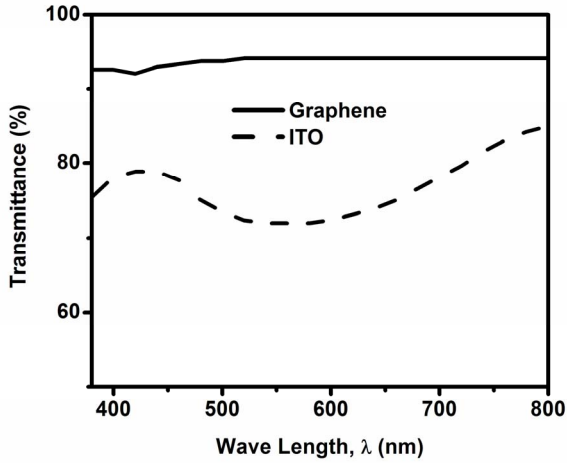


Fig.2: Transmittance data for graphene and ITO

Graphene is a two dimensional material made of a single layer of carbon atoms which are bonded together in a honeycomb pattern. It is relatively less expensive as it is made of abundant carbon. Some of its extraordinary characteristics are- it is much conductive, flexible, almost entirely transparent and extremely strong. In the visible and near infrared region graphene has a uniformly high transparency due to which it can be utilized to form ultra-thin transparent electrodes [11]. Graphene is a better charge collector and more transparent than other materials used in TC layer. Regular graphene's band gap is zero [12] and thus, it is considered as semi-metal material. Compared to ITO, graphene shows higher transmittance in the visible range as seen from Fig. (2) extracted from [13] and hence, graphene can transmit more light to emitter and more electrons will be produced.

In this work a graphene-TC layer based CZTS solar cell structure is proposed. The performance of proposed solar cell is investigated with ITO-TC layer based CZTS solar cells. As an emitter layer of ZnO and CdS materials are used, therefore, four CZTS based solar cells are investigated. Performance of these solar cells are compared based on current-voltage (J-V) characteristics curves and also, on various performance parameters which include short circuit current density, open circuit voltage, fill factor, conversion efficiency and optimum absorber width.

II. ANALYSIS

In this section, an analytical model has been developed to investigate the CZTS solar cell performance. As shown in Fig. (3), the doping levels of the emitter (heavily doped, of the order of 10^{17} cm^{-3}) and absorber layers (lightly doped, of the order of 10^{14} cm^{-3}) are chosen so that the absorber layer can be considered as fully depleted and the electric field in the absorber can be considered as uniform. The expression for this uniform field can be expressed as [14]

$$E = \frac{\epsilon_1 N_d}{\epsilon_1 N_d + \epsilon_2 N_a} \left[\frac{V_{bi} - (V_L - J_L R_s)}{W_a} \right] \quad (1)$$

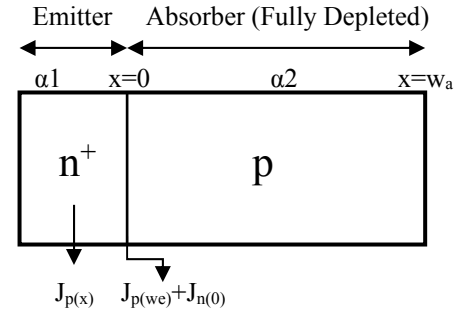


Fig. 3: Fully depleted solar cell

Where, ϵ_1 (ϵ_2) is the permittivity and N_d (N_a) is the doping level of the emitter (absorber) layer, R_s is the series resistance and V_{bi} is the built in voltage across the absorber layer, V_L and J_L are the load voltage and load current respectively.

Using the expression of electric field of Eqn.(1), the current and its continuity equation can be solved to determine the photo generated current. For absorber layer, these equations can be written as

$$J_n(x) = qn(x)\mu_n E + qD_n \frac{dn}{dx} \quad (2)$$

$$\frac{dJ_n}{dx} = \frac{qn(x)}{\tau_n} - q\alpha_2 \frac{\lambda I_0}{hc} T e^{-\alpha_2 x} e^{-\alpha_2 w_e} \quad (3)$$

Where, $n(x)$, $J_n(x)$ represent electron and electron current density respectively, μ_n (D_n) represents electron mobility (diffusivity), α_1 and α_2 represent the wavelength dependent absorption coefficients of emitter and absorber material respectively, I_0 is the sum intensity, T is the transmittance of the TC layer, q is the electronic charge, h is the Planck's constant, λ is the wavelength of incident light and c is the velocity of the light. In order to include the effect of surface recombination at the absorber edge, Eqn. (2) includes the diffusion term ($qD_n * dn/dx$), which is generally neglected in conventional models. Now combining Eqns. (2) and (3), a second order differential equation can be obtained as

$$\frac{d^2 J_n}{dx^2} + C \frac{dJ_n}{dx} - \frac{n}{L_n^2} = -G e^{-\alpha_2 x} \quad (4)$$

Where,

$$C = \frac{E}{V_T}, \quad G = \frac{G_0 \alpha_2}{D_n} e^{-\alpha_1 w_e}, \quad G_0 = T \frac{\lambda I_0}{hc}, \quad L_n = \sqrt{D_n \tau_n}$$

The required boundary conditions for this differential equation are given as

$$\left. \frac{dn}{dx} \right|_{x=0} = 0, \quad J_n(w_a) = qS_n n(w_a)$$

where the first one comes from the assumption that at $x = 0$

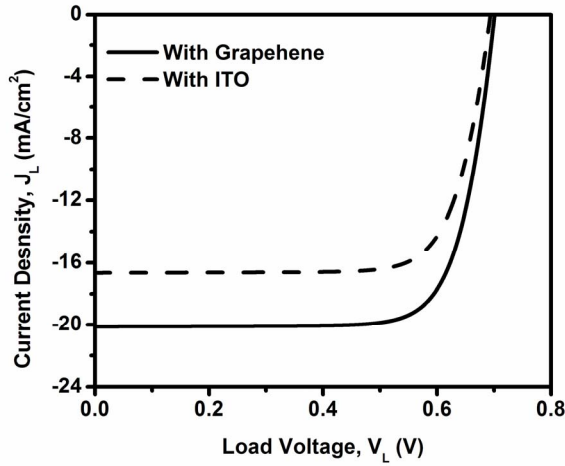


Fig. 4: J-V characteristic curves of the CdS/CZTS based solar cell using ITO and Graphene as different TC layer. Here $W_e = 0.1 \mu\text{m}$, $W_b = 3.5 \mu\text{m}$

TABLE 1: PARAMETERS OF THE CdS/CZTS BASED SOLAR CELL USING ITO AND GRAPHENE AS DIFFERENT TC LAYER. HERE $W_e = 0.1 \mu\text{m}$, $W_b = 3.5 \mu\text{m}$

Parameters	Graphene	ITO
$J_{sc} (\text{mA}/\text{cm}^2)$	20.1246	16.6610
$V_{oc} (\text{V})$	0.7012	0.6939
FF (%)	76.5592	76.3803
η (%)	10.8036	8.8304

the diffusion component of the current in the absorber region is negligible as compared with the drift component, since the electric field is maximum at $x = 0$. The second boundary condition comes from the surface recombination at the absorber edge. Using these boundary conditions, the photogenerated electron current at $x = 0$ can be solved as

$$J_n(0) = qD_n n(0)C$$

The total photo generated current at $x = 0$ and hence, for the solar cell can be obtained as

$$J_{ph} = J_p(w_e) + J_n(0)$$

Where $J_p(W_e)$ is the photogenerated hole current obtained from the solution of hole current in the emitter region [15]. Finally, the load current density J_L against the load voltage V_L expressed including effect of series resistance (R_s) following the procedure outlined in [16,17] as

$$J_L(V_L) = -J_{ph}(V_L) + J_{01} \left(e^{\frac{V_L - J_L R_s}{V_T}} - 1 \right) + J_{02} \left(e^{\frac{V_L - J_L R_s}{nV_T}} - 1 \right) + \frac{V_L - J_L R_s}{R_{sh}}$$

where, J_{01} , J_{02} are the dark saturation currents in the emitter and absorber layer respectively, n is the ideality factor, V_T is the thermal voltage and R_{sh} is the shunt resistance.

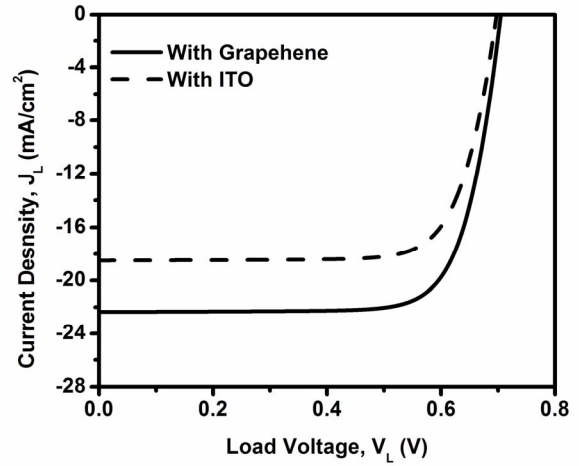


Fig. 5: J-V characteristic curves of the ZnO/CZTS based solar cell using ITO and Graphene as different TC layer. Here $W_e = 0.1 \mu\text{m}$, $W_b = 3.5 \mu\text{m}$

TABLE 2: PARAMETERS OF THE ZnO/CZTS BASED SOLAR CELL USING ITO AND GRAPHENE AS DIFFERENT TC LAYER. HERE $W_e = 0.1 \mu\text{m}$, $W_b = 3.5 \mu\text{m}$

Parameters	Graphene	ITO
$J_{sc} (\text{mA}/\text{cm}^2)$	22.3734	18.5033
$V_{oc} (\text{V})$	0.7051	0.6977
FF (%)	76.1436	75.9726
η (%)	12.0120	9.8079

III. RESULTS AND DISCUSSION

A number of simulations have been carried out for the CZTS based thin film solar cell with two different emitter layer materials (CdS and ZnO) and two different TC materials (ITO and graphene) using MATLAB. The results of those simulations have been presented and discussed in this section. From the ASTM G-173-03 standard, the air mass (AM) 1.5 global spectrum is obtained which is the incident photon flux [18]. The material parameters like bandgap, electron affinity, permittivity etc. as well as the absorption coefficients of CdS, ZnO and CZTS are obtained from [19].

Fig. (4) shows the current-voltage (J-V) characteristics of two CdS/CZTS based solar cells where one uses ITO and the other uses graphene as TC layer. Since graphene has higher transmittance than that of ITO over the visible range, more photons can be absorbed and hence, more photo-generated carriers are obtained in case of graphene TC layer. Therefore, graphene TC layer based CdS/CZTS solar cell is expected to show improved J-V characteristics as well as to show higher performance parameters such as short circuit current density

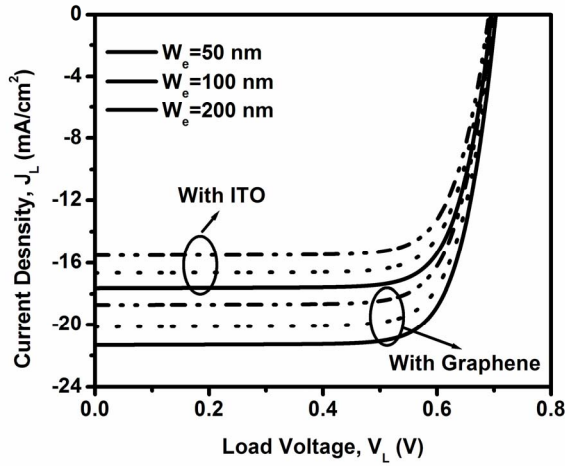


Fig. 6: J-V characteristic curves of the ZnO/CZTS based solar cell using two different TC layers (ITO and Graphene) for three different emitter width. Here $W_b = 3.5 \mu\text{m}$

TABLE 3: PARAMETERS OF THE ZnO/CZTS BASED SOLAR CELL USING GRAPHENE AS TC LAYER FOR THREE DIFFERENT EMITTER WIDTH. HERE $W_b = 3.5 \mu\text{m}$

Parameters	$W_e = 0.05 \mu\text{m}$	$W_e = 0.1 \mu\text{m}$	$W_e = 0.2 \mu\text{m}$
J_{sc} (mA/cm ²)	21.3234	20.1244	18.7246
V_{oc} (V)	0.7035	0.7012	0.6984
FF (%)	76.4715	76.3110	76.1870
η (%)	11.4715	10.7684	9.9632

TABLE 4: PARAMETERS OF THE ZnO/CZTS BASED SOLAR CELL USING ITO AS TC LAYER FOR THREE DIFFERENT EMITTER WIDTH. HERE $W_b = 3.5 \mu\text{m}$

Parameters	$W_e = 0.05 \mu\text{m}$	$W_e = 0.1 \mu\text{m}$	$W_e = 0.2 \mu\text{m}$
J_{sc} (mA/cm ²)	17.6452	16.6612	15.5188
V_{oc} (V)	0.6961	0.6939	0.6911
FF (%)	76.7417	76.5876	76.4728
η (%)	9.4261	8.8544	8.2018

(J_{sc}), open circuit voltage (V_{oc}), fill factor (FF) and conversion efficiency (η) than the solar cell based on ITO as TC layer. Fig. (4) and Table 1, which lists the performance parameters for the said solar cells, are in well agreement with this expectation.

Fig. (5) and Table 2 show the J-V characteristics and the performance parameters for graphene/ZnO/CZTS and ITO/ZnO/CZTS based solar cells respectively. Like graphene/CdS/CZTS based solar cell, these figure and table show the expected better J-V curve and higher performance parameter values for graphene/ZnO/CZTS based solar cell. Comparison of Figs. (4) and (5) as well as Tables 1 and 2 reveals that when ZnO is used as emitter layer instead of CdS, the J-V curve and the performance parameters become even better. This is due to the higher Bandgap of ZnO ($E_g = 3.3 \text{ eV}$) than that of CdS ($E_g = 2.4 \text{ eV}$) for which ZnO has higher absorption coefficients and hence, higher number of photogenerated carriers are obtained in case of ZnO-emitter layer. This results are consistent with those found in [20].

Fig. (6) shows the effect of variation of emitter width (W_e) on the J-V characteristics curves for ZnO/CZTS based solar

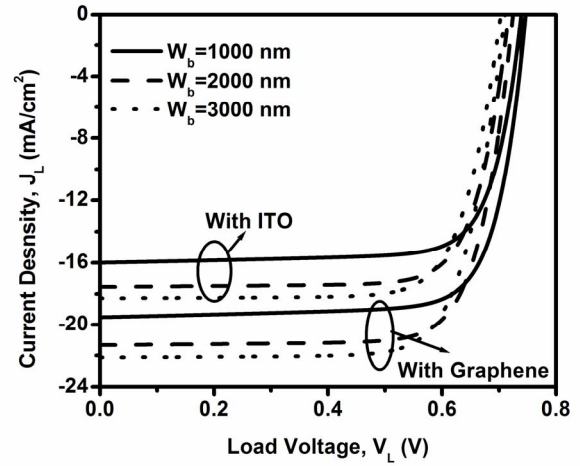


Fig. 7: J-V characteristic curves of the ZnO/CZTS based solar cell two different TC layers (ITO and Graphene) for three different absorber width. Here $W_e = 0.1 \mu\text{m}$

TABLE 5: PARAMETERS OF THE ZnO/CZTS BASED SOLAR CELL USING GRAPHENE AS TC LAYER FOR THREE DIFFERENT ABSORBER WIDTH. HERE $W_e = 0.1 \mu\text{m}$

Parameter	$W_b = 1 \mu\text{m}$	$W_b = 2 \mu\text{m}$	$W_b = 3 \mu\text{m}$
J_{sc} (mA/cm ²)	19.5176	21.3249	22.1287
V_{oc} (V)	0.7471	0.7247	0.7106
FF (%)	76.4427	76.3993	74.9324
η (%)	11.1466	11.8068	11.7828

TABLE 6: PARAMETERS OF THE ZnO/CZTS BASED SOLAR CELL USING ITO AS TC LAYER FOR THREE DIFFERENT ABSORBER WIDTH. HERE $W_e = 0.1 \mu\text{m}$

Parameters	$W_b = 1 \mu\text{m}$	$W_b = 2 \mu\text{m}$	$W_b = 3 \mu\text{m}$
J_{sc} (mA/cm ²)	16.0037	17.5685	18.2849
V_{oc} (V)	0.7393	0.7172	0.7032
FF (%)	76.3596	76.6238	75.4222
η (%)	9.0345	9.6547	9.6977

cells with TC layer used for one is graphene and for the other is ITO. In both cases, it is found that, increase in W_e results in more degraded J-V curves. Tables 3 and 4 list the performance parameters for these solar cells, which also shows that increase in W_e causes poor values for these parameters. This is due to the fact that when W_e increases, more number of photons will be absorbed in the emitter region causing a less number of photogeneration in the absorber region. A close observation from Fig. (6) reveals that J-V curves are more sensitive to increase in W_e in the case of graphene TC layer than the case when ITO is used as TC layer. This can be attributed to the higher transmittance of graphene than ITO, which causes higher number of photons to pass through the emitter layer and hence, to become more absorbed in the emitter layer.

Fig. (7) and Tables 5, 6 show the effect of absorber width (W_b) variation on the J-V curves and performance parameters for ZnO/CZTS based solar cells with ITO and graphene as different TC layers. Several observations can be made from these figure and tables. First, increase in W_b increases the

short circuit current density as more lights can be absorbed in the absorber (short-wavelength photons can be absorbed in large number). Second, open circuit voltage decreases as W_b increases. This is due to the fact that increase of W_b increases the probability of carrier recombinations in the absorber thereby causing the more loss of energy. The third observation is that both fill factor and conversion efficiency increases first, reach a peak and then decreases as W_b increases. This trend is expected as increase in W_b increases both the energy gain through absorption and the energy loss through recombination, which are not in equal rate. Therefore, an optimum absorber width can be found where the opposing and unequal rates of absorption and recombination results in the maximum conversion efficiency.

Based on the analysis in the previous paragraph, simulations have been carried out for “graphene / ZnO / CZTS”, “ITO / ZnO / CZTS”, “graphene / CdS / CZTS” and “ITO / CdS / CZTS” based solar cells to determine the optimum absorber width ($W_{b,opt}$) based on the maximum conversion efficiency. The results are plotted as conversion efficiency vs. absorber width in Fig. (8) and the corresponding $W_{b,opt}$ along with the performance parameters at this $W_{b,opt}$ are listed in Table 7. From Fig. (8) it is seen that all four solar cells peak the conversion efficiency for an absorber width close to 2 μm and the higher conversion efficiency is achieved if emitter layer is ZnO instead of CdS and if TC layer is graphene instead of ITO. Therefore, the highest conversion efficiency is obtained for “graphene / ZnO / CZTS” based solar cell and the lowest for “ITO / CdS / CZTS” based solar cell. An interesting finding is obtained from Table 7, which indicates that the lowest optimum absorber width is found for “graphene / ZnO / CZTS” based solar cell whereas the highest value for this optimum value is found for “ITO / CdS / CZTS” based solar cell. Indeed, this finding is significant, since this reveals that the best optimized solar cell based on “graphene / ZnO / CZTS” is economically becomes the cheapest owing to the need for less amount of material.

IV. CONCLUSION

In this work a simulation based analysis has been presented to explore the use of graphene as TC layer as an alternative of existing ITO in the CZTS based thin film solar cells. Simulations results found are very optimistic and significant. Apart from the fact that graphene is cheaper than the ITO, this work founds that using graphene as TC layer instead of ITO not only results in higher conversion efficiency along with other higher performance parameters such as short circuit current density, open circuit voltage and fill factor but also shows the need of less amount of absorber material to obtain this higher efficiency and henceforth, further reduces the cost of solar cell. Although this work is based on CZTS based thin film solar cells, this work can be a clue of using graphene as TC layer instead of ITO for other thin film solar cells.

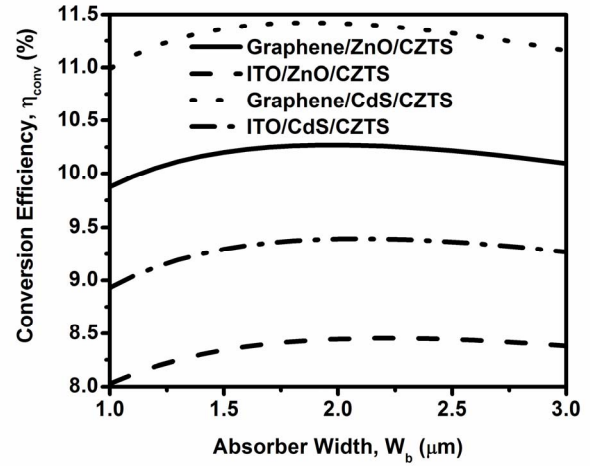


Fig 8: Conversion efficiency variation against the absorber width of the CZTS based solar cells for different TC/emitter combinations. Here $W_e = 0.1 \mu\text{m}$

TABLE 7: PARAMETERS OF THE CZTS BASED SOLAR CELLS UNDER OPTIMUM ABSORBER WIDTH CONDITION ($W_{b,opt}$) FOR DIFFERENT TC/EMITTER COMBINATIONS. HERE $W_e = 0.1 \mu\text{m}$

Parameter	Gr/CdS	ITO/CdS	Gr/ZnO	ITO/ZnO
$W_{b,opt} (\mu\text{m})$	2.0	2.2	1.9	2.1
$J_{sc} (\text{mA}/\text{cm}^2)$	19.1213	15.9412	21.2003	17.6593
$V_{oc} (\text{V})$	0.7207	0.71	0.7265	0.7155
FF (%)	74.5302	74.6885	74.16	74.3345
$\eta (\%)$	10.2708	8.4534	11.4221	9.3923

REFERENCES

- [1] H. Katagiri, “ $\text{Cu}_2\text{ZnSnS}_4$ thin film solar cells”, *Thin Solid Films*, vol. 480, pp. 426- 432, 2005.
- [2] Wilson Da Silva, “UNSW Takes Lead in Race for Non-toxic, Thin-film Solar Cells”, *Newsroom, The University of New South Wales (UNSW)*, Australia, 27 April 2016.
- [3] L. Schmidt-Mende and J. L. MacManus-Driscoll, “ZnO—nanostructures, defects, and devices,” *Materials Today*, vol. 10, no. 5, pp. 40–48, 2007.
- [4] Chetan Singh Solanki, *Solar Photovoltaics: Fundamentals, Technologies and Applications*, Third edition, Prentice Hall, 2015.
- [5] Qingbin Zheng, Zhigang Li, Junhe Yang and Jang-Kyo Kim, “Graphene Oxide-Based Transparent Conductive Films”- 14 April 2014, *Progress in Materials Science*, vol. 64, pp. 200-247, 2014.
- [6] Yasin Khatami, Wei Liu, Jiahao Kang, and Kaustav Banerjee, “Prospects of graphene electrodes in photovoltaics” Vol. 8824 88240T-6, 2014.
- [7] O’Connor, B., Haughn, C., An, K.-H., Pipe, K. P., Shtein, M., “Transparent and conductive electrodes based on unpatterned, thin metal films,” *Appl. Phys. Lett.*, vol. 93, no. 22, pp. 223304, 2008.
- [8] Bhosle, V., Prater, J. T., Yang, F., Burk, D., Forrest, S. R., Narayan, J., “Gallium-doped zinc oxide films as transparent electrodes for organic solar cell applications,” *J. Appl. Phys.*, vol. 102, no. 2, pp. 023501, 2007.
- [9] Tvingstedt, K., Inganäs, O., “Electrode grids for ito-free organic photovoltaic devices,” *Advanced Materials*, vol. 19, no. 19, pp. 2893, 2007.
- [10] Xuefeng Wang, Yong P. Chen and David D. Nolte, “Strong Anomalous Optical Dispersion of Graphene: Complex Refractive Index Measured by Picometrology”, *Optics Express*, vol. 16, issue 26, pp.22105-22112, 2008.

- [11] Hyesung Park, Jill A Rowehl, Ki Kang Kim, Vladimir Bulovic and Jing Kong, "Doped Graphene Electrodes for Organic Solar Cells", *Nanotechnology*, vol. 21, pp. 505204, 2010.
- [12] Jon Cartwright, "Graphene Band Gap Heralds New Electronics", *Chemistry World*, 29 September 2015. Available: <http://www.rsc.org/chemistryworld/2015/09/graphene-band-gap-electronics-transistors-semiconductor>.
- [13] Daniel R. Cooper, Benjamin D'Anjou, Nageswara Ghattamaneni, Benjamin Harack, Michael Hilke, Alexandre Horth, Norberto Majlis, Mathieu Massicotte, Leron Vandsburger, Eric Whiteway and Victor Yu, "Experimental Review of Graphene", *ISRN Condensed Matter of Physics*, vol. 2012, article id 501686, pp. 1-56, 2012.
- [14] M. S. Anjan and M. Z. Kabir, "Modeling of Current-Voltage characteristics of CdS/CdTe Solar Cells," *Phys. Stat. Solidi A*, vol. 208, pp. 1813–1816, 2011.
- [15] S. M. Sze, *Physics of Semiconductor Devices*, 2nd Edition, Wiley & Sons., 1981.
- [16] M. M. Haque, M. M. Rahman and M. I. B. Chowdhury, "Current-Voltage Characteristics of CdS/CIGS Thin Film Solar Cells: An Analytical Approach," *Proceedings of 2014 1st International Conference on Non Conventional Energy (ICONCE 2014)*, pp. 28-31, January 16 - 17, 2014, Kalyani, West Bengal, India.
- [17] Ashiq Been Rahim, Abdullah Sakib Hasan, Protik Biswas, Aman Ullah, Md. Iqbal Bahar Chowdhury, "Analytical Modeling of J-V Characteristics of CIGS Based Thin Film Solar Cell Considering Voltage and Space Dependent Electric Field in the Absorber Layer", *3rd International Conference on Advances in Electrical Engineering (ICAEE 2015)*, pp. 368-371, December 17-19, 2015, Dhaka, Bangladesh.
- [18] National Renewable Energy Laboratory (NREL), <http://rredc.nrel.gov/solarspectra/am1.5/ASTMG173/ASTMG173.html>.
- [19] Ramprasad Chandrasekharan, "Numerical Modeling Of Tin-Based Absorber Devices For Cost-Effective Solar Photovoltaics", Ph. D. Thesis, The Pennsylvania State University, May 2012.
- [20] Ayesha, Kulsum, Shamara Kabir Silvi and Md. Iqbal Bahar Chowdhury, "Modeling of JV characteristics of CZTS based thin film solar cells including voltage and space dependent electric field in the absorber layer", *4th International Conference on the Development in the Renewable Energy Technology (ICDRET 2016)*, pp. 1-4, January 7-9, 2016, Dhaka, Bangladesh.

Boost Hybrid Converter with Multiple Inductor Multiplier Cell

Md. Ali Azam Khan
Associate Professor, Department of EECE, MIST
email: azam032002@yahoo.com

Mohammad Ali Chowdhury
Department of EEE
email: mac@buet.ac.bd

Abstract – Switch mode dc-dc converters are attractive for their small size, ease of control and efficient power conversion. Output voltage is regulated by duty cycle control of semiconductor switch of switch mode dc-dc converters. The voltage gain and efficiency of practical switching regulators deviate from ideal values at extreme duty cycles. Also, desired gain /attenuation is not achievable at high/low duty cycles. In applications where high gain or high attenuation of voltage is desired with acceptable energy conversion efficiency, hybrid dc-dc switching converters are used. Hybrid dc-dc converters are combination of voltage multiplier/division circuit with appropriate SMPS circuits. By incorporating voltage multiplier/division cell with conventional converters like Buck, Boost, Buck-Boost, Ćuk and SEPIC converters, desired voltage gain (either very low or very high) may be achieved at acceptable energy conversion efficiency. In the present work with an aim to attain very high voltage gain by conventional Boost topologies, a new voltage multiplier cell consisting of multiple inductors and diodes is proposed. The proposed voltage multiplier cell can be duplicated in multiple number.

Keywords – Switch mode dc-dc converters are attractive for their small size, ease of control and efficient power conversion. Output voltage is regulated by duty cycle control of semiconductor switch of switch mode dc-dc converters. The voltage gain and efficiency of practical switching regulators deviate from ideal values at extreme duty cycles. Also, desired gain /attenuation is not achievable at high/low duty cycles. In applications where high gain or high attenuation of voltage is desired with acceptable energy conversion efficiency, hybrid dc-dc switching converters are used. Hybrid dc-dc converters are combination of voltage multiplier/division circuit with appropriate SMPS circuits.

I. INTRODUCTION

The DC-DC converters are widely used in industrial applications and computer hardware circuits, DC-DC conversion techniques have been developed very quickly. Modern electronic systems require high-quality, small, lightweight, reliable, and efficient power supplies. Linear power regulators, whose principle of operation is based on a voltage or current divider, are inefficient. This is because they are limited to output voltages smaller than the input voltage, and also their power density is low because they require low frequency (50 or 60 Hz) line transformers and filters. Linear regulators can, however, provide a very high-quality output voltage. Their main area of application is at low power levels. Electronic devices in linear regulators operate in their active (linear) modes, but at higher power levels switching regulators are used. Switching regulators use power electronic semiconductor switches in on and off states. Because there is a small power loss in those states (low voltage across a switch in the on state, zero current through a switch in the off state), switching regulators can achieve high energy conversion efficiencies.

Modern power electronic switches can operate at high frequencies. The high frequency DC-DC chopper-converters are known as switch mode power supplies (SMPS). SMPS are extensively used as power supplies of computers and electronic equipment. According to the type of voltage regulation (step-down/step-up/step-up-down) SMPS has three basic (Buck, Boost and Buck-Boost/Ćuk/Sepic) and many derived configurations. The voltage regulation is achieved by duty cycle (D) control of the switching signal of the semiconductor switch of the SMPS. In a practical SMPS, the efficiency and the voltage-gain are function of the duty cycle D. Each type of SMPS has high value of

efficiency for limited range of voltage gain (within a range of duty cycle D). In case of extreme gains (either high or low), hybrid SMPS converters may be used to maintain high conversion efficiency. Hybrid SMPS converters are combination of voltage multiplier/divider circuits with appropriate SMPS circuits. In this Thesis work the boost, the hybrid boost dc-dc converter with two inductors, the hybrid proposed boost converter with two inductors and the proposed hybrid boost dc-dc converter with three inductors are studied by simulation at variable duty cycles.

II. MULTIPLE INDUCTOR MULTIPLIER CELL

Proposed two inductor cell for boost dc-dc hybrid converter is shown in Figure-1. The circuit of Figure-2 takes the form of Figure-2 when the switch is ON. Currents flow through the inductors according to current division rule and rises with time. Instantaneous voltage across

the inductor L_1 and L_2 are $v_{L1} = L_1 \frac{di_{L1}}{dt}$ and $v_{L2} = L_2 \frac{di_{L2}}{dt}$

respectively. If the two inductors have half the value of the inductor of single inductor conventional dc-dc converter, then i_{L2} and i_{L1} will be larger than i_L of the L of conventional dc-dc converter. As a result v_{L2} and v_{L1} will be higher than v_L of conventional dc-dc converter.

When the switch of the proposed cell turns OFF, $V_{dc} + v_{L1}$ or $V_{dc} + v_{L2}$ appear across the switch providing the boost voltage to the next stage of the converter. Since v_{L1} or v_{L2} stage of the converter,

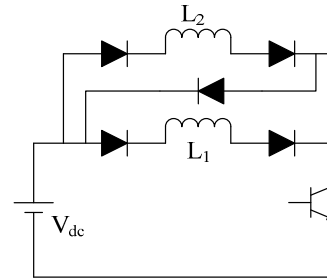
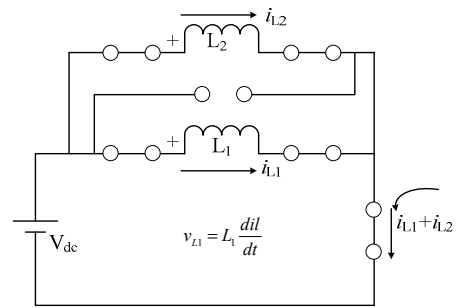
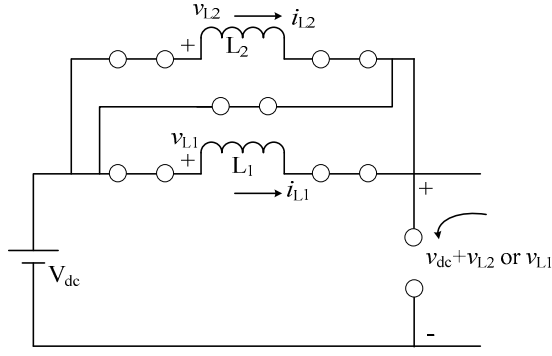


Figure - 1 Proposed Inductor cell for Boost converter



(a) Circuit for Figure-1, when switch is ON



(b) Circuit for Figure-1, when switch is OFF

Figure 2.2 Circuit of Figure 2.1 when is turns ON and OFF is higher than v_1 of conventional dc-dc converter, the proposed cell will provide higher Boost voltage to the next stage of the dc-dc converter at a lower duty cycle then the conventional dc-dc Boost converter. The proposed cell is different than the Boost cell of Figure 2.3-2.4 made of two inductors of reference [1]. It can be multiplied several times as shown in Figure 2.5.

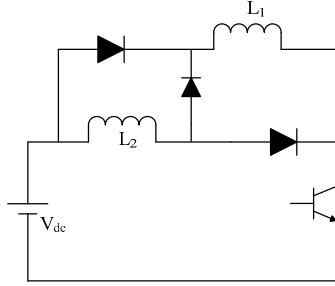
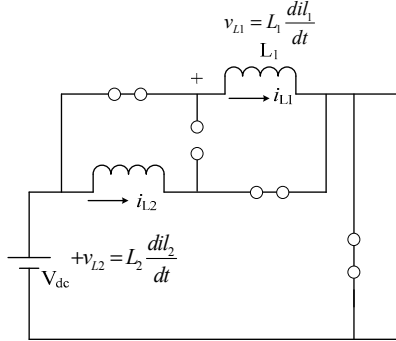
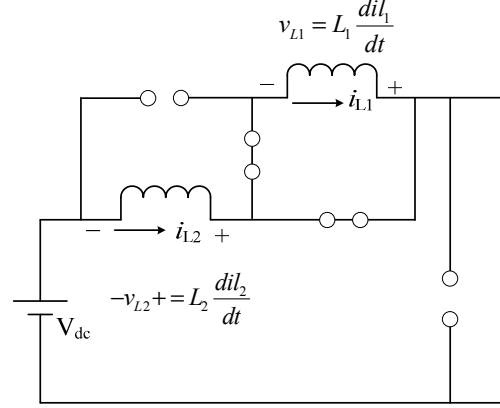


Figure-2.3 Boost cell made two inductors as reported in reference [1]



(a) Boost cell of Figure 2.3 when switch is ON



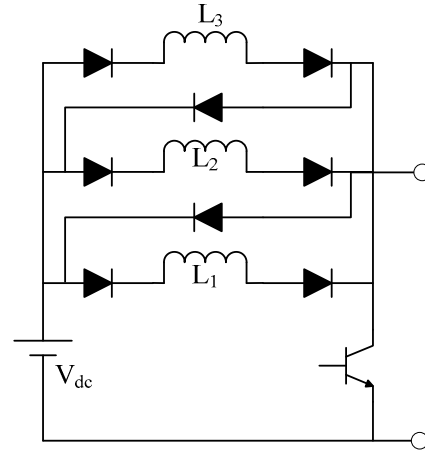
(b) Boost cell of Figure 2.3 when switch is OFF

Figure 2.4 Boost cell made of two inductors as reported in reference [1].

(a) The circuit of Figure 2.3 when switch is ON

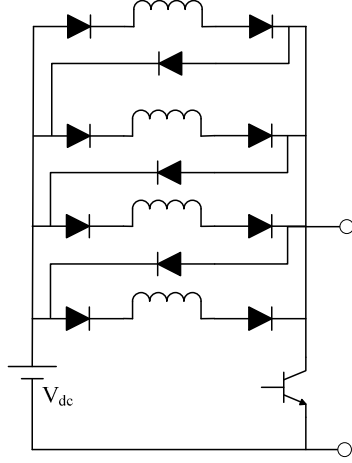
(b) The circuit of Figure 2.3 when switch is OFF.

Three and four inductor cells for example



a) Proposed boost cell with three inductors

III. RESULTS AND DISCUSSIONS



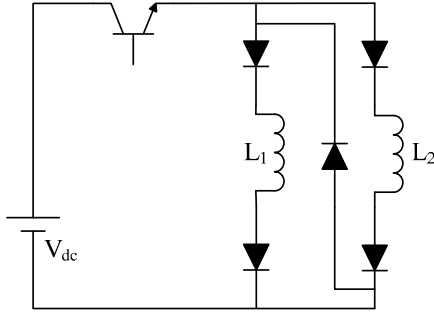
b) Proposed Boost cell with four inductors.

Figure 2.5 Proposed multi-inductor Boost cell for Boost, Ćuk and SEPIC hybrid high gain dc-dc converter.

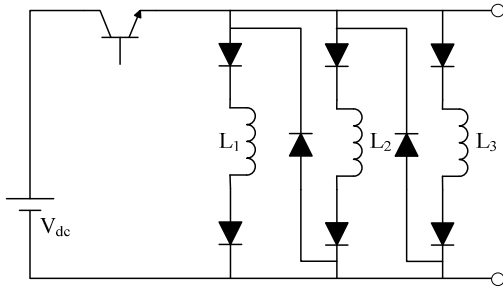
a) Proposed boost cell with three inductors

b) Proposed boost cell with four inductors.

Same proposed multi-inductor boost cell can be adopted for Buck-boost, high gain dc-dc converter as shown in Figure-2.6.



a) Two inductor Boost cell for Buck-Boost high gain hybrid dc-dc converter.



b) Three inductor boost cell for Buck-boost high gain hybrid dc-dc converter

Figure 2.6 Proposed multi-inductor boost cell for high gain hybrid buck-boost DC-DC converter.

a) Boost cell with two inductors

b) Boost cell with three inductors.

The boost, the hybrid boost dc-dc converter with two inductors [1], the hybrid proposed boost converter with two inductors and the proposed hybrid boost dc-dc converter with three inductors are studied by simulation at variable duty cycles. The total inductors in all the circuits are kept at 1 mH for the purpose of comparison on the same component value. The conventional Boost circuit is shown in figure 2.11 and its typical waveforms are shown in figure 2.12. The hybrid boost circuit of reference [1] is shown in figure 2.15 and its typical waveforms is shown in Figure 2.16. The proposed boost hybrid circuit with two and three inductor cells are shown in Figures 2.19 and 2.23 respectively. Typical waveforms of Figures 2.19 and 2.23 are shown in Figures 2.20 and 2.24 respectively. The input voltage, output voltage, voltage gain, input power, output power and efficiency are tabulated in Tables 2.1-2.4 and depicted in graphs of Figure 2.13-2.14, Figure 2.17-2.18, Figure 2.21-2.22 and Figure 2.25-2.26 for circuits of Figure 2.11, Figure 2.15, Figure 2.19, Figure 2.23 respectively. Table 2.A compares the results of four configurations for voltage gain of 3.5 in the first row, which indicate that for 1mH total inductor the voltage gain 3.5 takes place at 0.7, 0.4, 0.4 and 0.2 duty cycle for normal, hybrid boost [1], proposed hybrid boost of 2 inductors and proposed hybrid boost converter of 3 inductor. The efficiency of the paper hybrid circuit is higher than normal and proposed hybrid circuit for two inductors. But for 3 inductor circuit the efficiency is better than normal and hybrid circuit of [1].

Also in the second row of the Table 2.A it is evident that voltage gain of the proposed multiple inductors can be made higher than any other topology. The increase of inductor number keeping the total inductor value same is not possible in the hybrid circuit of the paper [1].

Table 2.A Comparison of Normal Boost, Boost Topology of [1], Proposed Boost Topology, Proposed Topology with Three Inductor

Normal Boost			Hybrid Boost of [1] two Inductor		
Duty cycle	Voltage Gain	Efficiency %	Duty cycle	Voltage Gain	Efficiency %
0.7	3.5	80	0.4	3.5	85
0.875	7.77	70.11	0.875	11.82	55.84

Table 2.A (Continued)

Hybrid Boost of Proposed Topology with 2 Inductor			Hybrid Boost of Proposed Topology with 3 Inductor		
Duty cycle	Voltage Gain	Efficiency %	Duty cycle	Voltage Gain	Efficiency %
0.4	3.5	74	.2	3.5	85.84
0.875	10.55	49.64	0.875	13.358	52.56

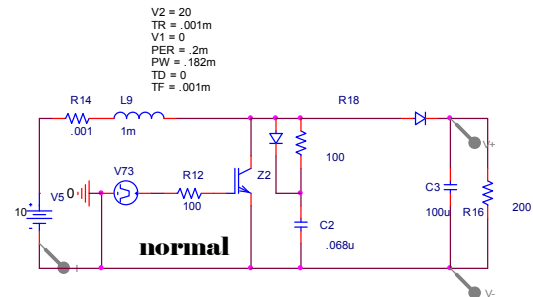


Figure-2.11: Boost DC-DC Converter

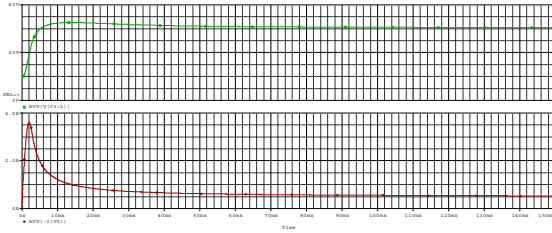


Figure- 2.12: Typical Waveforms of Boost DC-DC Converter of Figure-2.11

Table-2.1: Performance of Boost DC-DC Converter of Circuit of Figure-2.11

Duty Cycle	Vin (V)	Iin (A)	Vout (V)	Iout (A)	Pin (W)	Pout (W)	Voltage gain	Efficiency in percentage
0.125	10	0.084	12.46	0.062	0.84	0.77	1.24	91.66
0.25	10	0.156	16.82	0.084	1.56	1.41	1.68	90.38
0.375	10	0.261	22.49	0.112	2.61	2.51	2.24	80.45
0.50	10	0.40	26.56	0.127	4.00	3.37	2.65	84.25
0.625	10	0.544	30.54	0.152	5.44	4.64	3.05	85.29
0.75	10	1.00	39.75	0.198	10.00	7.87	3.97	78.70
0.875	10	4.30	77.72	0.388	43.00	30.15	7.77	70.11

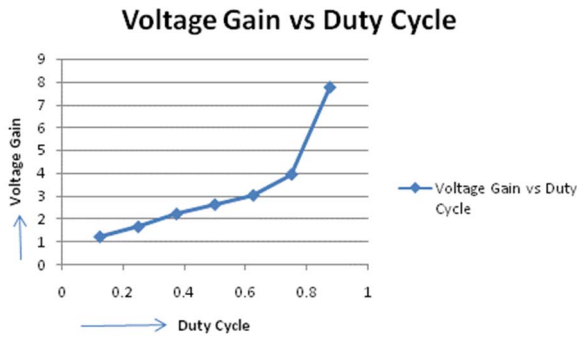


Figure-2.13: Voltage Gain vs Duty Cycle curve of Boost DC-DC Converter of Figure-2.11

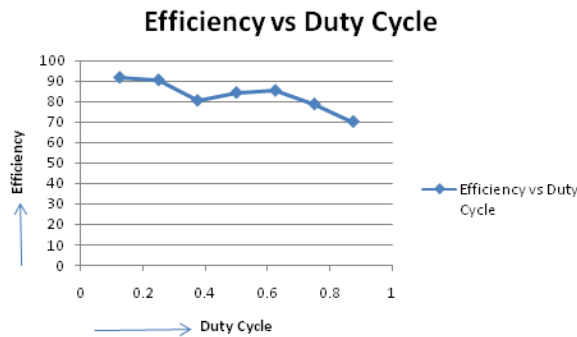


Figure-2.14: Efficiency vs Duty Cycle curve of Boost DC-DC Converter of Figure-2.11

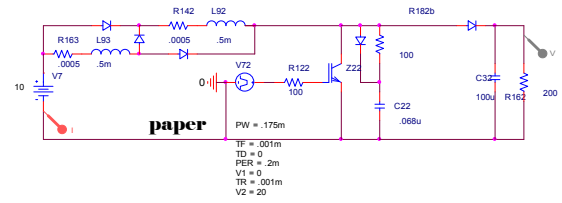


Figure-2.15 High Gain Hybrid Boost DC-DC Converter [1]

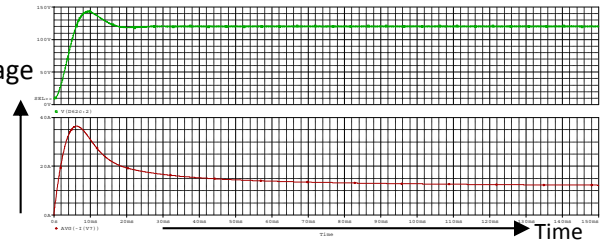


Figure-2.16: Typical Waveforms of High Gain Hybrid Boost DC-DC Converter of Figure-2.15

Table-2.2: Performance of High Gain Hybrid Boost DC-DC Converter of Figure-2.15 []

Duty Cycle	Vin (V)	Iin (A)	Vout (V)	Iout (A)	Pin (W)	Pout (W)	Voltage Gain	Efficiency in percentage
0.125	10	0.152	16.18	0.080	1.52	1.308	1.618	86.115
0.25	10	0.377	24.7	0.123	3.77	3.050	2.47	80.913
0.375	10	0.704	34.95	0.174	7.04	6.107	3.495	86.754
0.5	10	1.164	42.09	0.210	11.64	8.857	4.209	76.098
0.625	10	1.763	51.485	0.257	17.63	13.253	5.1485	75.175
0.75	10	2.818	61.692	0.308	28.18	19.029	6.1692	67.528
0.875	10	12.52	118.248	0.591	125.2	69.912	11.8248	55.841

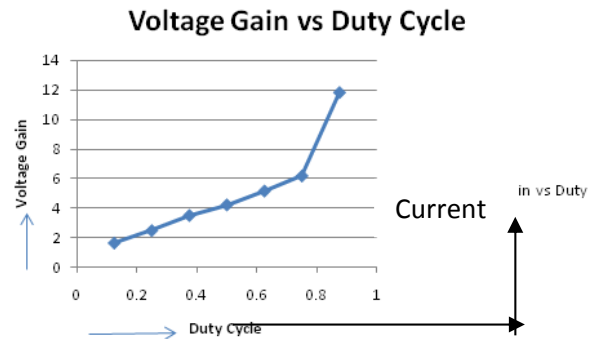


Figure-2.17: Voltage Gain vs Duty Cycle curve of High Gain Hybrid Boost DC-DC Converter of Figure-2.15

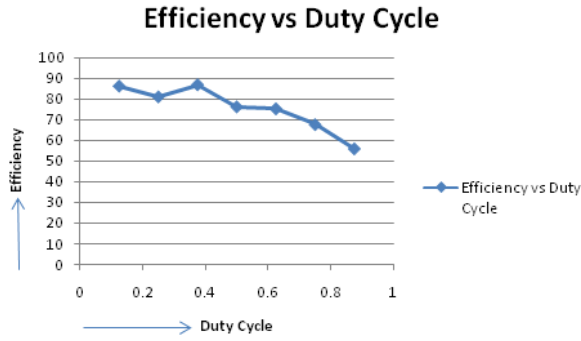


Figure-2.18: Efficiency vs Duty Cycle curve of High Gain Hybrid Boost DC-DC Converter of Figure-2.15

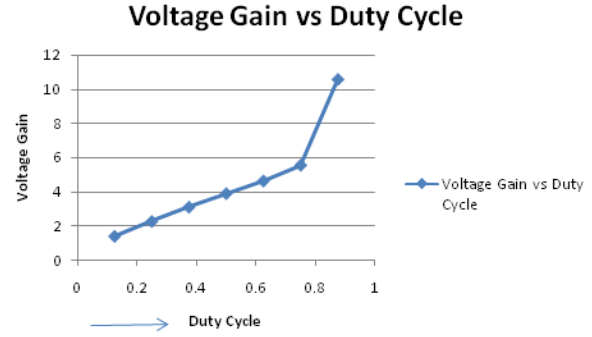


Figure-2.21: Voltage Gain vs Duty Cycle curve of High Gain Hybrid Boost DC-DC Converter of Figure-2.19

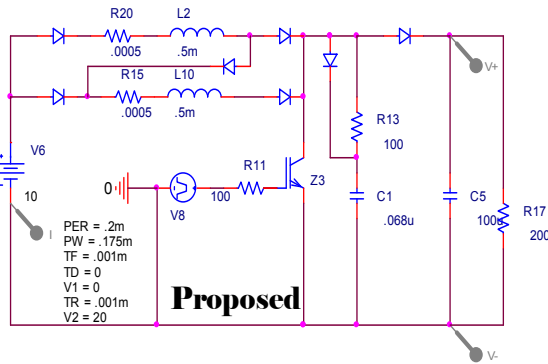


Figure -2.19: High Gain Hybrid Boost DC-DC Converter (proposed)

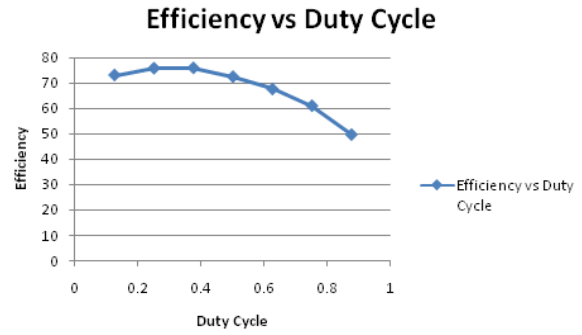


Figure-2.22: Efficiency vs Duty Cycle curve of High Gain Hybrid Boost DC-DC Converter of Figure-2.19

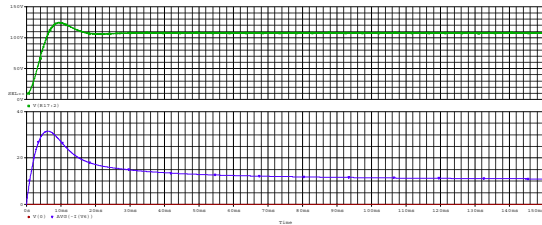


Figure-2.20: Typical Waveforms of High Gain Hybrid Boost DC-DC Converter of Figure-2.17(proposed)

Table-2.3: Performance of High Gain Hybrid Boost DC-DC Converter of Figure-2.19

Duty Cycle	Vin (V)	Iin (A)	Vout (V)	Iout (A)	Pin (W)	Pout (W)	Voltage Gain	Efficiency in percentage
0.13	10.00	0.14	14.25	0.07	1.39	1.02	1.43	73.04
0.25	10.00	0.35	22.92	0.12	3.47	2.63	2.29	75.70
0.38	10.00	0.64	31.25	0.16	6.44	4.88	3.13	75.82
0.50	10.00	1.06	39.12	0.20	10.57	7.65	3.91	72.39
0.63	10.00	1.60	46.49	0.23	16.00	10.81	4.65	67.54
0.75	10.00	2.54	55.65	0.28	25.40	15.49	5.57	60.96
0.88	10.00	11.22	105.55	0.53	112.20	55.70	10.56	49.65

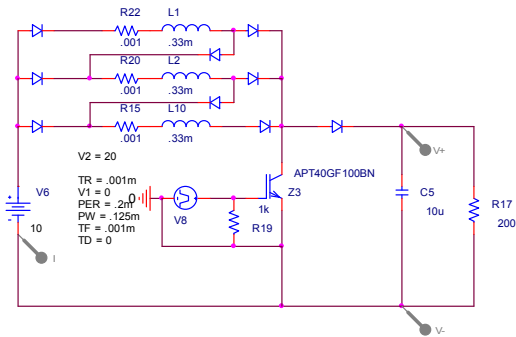


Figure-2.23: High Gain Hybrid Boost DC-DC Converter (proposed)

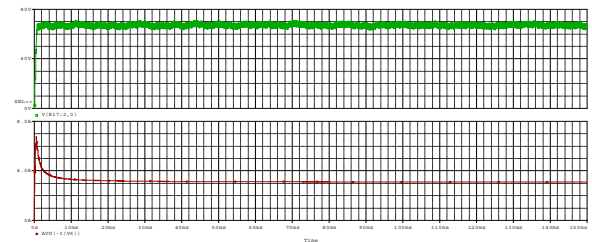


Figure-2.24: Typical Waveforms of High Gain Hybrid Boost DC-DC Converter of Figure-2.23 (proposed)

Table-2.4: Performance of High Gain Hybrid Boost DC-DC Converter of Figure-2.23

Duty Cycle	Vin (V)	Iin (A)	Vout (V)	Iout (A)	Pin (W)	Pout (W)	Voltage Gain	Efficiency in percentage
0.13	10.00	0.22	19.47	0.10	2.21	1.89	1.95	85.84
0.25	10.00	0.63	32.44	0.16	6.30	5.19	3.24	82.38
0.38	10.00	1.25	45.17	0.22	12.50	9.93	4.52	79.49
0.50	10.00	2.10	56.72	0.28	21.00	15.88	5.67	75.62
0.63	10.00	3.12	67.07	0.33	31.20	22.13	6.71	70.93
0.75	10.00	4.58	77.38	0.38	45.80	29.40	7.74	64.20
0.88	10.00	16.97	133.58	0.66	169.70	89.21	13.36	52.56

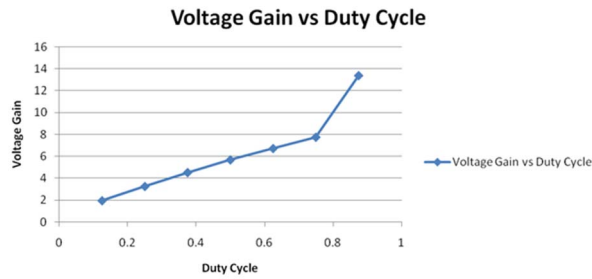


Figure-2.25: Voltage Gain vs Duty Cycle curve of High Gain Hybrid Boost DC-DC Converter of Figure-2.23

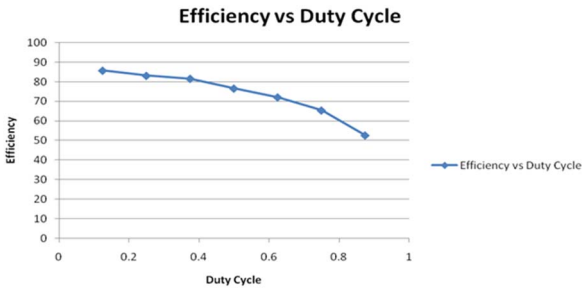


Figure-2.26: Efficiency vs Duty Cycle curve of High Gain Hybrid Boost DC-DC Converter of Figure-2.23

IV. CONCLUSIONS

A new multi-inductor voltage multiplier cell is proposed in the thesis to be incorporation with commonly used switch-mode dc-dc converters for their high voltage gain operation with acceptable efficiency of energy conversion. The proposed multiplier cell can use more than two parallel inductors with appropriate diode connections for voltage build up in them during ON time of the switch of the dc-dc converter. The charged inductors with voltage across them are connected in series during OFF time of the switch and charge the capacitor across the load to boost up the voltage of the load.

The proposed cell is investigated with conventional Boost, Ćuk, SEPIC and Buck-Boost dc-dc converters topologies with expected result of attaining same voltage gain than conventional dc-dc converter counterpart with higher efficiencies at lower duty cycle. The dc-dc converters were found to behave similarly for voltage gain and efficiency at higher duty cycles. It was also observed that the voltage gain of these converters in hybrid mode operations can be varied widely to higher gain levels than conventional converters.

Voltage multiplier cell made of parallel inductors and diodes at the input of conventional dc-dc converters have been found to operate satisfactorily for Boost, Ćuk and SEPIC converters. However, the hybrid Buck-Boost converter was found to exhibit the desired trend with very low efficiency. Hence the voltage multiplier cell made of inductors and diodes in its present form as reported in literature, and that proposed in this thesis is not suitable for hybrid operation of Buck-Boost dc-dc converters.

REFERENCES

- [1] B. Axelrod, Y. Berkovich, A. Ionovici: "Switched-capacitor/switched-inductor structures for getting transformerless hybrid DC-DC PWM converters", IEEE Transactions on Circuits and Systems I. vol. 55, no. 2, pp. 687-696. 2008.
- [2] A. Kishon Robert, C. Baskaran "Single switch high stepup dc-dc converters with voltage multiplier cell", International Journal of Advanced Trends in Computer Science and Engineering, Vol.2, No.2, Pages : 95-100 (2013) Special Issue of NCRTECE 2013 - Held during 8-9 February, 2013 in SMK Fomra Institute of Technology, OMR, Thaiyur, Kelambakkam, Chennai pp 95-100, February 2013.
- [3] A. Gopi and R. Saravanakumar, "High Boost Isolated DC-DC Converter with Controller. "World Applied Sciences Journal 23 (4): pp. 486-494, 2013 ISSN 1818-4952, 2013.
- [4] Daniel W Hart., "Power Electronics", McGraw Hill, 2010
- [5] R. W. Erickson, "Fundamental of Power Electronics", Chapman and Hall, USA, 1997
- [6] M. H. Rashid, "Power Electronics Handbook", Academic Press, USA, 2001
- [7] Timothy Skvarenina, "The Power Electronics Handbook", CRC Press 2002
- [8] Fang Lin Luo and Hong Ye, "Advanced Multi-Quadrant DC/DC Converters", CRC Press, 2006
- [9] Keng C Wu., "Switch Mode Power Converter", Elsevier Academic Press, 2006
- [10] Dorin O. Neaesu, "Power Switching Converters", CRC Press 2006

Analysis and Optimization of an Optical Frequency Comb for Multiple Channel Transmission in a Fiberoptic System

Rifat Nazneen

Department of Electrical and Electronic Engineering
American International University- Bangladesh (AIUB)
Dhaka, Bangladesh

Ebad Zahir

Department of Electrical and Electronic Engineering
American International University- Bangladesh (AIUB)
Dhaka, Bangladesh

Abstract— The proposed scheme was successful in generating equally spaced frequency combs by launching three continuous-wave pumps at 1 nm wavelength with 10 MHz linewidth using FWM in multiple highly non-linear fibers. Variations in pump power and frequency separations were explored to determine the effect of the input parameters on the frequency comb generation. The dispersion effect and EDFA length are shown to have a significant effect on the comb width as well. Results show that for an optimized number of pumps, the scheme is successful in generating the spectral lines for a specific bandwidth.

Keywords—*Optical Frequency Comb; Four Wave Mixing; Highly Nonlinear Fiber; Erbium Doped Fiber Amplifier; Fiberoptic Communication.*

I. INTRODUCTION

Rapid growth of subscribers' demand is putting a strain on the data traffic rate. Optical frequency comb (OFC) is one of the promising techniques that can increase the information carrying capacity for long distance fiber optical communication. It attracts the interest of researchers due to its wide applications such as high-accuracy photonic sensors, photonic or optical microwave signal processing, precise photonic metrology, photonic arbitrary waveform generation and ultrafast photonic signal processing [1], multi-wavelength ultra-short pulse generation [2], in optical communication systems like Wavelength Division Multiplexing (WDM), Dense Wavelength Division Multiplexing (DWDM) [2], Ultra Dense Wavelength Division Multiplexing (UDWDM), Nyquist Wavelength Division Multiplexing [3] etc. Its main advantage lies in the evenly spaced coherent spectrum lines as well as a stable phase response. OFC with its uniformly spaced frequency modes and its power tones have already been employed as ideal carriers in Wavelength Division Multiplexing (WDM) with a large number of channels [4]. Generally, the generation of OFC depends on the continuous laser source which modulates the radio frequency signal to generate a set of spectral lines. This spectrum has a central wavelength defined by laser source and lines are spaced by the RF oscillator frequency. Several methods have been employed to generate OFC such as mode locked laser, re-circulating shifter, using micro-resonators [5], cascading modulators, nonlinear effects in nonlinear medium etc [6]. The most traditional one relies on the use of passively mode locked

lasers (MLL) which generate high repetition rates but requires complexity of design, sophisticated thermal and acoustic isolation [7]. Such spectrum lacks both tune-ability and wavelength span [8]. The comb tone pitch cannot be effectively changed due to the inability of reconfigure and stabilize the MLL cavity at same time [9]. Another approach is by using a re-circulating shifter having a RFS loop, thus making the system more complex and costly. In the cascade modulator approach using resonant cavities, it's possible to generate an optical spectrum with a good OSNR but provides a narrowband spectrum for the comb lines [9]. Furthermore, it needs a DC bias control, which reduces the stability of the system [10]. Four-wave mixing (FWM) is the third order non linear effect in the fiber which provides an effective approach to generate OFC. This is a method where two or more wavelengths are combined to create a new wavelength. By using the wavelength conversion method in cascaded FWM, it is possible to generate equally spaced comb lines spanning a broad spectrum region [11]. Broadband generation of cascaded FWM generates traditionally in highly non-linear fibers and photonic crystal fibers [11]. Erbium doped fiber amplifiers (EDFA) which amplify the optical signal in long distance communication, also enhance the FWM effectively to generate broad bandwidth of spectrum. Research shows that there are several variables which have an impact on FWM and thus OFC.

In this paper, a scheme is proposed to characterize and optimize evenly spaced tunable frequency combs which are generated by launching three continuous-wave pumps at 1 nm spacing wavelengths into highly nonlinear fibers. The scheme shows the tuning of comb spacing have an impact on equally spaced spectral lines. Different pump powers and frequency separations are explored numerically to determine the effect of the input parameters on frequency comb generation. Variations of number of pump and EDFA length determined significant changes in the frequency comb generation due to the variation in FWM.

II. FOUR-WAVE MIXING

Four-Wave Mixing is a third order parametric nonlinear effect in the optical medium. It occurs when two or more light waves travel through the fiber and generate waves at new

wavelengths [12]. If three waves at different wavelengths pass through a longer length of fiber then it will generate the fourth wavelength. Then this effect can be defined by the equation,

$$f_{\text{new}} = f_1 + f_2 - f_3 \quad (1)$$

Where, f_1 & f_2 = pumped waves, f_3 = weak signal or probe light, and f_{new} = new photon at new wavelength or idler wavelength. This relationship is shown in Fig.1.

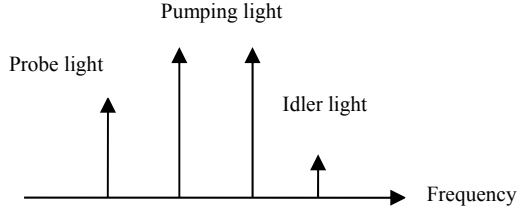


Fig.1. Four Wave Mixing in the frequency domain

Using this FWM the wavelength conversion occurs to provide high efficiency conversion and broad bandwidth in the fiber which is useful for the optical comb generation. Before converting the wavelength for comb generation using FWM some conditions must be satisfied:

- Pump wavelength must coincide with zero dispersion wavelengths.
- Chromatic dispersion variation in the longitudinal direction of the fiber should be minimized
- States of polarization of the pump and signals must be matched [12].

For efficient generation of FWM, the phase of this nonlinear polarization must match with the phase of the field wave at new wavelength as it propagates through the fiber. This phase-matching requirement occurs when the net wave-vector mismatch $\kappa = 0$, where κ is given by [13]:

$$\kappa = \Delta\kappa_M + \Delta\kappa_W + \Delta\kappa_{NL} = 0 \quad (2)$$

Where, $\Delta\kappa_M$, $\Delta\kappa_W$ and $\Delta\kappa_{NL}$ represent the mismatch occurring as a result of material dispersion, waveguide dispersion and nonlinear effects, respectively. Here,

$$\Delta\kappa_{NL} = \gamma (P_1 + P_2) \quad (3)$$

Where, γ is the fiber nonlinear coefficient, P_1 and P_2 are the incident power of both Lasers.

In highly non linear fiber the FWM generate efficiently to provide more optical spectral lines in broad bandwidth range. By decreasing effective area A_{eff} of nonlinear fiber enhance nonlinear co-efficient which satisfy the phase matching conditions of the spectral lines. Therefore, multiple FWM generates in fiber to enhance the optical frequency comb.

$$\gamma = \frac{2\pi n_2}{\lambda A_{\text{eff}}} \quad (4)$$

In the loop of non linear fiber induce more sidebands for the enhancement of FWM. The generation of number of sidebands depends on the equation,

$$N = 3*n - 2 \quad (5)$$

Where, n represents the number of pump lights launched into the feedback system and N means the total number of optical waves inside the fiber.

The nonlinear effects depend on transmission length. The longer length of fiber link provides the more the light

interaction and greater the nonlinear effect which is important for the optical comb generation.

$$L_{\text{eff}} = \frac{1}{\alpha} \quad (6)$$

III. PROPOSED SCHEME ANALYSIS

The proposed scheme is shown in Fig.2. The scheme is divided into two parts. In the first part, three continuous wave (CW) lasers with a linewidth of 10 MHz are coupled together using two 3db couplers to accumulate the pump power. It then passes through the phase modulator which modulates a 20 GHz bandwidth RF signal to suppress Stimulated Brillouin Scattering. The phase modulation of the CW laser source leads to the generation of many spectral sidebands by increasing the phase modulation depth and a wide range of OFC is obtained [14].

The modulated signal passes through a single mode fiber (SMF) of 1 km to reduce the chirp induced by the phase modulator. A short pulse train with a narrow linewidth is achieved after the SMF. This train is then amplified by an Erbium doped fiber amplifier (EDFA) with amplified spontaneous emission (ASE) to extend the transmission distance. Finally an optical filter is used to eliminate the ASE noise from the EDFA in order to get the desired signal.

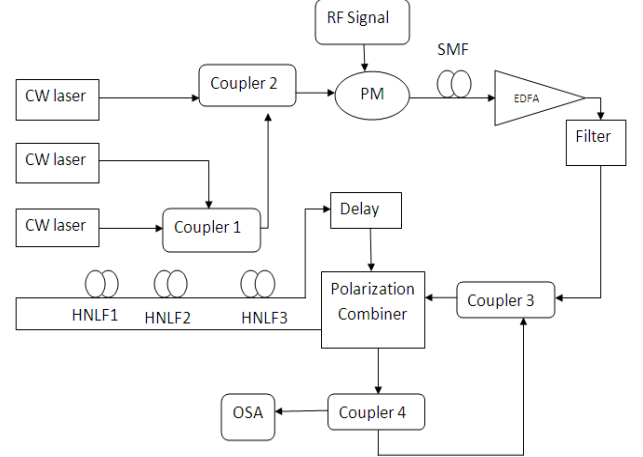


Fig.2. Proposed Scheme for the optical frequency comb generation

In the second part, the focus is on the effect of multiple highly nonlinear fibers (HNLF) on FWM generation. The feedback system contains three HNLFs which increase both multiple FWM and SPM effects in fibers to generate a broad range of optical frequency comb. In the feedback structure, a polarization combiner was used [15]. By using these two couplers, both are feeding back to the polarization combiner in different coupling ratios. As the coupler 3 and 4 are in 90:10 ratio, so the coupling ratio as well as feedback ratio is 9%. Recent experimental work validates the use of highly nonlinear feedback loop. Experimentally, it has been shown that with 9% feedback ratio and two laser sources can be used to generate frequency comb lines [16]. However, when the three pumps with same frequency spacing are injected into the feedback structure, second order equally spaced sidebands are produced by FWM in highly nonlinear fibers with a specific delay between each other. This delay reduces the overlapping of the spectral lines created with the FWM in the fiber loop.

IV. SIMULATION RESULTS ANALYSIS AND DISCUSSION

In the simulated scheme, three pumps at 1553 nm, 1552 nm, and 1551 nm are injected into a 3 db coupler with 10 MHz linewidth. An RF signal with 20 GHz bandwidth is modulated with the pulse waves to change the phase. A short pulse train is generated with 3 ps linewidth after the SMF having 1 ps/nm/km dispersion as shown in Fig.3.

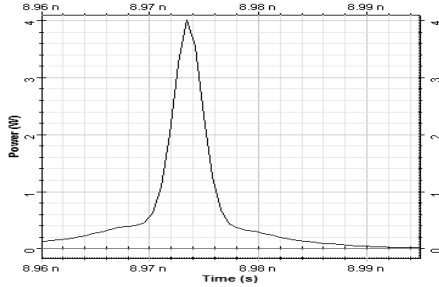


Fig.3. Pulse shape of pump wave after SMF

Using the EDFA, the pulse train is boosted by 1 W. Coupler 3 and Coupler 4 with 9% coupling ratios are fed back into the polarization combiner [15]. In the loop of nonlinear fibers are three HNLFs with zero dispersion wavelengths (ZDW): HNLF1 is 1600 nm, HNLF2 is 1550 nm, HNLF3 is 1500 nm. The length, nonlinear coefficient and dispersion slope of HNLFs are $L=0.03$ km, $\gamma=26$ W⁻¹.km⁻¹, slope=0.03 ps/nm/km². The layout was then simulated under different configurations to validate the effectiveness of the proposed scheme. Without using the SMF, results were acquired for 20 spectral lines in 10 dbm power deviation as shown in Fig.4. When the HNLF1 was removed from the scheme, results yielded 24 lines in 10 dbm power deviation as shown in Fig.5. In a third configuration, the whole feedback structure which contains HNLFs was removed to obtain 26 lines in 10 dbm power deviation as shown in Fig.6. In the final configuration which uses the proposed scheme, an evenly spaced 27 lines were acquired in 6 dbm power deviation as shown in Fig.7.

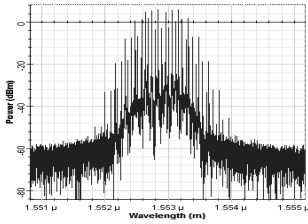


Fig.4. Response of the system without SMF

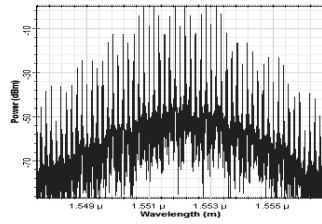


Fig.5. Response of the system without HNLF1

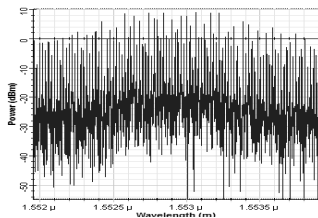


Fig.6. Response of the system without feedback structure

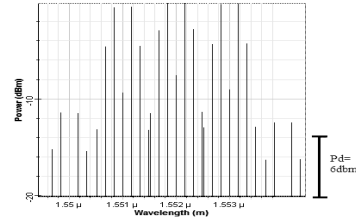


Fig.7. Response of the proposed scheme

V. CHARACTERIZATION AND OPTIMIZATION OF THE COMB RESPONSES

The impact of dispersion on the generation of FWM is evident in the results. As the dispersion of SMF increases, it reduces the number of comb lines as shown in Fig.8. Results show that as the number of pump sources with the same frequency spacing increases, the efficiency of FWM increases linearly. In this paper, the pumps are tuned with 1 nm and 0.1 nm frequency spacing to evaluate the performance of optical frequency comb shown in Fig.8 and Fig.9. By using 0.1 nm the numbers of lines are higher but they are not evenly spaced with each other. This implies that crosstalk can occur during the transmission system. Comparing Fig.8 and Fig.9 below, it is evident that while using 1 nm frequency spacing the number of lines increases linearly, but while using 0.1 nm spacing after five pump sources the spectral lines start reducing in number.

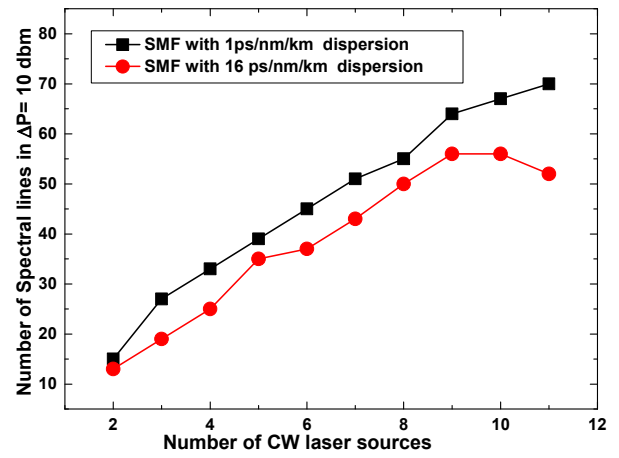


Figure 8: Number of spectral lines analysis for different pump sources in 1nm channel spacing and dispersion effects of SMF.

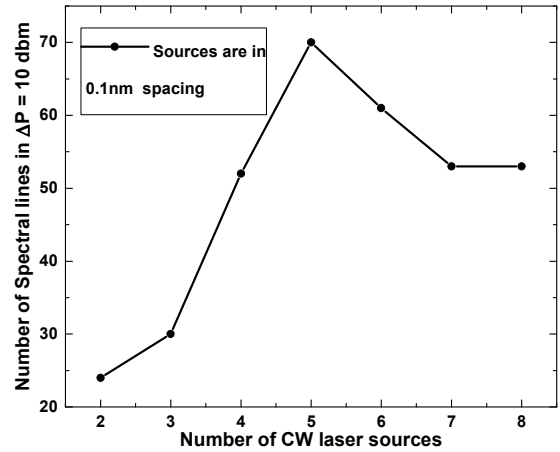


Figure 9: Number of spectral lines analysis for different pump sources.

Next, the impact of pump power on the system was explored. Results show that as the pump power increases, more sidebands are induced in the spectrum due to effective generation of FWM in highly non linear fibers. Fig.10 clearly shows the effect of varying the pump power from -30 dbm to

30 dbm. A sharp increase in the number of spectral lines were found at pump powers exceeding 20 dbm.

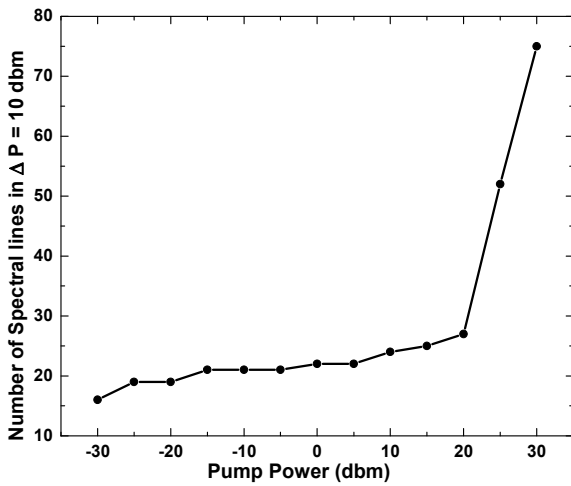


Figure 10: Correlation between the pump power and spectrum lines.

Last but not least, the effect of EDFA length on the generated spectral lines in Fig.11 was investigated as well. The number of frequency lines increase in 10 dbm power deviation as the EDFA length increases. When the EDFA length is set to 15 m, a maximum number of 27 spectral lines were achieved for the proposed scheme. Any further increase in the length did not have any major impact. A huge increase in length however showed a large drop in the number of spectral line generation. This is largely due to the increased output power for longer lengths of EDFA.

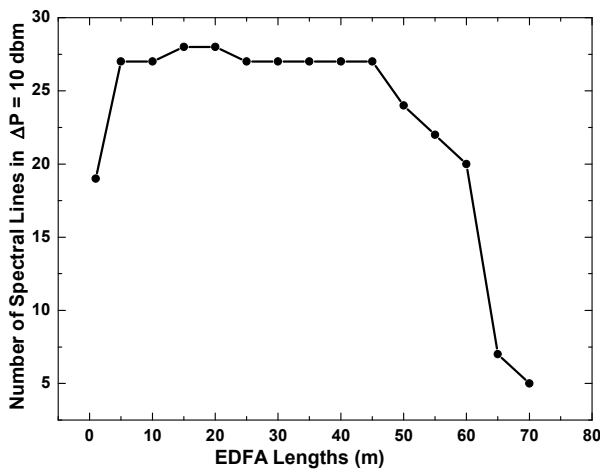


Figure 11: Variation in number of spectral lines for different EDFA lengths.

VI. CONCLUSION

In this paper, an OFC generation scheme was analyzed in detail. The OFC performance was characterized by investigating the impact of all the critical parameters. Once the parameters were optimized, the scheme yielded 27 comb lines with 6 dbm power deviation and 20 GHz modulation speed. The scheme was successful in generating evenly spaced

optical combs based on multiple four-wave mixing via a feedback structure of cascaded highly nonlinear fibers and a polarization combiner. This paper also successfully illustrates the correlation between variation of power, frequency spacing of pump sources, dispersion effect of SMF, and EDFA lengths on the optical frequency comb.

Further research can be done on the proposed scheme by using a Mach-Zender Modulator (MZM) to get a symmetrical optical comb. This will possibly improve the amplitude slope control characteristics of the different channels.

References

- [1] R. Ullah, B. Liu, Q. Zhang, Q. Tian, F. Tian, Z. Qu, C. Yan, M. Khan, I. Ahmad and X. Xin, "Optical frequency comb generation with high tone-to-noise ratio for large-capacity wavelength division multiplexed passive optical network", *Optical Engineering*, vol. 54, no. 11, p. 116102, 2015.
- [2] Yang, Ting et al. "Comparison Analysis Of Optical Frequency Comb Generation With Nonlinear Effects In Highly Nonlinear Fibers". *Opt. Express* 21.7 (2013): 8508.
- [3] Zhou, Rui et al. "Injection Locked Wavelength De-Multiplexer For Optical Comb-Based Nyquist WDM System". *IEEE Photonics Technology Letters*, vol.27, no.24, pp. 2595-2598, (2015)
- [4] C. Lei, H. Chen, M. Chen, S. Yang and S. Xie, "Recirculating Frequency Shifting Based Wideband Optical Frequency Comb Generation by Phase Coherence Control", *IEEE Photonics J.*, vol. 7, no. 1, pp. 1-7, 2015.
- [5] Nagarjuna, K.P, Shankar Kumar Selvarajaa, and V.R Supradeepa. "Generation Of Tunable, High Repetition Rate Frequency Combs With Equalized Spectra Using Carrier Injection Based Silicon Modulators". *Proc. of SPIE*, vol. 9752, *SPIE* (2016).
- [6] Jianrui Li, Jianrui Li et al. "Scheme for Generation of flat top and high signal-to-noise ratio optical frequency comb". *Chinese Optics Letters* 13.1 (2015): 010605-10610.
- [7] Li, Wei et al. "Generation Of Flat Optical Frequency Comb Using A Single Polarization Modulator And A Brillouin-Assisted Power Equalizer". *IEEE Photonics J.* vol 6. No 2 (2014).
- [8] A. Antikainen and G. Agrawal, "Dual-pump frequency comb generation in normally dispersive optical fibers", *Journal of the Optical Society of America B*, vol. 32, no. 8, p. 1705, 2015.
- [9] Melo, S.A.S. et al. "Frequency Comb Expansion Based On Optical Feedback, Highly Nonlinear and Erbium-Doped Fibers". *Optics Communications* 312 (2014): 287-291.
- [10] Yang, Yang et al. "Optical Frequency Comb Generation using two Cascaded Polarization Modulators". *Photonic Network Communications* (2015)
- [11] Arismar, Cerqueira et al. "Broadband Cascaded Four-Wave Mixing By Using a Three-Pump Technique In Optical Fibers". *Optics Communications* 282.22 (2009): 4436-4439..
- [12] O. Aso, M. Tadakuma and S. Namiki, "Four-Wave Mixing in Optical Fibers and Its Applications", *Furukawa Review*, no. 19, 2000.
- [13] A. Cerqueira Sodre, J. Chavez Boggio, A. Rieznik, H. Hernandez-Figueroa, H. Fragnito and J. Knight, "Highly efficient generation of broadband cascaded four-wave mixing products", *Opt. Express*, vol. 16, no. 4, p. 2816, 2008.
- [14] F. Zhang, J. Wu, Y. Li and J. Lin, "Flat optical frequency comb generation and its application for optical waveform generation", *Optics Communications*, vol. 290, pp. 37-42, 2013.
- [15] Dong, Qimeng et al. "Flat Frequency Comb Generation Based On Efficiently Multiple Four-Wave Mixing Without Polarization Control". *Photonics Sens* 6.1 pp.85-89, (2015).
- [16] Li, Jiabao et al. "Enhancement Of Cascaded Four-Wave Mixing Via Optical Feedback". *Opt. Express* 20.20 (2012): 21940.

Performance of Large Area Graphene FET Considering the Effect of Strain

Sabbir Ahamed*, Md. Nur Kutubul Alam, Md. Rafiqul Islam

Department of Electrical and Electronic Engineering

Khulna University of Engineering & Technology

Khulna-9203, Bangladesh

Email*: sabbir_kuet@hotmail.com

Abstract—In this work, strained large area graphene is used in the channel of Graphene-FET (G-FET) to realize the G-FET in logic application. The strain dependent quantum confinement effect in terms of quantum capacitance is investigated and a drain current model is proposed under tight binding approximation. The influence of strain on the performance of G-FET as a function of gate voltage, channel length and oxide thickness is studied. The results obtained from the present study indicate that the performance of G-FET is highly influenced by the strain. It is also found that the strained G-FET satisfactorily meet the requirements for logic applications.

Keywords—Large area G-FET; strained graphene; tight binding approximation; quantum capacitance

I. INTRODUCTION

Recently graphene has attracted numerous research attention because of its superior electronic and physical properties such as such as extremely high carrier charge mobility, high switching speed with ballistic transport behaviors, and anomalous quantum Hall effects [1-2]. However, due to the absence of bandgap, graphene is not suitable for logic devices but an excellent candidate for RF applications [3]. There are several ways to open bandgap in graphene such as graphene patterned into narrow ribbon which is called graphene nanoribbon (GNR) [4-6], biasing a bilayer graphene [7-8], using of strain engineering [9-10] etc. In GNR bandgap opens as a consequence of quantum confinement which is resulted from the finite width and edge geometry of graphene [5]. However, maintaining a uniform width in a practical device is challenging. Biasing can open a bandgap of 200-250meV which is hardly suitable for logic applications [7-8]. Recently strain engineering got significant attention in graphene electronics as it is found that the modification of band structure is possible by applying strain [9-10]. Strain changes the lattice structure a little bit, i.e. perturb the hexagonal structure of the graphene. As a consequence the orbital overlaps between neighboring carbon atom as well as the symmetry of the structure changes. And this causes some bandgap to open.

Two different kind of strain can be employed in graphene namely in-plane strain and biaxial strain. Although there are some reports on the presence of the in-plane strain confirmed

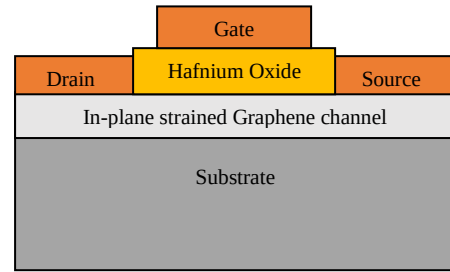


Fig. 1. Schematics of single-gate G-FET where single layer graphene under strain is used as a channel material.

from the G peaks and shifts in the Raman spectrum [11], the practical implementation of in-plane strain is limited till today.

In contrast biaxial strain can be achieved more easily as well as reproducibly from the lattice mismatch in heterostructures. It is found that placing the single layer graphene on the h-BN substrate at an optimum interlayer distance results a bandgap of 59meV [12-13]. However this bandgap is not significantly large. In addition any symmetric bilayer strain cannot open a bandgap in a graphene. It is because the symmetrical strain does not change the hexagonal structure, although it changes the lattice constant. Therefore attention should be concentrated to asymmetrical or in-plane strain engineering. It may be predicted that if graphene (single/multilayered) is possible to grow to a substrate having sufficient lattice mismatched, necessary band gap could be achieved.

In this work, the effect of in-plane strain is studied in graphene FET (G-FET) using tight binding approximation. Structure of the device under study is shown in Fig. 1. The quantum capacitance is considered to develop the current-voltage characteristics under the influence of strain. The study is performed by varying various parameters such as gate voltage, channel length, oxide thickness, dielectric strength etc. The results obtained in the present study for strained G-FET show satisfactory performance that leads to utilize it in logic devices.

II. MODELING

A. Structure of the Device

Structure of the large area G-FET device under study is shown in Fig. 1. It is a single gate structure in which $1\mu\text{m}\times 1\mu\text{m}$ undoped graphene sheet is used as the channel material. HfO_2 is used as the gate dielectric that has the relative dielectric constant $\kappa=25$. Thickness of the gate insulator is taken to be $t_{\text{ins}}=2\text{nm}$. This study is carried out at room temperature.

B. Mathematical modeling

Under tight-binding approximation, the quantum behaviour of graphene under strain were studied by self-consistently solving the energy band structure in an atomistic basic set. The energy dispersion throughout the entire Brillouin zone of graphene is given by [14]:

$$E(k) = \pm [j_1^2 + 4j_2^2 \cos^2(\frac{p\pi}{n_w+1}) + 4j_1j_2 \cos(\frac{p\pi}{n_w+1}) \cos(\frac{3}{2}k_x a_{cc})]^{1/2} \quad (1)$$

Where $j_1 = j_0 / (1+\varepsilon)^2$ and $j_2 = j_0 / (1+\frac{\varepsilon}{4})$ with ε is the magnitude of strain. The j_0 and a_{cc} are the tight binding orbital overlap energy of the nearest C-C atom and the C-C bond length, respectively, when no strain is applied. Values of j_0 and a_{cc} are -2.74eV and 0.142nm respectively [14]. k_x is the wave vector component along the x-direction, n_w is the number of parallel lines across the width of graphene and p is the band index. W_{ch} is the width of large area graphene which is $1\mu\text{m}$. This width is related with n_w by the expression [15],

$$W_{ch} = (n_w + 1) \frac{\sqrt{3}}{2} a_{cc} \quad (2)$$

Equation (1) shows the energy dispersion relation in strained graphene, and that obeys the time reversal symmetry. As a result, the energy is separated between the conduction band bottom and valance band top at Γ points ($k_x=0$, central Brillouin zone) [16]. Thus the band gap energy can be derived as

$$E_{bg} = [j_1 + 2j_2 \cos(\frac{p\pi}{n_w+1})] \quad (3)$$

(1) can be further simplified into the following equation:

$$E(k) = \pm \sqrt{\frac{E_{bg}^2}{2} + \frac{9}{2} j_1 j_2 \cos(\frac{p\pi}{n_w+1}) k_x^2} \quad (4)$$

The quantum capacitance can be calculated as [17]:

$$C_Q = \frac{e^2}{4\pi} \frac{\sqrt{E_{bg}}}{\sqrt{\frac{9}{2} j_1 j_2 \cos(\frac{p\pi}{n_w+1})}} x^{-1/2} e^\eta \quad (5)$$

where $\eta = (E_F - E_{bg} / 2) / k_B T$ and $x = (E - E_{bg} / 2) / k_B T$.

A series combination of quantum capacitance, C_Q and insulator capacitance, C_{ins} can be considered to determine the total gate capacitance. The C_{ins} for graphene can be obtained and calculated by the following simple expression [18]:

$$C_{ins} = N\kappa\varepsilon_0 (\frac{W_{ch}}{t_{ins}} + \alpha_0) \quad (6)$$

where N is the number of gates used in the device ($N=1$ when only the top gate is considered and $N=2$ when both top gate and back gate is considered), κ is the relative dielectric constant of the oxide material used as gate insulator, t_{ins} is the thickness of the gate insulator, W_{ch} is the channel width, and $\alpha_0 \approx 1$ is a fitting parameter which has no dimension. So, the total gate capacitance can be obtained by-

$$C_G = \frac{C_Q C_{ins}}{C_Q + C_{ins}} \quad (7)$$

Now the device drain current I_{DS} can be expressed as [17]:

$$I_{DS} = \frac{\mu_{eff} C_G}{2L} \left[\frac{2(V_{GS} - V_T) V_{Deff} - V_{Deff}^2}{1 + \frac{V_{Deff}}{V_{Cr}}} \right] \quad (8)$$

where μ_{eff} is the effective charge carrier mobility, L is the channel length, V_{GS} is the gate to source voltage, V_T is the threshold voltage, V_{Deff} is the effective drain voltage and V_{Cr} is critical voltage. The details of this equation is available in [17].

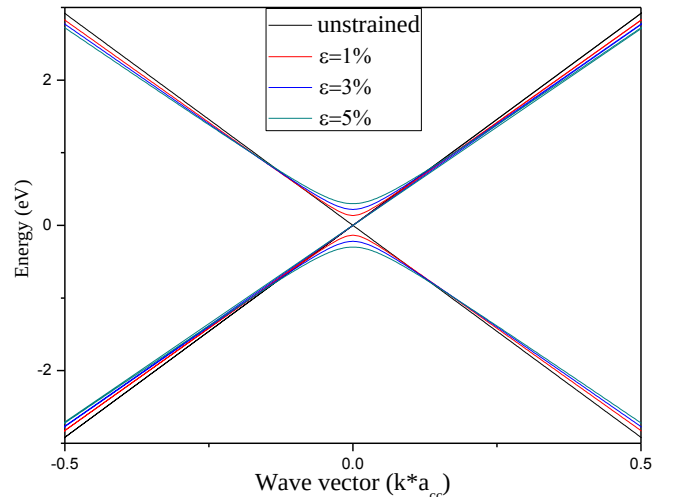


Fig. 2. Energy band structure of strained large area graphene.

III. RESULTS AND DISCUSSION

Fig. 2 shows the dispersion relation of large area graphene for different applied strain. It is seen that the band structure has a linear nature showing the presence of Dirac cone when no strain is applied. With increase in strain the band structure changes to parabolic shape. In addition to that the bandgap also increases with applied strain.

Fig. 3 shows the bandgap of large area graphene as a function of applied strain. It is seen that graphene has zero bandgap when no strain is applied. With the increase of in-plane strain the bandgap of graphene increases linearly. It is due to the fact that the Fermi level shifts in a direction perpendicular to k lines.

Fig. 4 shows the quantum capacitance as a function of the applied strain. As it is seen from the plot, quantum capacitance increases linearly with the increment of strain. The obtained values of quantum capacitance are low at lower strain because of low density of states that is characterized by the atomically thin quasi-one dimensional channel. And the density of states is further reduced due to quantum confinement boundary conditions in graphene's transverse direction.

Fig. 5 shows the effect of strain on I_D - V_D characteristics of the large area graphene FET. From (9), it can be seen that the drain current is proportional to the gate capacitance C_G . Since C_G is derived from the series combination of C_Q and C_{ins} and $C_Q \ll C_{ins}$, it can be said that C_G is almost equal to C_Q . And since C_Q is inversely proportional to j_1 and j_2 which are functions of strain, finally the drain current is inversely proportional to the applied strain. It is seen in Fig. 5 that when the applied strain is increased from 1% to 5% drain current gets reduced by almost 85%, from 2.8 mA to 0.4mA.

Fig. 6 shows the I_D - V_D characteristics at constant strain of 5% for different value of V_{GS} . It is seen that with the increment of the value of V_{GS} , the saturation current increases. This result is in agreement with the characteristic of conventional FETs.

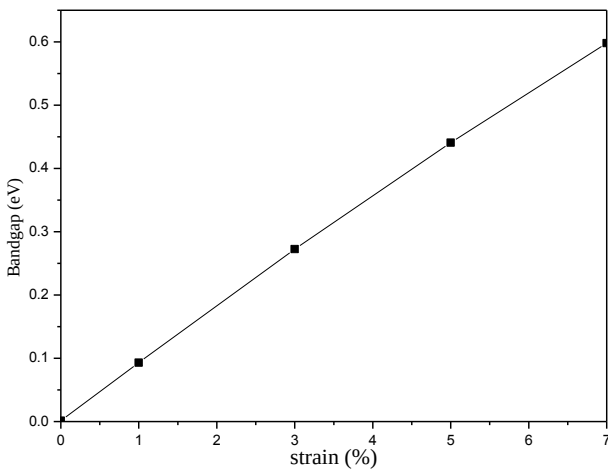


Fig. 3. Bandgap of large area graphene as a function of strain (%).

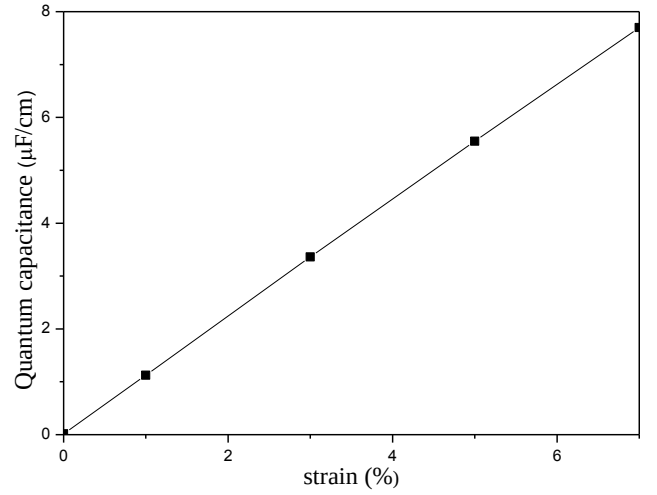


Fig. 4. The calculated quantum capacitance under in-plane strain.

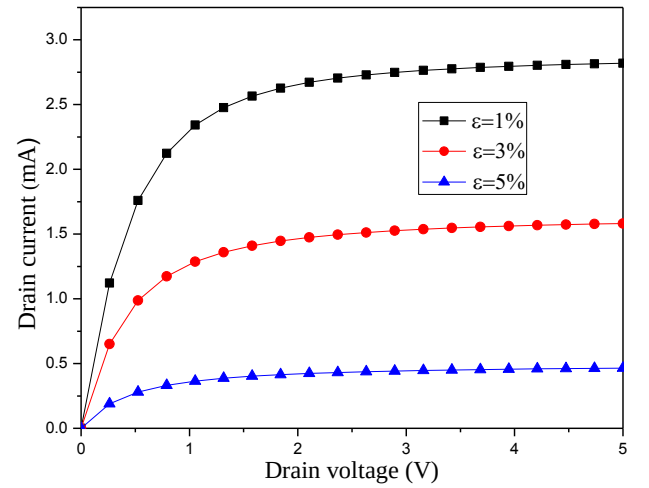


Fig. 5. I_D - V_D characteristics of strained GFET based on quantum calculation with $V_{GS}=0.9V$.

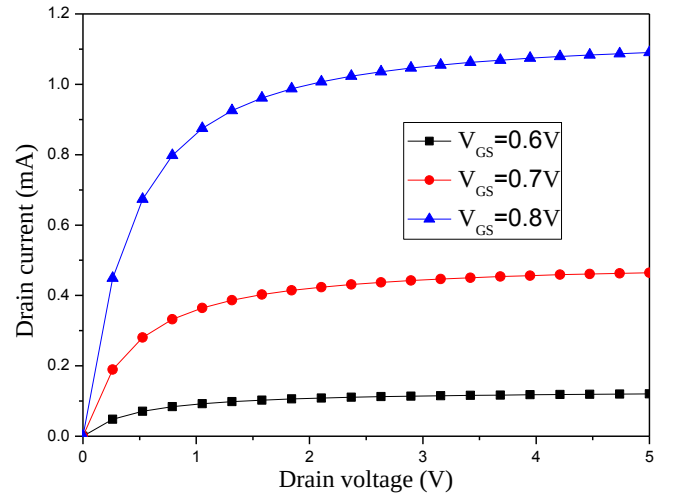


Fig. 6. I_D - V_D characteristics at constant strain of 5%.

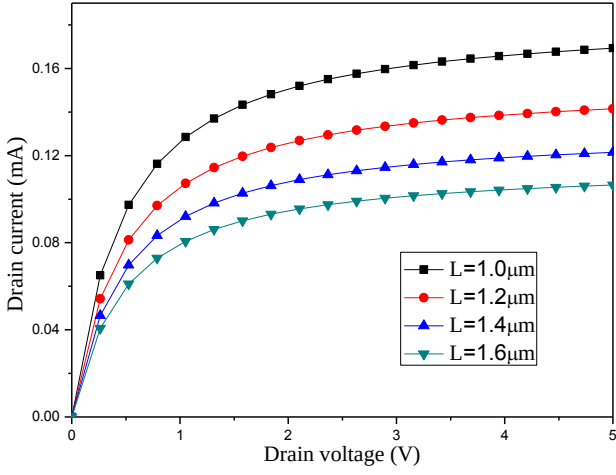


Fig. 7. I_D - V_D characteristics at $V_{GS}=0.9V$ with 5% strain for different channel length.

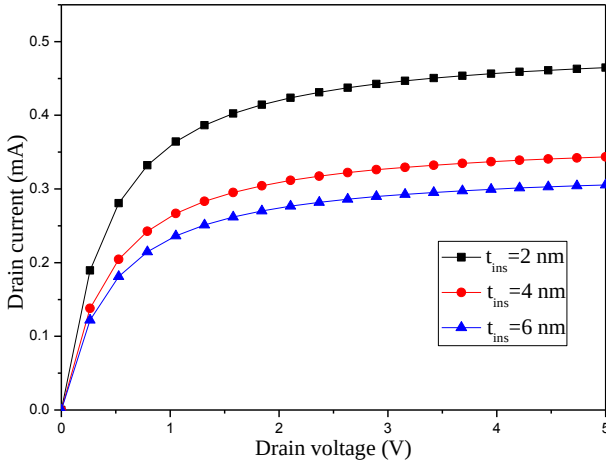


Fig. 8. I_D - V_D characteristics at $V_{GS}=0.9V$ with 5% strain for different oxide thickness.

Fig. 7 shows the effect of channel length on the I_D - V_D characteristic. The result is obtained at $V_{GS}=0.9V$ and at 5% applied strain. Since current is inversely proportional to channel length, the current decreases with the increment of length. In general, the reduction of channel length along with the implementation of ultrathin gate oxide would decrease the channel potential barrier that allows more electron injection from the source to the drain.

Fig. 8 shows the effect of oxide thickness on drain current. When the gate oxide becomes too thin, its capacitance becomes comparable with the quantum capacitance. Therefore, drain current is no longer dominated by quantum capacitance alone but by the series combination of C_G and C_Q . At that condition when the oxide thickness increases, the drain current decreases. Gate control of the inversion charges is weakened by the reduction of inversion layer capacitance in the gate capacitance model considering quantization effect.

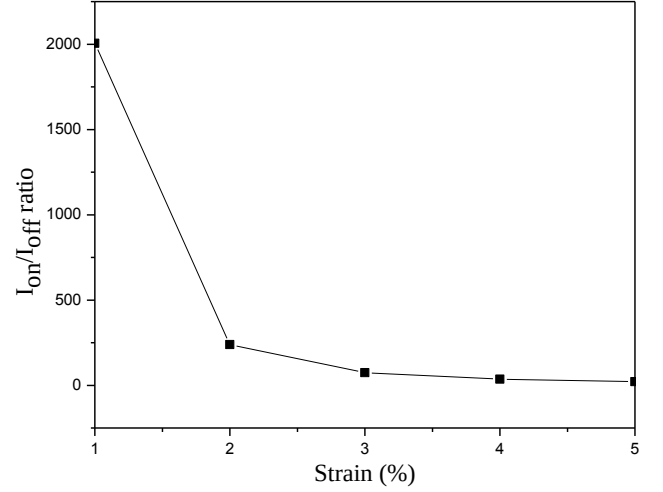


Fig. 9. I_{on}/I_{off} ratio of GFET device under in-plane strain.

Fig. 9 shows I_{on}/I_{off} ratio of GFET device under the influence of strain. As the strain increases, the on current is reduced. As a consequence, there is a significant reduction in the I_{on}/I_{off} ratio of G-FET device. It is found that when the value of strain is between 1% to 2%, the I_{on}/I_{off} ratio assumed to be acceptable to realize the logic device. Since the I_{on}/I_{off} ratio of large area G-FET is not reported previously, the comparison of I_{on}/I_{off} ratio is not included in this work.

IV. CONCLUSION

In this paper, the performance of large area G-FET is analyzed considering the effect of in-plane strain in a drain current model with quantum confinement. The dependence of drain current on gate voltage, strain, oxide thickness, channel length is discussed. The variation in the drain current with strain occurs due to the modification of electronic structure of single layer graphene channel. Introduction of strain increases bandgap and reduces drain current. The quantum capacitance is greatly dependent on strain i.e. bandgap. It is shown that strain plays the most significant role for promising performance of single layer graphene channel in logic devices.

REFERENCES

- [1] A.H. Castro Neto, F. Guinea, N.M.R. Peres, K.S. Novoselov, and A.K. Geim, "The electronic properties of graphene", *Reviews of Modern Physics*, vol. 81, no. 1, pp. 109–162, Jan. 2009.
- [2] M. F. Craciun, S. Russo, M. Yamamoto, and S. Tarucha, "Tunable electronic properties in graphene," *Nano Today*, vol. 6, no. 1, pp. 42–60, Jan. 2011.
- [3] M. Dragoman *et al.*, "Graphene for Microwaves," *IEEE Microwave Magazine*, vol. 11, no. 7, pp. 81–86, Dec. 2010.
- [4] M. Y. Han, B. Özyilmaz, Y. Zhang, and P. Kim, "Energy band-gap engineering of graphene nanoribbons," *Physical Review Letters*, vol. 98, no. 20, pp. 206805, May 2007.
- [5] M. Evaldsson, I. V. Zozoulenko, H. Xu, and T. Heinzel, "Edgedisorder-induced Anderson localization and conduction gap in graphene nanoribbons," *Physical Review B*, vol. 78, no. 16, pp. 161407, Oct. 2008.
- [6] F. Cervantes-Sodi, G. Csanyi, S. Piscanec, and A. C. Ferrari, "Edgefunctionalized and substitutionally doped graphene nanoribbons:

Electronic and spin properties,” *Physical Review B*, vol. 77, no. 16, pp. 165427, Apr. 2008.

- [7] E. V. Castro, K. S. Novoselov, S. V. Morozov, N. M. R. Peres, J. L. Santos, J. Nilsson and A. C. Neto (2007), “Biased bilayer graphene: semiconductor with a gap tunable by the electric field effect,” *Physical Review Letters*, vol. 99, no. 21, pp. 216802, Nov. 2007.
- [8] P. Gava, M. Lazzeri, A. M. Saitta, and F. Mauri, “Ab initio study of gap opening and screening effects in gated bilayer graphene,” *Physical Review B*, vol. 79, no. 16, pp. 165431, Apr. 2009.
- [9] 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, Sep. 2007.
- [10] A. Bostwick, T. Ohta, T. Seyller, K. Horn, and E. Rotenberg, “Quasiparticle dynamics in graphene,” *Nature Physics*, vol. 3, no. 1, pp. 36-40, Jan. 2007.
- [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, Oct. 2008.
- [12] M. H. Alam, K. I. Masood, and Q. D. M. Khosru, “Effect of biaxial strain on structural and electronic properties of graphene / boron nitride hetero bi-layer structure,” in *Proceedings of IEEE International Conference on Electrical and Computer Engineering (ICECE)*, Dec. 2014, p. 381-384.
- [13] Harihar Behera, and G. Mukhopadhyay, “Strain-tunable band gap in graphene/h-BN hetero-bilayer,” *Journal of Physics and Chemistry of Solids*, vol. 73, no. 7, pp. 818-821, Mar. 2012.
- [14] M. Han, Y. Zhang, and H.-B. Zheng, “Effect of uniaxial strain on band gap of armchair-edge graphene nanoribbons,” *Chinese Physics Letters*, vol. 27, no. 3, pp. 1-4, Aug. 2010.
- [15] M. T. Ahmadi, Z. Johari, D. C. Y. Chek, N. A. Amin, and R. Ismail, “Modelling of graphene nanoribbon Fermi energy,” *Journal of Nanomaterials*, vol. 2010, pp. 1-6, Mar. 2010.
- [16] D. Gunlycke and C. T. White, “Tight-binding energy dispersions of armchair-edge graphene nanostrips,” *Physical Review B*, vol. 77, no. 11, pp. 115116, Mar. 2008.
- [17] EngSiew Kang, and Razali Ismail, “Unified Drain Current Model of Armchair Graphene Nanoribbons with Uniaxial Strain and Quantum Effect,” *Journal of Nanomaterials*, vol. 2014, pp. 1-7, Apr. 2014.
- [18] J. Guo, Y. Yoon, and Y. Ouyang, “Gate electrostatics and quantum capacitance of graphene nanoribbons,” *Nano Letters*, vol. 7, no. 7, pp. 1935-1940, Jun. 2007.

Gate Oxide Dependent Performance of Graphene FET Using Quasi-Ballistic Transport Model

Md. Touhid Hossain*; Md. Nur Kutubul Alam; Md. Rafiqul Islam;

Department of Electrical and Electronic Engineering

Khulna University of Engineering & Technology

Khulna-9203, Bangladesh

Email*: touhid1503@gmail.com

Abstract— A quasi-ballistic transport model for graphene FET (GFET) is analyzed in this work. The effect of top and back gate oxides with different dielectrics are considered to analyze the device performance. A comparative study considering the equivalent oxide thickness (EOT) for the different oxide layers is done using the proposed model. The results obtained from the simulation demonstrate that the use of high-k dielectrics in the oxide layers and equivalent oxide thickness improve the output drain current significantly.

Keywords—Quasi-ballistic transport model, GFET, EOT, Drain current.

I. INTRODUCTION

One of the promising candidates as a material for future electronics is graphene which is structurally a thin sheet of graphitic carbon. With the continuous scaling down of metal oxide semiconductor field-effect transistor (MOSFET) it suffers from various short channel effects. Geometric design modification is no longer enough for solving the problem, new material with better transport properties is necessary. Graphene is considered to be one of the most favorable materials for the channel in FETs in the near future [1]. The major advantages are its high mobility, excellent chemical and mechanical stabilities and 2-D structure, that enable a perspective usage of graphene. Intensive experimental researches are being conducted worldwide on transistors using graphene as the channel material [2]-[4]. The 2-D structure makes extremely thin sheets of graphene, highly conductive, and it makes it possible to fabricate extended active areas of arbitrary length and width. Graphene is also unique in that it is a semimetal with a zero band gap and in that the electrons have a zero effective mass (m_e). While the zero m_e is advantageous for graphene transistors, the lack of a band gap is problematic, because it prevents the devices from turning off below a fairly high threshold conduction [3]. Since large-area graphene has no band gap, graphene MOSFETs with large-area graphene channels do not switch off and are not suited for logic applications. However, graphene could still be very useful in analog and radio-frequency (RF) applications where high ON/OFF current ratios are not required. For radio frequency (RF) transistors, on the other hand, complete switch

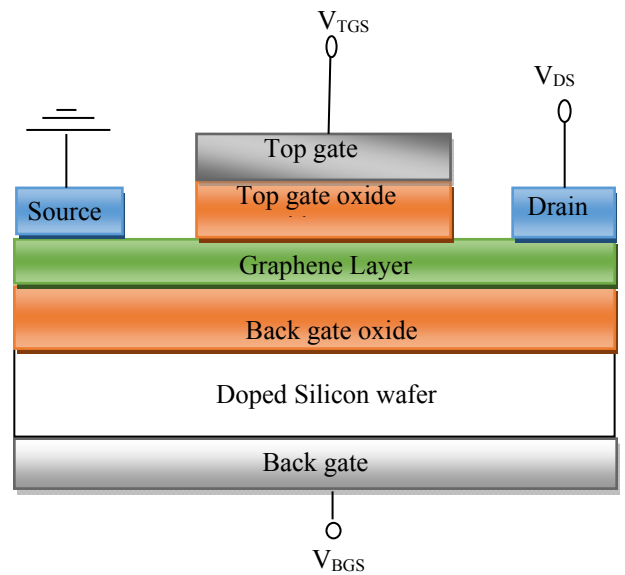


Fig. 1. Sketch map of a double-gate (DG) GFET.

off is not mandatory [5]. Recently, RF graphene MOSFETs with large-area graphene channels showing cutoff frequencies f_T in the Gigahertz range have been realized and a record f_T of 100 GHz has been reported for a 240-nm gate transistor [6].

In this paper, a quasi-ballistic transport model for graphene FET (GFET) is studied to analyze the device performance. The use of different oxide layers with different dielectric constants as back and top gate materials of the graphene FET is studied to analyze their effects on the drain current of the GFET. The equivalent oxide thickness (EOT) for the different oxide layers used in the model is considered later for a comparative study to see how it affects the device performance. The simulated results demonstrate that the use of high value dielectrics as oxide layers and using equivalent oxide thickness increase the drain current significantly.

The paper is organized as follows. In sec. II, a basic demonstration on a quasi-ballistic transport model of GFETs and theory of equivalent oxide thickness are presented and sec. III presents the simulated results and comparisons with discussion. Finally sec. IV concludes the paper.

II. PROPOSED TRANSPORT MODEL

A. Drain Current Model

This drain-current transport model is taken from [1]. In this proposed model, McKelvey's flux theory [8] and quasi-ballistic transport model (BTM) are used to obtain an explicit analytical drain current expression for GFETs. The BTM was used in the studies of double-gate (DG) and nanowire MOSFET, where the channel material was silicon [9], [10]. Further, the BTM approach was used to obtain a unified carrier transport model, which was applicable in both ballistic and diffusive-transport regimes [11]. We have used this model to study the dielectric strength dependent drain current of GFET.

The sketch of a GFET is shown in Fig. 1, where a layer of graphene with a channel length L , and a channel width W , is placed on a back-gate oxide. The biases applied to the top gate, back gate and the drain are V_{TGS} , V_{BGS} , and V_{DS} respectively, and the electric potential at the source is referenced to zero. A brief summary of the mathematical manipulation to obtain the drain current expression for the proposed model is as follows.

Flux is defined in this paper as the carriers pass through the channel per unit width per second, and in the units of $m^{-1}s^{-1}$. The drain-to-source current, I_{ds} , is given by charges carried by net fluxes between the positive- and the negative-going fluxes as follows [1]:

$$I_{ds} = Wq[f^+(0) - f^-(0)] \quad (1)$$

where q is the elementary charge, $f^+(0)$ and $f^-(0)$ are positive and negative-going carrier fluxes respectively, W is the width of the channel. The carriers' sheet density at the beginning of the channel is given by [11]

$$n(0) = \frac{[f^+(0) - f^-(0)]}{v_F} \quad (2)$$

where, v_F is the Fermi velocity, and $v_F = 9.71 \times 10^5 m/s$ [12]. Then, $f^+(0)$ and $f^-(0)$ can be solved by substituting them into (1), the drain current is as follows:

$$I_{ds} = \frac{qWv_F n(0)(1-r)[1 - \zeta_1(\eta f - U_{ds}) / \zeta_1(\eta f)]}{1 + r + (1-r)\zeta_1(\eta f - U_{ds}) / \zeta_1(\eta f)} \quad (3)$$

Where, r is the backscattering co-efficient [9]. Here, $\zeta_1(\eta f)$ is the Fermi integral. And U_{ds} can be found as

$$U_{ds} = \frac{qV_{ds}}{k_B T} \quad (3)$$

where, V_{ds} is the drain to source voltage, k_B and T have their usual meaning. The details about (3) can be found in [1]. To obtain I_{ds} , $n(0)$ is needed, which is expressed as follows [5],[6]:

$$n(0) = \frac{KV_C^2}{2q} + n_i \quad (4)$$

$$\text{where } K = \left(\frac{2q^3}{\pi}\right)(\hbar v_F)^2 \quad (5)$$

n_i is the induced carrier sheet density [6], treated as a fitting parameter and it is taken as $n_i = 10^{12} cm^{-2}$. V_C is the voltage drop across the quantum capacitance [5] and is given by [5], [6] as follows:

$$V_C = \frac{\sqrt{((c_t + c_b)^2 + 2K[v_t c_t + v_b c_b])} - (c_t + c_b)}{K} \quad (6)$$

where, c_t and c_b are the top and back oxide capacitance respectively. v_t and v_b are the effective voltages applied to the top and back gate, respectively. Details of these terms can be found in [1]. Equation (3) is used for the calculation of drain current.

B. Theory of Equivalent Oxide Thickness

An Equivalent oxide thickness (EOT) is a distance, usually given in nanometers (nm), which indicates how thick a silicon oxide film would need to be to produce the same effect as the high dielectric constant material being used. The same effect means the same value of capacitance produced. So it can be found through following mathematical manipulation:

$$C_{OX(SiO_2)} = C_{OX(high-k)} \quad (7)$$

where, C_{OX} is used to designate the capacitance of the oxide layer. That is,

$$\frac{\epsilon_{OX(SiO_2)}}{t_{OX(SiO_2)}} = \frac{\epsilon_{OX(high-k)}}{t_{OX(high-k)}} \quad (8)$$

where, ϵ_{OX} and t_{OX} are, respectively, the dielectric constant and oxide thickness. So finally the formula to obtain equivalent oxide thickness is as follows:

$$EOT(t_{OX[SiO_2]}) = t_{OX(high-k)} \frac{\epsilon_{OX(SiO_2)}}{\epsilon_{OX(high-k)}} \quad (9)$$

III. RESULTS AND DISCUSSION

At first, the proposed model is used to find the transfer and output characteristics of the GFET. The device parameters taken as $L=10 \mu m$ and $W=5 \mu m$. The transfer characteristics considering different dielectric oxide layers and thickness are shown in fig. 2 and fig. 3, respectively. It is observed in fig. 2 that using of oxides that have higher dielectric constant gives higher drain current. On the other hand fig. 3 shows that the lower oxide thickness gives higher drain current. Higher dielectric strength as well as small thick oxide layer gives better control of gate over the channel. Fig. 4 shows the output characteristics neglecting the back gate oxide capacitance. Fig. 5 shows the same taking the back gate oxide capacitance into account. The above two figures demonstrate that the consideration of back gate oxide capacitance influences the drain current by slightly reducing it. This reduction of current can be explained by (7), which shows that V_C reduces when back gate oxide capacitance is considered resulting in reduction of the drain current. Fig. 6 shows the direct analysis of how oxide thickness influences the device performance.

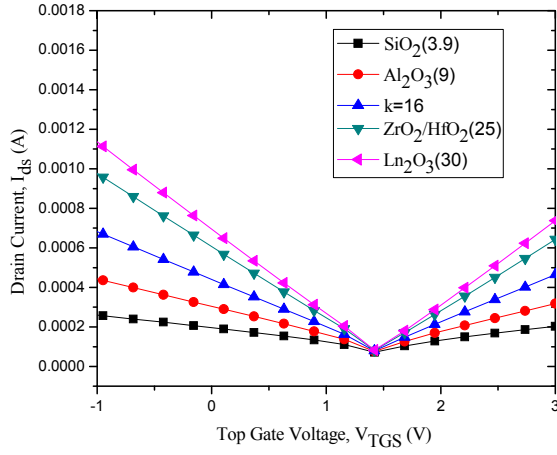


Fig. 2. Transfer characteristics for the GFET with different dielectric oxide layers ($V_{DS}=1$ V).

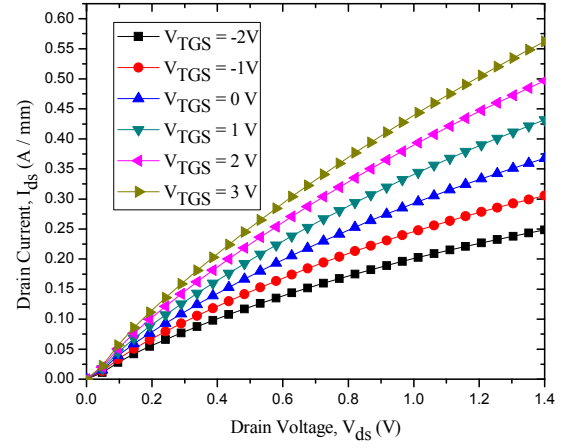


Fig. 5. Output characteristics of GFET considering back gate oxide capacitance.

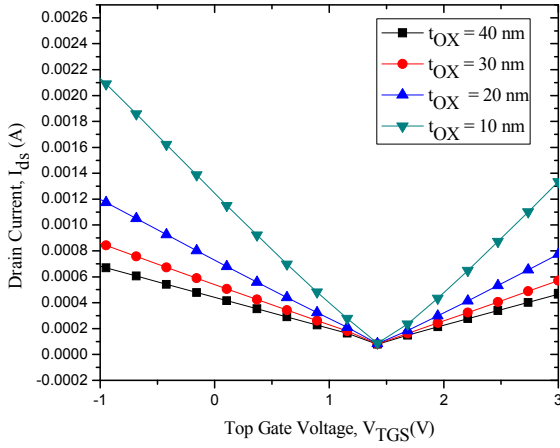


Fig. 3. Transfer characteristics for the GFET with different oxide thickness (t_{OX} = oxide thickness, drain voltage $V_{DS}=1$ V).

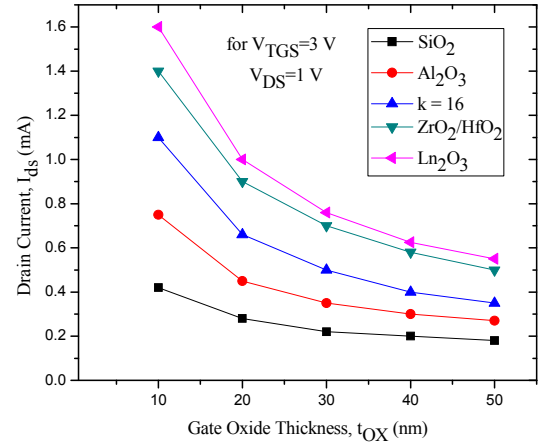


Fig. 6. Drain current (I_{DS}) versus gate oxide thickness (t_{OX}) of GFET (V_{TGS} =top gate to source voltage, V_{DS} =drain to source voltage).

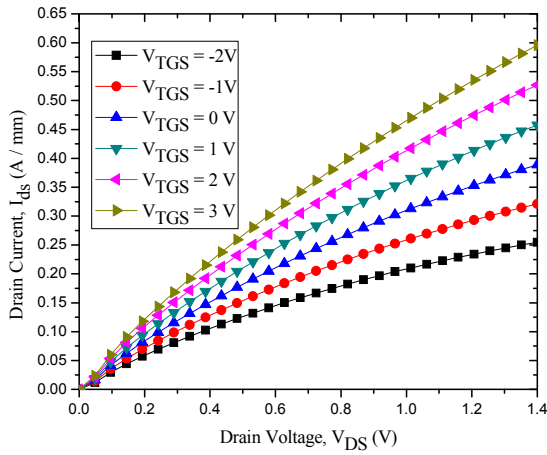


Fig. 4. Output characteristics of GFET neglecting back gate oxide capacitance.

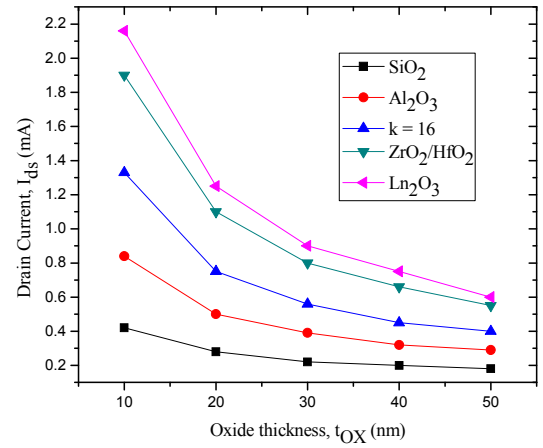


Fig. 7. Drain current (I_{DS}) versus oxide thickness (t_{OX}) for equivalent oxide thickness layer.

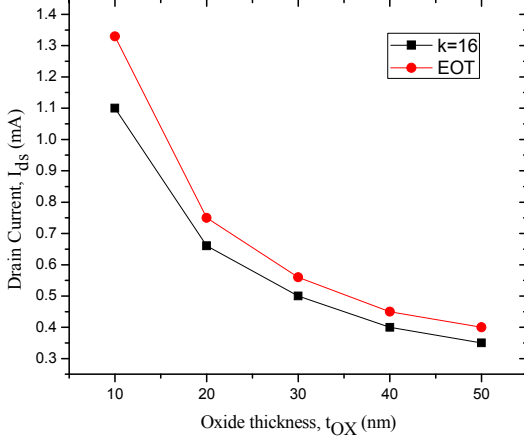


Fig. 8(a).

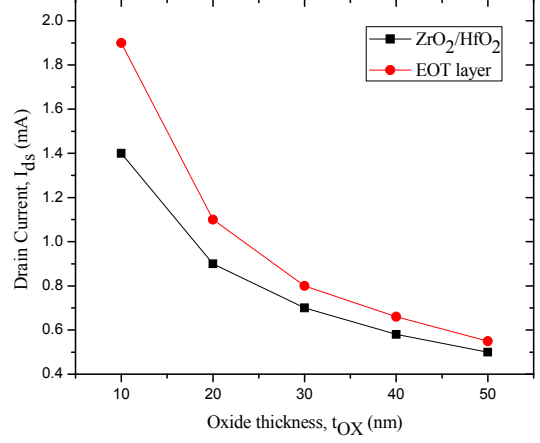


Fig. 8(c).

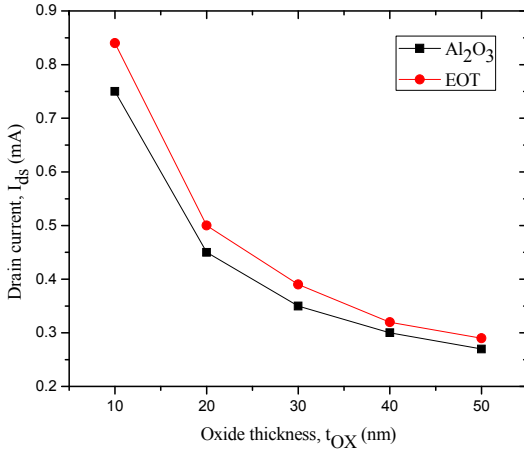


Fig. 8(b).

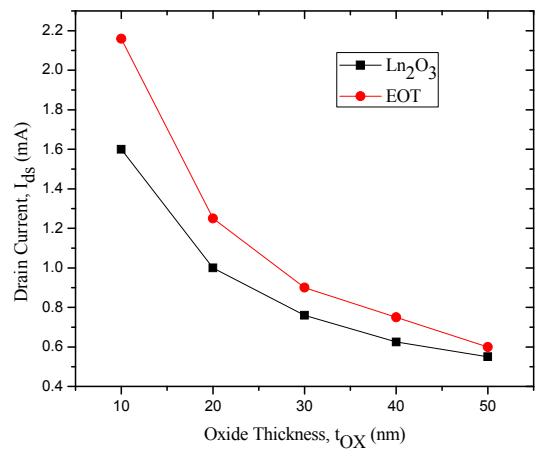


Fig. 8(d).

It is found that as the thickness increases the output drain current decreases. Also based on the analysis using different dielectrics it is found that high value of dielectric constant helps increasing drain current significantly.

Fig. 7 shows that when the oxide layer is replaced by its equivalent oxide thickness, more drain current is obtained. It is because high-k dielectric results the equivalent oxide with small thickness and when oxide thickness is less drain current is more which is also observed earlier.

A comparative analysis between equivalent oxide thickness and actual oxide thickness showing how it influences the drain current for different dielectric oxides is shown in figs. 8(a)-(d). The figures show that there is significant increase in the drain current for equivalent oxide thickness than with actual oxide layers. It is cleared from (3), (4), and (6) that the drain current is proportional to V_C which is inversely proportional to oxide thickness. When a particular oxide thickness having dielectric strength more than SiO_2 , its EOT will be smaller than the actual thickness. Therefore, drain current for high-k oxides will be higher when it is calculated with respect to EOT.

Fig. 8. Drain current vs oxide thickness using actual oxide thickness and equivalent oxide thickness for (a) $k=16$, (b) Al_2O_3 ($k=9$), (c) ZrO_2/HfO_2 ($k=25$) and (d) Ln_2O_3 ($k=30$).

IV. CONCLUSION

Performance analysis of GFET in this paper demonstrates satisfactory I-V characteristics which are needed for integrated circuit (IC) simulation and design. Also an insight on how the oxide layers influence the drain current is pointed out from the simulated results. The results demonstrate that for better drain current performance, the lower thickness of oxide layers are suitable. It is also found that the higher dielectrics as oxide layers increases the drain current significantly and using equivalent oxide thickness of oxide layers improves device performance. From the present study it can be concluded that the device performance is strongly related with GFET oxide material's dielectric strength and its EOT.

References

- [1] Guangxi Hu, Shuyan Hu, Ran Liu, Lingli Wang, Xing Zhou, and Ting-Ao Tang, "Quasi-ballistic transport model for graphene field-effect transistor," *IEEE Transactions on Electron Devices*, vol. 60, no. 7, July 2013.
- [2] I. Meric, M. Y. Han, A. F. Young, B. Ozyilmaz, P. Kim, and K. Shepard, "Current saturation in zero bandgap, top-gated graphene field effect transistors," *Nature Nanotechnol.*, vol. 3, no. 11, pp. 654–659, 2008.
- [3] J. Kedzierski, P.-L. Hsu, A. Reina, J. Kong, P. Healey, P. Wyatt, and C. Keast, "Graphene-on-insulator transistors made using C on Ni chemical-vapor deposition," *IEEE Electron Device Lett.*, vol. 30, no. 7, pp. 745–747, Jul, 2009.
- [4] W. J. Liu, X. W. Sun, Z. Fang, Z. R. Wang, X. A. Tran, F. Wang, L. Wu, G. I. Ng, J. F. Zhang, J. Wei, H. L. Zhu, and H. Y. Yu, "Positive bias-induced V_{th} instability in graphene field effect transistors," *IEEE Electron Device Lett.*, vol. 33, no. 3, pp. 339–341, Mar, 2012.
- [5] D. Jiménez and O. Moldovan, "Explicit drain-current model of graphene field-effect transistors targeting analog and radio-frequency applications," *IEEE Trans. Electron Devices*, vol. 58, no. 11, pp. 4049–4052, Nov, 2011.
- [6] S. A. Thiele, J. A. Schaefer, and F. Schwierz, "Modeling of graphene metal-oxide-semiconductor field-effect transistors with gapless large-area graphene channels," *J. Appl. Phys.*, vol. 107, no. 9, pp. 094505-1–094505-8, May, 2010.
- [7] H. Wang, A. Hsu, J. Kong, D. A. Antoniadis, and T. Palacios, "Compact virtual-source current-voltage model for top- and back-gated graphene field-effect transistors," *IEEE Trans. Electron Devices*, vol. 58, no. 5, pp. 1523–1533, May, 2011.
- [8] J. P. McKelvey, R. L. Longini, and T. P. Brody, "Alternative approach to the solution of added carrier transport problems in semiconductors," *Phys. Rev.*, vol. 123, no. 1, pp. 51–57, Jul. 1961.
- [9] A. Rahman and M. S. Lundstrom, "A compact scattering model for the nanoscale double-gate MSFET," *IEEE Trans. Electron Devices*, vol. 49, no.3, pp.481–489, Mar. 2002.
- [10] M. Lundstrom, Z. Ren, and S. Datta, "Essential physics of carrier transport in nanoscale MOSFETs," *IEEE Trans. Electron Devices*, vol. 49, no.1, pp.133–141, Jan. 2002.
- [11] G. Hu, J. Gu, S. Hu, Y. Ding, R. Liu, and T.-A. Tang, "A unified carrier-transport model for the nanoscale surrounding-gate MOSFET comprising quantum-mechanical effects," *IEEE Trans. Electron Devices*, vol.58, no.7, pp.1830-1836, Jul.2011.
- [12] G. W. Hanson, *Fundamentals of Nanoelectronics*. Upper Saddle River, NJ,USA:Pearson Education,Inc.,2008.

Analytical Modelling of Cds/CdTe Based Thin Film Solar Cells Considering Surface Recombination

¹Sanimul Alam Rivon, ²Protik Biswas, ³Abdullah Sakib Hasan, ⁴Ashiq Been Rahim, ⁵Md. Iqbal Bahar Chowdhury

Department of Electrical and Electronics Engineering (EEE)

United International University, Dhaka, Bangladesh

Email: ¹sanimulalamrivon@gmail.com, ²protik_biswas@yahoo.com, ³ovi_uui05@yahoo.com, ⁴real619@hotmail.com, ⁵ibchy@eee.uui.ac.bd

Abstract— This work develops an analytical model including the effect of series resistance as well as the surface recombination occurred at the edges. Mathematical intractability caused for the inclusion of series resistance has been resolved using the concept of the perturbation theory. Introducing a new boundary condition at the emitter edge of the fully depleted absorber region, the effects of surface recombination can be included in the analysis. The proposed model not only shows the importance of inclusion of surface recombination in the analysis but also explains the effects of variation of series resistance and absorber width and the effects of stress as well. Therefore, the proposed model demonstrates the underlying physics of CdS/CdTe solar cell better.

Keywords—Surface Recombination, Series Resistance, CdS/CdTe solar cell, Ni/Cu bilayer contact, Stress Effect, Absorber Width.

I. INTRODUCTION

Among the various thin film solar cells presently available in the world market, CdS/CdTe based solar cell is the most promising candidate for large-scale photovoltaic applications because of its high potential to realize low cost, high efficiency, reliable and stable solar cells. Owing to its direct bandgap of 1.45 eV, CdTe is suitable to absorb higher percentages of solar irradiations under AM1.0-AM1.5 conditions. Experimental and theoretical efficiency of these solar cells are reported as about 16-17% and more than 28% [3]. Usually, these solar cells are manufactured by depositing CdTe on glass as absorber and using a very thin ($\approx 0.1 \mu\text{m}$) [4], [5] heavily doped n-type CdS as emitter material.

In order to model the current-voltage characteristics as well as the behaviour of CdS/CdTe based thin film solar cells, several numerical and/or analytical models [6]–[10] have been devised in the literature. Of these, models in [6], [7] are fully numerical; semi-analytical models (as photogenerated current is deduced analytically, but numerical iterations are used to determine the load current) considering the effect of series

resistance are presented in [8], [9]. Fully analytical models are devised in [10]–[14] based on the concept of perturbation theory. Almost all these models considered the voltage-dependency of electric field in the absorber region; few [9], [12]–[14] considered space-dependency too. However, all these models failed to incorporate the effect of surface recombination in their analysis due to the neglect of diffusion current component in the absorber region for which an additional boundary condition is needed. Indeed, surface recombination need to be analysed in the modelling, as this degrades the solar cell performance severely.

In this work, the surface recombination effect is included in the modelling by considering the diffusion component of the photogenerated current in the field-dominated absorber region and also, by introducing a new boundary condition to resolve the mathematical intractability. The derived model is applied under various conditions to testify its validity. In doing so, the effect of surface recombination, series resistance, absorber width and stress on the J-V curves are analysed.

II. ANALYSIS

For any solar cell, the load current density (J_L) including the effect of series and shunt resistance as a function of the load voltage (V_L) can be expressed as

$$J_L(V_L) = -J_{ph}(V_L) + J_{01} \left(e^{\frac{V_L - J_L R_s}{V_T}} - 1 \right) + J_{02} \left(e^{\frac{V_L - J_L R_s}{nV_T}} - 1 \right) + \frac{V_L - J_L R_s}{R_{sh}} \quad (1)$$

where, J_{ph} is the photogenerated current density, J_{01} , J_{02} are the dark saturation current density in the emitter and absorber region respectively, R_s is the series resistance, R_{sh} is the shunt resistance, n is the ideality factor and V_T is the thermal voltage. The expressions for J_{01} and J_{02} are given as

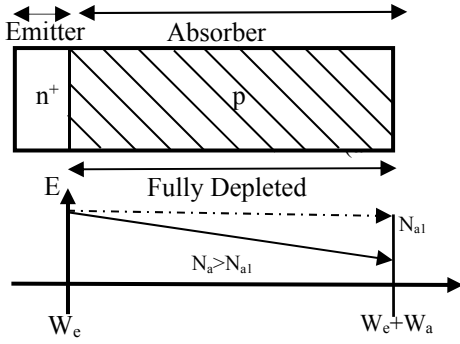


Fig. 1. Electric Field (E) Profile [11]

$$J_{01} = \frac{qn_i^2}{D_n}, \quad J_{02} = \frac{qn_i W_a}{2\sqrt{\tau_n \tau_p}}$$

where, q is the electronic charge, n_i is the intrinsic carrier concentration, D_n is the diffusivity of electrons, τ_n and τ_p are the lifetimes of electrons and holes in the absorber region respectively and W_a is the width of the absorber layer. The photogenerated current J_{ph} can be obtained as

$$J_{ph} = J_p(W_e) + J_n(0) \quad (2)$$

where, W_e is the emitter width, $J_p(W_e)$ is the hole photogenerated current density in the emitter region at the depletion layer edge (i.e. $x = W_e$) and $J_n(0)$ is the electron photogenerated current density in the absorber region at $x = 0$. Here $x = 0$ denotes the depletion layer edge of the absorber region. $J_p(W_e)$ can be obtained using the expression derived in [13]. For $J_n(0)$, an analytical model has been derived in this work.

Fig. (1) shows a structure of thin film CdS/CdTe based heterojunction solar cell. The emitter region of this cell is heavily doped (of the order of 10^{18} cm^{-3}) and the absorber layer is lightly doped (of the order of 10^{14} cm^{-3}). Therefore, the absorber region is fully depleted. In addition, the spatial variation of the electric field in the absorber region is negligible and hence, the electric field E in the absorber region can be assumed as constant. From [7] and considering the heterojunction structure, this field can be expressed as

$$E = \frac{\epsilon_1 N_d}{\epsilon_1 N_d + \epsilon_2 N_a} \left[\frac{V_{bi} - (V_L - J_L R_s)}{W_a} \right] \quad (3)$$

Where, ϵ_1 (ϵ_2) is the permittivity and N_d (N_a) is the doping level of the emitter (absorber) layer, R_s is the series resistance and V_{bi} is the built in voltage across the absorber layer. The basic equations needed to deduce the electron current density $J_n(x)$ due to photogeneration process can be given as

$$J_n(x) = qn(x)\mu_n E + qD_n \frac{dn}{dx} \quad (4)$$

$$\frac{dJ_n(x)}{dx} = \frac{qn(x)}{\tau_n} - q\alpha(\lambda)G_0(\lambda)e^{-\alpha(\lambda)x} \quad (5)$$

Where, $n(x)$ is electron density, E is the electric field given by Eqn. (3), μ_n and τ_n are the mobility and lifetime of electrons in

the absorber region, α is the wavelength-dependent absorption coefficient of the absorber material (CdTe) and G_0 is the optical generation rate given by

$$G_0(\lambda) = (1 - R) \frac{\lambda I_0(\lambda)}{hc} \quad (6)$$

Where, h is Planck's constant, c is the velocity of light, $I_0(\lambda)$ is the intensity of solar irradiation, $R(\lambda)$ is the reflection coefficient and λ is the wavelength of the incident light. Conventionally, the diffusion component of the electron current ($qD_n \frac{dn}{dx}$) is assumed as negligible owing to the presence of electric field therein. However, in this work, this diffusion component is considered negligible only at the depletion edge i.e. $x = 0$, as electric field becomes maximum at $x = 0$ [Fig. (1)]. This leads to a new boundary condition at $x = 0$ which can be given as

$$\frac{dn}{dx}(0) \approx 0 \quad (7)$$

Combining Eqns. (4) and (5), a second order differential equation (DE) of $n(x)$ can be obtained as,

$$\frac{d^2 n}{dx^2} + C \frac{dn}{dx} - \frac{n}{L_n^2} = -G e^{-\alpha_2 x} \quad (8)$$

Where,

$$C = \frac{E}{V_T}, \quad G = \frac{G_0 \alpha_2}{D_n} e^{-\alpha_1 W_e}, \quad G_0 = (1 - R) \frac{\lambda I_0}{hc}, \quad L_n = \sqrt{D_n \tau_n}$$

The solution of the DE given by Eqn. (7) can be obtained as

$$n(x) = C_1 e^{-\left(\frac{C}{2} - m_0\right)x} + C_2 e^{-\left(\frac{C}{2} + m_0\right)x} - A e^{-\alpha_2 x} \quad (9)$$

Where,

$$A = \frac{G_0(\lambda)}{\alpha_2^2 - C\alpha_2 - \frac{1}{L_n^2}}$$

and C_1 , C_2 are the constants to be determined from the boundary conditions. One boundary condition is given in Eqn. (7). For the other boundary condition, the surface recombination at $x = W_a$ can be considered instead of $n(W_a) = 0$ used in [8]-[12]. In this way, the effect of surface recombination can be included in the analysis which was not possible in the works [8]-[12]. This second boundary condition can be expressed as [13]

$$J_n(w_a) = qS_n n(w_a) \quad (10)$$

Using the boundary conditions given by Eqns. (7) and (10), C_1 and C_2 can be determined from which $n(0)$ and $J_n(0)$ can be calculated as

$$n(0) = C_1 + C_2 \quad (11)$$

$$J_n(0) = qD_n n(0) \cdot \frac{E_n}{V_T} \quad (12)$$

Therefore, J_{ph} can be calculated from Eqn. (2). However, consideration of series resistance R_s in the analysis makes the determination of J_L using Eqn. (1) intractable. Following the procedure based on the application of the concept of the

perturbation theory in the works [10]-[12], an analytical derivation of J_L at any load voltage V_L using Eqn. (1) can be made possible.

III. RESULTS AND DISCUSSIONS

This section presents and analyzes the results carried away by the MATLAB simulations based on the proposed model. AM1.5 global spectrum obtained from the ASTM G-173-03 standard [14] is used for the incident photon flux and the absorption coefficients of CdS and CdTe are taken from [15]. The widths of CdS and CdTe layers are chosen as 0.1 and 3.5 μm respectively.

Fig. (2) shows the current-voltage (J-V) characteristics curve obtained from the simulation using the proposed model for a CdS/CdTe based solar cell investigated by Hegedus *et al.* [6] having $R_s = 1.2 \Omega \text{ cm}^2$, $n = 1.7$ and $J_0 = 8 \times 10^{-8} \text{ mA-cm}^{-2}$. Simulations results show an open circuit voltage of 0.84 V and a short circuit current density of 17.82 mA-cm^{-2} which closely match with [6].

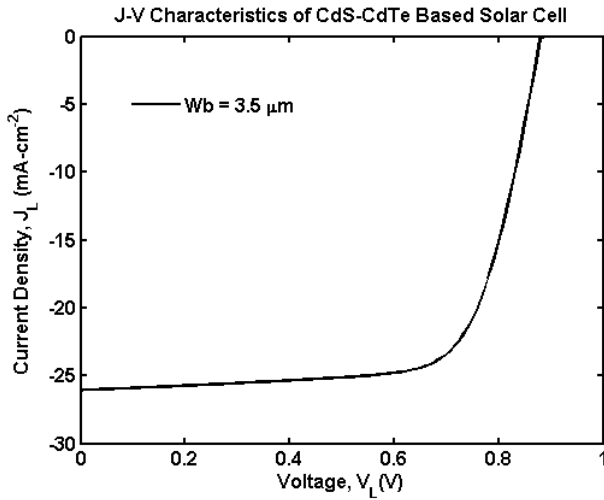


Fig. 2. J-V Curves obtained using the proposed model for the CdS/CdTe based solar cell investigated in [6].

Fig. (3) shows the effect of surface recombination (S_n) on the J-V characteristics of CdTe based solar cell. The J-V curve for $S_n = 0.0001 \text{ cm/sec}$ compared to other values of S_n can be considered as the J-V curve without considering surface recombination. From this figure, it is evident that J-V curves are strong function of surface recombination. This is expected as significant number of carriers are lost at the edge owing to the surface recombination and in order to alleviate this problem several technologies are employed at the present times which include the use of thin molybdenum layer at the edge, use of back surface field etc. The models of the previous works [8]-[12] lack the inclusion of this effect. Fig. (3) also shows that surface recombination has very little effect on the open circuit voltage (V_{OC}), as V_{OC} reduction is mainly due to the volume recombination processes like SRH and Auger recombination mechanisms occurred inside the absorber region. However,

from Fig. (3), it is evident that inclusion of surface recombination in the model is necessary.

Fig. (4) shows the effect of series resistance R_s on the J-V characteristics of CdTe based solar cell. This figure demonstrates that R_s causes decrease in the short circuit current, degrades the J-V curves close to the open circuit voltage and has no effect on the open circuit voltage. Indeed, R_s places an additional forward bias even under short circuit condition, which results in increase of the dark current and hence, causes a reduction in the overall current. Although this increase in dark current is insignificant under short circuit condition (hence, the decrease in short circuit current is small), this increment becomes severe as the load voltage increases. Therefore, an increased reduction is observed at load voltages close to the open circuit voltage. However, R_s has no effect on the volume recombination processes and hence, the open circuit voltage is not affected by the increase of R_s .

Fig (5) shows effect of absorber width variation on J-V curves of CdTe based solar cell. Increase in the absorber width causes more photogeneration as more photons can be absorbed in the increased absorber region and hence, photogenerated current will be increased. However, increase in the absorber width increases the probability of volume recombination in the absorber region thereby causing carrier loss as well as reduction in the open circuit voltage. Therefore, increase in absorber width surely decreases the open circuit voltage, but may increase or decrease in the short circuit current depending on whether the absorption process dominates the recombination process or not. In accordance with this analysis, Fig. (5) shows lower open circuit voltage as the absorber width increases and reveals that the recombination process is the dominant mechanism in the absorber region.

In CdTe based solar cells, front-end contact plays an important role, especially under stressed condition. Usually, a Cu/Ni bilayer contact (15 nm/50 nm) is used in the CdS/CdTe based solar cells. Under stressed condition this contact results in degradation of the solar cell performance. Therefore, the effect of stress should be investigated by the models. Fig. (6) shows this stress effect on the J-V curves of CdTe based solar cell having 7 μm thick CdTe layer. The stress condition is simulated in the work of Hegedus *et al.* [6] by leaving it open-circuited at 85 $^{\circ}$ C in dry air with 1 sun condition for 15 days. The change in the values of the ideality factor (n), series resistance (R_s) and the dark saturation current (J_0) are observed and simulated in the analysis ($n=1.66$, $R_s= 2.4 \Omega \text{ cm}^2$, $J_0=1.5 \times 10^{-7} \text{ mA cm}^{-2}$ before stress and $n=1.58$, $R_s= 3.42 \Omega \text{ cm}^2$, $J_0=2 \times 10^{-7} \text{ mA cm}^{-2}$ after stress). The simulation results shown in Table 1 for various performance parameters are in well agreement with those in [6]. Indeed, the decreased value of J_{sc} , V_{OC} , fill factor and conversion efficiency under stressed condition can be attributed to the increased value in R_s and J_0 under stress.

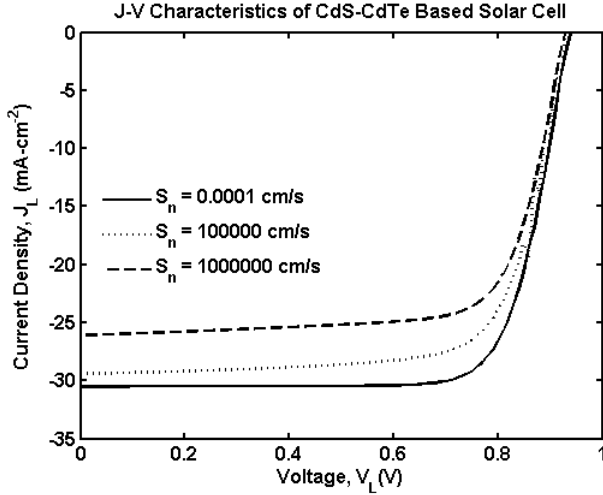


Fig. 4. J-V characteristics curve against the variation of surface recombination (S_n).

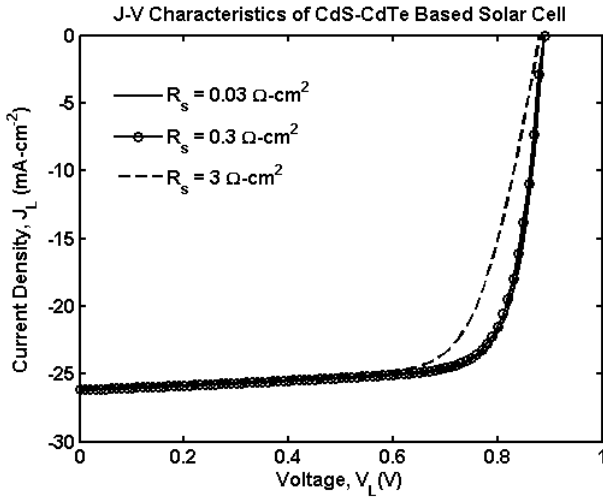


Fig. 3. J-V characteristics curve under the variation of series resistance (R_s).

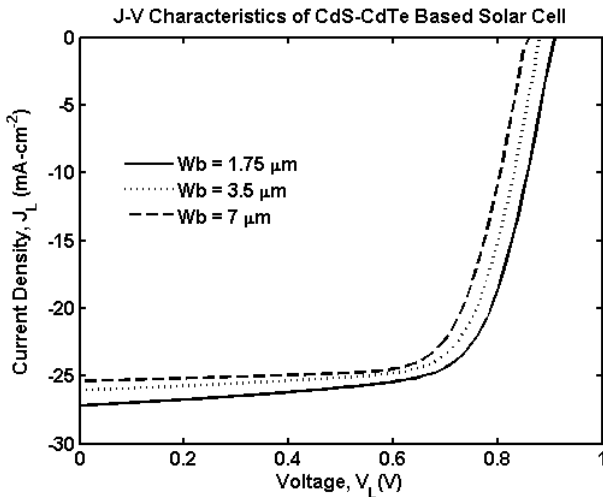


Fig. 5. J-V characteristics curve against the variation of absorber width (W_a).

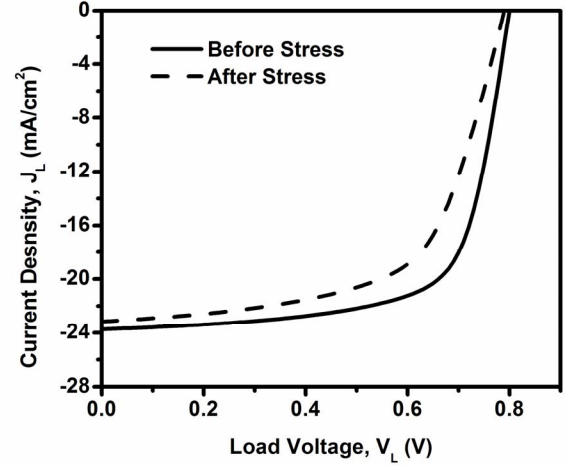


Fig. 6. J-V characteristics curve before and after applying stress.

TABLE 1: PARAMETERS OF THE CDS/CdTe BASED SOLAR CELL USING Ni/Cu BILAYER CONTACT BEFORE AND AFTER STRESS APPLIED. HERE $W_a = 7.0 \mu m$

Parameters	Before Stress	After Stress
J_{sc} (mA/cm ²)	23.7442	23.1561
V_{oc} (V)	0.8	0.79
FF (%)	69.6385	62.0329
η (%)	13.2281	11.3479

IV. CONCLUSION

An analytical model has been developed in this work to describe the current-voltage characteristics including the series resistance of a CdS/CdTe-based thin film solar cell. The model is able to include the voltage-dependent photocurrent collection. The main feature of the proposed model is the inclusion of the surface recombination effects in the analysis, for which a new boundary condition has been introduced. Simulation results of the proposed model show that the effects of surface recombination cannot be neglected. Using this model, the effects of the variation of the series resistance as well as that of the absorber width can be physically explained. The model is also able to explain the effect of stress on the CdS/CdTe based solar cells using Ni/Cu bilayer contact. Finally, the values of the open circuit voltage and the short circuit current density obtained by the proposed model for a solar cell built by Hegedus *et al.* [6] closely matches with the experimental data and hence, the proposed model can be primarily validated.

REFERENCES

- [1] T. Aramoto, *et al.*, "16.0% Efficient Thin-Film CdS/CdTe Solar Cells", *Jpn. J Appl. Phys.*, vol. 36, no. 10, pp. 6304-6307, 1997.
- [2] G. C. Morris and S. K. Das, "Some fabrication procedures for electrodeposited CdTe solar cells," *International Journal of Solar Energy*, vol. 12, no. 1-4, pp.95-108, 1992.

- [3] S. Ikegami, "CdS/CdTe solar cells by the screen-printing-sintering technique: Fabrication, photovoltaic properties and applications," *Solar Cells*, vol. 23, no. 1-2, pp. 89–105, 1988.
- [4] J. Nelson, *The Physics of Solar Cells*, Imperial College Press, Chap. 8, 2003.
- [5] S. Hegedus, "Current-Voltage Analysis of a-Si and a-SiGe Solar Cells including Voltage-dependent Photocurrent Collection," *Prog. in Photovoltaics: Research and Applications*, vol. 5, pp. 151–168, 1997.
- [6] S. Hegedus, D. Desai, and C. Thompson, "Voltage Dependent Photocurrent Collection in CdTe/CdS Solar Cells," *Prog. in Photovoltaics: Research and Applications*, vol. 15, pp. 587–602, 2007.
- [7] S. S. Hegedus and W. N. Shafarman, "Thin-Film Solar Cells: Device Measurements and Analysis," *Prog. in Photovoltaics: Research and Applications*, vol. 12, pp. 155–176, 2004.
- [8] M. S. Anjan and M. Z. Kabir, "Modeling of Current-Voltage characteristics of CdS/CdTe Solar Cells," *Phys. Stat. Solidi A*, vol. 208, pp. 1813–1816, 2011.
- [9] M. A. Mannan, M. S. Anjan and M. Z. Kabir, "Modeling of Current-Voltage Characteristics of Thin Film Solar Cells," *Solid-State Electron*, vol. 63, pp. 49–54, 2011.
- [10] M. M. Haque, M. M. Rahman and M. I. B. Chowdhury, "Current-Voltage Characteristics of CdS/CIGS Thin Film Solar Cells: An Analytical Approach," *Proceedings of 2014 1st International Conference on Non Conventional Energy (ICONCE 2014)*, pp. 28–31, January 16 - 17, 2014, Kalyani, West Bengal, India.
- [11] M. M. Haque, M. M. Rahman and M. I. B. Chowdhury, "Current-Voltage Characteristics of CdS/CdTe Thin Film Solar Cells: An Analytical Approach," *Proceedings of 2014 3rd International Conference on the Development in the Renewable Energy Technology (ICDRET 2014)*, pp. 1–4, May 29–31, 2014, Dhaka, Bangladesh.
- [12] Protik Biswas, Abdullah Sakib Hasan, Ashiq Been Rahim, Aman Ullah, Md. Iqbal Bahar Chowdhury, "Analytical Approach of J-V Characteristics of CdTe Based Thin Film Solar Cells Including Voltage and Space Dependent Electric Field in the Absorber Layer", *2nd International Conference on Electrical Information and Communication Technology (EICT 2015)*, pp. 446–450, December 10–12, 2015, Khulna, Bangladesh.
- [13] Ashiq Been Rahim, Abdullah Sakib Hasan, Protik Biswas, Aman Ullah, Md. Iqbal Bahar Chowdhury, "Analytical Modeling of J-V Characteristics of CIGS Based Thin Film Solar Cell Considering Voltage and Space Dependent Electric Field in the Absorber Layer", *3rd International Conference on Advances in Electrical Engineering (ICAEE 2015)*, pp. 368–371, December 17–19, 2015, Dhaka, Bangladesh.
- [14] Ayesha, Kulsum, Shamara Kabir Silvi and Md. Iqbal Bahar Chowdhury, "Modeling of JV characteristics of CZTS based thin film solar cells including voltage and space dependent electric field in the absorber layer", *4th International Conference on the Development in the Renewable Energy Technology (ICDRET 2016)*, pp. 1–4, January 7–9, 2016, Dhaka, Bangladesh.
- [15] S. M. Sze, *Physics of Semiconductor Devices*, 2nd Edition, Wiley & Sons, 1981.
- [16] National Renewable Energy Laboratory (NREL), <http://rredc.nrel.gov/solarspectra/am1.5/ASTMG173/ASTMG173.html>.
- [17] M. Burgelman in: J. Poortmans and V. Arkhipov (Editors.), *Thin film Solar Cells: Fabrication, Characterization and applications*, Wiley & Sons, Chap. 7, 2006.

Implementing ID3 Algorithm for Gender Identification of Bangladeshi People

Md. Sazzad Hossain; Md. Shamsuzzaman; Md. Ahsan Habib
Dept. of Computer Science and Engineering
Begum Rokeya University, Rangpur
Rangpur -5400, Bangladesh
sazzad.brur@gmail.com; sobuj@brur.ac.bd; md.ahsanhimel@gmail.com

Abstract—Data mining is the procedure of breaking down data from unlike perspectives and resuming it into useful information. It is very important in the field of classification of the objects. It has been fruitfully applied in expert systems to get knowledge. we can determine appropriate classification of unknown objects according to decision tree rules by applying inductive methods to the given values of attributes of those objects. In this paper a decision tree learning algorithm ID3 is applied to build a decision tree in achieving our goal to gender identification of unknown objects. Our experiments have used records of 50 individuals among which 26 were male and 24 were female subjects having age groups of 19 to 25 years. The classification related to the training sets is done by proper calculation. The output of the work is the classified decision tree and the decision rules. It has been observed that the proposed decision tree can recognize 45 subjects gender from 50 individuals. It is a faster process in recognition of individuals' gender and having accuracy level 85% to 90%.

Keywords—Data mining; patterns; expert system; inductive method; decision tree; ID3; algorithm;

I. INTRODUCTION

Gender identity is one's personal experience of one's own gender. Usually it describes one's private feel of being a woman or a man. In most Western societies, there exists a gender binary, a social dichotomy that enforces conformance to the ideals of masculinity and femininity in all aspects of sex and gender: biological sex, gender identity, and gender expression.

Though gender identification term is a medical term practiced to explicate sex reassignment surgery to the public. It is largely come up in psychology today, frequently as a core gender identity. It is an issue to verify gender in sports of an athlete to compete in an event that is limited to a single sex. Several techniques are used to identify gender. Most common methods are DNA testing, gait recognition, finger print based identification and so on, and all are expensive and time consuming.

Our research aim is to identify gender, this research is based on physical aspect. Biological aspect is not considered, Biological consideration is always takes long time to give the result. The main objective of our research is to determine result of identification of gender reducing time.

ID3 algorithm is primarily used for decision making. Ross Quinlan formulated ID3 (Iterative Dichotomiser 3) algorithm, used to generate a decision tree from a data set [14]. A defined set of examples are made to form a decision tree by ID3 classification algorithm. In order to draw the dependence of various attributes the resulting tree is used.

Data mining is a proficiency to elicit enshrouded predictive data from database, especially from large database. Data mining is a confirmative to examine the upcoming prediction with the help of historic behavior of data. Different techniques are used to mine data from database. One crucial technique is classification and segmentation, under which decision trees are created in order to predict data from the existing one.

A decision tree is a tree in which each leaf node acts a decision where each branch node acts as a choice between a numbers of choices. Decision tree are usually used for acquiring information for the aim of decision –making.

It is an abstraction to visually and explicitly represent the dependencies between attributes of elements in a set and their sorting on the tree determines a set of rules which explain and summarize the relations of all items throughout the set. Decision tree learning algorithms have been successfully used in knowledge discovery. To provide an appropriate assortment of objects Decision tree learning algorithms use induction in terms of their attributes, inferring decision tree rules. In their learning phase, explicit rules or interactions among relevant features are induced. [10]. Based on the characteristics of this algorithm, the main objectives of the research work are

- Construct the decision tree until the appropriate classification is reached.
- Another is to generate the decision rules for the problem.

Our research is based on Bangladeshi people. It can be applicable for Asian people also. It can be easily developed for different countries based on their people's age, height and weight.

II. RELATED WORK

For the classification and the pattern recognition, ID3 algorithm is the main method used in this paper. The first work on ID3 was done by J.R Quinlan in 1986. He

summarizes an approach to synthesizing decision trees that has been used in a variety of systems. He concludes with illustrations of current research directions. [1]. This is a learning method differs from non-linear classifiers such as support vector machines or neural networks where the learning phase is to determine the parameters of the non-linear kernel functions. This advantage of rule-based classifiers is increasingly used in the analysis of bio-medical data, financial data in banking securities, stocks trading data, policy and claim data in insurance and the data obtained from other sources. [2].

After Quinlan, ID3 has been used by many researchers in different fields. "Predictive Data Mining for Medical Diagnosis" this paper survey of the heart disease using Id3 algorithms of Naive Bayes, Decision Tree, Nearest Neighbors Algorithms (K-Nn). This survey gives the report the decision tree is the highly predictive for heart disease. [3]. "Decision Support System for Medical Diagnosis Using Data mining", these papers referred by the survey of compare 3 diseases are heart disease, diabetes and hepatitis. So this survey using algorithm was ID3, C4.5, and CART algorithms. [4]. Furthermore, data mining for breast cancers is finalized for mainly predicting the affinity between recurrence and other attributes of breast cancer. The result demonstrates the strength of Decision Tree in medical data mining which provide physicians with diagnostic assistance [7].

In the field of networks security, an interesting execution of ID3 was implemented. Victor H. Goreia (et.al) use ID3 to web attack detection. ID3 is sufficient to put apart a number of Web attacks, including a large proportion of their variants. It also surpasses existing methods: it portrays a higher true-positive detection rate and a lower false-positive one. The ID3 output classification rules that are easy to read and so computer officers are more likely to grasp the root of an attack, as well as extending the capabilities of the classifier. [5]. Another researcher Sonika Tiwari who used the improved ID3 for detecting Network Anomalies with horizontal partitioning based decision tree. She found the comparative analysis of different clustering algorithms and existing ID3 based decision tree. [6].

Some authors implemented and evaluated the ID3 algorithm with some interesting examples, like Anand Bahety applied the algorithm on the "Play Tennis" database. He classified suitable weather for playing tennis [8]. By using ID3 algorithm, Kumar Ashok(et.al) classified the "census 2011 of India" dataset. The information theory concept is applied in the field of data mining. [9].

An improved version called C4.5 of ID3 is now available. Lastly, there are vast amounts of data available and there are many ways it can be manipulated. So, using these algorithms is an iterative process where processes are always being improved (such as when new attributes are added for considerations – for example, there may other factors for risk of heart disease such as weight or family history). For the medical industry, these decisions can determine if a patient is a high risk for heart disease along with making conclusions as to what insurance coverage a company should give a person based on this risk.

III. DECISION TREE LEARNING ALGORITHM –ID3 BASICS

A general characteristic of a decision tree is to travel from root node to leaf node. In the case of ID3 decision tree, it first classifies instances by traversing node. It starts from the root node, checking the corresponding attribute of this node and then traveling down the tree branch based on the attribute value. This process is repeated at the child tree level. Here is the decision tree learning algorithm:

- All instances are represented as attribute - value pairs. For example, attribute 'Gender' and its value 'male', 'female'.
- The target function provides discrete output values such as 'true' or 'false', 'p(positive)' or 'n(negative)'. It is also possible to produce real valued outputs, but we will cover the issue in the later part of this paper.
- The training data may contain errors. This can be dealt with pruning techniques.

Basically, ID3 algorithm constructs the decision tree by applying a top-down greedy search through the given data sets to test each attribute at every tree node. In order to select the most useful attribute for classifying a given sets, we present a metric information gain. We need to minimize the depth of the tree i.e. we need to minimize the questions asked to find an optimal way to classify a learning set. Thus, we need some function to measure the most balanced splitting. The information gain metric is such a function which is worked based on entropy of a learning set.

Entropy --- measuring homogeneity of a learning set

$$\text{Entropy}(S) = - P(+ve)\log_2 P(+ve) - P(-ve)\log_2 P(-ve)$$

- $P(+ve)$: proportion of positive examples in S
- $P(-ve)$: proportion of negative examples in S

Information Gain --- measuring the expected reduction in Entropy

The information gain, $\text{Gain}(S,A)$ of an attribute A,

$$\text{Gain}(S,A) = \text{Entropy}(S) - \sum_{v \text{ from } 1 \text{ to } n \text{ of } (|S_v|/|S|)} \text{Entropy}(S_v)$$

The motives behind the ordering are:

- To creating small decision trees to identify records after a few decision splitting.
- To match a hoped for minimalism of the process of decision making

IV. RESULT AND DISCUSSION

In ID3 algorithm we need to classify gender based on some training data. Here we list all the training data that we have used.

Table 01: Training Data

No	Age(Year)	Weight(KG)	Height (Inch)	Sex
1	21	56	62	Male
2	25	64	68	Male
3	22	64	66	Male
4	22	54	63	Female
5	22	62	66	Male
6	24	56	65	Male
7	24	52	63	Male
8	21	60	71	Male
9	24	59	67	Male
10	22	62	66	Male
11	21	62	71	Male
12	21	48	63	Female
13	23	48	63	Female
14	21	42	60	Female
15	22	52	66	Male
16	24	50	58	Female
17	22	55	58	Female
18	24	55	60.5	Female
19	24	40	60	Female
20	22	56	64	Male
21	22	42	60	Female
22	22	50	60	Female
23	24	48	64	Male
24	24	45	62	Female
25	19	42	62	Female
26	24	49	63	Female
27	23	49	61	Female
28	21	42	65	Female
29	23	50	62	Female
30	20	51	63	Female
31	23	42	60	Female
32	24	45	62	Female
33	23	80	66	Male
34	21	49	61.5	Female
35	21	62	67.5	Male
36	24	56	59	Female
37	23	60	64	Male
38	25	67	68	Male
39	21	60	64	Female
40	22	57	68	Male
41	25	60	66	Male
42	21	61	69	Male
43	22	47	64	Female
44	25	72	68	Male
45	22	62	68	Female
46	22	62	68	Male
47	22	65	70	Male
48	19	60	64	Male
49	24	53	65	Male
50	21	80	66.5	Male

A. Learning Set

First categorize the value of the above table for better calculation result using ID3 algorithm. Table below indicates the acceptable value.

Table 02: Categorized attributes

Attribute	Possible Values				
Age	Young1	Young2	Young3	---	---
Weight	Very Thin	Thin	Average	Heavy	Very Heavy
Height	Very Short	Short	Average	Tall	Extra Tall
Gender	Female	Male	---	---	---

Into the following table we assign value to the categorized attributes and reflect the categorized training data.

Table 03: Values of the categorized attributes

Age (Year)	Young1 (Y1) (19-21)	Young2 (Y2) (22-23)	Young3 (Y3) (24-25)	---	---
Weight (Kg)	Very Thin (VT) (<=45)	Thin (T) (46-50)	Average (A) (51-55)	Heavy (H) (56-60)	Very Heavy (VH) (>=61)
Height (Inche)	Very Short (VS) (<=60)	Short (S) (61-62)	Average (A) (63-65)	Tall (T) (66-68)	Extra Tall (ET) (>=69)
Gender	Female	Male	---	---	---

Table 04: Categorized training data

No	Age	Weight	Height	Sex
1	Y1	H	S	Male
2	Y3	VH	T	Male
3	Y2	VH	T	Male
4	Y2	A	A	Female
5	Y2	VH	T	Male
6	Y3	H	A	Male
7	Y3	A	A	Male
8	Y1	H	ET	Male
9	Y3	H	T	Male
10	Y2	VH	T	Male
11	Y1	VH	ET	Male
12	Y1	T	A	Female
13	Y2	T	A	Female
14	Y1	VT	VS	Female
15	Y2	A	T	Male
16	Y3	T	VS	Female
17	Y2	A	VS	Female
18	Y3	A	S	Female
19	Y3	VT	VS	Female
20	Y2	H	A	Male
21	Y2	VT	VS	Female
22	Y2	T	VS	Female
23	Y3	T	A	Male
24	Y3	VT	S	Female
25	Y1	VT	S	Female
26	Y3	T	A	Female
27	Y2	T	S	Female
28	Y1	VT	A	Female
29	Y2	T	S	Female
30	Y1	A	A	Female
31	Y2	VT	VS	Female
32	Y3	VT	S	Female
33	Y2	VH	T	Male
34	Y1	T	S	Female
35	Y1	VH	T	Male
36	Y3	H	VS	Female
37	Y2	H	A	Male
38	Y3	VH	T	Male
39	Y1	H	A	Female
40	Y2	H	T	Male
41	Y3	H	T	Male
42	Y1	VH	ET	Male
43	Y2	T	A	Female
44	Y3	VH	T	Male
45	Y2	VH	T	Female
46	Y2	VH	T	Male
47	Y2	VH	ET	Male
48	Y1	H	A	Male
49	Y3	A	A	Male
50	Y1	VH	T	Male

B. Calculations

STEP 1: First calculate the Entropy(S) of the example set. Our example set contains 24 Female and 26 Male.

$$\text{Entropy}(S) = - (24/50)\log_2(24/50) - (26/50)\log_2(26/50) = 0.998845$$

STEP 2: Attribute Age

Age value can be Young1, Young2 and Young3. Total frequency of Young1, Young2 and Young3 are 14, 20 and 16 respectively. Where,

Age = Young1, 6 of subjects are 'Female' and 8 are 'Male'
Age = Young2, 11 of subjects are 'Female' and 9 are 'Male'
Age = Young3, 7 of subjects are 'Female' and 9 are 'Male'

$$\text{Entropy}(AY1) = - (6/14)\log_2(6/14) - (8/14)\log_2(8/14) = 0.985228$$

$$\text{Entropy}(AY2) = - (11/20)\log_2(11/20) - (9/20)\log_2(9/20) = 0.992774$$

$$\text{Entropy}(AY3) = - (7/16)\log_2(7/16) - (9/16)\log_2(9/16) = 0.988699$$

$$\begin{aligned} \text{Gain}(S, \text{Age}) &= \text{Entropy}(S) - (14/50) \times \text{Entropy}(AY1) - (20/50) \times \text{Entropy}(AY2) - (16/50) \times \text{Entropy}(AY3) \\ &= 0.998845 - (14/50) \times 0.985228 - (20/50) \times 0.992774 - (16/50) \times 0.988699 \\ &= 0.009487 \end{aligned}$$

STEP 3: Attribute Weight

Weight value can be Very Thin, Thin, Average, Heavy and Very Heavy. Where,

Weight = Very Thin, 8 of subjects are 'Female' and 0 are 'Male'

Weight = Thin, 9 of subjects are 'Female' and 1 are 'Male'
Weight = Average, 4 of subjects are 'Female' and 3 are 'Male'
Weight = Heavy, 2 of subjects are 'Female' and 9 are 'Male'
Weight = Very Heavy, 1 of subjects are 'Female' and 13 are 'Male'

$$\text{Entropy}(WVT) = - (8/8)\log_2(8/8) - (0/8)\log_2(0/8) = 0$$

$$\text{Entropy}(WT) = - (9/10)\log_2(9/10) - (1/10)\log_2(1/10) = 0.468995$$

$$\text{Entropy}(WA) = - (4/7)\log_2(4/7) - (3/7)\log_2(3/7) = 0.985228$$

$$\text{Entropy}(WH) = - (2/11)\log_2(2/11) - (9/11)\log_2(9/11) = 0.371480$$

$$\text{Entropy}(WVH) = - (1/14)\log_2(1/14) - (13/14)\log_2(13/14) = 0.371232$$

$$\begin{aligned} \text{Gain}(S, \text{Weight}) &= \text{Entropy}(S) - (8/50) \times \text{Entropy}(WVT) - (10/50) \times \text{Entropy}(WT) - (7/50) \times \text{Entropy}(WA) - (11/50) \times \text{Entropy}(WH) - (14/50) \times \text{Entropy}(WVH) \\ &= 0.998845 - (8/50) \times 0 - (10/50) \times 0.468995 - (7/50) \times 0.985228 - (11/50) \times 0.371480 - (14/50) \times 0.371232 \\ &= 0.581443 \end{aligned}$$

STEP 4: Attribute Height

Height value can be Very Short, Short, Average, Tall and Extra Tall. Where,

Height = Very Short, 8 of subjects are 'Female' and 0 are 'Male'

Height = Short, 7 of subjects are 'Female' and 1 are 'Male'

Height = Average, 8 of subjects are 'Female' and 7 are 'Male'

Height = Tall, 1 of subjects are 'Female' and 14 are 'Male'

Height = Extra Tall, 0 of subjects are 'Female' and 4 are 'Male'

$$\text{Entropy}(HVS) = - (8/8)\log_2(8/8) - (0/8)\log_2(0/8) = 0$$

$$\text{Entropy}(HS) = - (7/8)\log_2(7/8) - (1/8)\log_2(1/8) = 0.543564$$

$$\text{Entropy}(HA) = - (8/15)\log_2(8/15) - (7/15)\log_2(7/15) = 0.996791$$

$$\text{Entropy}(HT) = - (1/15)\log_2(1/15) - (14/15)\log_2(14/15) = 0.353359$$

$$\text{Entropy}(HET) = - (0/4)\log_2(0/4) - (4/4)\log_2(4/4) = 0$$

$$\begin{aligned} \text{Gain}(S, \text{Height}) &= \text{Entropy}(S) - (8/50) \times \text{Entropy}(HVS) - (8/50) \times \text{Entropy}(HS) - (15/50) \times \text{Entropy}(HA) - (15/50) \times \text{Entropy}(HT) - (4/50) \times \text{Entropy}(HET) \\ &= 0.998845 - (8/50) \times 0 - (8/50) \times 0.543564 - (15/50) \times 0.996791 - (15/50) \times 0.353359 - (4/50) \times 0 \\ &= 0.506829 \end{aligned}$$

STEP 5: Summary of results are listed below.

Entropy(S)	= 0.998845
Gain(S, Age)	= 0.009487
Gain(S, Weight)	= 0.581443
Gain(S, Height)	= 0.506829

To find the root node, just take the highest Gain value. Gain (S, Weight) = 0. 581443 is the highest among all attribute. Therefore, "Weight" is the root node of the decision tree among other attributes.

STEP 6: To find next decision node, "Weight" as root node has five possible values – Very Thin, Thin, Average, Heavy and Very Heavy. So root node has five branches. In our example data, every subject that has very thin weight is female. So we can declare that when weight is very thin then the subject will be female.

STEP 7: Now find the next node for thin subject.

$$\text{Entropy}(WT) = 0.468995$$

$$\text{Gain}(WT, \text{Age}) = 0.193506$$

$$\text{Gain}(WT, \text{Height}) = 0.180223$$

Hence, Age has the highest gain; therefore, it is used as the next decision node.

STEP 8: This process goes on until all data is classified perfectly or we run out of attributes.

V. CONCLUSION

The implementation of the algorithm and their studies conclude that as a decision tree learning algorithm, ID3 works well on any dataset classification related problems having discrete values. Related to the research work it concludes that the classification tree built using ID3 algorithm is efficient and optimized one. We can set some rules by focusing on the resultant decision tree towards the identification of gender.

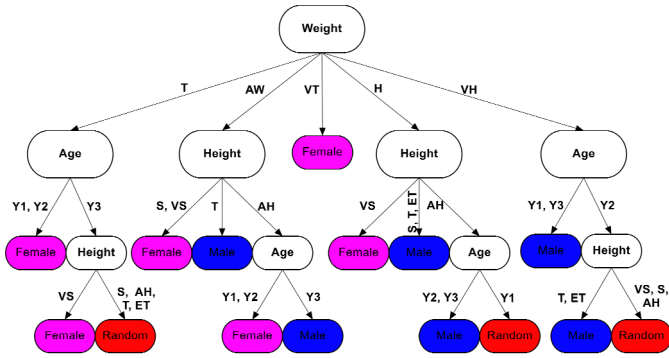


Fig. 1. Decision tree for gender identification.

Rules are:

1. When weight is VT, then the subject is female.
2. When weight is T, if age is Y1 OR Y2 then female, else if age is Y3 and height is VS then female, else random.
3. When weight is A, if height is S OR VS then female, else if height is T then male, else if height is A and age is Y1 OR Y2 then female, else male.
4. When weight is H, if height is VS then female, else if height is S OR T OR ET then male, else if height is A and age is Y2 OR Y3 then male, else random.
5. When weight is VH, if age is Y1 OR Y3 then male, else if age is Y2 and height is T OR ET then male, else random.

We can also conclude that although this system is developed to classify human gender, it can be used in many other fields where prediction and data classification must required.

REFERENCES

- [1] Quinlan, J.R., "Induction of Decision trees", Machine Learning 1: 81-106, 1986.
- [2] Md. N. Huda, M. Numao and T. Okada, "Multiple Instance Based Learning for Feature Extraction from Chemical Compound", 21st Century COE International Symposium, (2005).
- [3] J. Soni, U. Ansari, D. Sharma, S. Soni, "Predictive Data mining For Medical Diagnosis an Overview of Heart Disease Prediction", International Journal of Computer Applications (2011).
- [4] D.Senthil Kumar, Sathyadevi, G.; Sivanesh, S., "Decision Support System for Medical Diagnosis Using Data Mining", International Journal of Computer Science Issues (IJCSI); May 2011, Vol. 8 Issue 3, p147.
- [5] Victor.H.Garcia, Raul Monroy and Maricela Quintana, "Web Attack Detection Using ID3".
<http://homepage.cem.itesm.mx/raulm/pub/ifip06.pdf>
- [6] Sonika Tiwari and Prof. Roopali Soni, "Horizontal partitioning ID3 algorithm A new approach of detecting network anomalies using

decision tree", International Journal of Engineering Research & Technology (IJERT) Vol. 1 Issue 7, September – 2012.

- [7] Ruijuan Hu, (2011), "Medical Data Mining Based on Decision Tree Algorithm", Computer and Information Science, Vol. 4, No. 5; September 2011.
- [8] Anand Bahety, "Extension and Evaluation of ID3-Decision Tree Algorithm". University of Maryland, College Park.
- [9] Kumar Ashok, Taneja H C, Chitkara Ashok K and Kumar Vikas, "Classification of Census Using Information Theoretic Measure Based ID3 Algorithm". Int. Journal of Math. Analysis, Vol. 6, 2012, no. 51, 2511 – 2518.
- [10] C. Apte and S. Weiss, "Data Mining with Decision Trees and Decision Rules", Future Generation Computer Systems, vol. 13, (1997), pp. 197-210
- [11] L. S. Kumar and M. A. Padmapriya, "ID3 Algorithm Performance of Diagnosis For Common Disease", Int. Journal of Advanced Research in Computer Science and Software Engineering, vol. 2, (2012)
- [12] Priyadharsini and Dr. A. S. Thanamani, "Prediction of Missing Values in Blood Cancer & Occurrence of Cancer Using Improved Id3 Algorithm", Int. Journal of Innovative Research in Computer and Communication Engineering, vol. 2, (2014)
- [13] S. K. Pramanik, Dr. B. K. Pramanik, "Hybrid Classification Algorithm for Knowledge Acquisition of Biomedical Data", Int. Journal of Advanced Science and Technology, vol. 44, (2012)
- [14] Sweta Rai, Priyanka Saini, Ajit Kumar Jain "Model for Prediction of Dropout Student Using ID3 Decision Tree Algorithm", International Journal of Advanced Research in Computer Science & Technology, Vol. 2 Issue 1 Ver. 2 Jan-March 2014.

Detection of Breast Cancer from Histopathology image and Classifying Benign and Malignant State Using Fuzzy Logic

Fatema-Tuz Johra, Md. Maruf Hossain Shuvo
Department of Electronics and Communication Engineering
Khulna University of Engineering and Technology (KUET)
Khulna, Bangladesh
mimakuetian@gmail.com, maruf.shuvo1@yahoo.com

Abstract— Breast cancer is one of the major public health problem for women throughout the world. It has two states, known as benign and malignant. Benign state is slow growing, rarely spread to other parts of body and have well- defined borders. On the other hand, Malignant state has tendency to grow faster and it is life threatening. So, classification of this two state is crucial for proper diagnosis of a breast cancer patient. In this paper, we have introduced a new pipeline for breast cancer cell detection and feature extraction using an open source image analysis software named CellProfiler. We proposed an algorithm based on fuzzy inference system for classification of the benign and malignant state. Comparison using well known performance parameters such as accuracy, sensitivity and specificity shows that our proposed approach performs better than the Artificial Neural Network (ANN) and Support Vector Machine (SVM) based classification. The sensitivity, specificity, and accuracy of the proposed method is 95.6%, 90.63%, and 94.26% respectively.

Keywords—breast cancer detection; classification; artificial neural network; support vector machine; fuzzy logic

I. INTRODUCTION

Breast Cancer (BC) is one of the most common type of cancer among female population in the whole world. Death rate for BC is higher than other types of cancer. Diagnosis of Breast Cancer (BC) depends on the visual aspects of tissue or cells removed from a patient and then step by step evaluation are specimen preparation, selection of tumor cell nuclei, image preprocessing, feature extraction according to size, shape, texture and so on [1].

Proper diagnosis can help a patient to get rid of BC risk if the state of cancer type is benign (localized and non - invading) or malignant (invading and life threatening). For diagnosis, different methods like mammograms, CT scan, histopathological image analysis (biopsy image) are commonly used. If there is a positive result of mammograms about BC, then diagnosis with histopathological imaging can be done [2]. A histology image analysis system generally has a combination of hardware and software and it can be divided into two consecutive subsystems: (1) tissue preparation and image production and then (2) image processing analysis. To reduce the death rate among women two thing are very important that

are education about BC and proper screening that means diagnosis [3]. In this paper we have worked on the second type that plays a vital role to prescribe treatments for a breast cancer patient.

Fuzzy system is rule based and it has normally four components: first one is fuzzifier that transforms real input values in fuzzy values, second one is inference engine that helps to get fuzzy output and the defuzzifier that transforms fuzzy value to real value and the last one is fuzzy rule base [4]. Fuzzy logic is used in this work rather than ANN because fuzzy is simple to implement, knowledge based system, and able to use prior rule [5]. Neuro-fuzzy classifier is also used to classify the breast cancer and six selected features that are clamp thickness, uniformity of cell size, uniformity cell shape, marginal adhesion, single epithelial cell size, and bare nuclei can be used to gain good performance [5]. Random forest (RF) is another classifier that can be used with 10 features, these features are preprocessed by k-NN (K - nearest neighbor) algorithm before applying random forest and then the overall accuracy is 72 % [6]. In this paper we have worked with different features that gives satisfactory classification performance for fuzzy classifier. Main focus of this study is proper detection of cancer cell and classification of breast cancer state with higher accuracy, sensitivity and specificity than existing methods.

In this work, three prominent machine learning algorithms are used as classifier namely Artificial Neural Network (ANN), Support Vector Machine (SVM), and Fuzzy Logic. Comparison shows that the Fuzzy logic based classification is performing better than the other two classifiers. To classify benign and malignant state a set of extracted feature is very essential and in this study six features are used for overall analysis such as: area, compactness, eccentricity, major axis length, mean radius and the perimeter. These features shows significant variation from benign to malignant state that will be briefly discussed in the next Section.

This paper is arranged as follows: Section II introduces the overall methodology that contains with BC detection, feature extraction and classification. Section III describes results and discussion for each step of this work and gives a comparison

among the used classifiers. Finally, Section IV concludes the whole work and outcomes of this paper.

II. METHODOLOGY

For the analysis and classification of two types of cancer state we have used benign and malignant images of the dataset collected from the Center for Bio-Image Informatics, University of California, Santa Barbara, USA. The dataset contains 50 H&E stained histopathology images of two types of breast cancer. This is called H&E because routine histology uses the stain combination of hematoxylin and eosin.

Computational approaches to cell are attractive but it will be worthy if proper implementation with user-friendly software tools is used [7]. CellProfiler is such a useful and user friendly software that is used in this work to detect breast cancer cell or nuclei and to extract feature for cytology image analysis. This software consists sequence of modules i.e. a pipeline to complete any related target based task [2] such as quantitatively measuring phenotypes from many images. The flexible use of this can analyze images to generate accurate measurements [8]. Typical pipeline of this software consists of loading the images, cropping images, some image processing modules, identifying the objects and so on. These modules can automatically be added, removed, or rearranged within a pipeline by users [9]. The step by step working process in a flowchart for BC detection and classification of two state is shown in Fig.1.

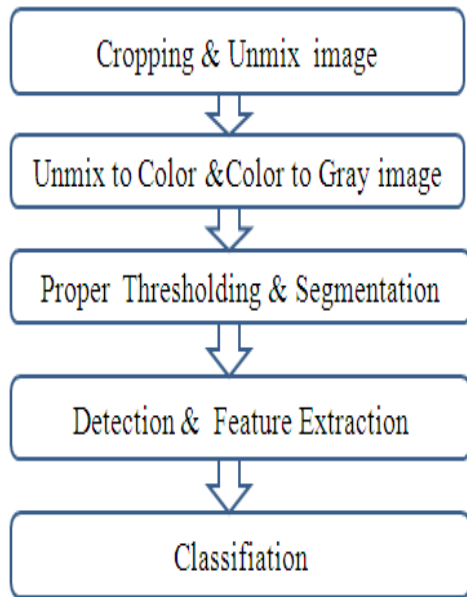


Fig. 1. Overall System Flow Chart for BC detection and classification of two state of BC.

Firstly, we have cropped the original image and then unmixed that image. The next step is to convert the unmixed image into color image followed by a color to gray scale conversion for thresholding and segmentation. After the breast cancer cell detection feature extraction from these images is done by using a pipeline. Finally, classification has been done with the ANN, SVM and Fuzzy Logic algorithm in MATLAB (Math works, Inc.) . In the next sections breast cancer detection,

feature extraction and classification process are discussed briefly.

A. Breast Cancer Detection

Normally a patient is suggested to do biopsy after ensuring that the person has cancerous cell or nearly cancerous cell that is unhealthy. It is very essential to know the cancer type in histopathological test. For this purpose, cancer cell or nuclei detection is very important. Proper segmentation with thresholding is used to detect the cancer cell or nuclei from images. Image segmentation is used to extract cells from the background and these cells are the focusing part for further disease identification and classification [2]. After detecting cell or nuclei in histopathological image we need to classify the detected cells whether benign or malignant. For this purpose extracting some useful features are necessary.

B. Feature Extraction

For disease classification and diagnosis of BC, object-level features can be classified into four categories: (1) size and shape, (2) radiometric and densitometric, (3) texture and (4) chromatin-specific [9]. Morphological large size, coarse chromatin texture, irregular shape of cancer nuclei are the characteristics that are observed and used for cancer detection in some methods [10]. In this work the feature extraction process is done by using the CellProfiler pipeline. From many extracted features we have used six important features such as: area, compactness, eccentricity, major axis length, mean radius and perimeter as shown in Table I. These features are chosen because of having significant variation with benign and malignant type images and their corresponding values are observed carefully for further classification so that the accuracy is high.

C. Classification

Classifiers are trained with features to successfully distinguish breast cancer state. Training the network is used to achieve a balance between the ability to respond properly to the input patterns. The ANN used to train has an input layer, output layer and an intermediate layer of hidden nodes [11]. The back propagation algorithm is used for ANN classifier. In SVM algorithm there is max-margin hyper plane and the margin for SVM that is trained with samples and then classify the two targeted classes from test data. SVM provides better generalization and requires fewer data than a passive learning procedure. It also has improved performance that yields less computation time [12]. The overall process of developing a fuzzy inference system for classification begins with defining the number of system inputs and outputs followed by the definition of input and output membership functions. Finally, the rules are made by using IF-THEN statements.

For each classifier, there are three measurements in percentage:

$$Sensitivity = \left(\frac{TP}{TP+FN} \right) \quad (1)$$

$$Specificity = \left(\frac{TN}{TN+FP} \right) \quad (2)$$

$$Accuracy = \left(\frac{TP+TN}{TN+TP+FN+FP} \right) \quad (3)$$

Where TP means the number of true positives that defined as malignant as malignant, TN is the number of true negatives that means benign as benign, FN is the number of false negatives that defines malignant as benign, and FP is the number of false positives that indicates benign as malignant. Since this study aims to estimate the performance of ANN, SVM and Fuzzy Logic classifier for classification of benign and malignant breast cancer state from histopathology image data, appropriate definition of the TP, FP, TN, and FN are very important [6].

III. RESULTS AND DISCUSSIONS

A. Simulation Results Using Cell Profiler for BC Detection

The simulation results using breast cancer images dataset are shown here by maintaining the workflow.

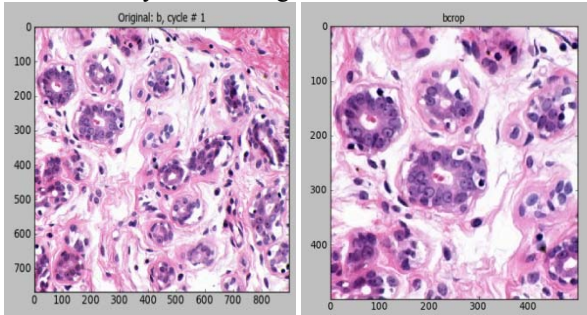


Fig. 2. Benign Breast Cancer Image (A) Original image (B) Cropped image.

Fig. 2(A) shows the original benign breast cancer image and the cropped image is shown in Fig. 2(B). These two image have certain borders for a portion and visually the cell or nuclei size and shape is not abnormally large and irregular shape. Fig. 3(A) shows the next step of this work that is unmixing colors of the cropped image and then in Fig. 3(B) thresholding is used for segmentation of the cell or nuclei.

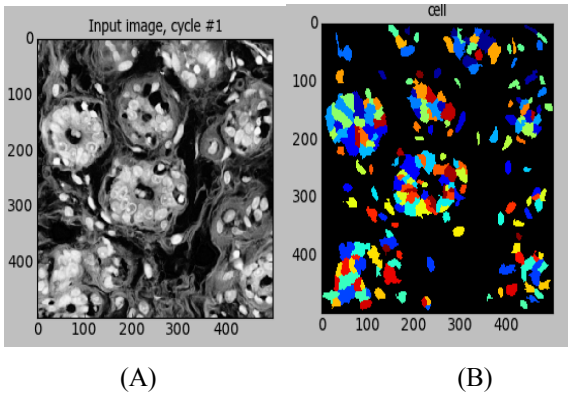


Fig. 3. Benign Breast Cancer (A) input image (B) Benign Breast Cancer image with primary object detection

There are two types of histopathological images used in this work that are benign and malignant. This two type of image are shown with unmix color in Fig. 4(A) and in Fig. 4(B) respectively.

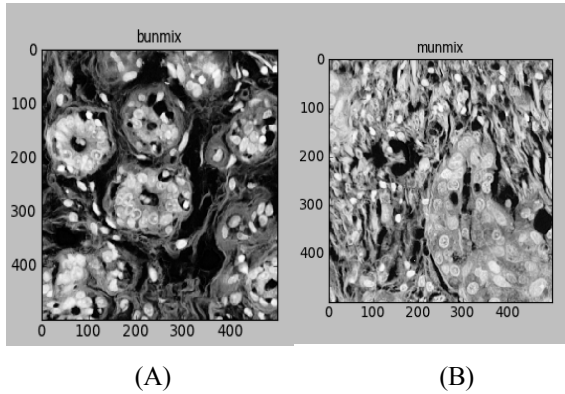


Fig. 4. (A) Benign Breast Cancer (unmix) image (B) Malignant Breast Cancer (unmix) image.

The final detection of cancer cell or nuclei with red outlines in Fig. 5(A) and comparison with the original image that is shown in Fig. 5(B) is accomplished step by step by using some image processing modules in CellProfiler.

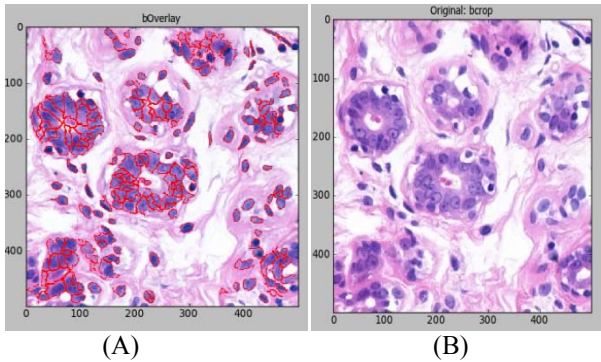


Fig. 5. (A) Benign Breast Cancer cropped image (B) Cell or nuclei detection with outlines.

Then features are extracted by using another modules in this software and within these extracted features significant six features are used to classify the benign and malignant state. These features and a sample data for each type of state are shown in Table I.

TABLE I. EXTRACTED SELECTED FEATURE NAME AND SAMPLE		
Feature Name	Benign data sample	Malignant data sample
Area	96	1109
Compactness	1.185696	1.243103
Ecentricity	0.704195	0.705471
Major axis length	13.9553	48.3275
Mean Radius	1.935762	4.400758
Perimeter	42.796	186.942

B. Classification Using ANN and SVM:

To classify the benign and malignant state we have used a neural network and back propagation algorithm. There are four types of confusion matrices that are training matrix, validation matrix, test matrix and all confusion matrix that is the combination of these three matrix as shown in Fig. 6. Simulated results shows that the test confusion matrix gives better performance.

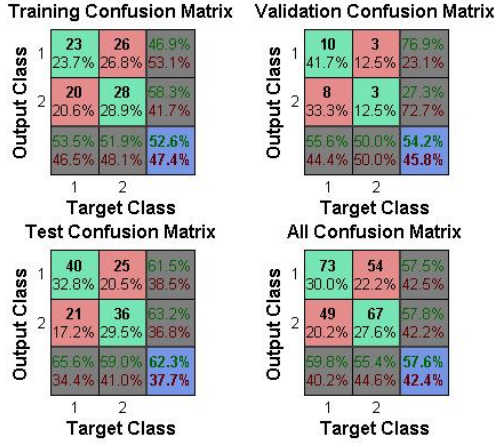


Fig. 6. Confusion Matrices for Classifier Performance Evaluation.

Then after the simulation of data the early stopping criteria is depicted in Fig. 7 and we can see that there is 6 epochs for this simulation.

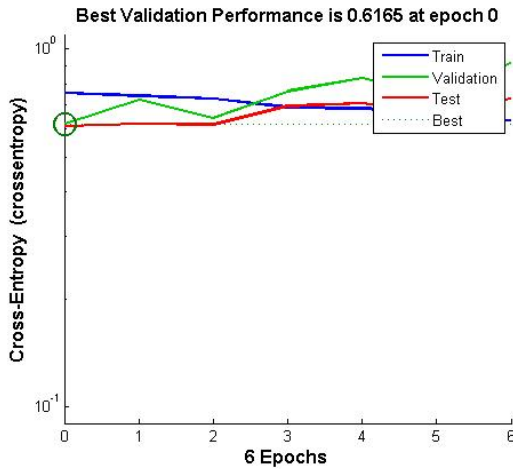


Fig. 7. The Early Stopping Criteria for ANN Algorithm.

The error histogram of this ANN algorithm is shown in Fig.8. Error means the subtraction of outputs from targets as indicated in the horizontal axis.

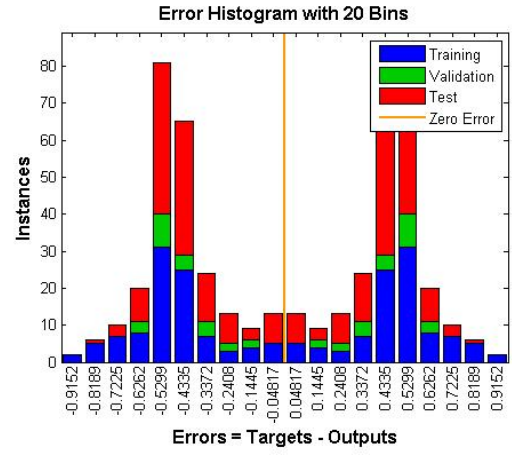


Fig. 8. Error Histogram of ANN Algorithm.

Then overall sensitivity, specificity and accuracy for ANN classifier is calculated as 61.5%, 63.2%, and 62.3% using equation 1, 2, and 3; for 50% of testing data from the confusion matrices as presented in Table II.

TABLE II. ACCURACY, SENSITIVITY AND SPECIFICITY COMPARISON IN ANN, SVM AND FUZZY LOGIC

Classifier	Sensitivity	Specificity	Accuracy
ANN	61.5%	63.2%	62.3%
SVM	83.7%	83.33%	86.9%
Fuzzy Logic	95.6%	90.63%	94.26%

For SVM based classification 50% of testing data is used. The overall result as shown in Table II, is better than the ANN classifier having sensitivity, specificity, and accuracy of 83.7%, 83.33%, and 86.9% respectively.

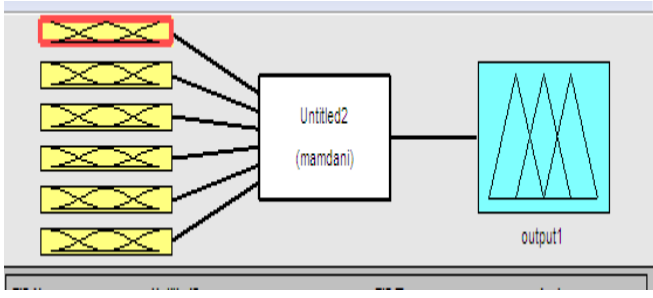
C. Classification using Fuzzy Logic:

The fuzzy inference system with input window and six inputs for six features is shown in Fig. 9(A) and the outputs with two range is shown in Fig. 9(B) for classifying two state of BC.

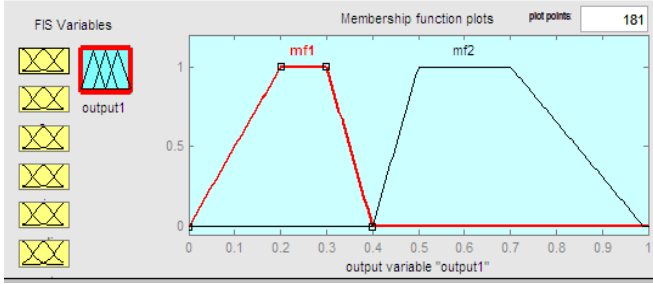
The crisp logic for the classification of benign and malignant state is presented in Table III.

TABLE III. CRISP LOGIC FOR FUZZIFICATION

Defuzzified Output	Classification of Cancer
(c>0) && (c<=0.4)	Benign
c>0.4) && (c<=1)	Malignant



(A)



(B)

Fig. 9. Benign and malignant classification using Fuzzy Inference System(A)input window (B) output window with membership function.

Fig. 10 shows the rule viewers which consist of the system's input and output and the defuzzification result column. The first to sixth columns are the area, compactness, eccentricity, major axis length, mean radius and perimeter. The last column is the classification output column. The last row of the seventh column shows the processing result that leads to defuzzification result to obtain benign or malignant state and in the bottom of this figure we can see the view of this unit.

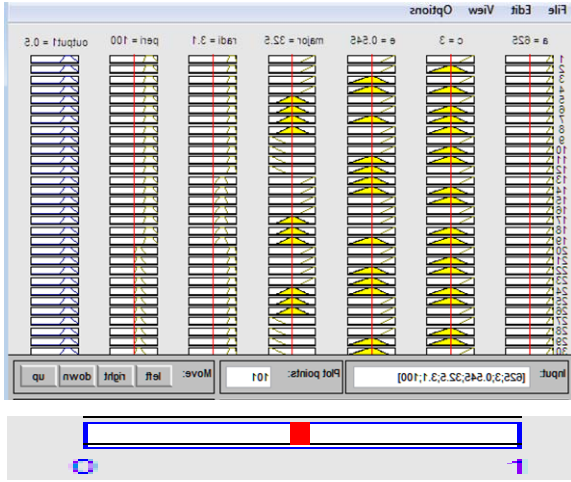


Fig. 10. Rule view of Breast Cancer Classification.

Classification using fuzzy logic has a sensitivity, specificity and accuracy of 95.6%, 90.63%, and 94.26% as shown in Table II. Table II reflects the overall accuracy

comparison among three classifiers used in this paper and accuracies are 62.3%, 86.9% and 94.26% for ANN, SVM and Fuzzy classifiers respectively. From this results we can say that the Fuzzy Logic is a useful classifier with higher accuracy, sensitivity and specificity for benign and malignant state classification than other classifiers.

IV. CONCLUSIONS

Detection of Breast Cancer cell and efficient classification of the benign and malignant state is presented here. For detection and feature extraction Cell Profiler is used with a new pipeline. In this study, ANN, SVM and Fuzzy Logic are three classifiers that are used to classify the breast cancer state. Comparison of the accuracy, sensitivity, and specificity of these three algorithm proves that the fuzzy logic based classification performs superior than the existing methods. The high performance of Fuzzy Logic as a classifier may help patient and pathologist for proper diagnosis of breast cancer.

REFERENCES

- [1] B. Stenkvis et. Al., "Computerized nuclear morphometry as an objective method for characterizing human cancer cell populations" Cancer research, vol.38, no.12, pp.4688-4697,1978.
- [2] L. He, L. R. Long, S. Antani, and G. R. Thoma, "Histology image analysis for carcinoma detection and grading", Computer methods and programs in biomedicine, vol.107, no.3, pp.538-56, 2012.
- [3] Eltoukhy, M. Meselhy, and I. Faye. "An Optimized Feature Selection Method For Breast Cancer Diagnosis in Digital Mammogram using Multiresolution Representation."Applied Mathematics ,vol.8, no.6, pp.2921-2928, 2014.
- [4] C.A. Pena-Reyes, and M. Sipper, " A fuzzy-genetic approach to breast cancer diagnosis", Artificial intelligence in medicine, vol.17,no.2, pp.131-155, 1999.
- [5] B. Fatima, and C.M. Amine, "A neuro-fuzzy inference model for breast cancer recognition", International Journal of Computer Science & Information Technology, vol.4,no.5,pp.163, 2012.
- [6] F. K. Ahmad, and N. Yusoff, "Classifying breast cancer types based on fine needle aspiration biopsy data using random forest classifier", in 13th International Conference on Intelligent Systems Design and Applications (ISDA),pp.121-125,2013.
- [7] E. Meijering, O. Dzyubachyk, and I. Smal, "Methods for cell and particle tracking", Methods Enzymol, vol.504, no.9, pp.183-200,2012.
- [8] M. A. Bray, M. S. Vokes, and A. E. Carpenter, "Using CellProfiler for automatic identification and measurement of biological objects in images", Current Protocols in Molecular Biology, pp.14-17, 2015.
- [9] M. N. Gurcan et.al., "Histopathological image analysis: a review," Biomedical Engineering, IEEE Reviews ,vol.2, pp.147-171, 2009.
- [10] M. Veta, J. P. Pluim, P. J. V. Diest, and M. A. Viergever, "Breast cancer histopathology image analysis: A review", IEEE Transactions on Biomedical Engineering, vol.61, no.5, pp.1400-1411,2014.
- [11] K. Gurney, "An introduction to neural networks",CRC press, ch.6, pp.41-56, 1997.
- [12] G. Schohn, and D. Cohn, "Less is more: Active learning with support vector machines", In International Conference on Machine Learning (ICML), pp.839-846,2000.

An Empirical Approach to Autonomous GSM BTS based on OSS and OSH

Md. Habibur Rahman

Department of Electrical and Electronic Engineering
Chittagong University of Engineering and Technology
Chittagong-4349, Bangladesh
habibur.habib.bd@gmail.com

Md. Mamunoor Islam

Department of Electrical and Electronic Engineering
Chittagong University of Engineering and Technology
Chittagong-4349, Bangladesh
mcm.mithu@gmail.com

Abstract— In this era of wireless engineering, Open source application has turned out to be an evolving concept in the development of GSM technology. Nowadays, GSM (Global System for Mobile Communications) networks are receiving increasing attention in the open-source community. Hence, aim of this paper is to illustrate the process of implementation of autonomous GSM BTS (Base Transceiver Station) by using OSH (Open Source Hardware) such as BladeRF device, a multipurpose signal processing board used for Software Defined Radio (SDR) and YateBTS, an OSS (Open Source Software) acts as a GSM access point which is software based. Generally, YateBTS permits available standard GSM phone to be employed as SIP (Session Initiation Protocol) endpoints in VoIP (voice over IP) network system. An experimental study has been conducted to test the feasibility and effectiveness of the implemented system. The outcome of experimental result is thoroughly examined in this paper.

Keywords— OSS; OSH; SDR; YateBTS; BladeRF

I. INTRODUCTION

GSM (Global System for Mobile communication) network has turned out to be the most broadly spread wireless network in this modern world within two decades [1]. The technology denotes a robust voice communication which is the most popular standard in the world covering 90% of the global population [2].

The number of mobile subscriptions is growing across all regions which have grown into 7.2 billion where 108 million are the new subscriptions in 2015 [3]. Hence the research and industry worlds have focused a lot on GSM sector for low-cost implementation, easy to use, customization and heightened security. For this purpose, reconfigurable open hardware and open source software are very much obligatory.

In open source software, Osmocom offers more network control and scalability but it needs commercial Base Transceiver Station (BTS) hardware and the source code should be closed. The price of that hardware is also comparatively very high [4]. Recently, Range Networks has developed OpenBTS where the source code is completely open but complexity arises when it is configured. Though it is most popular, OpenBTS does not permit multi-cellular network configuration [5]. Later, developer David Burgess along with Legba, Inc and Null Team has developed an OSS

named YateBTS which reduces the inadequacy of OpenBTS [6].

In the hardware section, Ettus (a software defined radio manufacturing company) products such as USRP series are more popular than other device. But the hardware is not completely OSH and it doesn't provide deep details rather than the basic block. Moreover, different applications require different daughter boards which increase the setup complexity and cost also [7]. The developer of OpenBTS named Range Networks also offers some hardware that may not be used as multipurpose rather than as BTS application [8]. Umtrx is another board available for this purpose but fixed application is required where frequency range is limited. MyridRF is also one of OSH but it is not commercially available at this moment. HackRF may be considered as a great choice for multi-application but has lacking of being only simplex [9]. Another company name as Naund provides this type of SDR boards (two models such as BladeRF x115 and BladeRF x40) which has extremely low priced compared with others. It has wide range of frequency and also suitable for multiparous application. Moreover, it doesn't need extra daughter board for different types of application [10].

This paper has been structured as follows: Section II shows the related works in the proposed area. Hardware and software sections are explained in Section III and Section IV. In Section V, the implementation has been discussed. Finally future work and conclusion about the whole work are drawn in Section VI and VII.

II. RELATED WORK

Numerous approaches have been conducted in recent years with the aim of investigating the potential of GSM BTS using SDR. Most of them are basis of OpenBTS and Ettus's USRP. H. Samra and David A. Burgess made the first approach in order to reduce the cost of GSM service [11]. In 2010, Range Networks has been launched as a startup company for the commercialization of OpenBTS based products in a DEMO conference [12]. Later, D. Burgess collaborated with the Null Team who were the developers of Yate and launched as OSS named YateBTS in 2014 [4].

Meanwhile, many works has been done in this sector. In 2010, the first installation approach has been made in a small island named Niue (about 1700 people) by a

telecommunication company where the OpenBTS system was introduced [13].

In 2012, A. Anand (et.al) has proposed a village cell concept using OpenBTS and Asterisk which was low cost cellular connectivity for rural areas [14]. In 2012, the team of Ninja Networks had set up a "NinjaTel Van" at Rio Hotel located in Las Vegas where OpenBTS system has been employed and 650 GSM phones were served by a small network with custom SIM cards [15].

Moreover, a cost effective GSM telephony project experimented in region of amazon has been proposed by M. Cabral (et.al) which is also constructed on open hardware and software system [16]. In 2014, Zhilin Li (et.al) has also proposed a GSM Base Station which was based on OpenBTS and USRP (Universal Software Radio Peripheral) [17].

III. HARDWARE

As previously described, BladeRF board is completely open source type hardware and there is no complexity for modification and the price is also too less than other duplex SDR device. The complete list of hardware is given below.

- BladeRF Board (OSH).
- Two GSM Antennas.
- General Computer.

A. BladeRF

It provides wide band frequency range (300MHz - 3.8 GHz) with fully bus-powered USB 3.0. From using experience, data rate is excellent when attached on USB 3.0 port. Some of the features of this board which push the young researchers are 12-bit RX/TX quadrature sampling which is independent, configurable 2x2 MIMO equipped with SMB cable which can be expanded up to 4x4. It is also capable of achieving 28MHz full duplex channels. Moreover it offers 16-bit DAC factory which is adjusted at 38.4MHz, On board 200MHz ARM9 equipped with 512KB embedded SRAM, On board 40KLE or 115KLE where JTAG port is also available. It also offers board design with modular expansion to add GPIO, Ethernet, expanding range of frequency and also power limits [18]. When the OS is based on Linux kernel, this device is really stable. In this board, an Altera Cyclone IV FPGA has been used which runs the interface in the middle of RF transceiver and FX3 [18]. The bladeRF can employ a bandwidth of 27.8 MHZ over its effective bandwidth with negligible spurs.

B. GSM Antenna

In this project, MULTICOMP IR-GSMPB-02-C5 antenna has been used. The frequency of this antenna is 824 to 960, 1,710 to 1,990 MHz with 0 dBi gain. The connector type is SMA and length is 87 mm which fulfills the requirement [19].

C. General Computer

A general computer with microprocessor such as Intel 1.6

GHz is enough for this operation. But for high speed operation and huge number of M2M connation, USB 3.0 included PC is essential. A USB 2.0 is suitable up to 13 mobile connections. For more connections, USB 3.0 is mandatory.

IV. SOFTWARE

Complete list of software for GSM BTS using BladeRF is given below.

- Yate telephone engine.
- YateBTS.
- Building and Installing BladeRF with FPGA.

A. Yate telephone engine

Yet Another Telephony Engine (Yate) which is known as open source software supports voice, videos as well as instant messaging [20]. The software is basically written in C++ language equipped with modular design which also allows other scripting languages such as Python, Perl and PHP for creating outward functionality [20]. At first, it is essential to install 'Yate'. As Linux kernel is chosen, every software/bundle/package is installed via terminal.

B. YateBTS

YateBTS is generally comprised of L1 PHY, L2 link layer and L3 radio resource manager of OpenBTS. Here, all functions are executed in Yate which offers some enormous advantages from the original OpenBTS SIP interface such as radically enhanced stability, direct support for all Yate supported protocols (ISDN, IAX, XMPP/Jabber/Jingle, MGCP, H.323 and others) and many other features of Yate such as SNMP etc. [20].

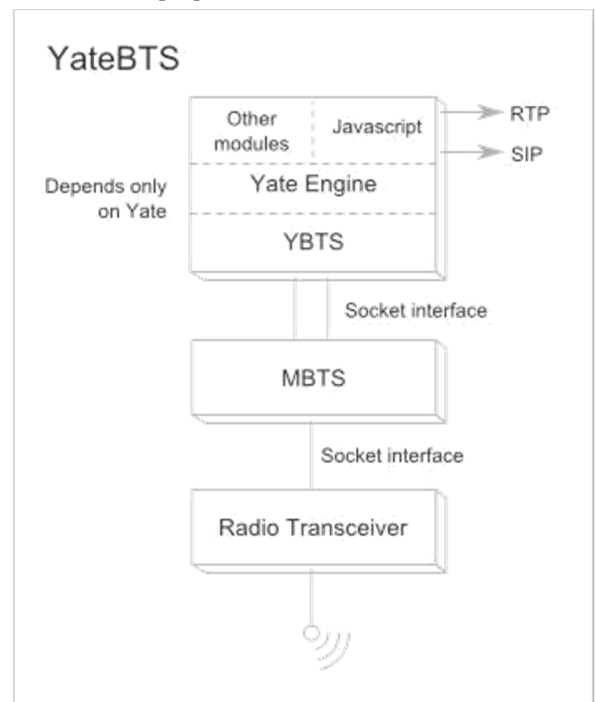


Fig. 1. Internal structure of a YateBTS base station [20]

Fig. 1 describes the internal structure of YateBTS base station. Firstly, YateBTS is composed of two major portions.

One is the lower layer comprised of radio transceiver and MBTS which usually handle the GSM part of the system. Another is the network layer controlled by Yate which consists of YBTS, the Yate module and Yate application modules such as JavaScript for outbound calls or SIP/IAX. Secondly, the MBTS links with the radio transceiver and the network layer through a socket interface. RTP and SIP can be handled by the JavaScript module [20].

C. *Building and Installing BladeRF with FPGA*

The next step is building and installing the BladeRF host libraries and tool in a Linux system. After that it needs quickly getting a device up and then to run using pre-built firmware and FPGA images. An important issue that should be remembered if anyone use libbladeRF v0.17.0 or newer must have use this patch but in lower version no need for patch.

V. IMPLEMENTATION

Inserting all hardware and installation software properly, the first task is to search the network in implementation process. For YateBTS, SIM (Subscriber Identity Module) card is not mandatory. Even with existing mobile operator network SIM, it is also possible to connect. It is another advantage over OpenBTS. When phone is connected, BTS replies with a welcome message and then allows the phone to be associated. After that BTS sends a desired 9 digits unique number through a SMS where first two digits define country code which is reconfigurable.



Fig. 2. SMS with 9 digits unique number

After connecting more than 1 mobile in the network, the system is ready for testing the voice call and SMS operations. Fig. 3 and Fig. 4 illustrate the practical implementations of voice call and SMS operation below.

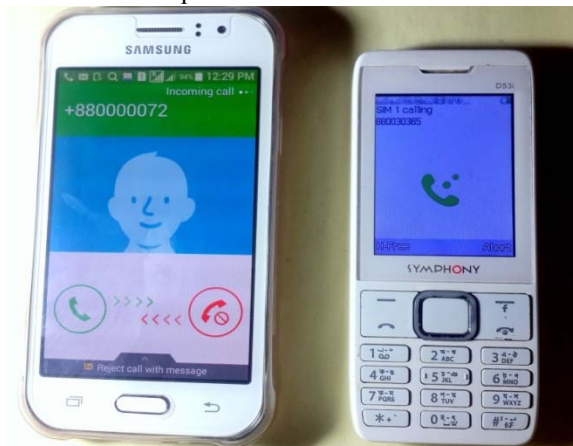


Fig. 3. Implementation of Voice Call within the network



Fig. 4. SMS Operation Implementation

In Yate terminal, different command is available for monitoring the number of listed phone within the network, the number of generated call, number of text message, mbts gain, noise, advanced option for GPRS and also controlling of SGSN/GGSN command. Fig. 5 illustrates Yate server terminal where list of registered phone can be observed.

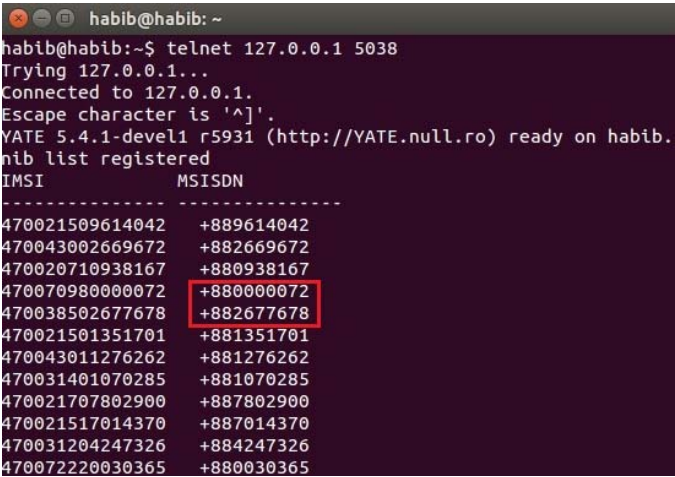


Fig. 5. Yate server terminal

Here, BTS functioning is crowded by electronic equipment and the power available is very limited in order not to overlap with the existing providers. Further issues were mostly related to the interference and the coverage of the BTS. Because of other systems interference, the ARFCN (Absolute Radio Frequency Channel Number) is initially chosen 992 (hence the value of $n=124$, uplink channel frequency is 883.6 MHz and downlink channel 928.6 MHz) which shows not strong enough to permit a standard phone call. Hence, mobile devices have to frequently re-search for available networks by wasting their batteries. When network is live, the related frequency is captured by rtl-sdr dongle and GNU radio 3.10 version. Both FFT (Fast Fourier Transform) and waterfall display of downlink frequency and uplink frequency is illustrated below which are exactly the calculated ones.

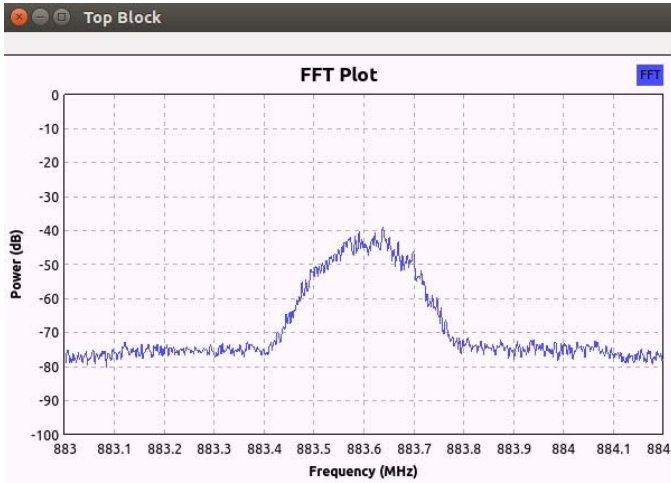


Fig. 6. FFT diagram of Uplink Frequency

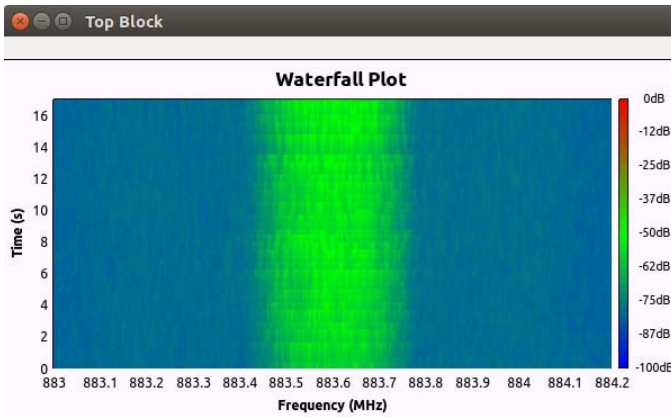


Fig. 7. Waterfall display of Uplink Frequency

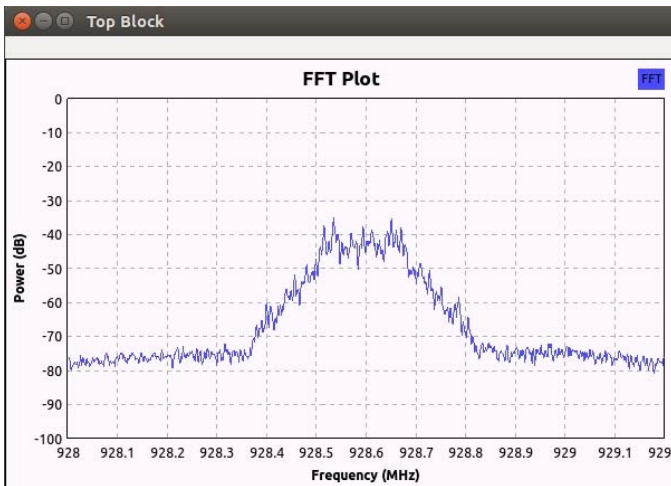


Fig. 8. FFT diagram of Downlink Frequency

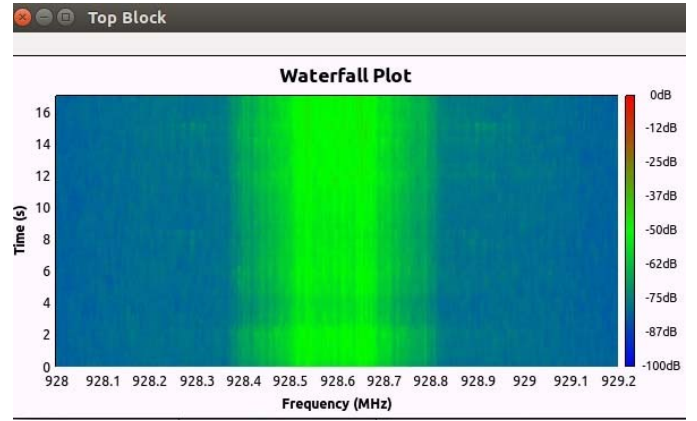


Fig. 9. Waterfall display of Downlink Frequency

The overall summary of the implemented results has been given below.

- Coverage Area = 78 square meter
- 7 GSM to GSM calls established and successfully terminated (Under USB 2.0)
- 17 GSM to GSM calls established and successfully terminated (Under USB 3.0)
- Bandwidth for one active GSM call as like as traditional.
- Delay for voice call= negligible. But .2 sec when number of call is more than 16
- Number of Text message delivered=43 (with 1% error)

VI. FUTURE WORK

The future target of this project is to cover a small village or any campus (such as Chittagong university of engineering and Technology) by using several BladeRF board and a core network system and also to implement handover feature. Another great scope of this project is to make a complete IMSI catcher for the monitoring of the conversation by active or passive cellular monitoring schemes.

VII. CONCLUSION

As the impact of open source hardware and software is effective in the development of cellular GSM technology, enormous opportunities are here to research to improve the GSM network to a great extent. This paper demonstrates such an example in the form of a GSM BTS with the use of new open source software YateBTS and BladRF, a low cost SDR board. These types of open hardware and software create new opportunity for young researchers to learn details about especially on GSM equipment and Protocol also which opens the door for analyzing the encryption and description of GSM data.

REFERENCES

- [1] T. Tsou, T. Cooper, R. McGwier, T. Charles Clancy, J. Reed, "Development of an Open-Source GSM Femtocell and Integrated Core Infrastructure", MILCOM 2012 - 2012 IEEE Military Communications Conference.
- [2] "Internet Society Global Internet Report 2015- Mobile Evolution and Development of the Internet", Published in 2015.
- [3] "Ericsson Mobility Report- on the pulse of the networked society", Published in June, 2015.
- [4] Thomas A. Cooper, "Integration of Open-Source GSM Networks", Published on April 27, 2012.
- [5] N. Prasannan, G. Xavier, A. Manikoth, R. Gandhiraj, R. Peter, K. P. Soman, "OpenBTS based micro telecom model: A socio-economic boon to rural communities", International Multi-Conference on Automation, Computing, Communication, Control and Compressed Sensing (iMac4s), 2013.
- [6] Klint Finley, "Out in the Open: This Super-Cheap Cellphone Network Brings Coverage Almost Anywhere", www.wired.com/2014/06/openbts/, Published on 6 September, 2014 (Retrieved on August, 2016).
- [7] Axelle Apvrille, Fortinet, "OpenBTS for dummies", published on August 31, 2011.
- [8] OpenBTS-UMTS 3G Commercial Desktop Kit www.rangenetworks.com/products/professional-development-kit (Retrieved on August, 2016).
- [9] Michael Ossmann, "HackRF, an open source SDR platform" www.kickstarter.com/projects/mossmann/hackrf-an-open-source-sdr-platform (Retrieved on August, 2016).
- [10] "Nuand - bladeRF Software Defined Radio" www.nuand.com/ (Retrieved on August, 2016).
- [11] Julie Bort, "Burning Man's open source cell phone system could help save the world" www.networkworld.com/article/2217442/wireless/burning-man-s-open-source-cell-phone-system-could-help-save-the-world.html (Retrieved on June, 2016).
- [12] Dean Takahashi, "DEMO: Range Networks rings in cell-phone service for \$2 a month", VentureBeat, September 14, 2010 (Retrieved on August, 2016).
- [13] Burgess, David, "FAKALOF A LAHI ATU, The OpenBTS Chronicles" March 7, 2010 (Retrieved on August, 2016).
- [14] A. Anand, V. Pejovic, D. L. Johnson, E. M. Belding, "Village Cell: Cost Effective Cellular Connectivity in Rural Areas", International conference on information and communication technologies and development (ICDT), Atlanta, GA, USA, March 2012.
- [15] "At Defcon, hackers get their own private cell network: Ninja Tel". Ars Technica. 2012-07-28 arstechnica.com/security/2012/07/ninja-tel-hacker-phone-network/ (Retrieved on August, 2016).
- [16] M. Cabral, I. Almeida, C. Melo, A. Klautau, "Low-cost GSM telephony in the Amazon region based on open-source / open-hardware projects", 2009 IEEE Latin-American Conference on Communications, Medellin, 10-11 Sept. 2009.
- [17] Zhilin Li ; Jian Tang ; Xuejie Zhu ; Caihong Kai, "Simple GSM base station based on universal software radio peripheral", International Conference on Computing, Communication and Networking Technologies (ICCCNT-2014), Hefei, 11-13 July 2014.
- [18] DKARS Dutch Kingdom Amateur Radio Society, Edition 6, published on, December 2014.
- [19] www.newark.com/multicom/ir-gsmpb-02-c5/antenna-gsm-umts-penta-band/dp/25R4657 (Retrieved on August, 2016).
- [20] wiki.yatebts.com/index.php/About_YateBTS (Retrieved on August, 2016).

Design and Implementation of Maximum Power Point Tracking Solar Charge Controller

Md. Rokonzaman ^{*} and Md Hossam-E-Haider ^β

Department of Electrical, Electronic and Communication Engineering (EECE)

Military Institute of Science and Technology (MIST), Mirpur Cantonment, Dhaka-1216

^{*}ragib_isme@yahoo.com and ^βhaider8400@yahoo.com

Abstract—Solar or photovoltaic (PV) system is an alternative clean energy resource that has received much attention in the research and industries. Solar charge controller (CC) is the heart of a solar system. Three common types of charge controller are ON/Off, pulse width modulation (PWM) and maximum power point tracking (MPPT). MPPT is getting very much popularity nowadays for its high capability of power extracting from solar panel. This paper is presenting here an improved design of MPPT solar charge controller using Arduino. The proposed technique significantly reduce system power loss and increase the efficiency. Experimented highest efficiency 97.75% is recorded from the proposed system. Additionally, some new features are added like smart device charging, wireless data logging and protections from high voltage disconnect (HVD), low voltage disconnect (LVD), over current protection, short circuit protection and reverse polarity protection.

Keywords—charge controller, MPPT technique, PV system, arduino, microcontroller, buck converter.

I. INTRODUCTION

With the advancement of semiconductor physics, its extraction process of energy has become more efficient these days. Particularly, perturb and observe (P&O) [1], incremental conductance (INC) [2], constant voltage (CV) [3], short current pulse [4], fuzzy logic control (FLC) [5], [6], artificial neural network (ANN) [7] and some other techniques [8] have been reported to offer an efficient energy extraction process. However, the efficiency of an energy extraction process can further be improved considering the Current-Voltage (I-V) characteristic of a solar cell.

The I-V characteristic of a solar cell is nonlinear and varies with irradiation and temperature [9]. There is a unique point on the I-V or Power-Voltage (P-V) curve of the solar array called Maximum power point (MPP) at which the entire PV system operates with maximum efficiency and produces its maximum output power. When a PV array is directly connected to a load, the systems operating point will be at the intersection of the I-V curves of the PV array and load. However, under most conditions this operating point is not at the MPP [10]. Therefore it is desirable to ensure that the load line passes through the MPP to continuously deliver the maximum power to the output. To achieve this maximum output power and thereby to improve the efficiency, maximum power point tracking (MPPT) technique can be employed in the PV system.

Different MPPT techniques have been developed and implemented [11]–[13]. It is also evident that none of those

have employed a microcontroller based MPPT technique using Arduino. It makes the system simpler, more efficient and highly flexible. This research therefore aims at designing and implementing a microcontroller based MPPT solar charge controller technique using Arduino. Solar panel is the source of nonlinear energy. TSPV-30P solar panel and 12V lead acid battery is used to experiment a PV system using this proposed charge controller. Electrical specification of this panel are mentioned in Table I.

TABLE I. ELECTRICAL SPECIFICATIONS OF TSPV-30P PANEL

Name	Rating
Maximum power	30W
Maximum voltage	17.10V
Maximum current	1.76A
Open circuit voltage, V_{oc}	21.16V
Short circuit current, I_{sc}	1.94A
Operating temperature	-40° C to +85° C

Table II depicted the developed specification against the older version of charge controller and specific uses of the proposed features.

TABLE II. DEVELOPED SPECIFICATIONS OF THE PROPOSED SYSTEM

Feature	Model/Rating	Specification
System	30W, 12V	Input source, Load
Aurduino Mega	ATmega1280	Microcontroller board
Current sensor	ACS712–20A	Protection for over current or short current flow
Buck converter	—	Convert panel voltage to battery voltage
MOSFET driver	IR2104 (10V to 600V)	High speed power MOSFET drivers
Wireless Fidelity (Wi-Fi) module	ESP8266	Data logging
USB connector	USB 2 (480 Mbit/s)	Direct charging of external devices
Display	DIS-00014 (20 × 4)	Monitoring system status

II. PROPOSED SYSTEM

Aurduino MEGA 2560 microcontroller is used to control the charge controller operation. This board is designed for more complex projects or research based on ATmega2560. Fig. 1 shows the pin-out configuration of Arduino MEGA 2560. It contains 54 digital input/output (I/O) pins, 16 analog

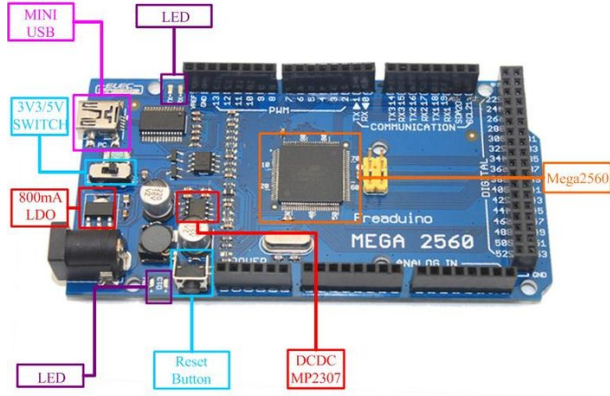


Fig. 1. Pin-out of Arduino MEGA 2560 [14].

inputs, 4 hardware serial ports, 16 MHz crystal oscillator, a power jack with universal serial bus (USB), an ICSP header and a reset button. It also facilitates a larger space to sketch desired design. Port A0 is used to solar voltage divider, port A1 to output of the ACS712 current sensor, port A2 for voltage divider of battery, A4 and A5 for liquid crystal display (LCD) display, port D2 for transmitting data to WiFi module ESP8266 and D3 for receiving. D5 for controlling back light control of the display, D6 for load control, D8 and D9 to control the MOSFET driver, and port D14-D16 is used for light emitting diode (LED) indication.

A. Experimental Setup

Fig. 2 shows block diagram of the proposed charge controller. This block diagram is divided into three portions namely source, control and load. 30W solar panel for source, 12V lead acid type rechargeable battery as load and a charge controller for control the both portion. Voltage sensor will sense the voltage and output is given to the microcontroller.

To sense current, ACS712-20A current sensor is used. Fig. 3 shows the pin diagram of ACS712. VOUT pin of this sensor is connected to analog pin (A1), VCC pin connected to 5V pin and GND pin to ground (GND) pin of Arduino. A buck converter circuit is used to convert panel voltage to load voltage. The schematic diagram of a buck converter is shown in Fig. 4, that is a composition of An inductor, capacitor and IRFZ44N MOSFET. When the MOSFET is ON, current will flow the inductor, load and output capacitor. Since the diode is reversed biased thus no current will flow through the diode. In this state magnetic energy will store in the inductor and electrical energy in the capacitor. When, this MOSFET is off stored energy in the inductor released and current complete its path through the diode and in the same time stored energy of capacitor will supply to the load and will control the current flow. IRFZ44N MOSFET is used for better and quick switching. Since it has very low amount of voltage drop it increases the total efficiency of the system. Liquid crystal display (LCD) with 20×4 character is used to display panel and battery voltage, load condition with pulse width modulation (PWM). To reduce wire density in the circuit a I2C connector is used. It only takes four wires to interface with the display and circuit. Finally a Wi-Fi module is used to get data from the system to send a fixed server through

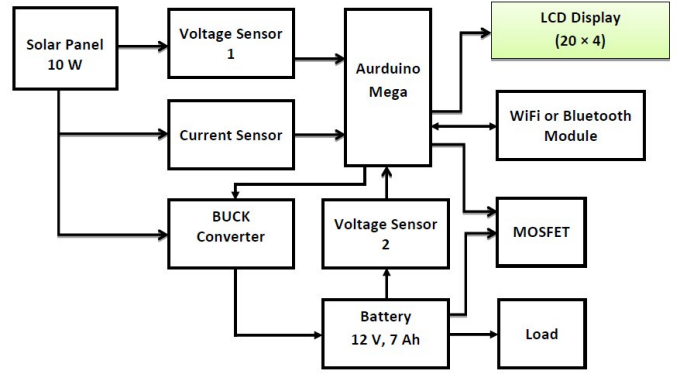


Fig. 2. Block diagram of the proposed solar charge controller.

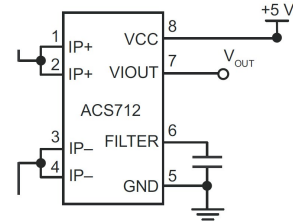


Fig. 3. Pin diagram of ACS712-20A

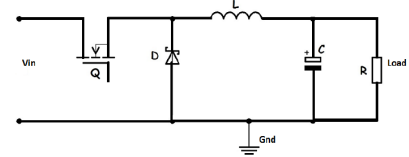


Fig. 4. Schematic diagram of a Buck converter

internet. This facility will enable remote access over the system to monitor.

B. Inductor Selection

Inductor has an important role to design any kind of converter. To design a buck converter an inductor will be used. Some parameters are assumed to calculate the inductance.

Let,

$$\text{Output voltage, } V_{out} = 12 \text{ V}$$

$$\text{Input voltage, } V_{in} = 17 \text{ V}$$

$$\begin{aligned} \text{Output current, } I_{out} &= \frac{\text{Panel power}}{\text{Output voltage, } V_{out}} \\ &= \frac{30 \text{ W}}{12 \text{ V}} \\ &= 2.5 \text{ Amp} \end{aligned}$$

$$\text{Switching frequency, } F_s = 50 \text{ kHz}$$

$$\begin{aligned} \text{Duty cycle, } D &= \frac{V_{out}}{V_{in}} \\ &= \frac{12 \text{ V}}{17 \text{ V}} \\ &= 0.7058 \\ &= 71\% \end{aligned}$$

Now, Inductor, $L = (V_{in} - V_{out}) \times D \times 1/F_s \times 1/\Delta I_L$ (1)

Where, ΔI_L = Inductor ripple current.

A good estimation for the inductor ripple current is 20% to 40% of the output current. Let,

$$\begin{aligned}\Delta I_L &= 40\% \text{ of the rated current} \\ &= 40\% \text{ of } 2.5 \text{ A} \\ &= 1.0 \text{ A}\end{aligned}$$

Now, from equation (1)

$$\begin{aligned}L &= (V_{in} - V_{out}) \times D \times 1/F_s \times 1/\Delta I_L \\ &= (17 - 12) \times 0.71 \times 1/50,000 \times 1/1.0 \\ &= 0.071 \text{ mH} \\ &= 71.0 \mu\text{H}\end{aligned}$$

$$\begin{aligned}\text{Inductor peak current} &= I_{out} + \frac{\Delta I_L}{2} \\ &= 2.5 + 1.0/2 \\ &= 3.0 \text{ A}\end{aligned}$$

C. Capacitor Selection

The output of buck converter has a amount of ripple. Capacitor is used to minimize the voltage overshoot and ripple.

$$\text{Output Capacitance, } C_{out} = \frac{\Delta I_L}{8 \times F_s \times \Delta V_{out}} \quad (2)$$

Where, Inductor ripple current, $\Delta I_L = 1.0 \text{ A}$

Switching frequency, $F_s = 50 \text{ kHz}$

Output voltage ripple, $\Delta V_{out} = 20 \text{ mV}$ (assume)

Now, from equation (2),

$$\begin{aligned}C_{out} &= \frac{\Delta I_L}{8 \times F_s \times \Delta V_{out}} \\ &= \frac{1.0}{8 \times 50000 \times 0.02} \\ &= 125 \mu\text{F}\end{aligned} \quad (3)$$

D. MOSFET selection

IRFZ44N MOSFET is used, its volatge rating is greater than 20% of rated voltage. Current rating is greater than 20% of rated current. It has minimum conduction loss.

III. CIRCUIT SIMULATION

Fig. 5 shows the simulation circuit diagram of the buck converter, designed by Proteus software. Q1, Q2 and Q3 are three IRFZ44N MOSFETs. IR2104 MOSFET driver is

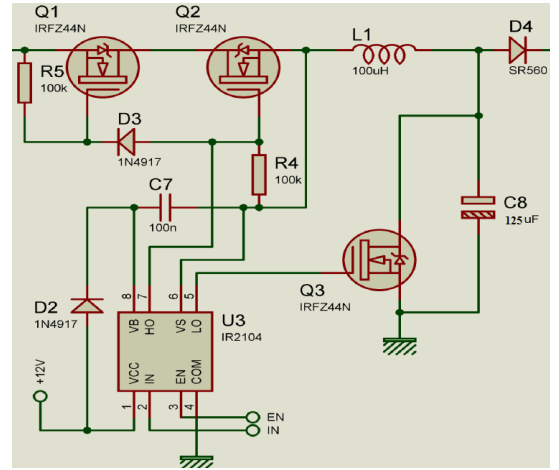


Fig. 5. Simulation circuit of buck converter

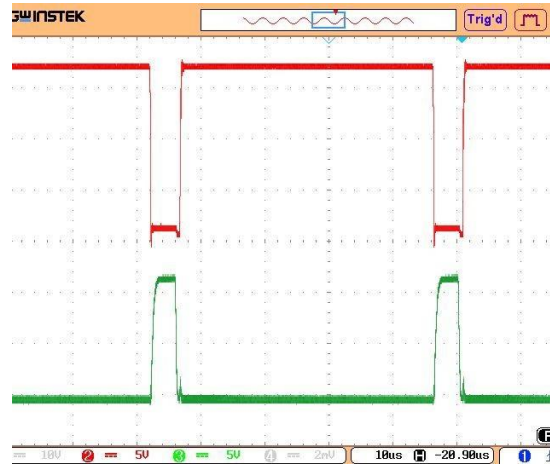


Fig. 6. PWM pulses from MOSFET

used that accept low power input from the microcontroller and produces a high current drive input for the gate of the MOSFETs. Fig. 6 shows the output of MOSFETs. This pulse width modulation (PWM) signals are represents the duty cycle.

Here, 80%-90% duty cycle is used that increased the efficiency of the system. For, hardware implementation an inductor bank is used to provide proper inductance. To design this system, desired inductor value is $250 \mu\text{H}$.

IV. HARDWARE DESIGN AND FIELD TEST

Fig.7 shows the experimental set up bench and circuit of charge controller, mentioning every equipment used for the system including solar panel, control circuit-Aurduino Mega, battery, inductor bank, 3W solar light as load and a display.

In addition Fig. 7 also shows the full hardware system in testing period and zoomed LCD display. Which represent the full status of the system. There are three columns of data. First column shows the solar panel voltage, current and power condition. Second column shows the battery voltage, current and charging state condition. Third column shows pulse width status with load on or off condition. The same status is directly transferable to a handset or laptop through a Wi-Fi device. This

TABLE III. EXPERIMENTAL DATA AND EFFICIENCY OF THE PROPOSED CHARGE CONTROLLER

Time	Panel			Load			Efficiency (%)	
	Voltage (V)	Current (A)	Power (W)	Voltage (V)	Current (A)	Power (W)	Session-wise	Average
8.00 AM	14.1	1.58	22.278	13.6	1.5	20.4	91.57	91.45
8.30 AM	14.3	1.8	25.74	13.8	1.63	22.49	87.38	
9.00 AM	17.1	1	17.1	13.6	1.21	16.45	96.23	
9.30 AM	16.8	1	16.8	13.6	1.14	15.5	92.28	
10.00 AM	16	1.5	24	13.9	1.16	16.12	67.18	
10.30AM	16.1	1.5	24.15	13.9	1.47	20.43	84.6	
11.00 AM	16.4	1.36	22.304	13.8	1.58	21.8	97.75	
11.30 AM	16.9	0.91	15.379	13.6	1.09	14.82	96.39	
12.00 PM	16.5	1.29	21.285	13.8	1.5	20.7	97.25	
12.30 PM	16.2	1.36	22.032	13.8	1.5	20.7	93.95	
1.00 PM	13.9	1.15	15.985	14.1	1.1	15.51	97.02	
1.30 PM	16.9	0.83	14.027	13.5	1	13.5	96.24	

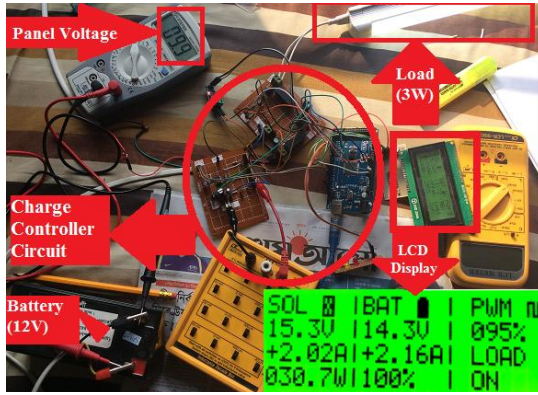


Fig. 7. Experimental Setup Bench.

will enable an advantage to maintain or observe the system from any place connected with internet.

V. EXPERIMENTAL RESULTS

Experimental data of proposed solar charge controller are showing in the Table. III. Which are represents the experimented data of voltage, current and power both of the panel and load. From this tested value the overall efficiency of the charge controller is determined. It's more than 90%. Fig. 8 shows the overall efficiency of the proposed charge controller. In the beginning of the day, At 8.00 AM, panel power and load power was 22.278W and 20.4W respectively. It is also observed that in the same time panel voltage and current was 14.1V and 1.58A respectively. In that time experimented efficiency was 91.57%. At 11.00 AM, the highest efficiency, 97.75% is recorded where in the lowest efficiency was 67.18% at 10.00 AM. Field test shows that the average efficiency of the proposed charge controller is 91.45%.

VI. CONCLUSIONS

In this paper a MPPT charge controller has been designed and tested. The proposed system demonstrate its low cost, low power consumption, microcontroller based control unit and high power efficiency. Microcontroller based alternative was chosen since it permits easy and customized system modifications. The proposed system also facilitates remote

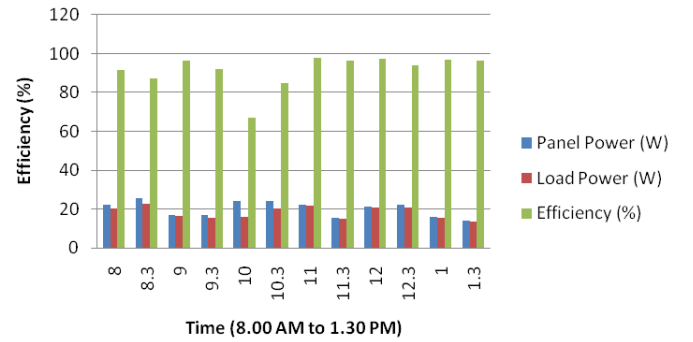


Fig. 8. Efficiency of the Charge Controller

monitoring with data storage, external device charging and continuous system status displaying. Abide by maximum standard of Infrastructure Development Company Limited (IDCOL), Bangladesh, Experimental results from a 30W prototype system have been presented and validate. Tests show that the average efficiency of the charge controller is 91.45%, which is fully satisfying the regulations of Technical Standard Committee (TSC) of IDCOL. Furthermore it can be coupled with more than one renewable source and interfaced to power grid through an inverter.

ACKNOWLEDGMENT

This work has been supported by Research and Development (R&D) fund of Military Institute of Science and Technology (MIST), Mirpur Cantonment, Dhaka-1216.

REFERENCES

- [1] N. Femia, D. Granozio, G. Petrone, G. Spagnuolo, and M. Vitelli, "Optimized one-cycle control in photovoltaic grid connected applications," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 42, no. 3, pp. 954–972, 2006.
- [2] K. S. Tey and S. Mekhilef, "Modified incremental conductance algorithm for photovoltaic system under partial shading conditions and load variation," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 10, pp. 5384–5392, 2014.
- [3] G. Yu, Y. Jung, J. Choi, and G. Kim, "A novel two-mode mppt control algorithm based on comparative study of existing algorithms," *Solar Energy*, vol. 76, no. 4, pp. 455–463, 2004.

- [4] T. Noguchi, S. Togashi, and R. Nakamoto, "Short-current pulse-based maximum-power-point tracking method for multiple photovoltaic-and-converter module system," *IEEE Transactions on Industrial Electronics*, vol. 49, no. 1, pp. 217–223, 2002.
- [5] R. B. Roy, E. Basher, R. Yasmin, and M. Rokonzaman, "Fuzzy logic based mppt approach in a grid connected photovoltaic system," in *8th International Conference on Software, Knowledge, Information Management and Applications (SKIMA 2014)*. IEEE, 2014, pp. 1–6.
- [6] T. L. Kottas, Y. S. Boutalis, and A. D. Karlis, "New maximum power point tracker for pv arrays using fuzzy controller in close cooperation with fuzzy cognitive networks," *IEEE Transactions on Energy conversion*, vol. 21, no. 3, pp. 793–803, 2006.
- [7] W.-M. Lin, C.-M. Hong, and C.-H. Chen, "Neural-network-based mppt control of a stand-alone hybrid power generation system," *IEEE Transactions on Power Electronics*, vol. 26, no. 12, pp. 3571–3581, 2011.
- [8] A. N. A. Ali, M. H. Saied, M. Mostafa, and T. Abdel-Moneim, "A survey of maximum ppt techniques of pv systems," in *IEEE Energytech*, 2012.
- [9] D. Hohm and M. Ropp, "Comparative study of maximum power point tracking algorithms using an experimental, programmable, maximum power point tracking test bed," in *Proc. 28th IEEE Photovoltaic specialists conference*. IEEE, 2000, pp. 1699–1702.
- [10] C. Hua and C. Shen, "Comparative study of peak power tracking techniques for solar storage system," in *Proc. APEC'98*, vol. 2. IEEE, 1998, pp. 679–685.
- [11] T. Esum, P. L. Chapman *et al.*, "Comparison of photovoltaic array maximum power point tracking techniques," *IEEE Transactions on Energy Conversion EC*, vol. 22, no. 2, p. 439, 2007.
- [12] B. Subudhi and R. Pradhan, "A comparative study on maximum power point tracking techniques for photovoltaic power systems," *IEEE transactions on Sustainable Energy*, vol. 4, no. 1, pp. 89–98, 2013.
- [13] B. Hauke, "Basic calculation of a buck converter's power stage," Dallas, Texas, Tech. Rep. SLVA477, Tech. Rep., 2011.
- [14] T. King, "The arduinoinfo.info wiki," [Online; Last accessed 18-May-2016]. [Online]. Available: <https://arduino-info.wikispaces.com/MegaQuickRef>

Depiction of an Interactive Prevarication System during Exigency Situation

Shayhan Ameen Chowdhury, Ahmed Imteaj, Bhaskar Ray and Muhammad Estiak Omar

Department of Computer Science and Engineering

International Islamic University Chittagong

Chittagong-4203, Bangladesh.

Email: shayhan@yahoo.com, imtu_1992@yahoo.com, raybhaskar.ray@gmail.com, estiak.omar@gmail.com

Abstract— Natural calamity is one of the major issues that works as a hindrance in the progression of an underdeveloped country like Bangladesh. The clutch of destructive natural disasters terribly attack millions of people and their property in almost every year. To mitigate such disruption, we developed an android based system through which people can get facilitation during natural calamities. The proposed system can help a user by generating weather update, nearby cyclone centers, hospitals, police and fire stations and also pointing those locations along with direction on the map. The user can also share his current location by SMS and the location map through image sending functionality. In addition, the user can see different reports (both in map and as a custom list view) submitted by the other users about unconventional incidents prevailed in any particular location and can also post a report on any escapade if he wants. The system has also a feature of emergency alert service through which an auto-message is dispatched to some emergency contact numbers.

Keywords—*Natural disaster; distress center; alert service; report; notify; android.*

I. INTRODUCTION

A natural disaster is an unexpected event which occurs due to natural processes of the earth; example includes cyclone, flood, earthquake, tsunamis and other geological processes. Every year, a lot of people died, severely injured or become homeless because of such disasters. During the natural calamities, the telecommunication systems merely work and people can hardly communicate with others. As a result, it becomes even more difficult to tackle the unexpected situation.

During a natural disaster, a lot of people need to shift in the nearby natural disaster shelter. Considering this, we have developed a system where the user can find out the nearby cyclone center along with distance and direction and other stuffs such as: nearby hospitals, police stations or fire stations in an efficient way. In addition, a user can share his location perfectly if the user faces any problem using auto-sending message features of our system. This system contains a reporting section. User can report an incident anytime from anywhere. Other users can observe the details of the incident through the map. This reporting section would help other users to become aware about any incident within an area.

Our proposed system has also a survival strategy section which would help users to get suggestion in different situations and they also get notified by the admin in an emergency. The main contribution of our system is to integrate

a lot of useful functionality in a common platform which works both in online and offline. The system could act as a survival assistant in any distressed situation as it has different informative functionalities such as: Weather update, Search for shelter during emergency, Search for police stations, hospitals and fire stations, Sharing location map by image sending, Incident report shows in the map as well as in custom list view, Incident report post from any location, Providing survival strategy, User's current location sharing through SMS, Sending alert notification during crisis, etc.

We have discussed on some relevant papers in section II. In section III we have explained the system functionality with proper demonstration. Later we have shown experimental result and survey summary focusing the necessity of our proposed system. Section V contains some future extension of our work and finally we conclude the paper.

II. LITERATURE REVIEW

The world is terribly affected by natural calamities every year. A lot of work has already done to reduce the loss due to such disasters. The author in [1] proposed a location based early disaster warning and evacuation system using Google Map. Updates of the disaster (tsunami, cyclone or flood) are put on DMS by the local weather office. To get automatic notification of upcoming disaster device user registers on Disaster Management Server (DMS) else the user can get manual notification. The user keeps in touch with DMS to get most renew data obtained by GSM. But the system does not work in offline and only shows the rescue team on the Google map. Rekha *et. al.* in [2] presented an Emergency Management System (EMS), which enables smart phone based ad-hoc communications at disaster times over Wi-Fi. The person in an emergency or anybody at the emergency site will call the EMS at avail service. Location Coordinates are sent on each request. The system works on the principles of client-Server system, wherein the server responds to the requests of the Clients. The Client and Rescue Application was Implemented as an Android Application and the Sever is implemented as a Web based Java Application. But their system only based on sending messages by the admin to the victim and they did not use any kind of map to show the location in their system. A multi-faceted information portal is developed in [3] for citizens and emergency personnel during disasters and emergency response. This portal is the result of a collaboration between the RESCUE project at UC-Irvine and the City of Ontario, California Fire Department (OFD). The portal provides a wide range of real-time information in crisis

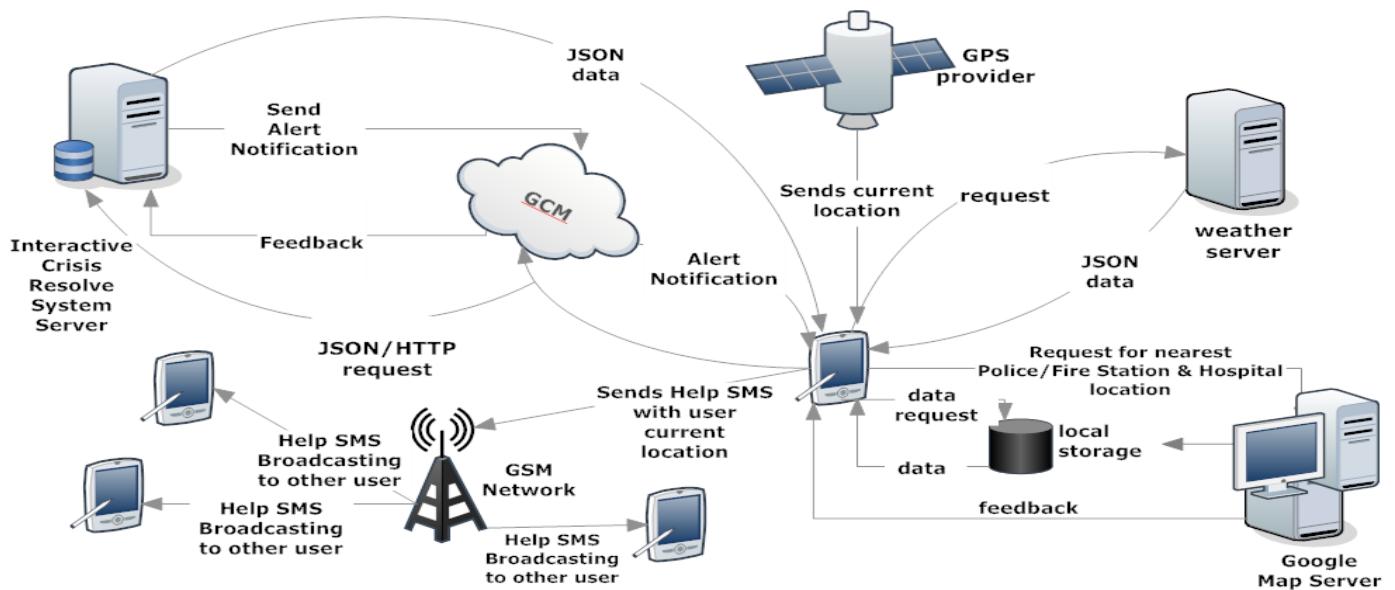


Fig. 1. Proposed system architecture.

situations, such as situation summaries, shelter info. etcshelter information, and aggregated services such as family reunification and donation management. But the paper has a unclear demonstration of their system and they do not provide any emergency alert service generated by the user.

The authors in [4] developed a system named HelpMe, a self-learning opportunistic ad-hoc system, which enables smartphone-based ad-hoc communications at disaster times over Wi-Fi. As the system is implemented for ad-hoc network, it has a huge disadvantage if we consider distance matter.

Catarchi *et. al.* [5] proposed a system which consists of a front- and a back-end layer. The front-end layer is composed of several front-end teams of first responders, and the back-end layer is an integrated peer-to-peer network that lets front-end teams collaborate through information exchange and coordination. But the system has very limited functionalities.

The authors of [6] proposed Area Mail disaster information service provided by NTT Do Como for tsunami alert and evacuation system for helping fishery workers. Area mail service capable of sending notification about damage due to disaster in limited areas, but for overall population it is an inefficient approach.

Australia and South Korea are planning to use satellite communication for disaster management when the failure of mobile network occurs [7], [8], [9].

After the tsunami that hit Sri Lanka on December 26 2004, the Sahana [10] FOSS (Free and Open Source Software) project was built. Key applications of Sahana projects are missing persons registry for finding missing persons, an organizational registry for a relief organization, coordination and tracking, request management system for tracking and matching specific aid or relief requests and offers, camp and shelter registry, volunteer management, inventory management to track supplies, and situational awareness. But

the problem is that it is a web based system and not responsive to use in portable device like mobile. Sahana Disaster Management System doesn't use any GPS system and doesn't support offline mode. So if any power disruption occurs during a disaster or crisis, the system will not work.

Similarly, the authors in [11] also proposed a web-based system named 'The RESCUE Disaster Portal for Disasters and Emergency Response' and has the similar drawbacks.

The author in [12] proposed a disaster management system which can help the user to send his current location using GPS through a message and can see nearby shelter zone on the Google map. But the system cannot work in offline, has limited feature and does not have the capability to display the nearby hospitals, fire service or police stations which are crying need during a disaster.

Shanthini *et. al.* [13] suggested an idea of sending alert service by a trusted party during a disaster. As a result, the registered user will be acknowledged about the upcoming disaster through a message or audio clip. But the user would not be able to report anything about any witness he observed and there is no opportunity for the user to generate alarm service if any problem arises.

A portable mobile phones can play an important role to resolve the crisis or disastrous situation in many ways, such as monitoring, communication, warning dissemination, evacuation, rescue and relief aid. Moreover, the approach of smart phones supporting GPS functionality can play a vital role in the crisis or disastrous situation management.

There are plethora of applications that will benefit users in an emergency on a specific service. But none of them actually provides the integration of various functionalities in a single platform that works both in online and offline. Through our proposed system, the user can get the opportunity to fetch necessary information both in online and offline.

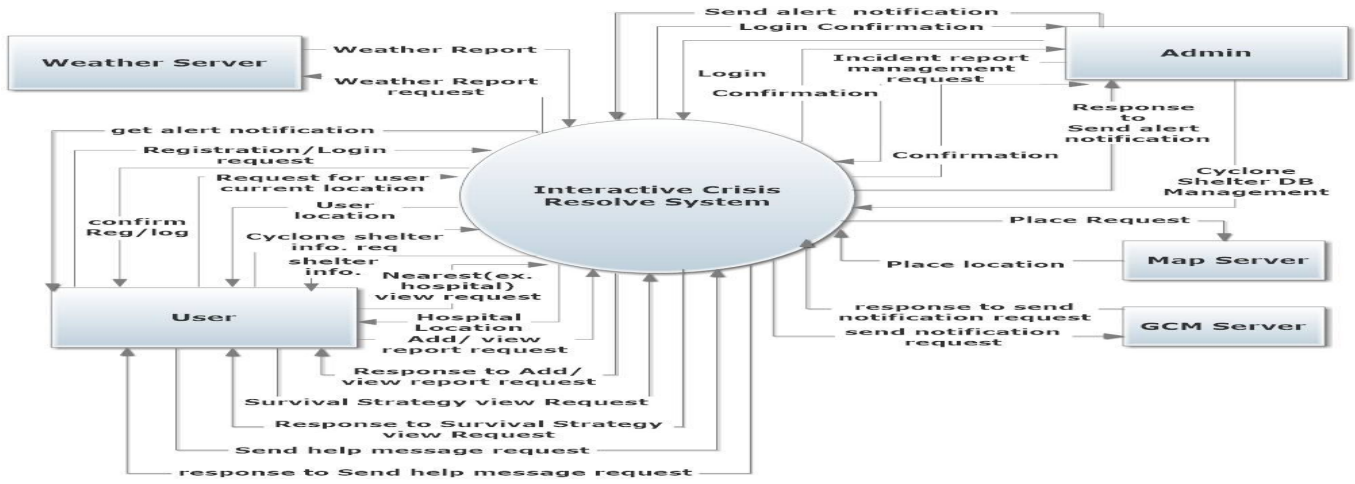


Fig. 2. Context diagram.

III. SYSTEM IMPLEMENTATION

A. System Architecture

The architecture displays the main components of the system (fig. 1). It shows the working of server database and android database, GPS and GCM cloud service. The server database information is fetched by sending HTTP request to the web server and the web-server provides the data in JSON format to android device. GPS satellite is used to get the current location.

B. Context Diagram

The context diagram of our proposed system is presented in fig. 2. Here we have propounded the overall data flow of our system among different entities. The main entities of our system are weather server, admin, user, map server and GSM and the data flows are shown by the arrow sign.

C. Mobile Application

Our proposed system is a mobile application named “Interactive Crisis Resolve System” and for the successful deployment of our system, we have used Eclipse IDE, JAVA, MySQL, SQLite, PHP, Java Script, JQuery, GCM (Google Cloud Message) service, GPS (Global Positioning System) and XML layouts.

1) *User registration and GCM registration process*: When a user does register for the first time in our application, on background a GCM registration process will issue a unique GCM id for user android device. After receiving a unique GCM registration id, our application will send GCM registration id along with user information to our server. Our server will store GCM registration id and user information in the database (fig. 3).

2) *Cyclone Center Module (Online Mode)*: To implement this module initially we have stored district, upzila, union and cyclone shelter information in server side relational database table. We have also designed an interactive user interface for our mobile application end. At first user has to select the

mode, online or offline. By providing district, upazila and union information a disaster victim can find all nearest cyclone shelter locations according to the victim’s present location via the google map (fig. 4, 5).

3) *Cyclone Center Module (Offline Mode)*: During an emergency or disaster situation power disruption or network failure can be occurred. So if there is no internet connection is available during crisis situations, the user can use our cyclone shelter offline module. If a user wishes to see the location of the shelter, an offline OpenStreetMap will display the shelter and user current location. To implement this module we have used SQLite offline database, Osmroid library, and Mobile Atlas Creator. Using Mobile Atlas Creator we have computed an offline OpenStreetMap for Chittagong region (fig. 6).

4) *Nearest Police/Fire Station and Hospital location*: During an emergency situation, this module will find out the nearest police stations, fire brigade stations and hospital location within a radius limit based on user current location. Besides, user can search nearby police/fire station and hospital location for a specific location from the different part of the country shown in fig. 7.

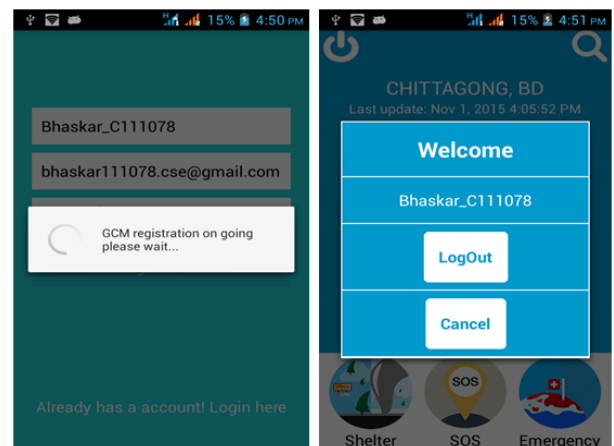


Fig. 3. User registration process.

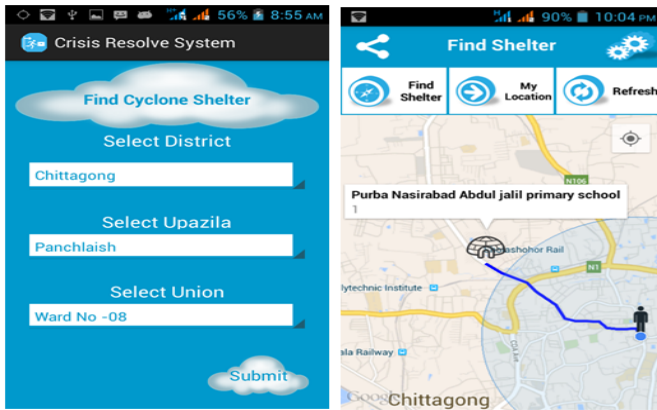


Fig. 4. Online cyclone center module.

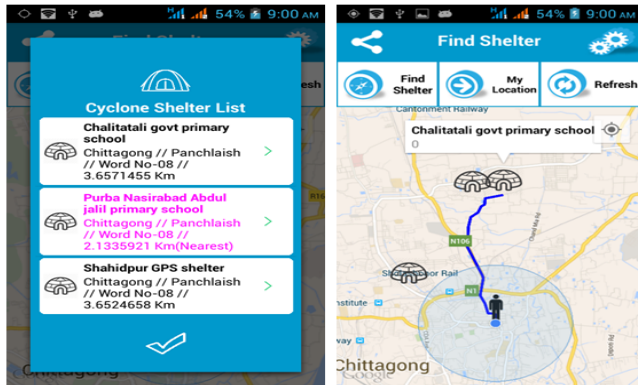


Fig. 5. Online Cyclone shelter details information with distance and path.

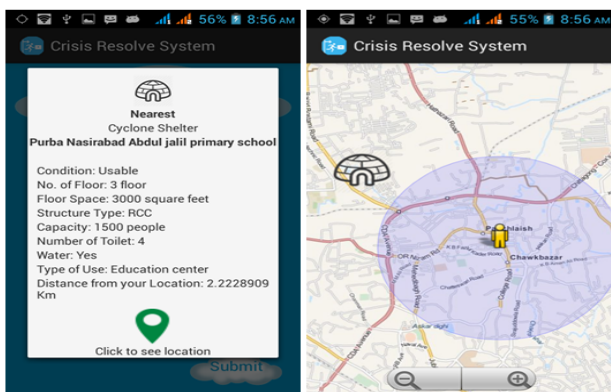


Fig. 6. Cyclone shelter details information with location (offline mode).

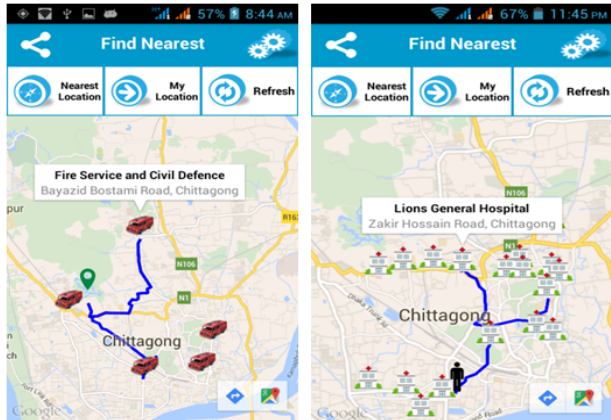


Fig. 7. Nearest Fire Station and Hospital location with path direction.

5) *Incident Report Section*: In this module, report data will be collected from users. This module gives power to the registered user. A user could report what he knows or sees and other user instantly gets informed. A map will show all the events occurred in different location using different marker. If a user wants to report on any incident they have to pick a location from map, then a report form will be appeared. Report form contains title, incident type, severity level, source, description, impact statistics, photo, location and more. User can also take an image of the event instantly and attach it with the report. The user can search incident for a specific location or nearby to user current location. From the searched incident list user can view a specific incident's detail information. User can comment to share his opinion and can see other user comment about any incident (fig. 8, 9,10).

6) *Help Message Sending Feature*: Help message sending feature will let user to save some of his friends or family's phone numbers. This feature sends an alert to those pre-stored contact with a message alerting them that the user needs assistance. Here user can store three contact numbers and an emergency SMS. User can also update this information if needed. We used unique shared preference memory to store this information. So if a user logout anytime and login later to our application, he/she will get this information preserved. And this feature will work in offline mode.

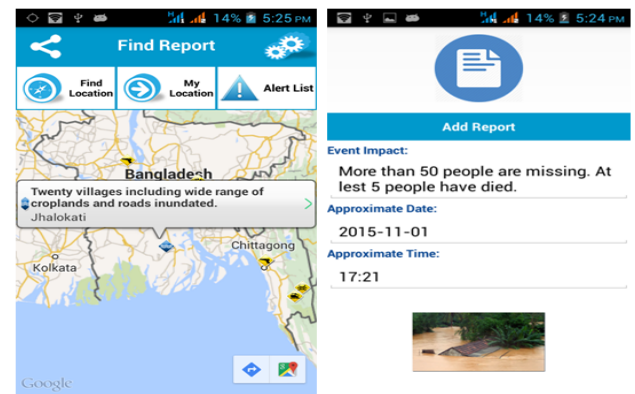


Fig. 8. (a) Incident report map (b) Reporting an incident.

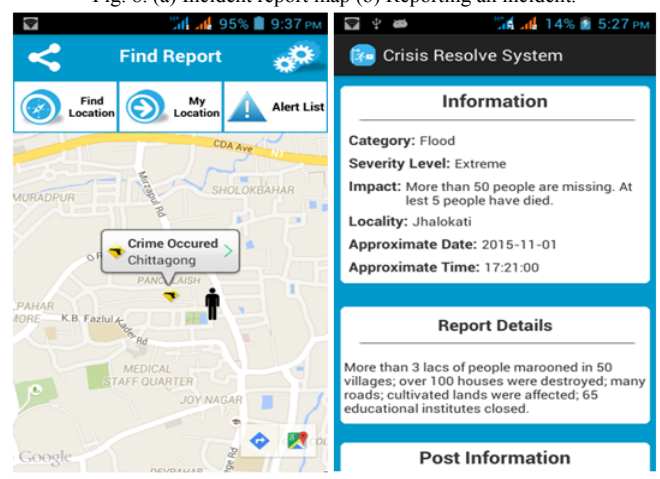


Fig. 9. (a) Report shows for a specific location (b) Detail of a report.

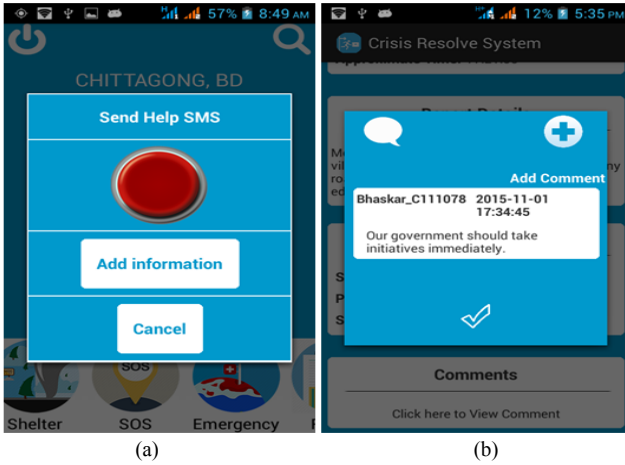


Fig. 10. (a) Help SMS Feature (b) User comment on a report.

7) *Push Notification Alert*: Govt. and institutions can send emergency alert to the citizens in emergency situations through our system. The push notification alert feature will help admin to send alert notification, such as alert signal, emergency news to registered user which is given in fig. 11. Whenever push notification is needed, our server will send a message to GCM(Google Cloud Message) server along with device registration id which is stored earlier in the databases. GCM server will deliver that message to all registered user mobile device using device unique registration ID. All registered user will receive this alert notification immediately. If any emergency contact number, web link is shared with an alert message, the user can directly call those numbers or visit those links from our application.

8) *Survival Strategy*: We know the world is constantly changing and becoming volatile each and every day. This feature includes instruction a person may use in those unexpected situations to save themselves. Here we have propounded a slight portion of this feature in fig. 12(a).

9) *Weather Update*: This feature brings all weather information available for the locations all around the world. The system generates current temperature, humidity, pressure, cloud condition for a specific location which is presented in fig 12(b).

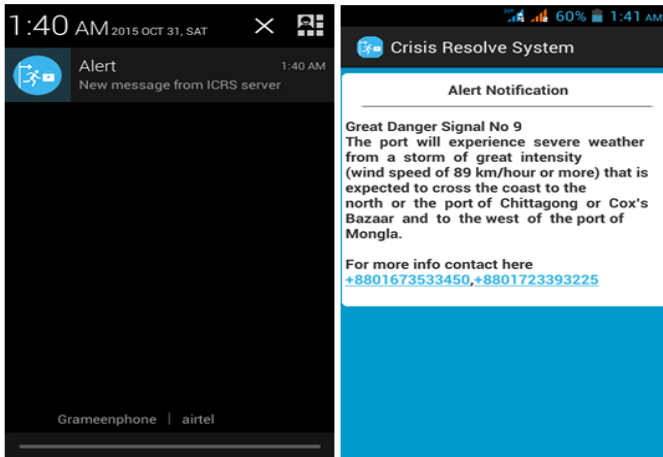


Fig. 11. Notification received by registered users.

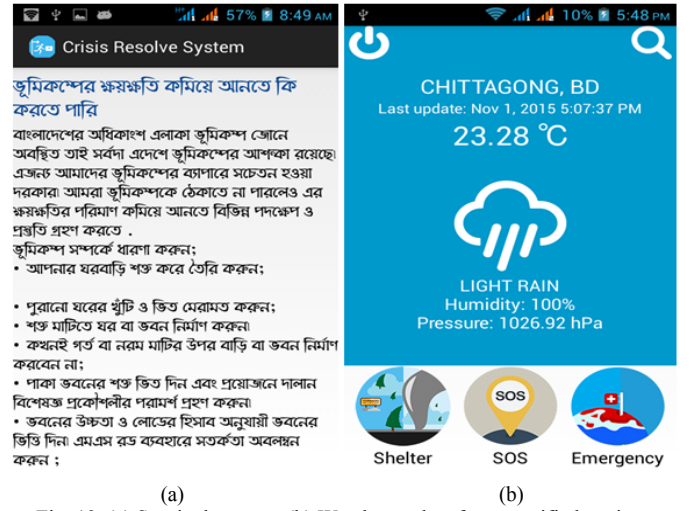


Fig. 12. (a) Survival strategy (b) Weather update for a specific location.

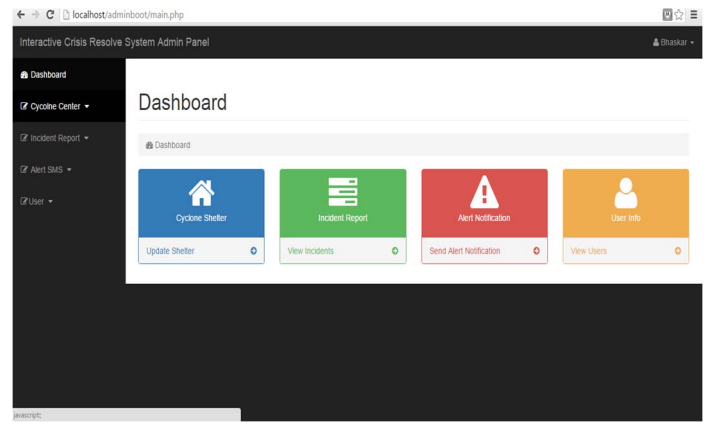


Fig. 13. Admin panel of the system.

10) *Admin Panel*: In admin panel, the admin will be able to login, perform insert, update, delete operation on the shelter information table. Admin can verify, view and delete the incident report submitted by the registered user, send push notification alert to the users which is shown in fig. 13.

IV. EXPERIMENTED RESULT

After developing our system, we have tested it by assuming that a cyclone will hit the coastal region of Chittagong. We have selected 15 different users and installed our application in their smartphones. The total scenario is given with time in table 1.

TABLE I. TESTING OF THE PROPOSED MODEL

Time	Scenario
5.00 pm	The experiment starts. Suppose weather office has announced about a new warning of a cyclone in the coastal region of Chittagong.
5.05 pm	Admin panel sends alert message to all registered users in Chittagong.
5.06 pm	Users received the alert message.
5.07 pm	User search for nearby shelter location. The system tracks user current location and generate all the nearby

	shelter location pointing on the map.
5.08 pm	The user sends an image of a current location map and emergency help message to the pre-stored three numbers and follow the map.
5.18pm	Three people injured and need to admit in a hospital immediately. The user search for the nearby hospital at once and the system computes all the nearby hospitals with approximate distance and direction on the map.
5.21 pm	Create a report about the disaster incident in Chittagong city to other users.
5.23 pm	Other users comment and share their opinions.

We have analyzed the other disaster management systems and find out their qualities. After that, we have compared our proposed system model advantages than the related works and the result is given in table 2.

TABLE II.
SOFTWARE QUALITIES IN SOME EXISTING DISASTER MANAGEMENT SYSTEMS AND OUR PROPOSED SYSTEM MODEL

Disaster Management Systems	Software Qualities
Envimon [14]	Availability Interoperability
Intergraph GeoMedia® [15]	Interoperability Modifiability
Sahana [10]	Modifiability Portability Scalability Usability
Decision Support Systems [16]	Modifiability
Interactive Crisis Resolve System (our proposed system)	Availability Modifiability Interoperability Portability Scalability Usability

After constructing the application, we displayed it to various users to observe their responses. We installed the application in the user's smartphone and ask them whether the application is helpful during disaster situations. The user responses are pictorially shown in fig. 14.

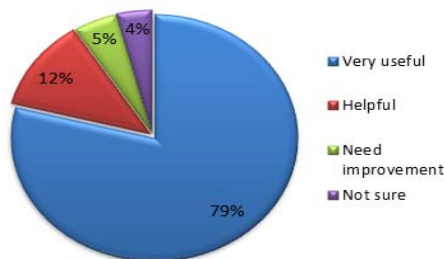


Fig. 14. User response.

V. CONCLUSIONS

Our proposed system can help a user to overcome in different crisis situations. The system not only help a victim in a disaster situation, but it also aware other people about any

unexpected incident. The push notification alert feature also helps administrator to send alert notification, such as, disaster alert signal, emergency news to the registered users. During natural calamities, when the network does not work properly, our system can work perfectly as it has the capability to work both in online and offline. In the future, we have a plan to make the system in Belgali and to implement the system in windows and iOS platform also. Hopefully, the system works as an rescue assistant and helps the user to tackle any hurdle situations.

REFERENCES

- [1] Amit Gosavi, S. S. Vishnu, "Disaster Alert and Notification System Via Android Mobile Phone by Using Google Map," Internal Journal of Emerging Tehnology and Advanced Engineering, Vol. 4, issue 11, Nov. 2014.
- [2] Rehka Jadhav, Jwalant Patel, Darshan Jain, Suyash Phadhtare, "Emergency Management System Using Android Application," International Journal of Computer Science and Information Technologies(IJCSIT), Vol. 5 (3), 2014.
- [3] Jay Lickfett, Naveen Ashish, Sharad Mehrotra, Nalini Venkatasubramanian and Jacob Green, "The RESCUE Disaster Portal for Disasters and Emergency Response," in Proceedings of the 5th International ISCRAM Conference, Washington, DC, USA, May 2008.
- [4] Mokryn, Osnat, et al. "Help me: opportunistic smart rescue application and system," in Proceedings of the 11th Annual Mediterranean Ad Hoc Networking Workshop (Med-Hoc-Net), pp. 98-105, Jun. 2012.
- [5] Catarci, Tiziana, et al. "Pervasive software environments for supporting disaster responses," Internet Computing, vol. 12, issue 1, pp. 26-37, 2008.
- [6] Teshirogi, Yasuaki, et al. "A proposal of tsunami warning system using area mail disaster information service on mobile phones," International Conference on Advanced Information Networking and Applications Workshops, pp. 890-895, 26-29 May, 2009.
- [7] Park, Dugkeun. "One of the Nowcasting Applications: Early Warning Systems for Natural Disasters in Korea." (2006).
- [8] Duke H. Jeong, "National Disaster Warning System in Korea." (2009).
- [9] Anas Aloudat and Katina Michael, "Toward the regulation of ubiquitous mobile government: a case study on location-based emergency services in Australia," Electronic Commerce Research, pp. 31-74, 2001.
- [10] Samaraweera, Isuru, and Sheran Corera. "Sahana victim registries: Effectively track disaster victims," in Proceedings of the 4th International Conference on Information Systems for Crisis Response and Management (ISCRAM), 2007.
- [11] Lickfett, Jay, et al. "The RESCUE disaster portal for disasters and emergency response." in Proceedings of the 5th International ISCRAM Conference, 2008.
- [12] Varsha S. Sonwane, "Disaster Management System on Mobile Phones Using Google Map," in Proceedings of International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 5, 2014.
- [13] S.Shanthini , Prof Mr.T. Jeba Moses, "Alert System from Disaster Management Server (DMS) By Trusted Disaster Party," in Proceedings of International Journal of Engineering Research And Advanced Technology (IJERAT), Volume. 02 Issue.03, March-2016.
- [14] T. Veijonen et al., "ENVIMON Disaster Monitoring System," International Disaster Reduction Conference, 2006.
- [15] Rocha, Artur, Bojan Cestnik, and Marco A. Oliveira. Interoperable geographic information services to support crisis management. Springer Berlin Heidelberg, 2005.
- [16] Wallace, William A., and Frank De Balogh. "Decision support systems for disaster management," Public Administration Review, pp.134-146, 1985.

A New Approach of Detection and Segmentation of Blood Vessels for the Classification of Healthy and Diseased Retinal Images

Bithi Barua, Md. Mehedi Hasan

Department of Electronics and Communication Engineering

Khulna University of Engineering & Technology

Khulna 9203, Bangladesh

bithi1109021@gmail.com, mehedi_ece@yahoo.com

Abstract— Retinal image analysis is increasing popularity for the detection of eye disease like diabetic retinopathy, glaucoma, cardiovascular disease. In modern ophthalmology, the structure of retinal image with proper segmentation has achieved much interest for disease detection. This paper presents a method with some new features extraction methodology after applying segmentation procedure in retinal images so that the retinal disease can be identified properly. At first we have extracted the vascular tree from the retinal image and then segmented the image for feature generation. These features have been used for the training of the ANN classifier by Backpropagation algorithm. The method described in this paper is applied to the High Resolution Fundus (HRF) image dataset which is a new publicly available one. The detection of the diabetic retinopathy, glaucoma and healthy condition of the image are 97%, 99% and 96% respectively.

Keywords- Retinal image; Vascular tree; Segmentation; Feature extraction

I. INTRODUCTION

The subdivision process of an image into its constituent parts such as objects, region containing pixels of similar properties is known as segmentation. The labeling of a pixel or group of pixels on basis of the grey values is known as classification. Segmentation of retinal image structures has gained great interest in modern ophthalmology. This process is important for recognizing the presence and vitality of retinal diseases such as diabetic retinopathy, hypertension, glaucoma. Early detection and treatment of retinal eye disease is important so that vision loss can be prevented. Analysis of the retinal mages has the potential to identify the retinal disease more accurately.

One person among ten gets affected by Diabetes, and it has complications such as vision loss, heart failure and stroke. Diabetic Retinopathy is the disease having abnormalities in the retina and in the worst case blindness or severe vision loss[1]. It affects the retina and the choroid. In this condition, a network of small blood vessels, called choroidal neovascularization (CNV), arises in the choroid and taking a

portion of the blood supplying the retina. As the amount of blood supplying the retina is decreased, the sight may be degraded and in the severe cases, blindness may occur. [2]. Glaucoma is the second most frequent cause of permanent blindness. It happens because of the permanent damage of the optical nerve connected with degeneration of retinal ganglia cells, axons (neural fibers) and gliocells (providing nutrition for the axons). From the center of the optic nerve, major blood vessels are generated through the retina. So because of the damage in the optic nerve, the number of the blood vessel lessens. Again, the retinal blood vessels have topological and geometrical properties which provide valuable clinical information to diagnose diseases such as diabetic retinopathy, glaucoma. So the proper tracing of the vessel trees from fundus images is essential for the accurate clinical diagnosis with the help of blood vessel tree structure[3].

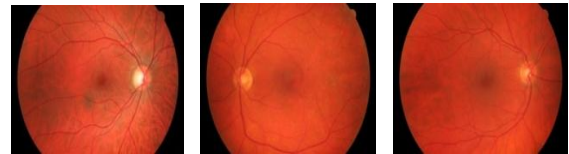


Fig. 1. An example of (a) healthy (b) diabetic retinopathy (c) glaucoma

Over the years, many methods have been applied for the classification of healthy and diseased retinal images. For this segmenting of the blood vessels has been approached by two categories: the pixel processing based approach and the tracking based approach.

Pixel processing based approach includes two operations. At first the detection of the retinal blood vessel by adaptive filtering is evaluated and then the morphological pre-processing techniques are used. Secondly the selection of a group of pixel for performing the operation is done to detect whether they are vessel pixels or background pixels. In this approach, every pixel is taken into account for multiple operations. There are various ways which have been shown

for pixel processing based approach. Chaudhuri [4] depicted the approach on the basis of two-dimensional matched filters. These filters are designed in such a way to get maximum response from vessel pixels having minimum response from the background pixels. Edge detection, matched filtering, and region growing are combined altogether [5] to acquire a fast solution for detecting the retinal blood vessels automatically. Niemeijer [6] presented a vessel segmentation algorithm which used a simple feature vector for pixel classification. Recently, Kawadiwale [7] used matched filtering with two different values of variance parameters. There are many other typical pixel processing approaches shown at Mendoca [8], Hoover [9], Staal [10], Zana[11]. Tracking based approach needs the detection of the vessel seed points so that vasculature tree can be easily traced with the help of some continuity property measurement of the blood vessel. This method uses exploratory algorithms. This algorithm is based on a single operation where the detection of the vascular tree and extraction of the various image features are simultaneously executed. There are two strategies in tracking based approach: semi automated tracing and automated tracing. Vessel seed points are manually selected by the user in semi automated tracing whether it is done by the algorithms automatically in fully automated tracing. The tracing method described in [12] is local gray-level minima which detects seed-tracking points automatically with a grid of one-pixel wide lines; after that 2-D correlation kernels is used to detect the vessel boundaries; Typical tracking based approaches are presented in Staal [10], Maritiner-perez [13] also.

Again Singh [14] proposed a new technique to detect retinal disease by broadening the retinal blood vessel applying the method morphology and noise removal algorithm having phase reversing method. Recently, Liskowski [15] developed a supervised segmentation technique that uses a deep neural network using the features global contrast normalization, zero-phase whitening, and augmented using geometric transformations and gamma corrections.

We have developed a technique that utilizes the segmentation process as well as some new features to identify the healthy and diseased retinal images with the help of ANN. The rest of this paper is organized as follows: Section II provides the motivation of this work. Section III discusses our proposed methodology. Section IV provides the visual results of our proposed method with some proper graph analysis. Section V concludes our work including the summary and future work.

II. MOTIVATION OF THE WORK

Before the detection of retinal image features and abnormalities, segmentation of retinal image must be done. The purpose of segmentation is to divide the whole image into so many small regions by removing the noisy and unwanted region.

Our method is based on the pixel processing based approach. We have segmented the blood vessel for the feature extraction. Multiple operations are applied to identify the features from

the healthy and diseased retinal image. In this work, we have proposed approaches for features extraction from the segmented retinal images so that these features can be applied for the pinpoint of the healthy and diseased retinal images. The four main steps for our novel framework:

- Vascular tree extraction from the retinal image.
- Segmentation of the image in the presence of retinal vessel tree.
- Feature generation from the segmented image.
- Construction of classifier which can classify the healthy and diseased retinal image.

Our work provides a new process for the identification of retinal disease with some new features. Classifier has been with the help of ANN classifier for discriminating the healthy and diseased retinal image. Our prototype has achieved average accuracy of 97%, 99% and 96% for identifying the diabetic retinopathy, glaucoma and healthy retinal images. This proposed scheme is applied on the HRF image dataset [16].

III. METHODOLOGY

The block diagram for our proposed methodology is illustrated in Fig.2. There are three stages which are preprocessing stage for feature extraction, feature generation stage and retinal image detection stage.

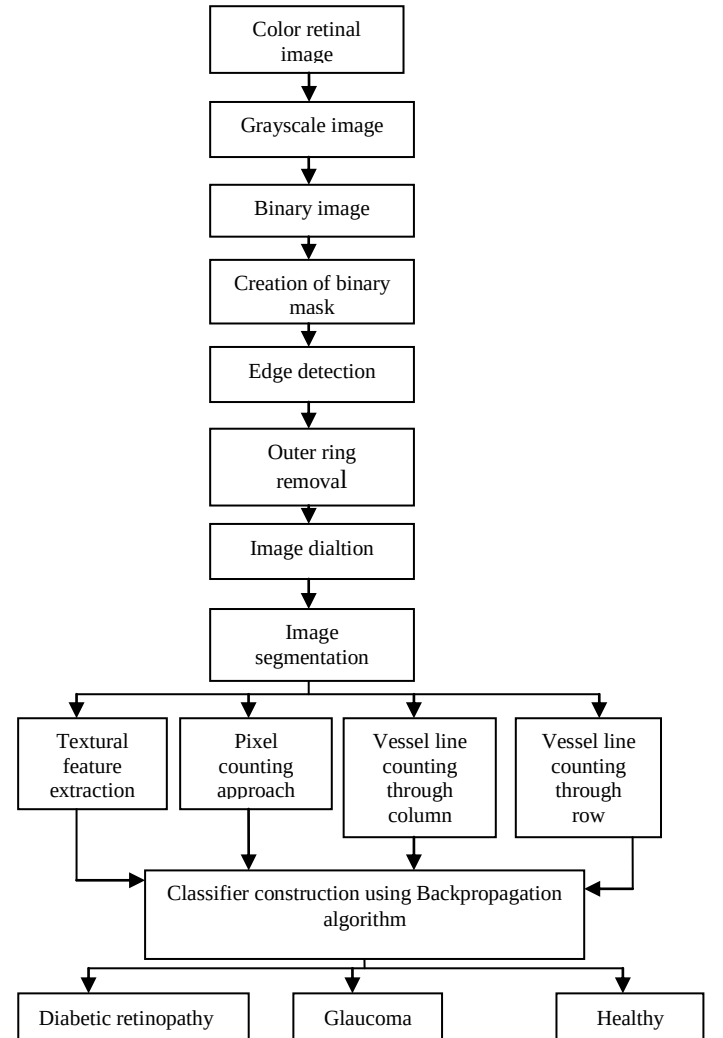


Fig. 2: Flow Diagram for Proposed Approach

Preprocessing stage: There are many subtasks involved in the preprocessing stage.

- The color retinal image which has been taken into account for the classification in the RGB version. This image has three channels: Red, Green and blue. Here the red, green, and blue intensities make a combination to form the color of each pixel.
- The conversion of the color retinal image to grayscale image has to be processed first. This image is known as 8-bit image as every pixel in the RGB image is expressed using same intensity. Grayscale image contains 256 possible shades of gray lying in between black and white.
- The grayscale image is then converted to binary image for the necessity of creating a binary mask which works as a filter.
- The edges of the blood vessels are determined by applying canny method.
- After removing the outer ring from the image, the whole image is divided into 500×500 blocks as the whole image is 3504×2336 pixels.

These are the subtasks which must be followed in the preprocessing before feature extraction.

Feature generation stage: There are four features which have been combined in our proposed methodology. Among them for generating three features, algorithms have been proposed. And the remaining one is the textural feature.

- 1) *Textural feature:* Textural features have been extracted using Haralick features [17] by Gray Level Co-occurrence Matrix (GLCM). GLCM provides information about the spatial relationship of pixels which is known as textures. It determines the number of the occurrence of a gray scale value i adjacent to a pixel value j . Four angles are used for observing the pixel adjacency i.e. $\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$. In Fig. 3(a) these directions are shown. There must be an offset value D which is used to define the pixel adjacency in a certain distance. The process of creating an GLCM using the image I is shown in Fig. 3(b). The textural features which we have extracted using GLCM matrix in our work are enlisted in the Table I.

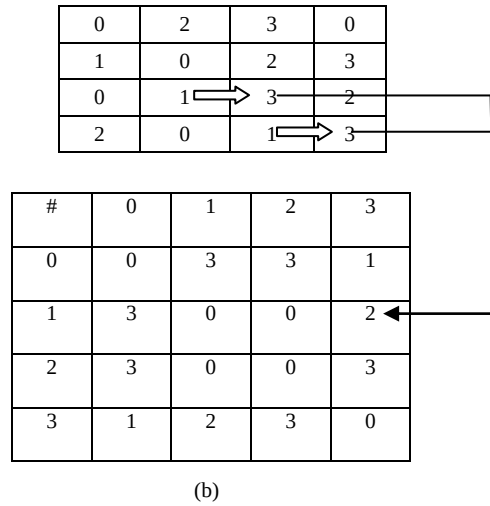
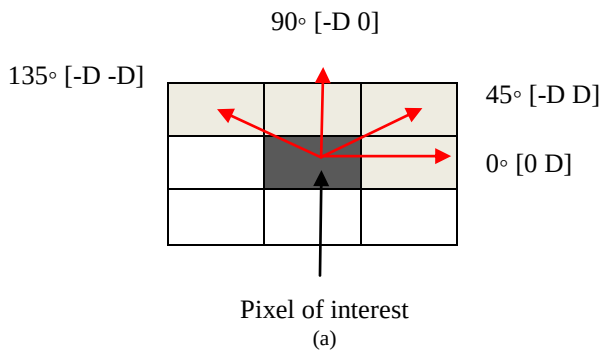


Fig. 3. (a) Input image I . (b) GLCM construction in 0° direction with a distance $d=1$.

TABLE I. Textural features extracted using GLCM

Property	Description	Formula
Contrast	Local variations to observe the texture fitness by measuring the intensity contrast between a pixel and its neighbor over the whole image.	$\sum_{i,j} i - j ^2 p(i,j)$
Correlation	Linear dependence in GLCM between different index by measuring of how correlated a pixel is to its neighbor over the whole image.	$\sum_{i,j} \frac{(i - \mu_j)(j - \mu_i)p(i,j)}{\sigma_i \sigma_j}$
Energy	Returns the sum of squared elements in the GLCM	$\sum_{i,j} p(i,j)^2$
Homogeneity	Returns a value that measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal.	$\sum_{i,j} \frac{p(i,j)}{1 + i - j }$

- 2) *Proposed algorithms for feature extraction:*

Algorithm 1: Pixel counting approach

- a. Input image is the dilated image of vascular tree of size 3504×2336 .
- b. Select a particular blocksize $n \times n$ and divide the whole image.
- c. Start the procedure of pixel counting from the 1st block.
- d. Count the number of black and white pixels for the block.
- e. Continue the procedure of pixel counting for every block in the image.

Algorithm 2: Vessel line counting through column in a block

- Input image is the dilated image of vascular tree 3504×2336
- Select a particular blocksize $n \times n$ and divide the whole image.
- Start the procedure from the 1st horizontal line of the block in 1st column regard as reference line.
- Check whether it is vessel line or not.
- If it is vessel line then count=1.
- If it is not, search for the vessel line from the reference line.
- When 1st vessel line is found, then start counting vessel lines from that line.
- This process continues until count=Total number of vessel lines per block.
- Repeat step (a) to (f) for the blocks through the 1st column in the image.
- Continue the procedure through every column in the image.

Algorithm 3: Vessel line counting through row in a block

- Input image is the dilated image of vascular tree 3504×2336.
- Select a particular blocksize $n \times n$ and divide the whole image.
- Start the procedure from the 1st horizontal line of the block in 1st row regard as reference line.
- Check whether it is vessel line or not.
- If it is vessel line then count=1
- If it is not, search for the vessel line from the reference line.
- When 1st vessel line is found, then start counting vessel lines from that line.
- This process continues until count=Total number of vessel lines per block.
- Repeat step (a) to (f) for the blocks through the 1st row in the image.
- Continue the procedure through every row in the image.

The proposed algorithms which we have developed can explain from the following Fig. 4. The retinal image containing the vascular tree which we have extracted is divided into small blocks shown in Fig.4. Applying Algorithm 1 the total number of black and white pixels is estimated for both the healthy and diseased retinal images. The number of vessel lines through the column and row is counted by Algorithm 2 and Algorithm 3 respectively. How these vessel lines are counted is shown in Fig.4.

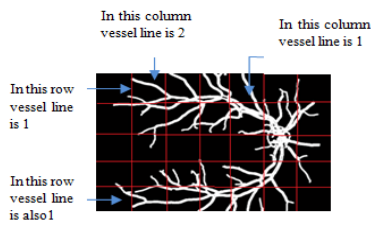


Fig. 4. Illustration of Algorithm 2 and Algorithm 3 with the division of the whole image.

Retinal image detection stage: This stage includes the process of classifier construction. We have applied Artificial Neural Networks (ANN). It presents the neural structure of the brain with a network based on “neurons”. Consider Fig. 5 which shows three basic layers of ANN i.e. input layer, hidden layer and output layer. Here the input layers take the features as input, hidden layers process the features and the output layer provides output.

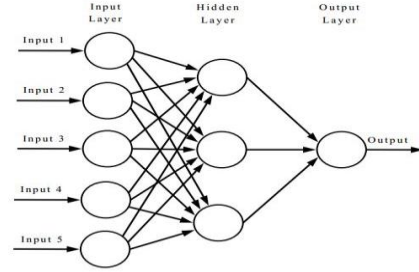


Fig 5: Artificial Neural Networks Diagram

Backpropagation is a common method for training a neural network. The training process uses Delta Rule, which calculates the difference between the actual outputs and the desired outputs. Using this error, connection weights in between the layers are increased in proportion to the error times.

The network which we have developed used 6 set of retinal images from the HRF image dataset for training and applied the extracted features to the input layer. Then the 1 set of images are used for testing to identify whether the retinal images are healthy or diseased.

IV. RESULTS

In order to identify the healthy and diseased retinal image accurately by segmentation method we combined our proposed algorithms with the textural features and evaluate the features applying on 7 set of image (i.e. 01_dr, 01_g, 01_h; 02_dr, 02_g, 02_h; 03_dr, 03_g, 03_h; 06_dr, 06_g, 06_h; 08_dr, 08_g, 08_h; 11_dr, 11_g, 11_h; 12_dr, 12_g, 12_h) in the HRF database.

The HRF database consists of 45 retinal images including three sets of fundus images: of healthy, glaucomatous, and diabetic retinopathy subjects (15 images in each group). The images were captured from a CANON CF-60 UVi camera having 60-degree FOV. The image size is 3504 × 2336 pixels. Every image contains 24-bits in each pixel and these images are in JPEG format. For each image, there is a binary mask to determining the FOV.

The subtasks which are described in the preprocessing stage have to be applied on the images before feature extraction. The results after applying subtasks on the image are shown for only 1 image group in Fig. 5. and Fig. 6.

Algorithm 1 measured the black and white pixels from every block at the image. After counting we have taken the average

pixels. From the average pixel analysis we have observed that the number of white pixels is larger in healthy images than the

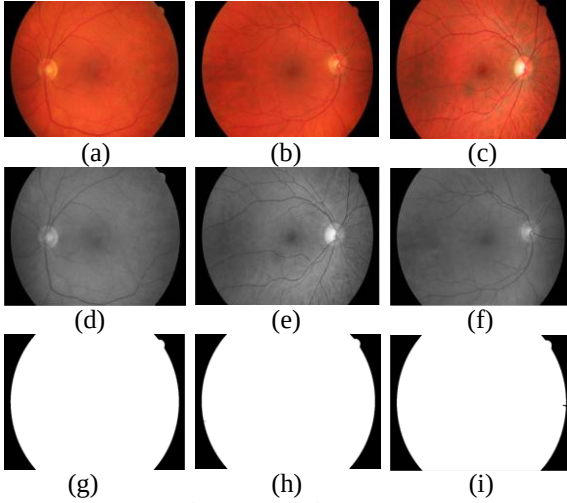


Fig. 5. An example of RGB image from HRF image dataset (a),(b),(c); Grayscale image (d),(e),(f); Binary mask (g),(h),(i) of diabetic retinopathy, glaucomatous and healthy conditions respectively.

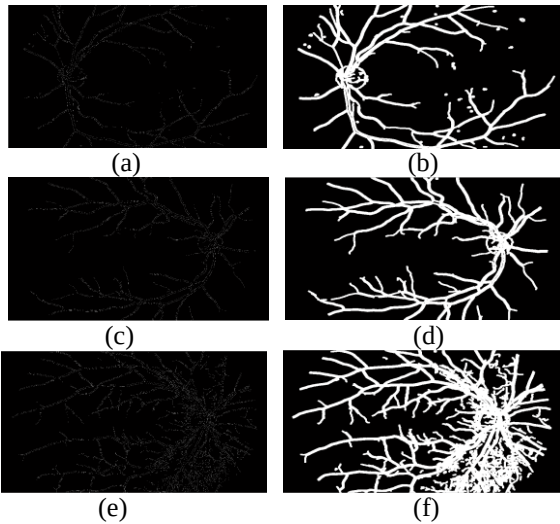


Fig. 6. Applying edge detection (a),(c),(e) and morphological processing (i.e dilation) (b),(d),(f) in diabetic retinopathy, glaucomatous and healthy conditions respectively.

diabetic retinopathy and glaucomatous. From the graphical observation, the variations in the average white pixels is shown for only 1 image group of HRF database. And in our case, this is true for the 7 image group on which we have applied our method.

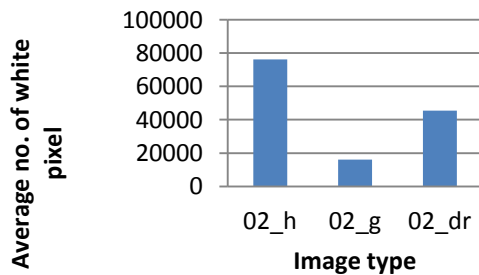


Fig. 7. Variations of average white pixel in healthy(02_h), glaucoma(02_g) and diabetic retinopathy(02_dr) retinal images in HRF image dataset.

the image. Each vessel line can be identified by the number of connected white pixels. The extracted features from Algorithm 1, 2 and 3 are used for training the ANN. The graph shown in Fig. 7 indicates that the mean square error during each of the iteration for training the ANN is gradually decreasing and is diminished completely after the iteration no. 500. That means ANN can accurately detect the healthy, diabetic retinopathy and glaucoma group retinal images after the iteration no. 500 of training.

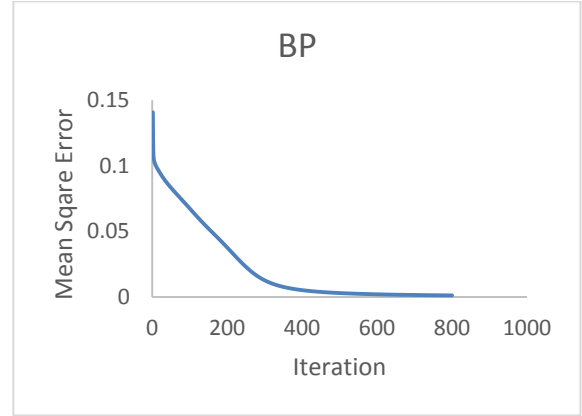


Fig. 7. Iteration vs Mean square error using Backpropagation algorithm for training ANN

The proposed method was primarily designed for high-resolution images and therefore, in this initial approach of ours, it is evaluated in the new HRF database. 7 set of retinal images from healthy, diabetic retinopathy and glaucoma group are considered and the mean value of accuracy for the detection the retinal diseases are listed in Table II.

TABLE II. Average Accuracy (ACC) measurement of detecting the diabetic retinopathy, glaucoma and healthy retina images with our proposed method using HRF image dataset.

Image type \ Image type	Diabetic retinopathy	Glaucoma	Healthy
Diabetic retinopathy	0.969	0.0574	0.0049
Glaucoma	0.019	0.9885	0.0098
Healthy	0.0315	0.0247	0.957

So it can be observed that the average accuracy of the detection of disease diabetic retinopathy, glaucoma and healthy are almost 97%, 99% and 96% respectively. So the proposed method using some new features provides a better result.

V. CONCLUSION AND FUTURE WORK

A new approach of feature generation for the retinal disease detection is proposed. For this, the initial steps include the taking of RGB image as input, converting the image to grayscale and then binary have been assessed. Then the combination of edge detection using canny operator and morphological processing has been applied. When the vascular tree is extracted, the features are generated by segmenting the image properly. These features are given as an input to the ANN which is trained by the Backpropagation algorithm and the detection of the diabetic retinopathy, glaucoma and healthy retinal image are accurately processed. Our proposed approach for feature generation is different from the existing methods and is applied on the HRF image dataset. This work can be extended in future by following procedures:

- Comparison of the proposed approach with the other approaches used for segmenting blood vessels for disease detection.
- In our approach, we have used only ANN classifier. So other classifiers such as KNN, SVM can be used for the evaluation of accuracy of disease detection.
- The proposed approach is applied only on the High Resolution Fundus (HRF) image dataset. It can be applied to the DRIVE, STARE image dataset to gain knowledge about the efficiency of this method.

References

- [1] E. J. Susman, W. J. Tsiaras, and K. A. Soper, "Diagnosis of diabetic eye disease", JAMA, vol. 247, no. 23, pp. 3231-3234, 1982.
- [2] Mai S. Mabrouk, Nahed H. Solouma and Yasser M. Kadah, "Survey of retinal Image Segmentation and Registration", GVIP Journal, vol. 6, no. 2, 2006.
- [3] Jaydeep De, Li Cheng, Xiaowei Zhang, Feng Lin, Huiqi Li, Kok Haur Ong, Weimiao Yu, Yuanhong Yu, and Sohail Ahmed, "A Graph-Theoretical Approach for Tracing Filamentary Structures in Neuronal and Retinal Images", IEEE Transactions on Medical Imaging, vol.35, no.1, pp. 257-272, 2016.
- [4] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum, "Detection of blood vessels in retinal images using two-dimensional matched filters", IEEE Transactions on Medical Imaging, vol. 8, no. 3, pp. 263-269, 1989.
- [5] Y. Wang and S. C. Lee, "A fast method for automated detection of blood vessels in retinal images," IEEE Computer Society Proc. Asilomar Conf., 1998, pp. 1700-1704.
- [6] M. Niemeijer, J. Staal, B. van Ginneken, M. Loog, and M. D. Abramoff, "Comparative study of retinal vessel segmentation methods on a new publicly available database," International Society for Optics and Photonics, pp. 648-656, 2004.
- [7] Ramish B. Kawadiwale, Vijay M. Mane, "Evaluation of algorithms for segmentation of retinal blood vessels", International Conference on Pervasive Computing (ICPC), 2015.
- [8] M. Mendonca and A. Campilho, "Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction," IEEE Transactions on Medical Imaging, vol. 25, no. 9, pp. 1200-1213, 2006.
- [9] A. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessel in retinal images by piecewise threshold probing of a matched filter response," IEEE Transactions on Medical Imaging, vol. 19, no. 3, pp.203-210, 2000.
- [10] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina," IEEE Transactions on Medical Imaging, vol. 2, no. 4, pp.501-509, 2004.
- [11] F. Zana and J.-C. Klein, "Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation," IEEE Transactions on Medical Imaging, vol. 11, no. 7, pp. 1111-1119, 2001.
- [12] Can, H. Shen, J. N. Turner, H. L. Tanenbaum, and B. Roysam, "Rapid automated tracing and feature extraction from retinal fundus images using direct exploratory algorithms," IEEE Transactions on Information Technology in Biomedicine, vol. 3, no. 2, pp. 125-138, 1999.
- [13] M. E. Martinez-Perez, A. D. Hughes, S. A. Thom, A. A. Bharath, and K. H. Parker, "Segmentation of blood vessels from red-free and fluorescein retinal images," Medical Image Analysis, vol. 11, no. 1, pp. 47-61, 2007.
- [14] Dalwinder Singh, Dharmveer, Birmohan Singh, "A new morphology based approach for Blood Vessel Segmentation in retinal images", Annual IEEE India Conference(INDICON), 2014.
- [15] Paweł Liskowski, "Segmenting retinal blood vessels with deep neural networks", IEEE Transactions on Medical Imaging, 2015.
- [16] <https://www5.cs.fau.de/research/data/fundus-images/> (Image dataset)
- [17] R. Haralick and L. Shapiro, Eds., Computer and Robot Vision. Addison Wesley Longman Publishing Co., Inc, 1991.

Analysis of Data Mining Techniques for Heart Disease Prediction

Marjia Sultana*, Afrin Haider and Mohammad Shorif Uddin

Department of Computer Science and Engineering

Jahangirnagar University

Dhaka, Bangladesh

*Email: mstoma11@gmail.com

Abstract— Heart disease is considered as one of the major causes of death throughout the world. It cannot be easily predicted by the medial practitioners as it is a difficult task which demands expertise and higher knowledge for prediction. This paper addresses the issue of prediction of heart disease according to input attributes on the basis of data mining techniques. We have investigated the heart disease prediction using KStar, J48, SMO, Bayes Net and Multilayer Perceptron through Weka software. The performance of these data mining techniques is measured by combining the results of predictive accuracy, ROC curve and AUC value using a standard data set as well as a collected data set. Based on performance factor SMO and Bayes Net techniques show optimum performances than that of the performances of KStar, Multilayer Perceptron and J48 techniques.

Keywords—Data mining; KStar; J48; SMO; Bayes Net; Multilayer Perceptron

I. INTRODUCTION

Data mining is defined as a way of extracting necessary and hidden predictive data from huge databases [1]. Data mining classifiers find wide applications in medical domain for diverse diagnosis. Different data mining techniques are used for prediction and decision making for different kinds of diseases like cancer, heart disease, diabetes etc. [2]. There are many types of data mining classifiers which can be applied to predict diseases. In this paper we are focusing on heart disease prediction.

The term heart disease includes several types of disorders which may damage the heart. Most common types of heart diseases are congenital heart disease, heart failure, hypertensive heart disease, cardiomyopathy, heart murmurs, rheumatic heart disease, pulmonary stenosis and coronary artery disease. These diseases are the main reasons of death for huge number of people all over the globe. WHO (World Health Organization) and CDC (Centers for Disease Control and Prevention) reported that heart disease is the major reason of death for the people of UK, Canada and Australia [3]. Medical journal The Lancet [4] reported that in Bangladesh about 1.78 lakh people died of strokes, 1.06 lakh died of ischemic heart disease and 28,000 died of hypertensive heart disease in 2013. Abdur Razzak [5] reported that 5.2% people of Kushtia and Jhenaidah districts in Bangladesh are suffering from heart disease. There are several causes of heart disease like smoking, age, gender, high blood pressure, high blood cholesterol, unbalanced diet, lack of exercise, over stress, poor

hygiene and family history of heart disease etc. We can use these causes as risk factors to predict heart disease.

Data mining tools predict future trends by allowing knowledge-driven decisions. The prediction of heart disease requires a huge size of data which is too complex and massive to process and analyze by conventional techniques. Diverse data mining techniques are being used by experts. Our objective is to find out the suitable data mining technique that is computationally efficient as well as accurate for the prediction of heart disease.

The rest of the paper is organized as follows. Literature review is explained in Section II. Section III includes methodology. Data mining techniques are described in Section IV. The results of our experiment are explained in Section V. Finally Section VI draws the conclusion of our findings.

II. LITERATURE REVIEW

Researchers use several data mining methods to address the present issue. Papers [2], [6], [7] studied the applications of different data mining techniques for prediction of various diseases. References [8], [9], [10], [11] specifically discussed heart disease prediction using data mining techniques.

Rupali and Patil [12] described an improved decision support system using two data mining classifiers namely Naive Bayes and Jelinek-Mercer smoothing for heart disease prediction. Chitra and Seenivasagam [9] proposed a hybrid intelligent algorithm for improvement of classifiers' accuracy to predict heart disease. Shamsher and Shukla [13] analyzed some classification techniques for prediction of heart disease with decreased number of attributes. Jyoti and Dipesh [14] compared the predictive performance of data mining techniques. Khaled and Das [2] evaluated the different data mining techniques to find frequent pattern based on cost, performance, speed and accuracy. Hlaudi and Mosima [11] applied J48, Naive Bayes, Bayes Net, REPTREE and CART for predicting heart attacks. In that paper, classification techniques J48, REPTREE and SIMPLE CART showed the best accuracy. Taneja [10] found that J48 that used selected attributes outperformed the Naive Bayes and Neural Network by achieving the highest accuracy.

But none of above researchers compared the prominent approaches, such as KStar, J48, SMO, Bayes Net and

Multilayer Perceptron (MLP) approaches to find an appropriate technique based on performance factors for heart disease prediction. Therefore, the current paper works on it to fill this gap.

III. METHODOLOGY

A. Experimental Setup

To conduct our experiment, we have selected five classifiers namely KStar, J48, SMO, Bayes Net and MLP. And then we have conducted experimentation on both our collected- and standard-data sets. The experiment is performed based on test mode of k-fold cross-validation with supplied test sets (where k=10). For the optimality analysis of classifiers, we focus on the performance factors: predictive accuracy and AUC value. Procedural architecture of our experiment is shown in Fig. 1. The experiment is conducted by the following process:

- **Preparing working data sets:** In our work, standard data set of heart disease is attained from UCI Machine Learning Repository [15]. It contains total 270 records where some records are for non-heart disease patients and some are for patients with heart disease. It comprises of total 13 attributes: age, sex, chest pain, resting blood pressure, serum cholesterol, fasting blood sugar, resting electrocardiographic results, maximal heart rate, exercise-induced angina, old peak ST depression induced by exercise relative to rest, the slope of the peak exercise ST segment, number of major vessels colored by fluoroscopy and thal. We have also collected patients' medical data from Enam Medical Diagnosis Centre, Savar, Dhaka, Bangladesh and prepare it in csv format. It contains total 100 records with 6 attributes: age, sex, chest pain, maximum blood pressure, diabetics and maximal heart rate. Here, some records are for non-heart disease patients and some are for patients with heart disease.
- **Experiment through Weka software:** The data values are entered to the Weka data mining software as training and then testing. Then we choose our proposed classifiers one by one and note down classifier output for each.
- **Comparative study of classifiers:** Based on performance factors, we have compared the performance of all classifiers.

IV. DATA MINING TECHNIQUES

We have used the following data mining techniques:

i. KStar

KStar [16], a lazy learner, is an instance-based classifier. KStar simply stores a training tuple and then

performs a little processing; finally waits for a test tuple. Then the test set is classified according to the similarity of the stored training set. KStar uses a function named entropy based distance which differs from other instance based learners [17]. KStar is similar to K-Nearest-Neighbor classifier.

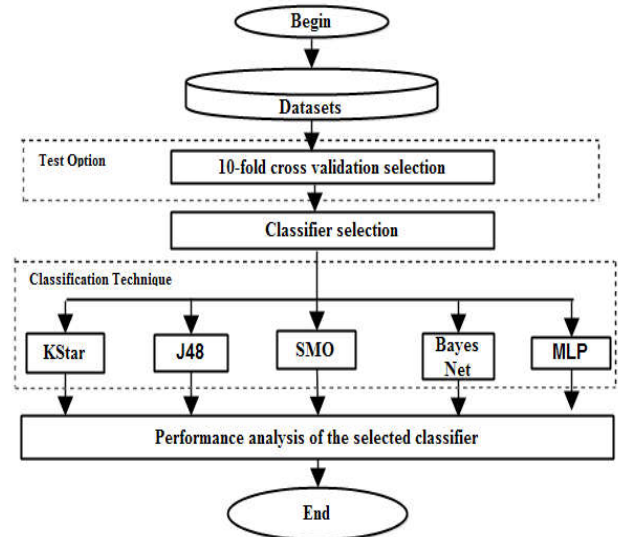


Fig. 1. Data mining classification process using Weka.

ii. Decision Tree

Decision tree [11] is suitable for experimental knowledge discovery because the tree structure depends on the data to perform prediction. In decision tree model, leaves indicate the outcome values and branches represent the attribute values of the data set. In data mining, decision tree does not make any decision; rather it simply visualizes the data for decision making [18]. Most commonly used decision tree algorithms are ID3, CART, C4.5, CHAID and J48 etc. In this paper, we apply J48 algorithm which is the extension of ID3 (Iterative Dichotomiser 3).

iii. SMO

SMO (Sequential Minimal Optimization) partitions the training set into smaller sets for solving Support Vector Machine (SVM) problems analytically. SMO is slow to solve the problem when the instances are not linearly separable. It is appropriate for linearly separable problem [19]. SMO can work with a large training set as it does not require any extra storage [20].

iv. Bayes Net

Bayesian Network or simply Bayes Net algorithm [21] is very powerful for disease prediction system because it can make decisions in case of missing input data. This classifier uses a directed graph model to represent relationships among

various attributes. It uses probability function to predict the presence or absence of different diseases.

v. Multilayer Perceptron

Multilayer Perceptron (MLP) [9], a classification technique, uses back propagation to classify instances. It is also known as feed-forward ANN (Artificial Neural Network). In MLP, the input data records are automatically separated into three data sets for the purpose of training, validation and testing. It is a good method to predict disease but this method requires more time to select the input characteristics for the construction of the network [9].

Machine learning software, WEKA (Waikato Environment for Knowledge Analysis) [22] written in Java is used for the purpose of data mining. We have used the WEKA software version 3.6.10 for the performance analysis of data mining methods.

V. RESULTS

Confusion matrix - a performance indicator gives the information about the number of correctly and incorrectly classified instances for each classifier. In TABLE I, 270 records of UCI standard data and 100 records of collected data are used to find confusion matrix of each technique. For two possible outcomes (one is Positive and another is Negative), each classifier generates a 2×2 confusion matrix.

There are four possible values of a confusion matrix for the two outcome values (Positive and Negative).

- True Positive (TP): TP is the total number of positive predictive instances which are classified correctly and truly positive.
- True Negative (TN): TN is the total number of negative predictive instances which are classified correctly and truly negative.
- False Positive (FP): FP is the total number of positive predictive instances which are classified incorrectly and truly negative.
- False Negative (FN): FN is the total number of negative predictive instances which are classified incorrectly and truly positive.

To analyze the confusion matrix we consider the following performance parameters:

$$\text{Accuracy (ACC)} = \frac{\Sigma \text{ True positive} + \Sigma \text{ True negative}}{\Sigma \text{ Total population}}$$

$$\text{True Positive Rate (TPR)} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{False Positive Rate (FPR)} = \frac{\text{FP}}{\text{FP} + \text{TN}}$$

TABLE II shows the accuracy in percentage and model building time in second for the investigated five classifiers (KStar, J48, SMO, Bayes Net and Multilayer Perceptron).

Fig. 2 shows the bar-chart for the overall accuracy of these methods.

TABLE I. CONFUSION MATRIX OF CLASSIFIERS

Classification	Confusion matrix for collected data		Confusion matrix for standard data	
	TP (53)	FN (16)	TP (82)	FN (38)
KStar	FP (9)	TN (22)	FP (29)	TN (121)
	TP (64)	FN (5)	TP (88)	FN (32)
J48	FP (9)	TN (22)	FP (31)	TN (119)
	TP (67)	FN (2)	TP (96)	FN (24)
SMO	FP (9)	TN (22)	FP (19)	TN (131)
	TP (67)	FN (2)	TP (94)	FN (26)
Bayes Net	FP (11)	TN (20)	FP (25)	TN (125)
	TP (63)	FN (6)	TP (92)	FN (28)
MLP	FP (8)	TN (23)	FP (33)	TN (117)

TABLE II. PERFORMANCE MEASURE OF CLASSIFIERS

Classification	Collected data		Standard data	
	Accuracy	Model building time(s)	Accuracy	Model building time(s)
KStar	75%	0	75.1852%	0
J48	86%	0	76.6667%	0
SMO	89%	0	84.0741%	0.02
Bayes Net	87%	0	81.1111%	0.02
MLP	86%	0.33	77.4074%	0.75

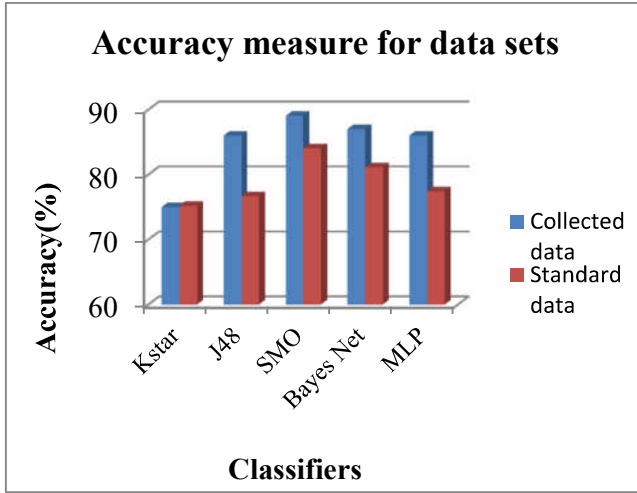


Fig. 2. Comparative study of accuracy measure for our collected and standard data sets.

From Table II and Fig. 2, for both data sets on the basis of overall accuracy SMO and Bayes Net perform better than the other three algorithms: KStar, MLP and J48.

Accuracy does not always give a true measure of the performance of a classifier, as FPR is a very dangerous error. For this reason the classifier performance is usually expressed using Receiver Operating Characteristics (ROC) curve (that is a plot of TPR vs. FPR). For an optimum classifier FPR = 0 in perfect case, which indicates the area under ROC curve (AUC) is 1 and for sub-optimal situations it is less than 1. Figs. 3 and 4 show the ROC curves for the data mining techniques using collected and standard datasets, respectively. The higher curve represents the higher performance. To evaluate the quantitative performance, we have calculated the area under ROC curve for each classifier. Table III shows the area under ROC curve (AUC) of the respective classifier.

TABLE III. AREA UNDER THE ROC CURVE FOR EACH CLASSIFIER

Classifier	AUC for Collected Data	AUC for Standard Data
KStar	0.8004	0.8367
J48	0.777	0.7443
SMO	0.8403	0.8392
Bayes Net	0.8808	0.902
MLP	0.8752	0.8527

From Table III, Figs. 3 and 4, we find that performance of Bayes Net classifier is the best and that of J48 classifier is the worst. The performances of SMO, MLP and KStar are also good. If you combine both accuracy and ROC performances, then it is better to make a decision on the basis of intersection operation of the classifiers. Based on this, we can conclude that Bayes Net and SMO are the optimum classifiers among the investigated five classifiers.

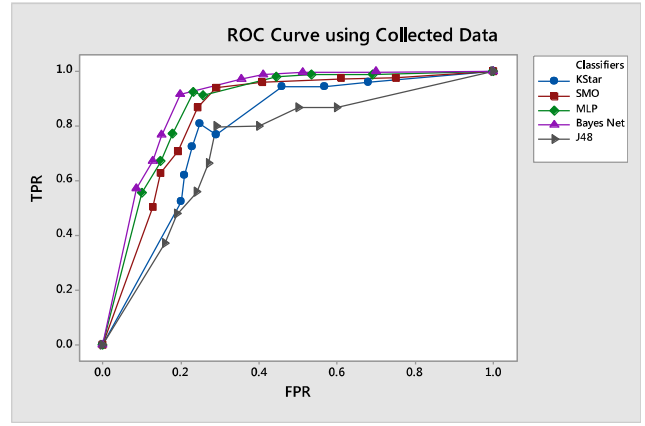


Fig. 3. ROC curves for data mining techniques using collected data set.

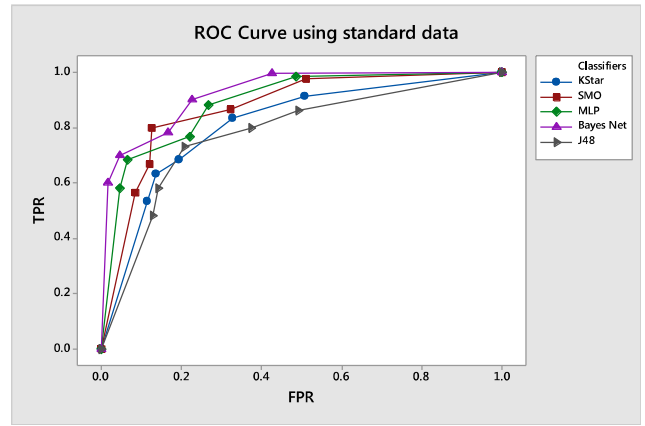


Fig. 4. ROC curves for data mining techniques using standard data set.

VI. CONCLUSION

This paper addresses the issue of prediction of heart disease according to some input attributes. The heart disease becomes a plague throughout the world. It cannot be easily predicted as it is a difficult task that demands expertise and higher knowledge for prediction. Data mining extracts hidden information that plays an important role in making decision. This paper performed an experiment using diverse data mining techniques to find out a more accurate technique for the heart disease prediction. In this paper, two data sets (collected and UCI standard) are used separately for each data mining technique.

Our findings show that for heart disease prediction the performances of Bayes Net and SMO classifiers are the optimum among the investigated five classifiers: Bayes Net, SMO KStar, MLP and J48.

REFERENCES

- [1] Micheline Kamber and Jian Pei Jiawei Han, *Data Mining Concepts and Techniques*, 3rd ed., 2012.
- [2] Mohammed Abdul Khaled, Sateesh Kumar Pradhan and G.N. Dash, "A survey of data mining techniques on medical data for finding locally frequent diseases," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3, pp. 149-153, August 2013.
- [3] "Heart Disease: Definition, Causes, Research," Available online: <http://www.medicalnewstoday.com/articles/237191.php> [Accessed: 29-Jan-2016].
- [4] "Heart diseases now a plague in Bangladesh," Available online: <http://www.thedailystar.net/heart-diseases-now-a-plague-in-bangladesh-57098> [Accessed: 26-Jan-2016].
- [5] Abdur Razzak, "A questionnaire survey on infectious disease among hospital patients in Kushtia and Jhenaidah, Bangladesh," *International Journal of Genetics and Molecular Biology*, vol. 3, no. 9, pp. 120-xxx, October 2011.
- [6] D. P. Shukla, Shamsher Bahadur Patel and Ashish Kumar Sen, "A literature review in health informatics using data mining techniques," *International Journal of Software & Hardware Research in Engineering*, vol. 2, no. 2, February 2014.
- [7] V. Gayathri, M.Chanda Mona and S.Banu Chitra, "A survey of data mining techniques on medical diagnosis and research," *International Journal of Data Engineering (IJDE) Singapore Journal of Scientific Research (SJSR)*, vol. 6, pp. 301-310, 2014.
- [8] M.Akhil Jabbar, Priti Chandra and B.L. Deekshatulu, "Heart disease prediction system using associative classification and genetic algorithm," in *International Conference on Emerging Trends in Electrical, Electronics and Communication Technologies-ICECIT*, 2012.
- [9] R. Chitra and V.Seenivasagam, "Review of heart disease prediction system using data mining and hybrid intelligent," *ICTACT Journal on Soft Computing*, vol. 03, no. 04, July 2013.
- [10] Abhishek Taneja, "Heart disease prediction system using data mining techniques," *Oriental Journal of Computer Science & Technology*, vol. 6, pp. 457-466, December 2013.
- [11] Hlaudi Daniel Masethe and Mosima Anna Masethe, "Prediction of heart disease using classification algorithms," in *Proceeding of the World Congress on Engineering and Computer Science*, vol. II, San Francisco, USA, 2014.
- [12] Rupali, R.Patil, "Heart disease prediction system using Naive Bayes and Jelinek-mercer smothing," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 3, no. 5, May 2014.
- [13] Shamsher Bahadur Patel, Pramod Kumar Yadav and Dr. D. P. Shukla, "Predict the diagnosis of heart disease patients using classification mining Techniques," *IOSR Journal of Agriculture and Veterinary Science (IOSR-JAVS)*, vol. 4, no. 2, pp. 61-64, Jul.-Aug. 2013.
- [14] Jyoti Soni, Ujma Ansari and Dipesh Sharma, "Prediction data mining for medical diagnosis: An overview of heart disease prediction," *International Journal of Computer Applications (0975-8887)*, vol. 17, March 2011.
- [15] "Machine Learning Repository," Available online: repository.seasr.org/Datasets/UCI/csv
- [16] John G. Cleary and Leonard E. Trigg, "K*: An Instance-based learner using an entropic distance measure," *Proc. 12th International Conference on Machine Learning*, pp. 108-114, 1995.
- [17] S. Vijayarani and M. Muthulkshmi, "Comparative analysis of Bayes and Lazy classification algorithms," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 2, no. 8, August 2013.
- [18] R. Vijaya Kumar Reddy, K. Prudvi Raju, M. Jogendra Kumar, CH. Sujatha, P. Ravi Prakash, "Prediction of heart disease using decision tree approach," *International Journal of Advanced Research in Computer Science and Engineering*, vol. 6, no. 3, March 2016.
- [19] Promad Kumar Yadav, K. L. Jaiswal, Shamsher Bahadur Patel, D. P. Shukla, "Intelligent heart disease prediction model using classification algorithms," *IJCSMC*, vol. 3, no. 08, pp. 102-107, August 2013.
- [20] Gaurav Taneja and Ashwini Sethi, "Comparison of classifiers in data mining," *International Journal of Computer Science and Mobile Computing*, vol. 3, pp. 102-115, November 2014.
- [21] Sheweta Kharya, "Using data minig techniques for diagnosis of cancer disease," *IJCSEIT*, vol. 2, no. 2, April 2012.
- [22] M. Hall, E. Frank, G.Holmes, B. Pfahringer, P. Reutemann and I. H. Witten, "The weka data mining software: An update," *SIGKDD explorations*, vol. 11, 2009.

Determining 2D Shape of Object Using Ultrasonic Sensor

Rechie Ranaisa Daru¹, Hridoy Biswas², Shuvrodeb Barman⁴,
Department of Aeronautical Engineering
Military Institute of Science and Technology (MIST)
Dhaka, Bangladesh
Email: shuvrodebbarman@gmail.com

Al-Quabid Ahmed³
Department of
Electrical, Electronic and Communication Engineering
Ahsanullah University of Science and Technology
Dhaka, Bangladesh
Email: alquabid@gmail.com

Abstract—Ultrasonic technology has been using for years. Different Ultrasonic sensors are used in homes, shops, to navigate, object detection and for what is known as map building. The aim of this work is to build an ultrasonic device which is basically one kind of a radar system to get exact distance and approximate 2D shape for fixed objects placed around the device based on trigonometric rules. Its uses distance of the Ultrasonic sensor with the trigonometric rules to get the shape of the object. Some sensors like laser or infrared are used to measure the distance of obstacles, and consequently enable a robot to take a decision depending on the obstacles, or lack of obstacles, around it. But they are sensitive to strong light sources containing a high level of infrared radiation, and also to the color of obstacles, cannot detect objects that reflect lasers (windows, chrome-plated objects, etc.). So we are building such kind of device that can detect the object that is neutral to high radiation and color of the object. Microcontroller, ultrasonic sensors and MATLAB software is used to achieve the aim. The result was a device that can show the approximate 2D shape in a dotted graph.

Keywords—*ultrasonic; sensor; 2d shape; trigonometry; microprocessor;*

I. INTRODUCTION

Ultra sound is a popular, cheap and reliable technique of sensing distance of objects for localization. It is fast, accurate and do not create disturbances in human activities. It is a technique that uses sound propagation to navigate, communicate or detect objects. This technique is known as sonar. For many years using sonar we are detecting objects from thousands of miles away below and above the surface of water. It is popular because of its easy principle of working. In this paper now we are trying to detect the shape of an object using the sonar. The reason and main benefit of this project is, we will able to get idea about an object shape from distance.

Our device use sonar principle to get the desired result. To create a 2D shape we need two planes to move the device body. We have used two servo motors and two sonars to create two planes of movement. We've got continuous data from the sonar and early data we had got was not sufficient to make the

2D shape. We've used Arduino and MATLAB to get the required data. While doing serial communication between Arduino and MATLAB extra power supply was required. After getting continuous real data from the sensor we transform the point value into axis parameters using trigonometric rules. The shape we get there may not be the actual shape as told earlier as there interrupt object.

II. ULTRASONIC SENSOR PRINCIPLES

Ultrasonic sensors are devices that use electro-mechanical energy transformation to measure distance from the sensor to the target object. Ultrasonic waves are longitudinal mechanical waves which travel as a sequence of compressions and rarefactions along the direction of wave propagation through the medium. It consists of 2 parts. One creates the beam waves and transmits it and other part receives the reflected wave. The object range r is computed from

Range $r =$

$$\frac{\text{sound speed} * \text{time of travel by the beam}}{2} \quad (1)$$

A. Creation of Map Area

Even we can create a map of an area using this technique. Sonars in can have three different purposes [6]:

- 1. Obstacle avoidance:** The first detected echo measures the range to the closest object. Robots use this information to plan paths around obstacles and to prevent collisions [6].
- 2. Sonar mapping:** A collection of echoes are used to construct a map of the environment. Similar to a radar display, a range dot is placed at the detected range along the probing pulse direction [8].
- 3. Object recognition:** A sequence of echoes or sonar maps is processed to classify physical objects. These are mainly combination of different surfaces.

The objective of this research is to present a sensor design that helps to estimate the shape of the object using the sonar in

air. Here the two output of the sonar will be combined. Sonar mapping and the Object recognition processes was followed to achieve the desired output. The performance is limited by limitations of sonar.

There various research and project is going on ultrasonic sensor. Obstacle avoidance system is one of the most popular and used system in robotics now-a-days [6]. Research on Ultrasonic rage finding for distance measurement in coal mining, development of radar using ultrasonic sensor, process monitoring and chemical analysis by ultrasonic sensor been done before[6]. In practical there are many use of multi beam sonar to get the accurate value and produce the 3-D shape of an unknown landscape under water.

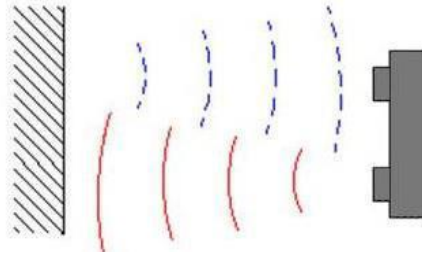


Fig. 1: Ultrasonic sensor reflecting and receiving beam

The factor of 2 converts the round-trip travel of beam distance to a range measurement. These are the main principal of sonar. Now we are using these principal to calculate the distance between the sonar and the object. Ultrasonic sensor can give three type of output using the same principle.

Ultrasonic sensor has programmed in a way that it will start to a timer from the moment it transmit the signal and stop the timer when it gets the reflected signal. Now as the sensor knows the velocity of the sound wave it divides the velocity with the half of the time it gets from the timer. And calculate the distance between the object and itself. And this is how we get the distance from the sonar.

For calculation we assume here the speed of sound is 340 m/s. The speed of sound varies significantly with atmospheric temperature, pressure and humidity and can affect the accuracy of a sonar system. The speed of sound in dry air at sea level air density and one atmosphere pressure is given by [6]

$$Ct = 20.05\sqrt{T_c} \times 273.16\text{m/s} \quad (2)$$

Here T_c is in degree Celsius. Under most conditions this is accurate to within 1%.

III. PREPARE YOUR PAPER BEFORE STYLING

Two ultrasonic sensors need to be used in order to show the two dimensional shape of the object. One ultrasonic sensor will move in XZ plane (Horizontal movement) and another ultrasonic sensor will move in XY plane (Vertical movement). While moving in XZ plane the ultrasonic sensor will provide the shape of the object about X axis. It will deliver the location

$(x_1, x_2, x_3 \dots x_n)$ along X-axis of the object. The ultrasonic sensor will give the distance of OB. As transmitter and receiver are considered to be in the vicinity of each other so the distance of transmitted beam and receiving beam is considered identical.

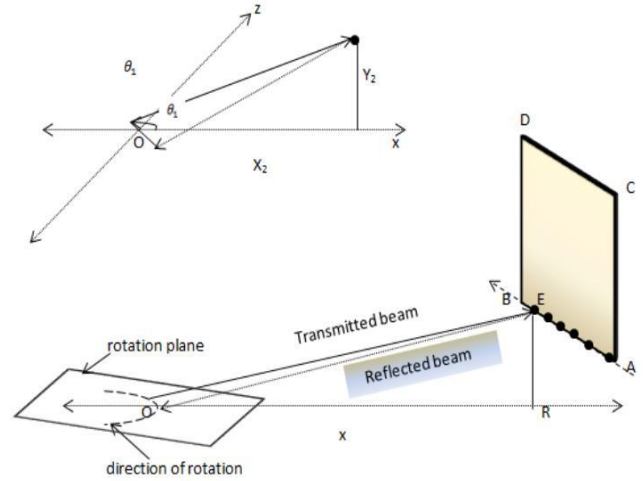


Fig. 2: Ultrasonic sensor1 moving in XZ plane from A to point B providing points along x axis of the object

From,

$$\cos \theta_1 = \frac{OB}{OR} \quad (3)$$

We can determine the position of point E of the object. Similarly every point on the object can be determined horizontally. After crossing the edge of the object the value of OB will be zero so $= 0$. While moving this ultrasonic sensor along XZ plane step by step for a fix value of x another ultrasonic sensor will keep moving in XY plane delivering $y_1, y_2, y_3, \dots, y_n$ along Y axis o of the object.

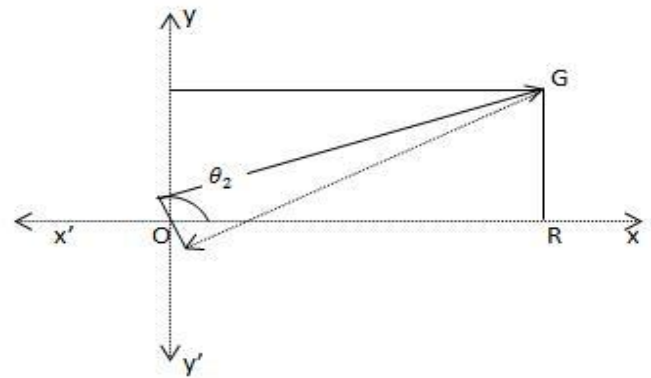


Fig. 3: Ultrasonic sensor2 moving in XY plane from E to point F providing points along y axis for a fix x point

$$\sin \theta_2 = \frac{RG}{OG} \quad (4)$$

We can determine the position of point G for a fix value E of the object. Similarly every point of the object along E to F can be determined vertically for a fix value of E. From the value of x and y of the object we can determine the shape of the object by plotting it in a graph.

IV. SYSTEM OVERVIEW

The designed RADAR system consists of the following parts.

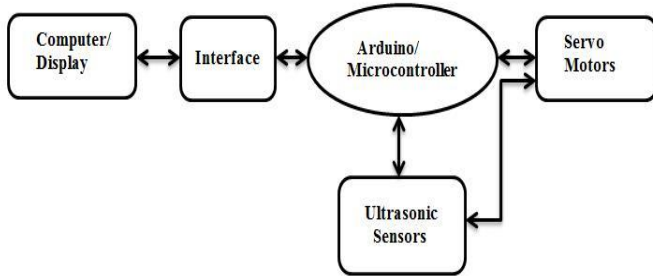


Fig. 4: Block Diagram of the System

A. Microcontroller Arduino

It is a highly integrated chip that contains all the components comprising a controller. It includes a CPU, RAM, some form of ROM, I/O ports, and timers [2]. The Arduino was introduced aiming to provide a low cost, easy way for novices and professionals to create devices that interact with their environment using sensors and actuators[3]. Arduino also supplies here the present angle and the distances that were coming from the sensors.

B. Servo Motors

Servomotor is a servomechanism. It is a closed-loop servomechanism that uses position feedback to control its motion and final position [4]. Here we use two servos.

One servo will move the sensor1 around the XZ plane. It will move 10 degrees with each loop after the sensor2 completes its cycle. This Servo needs a heavy base to stand on with the whole device. After rotating right to left the servo will follow the same path in reverse by 10 degrees of decrement.

Another servo will move the servo in XY plane. It will move stepping 2 degrees in each cycle.

Both were program to rotate from 45-165 degree. From this the angle of receiving beam can be determined. Servo with larger torque is required to build this device as the horizontal servo will be rotating the whole device. The servo we used is SpringRC SM-S4315M that has torque 4.8V: 213.4 oz-in (15.37 kg-cm).

C. Ultrasonic Sensor

Ultrasonic sensor creates sound waves and sends the waves in a direction and then it receives the reflections of the waves. By receiving the reflections it can tell about the distance of the object from where the waves are reflected. In this project HC - SR04 is used. It triggers beam of 40Hz. Tested highest distance was 500cm. Triggers Input Signal 10uS TTL pulse. The Module automatically sends eight 40 kHz and detect whether there is a pulse signal back. We are using two sensors in this purpose one vertically and one horizontally.

D. Interface

In the interface used the data we get from the servo and ultrasonic sensor to create the shape of the object. From the servos we get the angles (θ_1 and θ_2) and from the sensors we get the distances (OB and OG). Then, we just put the values to the equation (1) and equation (2) to get the points. Then we plot the points to get the 2D shape.

E. Computer Screen

We use the computer screen to see the result that we get from the interface. In computer screen we see the 2D graph of the object that we got from the interface.

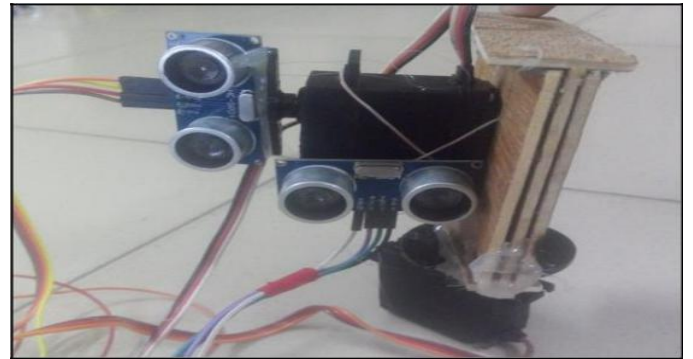


Fig. 5: Overall system

V. ALGORITHM

Total four loops are used here. Two loops for controlling the servos and two loops for controlling the ultrasonic sensors. Loops were designed such that if one sensor or servo is not working then the total system will not work or not start. There is a possibility of creating false alarm or no detection if the transmitting and receiving beams overlap with each other. So enough delay had given to transducer to trigger in each cycle. For sensor2 full rotation was count starting from 45 degree, reached to 165 degree and again returned to 165 degree. From the Arduino the values of distances and angle were sent by creating an array of the system.

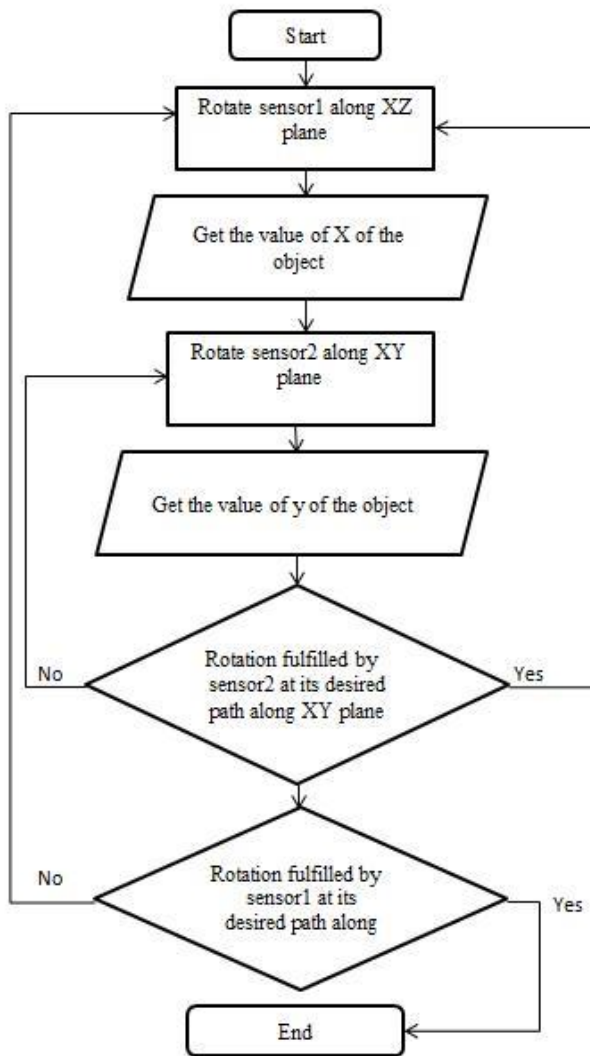


Fig. 6: Overall system Flow chart

VI. SYSTEM TEST AND RESULT

The system was connected as shown in figure-4. The output was shown in two manners. one is shown in serial monitor of the Arduino screen and other is shown in the MATLAB by plotting the graph. The output of the object can be shown better way using the software processing.

A. Testing the System

The system was checked by passing only the distance with the serial port to the MATLAB to run it. The output graph shows the correct result and a successful interference between two software. The sonar was tested by distance measurement. While device running there is always the serial monitor opened to check the system working

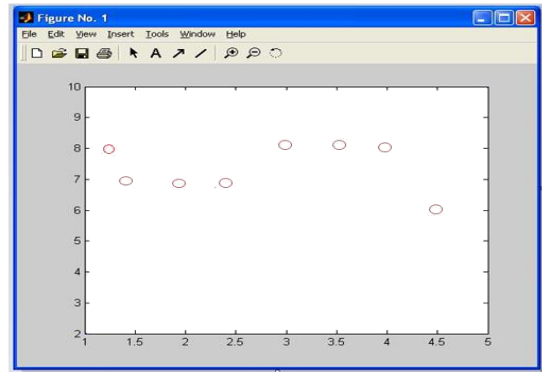


Fig. 7. MATLAB Graph distance Vs Time

B. Arduino output

After connecting with the Arduino with Computer the power was supplied and the serial monitor was showing the x and y values it getting from the system output. This serial monitor was kept on the programming to check that every component connected with the microcontroller is working

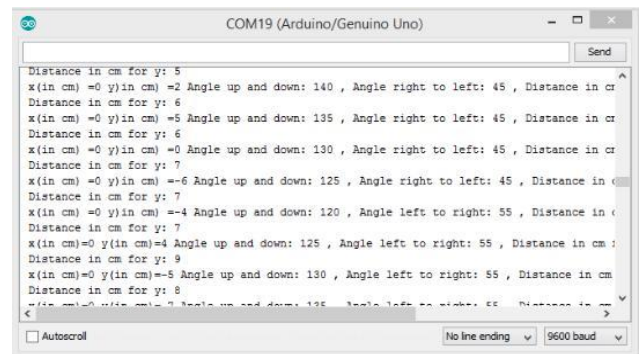


Fig. 8. Arduino Serial Monitor output

C. MATLAB output and testing

After getting the output from the MATLAB the size of the object was testified. The output was not so smooth and the exact value of the object because we only take integers value to avoid criticality of calculation during process. The shape was rectangular and was $12.5 \times 11 \text{ cm}^2$.



Fig. 9: Object as model

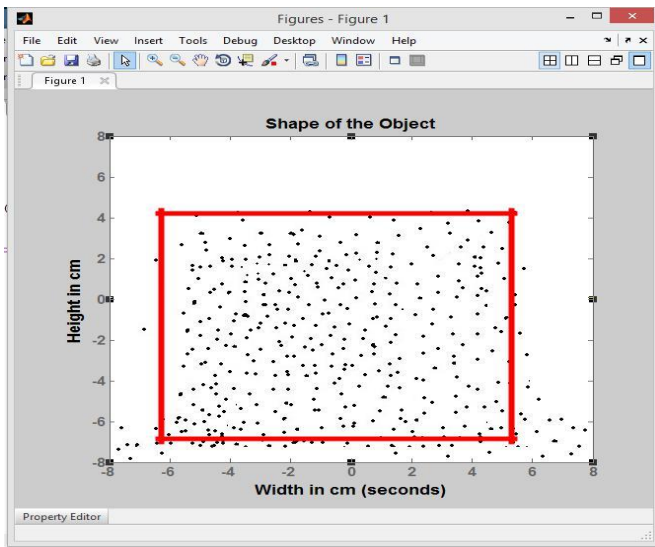


Fig. 10: MATLAB graph showing the shape of the object in dotted form we put a red rectangular dotted surface to show the approximate shape it's determining from the Ultrasonic sound output.



Fig. 11: Round Shape object

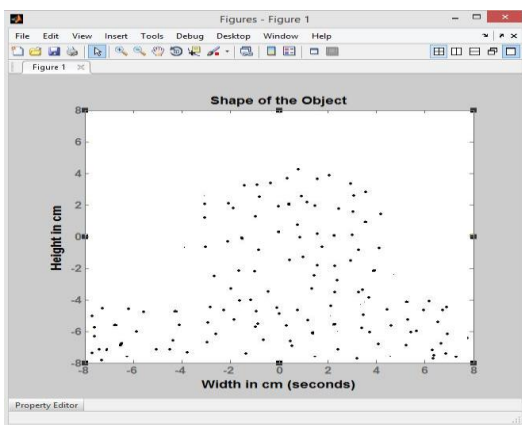


Fig. 12: The output of a round shaped object

The output was the device makes a cluster of dots in area 11x11. The dots showing below the red dotted plot is the output coming from the ground reflection as the object was put in the ground and ground was also giving false alarm. With a better graphics program the base or ground of the object can also be shown in the screen.

VII. CONCLUSION

Major outcome of the project is that it can replace many costly circuits or equipment. It works with almost any surface type. Resistant to vibration, radiation, background light and noise. It's unaffected by dust, dirt or high humidity. Speed of detection and measurement updates fast, works same both in air and water medium and easy to operate. System is easy to manufacture and reliable. It also has high Penetration capability. Its use is ecofriendly. Its flexibility due to software used.

Though considerable drawbacks are that for smaller object the conceptual shape of the object might only be replaced by a single dot, subsequently eliminating any chance of getting any idea of the shape of the object. A more complicated algorithm is needed to be used in this case.

VIII. FUTURE SCOPE

This project can be further enhanced by embedding 3D mapping with drone technology currently available. Same unit can be used as detecting the moving object and find their speed and range in near future. Moreover, it can measure the bearing and distance of ships along with its 3D shape to prevent collision with other ships.

REFERENCES

- [1] C. Singh, A. Agnihotri, P. Patel, "Development Of Radar Using Ultrasonic Sensor," IJARSE, Vol. No.2, Issue No.10, October 2013
- [2] J.E. Piercy, "American National Standard: Method for calculation of the absorption of sound by the atmosphere," ANSI SI-26-1978 (Acoust. Soc. Am., 1978)
- [3] H.K. Ali, J.S. Abdaljabar, S.M. Abdullah, "Design of Ultrasonic Radar," IJESSE, ISSN: 2319-6378, Vol.No.3 Issue.No.7, May 2015
- [4] H.P. Moravec and A. Elfes, "High resolution maps from wide angle sonar," presented at the IEEE conf. Robotics and Automation, 1985.
- [5] L. Kleeman and R. Kuc, "Mobile Robot Sonar for Target Localization and Classification," International journal on Robotics research, volume-14, Number-14, August-1995, pp 295-318.
- [6] S. Punia, R. Singh, "Ultra Sonic Range Finding For Distance Measuring In Coal Mining," IJRET, Eissn:2319-1163 | Pissn: 2321-7308K.

A Wireless Vehicular Traffic Service Management Scheme

M. A. Moyeen¹, G. G. Md. Nawaz Ali^{2,3}, Peter Han Joo Chong⁴, Nazmul Islam¹

¹Institute of Information and Communication Technology, Khulna University of Engineering & Technology,

²Department of Computer Science and Engineering, Khulna University of Engineering & Technology, Bangladesh

³School of Electrical & Electronic Engineering, Nanyang Technological University, Singapore

⁴Department of Electrical and Electronic Engineering, Auckland University of Technology, New Zealand

Email: {moyeen.kuet, nazmul.kuet}@gmail.com, nawaz.ali@ntu.edu.sg, peter.chong@aut.ac.nz

Abstract—Though, an intensive research on traffic management over the decade has eradicated much of its limitations, yet it remained challenging. Inefficient traffic signaling incorporates huge delay that hinders not only our daily activities but can be life threatening. Besides, inefficient traffic signaling and fraudulent vehicles impedes normal traffic system by introducing several uneven phenomena. Also tracking of individual vehicle for citizen safety applications demands an efficient and accurate mechanism. Thus traffic management requires an intelligent system that will be able to handle normal traffic flow as well as emergency vehicles and fraudulent vehicles. This paper proposes a hierarchical wireless traffic entity network driven adaptive traffic signaling mechanism that not only considers normal traffic signaling but also serves emergency vehicles with higher precedence. In addition the proposed approach considers urban traffic lighting scenarios and provides fraudulent vehicle detection mechanism.

Keywords—component; Adaptive signaling, Traffic Entity, Emergency vehicle, Fraudulent Vehicle, Urban Traffic lighting.

I. INTRODUCTION

Traffic congestion is up to now a challenging problem in our day to day life. Extra delays arose by traffic congestion problem can be life threatening when emergency vehicles, e.g., ambulance, fire service vehicles, police cars, gets immobile in to the middle of the road. Traffic delays has substantial impact on the financial activity [1] as business environment requires speed in all aspect. Investors around the globe do not want to spend their precious time waiting in the middle of the road. Besides longer waiting time has significant impact on both physical and mental health [1] which can drastically reduce the productivity. Along with inefficient traffic light control, fraudulent vehicles (e.g., unregistered, non-permitted, malfunctioning vehicles, vehicle containing terrorists or kidnappers) imposes extra unfair situations (e.g., slow movement of vehicle, road accidents, sudden stop of vehicle due to engine failure etc.) resulting in a longer traffic delay.

To address these issues several methods have already been proposed. The most common traffic signaling includes fixed time signaling [2]. With fixed time signaling vehicle has to wait a fixed amount of time even though no vehicle on its cross road. Several Adaptive traffic control systems proposed up to now can handle traffic signaling with adjusting the time frame based on the congestion of the route. Some of them provides facility to handle emergency vehicle. But no precise fraud vehicle detection technique is discussed so far. Proposed approach considers vehicle reporting time and priority information to

mitigate the waiting time for normal and emergency vehicles. Vehicle arrived at an intersection point can only pass through if it has reporting time fallen within currently allowed time frame. This method performs well as it allows vehicle to immediately pass through the intersection point if there is no vehicle on its cross road. Traditional fixed time method and other method who determines green light duration in an adaptive manner will also fail due to their fixed waiting time even though there is no vehicle on the cross road. Again proposed approach only allows to pass the vehicle, which has least reporting time. Thus vehicle which arrived first at an intersection will mostly get served first. Fraudulent vehicle can be detected easily with this approach by allowing them to pass through a single line vehicle scanner. Proposed approach detects fraudulent vehicle even though there is no on board device installed into the vehicle or the on board device fails to communicate. This feature allows to detect fraudulent vehicle automatically. Citizen safety and comfort applications could improve the life style of human being. Proposed approach could assist in this circumstances as it allows V2V and V2I communication [3].

The rest of the paper is organized as follows. Section II discusses some related work. Section III discusses the system model of the proposed approach. Section IV describes the proposed approach. Section V provides some analysis of the proposed approach, and finally Section VI draws the conclusion of the paper.

II. RELATED WORK

As research on traffic congestion problem has been sustained more than a decade, lots of solution have already been proposed. A fuzzy logic based adaptive traffic signaling [4] considers emergency vehicle handling in smart cities. But it introduces some extra delay to make the way for the emergency vehicle.

Micro auction based traffic light control technique [5] can take intelligent decision based on priority and sensor data. This method has not mentioned any central control over the local intersection point. Thus accessing of individual vehicles information to provide the safety facility like fraudulent vehicle detection or reporting roadside robbery becomes impossible.

Adaptive traffic signaling with car to car communication [6] can in fact become cost effective solution, but say to truth, in the real world no one like to burn their own fuel for others, also no one would like to compromise when the question is to proceed

after all other vehicles crossed. It also lacks of fraudulent vehicle detection. The virtual traffic lighting approach [7] collects value of vehicle that desires to pass through the intersection. They makes one of the vehicle as their leader. Leader vehicle has to guide other vehicle to move on. But this mechanism is concerned with energy loss of leader vehicle. Also it does not provide fraudulent vehicle detection facility. Several anti-theft mechanism [8] [9] has been proposed up to now. But when the devices associated with the vehicle for theft control is damaged it becomes almost impossible to detect the vehicle. Sometimes it is almost impossible to detect vehicle from millions of vehicle that was theft. A stochastic modeled intersection has been proposed [10] that collects vehicle backlogs and does perturbation analysis to adjust the metric for controllable traffic cycles.

III. SYSTEM MODEL OF THE PROPOSED SCHEME

The overall architecture of the proposed approach as shown in Fig.1 employs a Central Control Unit (CCU) for the central controlling of traffic management, one or more Zonal Control Unit (ZCU), one or more Local Control Unit (LCU) at each intersection point, one or more vehicle scanner Unit (VSU) and an integrated vehicle On Board Unit (OBU) in each vehicle.

On Board Unit; OBU mounted on each vehicle can communicate with LCU and other vehicle's OBU taking the advantage of Dedicated Short Range Communication (DSRC) [11]. Each vehicle entering to the range of LCU authenticates itself to the LCU and expresses its desire to pass through the intersection point at a certain direction. OBU can pass information among other vehicle's OBU forming a V2V communication to seek help when it falls in danger. Rear vehicle can take precaution hearing information such as sudden break or break failure of front vehicle with the help of V2V communication. Programmable OBU can be enriched with a wide range of applications such as driving assistance, identification of sleepy, anxious, ill or mentally imbalanced driver [12]. The information obtained from other OBU can be used to warn driver for controlling speed of the car or stopping the car in case of prediction of serious vulnerabilities.

Local Control Unit; One or more LCU placed at each intersection point contains hardware that facilitates both DSRC and Long Range Wi-Fi communication. DSRC can be between LCU and OBU or LCU to LCU and Long Range Wi-Fi communication can be between LCU to LCU and LCU to ZCU. LCU can authenticate vehicle by querying the ZCU's database. Some fraudulent vehicle with working OBU can be detected with LCU by utilizing the information obtained from its parent node. LCU controls the traffic as its prime functionality with an intelligent traffic handling mechanism. Local database of LCU holds the currently waiting vehicle's information. LCU sends real time traffic information and other emergency message to ZCU to keep traffic network up to date.

Vehicle Scanner Unit; to aid LCU to detect vehicle not containing OBU or vehicle with faulty OBU, VSU's are placed at some of the intersection point. The placement of VSU can be done by choosing intersection point that represents a zonal corridor. One or more VSU's are placed on the chosen intersection point based on the traffic flow statistics. VSU can

communicate with LCU through DSRC and with ZCU through long range Wi-Fi communication.

Zonal Control Unit; ZCU with the capability of Long Range Wi-Fi communication can communicate with LCU, CCU and other ZCU. The placement of ZCU can be done with the division of the total area into several zones based on the traffic statistics. ZCU can fetch information from CCU about a car when it first enters into a zone. The information fetched from CCU can be shared by LCU and other ZCU to aid faster communication. Each ZCU maintains a local database and that is queried by LCU and other ZCU to get proper information about a vehicle. They also sends real time emergency messages to the CCU and replies with the result obtained for the query from the CCU about a traffic node such as - vehicle, LCU, VSU etc.

Central Control Unit; CCU along with central control over the traffic system maintains the vehicle database to provide real time vehicle information. Integration of criminal record with the central database can facilitate to identify fraud vehicles. CCU is capable of doing Long Range Wi-Fi communication with the ZCU to request about a traffic node or to provide query result of ZCU. They also push notice like road maintenance or road blockage to ZCU that is to be carried out to the corresponding LCU so that ZCU can relocate traffic to different direction. Fig.

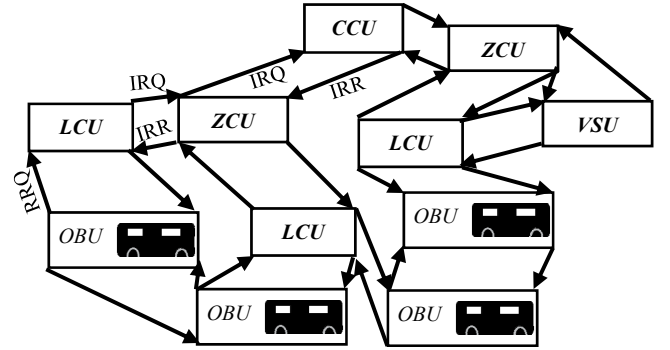


Fig. 1. Proposed system architecture.

1 shows the network entities and direction of communication. With the arrow direction it's apparent that CCU has bidirectional communication with ZCU's and ZCU has bidirectional communication with LCU's, VSU's and CCU. It's also clear that LCU can directly communicate with OBU's, VSU's and it's ZCU in bidirectional mode. VSU's can communicate in bidirectional mode with its nearby LCU and ZCU. OBU can communicate with LCU and other OBU's in bidirectional mode.

IV. THE PROPOSED APPROACH

For each vehicle V_j entering to the area of LCU_j reports its presence with sending a RRQ (Register Request) packet to the LCU_j . RRQ packet holds the vehicle id, reporting time, priority information (as shown in Table I), and an optional emergency code. LCU_j after receiving the request stores them in a temporary location and queries to the ZCU_j about the vehicle with an IRQ (Info Request) packet containing LCU id, vehicle id and emergency code. If vehicle is new to this zone ZCU_j will fetch information about the vehicle from CCU by sending an IRQ packet containing ZCU ID and vehicle ID to the CCU. CCU in reply to the IRQ sends IRR (Information Request Response) packet containing the fraud flag, action flag and vehicle ID.

ZCU_j upon receiving the IRR stores the information into its database and sends the information about the vehicle to LCU_j with an IRR packet containing vehicle ID and action flag. Besides sending IRR to LCU_j if fraud flag is set to True, ZCU_j sends a Fraud Vehicle Detected Message (FRVDM) containing vehicle ID, ZCU ID and LCU ID to the nearest police station so that they can capture the vehicle with the zone and intersection information. If vehicle is already added in the ZCU_j 's database IRR is sent immediately. After receiving the IRR, LCU_j permanently stores the vehicle information containing vehicle id, reporting time and priority information into its database from the temporary storage. Vehicle with higher priority gets privilege so that they experience smaller waiting time in the traffic queue than the lower priority vehicles.

If there is any emergency code in IRQ to the ZCU_j , then ZCU_j informs the emergency services, e.g., hospital, police station with an emergency request containing vehicle ID, ZCU ID and LCU ID. Emergency services upon receiving the request handles the issue as early as possible.

A. Handling Traffic Light:

In order to handle traffic light proposed approach takes the advantage of reporting time of vehicle stored in LCU 's database. LCU allows those vehicle to move on who has reporting time fallen in range determined statistically based on traffic condition.

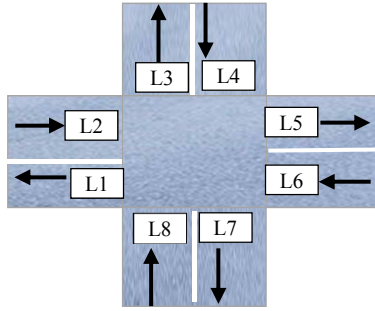


Fig. 2. 4-way Traffic intersection scenario.

Consider a 4-way intersection scenario as shown in Fig. 2. Where L1, L2, L3, L4, L5, L6, L7, L8 represents each individual lane. Note that a vehicle in the junction can take any of the four possible lanes (e.g. vehicles at lane L2 can follow L1, L3, L5 and L7). While consideration of vehicles of lane L2, they are said to be current phase P1 vehicle. Vehicles on other lanes are said to be cross phase P2 vehicle if they have possibility of

colliding with P1 vehicles. If n vehicle $\sum_{j=1}^n V_j$ approaches at P1

and the vehicle on the phase P2 is empty then the vehicles who reported to LCU are free to move on. But if there is already moving vehicle from phase P2 then vehicle of phase P1 will experience stop signal at their side. Vehicles from phase P1 will normally wait until its reporting time to LCU_j falls into range of currently serving time frame. But if emergency vehicle or other priority vehicle appears in phase P1 then LCU_j will find the vehicle with higher priority between phase P1 and phase P2. If vehicle of P2 has higher priority normal traffic flow will commence. When P1 has higher priority vehicle then LCU will

immediately stop traffic from P2 and will start traffic flow from P1. This indeed provides greater flexibility to handle priority vehicles thus saving life and time for priority user. In case of P1 having priority vehicle and P2 serving normal vehicle then P2 will stop traffic flow and P1 will start traffic flow until priority vehicle gets passed. A situation here can arise where vehicle with larger reporting time can get passed before or along with the priority vehicle. This situation can be handled by resetting currently serving time frame to the previously served time frame. Also if all the vehicle of a particular reporting time get passed along with priority vehicle then that time is discarded from the future time frame sequence. If P2 contains priority vehicle and P1 is serving normal traffic flow then P1 will stop service and P2 will start serving traffic flow until priority vehicle get passed through the intersection point. While P1 at service, if both P1 and P2 observes same priority vehicle then P1 will continue serving until priority vehicle get passed from P1. After that P2 will take control and will serve until its priority vehicle get passed. The overall scheme can be thought as Release Reported First Vehicle But Priority (RRFVBP). The sample priority assignment of some type of service is presented in Table I.

Table I. Priority assignment (A sample priority mapping).

Vehicle ID	Type of Service	Priority flag value
V001	Ambulance	1
V002	Fire service	1
V003	Prime Minister/ Guest like president of other country	2
V004	Normal Vehicle	3

In order to handle traffic signaling in urban areas, proposed approach utilizes the previously mentioned RRFVBP technique. Thus the major problem of waiting longer time in intersection point even though there is no vehicle on cross side has been eradicated with the proposed approach. This becomes possible by RRFVBP technique because it keeps open traffic flow if there is no vehicle on cross side.

Pedestrians are allowed to cross road from the phase experiencing stop signal. A fixed amount of time can be allocated for them before starting traffic flow on that phase.

B. Handling fraud vehicle:

VSU 's are designed such that only one vehicle can pass from each VSU . Thus vehicle failed to communicate with LCU can be easily detected and marked as fraudulent vehicle. Also VSU can capture vehicle that is reported fraudulent. VSU can communicate with OBU , LCU and ZCU to detect fraudulent vehicle. If VSU finds a vehicle with ID already marked fraudulent then it can detect it as only one vehicle passes through each VSU . If no OBU exists on the vehicle, then the vehicle is marked as fraudulent. Automatic lane switching can be used to relocate fraudulent vehicle aside so that relevant authority can take action.

V. RESULTS AND DISCUSSION

To evaluate the proposed scheme overall phenomenon has been simulated. The simulation environment consists of a

computer with 8GB ram, core i7 processor, 240GB SSD. Python has been chosen for the programming language.

In order to simulate traffic signaling scenario a vehicle is considered to be present at P1 and total number of vehicle entering P1 and P2 is drawn from Poisson distribution. Simulation are done for fixed time traffic signaling and proposed RRFVBP traffic signaling mechanism. For RRFVBP, a vehicle V_n sitting behind n vehicles will get pass from the intersection if reporting time t_n of V_n falls into the range of currently allowed time frame $t_1, t_2, \dots, t_{n+1}$. Thus total delay experienced by V_n can be obtained from the equation (1),

$$\text{Total Delay, } T_d \approx C_f \times A_d \quad (1)$$

Where C_f means total vehicle in front of vehicle V_n and A_d means average delay introduced while passing each vehicle through the intersection.

Table II. No of Vehicles per reporting time.

Phase P1			Phase P2	
Reporting Time	No. of Vehicle		Reporting Time	No. of Vehicle
t_1	V_{N1}		t_3	V_{N3}
t_2	V_{N2}		t_4	V_{N4}

If t_n does not fall into the range of currently allowed time frame and P1 has vehicles with reporting time t_1, t_2 which are currently being served then equation (1) still persists if there is no vehicle on P2. But if there is vehicle with earlier reporting time t_3, t_4 than t_n then number of vehicles on each reporting time are considered separately to estimate the total delay experienced by V_n . Considering the Table II it should be clear that number of vehicles at each reporting time are taken into account. The total delay can be estimated using the following function,

$$\text{Total Delay, } T_d \approx Q(t) + Q(t') \quad (2)$$

Where t represents range of reporting time allowed at phase P1 and t' represents range of reporting time allowed at phase P2. Time consumed for each individual reporting time follows the equation (1). Thus total time consumed is the sum of delay experienced for each reporting time. Let's consider the table II to calculate total delay for observing vehicle who has reporting time t_n as follows-

$$Q(t) = (V_{N1} + V_{N2})A_d \quad (3)$$

$$Q(t') = (V_{N3} + V_{N4})A_d \quad (4)$$

Using equation (2), (3), (4) total delay experienced by vehicle with t_n reporting time should be

$$\text{Total Delay, } T_{dn} \approx (V_{N1} + V_{N2} + V_{N3} + V_{N4})A_d \quad (5)$$

There is some switching delay that can be ignored due to being small.

From Fig. 3, it can be seen that fixed time traffic signaling introduces fixed delay for any vehicle. If there is no vehicle on

cross side fixed delay on cross side will commence. Thus inefficient traffic control occurs. RRFVBP has lowest delay experienced by each individual vehicle compared to the fixed time method. RRFVBP and FD sometimes intersects as shown in Fig. 3. This is possible if both serves same amount of vehicle on both current side and cross side. The lowest peak of RRFVBP suggests that vehicle in consideration serves as priority vehicle during that time. It is apparent that fixed time method takes same amount of time as it takes for other type of vehicle while handling priority vehicle. But RRFVBP takes smallest possible time to handle priority vehicles.

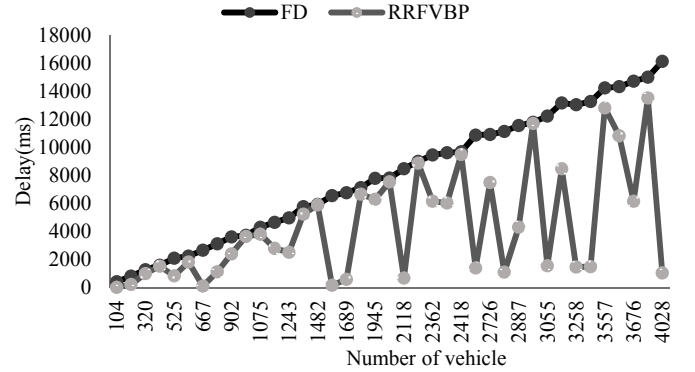


Fig. 3. Delay experienced by individual vehicle with increase in number of vehicle.

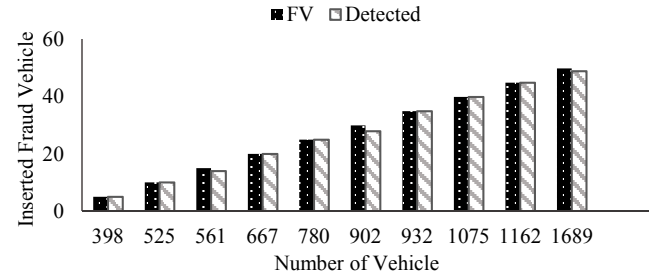


Fig. 4. Number of fraudulent vehicle inserted and detected while increasing the number of the vehicle.

From Fig. 4, it can be seen that while increasing the number of vehicle, total number of inserted fraud vehicle still can be detected with high accuracy by the proposed scheme. Though inserted fraud vehicles are passed through the single line VSU, a very few of them fake their identity to the VSU.

Intelligent management of traffic service not only aids in traffic congestion but also aids in handling fraudulent vehicles. Along with them other citizen safety applications improve the lifestyle of the mass people.

VI. CONCLUSION AND FUTURE WORK

Traffic management in an optimized way could improve the living standard of human being. Proposed approach considers several aspects of traffic service covering traffic signaling, citizen comfort applications, fraud detection mechanism and

other citizen safety applications. Traffic signaling mechanism proposed in this approach has significant improvement over traditional fixed time traffic signaling. Also vehicle on urban areas where less car moves through intersection has been handled with this smart RRFVBP technique. Fraudulent vehicle detection considering this approach has highest success rate as proposed approach works perfectly even if there is no OBU into the fraud vehicle. Road side emergencies handled with the proposed approach are done in an automatic fashion so that public safety is properly assured.

Analysis of the proposed approach considers only 4-way intersection. This guides future research direction to consider n-way intersections, where $n \geq 2$. Also proposed approach did not consider the packet loss during wireless communication between traffic entities. Comparison with other efficient traffic light control system to identify the pitfalls could be another future research direction.

REFERENCES

- [1] Falcocchio, John C., and Herbert S. Levinson. "The Costs and Other Consequences of Traffic Congestion." *Road Traffic Congestion: A Concise Guide*. Springer International Publishing, 2015. 159-182.
- [2] Fayazi, S. Alireza, et al. "Traffic signal phase and timing estimation from low-frequency transit bus data." *IEEE Transactions on Intelligent Transportation Systems* 16.1 (2015): 19-28.
- [3] Dey, Kakan Chandra, et al. "Vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication in a heterogeneous wireless network—Performance evaluation." *Transportation Research Part C: Emerging Technologies* 68 (2016): 168-184.
- [4] Djahel, Soufiene, et al. "Reducing emergency services response time in smart cities: An advanced adaptive and fuzzy approach." *Smart Cities Conference (ISC2)*, 2015 IEEE First International. IEEE, 2015.
- [5] Covell, Michele, Shumeet Baluja, and Rahul Sukthankar. "Micro-Auction-Based Traffic-Light Control: Responsive, Local Decision Making." *Intelligent Transportation Systems (ITSC)*, 2015 IEEE 18th International Conference on. IEEE, 2015.
- [6] Gradinescu, Victor, et al. "Adaptive traffic lights using car-to-car communication." *Vehicular Technology Conference, 2007. VTC2007-Spring*. IEEE 65th. IEEE, 2007.
- [7] Shi, Jingmin, et al. "There is a Will, There is a Way: A New Mechanism for Traffic Control Based on VTL and VANET." *High Assurance Systems Engineering (HASE)*, 2015 IEEE 16th International Symposium on. IEEE, 2015.
- [8] Singh, Pritpal, et al. "Advanced vehicle security system." *Innovations in Information, Embedded and Communication Systems (ICIECS)*, 2015 International Conference on. IEEE, 2015.
- [9] Apasiba, Abeo Timothy, et al. "Enhancing the Security of Vehicles and the Work of the DVLA Using GSM/SMS Technologies." *International Journal of Science and Engineering Applications* 4.5 (2015).
- [10] J. L. Fleck, C. G. Cassandras and Y. Geng, "Adaptive Quasi-Dynamic Traffic Light Control," in *IEEE Transactions on Control Systems Technology*, vol. 24, no. 3, pp. 830-842, May 2016.
- [11] Ali, GG Md Nawaz, et al. "Efficient data dissemination in cooperative multi-RSU vehicular ad hoc networks (VANETs)." *Journal of Systems and Software* 117 (2016): 508-527.
- [12] Ali, GG Md Nawaz, Chan, Edward Chan, and Li, Wenzhong, "On scheduling data access with cooperative load balancing in vehicular ad hoc networks (VANETs)." *The Journal of Supercomputing* 67.2 (2014): 438-468.

Performance Analysis of a Miniaturized Implantable PIFA Antenna for WBAN at ISM Band

Nasim Al Islam* and Farhadur Arifin

Department of Electrical and Electronic Engineering (EEE)

American International University-Bangladesh (AIUB)

Dhaka, Bangladesh

*nasim.islam.bd@ieee.org

Abstract— A flexible implantable PIFA (Planar Inverted F Antenna) antenna for biomedical applications is proposed in this study. The main notability of this design refers to its subtle dimension, flexibility and subordinate thickness that makes it perfectly suitable for implementing inside human or animal tissues. The antenna is aimed to operate in the Industrial, Scientific and Medical (ISM) band (2.4–2.4835 GHz). The thickness of this antenna is only .735 mm, which implies that this antenna is reliable to perform under bent conditions. The antenna offers a compact design with a dimension of 9.48 mm × 7.8 mm × .735 mm (54.348 mm³). Copper and Rogers R03010 are chosen as the patch material and substrate material accordingly. The antenna is encapsulated inside biocompatible material Rogers R03010 for safety concern in skin or muscle tissues. Several types of calculations and performance measurement of this antenna have been done by using CST Microwave studio in both planar and bent conditions by maintaining all the electrical properties of human skin tissue. Finally, Specific Absorption Rate (SAR) and thermal loss are evaluated to comply with the antenna safety issues.

Keywords- PIFA, ISM Band, Implantable Antenna, Human Tissue Model, Specification Absorption Rate (SAR).

I. INTRODUCTION

Implantable devices have been developed enough to enhance patients nursing quality and to affirm patients safety now a days. In this modern era, it is impossible for most patients to afford long-term hospital stays due to financial confinements, work and other reasons, even though their health status must be monitored in a real-time or short periodic time mode. In order to keep pace with the technological advancement in healthcare sector, remote diagnostic systems can play emergent role by coordinating with wireless communication technology. As a result, implantable antenna is becoming a vital part of portable medicinal devices focusing on real-time monitoring [1].

Miniaturization is a key challenge to be proficient for implantable medical devices in the human body. For designing implantable antenna, variation of electromagnetic properties in the vicinity of human tissue is to be considered. These properties alter the antenna characteristics such as resonance frequency and impedance matching [2-3]. It is also necessary to encapsulate the antenna with biocompatible material to ensure the protection of human tissue from radiation [4].

Wireless Body Area Network (WBAN) provides connectivity between different nodes of body with the help of sensors or antenna to a remote health monitoring system

through wireless communication channel. WBANs supporting healthcare applications are in early development stage but offer important commitments at monitoring, diagnostic and therapeutic levels. The RF technology is improvised enough that data retrieved from a medical device can be sent into a receiver module which might be far away from patient and thus an effective way of healthcare improvement is possible. However in this structure, the data exchange capability is limited and the range between internal and external system is confined. Also in practical cases, for establishing communication of WBAN through an in body implantable antenna is always challenging due to the size constraint and safety regulations [5]. This is where the main motive of this research work comes in.

For designing a biocompatible antenna, several bands are mostly considered like Medical Implant Communication Service (MICS), Wireless Medical Telemetry Service (WMTS) and Industrial, Scientific and Medical (ISM). These bands provide a wide range various frequency range for biomedical applications. ISM band has been more suitable for biomedical applications due to its high frequency bandwidth, which allows making the wavelength smaller as well to obtain higher bit rate after implantation [6-7].

Now a days, different kind of artificial body implants like cochlear implants, artificial eyes, brain and cardiac pacemakers [8-10], nerve signal recorders [11] etc. are being controlled and monitored through RF technology. Recent researches in the design of implantable antennas cover several factors like size reduction, bending capability, bandwidth requirement, SAR evaluation, biocompatibility, radiation and coupling effects [12]. With improvement of safety requirements as well as reduced cost and size, usage of these kinds of devices is opening a new era in biomedical device arena. The proposed antenna in this context can offer valuable contribution in biomedical and healthcare applications. One of the key features of this antenna is its miniscule dimension with proper evaluation of safety issues.

The organization of this paper is described as follows. In Section II, modeling and design procedure of the antenna is narrated. In Section III, several parameters of the antenna have been calculated and analyzed for both planar and bend condition. In Section IV, comparison in different context is demonstrated to prove the versatility of this antenna. Finally, Section V covers the concise summary of this work.

II. MODELING OF THE ANTENNA

In this paper, a flexible rectangular patch shaped miniaturized antenna is designed for different biomedical applications. Configuration of this antenna is mainly based on PIFA type. In order to make antennas suitable for in body applications, there are two major issues that need to be considered: miniaturization and biocompatibility. Shrinking the dimension of the ground plane and placing radiating material on both sides of the substrate material, size reduction is achieved. The entire antenna was encapsulated by biocompatible material Rogers R03010 for maintaining the safety regulations properly. For designing this antenna, Rogers R03010 is used as the substrate material which has a dielectric constant of 10.2, $\tan \delta$ of 0.0035 and thickness of 0.635 mm. This kind of low thickness of antenna proves its flexibility to bend as well as reduces the total electric size of the antenna. Copper with a thickness of 0.1 mm is used as the patch and ground material for lessening the cost as well as due to its availability. Meandering slots were cut according to the current distribution to keep the antenna working accurately under ISM Band. ISM band allows the diversity for shortening the wavelength with a high bit rate [13-14].

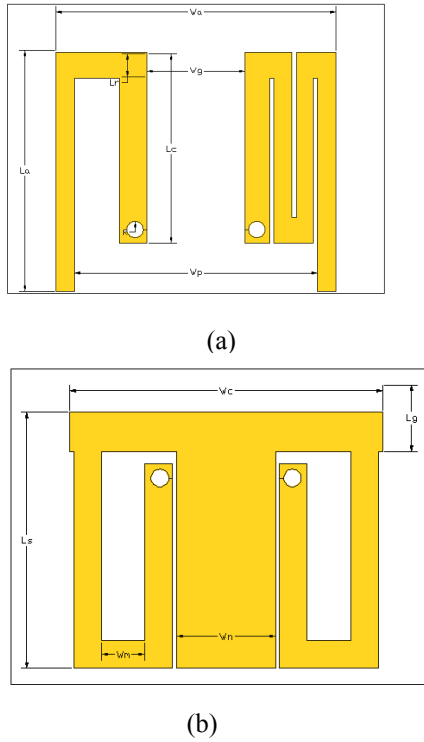


Fig. 1. Model of the Antenna: (a) Top layer; (b) Bottom layer

The entire antenna is wrapped by a superstrate material of Rogers R03010 which prevents direct contact between radiating material with human skin tissue and also acts as a buffer between these two layers for lessening the radiation loss inside human tissue [15]. Fig. 1 shows the top and bottom layers of the antenna with their size parameters. The whole antenna was designed and simulated inside human skin tissue

model using CST Microwave Studio. Final geometrical size of this antenna is $9.48 \text{ mm} \times 7.8 \text{ mm} \times .735 \text{ mm}$ (54.348 mm^3).



Fig. 2. Cross-sectional view of one layer skin model geometry of the antenna

In Fig. 2, the antenna in the vicinity of human skin tissue is presented for a better understanding of the placement of antenna. Table I summarizes the antenna size parameters.

TABLE I. Detailed Antenna Size Parameters

Parameter	Size (mm)
Wa	9.240
La	7.862
Wg	3.240
Wp	8.040
Lc	6.178
Ln	0.737
R	0.270
Wc	9.480
Ls	7.800
Wm	1.320
Wn	3.000
Lg	2.0278

TABLE II. Dielectric Values of Human Skin Tissue at Different Frequencies

Frequency (GHz)	ϵ_r	σ (S/m)
2.0	39.29	1.45
2.4	38.79	1.70
2.45	38	1.46
2.5	37.66	1.77
3.0	37.05	2.14

Table II visualizes the dielectric properties like permittivity and conductivity of the human skin tissue which needs to be maintained according to the operating frequency of the antenna [16].

III. SIMULATION AND COMPARISON ANALYSIS

Several antenna parameters were simulated for strengthening the reliability of the antenna for biomedical applications. The antenna performance was measured inside human skin tissue model where the encapsulated antenna is immersed inside the skin bio tissue of thickness 2.682 mm at CST Microwave Studio simulation environment. Fig. 3 shows the antenna inside the skin tissue model.

Fig. 4 demonstrates the S11 parameter of the antenna inside skin tissue model. S11 parameter or reflection coefficient of an antenna exhibits the amount of power which is being radiated from the antenna. It can be seen from Fig. 4 that the antenna has a resonant frequency 2.48GHz with a return loss of -17.81dB which satisfies the ISM band. The bandwidth of the antenna is around 198.4 MHz.

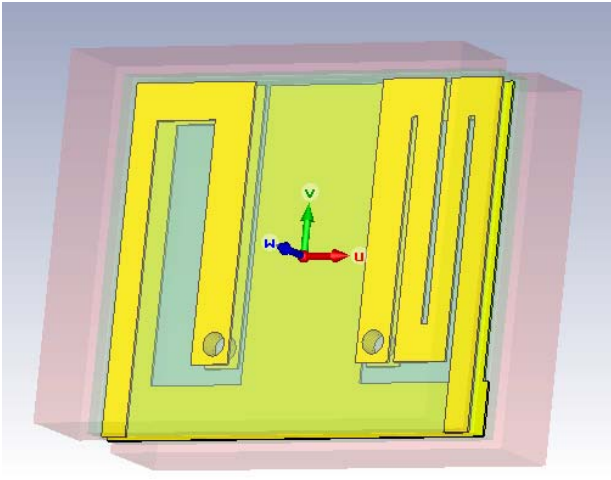


Fig. 3. Antenna inside Skin Tissue Model

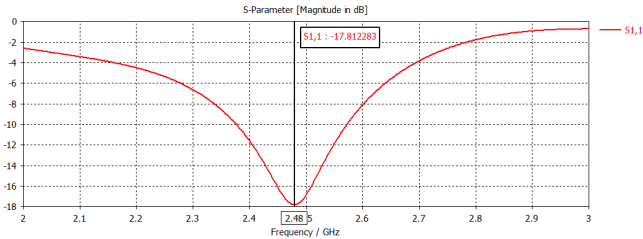


Fig. 4. S11 parameter of the antenna inside Skin Tissue Model

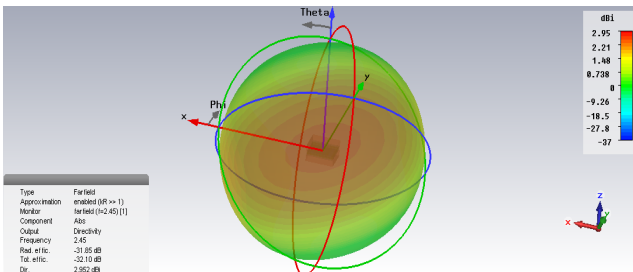


Fig. 5. 3D view of Far-Field Region of the antenna inside Skin Tissue Model

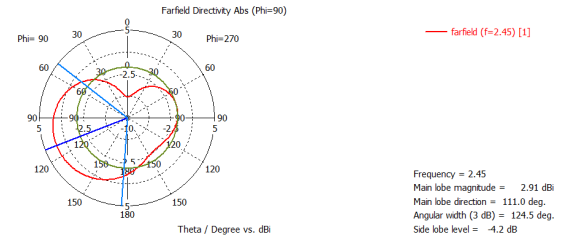


Fig. 6. 2D view of Far-Field Region of the antenna inside Skin Tissue Model

Radiation characteristics of the antenna can be manipulated from 3D and 2D view of the Far-Field radiation pattern. From Fig. 5 it is observed that the antenna has a directivity of 2.952 dBi. From Fig. 6, it can be observed that the main lobe magnitude is 2.91 dBi which is centered at 111 degree with an angular width of 124.5 degree. The maximum gain of the antenna is -28.898 dBi at $\theta = 75$ degree and $\phi = 90$ degree.

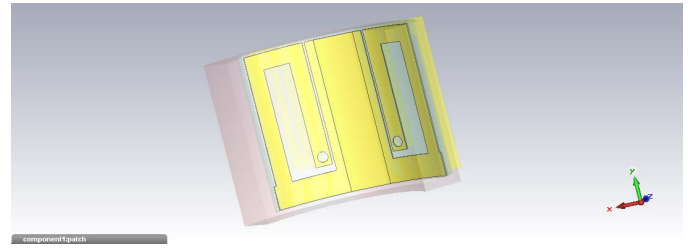


Fig. 7. Antenna undergoing Bent Test

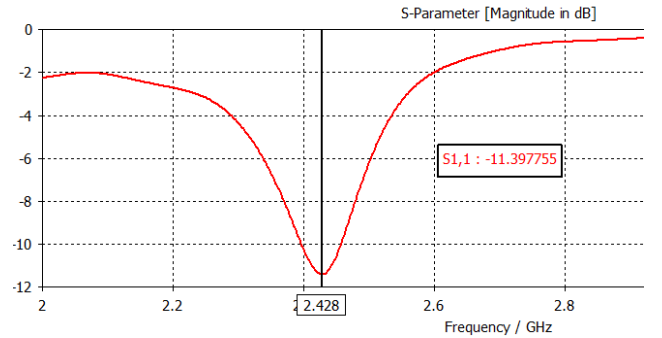


Fig. 8. S11 parameter of the antenna at Bent Condition

As this antenna is subjected to work inside the body, it should work perfectly under bent condition as there are many curvature in human body. Fig. 7 shows the antenna in bent condition when the antenna is convexed around the x-axis around a 6 cm diameter cylinder. Thickness of the antenna allows it to be bent very easily within the human skin tissue. Fig. 8 represents the S11 parameter of the antenna while bent condition and from the graph it is visible that it covers the ISM band though the return loss is quite low.

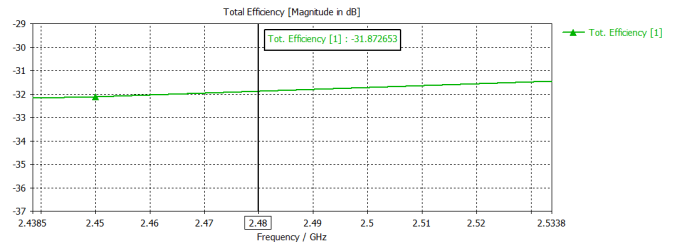


Fig. 9. Total Efficiency of the antenna inside Skin Tissue Model

From Fig. 9, we can observe that radiation efficiency of this antenna is achieved around -31.85 dB at the resonant frequency.

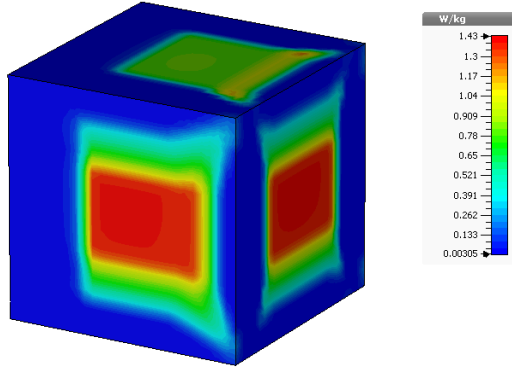


Fig. 10. 3D SAR Distribution subjected to 13.5 mW of input power

Specific Absorption Rate (SAR) is that parameter of antenna which measures the amount of radiation that is being observed by surrounding tissue. SAR should be under certain limiting value so that the radiation cannot harm human body. The standards for SAR calculation as regulated by IEEE/IEC 62704-1 standard states that the 10-g averaged SAR should not exceed 2W/kg as set by FCC and ICNIRP guidelines [17-18]. From Fig. 10, we can see that the peak value of SAR is 1.427W/Kg for an input power of 13.5 mW which is safe under safety regulations.

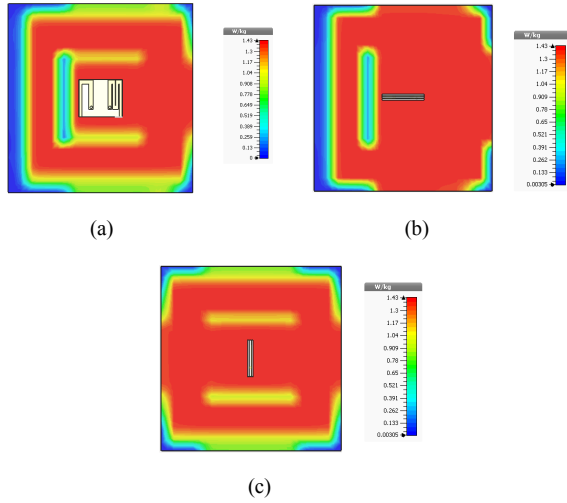


Fig. 11. SAR Distribution in (a) X-Y plane, (b) X-Z plane and (c) Y-Z plane

In Fig. 11(a), (b) and (c), the SAR distribution in different planes are shown. Each of them exhibits a maximum SAR value less than 2W/Kg, which makes the antenna suitable for in body implantation. Another important parameter regarding the safety of WBAN device is the thermal loss. Sometimes thermal loss can be very high even with a lower SAR value which cannot be tolerated [24]. Fig. 12 shows the thermal loss calculation of this antenna which reflects a lower value of thermal loss.

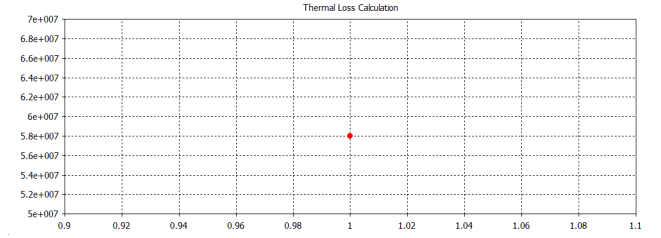


Fig. 12. Thermal Loss of the antenna inside Human Skin Tissue

IV. COMPARISON ANALYSIS

TABLE III. Performance Comparison of Different Antennas

References	Antenna Type	Operating frequency	Dimensions (mm ³)	Miniaturization method
[11]	Cavity slot	2.4 GHz	17.92	H-shaped slot
[12]	PIFA	402 MHz/ 2.4 GHz	135.45	PIFA
[15]	PIFA	402 MHz/ 2.4 GHz	3072	PIFA
[19]	PIFA	402 MHz/ 2.4 GHz	1265.63	Meandered PIFA
[20]	Patch	402 MHz/ 2.4 GHz	857.25	Split ring resonator
[21]	PIFA	402 MHz/ 2.4 GHz	534.48	Multi-layer PIFA
[22]	PIFA	402 MHz/ 2.4 GHz	691.52	Spiral PIFA
[23]	PIFA	402 MHz/ 2.4 GHz	468.12	PIFA, open-end slots
Proposed	PIFA	2.45 GHz	54.35	Meandered PIFA

TABLE III draws the comparison analysis in the aspect of antenna type, dimension, operating frequency, bandwidth and miniaturization method among the proposed antenna and several reported antennas. The proposed antenna in this paper exhibits a very thinner patch compare to its counterparts. Fig 13. represents the comparison between S_{11} parameter in planar and bent conditions. In both conditions, the antenna works perfectly in ISM band though the resonant frequency has shifted a bit for bent condition.

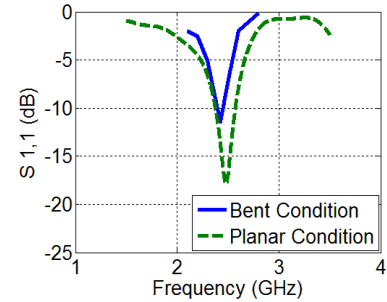


Fig. 13. Comparison of S_{11} parameter in Planar and Bent condition

V. CONCLUSION

A miniaturized flexible meandered patch implantable PIFA antenna for Wireless Body Area Network (WBAN) operating at ISM Band is designed. The miniaturization is achieved by cutting slots and by introducing a novel top layer design, all while maintaining the resonant frequency at the ISM band. After completing the modeling part, the whole antenna was first inserted inside a biocompatible Rogers R03010 superstrate and then the whole encapsulated antenna was placed inside a one layer skin model to mimic an actual real world scenario of such implantation inside human body. All the performance measurements such as resonant frequency, far-field radiation pattern, total efficiency, SAR and thermal loss were then taken with the antenna inside the skin tissue model. After that, the antenna was subjected to bending and its performance parameters were again measured and compared with its planar condition. Some of the main points that can be drawn from this antenna are-

- Thin, tiny and bendable implantable design.
- High bandwidth and good return loss at 2.48GHz resonant frequency.
- Ability to bend without much change in its performance, meaning it has the possibility of taking the shape of a capsule like pill.
- SAR lower than the maximum limit regulated by IEEE/IEC 62704-1 standard.

Overall, the results indicates that the antenna exhibits all the performance parameters needed to be called implantable. Future work should include fabricating the antenna and testing it out with a human mimicking gel and comparing it with the simulated results. Furthermore a total biomedical system can be incorporated with the antenna to test it out in a more real world scenario.

REFERENCES

- [1] S. Zhang, Y. P. Qin, P. U. Mak, S. H. Pun, and M. I. Vai, "Real-time medical monitoring system design based on intra-body communication," *Journal of Theoretical and Applied Information Technology*, vol. 47, no. 2, pp. 649–652, 2013.
- [2] T. Karacolak, R. Cooper, E. S. Ünlü, and E. Topsakal, "Dielectric properties of porcine skin tissue and in vivo testing of implantable antennas using pigs as model animals," *IEEE Antennas Wireless Propag. Lett.*, vol. 11, pp. 1686–1689, 2012.
- [3] D. C. Garrett, B. A. Besler, J. Bourqui, E. C. Fear, and A. M. Setup, "Estimating Bulk Dielectric Properties of Biological Tissues at Microwave Frequencies," no. 1, pp. 2–3, 2016.
- [4] F. Merli, B. Fuchs, J. R. Rosig, and A. K. Skrivervik, "The effect of insulating layers on the performance of implanted antennas," *IEEE Trans. Antennas Propag.*, vol. 59, no. 1, pp. 21–31, Jan. 2011.
- [5] K. Agarwal and Y. Guo, "Interaction of Electromagnetic Waves with Humans in Wearable and Biomedical Implant Antennas," pp. 154–157, 2015.
- [6] D. C. Sirait, Basari, F. Y. Zulkifli, and E. T. Rahardjo, "An implanted dipole antenna for RFID-based patient monitoring system," 2013 Int. Conf. Qual. Res. QiR 2013 - Conjunction with ICCS 2013 2nd Int. Conf. Civ. Sp., pp. 142–145, 2013.
- [7] W. Xia, K. Saito, M. Takahashi, and K. Ito, "Performances of an implanted cavity slot antenna embedded in the human arm," *IEEE Trans. Antennas Propag.*, vol. 57, no. 4, pp. 894–899, Apr. 2009.
- [8] Vorobyov, C. Hennemann, a. Vasylenko, J. D. Decotignie, and J. Baumgartner, "Folded loop antenna as a promising solution for a cochlear implant," 8th Eur. Conf. Antennas Propagation, EuCAP 2014, no. EuCAP, pp. 1735–1738, 2014.
- [9] Z. Psenakova, M. Smondrk, J. Barabas, G. Lo Sciuto, and M. Benova, "Simulation and Assessment of Pacemaker RF Exposure (2 . 4 GHz) by PIFA Antenna," pp. 569–573, 2016.
- [10] M. K. Hosain, A. Kouzani, and S. Tye, "Multi-layer implantable antenna for closed loop deep brain stimulation system," 2012 Int. Symp. Commun. Inf. Technol. Isc. 2012, pp. 643–648, 2012.
- [11] A. Al-Armaghany, B. Yu, T. Mak, K. F. Tong, and Y. Sun, "Feasibility study for future implantable neural-silicon interface devices," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, pp. 3009–3015, 2011.
- [12] Z. Duan, Y. X. Guo, M. Je, and D. L. Kwong, "Design and in vitro test of a differentially fed dual-band implantable antenna operating at MICS and ISM Bands," *IEEE Trans. Antennas Propag.*, vol. 62, no. 5, pp. 2430–2439, 2014.
- [13] H. Usui, M. Takahashi, and K. Ito, "Radiation characteristics of an implanted cavity slot antenna into the human body," in *Proc. IEEE Antennas and Propag. Soc. Int. Symp.*, Albuquerque, NM, 2006, pp. 1095–1098.
- [14] J. L. Buckley, K. G. McCarthy, L. Loizou, B. O'Flynn, and C. O'Mathuna, "A Dual-ISM-Band Antenna of Small Size Using a Spiral Structure With Parasitic Element," *IEEE Antennas Wirel. Propag. Lett.*, vol. 15, no. c, pp. 630–633, 2016.
- [15] J. Kim and Y. Rahmat-Samii, "Implanted antennas inside a human body: Simulations, designs, and characterizations," *IEEE Trans. Microw. Theory Tech.*, vol. 52, no. 8, pp. 1934–1943, Aug. 2004.
- [16] S. Gabriel, R. W. Lau, and C. Gabriel, "The dielectric properties of biological tissues," *Phys. Med. Biol.* pp. 2231–2293, Oct. 2004 [Online]. Available: <http://niremf.ifac.cnr.it/tissprop/>, (1996)
- [17] Safety, E., Committee, S. C., Radiation, N., & Board, I. S. (2008). *IEEE Recommended Practice for Measurements and Computations of Radio Frequency Electromagnetic Fields With Respect to Human Exposure to Such Fields , 100 kHz –300 GHz. Measurement* (Vol. 2002).
- [18] ICNIRP, "Guidelines for limiting exposure to timevarying electric, magnetic, and electromagnetic fields," *Health Phys.*, vol. 74, pp. 494–5229, 1998.
- [19] T. Karacolak, A. Z. Hood, E. Topsakal, "Design of a dual-band implantable antenna and development of skin mimicking gels for continuous glucose monitoring," *IEEE Trans. Microw. Theory Tech.*, Vol.56, pp. 1001–1008, Apr 2008.
- [20] C.J. Sa'nchez-Ferna'ndez, O. Quevedo-Teruel, J. Requena-Carrio'n, L. Incl'a'n-Sa'nchez, E. Rajo-Iglesias, "Dual-band microstrip patch antenna based on short-circuited ring and spiral resonators for implantable medical devices," *IET Microwaves, Antennas Propag.*, vol.4, no.8, pp. 1048-1055, 2010.
- [21] F. Merli, L. Bolomey, J.-F. Zürcher, G. Corradini, E. Meurville and A.K. Skrivervik, "Design, realization and measurements of a miniature antenna for implantable wireless communication systems," *IEEE Trans. on Antennas and Propagat.*, vol. 59, no. 10, pp. 3544–3555, Oct. 2011.
- [22] C. Liu, Y. X. Guo, and S. Xiao, "Compact dual-band antenna for implantable devices," *IEEE Antennas and Wireless Propag. Lett.*, vol. 11, pp. 1508-1511, 2012.
- [23] L.J. Xu, Y.X. Guo, W. Wu, "Dual-Band Implantable Antenna with Open-End Slots on Ground," *IEEE Antennas and Wireless Propag. Lett.*, vol. 11, pp. 1564-1567, 2012.
- [24] S. Caporal Del Barrio, G. F. Pedersen, and A. S. Morris III, "Thermal loss becomes an issue for narrow-band tunable antennas in fourth generation handsets," *IET Microwaves, Antennas Propag.*, vol. 9, no. 10, pp. 1015–1020, 2015.

An Automatic Layout Faults Detection Technique in Responsive Web Pages Considering JavaScript Defined Dynamic Layouts

M. A. Moyeen¹, G. G. Md. Nawaz Ali^{2,3}, Peter Han Joo Chong⁴, Nazmul Islam¹

¹Institute of Information and Communication Technology, Khulna University of Engineering & Technology,

²Department of Computer Science and Engineering, Khulna University of Engineering & Technology, Bangladesh

³School of Electrical & Electronic Engineering, Nanyang Technological University, Singapore

⁴Department of Electrical and Electronic Engineering, Auckland University of Technology, New Zealand

Email: {moyeen.kuet, nazmul.kuet}@gmail.com, nawaz.ali@ntu.edu.sg, peter.chong@aut.ac.nz

Abstract—Revolutionary changes in technology has brought a lots of different types of devices towards the Internet enabled services. It has become challenging to keep web sites functional and user friendly in a wide range of devices. Responsive Web Design (RWD) steps up to facilitate the necessary support to overcome those challenges with complex cascading style sheet (CSS) and JavaScript defined layouts. Fluidly modified layout that controls appearance of Responsive Web Pages (RWP) is subjected to erroneous when faulty CSS and JavaScript modifies the actual look of the web page. Though several methods have already been proposed to detect layout faults in web pages, they are error-prone and time consuming in nature. Thus an automatic approach considering both CSS and JavaScript defined dynamic layout fault detection technique have been proposed in this paper. Simulation result demonstrates that proposed technique outperforms the existing methods in case of time requirement, memory requirement and fault detection rate.

Keywords—Responsive Web Design, Layout Fault, Responsive Layout Graph

I. INTRODUCTION

With the tide of technological advances, lots of portable and non-portable computing devices are being used nowadays. A recent statistic suggests that every day over 3.5 million new devices are being activated worldwide [1]. A majority of them demands web enabled services. Web enabled services are no longer limited to information services, rather a lots of other types of services (e.g., E-commerce, social communication, Medical services, on-line games and many more) are being used at these days. Devices with different view port width and height makes it difficult to represent website in a uniform manner. One solution to this problem involves making different version of the website for different device types and serve the request based on viewers viewing platform [2]. 67% respondents in a Google survey prefers purchase from a mobile friendly website [3]. Though some giant companies (e.g., Facebook, twitter etc.) maintains separate versions of mobile sites for different device types. Nonetheless, it is tough to maintain each of them separately [4]. Being an optimized version of the main website [5], these versions are less capable of fulfilling the visitors' desire.

RWD (Responsive Web Design), a recent approach, allows to create a single website, and controls the appearance of the

website across a wide range of devices with varying viewport size and resolutions by adapting contents and layouts [6]. The key idea behind this is to fluidly distribute web elements such that they can organize their positions in an adaptive fashion. RWD allows any kind of devices to join and take the taste of the internet whereas statistically dependent fixed width designs limit the access from devices disjunctive to statistical results [5]. RWD with fluid grid and flexible images is capable of tackling all the possible resolutions [7].

Keeping all the challenges in mind, RWD introduces a concept- "media query" which can detect current viewport width, orientation, height and repositions and resizes web elements such that spaces are used in an optimum way insuring better user experience [7]. Using this powerful CSS (Cascading Style Sheet) feature not only the device information are obtained, but also the device dependent user demands can be served.

Though RWD has huge benefits, it is very complex to test the responsively designed web pages. A small change in CSS can result in a huge damage to the existing layout. Also changes introduced by JavaScript are hard to track due to the dynamic nature of JavaScript driven actions. Several methods have already been proposed but most of them needs manual effort to detect error introduced in the layout. Again some of them are image based techniques which have performance issues in case of both time and memory consumption. Moreover, image based techniques do not provide actual error message. So it is often hard to locate the actual position of error in the source code. An automatic Responsive Layout Graph (RLG) based approach REDECHECK [3] can detect CSS introduced layout faults. Our proposed technique is an extension of REDECHECK technique. The prime concern of the proposed technique is the detection of JavaScript defined layout faults that is not considered in the RLG based automatic technique.

To evaluate the proposed technique four responsive websites are considered for simulation. Simulation result suggests that proposed technique has much more fault coverage than the exiting technique REDECHECK.

The rest of the paper is organized as follows. Section II introduces some related works. Section III describes the background of the proposed technique. Section IV describes

the proposed technique. Section V presents results and discussion, and finally Section VI draws conclusion with some future guidance.

II. RELATED WORKS

To test layout of responsively designed website several methods have already been proposed. An automatic potential layout fault detection technique [3] can detect potential layout fault introduced by changes in CSS defined layout. They have introduced a RLG based technique that has several advantages over traditional image comparison technique [8].

Record and replay based techniques stores the user interaction component in a script and replay later [8] [9] [10]. These techniques can store user interaction coordinate and later use them for testing. But a change in GUI component needs to change the whole script. Another cross browser testing approach implemented called X-PERT [11], focuses on single resolution cross browsing testing. They do not consider RLG based technique and suffers from time and memory consumption issue.

A recent technique aims at automatically finding differences between how a web page should look and its final design [12]. But it requires a lots of image to be saved for testing the fluidity of the web page under test. Currently several developer tool allows testing web pages in different viewport width. Multi-screenshot tool [13] allows a tester to view web page on multiple viewport width. But they are not automatic like the proposed approach.

III. BACKGROUND

A. Responsive Web Design

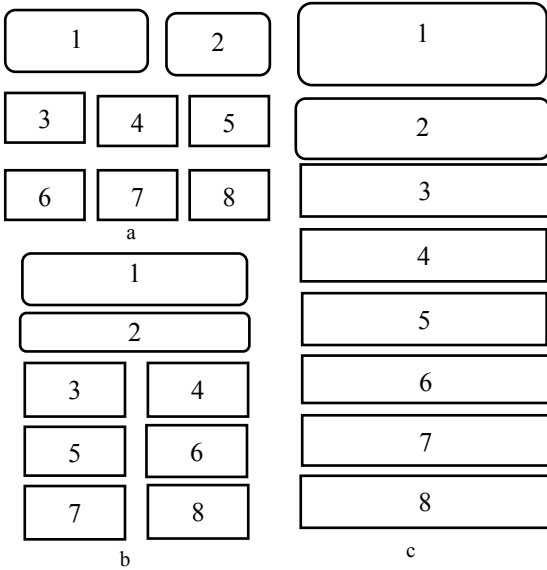


Fig. 1. Web element distribution on (a) 1024 pixel width, (b) 768 pixel width and (c) 320 pixel width.

In RWD, no matter which device is in use, web elements are automatically adjusted and distributed evenly so that every element and associated contents of the web page fits into the device's viewport area. This can be illustrated with the Figs. 1(a), 1(b) and 1(c).

From Fig. 1 (a) it can be seen that when a device width is 1024 px, one of the web element numbered as 1 occupied more than 50 percent of the viewport width, web element numbered 2 takes rest of the available space, web elements numbered from 3 to 8 takes their position in a three column representation. From Fig. 1 (b) when viewport width is 768 pixel then web elements numbered 1 and 2 occupy the full width of the device. In this case web element numbered 1 has much more height than numbered 2. Web elements numbered 1 through 8 take full width of the viewport when viewport width is 320 pixel as shown in Fig. 1 (c). In this circumstance web element numbered 1 has more height than others to fit its content to the viewport area. Web element numbered 2 has also more height for the same purpose than web elements numbered 3 to 8. From these three figures it is apparent that web elements are repositioned and reshaped to hold responsiveness of the web pages. This repositioning and reshaping is done with the help of media query, one of the blessings of CSS.

B. Layout Faults

When faulty CSS rule breaks the responsiveness of the web layout, layout faults occurs. For example, for a viewport width of 320px when any of the web element size is greater than 320px, then it will not fit into the viewport width. Thus layout fault happens. From Fig. 2 it can be seen that when faulty CSS

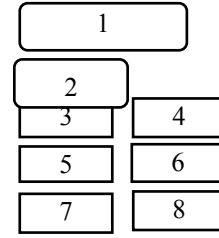


Fig. 2. Faulty CSS rules breaks the layout by throwing down the web element 2 in the second line.

rule enhances the width of the web element numbered 1 then the web element numbered 2 breaks down to the second line. Another faulty CSS rule aligns web element numbered 2 over web element numbered 3. This undesirable phenomenon breaks the responsiveness of the web page.

IV. PROPOSED APPROACH

Proposed approach utilizes the novel concept Responsive Layout Graph (RLG) [3] that considers visibility, width and relative alignment of web elements to aid in automatic layout fault detection. A RLG is a graph in which each node representing a web element, holds additional information about visibility constraint, width constraint and alignment constraint. For a web element E_i visibility constraint V_c represents a set of tuple $\{(x_1, x_2), (x_3, x_4), \dots, (x_n, x_{n+1})\}$, where each tuple signifies a pair of viewport width (e.g., (x_1, x_2)) for which E_i is visible in the DOM of a web page W .

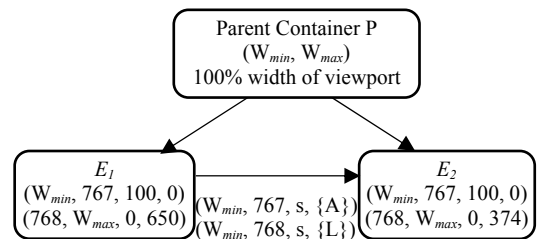


Fig. 3. An example of RLG fragment for the layout as shown in Fig. 1. (a) and (b).

For a web element E_l width constraint W_c is a 4-tuple (x_1, x_2, m, c) that denotes varying nature of width WE_l of web element E_l in response to the variation of width WE_p of parent element E_p . WE_l can be obtained from the linear equation $WE_l = \frac{(m \times WE_p)}{100} + c$, where m and c defines percentile

value of parent width WE_p and fixed width offset respectively. For example, for a 4-tuple (Wmin, 767, 100, 0) web element E_l occupies full width of the parent element WE_p .

Alignment constraints concept is directly drawn from alignment graph [3], where parent child or siblings relationship and set of alignment attributes $Q = \{L, R, A, B, \dots\}$ are considered. A parent child relationship pc exists between two web elements if one of the element is direct container of another element. A sibling relationship s exists between two web elements if both of them belong to a same parent container. The alignment attribute L/R/A/B denotes a web element is Left/Right/Above/Below of another web element.

An example RLG fragment shown in Fig. 3 denotes E_1 and E_2 are sibling. When viewport max width is 767px, E_1 is aligned above E_2 . But when max width is $W_{max} \approx 1024$ px, E_1 resides left side of E_2 , E_1 occupies 650px of width, and E_2 holds 374pixel of width.

Proposed approach presented here is organized such that Subsection A states the RLG generation process, and Subsection B states the automatic fault detection technique.

A. RLG Generation

RLG of stable version of web pages that works as reference RLG, and RLG of web page while web page under test, are the two cases of RLG generation of web pages. Overall RLG

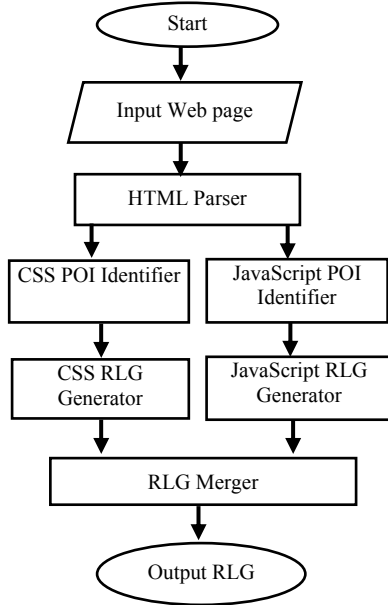


Fig. 4. RLG generation process of web pages considering both CSS and JavaScript defined layout.

generation process are described in Fig. 4. The process of RLG generation begins with taking web page as input.

The input data from web page consists of HTML, CSS and JavaScript programs. With the input data HTML parser at first generates tree like structure where each node N holds additional information such as id selector, class selector, inline styles and JavaScript functions. After the generation of the tree like structure two parallel process denoted by CSS POI Identifier, and JavaScript POI Identifier starts identification of point of interest (POI) from the output of HTML parser. The point of interest here is identified through relevant HTML layout controlling tags and CSS selectors. For JavaScript RLG generation function calls inside HTML element is taken into account with high importance. CSS RLG generator process followed by CSS POI Identification begins parsing of CSS element from inline styling, internal style sheets and external style sheets. Relevant values of layout controlling properties are extracted, and RLG is generated by gaining information of relative alignment of each of the layout controlling web elements from the parsed HTML tree like structure.

JavaScript defined dynamic layout has potential difficulties in generation of RLG due to lots of possibility of JavaScript functions that changes web elements properties. Proposed approach taking those difficulties into consideration, limits the range of layout controlling functions and attributes by identifying point of interest. In JavaScript RLG generation process, elements those are included in the POI are located inside JavaScript program segments. After that relevant actions controlling layout of the webpage are identified. A RLG is generated by calculating relevant attributes such as- width, height, padding, margin etc. For example, a JavaScript code segment doubles the width of a web element, so that calculation of exact width of that element after the JavaScript action is done and reflected in the generated RLG. In RLG generation process visibility attributes such as 'display', 'hide()', 'show()', 'hidden' etc. are also taken into account.

After both CSS and JavaScript RLG are generated, a sophisticated merge process initiates. There JavaScript RLG overrides CSS RLG. Resultant RLG is returned for storage or comparison with the benchmark RLG.

B. Automatic Layout Fault Detection

Automatic Layout fault detection process incorporates the

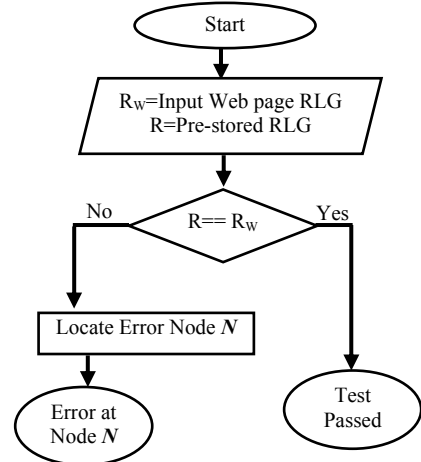


Fig. 5. Automatic Layout Fault Detection Process.

above mentioned RLG generation process. Overall process of automatic layout fault detection is shown in Fig. 5.

- In order to detect layout faults, automatically at first RLG of input web page to be test is generated.
- After that pre-stored RLG of stable version of web page is compared with the RLG that is generated newly with the input web page.
- If no difference is experienced within the comparison result, then current version of web page is fault free in case of layout.
- But if there is difference in between the comparison result, then at first node that differs is located. After that error message representing the node is prepared and notified the user with test failure message. This type of error message noticeably pinpoints the error location.

V. RESULTS AND DISCUSSION

To evaluate the performance of proposed Responsive Layout Graph Based JavaScript CSS Defined Layout Fault detection Technique (RLGJCT), overall phenomena has been simulated over four responsively designed website mentioned in Table I. The result of simulation has been compared with the image based technique in terms of time and memory consumption. The simulation environment consists of 8GB RAM, 240 GB RAM, UBUNTU 14.04 LTS operating system, python selenium web driver, python tinycss parser, Python Imaging Library (PIL), BeautifulSoup and custom python based JavaScript Parser.

In order to test fault detection capability faults are inserted manually and later compared proposed technique with REDECHECK. From Fig. 6 it can be seen that proposed technique detects much more fault than the REDECHECK. From intersection of two curves it is also apparent that at some point both REDECHECK and proposed technique detects same amount of fault.

Table I. Web Page Chosen for evaluation.

Name	Page	Abbreviation (Used in this paper)	URL
Skinny Ties	Home Page	STH	http://skinnyties.com/
Bootstrap	Home Page	BOOTH	http://getbootstrap.com
Bootstrap	CSS Page	BOOTBCS	http://getbootstrap.com/css/
World Wild Life	Home Page	WWLFH	http://www.worldwildlife.org
Starbucks	Home Page	STARBH	http://www.starbucks.com

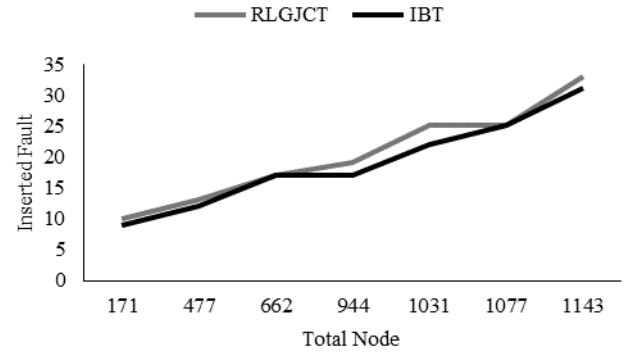


Fig. 6. Layout fault insertion with the increase in total HTML DOM node and detection by REDECHECK and proposed RLGJCT.

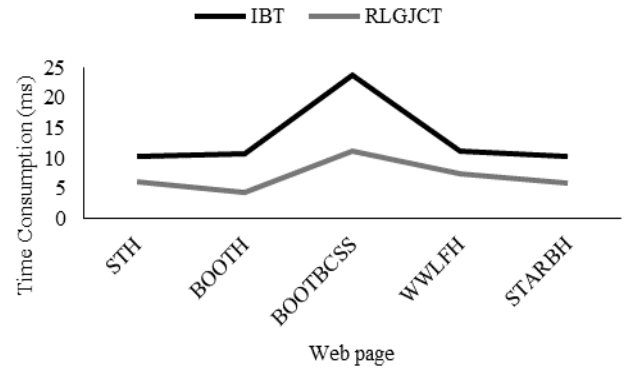


Fig.7. Time consumption by Image Based Techniques and proposed RLGJCT technique while testing the web pages mentioned in Table I.

Simulation has been conducted to compare proposed RLGJCT with the traditional Image Based Comparison technique (IBT). Fig. 7 suggests that proposed RLGJCT takes less time while testing the web pages mentioned in Table I than the IBT. IBT takes longer processing time due to the consideration of large set of pixel data. As they process pixel data they lack of pinpointing exact problem in the layout. Also RLGJCT uses a tree like structure that has time complexity $O(\log(n))$ whereas IBT considering each pixel has time complexity $O(n)$ at best case. Fig. 8 shows that IBT requires much more memory space than the proposed RLGJCT. While testing a web page all the traditional IBTs need to save one stable version of the web page as image and generate another test version image. Both require much storage space than the RLG generated by RLGJCT.

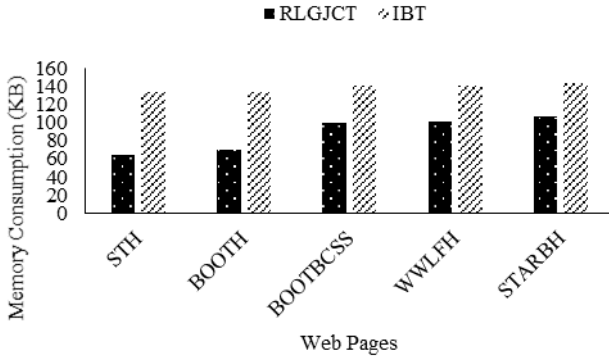


Fig. 8. Memory consumption by Image Based Techniques and proposed RLGJCT technique while testing the web pages mentioned in Table I.

VI. CONCLUSION AND FUTURE WORK

Automatic detection of layout faults in web page could save lots of manual effort needed for testing of web layouts. Layout faults triggered by JavaScript is hard to detect due to the dynamic nature of JavaScript. Proposed approach considers both JavaScript and CSS defined layout to aid in automatic detection of potential layout faults. Production of interactive error message following the proposed approach could pinpoint the exact error location.

Proposed approach considers only a few JavaScript functions and CSS attributes. One of the future research direction could be to extend the proposed approach by considering more JavaScript functions and CSS attributes that controls web layout. Consideration of stack order of web elements could be another future research direction.

REFERENCES

- [1] Koskimies, Oskari, et al. "EDB: a multi-master database for liquid multi-device software." Proceedings of the Second ACM International Conference on Mobile Software Engineering and Systems. IEEE Press, 2015.
- [2] Peng, Wenhui, and Yaling Zhou. "The Design and Research of Responsive Web Supporting Mobile Learning Devices." 2015 International Symposium on Educational Technology (ISET). IEEE, 2015.
- [3] Walsh, Thomas A., Phil McMin, and Gregory M. Kapfhammer. "Automatic Detection of Potential Layout Faults Following Changes to Responsive Web Pages (N)." Automated Software Engineering (ASE), 2015 30th IEEE/ACM International Conference on. IEEE, 2015.
- [4] Gibbs, Chris, and Ulrike Gretzel. "Drivers of responsive website design innovation by destination marketing organizations." Information and Communication Technologies in Tourism 2015. Springer International Publishing, 2015. 581-592.
- [5] Gardner, Brett S. "Responsive web design: Enriching the user experience." Sigma Journal: Inside the Digital Ecosystem 11.1 (2011): 13-19.
- [6] Lee, Jeonghyun, et al. "Responsive Web Design According to the Resolution." 2015 8th International Conference on u-and e-Service, Science and Technology (UNESST). IEEE, 2015.
- [7] Marcotte, Ethan. Responsive web design. No. 4. Editions Eyrolles, 2013.
- [8] A. Adamoli, D. Zaparanuks, M. Jovic, and M. Hauswirth, "Automated gui performance testing," Software Quality Journal, pp. 1-39, 2011.
- [9] Borjesson, Emil, and Robert Feldt. "Automated system testing using visual gui testing tools: A comparative study in industry." 2012 IEEE Fifth International Conference on Software Testing, Verification and Validation. IEEE, 2012.
- [10] A. Memon, "Gui testing: Pitfalls and process," IEEE Computer, vol. 35, no. 8, pp. 87-88, 2002.
- [11] S. Roy Choudhary, M. R. Prasad, and A. Orso, "X-PERT: Accurate identification of cross-browser issues in Web applications," in Proc. Of the 35th ICSE, 2013.
- [12] S. Mahajan and W. G. J. Halfond, "Finding HTML presentation failures using image comparison techniques," in Proc. of the 29th ASE, 2014.
- [13] Kersley, "Responsive design testing," accessed: 2016-06-12. [Online]. Available: <http://mattkersley.com/responsive/>

Detecting Jute Plant Disease Using Image Processing and Machine Learning

Zarreen Naowal Reza¹, Faiza Nuzhat², Nuzhat Ashraf Mahsa³, Md. Haider Ali⁴

Department of Computer Science and Engineering

BRAC University

Dhaka, Bangladesh

auroraprova@gmail.com¹, faiza.joyee@gmail.com², nuzhat.mahsa@gmail.com³, haider@bracu.ac.bd⁴

Abstract— Detecting stem diseases of plants by image analysis are still in an inchoate state in the research field. This research has been conducted on detecting the stem diseases of jute plants which is one of the most important cash crops in some of the Asian countries. An automated system based on an Android application has been implemented to take pictures of the disease affected stems of jute plants and send them to the dedicated server for assaying. On the server side, the affected portion from the image will be segmented using customized thresholding formula based on hue-based segmentation. The consequential feature values will be extracted from the segmented portion for texture analysis using color co-occurrence methodology. The extracted values will be compared with the sample values stored in the pre-defined database which will lead the disease to be identified and classified using Multi-SVM classifier. At the final step, the classification result along with the necessary control measures will be sent back to the farmer within three seconds through the application on their phone.

Keywords— *hue based segmentation, color co-occurrence methodology, texture analysis, Multi-SVM classifier, Android application*

I. INTRODUCTION

Plant disease detection using image processing techniques has been playing a role of immense importance in agricultural industries for the recent years. Jute is considered as the Golden Fiber of Bangladesh for its huge contribution to our economy in past though the production of jute has been hurdled by various factors from the recent years. One of the factors is the inaccessibility of expert's opinion to the root level cultivators or farmers. This system will help the farmers to get accurate solution in order to cultivate healthy crops. Although visual supervision by experts has been considered as the primary way of detecting crop diseases, the use of technology can alleviate the level of reliability as well as save valuable crop of the diligent cultivators by providing correct information in the fastest way.

This research will be beneficial to the farmers who struggle to get expert's guidance and suggestions to keep their crop safe from several fatal diseases. The mobile application interface of this system is very user friendly and people with no prior technological knowledge can use this system and get benefited by it. In this research, the five most common diseases have been detected followed by providing respective

control measures. The diseases are named as Anthracnose (*Colletotrichum Corchori*), Black band (*Botryodiplodia Theobromae*), Die back (*Glomerella Cingulate*), Stem rot (*Macrophomina Phaseolina*) and Soft rot (*Sclerotium Rolfsii*). By using various image processing techniques, the system can recognize these diseases from the image of the affected stem and provide users with disease identification and proper control measures within three seconds. Therefore, the system has every scope to serve the emergent field of plant disease detection by providing the classification of stem diseases within a short period of time with higher level of accuracy and a very simple user interface.

The rest of this paper is arranged as follows. Section II discusses related work. Section III explains the entire system overview. In Section IV the system implementation is discussed step by step. Section V demonstrates experimental results and analysis for detecting jute diseases. In section VI outcome of a survey conducted on farmers is presented and discussed. Section VII concludes this paper with a brief summary.

II. LITERATURE REVIEW

Though ample number of researches has been done on detecting plant diseases using image processing and computer-vision most of them have failed to detect the stem diseases efficiently. As example, S. Arivazhagan and his co-authors have discussed about leaf disease detection using image processing and neural network where they used the Color Co-occurrence Methodology for extracting the features for texture analysis [1]. For disease classification purpose, they have used multi-class SVM (Support Vector Machine) classifier with a winner-takes-all strategy where the classifier with the highest output function assigns the class [1]. Besides, referring to the integration with mobile applications Amos Gichamba and Ismail Ateya Lukandu have described different employments of mobile applications for agricultural purpose and presented a general model for designing such systems [2]. However, their focus of interest was detecting the leaf diseases of plants rather than the stem diseases. Another mobile-application based system named 'Beetles' has been presented by Rahat Yasir and Nova Ahmed which has been designed to support farmers in rural area to detect crop diseases from the image captured by a cell phone in real-time using histogram and color information of the image [3]. Having said that, their

system detects diseases based on color compromising the accuracy of results in case of diseases with same or almost same color. On the other hand, H. Al-Hiary and his co-authors have come up with the detection process of stem disease in such way that their system requires the background of the picture to be plain or unicolor [4]. Regrettably, as their segmentation method is based on masking the green pixels the efficiency has been compromised for most of the cases where stems turn brownish in color at matured stage [4]. To overcome this defect Y. Sanjana and her co-authors have proposed a system to detect different leaf diseases of crops using image analysis consisting of mathematical morphology for segmentation, texture, shape and color feature extraction for classification of diseases [5]. Nevertheless, their system needs the intervention of respective experts to give review and mention the level of severity by examining the extracted features which can increase the period of time needed for getting the ultimate result [5].

III. SYSTEM OVERVIEW

This system is designed in such way that the client does not require much technical knowledge to operate it. The instructions are pretty straight forward and will lead the user to the desired solution within short time. This automated system is mostly combined of press and send techniques.

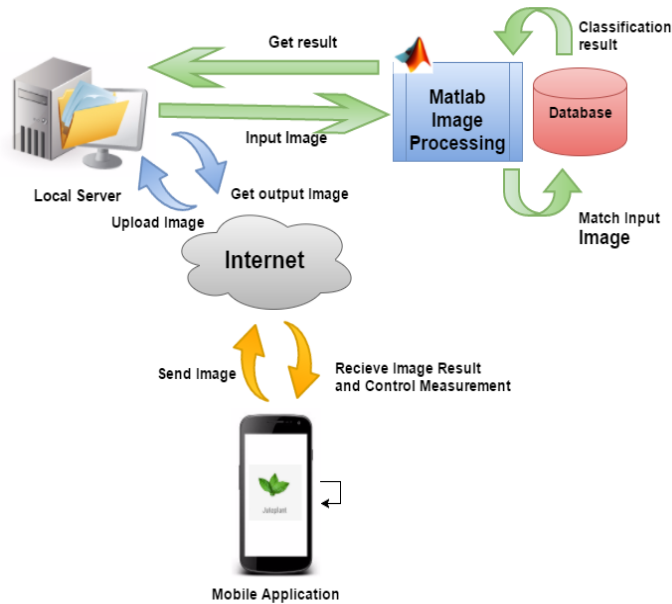


Fig. 1. System Overview

Farmers or users will take image using mobile phones and send it to server via mobile application. In server, the image will be analyzed based on visual and textural attributes using different image processing algorithms and the system will yield the disease name which will be sent to the mobile application on the client's end.

IV. SYSTEM IMPLEMENTATION

A. Mobile Application Development

Android studio with Android SDK tools (maximum level API 23 support) was used in creating and implementing the functionalities, this application supports all versions of Android operating system. This mobile application has four major functionalities explained below-

1) *Image capture/selection:* At first, after clicking the icon, the application shows a basic layout having one image view feature and two buttons. One button is for uploading the image to the server and the second one is for downloading the result from the server. At the very beginning, an intent ACTION_PICK is sent to get the image from external storage of the phone. Then that bitmap image is set on the image view.

2) *Upload image to the server:* The upload image button for uploading the image is pressed in order to upload the image to server. Image uploading is done with the help of AsyncTask class. Asynchronous task class is a special class which allows the application to perform background work and publish the result to UI threads without interrupting the UI thread or handlers. Asynchronous programming helps the app stay responsive when it does work in the background. Firstly the bitmap image is converted to string using the built-in method method. The bitmap image is firstly compressed into byte array output stream and then put into a byte array. Then this array is converted into base 64 string format. The image in string form is put into a Hashmap data structure which will be sent as data using Http URL connection. The connection between the application and the server is maintained by using HttpURLConnection class in java.

3) *Download the result from the server:* Once the image processing is completed and the disease is detected. The server sends back the result in form of string image and text string. After uploading is completed, the Download the result button is pressed and the layout is set to new page. Here also we use asynchronous programming for the same reason as we need the task to operate in background. The input stream is decoded from string to bitmap format. The next layout contains two buttons, one image view and one text view. The top button is for getting the segmented image which shows only the affected area of the plant and the bottom button is for navigating the user to solution needed for the disease. The image view shows the segmented image and the text view updates with the name of the disease the plant is diagnosed with.

4) *Getting solution:* After clicking the button, the user is directed to solution.xml layout where he or she can see the control measurement recommended by the specialist in Bangladesh Jute Research Institute (BJRI). The instructions are simple enough to understand and the occurrence period of the diseases is also included

B. Client and server interaction

The system will be executed in the server as soon as the client sends the image via mobile application. When the

mobile application sends data to the server, it is received using a PHP script using methods and the fetched images are stored in the assigned directory and our image processing will start on the assigned image automatically.

C. Shell execution

‘.exe’ file of the MATLAB code has been created and saved in the local server. From the PHP script with an function called exec(), this ‘.exe’ file of MATLAB is run through shell execution. The exec () method in the PHP is called just after receiving and saving the image in the server.

D. Parsing the result

Once the MATLAB code is executed through shell execution and the result is saved in the pre-assigned directory in our server, the PHP script fetches the data for the mobile application for request method type GET and sends it to our client. The result consists of two types of data. One is the segmented image which highlights the affected portion of the plant in Base 64 string form and the other part of the result is the disease name the MATLAB code generates after analyzing the image in string form. Both of them are fetched by PHP script and sent over the internet to our mobile application.

E. Server Side Operation

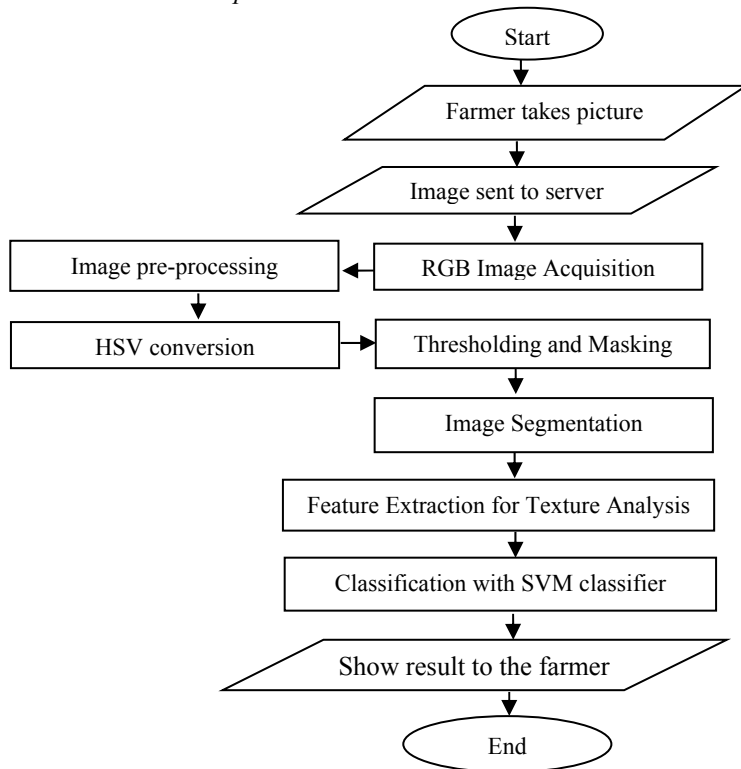


Fig. 2. Flowchart of overall process

The image received by the Android application has been sent to the system server where the image would be gone through several levels of analysis.

The stages from segmenting the image to identifying the disease are described as followed for Anthracnose.

1) *Image Preprocessing*: Before proceeding towards image analysis the image should be processed in order to acquiring better result. Images taken from camera phones contains different factors which alters the result of the analysis. The image preprocessing procedure is performed by following certain steps consists of image resizing, image enhancement and noise removal.

a) *Resize Image*: To perform the classification, the size of the input image must match the size of the images stored in the database. Thus, the input image must be resized to a fixed dimension at the very first stage.

b) *Enhance Image*: In this step the image intensity values or colormap has been enhanced by adjusting the contrast of the image by defining the upper and lower limits of the pixel values that will be used for stretching the image. The limits are specified by considering the bottom 1% and the top 1% of all pixel values of the image [6].



Fig. 3. A jute stem infected by anthracnose a) Original Image b) Enhanced Image

c) *Noise Removal*: The uploaded images may contain noise. Noise can alter the features of an image and leads to obtain unexpected and deceptive result. Thus it is very important to remove the noises from the image. In this research, a bilateral smoothing filter has been used for noise cancellation. The bilateral filter is the technique which smoothens the image by replacing the intensity value of each pixel with the weighted average of the intensity values of neighboring pixels [7].

2) *Hue-based segmentation*: In this paper, the hue-based segmentation method has been used to segment only the affected portion from the image. As the entire process has been conducted on stem diseases it is not possible to simply mask the green pixels from the image likewise the ones done in case of detecting leaf diseases. Therefore, the hue-based segmentation method has been applied along with a customized thresholding formula.

a) *HSV Conversion*: At first, the RGB spaced image has been converted to HSV color space. In HSV, hue image conveniently represents the original color whereas saturation works well to mask the image and extract the region of interest [8]. Then the individual channels have been extracted to separate the hue, saturation and intensity images.

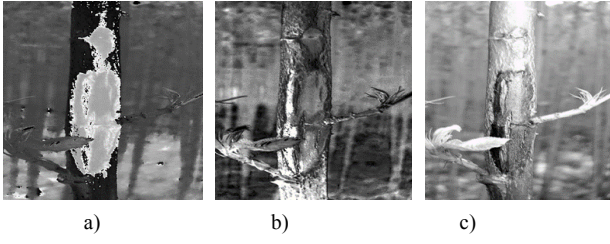


Fig. 4. a) Hue, b) Saturation and c) Intensity Image

b) Thresholding: Image thresholding is an effective way of dividing an image into a foreground and background. In this research, a threshold formula has been introduced to mask only the disease affected part of the stem from the image. At first, the saturation image has been manipulated with a theory described by R.C Gonzalez where a binary mask is generated by thresholding the saturation image with a threshold value which is equal to ten percent of the maximum value of the image [8]. Any pixel value greater than the threshold value is set to 1 (white) and all others are set to 0 (black) [8]. After that the masked saturation image has been multiplied with the hue image. Now at this point the region of the stem from the image should be separated. However, this approach solely does not accomplish the goal to extract only the disease affected portion of the stem and discard the rest. To achieve the goal, the product image has been masked again with a threshold value of 0.5 as the pixels with value greater than 0.5 prominently identify the disease affected region. Therefore, the comprehensive formula for thresholding appears as followed.

$$\text{MaskedImage} = M > 0.5$$

where $M = \text{Hue} * (\text{Saturation} > (\text{Max}(\text{Saturation}) * 10\%))$



Fig. 5. Masked Image

c) Blob Detection: The masked image still contains some unwanted regions which are definitely needed to be get rid of. For that purpose, the morphological analysis has been performed which includes erosion and dilation and for the largest connected component has been extracted using eight-way local neighborhood measurement.

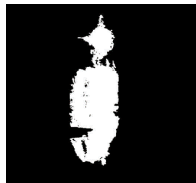


Fig. 6. Segmented Blob

d) RGB Conversion: For further analysis, the segmented portion needs to get back to its original color by being converted to RGB color space.

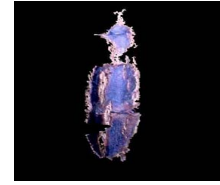


Fig. 7. Segmented RGB Image

3) Feature Extraction: The significant features from the image need to be extracted in order to perform the texture analysis. For that purpose, the color co-occurrence methodology has been used which is developed through the GLCM (Grey-level Co-occurrence Matrices). GLCM is used for sampling an image statistically in a way certain grey-levels occur in relation to other grey-levels [1]. It also provides feature information about the position of the pixels of the image in relation to their neighboring pixels. For this research, thirteen [9] feature values have been reckoned for each of the training images and the input image to perform texture analysis. The features are mentioned below:

1. Contrast
2. Correlation
3. Energy
4. Homogeneity
5. Mean
6. Standard Deviation
7. Entropy
8. RMS (root mean square) contrast
9. Variance
10. Smoothness
11. Kurtosis
12. Skewness
13. IDM (Image difference-measure)

The above features are calculated from the GLCM using their corresponding formulas.

4) Classification: After the extraction, the necessary features are compared with the pre-calculated dataset stored in a .mat file. The SVM (Support Vector Machine) classifier is used for classifying the disease. Support vector machines (SVMs) are a set of related supervised learning methods used for classification and regression [1]. Supervised learning refers to the process where a given set of labeled observations are stored in the database and an unknown sample has been analyzed and compared to assign it with a label in accordance. As there are five diseases to be dealt with there must have to be five classes. Thus a multi-class SVM classifier has been used. In this case, the classification of new instances or input images has been performed where the classifier with the highest output function assigns the class and detects the particular disease.

5) Create Database and Train Classifier: In this research, two .mat files have been generated, namely: (i) Training data, and (ii) Result.

The dimension the of training dataset can be defined as 150x13 where 150 is the number of rows and 13 is the number of columns. The 13 columns contain thirteen feature values of each of the training images. Similarly, result file contains 1 row with 150 columns each holding a digit from 1 to 5 representing the five diseases - Anthracnose, Black band, Die back, Soft rot and Stem rot respectively.

After that, a software program written in MATLAB loads .mat files representing the datasets, train the classifier and then use it to execute the classification task on the test data.

V. EXPERIMENTAL RESULTS AND OBSERVATION

For our research, we have examined with two different cases for extracting the features for texture analysis from the training images. In case 1, we have calculated the features from the training images without applying the hue-based segmentation method on them. In this case the diseased portion is not segmented from the image and the features are calculated along with the background, unwanted portions and noise of the images. As a result, when the segmented part of the test images is classified based on those feature values, it provides a dissatisfactory classification outcome. On the other hand, in case 2 at first the diseased portion is segmented from the training images after being preprocessed and then the features are extracted. The latter case succeeds in providing a satisfactory outcome in classifying the diseases.

From the observation of both cases, we found that case 1 provides us with 60 percent of accuracies in detecting the diseases whereas case 2 provides an accuracy of around 86 percent which is far better than case 1. Comparing both the accuracy results, we have decided to proceed with the method applied in case 2.

The comparison diagram between two cases based on detecting five diseases is shown in Fig.8. In the figure, for case 1 the percentage of detecting Anthracnose as Anthracnose is 62.50%, detecting Black Band as Black Band is 33.33%, detecting Die Back as Die Back is 60%, detecting Stem Rot as Stem Rot is 98% and detecting Soft Rot as Soft Rot is 50%. On the contrary, for case 2 the percentage of detecting Anthracnose as Anthracnose is 87.90%, detecting Black Band as Black Band is 66.66%, detecting Die Back as Die Back is 80%, detecting Stem Rot as Stem Rot is 100% and detecting Soft Rot as Soft Rot is 87.50%. Fig.9 displays the total amount of accuracy in percent based on detecting all the diseases as they are expected to be. In the figure, case I provides an accuracy level of 60% whereas case 2 performs with an accuracy of 86%.

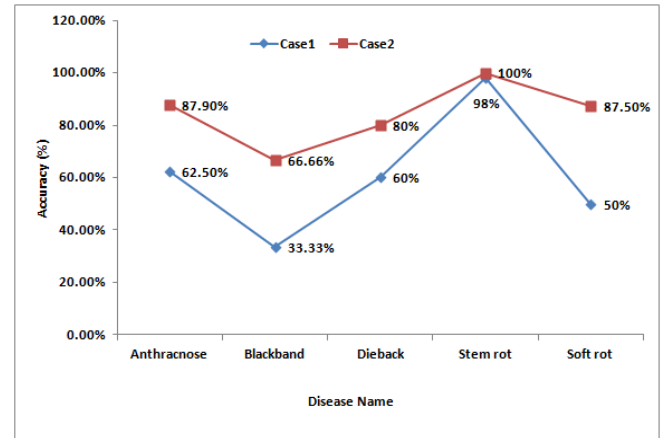


Fig. 8. True positive comparison between two cases

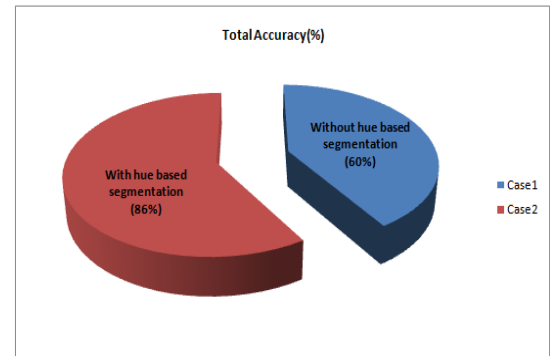


Fig. 9. True positive comparison between two cases based on total outcome

In both cases the following formula is used

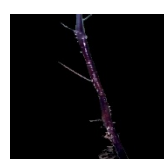
$$\text{Total percentage of accuracy} = (C \times 100) / N$$

Where, C = no. of correct outcome

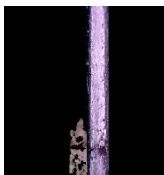
N = no. of test image



a) Soft Rot



b) Black Band



c) Die Back



d) Stem Rot

Fig. 10. the original images of a) soft rot, b) black band, c) die back and d) stem rot along with their segmented image



Fig. 11. a) select image from gallery b) upload image c) acquire result

VI. EVALUATION

We have conducted a short evaluation test on five farmers from BJRI (Bangladesh Jute Research Institute) where we asked them to use the application and rank it out of 5 based on the following criteria. The outcome is as follows.

TABLE I. FARMER FEEDBACK

User no.	Evaluation Criteria			
	User-friendly (0-5)	Preference over/along with manual inspection (0-5)	Bug/Error (0-5)	No. of correct outcome (5 images) test
1	3	2	4	4
2	4	4	4	4
3	3	3	3	3
4	2	2	4	4
5	4	4	5	5
Avg	3.2	3.0	4.0	4.0

The average outcome gained from the evaluation test provides an overall impression about the efficacy of the system where three out of five farmers preferred using this application to detect diseases along with manual inspection. Most importantly, the accuracy level obtained from the hands-on use (80%) almost matches the accuracy level measured from the test images (86%).

VII. CONCLUSION

In this paper, we have built an automated system to deduce stem oriented diseases for jute plants using image segmentation and feature extraction with along with potential machine learning. Application of machine learning specially image analysis and texture analysis in practical cases are now more common and encouraged than ever before. Although visual analysis done by human is simpler technique but it cannot be accessible always. On the other hand, providing the farmers with an automated and reliable system for crop disease detection to be used from their mobile phone can bring an insurgency for agricultural industry. The research work does not contain disease detection technique for the leaves. In future, it will be a great challenge for us to build a universal app that can be used to detect any sort of disease of the jute plants considering both stems and leaves.

ACKNOWLEDGMENT

We want to thank Dr. Md. Mahbubul Islam, Chief Scientific Officer and Dr. Saleh Mohammad Ashraful Haque, Senior Scientific Officer, Bangladesh Jute Research Institute (BJRI) for providing us with necessary dataset and suggestions.

REFERENCES

- [1] S. Arivazhagan, R. Newlin Shebiah, S. Ananthi, and S. Vishnu Varthini, "Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features," *Agric Eng Int: CIGR Journal*, vol. 15, no. 1, pp. 211–217, Feb. 2013. [Online]. Available: http://jp.mathworks.com/matlabcentral/answers/uploaded_files/25743/FinalAlgorithm.pdf. Accessed: Mar. 27, 2016.
- [2] Amos Gichamba, Ismail A. Lukandu "A Model for designing M-Agriculture Applications for Dairy Farming", *The African Journal of Information Systems*, Volume 4, Issue 4, 2012.
- [3] R. Yasir and N. Ahmed, "Beetles: A Mobile Application to Detect Crop Disease for Farmers in Rural Area," presented at the Workshop on Human And Technology (WHAT), Khulna, Bangladesh, Mar. 8, 2014, Academia, pp. 11–14. [Online]. Available: http://www.academia.edu/6370973/Beetles_A_Mobile_Application_to_Detect_Crop_Disease_for_Farmers_in_Rural_Area. Accessed: Mar. 21, 2016.
- [4] H. Al-Hiary, S. Bani Ahmad, M. Reyat, M. Braik, and Z. AlRahamneh, "Fast And Accurate Detection And Classification Of Plant Diseases," *International Journal Of Computer Applications (0975-8887)*, vol. 17, no. 1, pp. 31–38, Mar. 2011.
- [5] Y. Sanjana, AshwathSivasamy, and SriJayanth, "Plant Disease Detection Using Image Processing Techniques," *IJIRSET*, vol. 4, no. 6, pp. 295–301, May 2015.
- [6] Available at: <http://www.mathworks.com/> Accessed: Mar. 2, 2016
- [7] S. Paris, P. Kornprobst, J. Tumblin, and F. Durand, "A Gentle Introduction to Bilateral Filtering and its Applications,". Available at: https://people.csail.mit.edu/sparis/bf_course/course_notes.pdf Accessed: Mar. 19, 2016
- [8] R. C. Gonzalez, R. E. Woods, D. J. Czitrom, and S. Armitage, *Digital image processing*, 3rd ed. United States: Prentice Hall, 2007.
- [9] "Plant disease classification" Available at: <http://www.mathworks.com/matlabcentral/fileexchange/50624-plant-disease-classification/> Accessed: Mar. 5, 2016

Mobile Health Applications in Bangladesh

A State-of-The-Art

Md Mahboob Karim, Muhammad Nazrul Islam, Annita Tahsin Priyoti, Warda Ruheen, Nusrat Jahan,
Progga Laboni Pritu, Tanu Dewan, Zaki Tasnim Duti

Department of Computer Science and Engineering
Military Institute of Science and Technology (MIST)
Mirpur Cantonment, Dhaka – 1216, Bangladesh
E-mail: nazrulturku@gmail.com

Abstract— The social and economic development of a country is greatly influenced by providing effective health care services. In recent years, mobile health applications (mHealth apps) becoming an increasingly important means for delivering health services in developing countries like Bangladesh. This paper provides an overview of current status of mHealth apps development in Bangladesh. All mHealth apps developed for the users from Bangladesh were explored meticulously to synthesis their services/functionalities, and to depict their development platform and focused end-users. Finally, we outline the research and practical implications of our findings; that could increase our understanding of functional and development requirements of mHealth apps to provide effective health services in Bangladesh.

Keywords— mHealth; exploratory study; Bangladesh; health informatics; operating systems for mobile devices; and end users.

I. INTRODUCTION

Mobile Health (mHealth) is a component of electronic health (eHealth) and an emerging field in medical and public health. Over the last decade mHealth received important attention to provide effective health services. The World Health Organisation (WHO) defined mHealth as “medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants (PDAs), and other wireless devices” [1].

Providing effective health care service greatly impact the social and economic development of a country [2]. The mHealth field has emerged to support health services, which is considered as the basic requirement to improve the quality of life [3]. Applications developed for providing mobile health services (mHealth apps) include the use of mobile devices in real time health monitoring, collection of clinical and community health data, diseases prevention and detection, communicating between health-care professionals, basic diagnosis, and the like [4,5]. Now days, mHealth services are also becoming prevalent in the developing countries, like Bangladesh [2].

Bangladesh is a highly populated country. The current population in Bangladesh is about 163 million [6]. In Bangladesh, at the end of March 2016, the mobile phone subscribers were about 130 million [7] and internet subscribers were 61 million in Bangladesh [8]. Again, according to WHO, the recommended ratio of doctors to nurses to technologists is 1:3:5 but in Bangladesh the present ratio is 1:0.4:0.24 [9]. This clearly indicates that, Bangladesh has a staggering shortage of over 90,000 doctors, 273,000 nurses and 455,000 technologists [9]. Moreover, a large number of people in Bangladesh live in rural and remote areas, where healthcare facilities are frequently remote and inaccessible [2]. Though in Bangladesh, a number of initiatives relating to mHealth have been taken that includes, for example, established free teleconsultation, mobile application development, SMS services, sending Bangla voice messages, among others, but still there is much lacking within the health sectors in Bangladesh. Moreover, the health system of Bangladesh is haunted by challenges of increasing health coverage, access, and affordability [10]. Developing and implementing mHealth apps thus can play vital roles to overcome this limitation to a great extent in Bangladesh [11, 12].

Furthermore, previous research on health services in Bangladesh mainly strive to assess the opportunities, issues, and challenges of implementing ICT based health services in Bangladesh [13, 14]. No study has been conducted that explicitly focus on mobile health applications and depict the present state of mHealth applications, its uses and the development platform.

Therefore, the objective of this paper is to understand the present state of the art of mHealth applications in Bangladesh. The research question can be formulated as: ‘What is the current state of mHealth apps development in Bangladesh?’ To attain this research objective, we have conducted an exploratory study to analysis the all mobile Health applications (mHealth apps) that are preliminary developed for the users from Bangladesh with respect to its services/functionalities, development platform, and focused end users.

The remainder of this paper is organized as follows. Section II elaborates the related work. The study methodology is discussed in Section III. Section IV presents the study findings. The summary findings, the study implications for the research and practice, and the study limitations are briefly discussed in section V. A concluding remark and our idea for future work are discussed in the final section VI.

II. RELATED WORK

During the last decade, the study, development and implementation of mHealth services and applications become an essential point of attention in public health. Though a number studies have been conducted on mHealth services and reported the key contribution and importance of mHealth [4], but a limited number of studies have been conducted that are explicitly focused on mHealth services in Bangladesh. This section will briefly discuss the related articles on mHealth services in Bangladesh.

Ahmed et al. [13,15] highlighted the eHealth and mHealth initiatives taken in Bangladesh and assessed their true prospective. They found that tele-consultation, prescription, and referral are most common initiatives. They also suggested that trained health care professionals, capacity building, research support, and experience sharing are key demands for promoting and implementing health services in Bangladesh. In another study, Hoque et al [14] depicted the current status and highlighted the challenges of e-health services in Bangladesh. This study found that both government and private sector are committed to provide ICT based health service in Bangladesh, but these services are not well known to the most of the citizens (or service recipients) of Bangladesh. They also found that the overall status of e-Health in Bangladesh is good but not passable; thus they suggested that researcher, developer and government should work together to provide easily accessible and widely available ICT based health service in Bangladesh.

A readiness evaluation in some selected rural and urban hospitals and clinics of Bangladesh was carried out by Chowdhury & Jahan [16] to assess the potentialities, issues and challenges; as well as to recommend a set of guidelines for implementing mobile healthcare system in Bangladesh. Additionally in another study Jahan & Chowdhury [6] evaluated the opportunities, issues and threats of developing mHealth system in the developing world. They also proposed a potential mHealth model and a set of guidelines for implementing successful mHealth system in a developing country like Bangladesh.

Khatun et al. [17] interviewed 4,915 randomly-chosen respondents in a rural sub-district in Bangladesh to examine their knowledge, perception, use, cost and compliance with mHealth services. They found that 31% of the respondents were aware about mHealth services, while a few (mostly males, younger age group, better educated, and those from richer households) of the total respondents were aware of the available mHealth services in Bangladesh.

A few other studies have conducted focusing on specific users group as well as for specific disease. For examples, Hoque & Sorwar [18] derived a set of factors of influencing

the use and adoption of mHealth services by the elderly peoples in Bangladesh; that includes gender, experiences with mobile phone use, performance expectancy, effort expectancy, social influence, facilitating condition, hedonic motivation, price value, and habit. Prue et al. [19] found that mobile phone services significantly helps health works for rapidly detecting and treating patients with malaria in a remote area in south-eastern Bangladesh.

A review study has been conducted by Basar et al. [20] to synthesis the related articles and the existing mobile applications for the diabetic patients. As results, they reported the most common and important opportunities and features for developing mobile apps for the diabetic patients of the developing or the least developed countries like Bangladesh.

In sum, the previous work mainly focused on the challenges and opportunities of implementing ICT based health services in Bangladesh. A few papers were explicitly focusing on mobile health systems and one paper strives to depict the factors of adopting mHealth system by older people in Bangladesh. Apart from these, a few other studies have been conducted to understand the state of the art of mHealth systems in the world, for example, Boulos [21] examine the state of the art in mobile clinical and health related apps. However, based on this literature review and to the best our knowledge, no study has been conducted to present the state of the art of mHealth apps that are explicitly developed for the peoples from Bangladesh. This study will strive to fill this research gap to some extent by providing current status of mHealth apps in Bangladesh.

III. STUDY METHODOLOGY

An exploratory study was conducted to understand the state of the art of mHealth apps in Bangladesh. The study has been conducted in the following three sequential steps:

A. Selecting the Realted Apps

In this step, we strived to find out all mobile apps relating to the health care services in Bangladesh. Four major mobile platforms used in Bangladesh were considered in the apps selection process that includes Windows, Android, Blackberry, and iOS. Firstly, we searched all the popular and best apps stores [22] for each platform using a set of keywords. Apps stores *Google Play* for Android, *App Store* for iOS, *Windows Phone App Marketplace* for Windows, and *Blackberry App World* for Blackberry were searched using the search strings like 'Mobile health Bangladesh', 'Apps for healthcare Bangladesh', 'Fitness Bangladesh', 'Doctor Bangladesh', 'Health Tracker Bangladesh', 'Bangladesh Health Doctor', 'Healthy Bangladesh', 'Digital Health Bangladesh', 'Bangla Health Guide', 'Healthy Bangladesh Citizens', 'Pocket Medicine Bangladesh', and 'Bangladesh health care'.

Secondly, we meticulously check each of the apps (showed as a search result) to ensure whether the apps is for health care service and for the users from Bangladesh. In this phase, we carefully handle the all duplicate findings. The searching activities for selecting the related apps were performed during the period of March – April, 2016. Finally, he selected apps

and their platforms are listed along with their search keywords.

B. Extracting the Study Data

After completing the apps selection, we extracted data related to the services or functionalities provided by each app, target users of the selected apps, and the platform or operating systems for which the apps was developed.

C. Analyzing the Study Data

Finally, the study data were analyzed using the brainstorming and descriptive statistics. The apps were grouped in different clusters based on their services or functionalities in health care. For clustering, we have used brainstorming technique [23]. Authors in two groups are independently performed the job clustering in order to maintain the research rigor and to avoid the possible biasness. We compared the findings of two groups, and made some adjustment to finalize the clustering.

IV. STUDY FINDINGS

The study found a total of 180 mobile health applications that are preliminary developed for the users from Bangladesh. These apps were analyzed to understand the present status of mobile health applications in Bangladesh in three perspectives: health services covered by mobile application, platform used to develop the applications, and focused end users of these applications. The findings are briefly discussed below:

Clustering mHealth apps – We found a total of seven clusters (Health Care Information, Health/Medical Institute, Body Fitness, Doctors’ Information, Mother, Pregnancy & Child, Diseases Specific, and Nutrition) of apps based on their

services and /or functionalities and two clusters (i.e., Herbology and Homeopathic) of apps for alternative treatment systems (see table 1). Table 1 briefly introduces the clusters, and provides the example(s) name(s) of mHealth apps and the number of total apps found to each cluster. The results (see Figure 1) showed that maximum numbers of apps (36%) were developed to provide information relating to health care, while a minimum numbers (3%) of apps were developed for homeopathic treatment. An average percent (9-10%) of total apps were developed for body fitness, health/medical institutes, doctors’ information, herbology, and mother, pregnancy & child. A limited number of apps (6%) were developed for the disease specific health service.

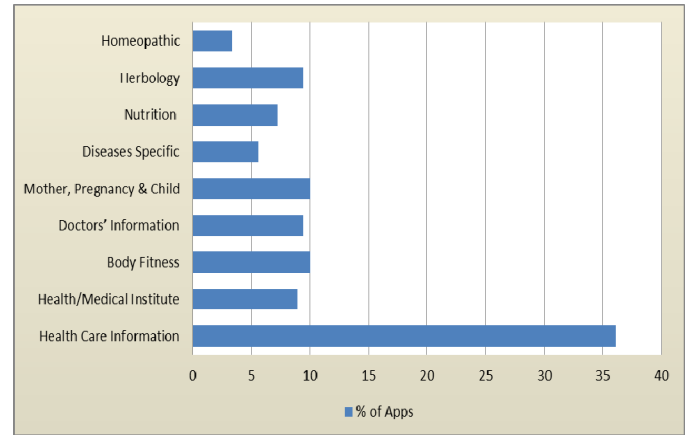


Fig. 1. Percent of apps belongs to different cluster

TABLE 1. CLUSTER OF MHEALTH APPLICATIONS

Sl	Cluster	Brief Description	Total no	Example of apps
1	Health Care Information	Apps belong to this cluster provide information related to health care and health care services in Bangladesh like first aid, blood finder, drugs, medicine, etc.	65	'Bangla Health Guide', 'MSডাক্তার' <MS Doctor>
2	Health/Medical Institute	Apps that provides important information related to a specific hospital & their services, hospital's phone number, timings, contact address, emergency hot line numbers, locations, and the like are belongs to this cluster.	16	'BD Hospital Information', 'BSMMU'
3	Body Fitness	Apps of this cluster provide assistance for physical exercise, weight loss, chest exercise, biceps exercise, triceps exercise, six pack abs exercise, etc. and information about gyms, gyms facilities, etc.	18	'Bangla Gym Guide', 'Gym Now: Gyms Clubs of Bangladesh'
4	Doctors' Information	Apps belong to this cluster provide assistance to doctors; and information about doctors in Bangladesh like their expertise, appointment procedure, contact information, hospital address, visiting time, and the like.	17	'DoctorsBD', 'Doctorola'
5	Mother, Pregnancy & Child	This cluster is created for those apps which are made on the purpose to support the women, pregnant women and their babies by giving all sorts of medical information and support.	18	'Mommy's Book Bangladesh', 'Baby Tika'
6	Diseases Specific	Apps belong to this cluster are made on the basis of a particular disease. People will find the details about that specific disease including the remedies and precautions.	10	'Diabetes Info in Bangla',
7	Nutrition	Apps of this cluster provide information related to the nutrition support for a robust health.	13	'ফলের উপকারিতা' <fruits benefits>
8	Herbology	All kinds of herbal treatment, herbal medicine and useful (herbal) information related apps are belongs to this cluster. These apps also provide all sorts of herbal remedies and their preventive and curative properties that enlighten peoples with some traditional methods and their qualities.	17	'ভেষজ চিকিৎসা' <herbal treatment>, 'Herbal Medicine Bangla'
9	Homeopathic	Apps of this cluster provide homeopathic treatment related information in Bangladesh context.	06	'Homeopathic Quick References'

Development platforms of mHealth apps – We found that mHealth applications in Bangladesh were developed almost for all the major mobile platforms including Android, iOS, Windows, and BlackBerry. For example, ‘Patient Aid’, ‘Omlooto’, ‘Health vault’, and ‘Homeopathic Guide’ were developed for Android, iOS, Windows, and BlackBerry phone, respectively. The results (see Table 2 and Figure 2) showed that about 33% of total apps were developed for multiple platforms; and 67% apps were developed for a unique platform. The results (see Figure 2 and Table 2) showed that more than two-third percent of all apps were Android based; and almost an equal percent of apps were developed for iOS, Windows, and BlackBerry platforms.

TABLE 2. APPS DEVELOPED IN DIFFERENT PLATFORMS

Platform	No of Apps (develop for a unique platform)	No of Apps (including the common apps across other platforms)
Android	101	149
iOS	10	42
Blackberry	3	43
Windows	7	41
For all Platforms		18 (10%)
For multiple (but not for all) platforms		41 (23%)

If we consider the unique platform (excluding the common apps across other platforms), the results showed that maximum number of apps was Android based and lowest number of apps was Blackberry based. If we considering the apps developed for a particular platform (including the common apps across other platforms), the results showed that almost an equal number of apps were developed for iOS, Windows, and Blackberry; and a maximum number of apps for Android. For example, ‘Bangla Health’ is developed for all platforms. However, the results are not surprising in context of Bangladesh. Since, in Bangladesh, most of the smartphone users use the Android phone; and comparing to the Android phone users a small percent of smartphone users use the iOS, Windows, and Blackberry phone [24].

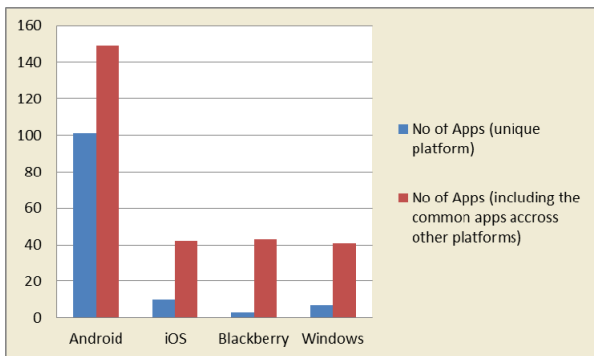


Fig. 2. Number of apps developed in different platforms

Focused Users of mHealth apps – We found that mHealth applications were developed in Bangladesh mainly for the healthcare professionals (e.g. doctors), mom & kids, patients, and other general users (e.g., health-conscious people). The results of our analysis found that about 5%, 13%, 18%, and 64% apps were developed considering the focused users as

healthcare professionals, mom & kids, patients, and other general users respectively (see Figure 3). These findings are consistent with the results of clustering mHealth apps. Since, most of the applications were developed for the health care information, which are useful to mainly the general users. A limited number of apps were developed for a specific disease (e.g., for diabetic patients) and for doctors.

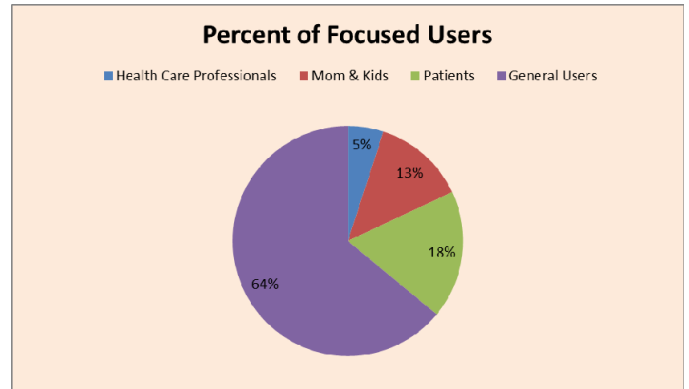


Fig. 3. Percent of focused users of mHealth Apps

V. DISCUSSION

A. Main Findings

The results indicated the extent of scarcity for a specific cluster of mHealth apps in context of Bangladesh. By seeing which cluster is lacking behind and which is not, it becomes easy to identify what types of mobile apps are necessary for the people of Bangladesh now. Again, a consistence was found between the mobile platform used for developing the mHealth apps and the type of mobile phone (platform) used by citizens of Bangladesh. For example, maximum number of apps was developed for Android phone and the maximum numbers of smartphone users in Bangladesh use the Android phone. The study also found that healthcare professionals, mom & kids, patients, and other general users are the focused end users for the mHealth apps in Bangladesh, while maximum apps focused on general users and limited number of apps focused on healthcare professionals.

B. Research Implications

This research has implication for future research and practice as well. For Bangladesh government, these results will provide an up-to-date view of current mHealth apps in Bangladesh that will help them to take further initiative in public health; e.g., taking necessary initiative to promote, use and adoption of mHealth services by the government health institutes and health workers. To information technology (IT) entrepreneur and Apps development organization, our outcomes will provide effective indication for developing new and innovative mHealth apps in Bangladesh. Private health institutes will also get a privilege to learn about the present situation of mHealth apps in Bangladesh, which in turn will help them to move forward. Finally, though a good number of mHealth apps were developed in Bangladesh but still to the best of our knowledge, most of the citizens of Bangladesh are not so aware about this service. Therefore further research

should be carried out on mHealth to understand the underlie phenomenon of mHealth apps usages in Bangladesh; to assess how the usability, accessibility, and interoperability issues are considered in mHealth apps; to investigate the challenges and opportunities of mHealth uses in Bangladesh; to assess the users trustiness on mHealth services in Bangladesh; and the like.

C. Study Limitations

The study presented in this paper has a few limitations that are important to acknowledge. Firstly, strings/keywords used for searching the relevant apps may not cover all the available mHealth applications. Secondly, since brainstorming (qualitative analysis) technique was used for analyzing the selected apps for clustering, and since qualitative analysis is subjective and partly depends on analyzers' skills, expertise, knowledge, therefore there might be some flaws in clustering analysis. Finally, the criteria chosen for selecting the relevant apps for this study might also fail to notice some other relevant mHealth application too. However, we tried to alleviate these limitations by employing a second check very meticulously for apps selection and apps clustering, and made some adjustment where necessary.

VI. CONCLUSIONS

In this paper, a systematic review was conducted on mHealth apps in Bangladesh. The results map out a clear picture of current status of mHealth apps in Bangladesh in terms of their services/functionalities; platform used by mHealth apps, and focused end users. The paper also highlighted the possible research implications in practice and research. Our future work will focus on the actual usages of mHealth apps in Bangladesh to provide underline theories for the design, development and adoption of mHealth services in Bangladesh to provide effective health services.

REFERENCES

- [1] WHO report, mHealth: New horizons for health through mobile technologies, available at www.who.int/goe/publications/goe_mhealth_web.pdf 2011
- [2] S. Jahan, and M. Mozammel Hoque Chowdhury. "mHealth: A Sustainable Healthcare Model for Developing World." American Journal of Modeling and Optimization 2, no. 3. pp.73-76, 2014.
- [3] J. A. Vergara, Y. M. Ramírez and J. E. Camargo, "A pervasive and ubiquitous mobile health application for tracking people with disabilities," in Proceeding sof the 10th Colombian Conference on Computing, Bogota, 2015, pp. 206-213.
- [4] B. M. C. Silva, J. J. P. C. Rodrigues, I. M. C. Lopes, T. M. F. Machado and L. Zhou, "A Novel Cooperation Strategy for Mobile Health Applications," in IEEE Journal on Selected Areas in Communications, vol. 31, no. 9, pp. 28-36, September 2013.
- [5] A.H. Rashidee, Emerging mobile health in Bangladesh, available at <http://www.thedailystar.net/news/emerging-mobile-health-in-bangladesh> (Accessed on March, 2016)
- [6] World Population Prospects: The 2015 Revision.(Medium-fertility variant). "worldometers". Elaboration of data by United Nations, Department of Economic and Social Affairs, Population Division. Available at <http://www.worldometers.info/world-population/bangladesh-population/> (Accessed on April, 2016)
- [7] Bangladesh Telecommunication Regulatory Commission (BTRC) report, Available at <http://www.btrc.gov.bd/telco/mobile> , April 30, 2016 (Accessed on May 2016)
- [8] Bangladesh Telecommunication Regulatory Commission (BTRC) Report, Available at <http://www.btrc.gov.bd/telco/internet> (accessed on May, 2016).
- [9] Health Bulletin (March, 2018), Ministry of Health and Family Welfare, Directorate general Of health Service DGHS, Available at <http://www.dghs.gov.bd/index.php/en/publications/health-bulletin/dghs-health-bulletin> (Accessed on April, 2016).
- [10] The socio-economic impact of mHealth, Report by Boston Consulting Group and Telenor group. Available at <https://www.telenor.com/media/in-focus/the-socio-economic-impact-of-mhealth/> (Accessed on May, 2016).
- [11] S. Akter, J. D'Ambra, and P. Ray, "User perceived service quality of mhealth services in developing countries: Research paper," in European Conf. Inf. Syst. (ECIS 2010), Pretoria, South Africa, 2010.
- [12] D. Vatsalan et al., "Mobile technologies for enhancing ehealth solutions in developing countries," in Second International Conf. eHealth, Telemedicine, Social Medicine (eTELEMED), St. Maarten, Netherlands Antiles, 10-16 Feb. 2010, pp. 84-89
- [13] T. Ahmed, G. Bloom, M. Iqbal, H. Lucas, S. Rasheed, L. Waldman, A.S. Khan, R. Islam, R. and A. Bhuiya. "E-health and M-health in Bangladesh: Opportunities and Challenges", IDS Evidence Report 60, March, 2014.
- [14] M.R. Hoque, M.F.A. Mazmum and Y. Bao. "e-Health in Bangladesh: Current Status, Challenges, and Future Direction", Int Tech Manag Rev 4(2), pp. 87-96, 2014
- [15] T. Ahmed, H. Lucas, A. Sadun Khan, R. Islam, A. Bhuiya and M. Iqbal. "eHealth and mHealth initiatives in Bangladesh: A scoping study", BMC Health Services Research, Vol. 14, p.260, 2014.
- [16] H.Chowdhury, and S. Jahan. "Applicability of mHealth for Healthcare Management in Developing Countries: A Study in Bangladesh", International Journal of Bio-Science and Bio-Technology 6(4), pp.113-122, 2014.
- [17] F. Khatun , S. Hanifi, M. Iqbal, S. Rashee, M.S. Rahman, T. Ahmed, et al. "Prospects of mHealth Services in Bangladesh: Recent Evidence from Chakaria", PLoS ONE 9(11), 2014.
- [18] M. R. Hoque and G. Sorwar. "Factors Influencing mHealth Acceptance among Elderly People in Bangladesh", Australasian Conference on Information Systems, 2015, pp.1-11.
- [19] C. S. Prue, K. L. Shannon, J. Khyang, L. J. Edwards, S. Ahmed, M. Ram, T. Shields, and M. S. Hossain, G. E. Glass, M. M. Nyunt, D. A. Sack, D. J. Sullivanjr, and W. A. Khan. "Mobile phones improve case detection and management of malaria in rural Bangladesh", Malaria Journal, 12 (1), pp. 1-7, 2013.
- [20] M. A. Basar and H. N. Alvi and G. N. Bokul and M. S. Khan and F. Anowar and M. N. Huda and K. A. A. Mamun. "A review on diabetes patient lifestyle management using mobile application". In proceeding of the 18th International Conference on Computer and Information Technology, pp. 379-385, Dhaka, Bangladesh, December, 2015.
- [21] N. Maged, K. Boullos, A. C. Brewer, C. Karimkhani, D. B. Buller, R. P. Dellavalle. "Mobile medical and health apps: state of the art, concerns, regulatory control and certification", Online J Public Health Inform,5(3):229, 2014.
- [22] R. Trenholm, Best app store: Android, Apple, Windows or BlackBerry?, Available at <http://www.cnet.com/news/best-app-store-android-apple-windows-or-blackberry/> (Accessed on June, 2016)
- [23] A. Oslapas, "Beyond brainstorming: Creative problem-solving techniques," Frontiers in Education Conference, 1993. Twenty-Third Annual Conference. 'Engineering Education: Renewing America's Technology', Proceedings., Washington, DC, 1993, pp. 837.
- [24] Z. Amin. "Market Insight: The Smartphone Industry in Bangladesh". Light castle Blog, Available at <http://www.lightcastlebd.com/blog/2015/01/market-insight-market-smartphones-bangladesh> (accessed on March, 2016).

A Framework for Detecting Arsenic Disease

Md. Ariful Islam and Mohammad Shamsul Arefin
Department of Computer Science and Engineering
Chittagong University of Engineering and Technology
Chittagong, Bangladesh
E-mail: arifnislam@gmail.com and sarefin_406@yahoo.com,

Abstract— Humans are exposed to arsenic (As) primarily from air, food and water. Anyone can develop arsenic toxicity as a result of arsenic exposure. Most of the reports of chronic as well as toxicity in human focus attention on skin manifestations because of its diagnostic specificity. A dermal manifestation such as hyperpigmentation is diagnostic of chronic arsenicosis. This paper focuses on designing and modelling a system that will collate Pigmented Skin Lesion (PSL), their analysis, corresponding observations and conclusions by medical experts using prototyping methodology. The system uses computational intelligence technique to analyse, process, and classify the image library data based on texture and morphological features of the images. A user in a remote location can use mobile data acquisition devices (such as smartphone) to generate images of PSL, supply such images as input to the proposed system, which in turns should intelligently be able to specify the arsenic poisoned hyperpigmentation status of the imaged PSL. The system has been evaluated with a set of images and found that the system can almost accurately detect arsenic.

Keywords— Arsenic disease, medical imaging, automated diagnosis, skin disease.

I. INTRODUCTION

Arsenic is toxic substance to human health and toxicity depends on the amount of arsenic intake. Arsenicosis is a dangerous disease which may cause cancer. It can be prevented if arsenicosis detected in its early stage. This study can be described as a technological tool that can detect arsenicosis. The naked eye observation of experts is the main approach adopted in practice for detection.

The importance of such a framework is to detect arsenic disease in human body which let people aware. Arsenic disease generally develops in rural area because of drinking ground water directly. Rural people have an inherent apathy to go to doctor. Chronic arsenicosis may develop cancer and certain death. This framework has been implemented in smartphone. Since, smartphones i.e. android, is popular in rural area, people can easily check for arsenic poisoning and take necessary step to cure it. An early detection can save a human life.

Although there are many systems for detecting arsenicosis detection, most of them are based on manual process and needs trained medical personnel and none of them consider digital detection system. Considering this face, in this paper, we develop a system that can detect arsenicosis using technological support like digital image processing.

We decompose our arsenicosis detection task into three parts. These are proper segmentation, feature extraction and selection, and lesion recognition to identify arsenic hyperpigmentation. Lab test result of arsenic concentration in urine, hair and nail may require for more accurate result. This framework may help a human to take medicine and aware

him before it goes to riskiest level and can reduce the possibilities of cancer, partial paralysis or blindness etc.

The remainder of this paper is organized as follows. Section II provides a brief review of related work. In section III, we detail the computation framework of our proposed approach. Section IV presents the experimental results. Finally, we conclude and sketch future research directions in Section V.

II. RELATED WORK

Detecting dermatological 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 and detecting them with a high accuracy rate is the primary purpose of this work.

Mazumder [1] describes the symptoms of arsenic disease i.e. hyperpigmentation. He shows the computational characteristics of pigments. Tushabe et al. [3] proposed an image-based diagnosis method where images of skin disorder were used to classify skin diseases into broad category of either viral infected or bacterial infected. Yasir et al. [4] used different types of image processing algorithms for feature extraction and feed forward artificial neural network for training and testing purpose. They proposed a system where extracted 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. Arifin et al. [5] proposed an automated dermatological diagnostic system. They have used different pre-processing algorithms and used feed forward back propagation artificial neural networks for training and testing purpose. In the last decade, usage of Dermoscopy or Epiluminiscence Light Microscopy (ELM) changed the dermatologist's approach to suspicious PSL. However, the analysis using ELM are extremely complex and subjective [6]. To prevent aforementioned challenge of quantitative interpretation, methods based on Computer-Aided Diagnosis (CAD) have been introduced towards automating the diagnosis procedures [6, 7].

Wahab et al. [8] presented a texture classification system for three skin diseases. Texture analysis techniques they used are: grey level co-occurrence matrices, wavelet analysis, Laws' texture energy measurements, grey level gradient distribution and statistical moments (mean, variance, kurtosis, skewness). Nachbar et al. [9] proposed a methodology for diagnosis of doubtful melanocytic skin lesions. The used ABCD (Asymmetry, Border irregularity, Colour variance, Differential structure) rule of dermatoscopy.

While the ABCD rule has become a standard used by many dermatologists, its diagnostic use by dermatologist based on visual and qualitative evaluation of such criteria are often subjective. Mazumder [1] says, arsenic disease

pigments are symmetrically distributed and have 2-4 mm diameter.

III. FRAMEWORK FOR ARSENICOSIS DETECTION: SYSTEM ARCHITECTURE AND DESIGN

The methodology of this work is based on soft science design for the diagnosis of arsenic disease represented by a skin image. Fig. 1 shows the overall systematic steps of the proposed framework for arsenic disease diagnosis. Each important step of the framework is thereafter discussed.

We are typically following a procedure that involves four main steps in processing the PSL:

- Image Formation and Pre-processing
- Segmentation
- Feature Extraction
- Diagnosis Accuracy Assessment

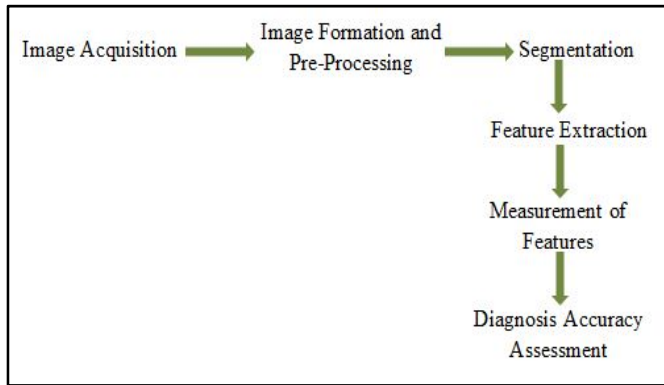


Fig.1 System architecture of the framework

After the acquisition and pre-processing of colour images of the PSL, we intend performing segmentation of the lesion. The lesion areas and boundaries would be clearly identified, and various attributes of the lesion characteristics would be measured.

A. Image Formation and Pre-processing

Images to be used would be standardized into a chosen resolution. It is expected that some of the images to be used might have features such as not hand. These features are typically regarded as noise, and thus need to be filtered off in order to facilitate separation of the lesion area from the surrounding skin. A skin detection algorithm by Tabassum, M. R. et al. [10] is used to crop the hand area from the image.

B. Segmentation Process

The segmentation is one of the most important stages for analysing image properly since it affects the accuracy of the subsequent steps. However, proper segmentation is difficult because of the great varieties of the lesion shapes, sizes, and colours along with different skin types and textures. In addition, some lesions have irregular boundaries and in some cases there is smooth transition between the lesion and the skin. To address this problem, several algorithms have been proposed. They can be broadly classified as thresholding, edge-based or region-based methods. In this framework we used contour based image segmentation algorithm using morphological edge detection by Hsiao, Y. T et al. [11].

C. Feature Extraction

This framework would characterize the ABCD rule into quantitative attributes measured by image analysis, and implore texture analysis technique. ABCD represent the asymmetry, border structure, colour variation, and

dermatoscopical structure so called diameter of the lesion and define the basis for a diagnosis by a dermatologist. We would characterize symmetry instead of asymmetry since arsenic pigments are symmetrically distributed, described by Mazumdar D.G. [1].

First feature extraction is the lesion symmetry. The procedure used hereafter starts with finding major axis of the lesion mask L applying the minimum bounding box method. Then L is flipped around the major axis to find its symmetry mask, S_L . Denote by $A = L \cup S_L$. One may tell that A includes L and non-overlapping regions of S_L . These non-overlapping regions are called false symmetry, FS . The symmetry is calculated by:

$$Sym = 1 - FS/A$$

and takes values between 0 and 1. The higher the value, the more symmetric is the lesion.

Second is border irregularity. We used compact index and edge abruptness. Density index (compact index/CI) is the measurement of the most popular form of barrier which estimates unanimous 2D objects. This can be determined by using the following equation

$$CI = \frac{P_L^2}{4\pi A_L}$$

Where, P_L = Perimeter of the Lesion. A_L = Area of the Lesion.

Lesion with irregular boundaries (abruptness edge) has a large difference in radial distance. Barrier irregularity is estimated by analysing the distribution of radial distance difference:

$$C_r = \frac{\frac{1}{P_L} \sum_{P \in C} d_2((P, G_L) - m_d)^2}{m_d^2}$$

Where, m_d is the mean distance of d_2 between the centred point and the barrier G_L .

Third is colour variation. One early sign of hyperpigmentation is the emergence of colour variations in colour. The descriptors of colour are mainly statistical parameters calculated from different colour channels, like average value and standard deviation of the RGB or HSV colour channel. In this framework, colour variance of the RGB image has been calculated using HSV channel. To calculate colour variance, first step is to convert the RGB colour map into an HSV colour map. The column of the output matrix represents hue, saturation, and value respectively. In second step, the colour variance of the HSV colour map has been calculated which represent the actual colour variance for an image.

Fourth is diameter. The raindrop appearance results from the presence of numerous rounded hyperpigmented, typically 2-4 mm in diameter. To find the diameter of the lesion we calculated equivalence diameter of same area circle.

We have measured the threshold of extracted features using OpenCV library. Threshold parameters are shown in TABLE-I.

TABLE I
THRESHOLD PARAMETERS

Symmetry	Compact Index	Edge Abruptness	Color Variance	Diameter	No of Pigments
> 0.75	2.0 - 2.5	0.02 - 0.10	0.10 - 0.15	2 - 4	> 45

IV. SYSTEM IMPLEMENTATION AND EXPERIMENTS

In this section, we provide the implementation procedure and performance analysis of our developed system.

A. Experimental Setup

The Framework for Detecting Arsenic Disease has been developed on a machine having the windows 8.1, 2.50GHz, Core i5-3210M processor with 4GB RAM. The system has been developed for Android OS. Minimum required SDK 16 (Android 4.1). Java is used in the front-end. OpencvCV-2.4.9 for android is used to implement various computer vision algorithms. For coding we have used Android Studio 1.2.1.1.

B. Implementation

The system architecture following illustrates the internal and external structure of system modules integrated together in one package to form one system. The following subsections provide a brief background overview of the tools used and the implementation details of the different modules of the developed system. The whole system was developed for android mobile operating system platforms because it is open source.



Fig. 2 Selected image from gallery

Fig. 2 shows the detection screen of our system where an image is chosen from gallery. Users can also take a new picture using camera. Fig. 3 shows the binary masked image of the selected image.



Fig. 3 Binary masked image

Binary masked image is used to crop skin area from the image. This step will help to eliminate unwanted pixels from the image which will increase the performance and reduce error rate. Fig. 4 shows the cropped hand area of the image.

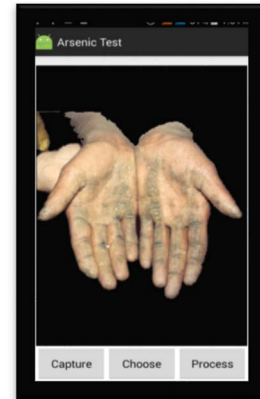


Fig. 4 cropped skin area

Segmentation process is shown in the next step. Contour based morphological edges are detected. In this step, we get the entire segmented lesions. Fig. 5 shows the segmented lesion image.



Fig. 5 Segmented lesions

For each segmented pigment we calculate the property like, symmetry, border irregularity, colour variance and diameter of the pigment. A lesion is suspected if its feature value fall within the range and result shown based on number of suspected lesion. Fig. 6 shows the result screen of the arsenic disease detection test.

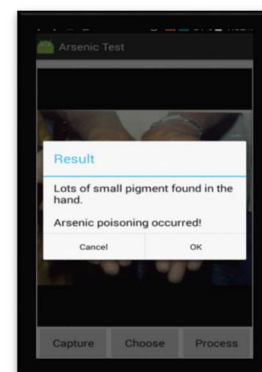


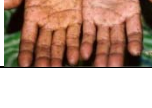
Fig. 6 Result screen

C. Experimental Results

In our experiment, we selected two categories of images, i.e. true positive, true negative. An image which contains arsenic caused hyperpigmentation is true positive and an image which does not contain arsenic caused hyperpigmentation is true negative. We used 19 true positive and 12 true negative images. TABLE II shows the results of our system about detection of arsenic in each of these images.

TABLE II
RESULT OF THE SYSTEM

Test No.	Image	Actual Arsenic Poisoning?	Detected Arsenic Poisoning by the System
01		Yes	Yes
02		Yes	Yes
03		Yes	Yes
04		No	No
05		Yes	Yes
06		Yes	Yes
07		Yes	Yes
08		No	Yes
09		No	No
10		No	No
11		Yes	Yes
12		Yes	No
13		Yes	Yes
14		No	No

15		Yes	Yes
16		No	No
17		Yes	Yes
18		Yes	Yes
19		No	No
20		No	No
21		No	No
22		No	No
23		Yes	Yes
24		No	No
25		Yes	Yes
26		Yes	No
27		Yes	Yes
28		No	Yes
29		Yes	Yes
30		Yes	Yes
31		Yes	No

From the result of TABLE II, we can find that among 19 true positive images, our system detects 16 images having arsenic and among 12 true negative images, our system accurately identifies that 10 images has no arsenic. So, we can see that our system misclassified three true positive images and two true negative images.

The results of our experiments are summarized in TABLE III.

TABLE III
SUMMARY OF THE EXPERIMENTS

Type of Image	Number of Images	Number of Accurately Detected Images	Detection Rate (%)
True Positive	19	16	84.21
True Negative	12	10	83.33

Considering the results of TABLE III, we computed the performance of our system using following formula.

$$\text{Accuracy} = \frac{\text{True positive accuracy} + \text{True Negative accuracy}}{2}$$

Using above formula we found that the accuracy of our system is $\frac{84.21+83.33}{2} = 83.77\%$. We also computed the performance of the system in terms of its precision and recall using following formulas.

$$\text{precision} = \frac{|\{\text{relevant images}\} \cap \{\text{retrieved images}\}|}{|\{\text{retrieved images}\}|}$$

and

$$\text{recall} = \frac{|\{\text{relevant images}\} \cap \{\text{retrieved images}\}|}{|\{\text{relevant images}\}|}$$

As we stated earlier that we used total 31 images, where 19 were arsenic poisoned patient hands. Our developed system identified 18 images as arsenic poisoned hand, where 16 were correct. So, precision is $16/18 = 0.88$ and recall is $16/19 = 0.84$.

V. CONCLUSION

Our framework provides a detection system of arsenicosis. It can be used in android supported devices such as smart phones, tablet computers etc. In this framework, detection system is done by analysing visible symptoms such as melanosis hyperpigmentation in hands. Segmentation of pigments is done by contour based morphological edge detection. Pigments are analysed by the features like, symmetry, regularity, colour variance etc. Due to the lack of sufficient images, this time we could not conduct the experiments on a large number of images. Next time we plan to conduct experiments on large number of images that will give more insight about the accuracy of the system. In addition, we plan to adapt related pathological tests results in the system so that we can consider both images and pathological tests results to improve the accuracy of the system.

REFERENCES

[1] D. N. G. Mazumder, "Chronic arsenic toxicity & human health", Indian Journal of Medical Research, vol. 128, no. 4, pp. 436-447, 2008.

[2] D. N. G. Mazumder, "Diagnosis and treatment of chronic arsenic poisoning", United Nations Synthesis Report on Arsenic in Drinking Water, 2000.

[3] F. Tushabe, F. Mwebaze, and F. N. Kiwanuka, "An image-based diagnosis of virus ad bacterial skin infections", Proceedings of ICCIR, 2011.

[4] R. Yasir, M. A. Rahman, and N. Ahmed, "Dermatological disease detection using image processing and artificial neural network", In Proc. of International Conference on Electrical and Computer Engineering (ICECE), pp. 687-690, 2014.

[5] M. S. Arifin, M. G. Kibria, A. Firoze, M. A. Amin, and H. Yan, "Dermatological disease diagnosis using color-skin images", In ICMLC, pp. 1675-1680, 2012.

[6] P. Rubegni, G. Cevenini, M. Burrioni, R. Perotti, G. Dell'Eva, P. Sbrano, C. Miracco, P. Luzzi, P. Tosi, P. Barbini, and L. Andreassi, "Automated diagnosis of pigmented skin lesions". International Journal of Cancer, vol. 101, no. 6, pp.576-580, 2002.

[7] I. Stanganelli, A. Brucale, L. Calori, R. Gori, A. Lovato, S. Magi, B. Kopf, R. Bacchilega, V. Rapisarda, A. Testori, A. Ascierto, E. Simeone, and M. Ferri, "Computer-based diagnosis of melanocytic lesions", In Anticancer Research, vol. 25, pp.4577-4582, 2005.

[8] N. A. Wahab, M. A. Wahed, and A. S. Mohamed, "Texture features neural classifier of some skin diseases". In Proc. of IEEE 46th Midwest Symposium on Circuits and Systems, vol. 1, pp. 380-382, 2003.

[9] F. Nachbar, W. Stolz, T. Merkle, A. B. Cognetta, T. Vogt, M. Landthaler, and G. Plewig, "The ABCD rule of dermatoscopy: high prospective value in the diagnosis of doubtful melanocytic skin lesions". Journal of the American Academy of Dermatology, vol. 30, no. 4, pp. 551-559, 1994.

[10] M. R. Tabassum, A. U. Gias, M. Kamal, H. M. Muctadir, M. Ibrahim, A. K. Shakir, and M. Islam, "Comparative study of statistical skin detection algorithms for sub-continental human images". Information Technology Journal, vol. 9, issue. 4, pp.811-817, 2010.

[11] Y. T. Hsiao, C. L. Chuang, J. A. Jiang, and C. C. Chien, "A contour based image segmentation algorithm using morphological edge detection". In Proc. of IEEE International Conference, Systems, Man and Cybernetics, vol. 3, pp. 2962-2967, 2005.

Feature Extraction and Classification of EEG signal for Different Brain Control machine

Sheikh Md. Rabiul Islam¹, Ahosanullah Sajol¹, Xu Huang², and Keng Liang Ou³

¹Dept. of Electronics and Communication Engineering

Khulna University of Engineering & Technology, Bangladesh.

²Faculty of ESTeM, University of Canberra, Australia,

³College of Oral Medicine, Taipei Medical, University, Taiwan,

E-mail:robi@ece.kuet.ac.bd; sajol1993@gmail.com; Xu.Huang@canberra.edu.au, klou@tmu.edu.tw

Abstract—Brain computer interface is used for human and machine learning analysis. This paper represents the EEG datasets that are built with different cognitive task such as left, right, back and front imaginary movement with eye open. We have used different feature extraction method to classify these EEG signal using Support Vector Machine (SVM), k-Nearest Neighbor (k-NN) and Artificial Neural Network (ANN). All these methods are compared with other work that have done with other datasets. The proposed work is obtained 95.21% accuracy 98.95% sensitivity for SVM and k-NN is 90.88% and ANN is 94.31%. The performance results have shown higher enough than all others.

Keywords—EEG datasets, SVM, ANN, k-NN, Accuracy.

I. INTRODUCTION

Electroencephalography (EEG) is a neural information processing technique for copying the brain's electrical potentials now a days. In medical diagnosis, EEG can be able to diagnose the brain activities by Brain-computer interfaces (BCI). It can translate the brain activities by a set of command of computer software or Neuro prosthesis [1]. In [1], developed an algorithm on spectral features and used neural network and probabilistic neural network (PNN) classifiers on the six basic emotions such as happiness, sadness, fear, anger, disgust and surprise. In [2] EEG database it has been developed for four emotional states and also classifiers used to classify of four states by K-NN algorithm, Multilayer Perceptron and SVM. They have also used for statistical features in time domain and frequency domain. The authors in [3] have developed a new filtering procedures before feature extraction and also developed a new features extraction methods for six different emotional states. They have higher classification rate upto 85.17 % by applying Genetic Algorithms.

The EEG datasets are developed from a patients with different electrodes from scalped [4]. Therefore, the researcher have grown up about evaluation and analysis of cognitive process. They realize how to process EEG data and classify it from EEG recorded data [5], [6]. The value of features are collected from recorded EEG dataset. This feature values are used in performance of EEG signal classification [5]. The EEG classification has always been impact on the performance based on inadequate feature extractions methodology. In this paper, we have tried to develop an EEG dataset from Biomedical Signal Processing laboratory. In this work the raw EEG signals have been collected using biopack data acquisition unit MP36 and classified them into left, right, back and front movement. In

our proposed framework, the raw EEG datasets are extracted of feature in time domain and frequency domain using Acknowledge 4.1 software. The coefficients are used as features for classification of EEG datasets. These features are training to support vector machine (SVM), k-nearest neighbor (kNN) and artificial neural network classifiers (ANN) separately for performance analysis of proposed framework.

This paper is organized as follows. A brief description on data processing of EEG signals based on our proposed framework is given in section II. In section III, obtained results are analyzed and consequent issues are discussed. Finally conclusions are described in section IV.

II. METHODOLOGY

We have described the details of proposed framework as shown in Figure 1 in this section. This frame work has raised some problems. How to collect EEG database? What should be the way of experimental setup to collect EEG signals? How could it be the possible way of feature extraction of it? After feature extraction, how to classify EEG datasets based on cognitive task. All these problems, we have tried to explain below.

A. Data acquisition of EEG signals

Data acquisition is a technique by sampling signals. This sample values are converted into digital numeric value by computer interface between software and machine. The dataset is recorded in this work and it is obtained from the Department of Biomedical Engineering, Khulna University of Engineering and Technology, Bangladesh. This dataset is recorded from a normal subject. Each subject sits in a relaxing chair with armrests. The way of cognitive tasks gives instruction to subject like left, right, back and front movements. The data was acquiesced in channel 1 with electrode and three leads black white and red using Biopack data acquisition unit MP36R. The MP36R have a four-channel data acquisition unit designed to work with Acknowledge 4.1 software for the research market. In this work, EEG dataset are taken from 15 human for different imaginary movement. They were graduate students and aged between 20 and 24 in Khulna University of Engineering and Technology in this work. The main condition of experimental cognitive task was that all the participants were free from any medication, drugs, neurological disorder and normal eye vision.

B. Data Analysis

Those acquired raw EEG datasets are processed using biopack acknowledge 4.1 software. First of all FIR band pass filter was used for reducing noise. Also hanning window technique was applied for reducing the discontinuity and better smoothness of the signal for better analysis. We have done features extraction and classification of them. These are described in section C and D.

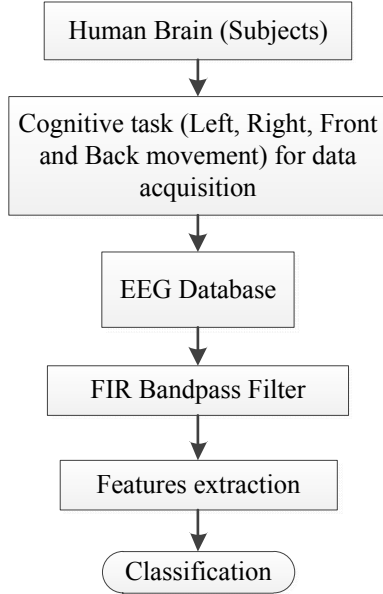


Fig.1 Flow chart of the proposed methodology of BCI.

C. Feature Extraction

Feature extraction is either time domain or frequency domain for EEG signal. In most of the cases we have used different features in bio potential signals, this is because of the characteristics of the EEG signal. However, there are different frequency band of EEG signals, such as alpha beta, delta, and gamma. We have work out of our EEG dataset based on time domain and frequency domain features extraction directly from the signal. The important time domain features such as, maximum value, mean value, standard deviation, skewness, kurtosis etc., were extracted from raw EEG datasets. These are described below:

i. Mean Value

The range of EEG potential is microvolts with time. The mean of EEG is constant and small values which change potentially. Definition of mean as described based on EEG signal is given by:

$$\mu_x = \frac{1}{N} \sum_{i=1}^N x_i$$

where x is EEG signal and N is total number EEG data.

ii. Standard deviation:

Standard deviation is a measurement to find the amount of variation or dispersion from the average. To determine standard

deviation of EEG signal, we have considered low and high variation. The standard deviation can be expressed as [7],

$$std = \sqrt{\frac{\sum (x - \mu_x)^2}{N - 1}}$$

iii. Skewness:

It is a measurement of the asymmetry of the probability distribution of an EEG signal and it's mean. The mean value can be positive, negative or undefined. The mean values can be zero based on completely symmetrical distribution and some nonzero values are asymmetrically distribution in respect of the baseline [7].

$$Skew = \frac{\sum_{i=1}^N (x_i - \mu_x)^3}{(n - 1)\sigma_x^3}$$

where μ_x is the mean of x , σ is the variance of x .

iv. Kurtosis:

It's the measurement of the peakness of the probability distribution of EEG signal. The signal contains transient spikes, isolated high-voltage wave group. The kurtosis of it presents high positive values or negative values. These values are observed when EEG with high or low frequency and amplitude modulation is analyzed. The moment co-efficient can be written as [7],

$$Kurt = \frac{\sum_{i=1}^N (x_i - \mu_x)^4}{(n - 1)\sigma_x^4}$$

D. Classifiers

Classification is an important task of EEG signal. It can operate based on the feature extraction. The features values are an input and it can predict the class for classifier technique. A classifier has a number of parameters that need to be learned from training data. The learned classifier is a model of the association between the features and the classes. For example, an input feature q of a class p , the model of classifier function $p = f(q)$ can predict new instances. These instances are not used in training data [7]. Those extracted data from EEG datasets were then classified using linear and nonlinear classifiers model for the analysis of better classifier for EEG data classification. These are described below in brief.

i. Support Vector Machine

Machine learning is learning process from source of data. There are many methods for learning structure. We just peak up one of the most prominent method as Support Vector Machines. This method is a group of supervised learning machines based on statistical learning theory which can be used for pattern recognition and regression [9] and the Vapnik-Chervonenkis (VC) dimension. For example, an input set is p and output set is q so the set of training as $\{(p_1, q_1), (p_2, q_2) \dots (p_m, q_m)\}$, the function of model of classifier: $q = f(p, \alpha)$, where α is the parameter of the function. The training dates have also a risk for classification which is also estimated by empirical risk and

true risk formula [9] on raw data's. We have used this method on recorded EEG datasets.

ii. Artificial Neural Network

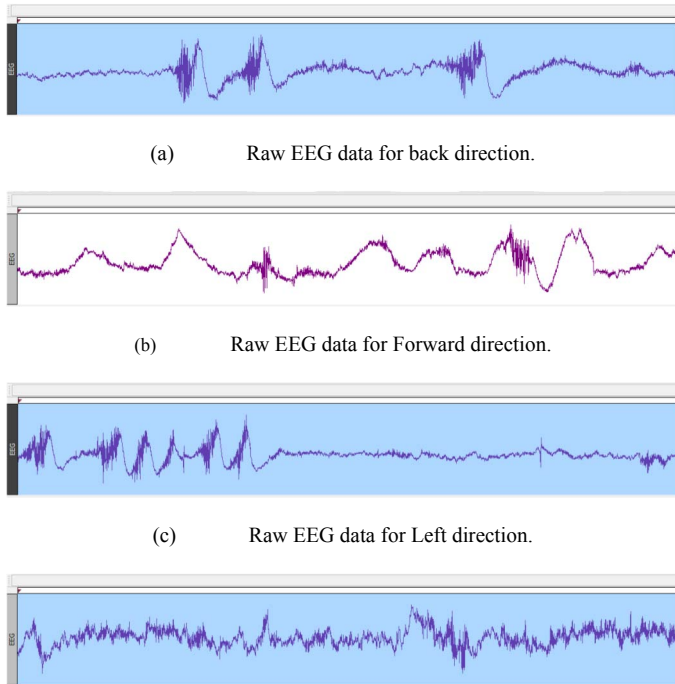
In artificial neural network input dates are trained and retrained for the suitable target value. In this work backpropagation algorithm is used for classification. In backpropagation algorithm and contains single hidden layer which reduce the data training time. Also the backpropagation algorithm reduces error by backpropagation from the output layer, until the ANN learns the training data [10]. Here 70% datasets are used for training, 15% validation and 15% testing.

iii. K - Nearest Neighbor Algorithm

In pattern recognition, K-nearest neighbor (k-NN) [11] is a non-parametric or linear method used classification and regression of EEG datasets as the third classifier in this research work. Different feature values are fed to the k-NN and the classification accuracy has been determined. The output of class by an object is classified by a majority vote of its neighbors. The object is assigned by its k nearest neighbors. If $k = 1$, then the object is locating the nearest neighbour in instance space and labelling the unknown instance with the same class label as that of the located (known) neighbour. A higher value of k results in a smoother, less locally sensitive, function

II. RESULTS AND DISCUSSIONS

In this experiment, we acquired EEG datasets form 15 subjects. For the brain control machine, we just setup four different conditions i.e., left, right, backward, forward have to collect EEG signals from the subjects. Those data are being processed through AcqKnowledge 4.1 software. For example four different conditions for only one subject's raw EEG data is shown in Figure 2.



(d) Raw EEG data for Right direction.

Fig. 2. Raw EEG signal from 1 subjects with different condition.

The brain activity of subject is measured by EEG machines and also records electrical activities from brain. This was converter from brain signal to electrical signal in microvolts. In this record, created artifacts by which generated from the patient body movement and, from outside the body (i.e., equipment, environment). These artifacts are filtered using the FIR band pass filter within the range of 0.2Hz to 48 Hz and also Blackman window. Filtered EEG data is showing in Figure 3.

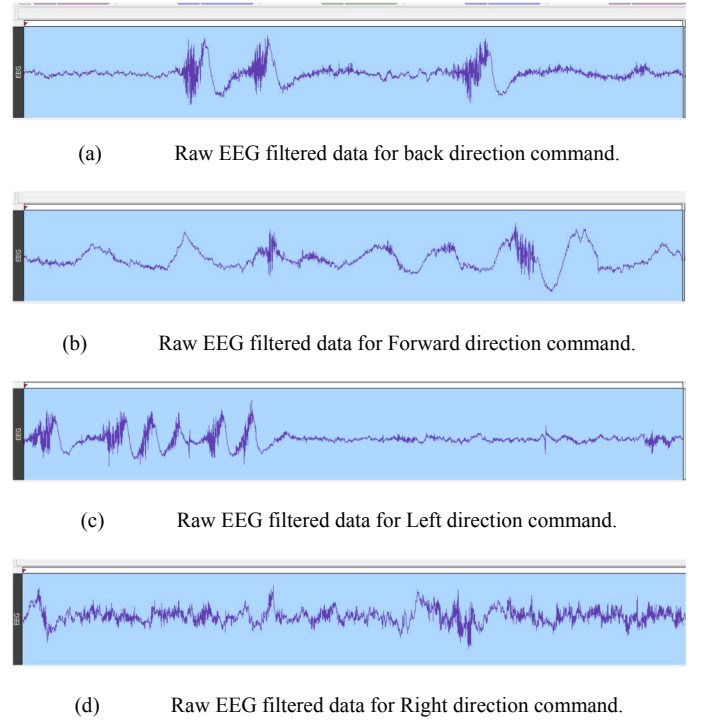


Fig.3. Filtered Raw EEG signal from 1 subjects with different condition.

For the classification results, we have used three different classifiers such as Artificial Neural Network (ANN) and Support Vector Machine (SVM) and K-nearest neighbor (k-NN) with $k = 3$ for EEG conditions, i.e., left, right, front, and back imaginary movement with eyes open. In case of ANN classification, classification is done using the whole EEG signal as well as with its individual frequency bands. However k-NN and SVM classification only EEG classification is done both time domain and frequency domain features. For ANN classification the features which are used for classification are maximum value, mean value, standard deviation, skewness, kurtosis in time domain and maximum frequency and maximum values in frequency domain are in the ranges as shown in Table IV. For SVM classification the features used for classification are maximum value, mean value, standard deviation, skewness, kurtosis in time domain and maximum value for alpha band, maximum value for beta band and maximum value for delta band are used as shown in Table IV.

The performance parameters for classifiers have used accuracy, sensitivity, specificity, precision, and Kappa statistic [7], [12]. These parameters are defined as follows:

$$\text{Accuracy} = \frac{\text{Total no. of correctly classified instances}}{\text{Total numbers of instances}} \times 100$$

$$\text{Sensitivity} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \times 100$$

$$\text{Specificity} = \frac{\text{True Negative}}{\text{True Negative} + \text{False Positive}} \times 100$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \times 100$$

$$\text{Kappa}(k) = \frac{(P_o - P_e^C)}{(1 - P_e^C)}$$

where P_o represents the probability of output between the classifier and the true process, and P_e^C denotes sum of the proportion of instances in a class multiplies the proportion of true process of that class in the dataset.

Table I. Classification results of different cognitive task.

Meth od	Cogniti ve task of subject s (No. of subject s =15)	Accura cy (%)	Sensitiv ity (%)	Specific ity (%)	Precisi on (%)	Kapp a statist ic
SVM	Left	95.21	98.95	97.45	96.95	0.97
	Right	95.21	98.95	97.45	96.95	0.97
	Front	95.21	98.95	97.45	96.95	0.97
	Back	95.21	98.95	97.45	96.95	0.97
k-NN	Left	90.88	95.77	94.66	93.32	0.94
	Right	90.88	95.77	94.66	93.32	0.94
	Front	90.88	95.77	94.66	93.32	0.94
	Back	90.88	95.77	94.66	93.32	0.94
ANN	Left	90.88	95.77	95.66	94.42	0.94
	Right	100	100	98.45	98.55	0.98
	Front	93.33	97.59	96.95	95.72	0.95
	Back	93.33	98.59	96.95	95.72	0.95

Table II. Overall Classification results of cognitive task.

Metho d	Accurac y (%)	Sensitivit y (%)	Specificit y (%)	Precisio n (%)	Kappa statisti c
------------	------------------	---------------------	---------------------	-------------------	------------------------

SVM	95.21	98.95	97.45	96.95	0.97
k-NN	90.88	95.77	94.66	93.32	0.94
ANN	94.39	97.74	97	95.82	0.95

The results in Table I of classifiers performance parameters for different features sets such as accuracy, sensitivity, specificity, precision, Kappa statistic have shown worthy values based on raw EEG datasets. The classifier of SVM have acquired the same values of classifier parameter for different cognitive task such as 95.21% accuracy, 98.95% sensitivity, 97.45% specificity, 96.95% precision, and 0.97 Kappa. However the classifier of k-NN have achieved the same values of classifier parameter such as 90.88% accuracy, 95.77% sensitivity, 94.66% specificity, 93.32% precision, and 0.94 Kappa. In ANN, 100% accuracy, 100% sensitivity for right movement. The overall classification of all three classifiers is shown in Table II. The SVM classifiers achieve 95.21 % accuracy, K-NN classifiers achieve 90.88 % accuracy and ANN classifiers achieve 94.39% in the classification using five different features methods and four different cognitive tasks. In Table I and II, we observe that SVM classifiers have higher accuracy, sensitivity, specificity, precision, Kappa statistic value. For the case of k-NN, if we give k=1 or k=2, its accuracy gives higher than SVM and ANN.

Many comparison results with the previous research work on EEG signals have not shown directly because of the variety of EEG datasets, features extraction types, subject availability, and the cognitive tasks used. We have presented a brief comparison of our proposed frame work and other existing research work. In Table III, we have tried to give the comparison based on the dataset, feature extraction methods, machine learning algorithm and the classification performance. Most of the researchers have used ANN and kernel-based SVM). The two classifiers are basically nonlinear and k-NN are linear classification model. The authors have done on low number instances in the classification because they feel over fitting problem of data's in classification [7], [12]. In this reaserch case, the experimental task acquired 60 instances for each class in the classification on 15 subjects. This is not high instances as compared with reference [7]. However this is not problem in classification using EEG datasets. Hence, both the linear (e.g.,K-NN) as well as the non-linear classifiers (e.g., SVM, ANN) along with 70% data are used for training, 15% validation and 15% testing used. In Table I, II and III, both linear (i.e.,k-NN) and non-linear (SVM, ANN) classifiers were found to be better than other classifiers that have done is previous work. More specifically in our EEG datasets, SVM have shown higher accuracy, sensitivity, specificity, precision, Kappa statistics than kNN and ANN. However, the proposed framework has shown higher results of performance parameters than others in Table III.

Table III. Comparison results of existing feature extraction, classifications and performance techniques for EEG dataset.

Ref.	Subject s	EEG scalp	Feature	Classifier	Accurac y (%)
------	--------------	--------------	---------	------------	------------------

		electrode s			
[13]	7	6	AR coefficient s	SVM,AN N, ELM	54-56
[14]	7	6	Frequency band power and AR coefficient s	SVM with RBF kernel	70
[11]	4	6	DWT with db4	K-NN	81.48- 89.58
[13]	3	8	Power spectral density feature	ANN, SVM	69.90- 68-35
[15]	7	6	Immune feature	SVM	85.4- 97.5
[16]	X	X	Frequency band power	ANN	88.89
[7]	8(560)	128	Wavelet relative energy	SVM, MLP, K-NN Naïve Bayes	88.89 98.75 98.21 83.57
Propose d work	15(60)	3	Maximum ,Mean, Standard deviation, Skewness, Kurtosis	SVM K-NN ANN	95.21 90.88 94.39

Table IV. EEG frequency band and their extraction region.

Band	Frequency (Hz)	Locations
Delta	<4	Frontal in adults, posteriorly in frontal in adults, posteriorly waves.
Theta	4-7	Found in locations not related to task at hand
Alpha	8-15	Both sides posterior regions of head, generally on dominant side there is higher amplitude. Central sites at rest
Beta	16-31	Both sides on head symmetrically distributed most evident frontally; low- amplitude waves
Gamma	32+	Somatosensory cortex

III. CONCLUSION

In this work, we presented frame work to collect EEG datasets using software Biopackwith interface MP36 unit, using Acknowledgement 4.1 software for filtering, smoothing and extracting different time domain features and frequency domain features with different band. Three classifiers were used for classifying the cognitive task in different direction EEG data. The nonlinear SVM classifier achieved the higher sensitivity and accuracy of the classifiers obtained as 95.21% accuracy, 98.95% sensitivity, 97.45% specificity, 96.95% precision, and 0.97 Kappa statistical approach respectively than artificial neural network and k- nearest neighbor. Some error occurred while classifying because they didn't identify any artifacts with use of only 3 leads and number of sample data limits for collecting the EEG signal. Those problems decrease the accuracy of these classifiers. We think this proposed approach is useful for clinical applications.

REFERENCES

- [1] Saadat.Nasehi and HosseinPourghassem, "An optimal EEG-based emotion recognition algorithm using gabor features," WSEAS Transactions on Signal Processing, Vol. 8, No. 3, July 2012, pp. 87-99.
- [2] Xiao-Wwi Wang, DanNie, and Bao-Liang Lu, "EEG-Based Emotion Recognition Using Frequency Domain Features and Support Vector Machines", Neural Information Processing, Volume 7062 of the series Lecture Notes in Computer Science pp 734-743, 2011.
- [3] Panagiotis C. Petrantonakis, Leontios J. Hadjileontiadis "Emotion Recognition from Brain Signals using Hybrid Adaptive Filtering and Higher Order Crossings Analysis", IEEE Transactions on Affective Computing, Vol. 1, No. 2, Aug.2010, pp. 81-97.
- [4] Ramaswamy Palaniappan, "Biological Signal Analysis", BoBoCoAe, RP & Ventus Publishing, 2010, ISBN: 8776815943 9788776815943.
- [5] Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM. "Brain- computer interfaces for communication and control," Clinical Neurophysiology, vol. 113, Issue 6, June 2002, pp. 767-791
- [6] M. A. Hassan, A.F. Ali, M. I. Eladawy, "Classification of the Imagination of the Left and Right hand Movements using EEG" 2008 Cairo International Biomedical Engineering Conference ,Dec. 2008, pp. 1-5.
- [7] HafeezUllah Amin, Aamir Saeed Malik, Rana Fayyaz Ahmad, Nasreen Badruddin, NidalKamel, Muhammad Hussain, Weng-Tink Chooi "Feature extraction and classification for EEG signals using wavelet transform and machine learning techniques" Australasian Physical & Engineering Sciences in Medicine, March 2015, Volume 38, Issue 1, pp 139-149.
- [8] Islam, S.M.R., Huang, X., Le, Kim: " A novel image quality index for image quality assessment" Neural Information Processing, Vol. No. 8228 in Lecture Notes in Computer Science, Springer Berlin Heidelberg (2013), pp. 549-556..
- [9] InanGuler, ElifDeryaUbeyli., "Multiclass support vector machines for EEG-signals classification" IEEE Transactions on Information Technology in Biomedicine, Vol. 11, Issue: 2 , Mar. 2007, pp. 117-26.
- [10] Lin C-J, Hsieh M-H (2009) Classification of mental task from EEG data using neural networks based on particle swarm optimization. Neurocomputing 72:1121–1130
- [11] Yazdani, Ashkan; Hoffmann, Ulrich; Ebrahimi, Touradj, "Classification of EEG signals using Dempster Shafer theory and a k-nearest neighbor classifier", The 4th International IEEE EMBS Conference on Neural Engineering, April 2009, pp. 327-330.

- [12] AbdulhamitSubasi, M. Ismail Gursoy, "EEG signal classification using PCA, ICA, LDA and support vector machines" Expert Systems with Applications, Vol. 37, Issuee 12, Dec. 2010, pp. 8659–8666.
- [13] Liang NY., Saratchandran P., Huang GB., Sundararajan N. "Classification of mental tasks from EEG signals using extreme e-learning machine". International Journal of Neural System, Vol. 16, Issue 1, Feb. 2006, pp. 29–38.
- [14] Sarah M. Hosni,Mahmoud E. Gadallah ,Sayed F. Bahgat,Mohamed SAbdelWahab, "Classification of EEG signals using different feature extraction techniques for mental-task BCI", International conference on computer engineering & systems, 2007. ICCES'07, Nov. 2007, pp 220–226.
- [15] Lei Guo , Youxi Wu ; Lei Zhao ; Ting Cao ; Weili Yan ; Xueqin Shen," Classification of mental task from EEG signals using immune feature weighted support vector machines. IEEE Transaction on Magnetics,Vol. 47 Issue 5,May 2011 pp. 866–869,.
- [16] Jahidin A, Ali MM, Taib MN, Tahir NM, Yassin IM, Lias S (2014) "Classification of intelligence quotient via brainwave subband power ratio features and artificial neural network", Computer. Methods and Programs Biomedicine, Vol.114, Issue 1, April 2014, pp. 50–59.

Evaluation of Geographic Locations for River Bridge Construction: An Approach for Multiple Criteria Decision Analysis with Evidential Reasoning

Nasrin Sultana

Department of Computer Science & Engineering
University of Chittagong, Bangladesh
Email: nasrin_cse@cu.ac.bd

Risul Islam Rasel

Department of Computer Science & Engineering
International Islamic University Chittagong, Bangladesh
Email: risul-islam@cse.iuic.ac.bd

Abstract—Geographic location selection for river bridge construction is one of the most challenging tasks because it has to face with statistical data and qualitative information with improbability. Organizational assessment, safety assurance, risk measurement, contractor appraisal and many other things depend on a proper Geo-location selection. River bridge site selection is such an example in which multiple attributes; both qualitative and quantitative criteria have to be analyzed for selecting a proper bridge location. This paper signifies the use of a novel technique called Evidential Reasoning (ER), is a Multi Criteria Decision Analysis method applied for decision making where many criteria have to be explicitly considered for making decision. ER approach is competent for focusing the improbability of multi-criteria decision problem, where both subjective and quantitative natures are exists. It manages uncertainties using a degree belief formation. It also helps in aggregating degrees of belief from lower level to higher level attributes. To conduct the experiments of this work, Geo-location selection for The Padma Bridge construction in Bangladesh is taken as research domain and the experiment results and analysis are discussed in this paper.

Keywords— *Assessment, Evidential Reasoning, Multi Criteria Decision Analysis, Numerical data and Qualitative information, Uncertainty, Degree of belief*

I. INTRODUCTION

Rivers are natural flow of water. But sometimes it mires our usual flow of movements. Transportations and freight across the river by ferries, launches or manually-operated boats are time consuming and tedious process. But a bridge can eradicate that persecution. However selecting the most suitable site for bridge construction is a critical problem. Large numbers of criteria have to be considered and evaluated. Moreover, many criterions are correlated to each other in a very intricate way. Consequently, step up in one often consequences in refuse to another. Besides some criteria are qualitative (river stream, River training Works, Bank Erosion etc.) and some are quantitative (Bridge length, Traffic demand, Land Acquisition, estimated cost). So decision-makers have to scrutinize the both quantitative data and Quantitative data to make a decision. Quantitative data can be measured but the qualitative data is descriptive in nature only. As an example, nature of bank erosion can't be measured. It can only be described as the river bank erosion is favorable, unfavorable,

no erosion (excellent) or severe erosion (worst). As a result, qualitative data occupies uncertainty as it is hard to measure the concepts or distinctiveness or superiority of a thing with 100% conviction. So, the qualitative criteria which are concerned in selecting proper site for bridge construction face lots of uncertainties should be deal with apt method.

There are many techniques to select suitable geographical site for river bridge construction. GIS and Remote sensing approaches have been using to evaluate suitable site for river bridge construction based on river bank erosion and deposition [1]. That approaches handle only two criteria but selecting suitable and sustainable geo-location is a multi-criteria problem. AHP method also able to handle multiple criteria decision making problem [2],[3]. Those approaches used a pair wise evaluation matrix for identifying the impact between two features or data [4],[5]. But it can not address the uncertainty such as incompleteness or vagueness. Moreover, when another alternative will add, it can be perceived that the grade of the alternatives will be altered in terms of their significance. These sort of problems allied with AHP generate severe problems in decision taking. If a belief structure is used for evaluating the grades of attributes or features, the incompleteness may be focused and therefore the vagueness will also being addressed [6]. The issues are mentioned in prior can be handled also by applying Evidential Reasoning Approach (ER), which is a multiple criteria decision analysis (MCDA) method [7],[8]. This approach was introduced based on decision theory, which is particularly known as utility theory [9],[10] and artificial intelligence, which is particularly known as theory of evidence [6][11]. This approach deals with MCDA problems, containing both quantitative and qualitative criterions having different unpredictability, like as imperfect information, vagueness, ambiguity etc. This technique has handled uncertainty by suggesting a belief structure assigning a degree of belief in different assessment grades of the attributes, that is absent in AHP or any other multiple criterions decision analysis techniques. For the management facility of ER approach in managing and addressing uncertainty, it has been applied to numerous research areas, especially in environmental impact measurement [12], pipeline leak detection [13], and system reliability prediction [14]. It also has been applied to carry out safety and threat analysis [9]. In this research ER approach is applied to identify sustainable geographical location for river bridge construction.

II. METHODOLOGY

A. Evidential Reasoning

Evidential Reasoning Approach (ER) had been introduced by Yang and Singh for multi-criteria decision and design problem analysis under uncertainty [9],[15]. It is a multiple criteria decision analysis (MCDA) technique for dealing with decision and design problems in decision making environment having both qualitative and quantitative criteria with different uncertainties; like as deficient statistics and vagueness. It has been applied for propping up different decision investigation, opinion judgments and assessment of actions [7][8]. The evidential reasoning approach was introduced with the help of decision theory (utility theory) and artificial intelligence (theory of evidence). It applies a belief structure to form a decision having ambiguity, a belief decision matrix to handle MCDA problems with uncertainty, ER algorithm to combine criteria for producing distributed evaluations, and the idea of belief and creditability functions is applied to produce a utility space for assessing the level of ignorance. A conventional decision matrix was also introduced to model an MCDA problem.

B. Multi-criteria decision analysis (MCDA)

Multi-criteria decision analysis (MCDA) is a discipline which is used for decision making where multiple criteria are explicitly considered. The criteria can be qualitative or quantitative or both that are to be assessed for making decisions. Material cost and price evaluation, sustainability etc are the major factors that should be evaluated before selecting a Geo-location for river bridge construction. There are two types of MCDA problems and methods. They are: *Multi-criteria evaluation problem* and *Multi-criteria design problem*. In multiple criteria evaluation problem there will be some specific alternatives. And based on the analysis using the multiple criteria it finds the best alternative for decision maker. In multiple criteria design problem alternatives are not known explicitly. A mathematical model is applied to get the solution. Alternative numbers can be either infinite or not countable. Here in this study first approach is followed and analyzed some specific number of alternatives in order to get the preminent solution of selecting sustainable geo-location for river bridge construction.

III. EXPERIMENT DESIGN

The process of criteria selection can be varied from site to site and also depends on the survey authority of the contraction company. Here only the basic criteria are taken into account for the experiment. There are two types of criteria mostly: one is qualitative criteria and another is quantitative criteria. In this study both are separated to do analysis as because of the convenience of the experiments.

Among all these criteria for simplicity only 6 important criteria are considered to undertake the experiments; this is because of making the numerical study convenient for the study. They are River Training Work (RTW), River Stream (RS), Maximum Bridge Span (BS), Bank Erosion (BE), Land Acquire & Resettlement Cost (LA) and Traffic Demand (TD).

Here RTW, RS, BS, BE are qualitative criteria and LA and TD are quantitative criteria. Figure 1 shows the process of multi-criteria decision analysis using evidential reasoning. Here $C_1, C_2, \dots, C_n$ indicates the criteria that are selected for the analysis and $H_1, H_2, \dots, H_n$ indicates the evaluation grades and each grade value contain any value between 0 and 1. These criteria can be qualitative or quantitative or both. $\omega_1, \omega_2, \dots, \omega_n$ indicates the weights come from the respective selected criteria. Final step of this analysis produces the evaluation grade values for a specific bridge construction site. And this analysis is repeatable for all bridge construction sites that are selected primarily for choosing the best location for constructing a river bridge.

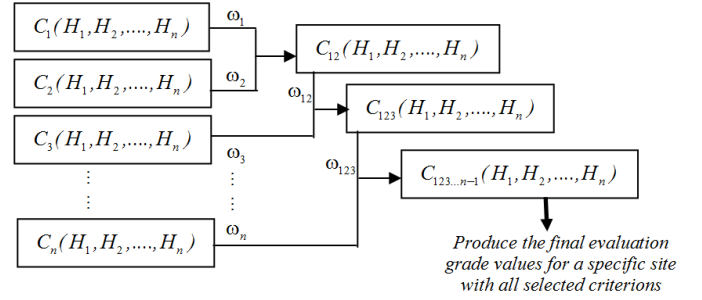


Fig. 1. Alternative selection process using multi-criteria decision analysis

As it is mentioned in methodology section that multi-criteria evaluation problem is considered in this work for analysis, so there are some alternatives; specifically four geo-location sites are selected priory. The aim of this work is to show how to identify the best geo-location for a river bridge construction based on analyzing the qualitative and quantitative criteria which should be considered in site selection decision making. Every criterion has its own evaluation grade and based on that evaluation grade all criteria have been analyzed together in order to evaluate the all alternative geo-locations for river bridge construction. Figure 1 shows the process of finding the best sustainable alternative.

A. The qualitative criteria list

- Traditional crossing point
- Maximum bridge span
- Minimum free board
- Space for the bridge foundations
- Slope profile
- Level difference between two banks.
- River Stream should be straight and narrow
- River bed should be stable
- Well define bank
- River training works
- River bank erosion
- Toppling instability of river bank
- Erosion of river slant
- Land sliding
- Erosion/ Siltation
- Agricultural/ vegetation loss
- Fisheries
- Air, noise and vibration pollution

B. The quantitative criteria list

- Bridge length
- Land Acquisition
- Resettlement
- Traffic demand
- Economic growth
- Cost estimates
- Population affected

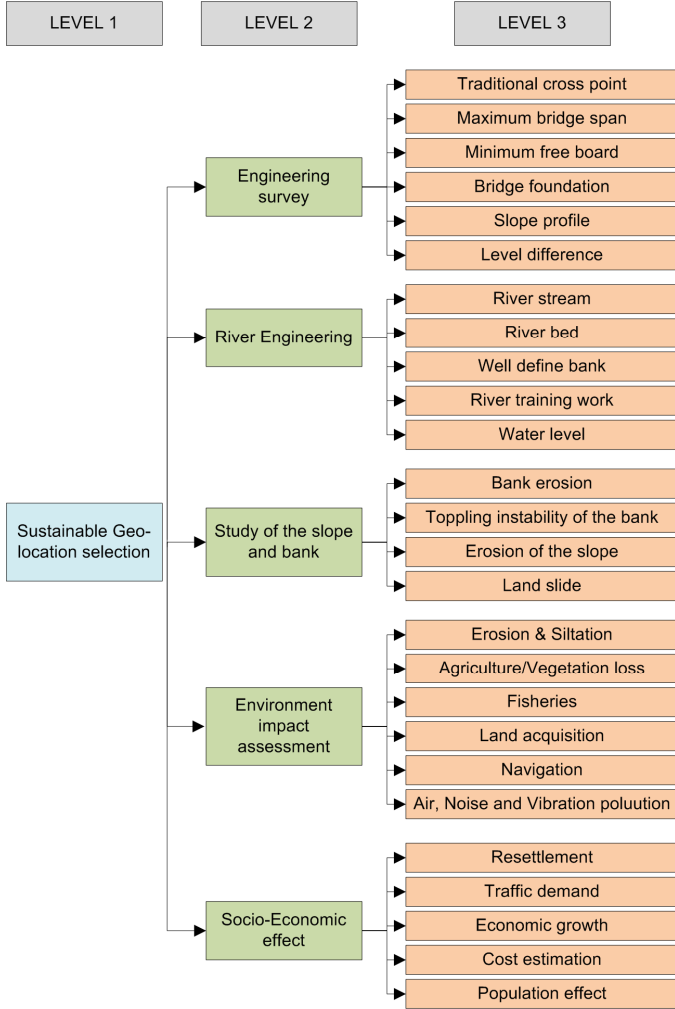


Fig. 2. Criteria list for river bridge site selection

IV. EXPERIMENT AND RESULT

In previous section ER algorithm has been described and also have discussed about how to apply it. Now this section will represent the outcomes of the experiments. The Padma River [22], is one of the major rivers of Bangladesh, divides the south-west zone from the northern and eastern regions of the country. Proposed Padma Bridge is a multipurpose bridge planned to contain four lanes highway traffic, a single freight rail way, a gas line and different communication amenities. Feasibility studies were done by an expert group from Japan International Cooperation Agency (JICA) [23]. They proposed four locations for the construction of the Padma Bridge [24]. The four locations are Geo-Site A: Paturia-Goalundo, Geo-Site

B: Dohar-Charbhadrasan, Geo-Site C: Mawa-Janjira and Geo-Site D: Chandpur-Bhedarganj. Now in this study, ER algorithm has applied to select suitable site for Padma Bridge.

The alternative set = {Geo-Site-1, Geo-Site-2, Geo-Site-3, Geo-Site-4}

TABLE I. COMBINE DEGREE OF BELIEF FOR ALL ALTERNATIVES

Alternatives	Worst	Unfavourable	Favourable	Excellent
Site-1	0.045675	0.479918	0.460298	0.014109
Site-2	0	0.582845	0.358997	0.058158
Site-3	0	0.458043	0.47642	0.065537
Site-4	0	0.507826	0.460596	0.031578

TABLE II. EXPECTED UTILITY FOR ALL ALTERNATIVES

Alternative	Minimum Utility	Maximum Utility	Average Utility
Site-1	0.61071	0.61071	0.61071
Site-2	0.618828	0.618828	0.618828
Site-3	0.651873	0.651873	0.651873
Site-4	0.630938	0.630938	0.630938

From the TABLE II, it can be seen that the average utility value of site-3 is larger than Site-4 and which is larger than Site-2 and Site-2 is larger than Site-1. It also depicts the maximum utility value of site-3 is greater than the maximum utility value of all other sites. So rank of the sites will be:

GeoSite-3 > GeoSite-4 > GeoSite-2 > GeoSite-1, which means: Mawa-Janjira > Chandpur-Bhedarganj > Dohar-Charbhadrasan > Paturia-Goalundo

Hence it is clear that according to ER approach site-3 (Mawa – Janjira) is the best suitable site for constructing the Padma Bridge and then the three other are preferred sites.

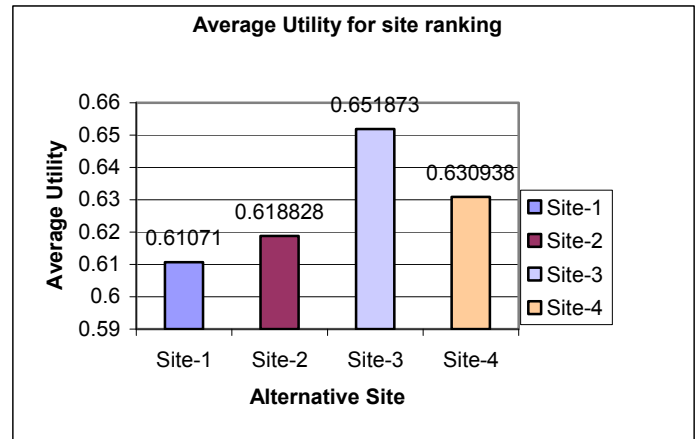


Fig. 3. Average Utility for site ranking

V. SCOPE AND LIMITATION

To construct a bridge over river, proper site selection is an important decision that has to be determined. Site selection is a multi-criteria decision-making problem that should be taken care off very carefully. Moreover determining all criterions is a complex issue. Though this study suggests some good criterions, but still there have some space for doing some further research on that issue to identify each and every important criterion to evaluate better geo-location selection. Hierarchical structure of criteria is not considered in this study, but some criteria may contain several sub criteria themselves. The study is done on only one domain, the Padma Bridge; but it can be applied to some other river bridge construction sites in order to evaluate the outcomes from the experiment. And thus there will be a comparison whether the model can apply to the all around the globe for sustainable geographical location selection for river bridge construction. Though this work is related to only river bridge construction but it can be applied to the other engineering constructions domain where multi-criteria decision problem should be explicitly considered.

VI. CONCLUSION

The objective of this research is to assess the sustainable geographical location for the River Bridge construction. Selecting a sustainable geographical location is a complex decision making problem since many criteria have to be analyzed unambiguously. At past many approach were used to select suitable geo-location for river bridge construction but most of them were based on some specific criteria. In this study, ER approach is applied for selecting suitable geo-location for river bridge construction which is more feasible than any other approaches. Geo-location selection for river bridge construction is a multifarious decision making problem because both qualitative and quantitative criteria are involved. So ER approach is selected for analysis since it can handle the both criteria. It also works for any number of criteria and any number of alternative sites which make it as a pioneer in such kind of complex decision making environment. Moreover ER approach can handle uncertainty and incompleteness of data and information, which is the most effective feature of ER to think of picking it up for such kind of experiments. In future further studies will be done to solve the limitations that are investigated in this study and better methodology will be developed to select a sustainable geographical location for river bridge construction.

REFERENCES

- [1] Islam, M T (2010) "Evaluation of sites for a river bridge : A GIS and remote sensing approach" TRITA-LWR-EX-10:13
- [2] Saaty, Thomas L. (2008). Decision Making for Leaders: The Analytic Hierarchy Process for Decisions in a Complex World. Pittsburgh, Pennsylvania: RWS Publications. ISBN 0-9620317-8-X
- [3] Dey, Prasanta Kumar (November 2003). "Analytic Hierarchy Process Analyzes Risk of Operating Cross-Country Petroleum Pipelines in India". *Natural Hazards Review* 4 (4): 213–221. DOI:10.1061/(ASCE)1527-6988(2003)4:4(213)
- [4] Saaty, T.L. (1980). The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation. New York: McGraw-Hill.
- [5] Atthirawong, Walailak; Bart McCarthy (September, 2002). "An Application of the Analytical Hierarchy Process to International Location Decision-Making". In Gregory, Mike. *Proceedings of The 7th Annual Cambridge International Manufacturing Symposium: Restructuring Global Manufacturing*. Cambridge, England: University of Cambridge. pp. 1–18.
- [6] L. Zadeh, A simple view of the Dempster-Shafer Theory of Evidence and its implication for the rule of combination, *The AI Magazine*, Vol. 7, No. 2, pp. 85-90, Summer 1986.
- [7] D. L. Xu and J. B. Yang, "Introduction to multi-criteria decision making and the evidential reasoning approach", Working Paper Series, Paper No.:0106, ISBN: 1 86115 111 X (<http://www.umist.ac.uk/management>), Manchester School of Management, UMIST, pp. 1-21, 2001.
- [8] J. B. Yang, "Rule and utility based evidential reasoning approach for multiple attribute decision analysis under uncertainty", *European Journal of Operational Research*, Vol. 131, No.1, pp.31-61, 2001.
- [9] Köksalan, M.M. and Sagala, P.N.S., M. M.; Sagala, P. N. S. (1995). "Interactive Approaches for Discrete Alternative Multiple Criteria Decision Making with Monotone Utility Functions". *Management Science* 41 (7): 1158–1171
- [10] Bragge, J.; Korhonen, P., Wallenius, H. and Wallenius, J. (2010). "Bibliometric Analysis of Multiple Criteria Decision Making/Multiattribute Utility Theory". *IXX International MCDM Conference Proceedings*, (Eds.) M. Ehrgott, B. Naujoks, T. Stewart, and J. Wallenius,. Springer, Berlin 634: 259–268.
- [11] Kari Sentz and Scott Ferson (2002); *Combination of Evidence in Dempster-Shafer Theory*, Sandia National Laboratories SAND 2002-0835
- [12] Y. M. Wang, J. B. Yang and D. L. Xu, "Environmental Impact Assessment Using the Evidential Reasoning Approach", *European Journal of Operational Research*, Vol.174, No.3, pp.1885-1913,2006.
- [13] Zhi-Jie Zhou, Chang-Hua Hu, Dong-Ling Xu, Jian-Bo Yang and Dong-Hua Zhou, "Bayesian reasoning approach based recursive algorithm for online updating belief rule based expert system of pipeline leak detection" *Expert Systems with Applications* 38(2011) 3937-3943.
- [14] Chang-Hua Hu, Xiao-Sheng Si and Jian-Bo Yang, "Dynamic Evidential Reasoning algorithm for systems reliability prediction", *International Journal of Systems Science* Vol. 41, No. 7, July 2010, 783-796.
- [15] Köksalan, M., Wallenius, J., and Zionts, S. (2011). *Multiple Criteria Decision Making: From Early History to the 21st Century*. Singapore: World Scientific.
- [16] B. Yang and P. Sen, "Multiple attribute design evaluation of large engineering products using the evidential reasoning approach", *Journal of Engineering Design*, Vol.8, No.3, pp.211-230, 1997.
- [17] A. Taroun and J. B. Yang, "Dempster-Shafer theory of evidence: potential usage for decision making and risk analysis in construction project management." *Journal of the Built and Human Environment Review* 4, no. 1(2011): 155-166
- [18] J. B. Yang and D. L. Xu, "Nonlinear information aggregation via evidential reasoning in multi-attribute decision analysis under uncertainty," *IEEE Trans. Syst. Man Cybern., A, Syst. Humans*, vol. 32, no. 3, pp. 376-393, May 2002.
- [19] Tanjim Mahmud and Mohammad Shahadat Hossain, "An Evidential Reasoning-based Decision Support System to Support House Hunting", *International Journal of Computer Applications* 57(21):51-58, November 2012
- [20] P. Sen and J. B. Yang, "Design decision making based upon multiple attribute evaluation and minimal preference information", *Mathematical and Computer Modelling*, Vol.20, No.3, pp.107-124, 1994.
- [21] Saaty, Thomas L. (1996). *Decision Making with Dependence and Feedback: The Analytic Network Process*. Pittsburgh, Pennsylvania: RWS Publications. ISBN 0-9620317-9-8
- [22] J Masud Hasan Chowdhury, "Padma River" *Banglapedia: National Encyclopedia of Bangladesh*.
- [23] Japan International Cooperation Agency (JICA) "THE FEASIBILITY STUDY OF PADMA BRIDGE" JICA,2005.
- [24] W.K. Wheeler, R.J. Aves & C.J. Tolley,Mohammad Zaman,Md. R Islam" Detailed design of the Padma Multipurpose Bridge, Bangladesh – An Overview" *IABSE-JSCE Joint Conference on Advances in Bridge Engineering-II*, August 8-10, 2010, Dhaka, Bangladesh.

Developing a Block-wise Interpolation based Adaptive Data Embedding Scheme

Md. Abdul Wahed*

Hussain Nyeem[†]

Department of Electrical, Electronic and Communication Engineering
Military Institute of Science and Technology (MIST), Mirpur Cantonment, Dhaka-1216
*wahedruet@gmail.com [†]h.nyeem@eece.mist.ac.bd

Abstract—Data hiding schemes based on quadratic interpolation have the potential to offer better image quality and higher embedding capacity. Being reversible, such schemes also allow user to restore the embedded image back to its original version, which is the key requirement of many applications like medical and military imaging. This paper presents a simplified data hiding scheme based on quadratic interpolation. Particularly, we apply quadratic interpolation in non-overlapping blocks of an input image. Instead of calculating multiple interpolated values considering neighborhood pixels in multiple directions, we simply consider only one (*i.e.*, either horizontal, vertical or diagonal) direction for interpolation and correlate the interpolated values both with and without corresponding local mean. This significantly reduces the complexity and offers high embedding capacity as high as 2.5 Mbits with better or similar image quality, while compared with a recently published interpolation based scheme.

Keywords—adaptive embedding; block-wise interpolation; reversible embedding; data hiding; digital watermarking

I. INTRODUCTION

Data hiding is an emerging technology that embeds a set of *data* (or more specifically *message* or *watermark*) into a *cover-image* (or other media like audio and video). It has different objectives such as image content-authentication, copy-right protection, integrity establishment and annotation. These objectives require different properties, which mainly define desired data *generation* and *embedding* processes [1]. In *generation*, data having required properties are generated for an intended application. The *embedding* process on the other hand deals with how and where the data are to be embedded in the cover image. With employing a suitable *generation* function, development of the *embedding* schemes has attracted much attention in recent data hiding research [2]–[7].

Data embedding schemes are evolving to attain two main requirements: (*i*) better image quality and (*ii*) higher embedding capacity. Data embedding introduces an inevitable distortion to the cover-image. Minimizing this distortion also reduces the embedding capacity. As a result, achieving high embedding capacity with low distortion has remained challenging, addressing which underpins many data hiding schemes. The early developments of data hiding schemes were mainly *irreversible* [8]–[10]. Those schemes have permanent degradation and relatively limited capacity, which are not desirable in applications like medical and military imaging. Therefore many reversible data hiding schemes have been developed that can restore an embedded image to the original version, and offer better image quality and higher embedding capacity.

Reversible embedding schemes can be of different types; for example, Difference Expansion (DE) schemes [11], [12],

Histogram Shifting (HS) schemes [13], [14]; Reversible Contrast Matching (RCM) schemes [15], [16], Prediction Error Expansion (PEE) schemes [3], [17]. Recently interpolation based data embedding schemes have created a new paradigm of reversible data hiding [4]–[7], [18]. Such schemes allow the data to be embedded only in the interpolated pixels of a cover image keeping the original pixels untouched. The embedded data then can be extracted and the cover image can be restored to the original version When required. These schemes therefore have the potential to be used in many applications like medical and military imaging.

Different interpolation techniques are employed for up- (or down-) sampling the cover image during embedding; for example, a neighbor mean interpolation [18], a neighbor pixels interpolation [4], combination of interpolation and HS [5], combination of maximized difference between neighbor pixels and an extended interpolation technique [6], and quadratic interpolation in all possible directions of an up-sampled pixels in an image [7]. Those schemes although improved interpolated image quality for high embedding capacity, they do not account for any varying capacity requirements. This means that, within the maximum embedding capacity limit, embedding of varying size payloads may result in the same level of distortion, which is not desirable and can be effectively managed using a suitably designed adaptive data embedding process.

In this paper, we have developed and presented an adaptive data embedding scheme based on quadratic interpolation. Unlike an existing interpolation based data hiding scheme [7], we have introduced a simplified up-sampling process both with and without the consideration of image blocks' mean values. Additionally, instead of always considering a maximum possible embedding capacity of an image, an adaptive embedding approach has been developed and employed. This adaptive embedding determines how many bits of an up-sampled pixel can be used for the best possible embedded image quality.

This paper is structured as follows. Sec. II presents a general framework of quadratic interpolation for up-sampling an image. Our proposed scheme is presented in Sec. III. Experimental setup, preliminary results and performance analysis are presented in Sec. IV. Conclusions are given in Sec. V.

II. INTERPOLATION BASED IMAGE UP-SAMPLING

This section presents a general framework of quadratic interpolation for up-sampling an input image. As discussed in Sec. I, Zhang *et al.* [7] reported that the quadratic interpolation offers better up-sampled image quality. Therefore, in our proposed data embedding scheme, we use quadratic interpolation to compute the up-sampled image. However, unlike

pixels :	P_1	P_2	P_3	P_4	P_5
pixel values :	y_1	y_2	y_3	y_4	y_5
pixel positions :	$x_1=1$	$x_2=2$	$x_3=3$	$x_4=4$	$x_5=5$

Fig. 1. Pixel-frame for quadratic interpolation.

the up-sampling process used in [7], we simplify the process considering image blocks of size 5×5 with the correlation of respective neighbor pixels' mean values. This is explained in Sec. III.

For a general framework of quadratic interpolation, let us consider a set of 5 consecutive pixels, $\{P_i = (x_i, y_i)\} : i \in \{1, 2, \dots, 5\}$ in a given direction. The direction can be horizontal, vertical or diagonal. Here, each pixel, P_i consists of its position, $x_i = i$ and intensity value $y_i \in \{0, 2^n - 1\}$, where n is the bit-depth of given image, as shown in Fig. 1. Also note that $\{P_i\} : i \in \{1, 3, 5\}$ are the original pixels and the other two pixels, i.e., $\{P_i\} : i \in \{2, 4\}$ are up-sampled pixels that are to be computed as follows.

To compute the interpolated values (i.e., up-sampled pixels, $\{P_2, P_4\}$), we first formulate a set of three parabolic equations for the three known/original pixels as in (1). These equations can be solved to obtain the values of a, b , and c , which can be later used to calculate the interpolated values, y_2 and y_4 using (2).

$$\left. \begin{aligned} ax_1^2 + bx_1 + c &= y_1 \\ ax_2^2 + bx_2 + c &= y_2 \\ ax_3^2 + bx_3 + c &= y_3 \end{aligned} \right\} \quad (1)$$

$$\left. \begin{aligned} y_2 &= ax_2^2 + bx_2 + c \\ y_4 &= ax_4^2 + bx_4 + c \end{aligned} \right\} \quad (2)$$

III. PROPOSED ADAPTIVE EMBEDDING

This section presents our proposed data embedding scheme based on quadratic interpolation. The main feature of this scheme, different from existing interpolation based data embedding schemes, is an adaptive embedding approach that aims to provide the required embedding capacity with minimum possible distortion and thus to output high quality embedded images. We have also simplified the block-based quadratic interpolation used for up-sampling an input image by considering only one directional interpolation either in horizontal, vertical or diagonal direction. This up-sampling process has also a provision for considering local mean for computing an up-sampled value. Therefore, the proposed scheme has two main parts: (i) up-sampling using quadratic interpolation, and (ii) adaptive data embedding, which are discussed below.

A. Image Up-Sampling

Let us consider an input image, I of size $M \times N$. The quadratic interpolation based image up-sampling is carried out as per the following steps.

Input: An input image, I

Output: An up-sampled image, $\{I_p\}$

Step 1: I is divided into 3×3 non-overlapping blocks, $\{I_k\}$ with necessary padding, where k is the total number of 3×3 blocks.

Step 2: Each block, I_k is up-scaled to a block, I_b of size 5×5 , such that

$$I_b(i, j) = \begin{cases} I_k(\frac{i+1}{2}, \frac{j+1}{2}) & \text{if } i, j \in \{1, 3, 5\} \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

Step 3: For all the up-sampled pixels in a block, I_b , interpolated values are calculated as discussed in Sec. II.

Step 4: For all the up-sampled pixels in a block, I_b , the mean of their nearest pixels are individually calculated.

Step 5: Each up-sampled pixel in a block, I_b is replaced with the mean of respective interpolated value (Step 3) and neighborhood mean value (Step 4). Thus we get a set of up-sampled blocks, $\{I_p\}$ where each block is of size 5×5 .

The above steps simplify the interpolation and local-mean based up-sampling process. In this up-sampling process, a set of interleaving pixels are added between each consecutive rows and columns, whose initial values are set to '0'. The interpolated value for each of these interleaving pixels is calculated as per the process discussed in Sec. II.

To calculate local mean value for the interleaving pixel, p , we consider the set of its known (or original) neighbor pixels, $\{q\}$, i.e., $q \in \{I_b\} \cap \{I_k\}$ that satisfies condition given in (4). Here, $d(p, q)$ is the euclidean distance between p and q . We then calculate the mean of the neighbor pixels $\{q\}$ for each p , which we denote as the local mean of p .

$$\forall p \in \{I_b\} - \{I_k\}, \forall q \in \{I_b\} \cap \{I_k\}, d(p, q) \leq \sqrt{5} \quad (4)$$

For each interleaving pixel, p , once we obtain the interpolated and local-mean values, we compute their average value and then replace p with that value to finally obtain the up-sampled image as a set of up-sampled blocks, $\{I_p\}$.

Additionally, the proposed up-sampling process can optionally disregard the local mean to compute the interleaving pixels. In this case, instead of Step 4 and Step 5 in the above up-sampling process, only the pre-allocated interleaving pixels (with '0's) are replaced by the interpolated values calculated in Step 3. In Sec. IV, we present experimental results to compare these two up-sampling processes.

B. Adaptive Embedding

In embedding, the data or secret message is embedded into the cover image without changing the original pixels of the up-sampled image. The up-sampled pixels, with/without the consideration of the local mean as discussed above, are only used to embed the secret data.

The same 5×5 blocks, output of the up-sampling process, are used for embedding. Fig. 2 shows a basic 5×5 block of the interpolated image, where the cross-marked pixels denote the unchanged pixels, and the dot-marked pixels denote the

×	•	×	•	×
•	•	•	•	•
×	•	×	•	×
•	•	•	•	•
×	•	×	•	×

Fig. 2. Structure of up-sampled image blocks of size 5×5 .

interpolated pixels. We first consider the up-sampled pixels in each block of the image, $\{I_p\}$. We find the number of embeddable bits, n in each up-sampled pixel. Here, n is determined from (5), where y is the intensity value of the up-sampled pixel.

$$n = \begin{cases} \lceil \log_2(y) \rceil & \text{if } y > 1 \\ 1 & \text{otherwise} \end{cases} \quad (5)$$

Let us define a threshold, T for controlling the embedding capacity. If $n \leq 1 + T$, we replace the LSB of that pixel, y with 2 bits of the secret data. It can be easily accomplished by resetting the n -LSBs of that pixel and adding its decimal value with the decimal value of the n -bit secret data.

For any other values of n , *i.e.*, if $n > 1 + T$, a total of $n - T - 1$ secret-bits are embedded, keeping the T most significant bits (MSBs) of the embedding (*i.e.* up-sampled) pixels unchanged, and one LSB reserved for a flag bit. Here, the flag bit is used to determine whether the embedded data bits are original or complemented. Uses of the flag bit and T are explained below.

Let an embedding pixel is y . We first calculate two versions of the embeddable pixel, y_d and $y_{\bar{d}}$, and then replace y with one of them that is closer to y . The decimal value of $n - T$ bit data (consists of $n - T - 1$ bit secret data and 1-bit flag) is d and its complement value is $\bar{d} = 2^{n-T} - d - 1$. Now, embeddable pixel y_d is computed by embedding d in y . Similarly, another embeddable pixel, $y_{\bar{d}}$ is computed by embedding $\bar{d}$ in y . The embedding condition is given in (6), which means that an up-sampled pixel is replaced with an embeddable pixel, y_d or $y_{\bar{d}}$, for which the difference between the up-sampled and embeddable pixels remain smaller. Note that the flag-bit (the LSB) in d is kept '0' such that $\bar{d}$ contains the flag-bit 1 to indicate that the d is complemented. The idea of using this flag-bit is to further minimize the distortion caused by the embedding.

$$y = \begin{cases} y_d & \text{if } |y - y_d| \leq |y - y_{\bar{d}}| \\ y_{\bar{d}} & \text{otherwise} \end{cases} \quad (6)$$

$$C_m \geq C_r \quad (7)$$

Adapting embedding is achieved with the capacity controlling threshold T . The lower the value of T , the higher the embedding capacity. Thus, with $T = 1$, the maximum embedding capacity, C_m is determined. If we have the required capacity C_r (*i.e.*, the bit length of the payload, *i.e.*, the embeddable

15	116	197	209	217
65	105	188	198	56
47	162	145	194	146
189	132	160	180	161
17	111	132	59	245

(a)

15	116	197	209	217	217
65	105	188	198	56	56
47	162	145	194	146	146
189	132	160	180	161	161
17	111	132	59	245	245
17	111	132	59	245	245

(b)

Fig. 3. Example of the proposed image up-sampling process: (a) original image, and (b) image with non-overlapping blocks of size 3×3 (Step 1).

secret data plus any side information), embedding function determines the minimum value of T to satisfy the capacity condition in (7). In this setting, the value of T lies in the range of $[1, n - 2]$. Note that, with a largest value of T , the capacity condition may not be fulfilled for some large C_r . This means that the payload size may sometimes exceed the maximum embedding capacity, which may also be true for any other embedding schemes. In that case, employing a suitable payload compression technique or multilevel embedding approach may be considered [19], which is beyond the scope of this paper.

C. A Working Example

We now illustrate the up-sampling of an input image and data-embedding into the up-sampled image with an example. We consider a typical image of size 5×5 shown in Fig. 3a, which is subdivided into 3×3 non-overlapping blocks with necessary padding. We use here symmetric padding, *i.e.*, add symmetrically the last column(s) and last row(s) to make total number of columns and rows of input image multiple of 3 as shown in Fig. 3b.

Each 3×3 block is up-sampled to a block of size 5×5 by interleaving pixels both in rows and columns in the block as explained in Sec. III-A. These interleaving pixels are initially set to '0' values as shown in Fig. 4a, which are later replaced by the interpolated values as shown in Fig. 4b. If we consider the local mean, interleaving pixels are replaced by the average of the local mean and the interpolated values. Thereby the up-sampled image is computed as shown in Fig. 4c.

As the embedding process explained in Sec. III-B, using a random data (*e.g.* 1101010101010101101011010101001010...) to be embedded in any image block by block, we obtain the embedded image accordingly.

IV. EXPERIMENTAL RESULTS

In this section, we present experimental results for performance evaluation of our proposed adaptive embedding scheme. A set of common test images of size $512 \times 512 \times 8$ are used in our preliminary experimental setup. We use a computer having 1.3 GHz Intel core i5 processor and 4 GB memory, and MATLAB R2012b for the implementation of the proposed scheme.

The test images are up-sampled to a size of $855 \times 855 \times 8$ using block based quadratic interpolation, both with and without the consideration of local mean. A secret data or

15	0	116	0	197	209	0	217	0	217
0	0	0	0	0	0	0	0	0	0
65	0	105	0	188	198	0	56	0	56
0	0	0	0	0	0	0	0	0	0
47	0	162	0	145	194	0	146	0	146
189	0	132	0	160	180	0	161	0	161
0	0	0	0	0	0	0	0	0	0
17	0	111	0	132	59	0	245	0	245
0	0	0	0	0	0	0	0	0	0
17	0	111	0	132	59	0	245	0	245

(a)

15	68	116	159	197	209	214	217	218	217
49	70	102	144	197	203	142	105	93	105
65	80	105	141	188	198	109	56	38	56
65	97	125	150	171	195	117	70	54	70
47	121	162	170	145	194	164	146	140	146
189	150	132	135	160	180	168	161	159	161
82	102	119	132	143	104	173	214	227	214
17	73	111	131	132	59	175	245	168	245
0	64	108	130	129	44	176	256	282	256
17	73	111	131	132	59	175	245	268	245

(b)

15	90	116	137	197	209	215	217	155	217
57	89	123	131	161	208	159	159	95	92
65	89	105	140	188	198	121	56	108	56
92	94	122	130	148	149	90	103	114	114
47	125	162	143	145	194	161	146	116	146
189	152	132	114	160	180	167	161	179	161
121	145	118	117	122	137	193	175	218	214
17	94	111	121	132	59	182	245	227	245
41	81	98	101	113	113	192	227	226	227
17	83	111	112	132	59	187	245	233	245

(c)

Fig. 4. Example of the proposed image up-sampling process: (a) up-sampled image blocks of size 5×5 with 0 up-sampled values using (3), (b) up-sampled image blocks with interpolation and neighborhood-mean-based up-sampled values, and (c) up-sampled image blocks without considering neighborhood-mean.

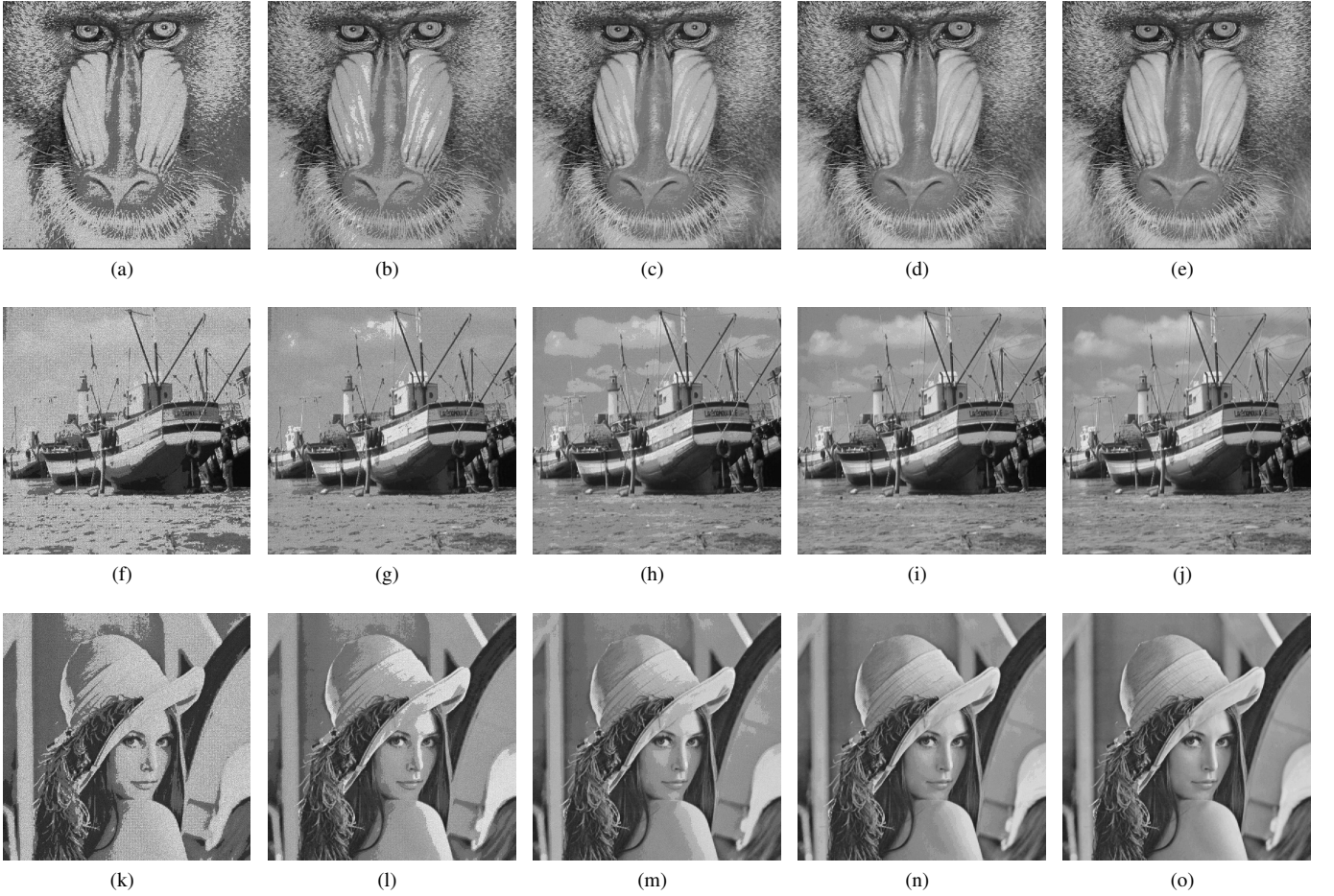


Fig. 5. Example of embedded image considering local mean: Baboon (first row), Boat (second row), and Lena (third row) image for: (a,f,k) $T = 1$, (b,g,l) $T = 2$, (c,h,m) $T = 3$, (d,i,n) $T = 4$, and (e,j,o) $T = 5$.

message is embedded into the up-sampled images as per the process discussed in previous sections for different values of T . Here, T deterministically controls embedding capacity for the required size of payloads.

We compare the embedded image quality in terms of two key measures: peak signal to noise ratio (PSNR) and structural similarity (SSIM). The PSNR values for the embedded images are calculated in terms of mean square error (MSE) as given in (8) and (9). Here $x_{i,j}$ and $y_{i,j}$ are the pixel value of position

(i, j) in an original image and its embedded version respectively of size $M \times N$. Similarly, SSIM values are computed using (10), where μ_x and μ_y are the average values of x and y , and σ_x^2 and σ_y^2 are the variance of x and y , respectively; σ_{xy} is the covariance of x and y ; and $c_1 = (k_1 L)^2$ and $c_2 = (k_2 L)^2$ are two regularization constants to stabilize the division with weak denominator for the L dynamic range of the pixel values and a set of small constants, k_1 and k_2 [20].

The PSNR and SSIM are calculated for different em-

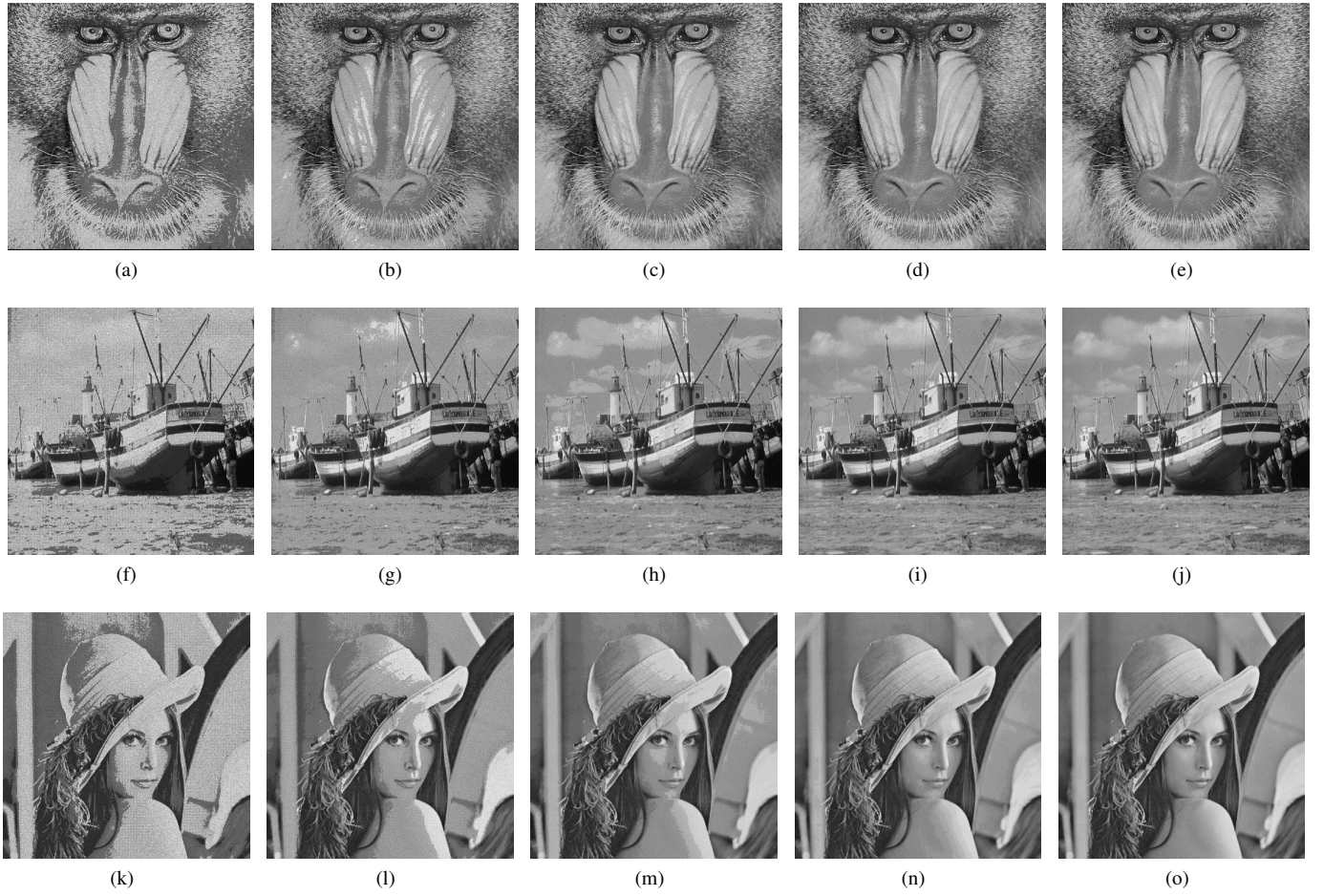


Fig. 6. Example of embedded image without considering local mean: Baboon (*first row*), Boat (*second row*), and Lena (*third row*) for: (a,f,k) $T = 1$, (b,g,l) $T = 2$, (c,h,m) $T = 3$, (d,i,n) $T = 4$, and (e,j,o) $T = 5$.

TABLE I. IMAGE QUALITY AND EMBEDDING CAPACITY COMPARISON FOR DIFFERENT IMAGES AND VALUES OF T

Image	Mean	T=1			T=2			T=3			T=4			T=5		
		PSNR (dB)	SSIM	Capacity (bits)	PSNR (dB)	SSIM	Capacity (bits)	PSNR (dB)	SSIM	Capacity (bits)	PSNR (dB)	SSIM	Capacity (bits)	PSNR (dB)	SSIM	Capacity (bits)
Lena	With	17.9581	0.2531	2496823	24.1858	0.4643	2028967	30.095	0.7035	1561111	36.0534	0.8785	1094409	42.033	0.9611	778182
	Without	17.965	0.2553	2496373	24.1758	0.4663	2028517	30.1131	0.704	1560661	36.0532	0.878	1094699	42.0273	0.9611	779938
Baboon	With	17.8293	0.3967	2540692	23.7568	0.6288	2073686	29.959	0.8302	1607676	35.9768	0.9402	1146906	41.9331	0.9829	739488
	Without	17.9299	0.4181	2535078	23.7066	0.6429	2068616	29.9526	0.8388	1602798	35.9479	0.9439	1143604	41.8085	0.9843	743699
Goldhill	With	19.1545	0.3148	2391534	25.5142	0.5731	1923678	31.3723	0.7958	1455822	37.5558	0.9271	1004072	43.2614	0.9766	685003
	Without	19.1137	0.3209	2391728	25.4854	0.5791	1923872	31.3435	0.7992	1456018	37.5484	0.9287	1004987	43.2099	0.9769	689064
Boat	With	16.7191	0.2632	2561225	23.7948	0.477	2093467	29.3591	0.7064	1637388	35.2131	0.8796	1211058	41.1438	0.9614	849866
	Without	16.7469	0.269	2557790	23.7844	0.4814	2090321	29.3276	0.7094	1634770	35.2577	0.8834	1210032	41.128	0.963	850905

bedding capacity controlling threshold T . Table. I illustrates experimental results for different values of the said parameters. As evident in the table, for different test images, (e.g. Lena, Baboon, Goldhill and Boat) the values of PSNR and SSIM increase (*i.e.*, image quality improves) with the increasing values of T . The embedding capacity (in bits) however reduces with the increasing values of T as it let the number of unchanged MSBs increase (*i.e.*, the no. of LSBs to be used for embedding is reduced).

$$\text{MSE} = \frac{\sum_{j=1}^N \sum_{i=1}^M (x_{i,j} - y_{i,j})^2}{MN} \quad (8)$$

TABLE II. ZHANG *et al.* SCHEME [7] VS OUR SCHEME

Image	Scheme	Payload (bpp)	PSNR (dB)	SSIM
Lena	Zhang <i>et al.</i> [7]	0.7631	33.3601	0.9712
	Ours	1.0645	42.0330	0.9611
Baboon	Zhang <i>et al.</i> [7]	1.6309	22.7899	0.8612
	Ours	2.1992	29.9590	0.8302
Goldhill	Zhang <i>et al.</i> [7]	0.9689	30.4157	0.9439
	Ours	1.3735	37.5558	0.9271
Boat	Zhang <i>et al.</i> [7]	0.8223	30.3440	0.9589
	Ours	1.1626	41.1438	0.9614

$$\text{PSNR} = 10 \log \frac{L^2}{\text{MSE}} \quad (9)$$

$$\text{SSIM} = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (10)$$

Our proposed scheme offers maximum possible embedding capacity for $T = 1$. In Table I, for example, the maximum embedding capacity, 2496823 bits and 2496373 bits (with and without considering the local mean respectively) are obtained for Lena image with $T = 1$. This high embedding capacity of 2.5 Mbits also means to have maximum possible distortion in the embedded image (*i.e.* PSNR of 17.95 dB and SSIM of 0.25). Lower capacity requirement however may result in better image quality. For example, the minimum capacities are 778182 bits and 779938 bits respectively with a PSNR of 42.03 dB and SSIM value of 0.96 for Lena image with $T = 5$. Thus, our proposed scheme offers improvement of image quality, PSNR of 17.95 dB to 42.03 dB and SSIM of 0.25 to 0.96 for the decreasing embedding capacity of 2.5 Mbits to 0.7 Mbits with an increasing value of T from 1 to 5. The similar trend is also evident for other test images.

The embedded versions of different test images are shown in Fig. 5 and Fig. 6 for values of T from 1 to 5. It is observed that, with the increment of T , visual artifacts are reduced. Our scheme thus offers a tradeoff between these two conflicting requirements, the embedded image quality and embedding capacity, for a varying size of payload by adaptive selection of a suitable value of T in a particular application scenario.

We also compare the performance of our proposed scheme with that of a recently published interpolation based scheme [7] shown in Table II. It is noticed that, for significantly higher payload (bit per pixel, bpp), our scheme offers better or similar performance in terms of PSNR and SSIM than Zhang *et al.* [7] scheme. For example, in case of Boat image, Zhang *et al.* [7] scheme provides a PSNR of 30.3440 dB and SSIM of 0.9589 for a payload of 0.8223 bpp, while our scheme offers better PSNR and SSIM values of 41.1438 dB and 0.9614 respectively for a larger payload of size 1.1626 bpp. This is also true for our other test images.

V. CONCLUSIONS

In this paper, we have developed and presented an adaptive reversible data hiding scheme based on quadratic interpolation. We have also presented the preliminary result of this scheme's performance evaluation. This result demonstrates that it can adaptively embed upto 2.5 Mbits of data into an input gray-scale image of size $512 \times 512 \times 8$. While compared with a recent interpolation based scheme, our proposed scheme offers significantly higher embedding capacity with better or similar image quality. The embedded image can also be instantly restored to the original version when required, by down-sampling the the embedded image. This means that the proposed scheme can be used to effectively hide different size of data keeping the original pixels of an image untouched, and thus it has the potential to be used in medical and military imaging.

Our research continues to further development of the proposed scheme by employing other interpolation techniques.

Performance evaluation and benchmarking of the scheme will also be carried out in an extended experimental setup with consideration of larger set of test-images and more number of related schemes.

REFERENCES

- [1] H. Nyeem, W. Boles, and C. Boyd, "Digital image watermarking: its formal model, fundamental properties and possible attacks," *EURASIP Journal on Advances in Signal Processing*, vol. 2014, no. 1, pp. 1–22, 2014.
- [2] L. Luo, Z. Chen, M. Chen, X. Zeng, and Z. Xiong, "Reversible image watermarking using interpolation technique," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 1, pp. 187–193, 2010.
- [3] Y. Hu, H.-K. Lee, and J. Li, "DE-based reversible data hiding with improved overflow location map," *IEEE Trans. CSVT*, vol. 19, pp. 250–260, 2009.
- [4] C.-F. Lee and Y.-L. Huang, "An efficient image interpolation increasing payload in reversible data hiding," *Expert Systems with Applications*, vol. 39, no. 8, pp. 6712–6719, 2012.
- [5] W.-J. Wang, Y.-H. Zhang, C.-T. Huang, and S.-J. Wang, "Steganography of data embedding in multimedia images using interpolation and histogram shifting," in *Proc. IHMSP'13*. IEEE, 2013, pp. 387–390.
- [6] J. Hu and T. Li, "Reversible steganography using extended image interpolation technique," *Computers & Electrical Engineering*, vol. 46, pp. 447–455, 2015.
- [7] X. Zhang, Z. Sun, Z. Tang, C. Yu, and X. Wang, "High capacity data hiding based on interpolated image," *Multimedia Tools and Applications*, pp. 1–24, 2016.
- [8] I. J. Cox, J. Kilian, T. Leighton, and T. Shamon, "Secure spread spectrum watermarking for images, audio and video," in *Proc. IEEE ICIP'96*, vol. 3. IEEE, 1996, pp. 243–246.
- [9] C.-Y. Lin, M. Wu, J. A. Bloom, I. J. Cox, M. L. Miller, and Y. M. Lui, "Rotation, scale, and translation resilient watermarking for images," *IEEE Trans. Image Processing*, vol. 10, no. 5, pp. 767–782, 2001.
- [10] C.-K. Chan and L.-M. Cheng, "Hiding data in images by simple lsb substitution," *Pattern recognition*, vol. 37, no. 3, pp. 469–474, 2004.
- [11] J. Tian, "Reversible data embedding using a difference expansion," *IEEE Trans. CSVT*, vol. 13, pp. 890–896, 2003.
- [12] A. M. Alattar, "Reversible watermark using the difference expansion of a generalized integer transform," *IEEE Trans. Image Processing*, vol. 13, pp. 1147–1156, 2004.
- [13] Z. Ni, Y.-Q. Shi, N. Ansari, and W. Su, "Reversible data hiding," *IEEE Trans. CSVT*, vol. 16, pp. 354–362, 2006.
- [14] V. Sachnev, H. J. Kim, J. Nam, S. Suresh, and Y. Q. Shi, "Reversible watermarking algorithm using sorting and prediction," *IEEE Trans. CSVT*, vol. 19, pp. 989–999, 2009.
- [15] D. Coltuc and J.-M. Chassery, "Very fast watermarking by reversible contrast mapping," *IEEE Sig. Proc. Lett.*, vol. 14, no. 4, pp. 255–258, 2007.
- [16] X. Chen, X. Li, B. Yang, and Y. Tang, "Reversible image watermarking based on a generalized integer transform," in *Proc. ICASSP*. IEEE, 2010, pp. 2382–2385.
- [17] D. M. Thodi and J. Rodríguez, "Expansion embedding techniques for reversible watermarking," *IEEE Trans. Image Processing*, vol. 16, pp. 721–730, 2007.
- [18] K.-H. Jung and K.-Y. Yoo, "Data hiding method using image interpolation," *Computer Standards & Interfaces*, vol. 31, no. 2, pp. 465–470, 2009.
- [19] H. Nyeem, W. Boles, and C. Boyd, "Watermarking capacity control for dynamic payload embedding," in *Recent Advances in Information and Communication Technology 2015*. Springer, 2015, pp. 143–152.
- [20] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

Design of Highly Stable and Efficient Molybdenum Telluride PV Cells with Arsenic Telluride BSF

Mrinmoy Dey^{1,2*}, Maitry Dey³, M. A. Matin^{1,2} and Nowshad Amin^{4,5}

¹Department of Electrical and Electronic Engineering, Chittagong University of Engineering and Technology,
Chittagong-4349, Bangladesh

²Renewable Energy Laboratory (REL), Chittagong University of Engineering and Technology,
Chittagong-4349, Bangladesh

³Department of Electrical and Electronic Engineering, Premier University,
Chittagong-4203, Bangladesh

⁴Solar Energy Research Institute (SERI), The National University of Malaysia,
43600 Bangi, Selangor, Malaysia

⁵Department of Electrical, Electronic and Systems Engineering, The National University of Malaysia,
43600 Bangi, Selangor, Malaysia

*mrinmoycuet@gmail.com

Abstract—For high efficiency and better thermal stability, Molybdenum Telluride (MoTe₂) is remarkable as potential photovoltaic (PV) cell. AMPS (Analysis of Microelectronic and Photonic Structures) simulator is used to investigate the cell performance parameters to design the highly efficient ultra-thin MoTe₂ PV cell. In this research work, it has been explored that the cell conversion efficiency of MoTe₂ PV cell is improved with the insertion of Arsenic Telluride (As₂Te₃) as back surface field (BSF) above the back contact metal. The highest conversion efficiency of 25.08% was found for As₂Te₃ BSF with only 0.9 μm of absorber layer whereas it was 17.06% for no BSF with 1 μm thickness of absorber layer. The thermal stability of MoTe₂ PV cell with As₂Te₃ BSF showed better stability also.

Keywords— Thermal Stability, Ultra-thin, MoTe₂, Photovoltaic cell, AMPS, As₂Te₃ BSF.

I. INTRODUCTION

Molybdenum Ditelluride (MoTe₂) is commonly known as Molybdenum Telluride and it is considered by the researchers as potential candidate for cost effective, highly efficient, stable PV cells and green energy for thin film technology in earth. Subsequently, some research groups are doing extensive research work to explore the hidden potentiality on MoTe₂ PV cells. The binary semiconductor material MoTe₂ has direct band gap 1.1 eV and the work function is 4.7 eV [1]. The percentage of mass distribution for MoTe₂ material is 27.32% molybdenum (Mo) and 72.68% tellurium (Te).

The higher performance of PV cell, preservation of absorber layer material and better thermal stability are the challenges for MoTe₂ PV cell research work. During the production of efficient MoTe₂ PV cells, tellurium pressure or vapour is often play significant role. In this study, Cadmium Sulphide (CdS) window layer is selected with MoTe₂ as

absorber layer to form CdS/MoTe₂ PV cells. In this research work, the possibility of ultra-thin MoTe₂ PV cell is investigated with the addition of potential BSF layer at the back contact for higher performance and better cell stability. Many research groups are researching extensively on CdTe solar cells to explore the hidden potentiality by the insertion of back surface field between the absorber layer and back contact metal. The performance of CdS/CdTe solar cell is improved through adding ZnTe BSF discussed by Dey et. al [2], As₂Te₃ BSF as discussed by Dey et. al [3], GeTe BSF as discussed by Dey et. al [4], PbTe BSF as discussed by Matin and Dey [5] and SnTe BSF as discussed by Dey et. al [6]. The insertion of suitable BSF layer placement between the back contact and absorber layer in the proposed cell showed better performance to design this ultra-thin high performance MoTe₂ PV cell.

In this research work, numerical analysis is done by well known AMPS simulator to examine the performance of ultra-thin CdS/MoTe₂ cell structure with As₂Te₃ BSF layer to explore the hidden potentiality of MoTe₂ PV cells with improved cell performance parameters. Both the higher cell conversion efficiency (25.08%) and better thermal stability were achieved in the modified proposed cell structure (SnO₂/Zn₂SnO₄/CdS/MoTe₂/As₂Te₃/Ni) with less than 1 μm of absorber layer.

II. MODELING AND SIMULATION

To represent the complicated system in a simplified mode, PV cell model is supportive by avoiding the non useful features. Consequently, all the significant parameters are considered and the non important ones are ignored, then a good model can be constructed for analysis. AMPS-1D simulator is widely used more than 200 research groups for designing PV cell structure [7]. The conventional structure of

CdS/MoTe₂ PV cell is shown in Fig. 1 for numerical analysis to investigate the different parameters of cell performance.

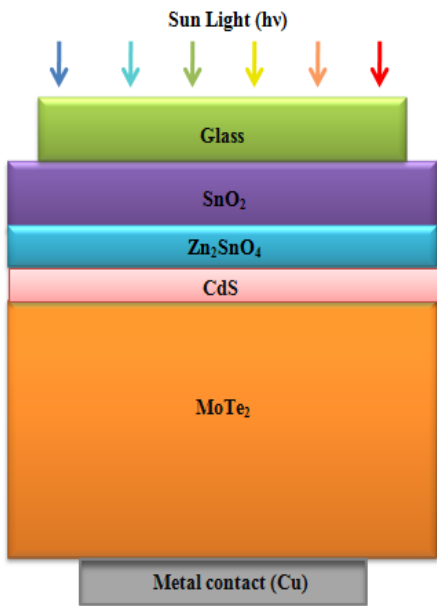


Fig.1 Modeling of conventional MoTe₂ PV cell

In this research work, the model of conventional MoTe₂ PV cell consists of Stannic Oxide (SnO₂) as TCO (Transparent conducting oxide) layer, Zinc Stannate (Zn₂SnO₄) as buffer layer, CdS as window layer and finally MoTe₂ as absorber layer. The hidden potentially of CdS/MoTe₂ cell structure can be explored by adding the potential As₂Te₃ BSF layer in between absorber layer and back contact metal.

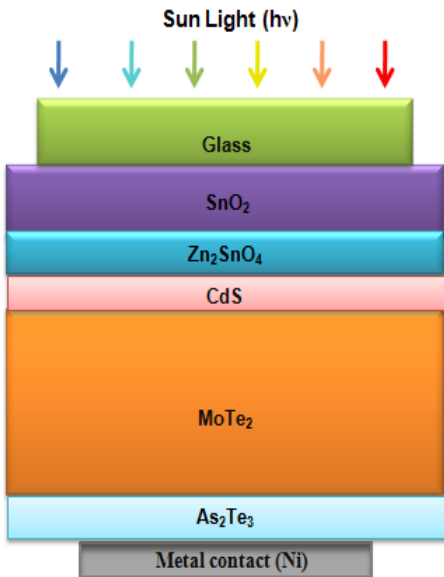


Fig.2 Proposed cell structure of MoTe₂ PV cell

In this study, the used materials with different parameters are shown in Table I which are chosen based on experimental data, literature values or in some cases reasonable estimations [3, 8].

TABLE I THE MATERIAL PARAMETERS USED FOR THE NUMERICAL ANALYSIS OF THE PROPOSED CELLS

Parameters	n-CdS	p-MoTe ₂	p-As ₂ Te ₃
Thickness, W (μm)	0.01-0.1	0.1-10	0.1
Permittivity, ε/ε ₀	9	13	20
Electron mobility, μ _c (cm ² /Vs)	350	110	500
Hole mobility, μ _p (cm ² /Vs)	50	426	210
Carrier concentration, n, p (cm ⁻³)	1×10 ¹⁷	1×10 ¹⁷	7.5×10 ¹⁹
Band gap, E _g (eV)	2.42	1.10	0.60
Density of state in conduction band, N _c (cm ⁻³)	1.8×10 ¹⁹	1×10 ¹⁵	1×10 ¹⁶
Density of state in valence band, N _v (cm ⁻³)	2.4×10 ¹⁸	1×10 ¹⁷	1×10 ¹⁷
Electron affinity, χ (eV)	4.50	4.20	4.0

III. RESULTS AND DISCUSSION

A. Optimization of Conventional Cell

The numerical analysis was done by AMPS-1D simulator in conventional CdS/MoTe₂ PV cell to optimize the thickness of CdS window layer as well as thickness and doping concentration of MoTe₂ absorber layer to overcome the challenges of thin film PV cells. Fig. 3 shows the variation of thickness for CdS window layer from 10 nm to 100nm to investigate the performance of MoTe₂ PV cells without BSF layer. From the numerical analysis considering the conversion efficiency and current technology, CdS layer is reduced to 70 nm in the cell structure of CdS/MoTe₂ PV cell for optimization of window layer along with 500 nm TCO layer and 100 nm Zn₂SnO₄ buffer layer.

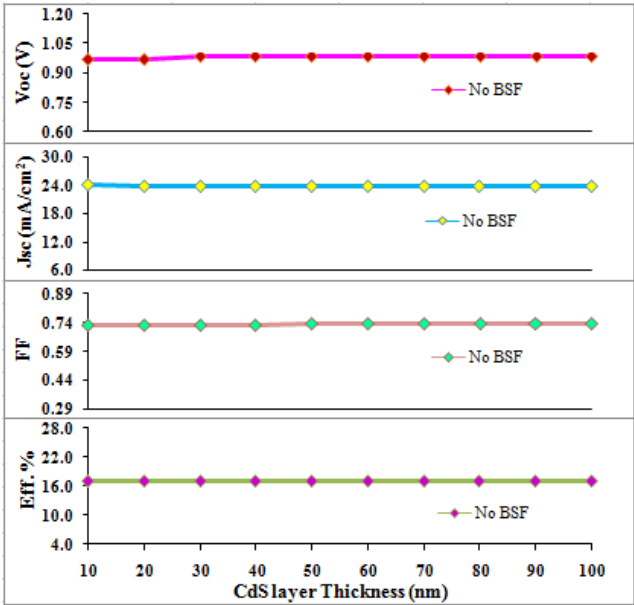


Fig.3 Effect of the CdS thickness variation on cell performance of the conventional cell

The main challenge of thin film PV cell is to preserve the absorber layer material for GW range solar power generation. Fig. 4 illustrates the thickness variation of absorber layer from 100 nm to 10 μm to examine the performance of MoTe₂ PV cells. The MoTe₂ absorber layer is optimized to 1 μm with 500 nm TCO layer, 100 nm Zn₂SnO₄ buffer layer and 70 nm CdS window layer in the CdS/MoTe₂ PV cell structure to design in ultra-thin range.

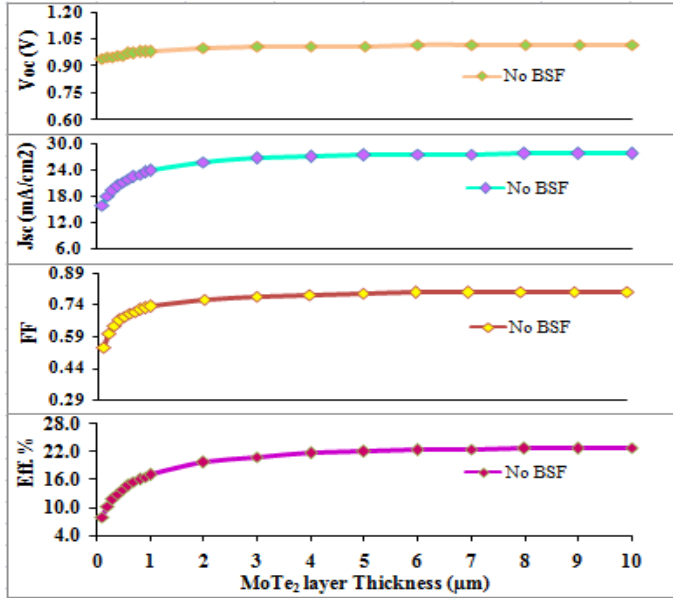


Fig.4 Effect of the MoTe₂ thickness variation on cell performance of the conventional cell

The performance parameters of PV cell are depends on the amount of doping concentration. The effect of doping concentration variation for absorber layer is shown in Fig. 5. For this purpose, the doping concentration of MoTe₂ layer is varied from 10¹³ cm⁻³ to 10¹⁷ cm⁻³ and the highest cell conversion efficiency is found for higher doping concentration of 10¹⁷ cm⁻³ which is now practically achievable.

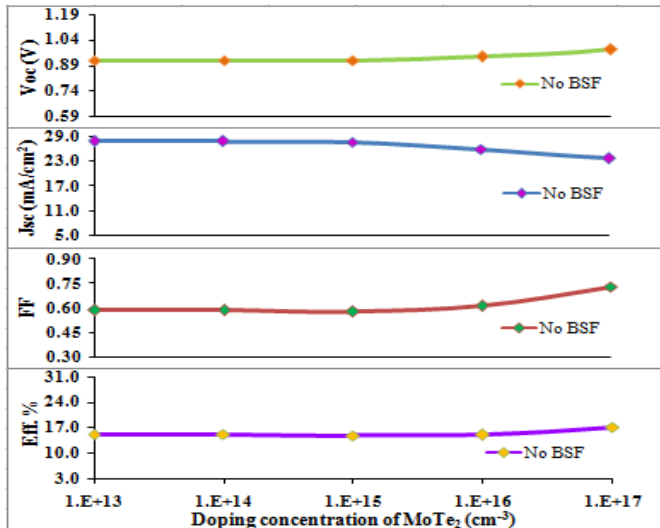


Fig.5 Effect of the doping concentration of MoTe₂ on cell performance of the conventional cell

B. Effects of As₂Te₃ BSF Insertion

In this research work, the conventional MoTe₂ PV cell is modified into proposed cell which consists of SnO₂ layer, Zn₂SnO₄ buffer layer, CdS window layer as well as 1 μm thickness of absorber layer with 10¹⁷ cm⁻³ doping concentration. The modification of conventional cell is completed by adding a low band gap material As₂Te₃ ($E_g = 0.60$ eV) of 100 nm thickness placed above the back contact metal to improve the cell conversion efficiency, thermal stability and preserve the absorber layer also by reducing back surface recombination velocity in between the absorber and BSF layer. Table II shows the effects of As₂Te₃ BSF insertion on MoTe₂ PV cell with 1 μm thickness of absorber layer. With the insertion of As₂Te₃ BSF layer, the magnitude of cell performance parameters is higher compared to conventional cell structure.

TABLE II THE OUTPUT PARAMETERS OF PROPOSED CELL WITHOUT BSF LAYER AND WITH BSF LAYER

Structure/ parameters	V _{oc} (V)	J _{sc} (mA/cm ²)	FF	Eff. (%)
Base case	0.98	23.740	0.730	17.06
As ₂ Te ₃ BSF	1.07	27.79	0.842	25.06

Numerical analysis has been done in the proposed cell to investigate the possibility to lower thickness of absorber layer. The simulation results are shown in Fig. 6 that the cell performance parameters are varied with the variation of absorber layer thickness from 0.1 μm to 5 μm . It is depicted in Fig. 6 that the cell conversion efficiency is getting increased in the 1 μm range of absorber layer with highly doped As₂Te₃ BSF. This is because for the improved values of open circuit voltage, fill factor and current density.

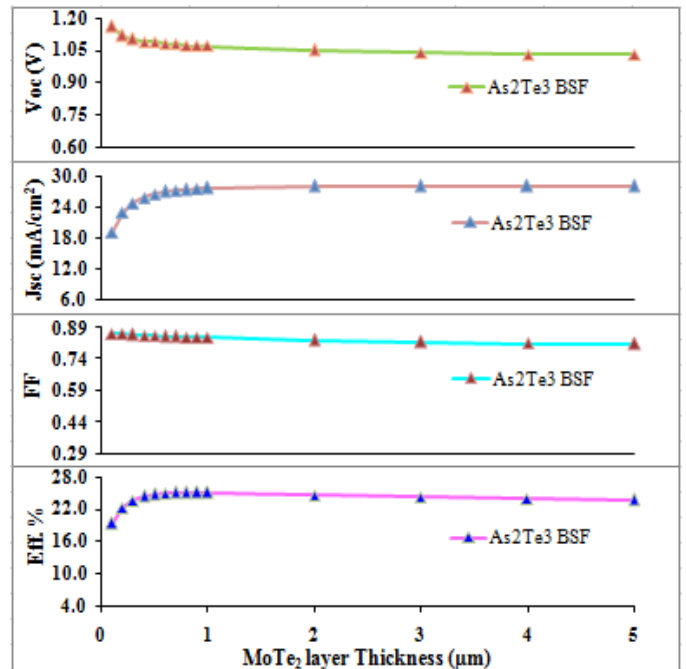


Fig.6 Variation of absorber layer thickness in the proposed cell with As₂Te₃ BSF

It is found from the numerical analysis that the conversion efficiency is 17.06% for conventional structure of MoTe₂ PV cell without BSF layer. The highest conversion efficiency 25.08% (FF = 0.843, V_{oc} = 1.07 V and J_{sc} = 27.69 mA/cm²) is achieved at only 0.9 μ m of absorber layer with 100 nm of As₂Te₃ BSF layer which is approximately 8% greater than the conventional cell structure. These results of cell performance with the addition of As₂Te₃ BSF layer are agreeable to the related published work [8]. The cell performance is improved by reducing back surface recombination velocity through highly doped p-type BSF layer.

C. Effects of Operating Temperature

The thermal stability is important parameter for PV cell and it is determined by the performance of cell under higher operating temperature. In this case, the operating temperature plays a significant role to test the cell performance parameters of the proposed cell with compared to conventional cell structure. From the numerical analysis, the potentiality of As₂Te₃ BSF layer insertion in the proposed cell structure of CdS/MoTe₂ PV cell is observed. The effect of operating temperature is shown in Fig. 7 for both MoTe₂ PV cell without and with As₂Te₃ BSF. The performance of MoTe₂ PV cell is examined for both without and with As₂Te₃ BSF at operating temperature ranging from 25°C to 160°C. The cell performance parameters are decreased at higher operating temperature and the degradation of conversion efficiency is decreased in the proposed MoTe₂ PV cell by adding the potential As₂Te₃ BSF.

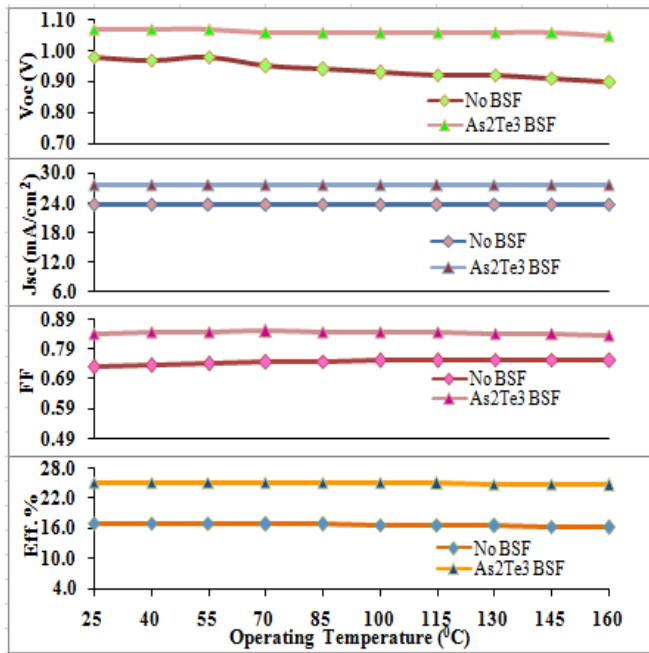


Fig. 7 Effect of temperature on cell performances of MoTe₂ cell with and without As₂Te₃ BSF

The normalized efficiency has been calculated for MoTe₂ PV cell with and without As₂Te₃ BSF. Fig. 8 shows that the normalized efficiency is decreased at higher operating temperature for both cases. For the case of without BSF in the

MoTe₂ PV cell, the normalized efficiency is linearly decreased with the increased of operating temperature ranging from 25°C to 105°C. By adding the As₂Te₃ BSF, the normalized efficiency of the proposed cell is almost constant in the operating temperature ranging from 25°C to 55°C with a TC of 0.27%/°C and then from 55°C to 105°C, the normalized efficiency is linearly decreased at a temperature co-efficient of -0.33%/°C which shows better stability in the proposed cell with As₂Te₃ layer.

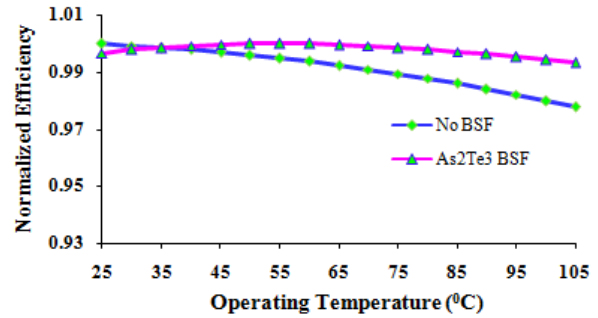


Fig. 8 Effect of operating temperature on normalized efficiency

IV. CONCLUSION

The effect of As₂Te₃ BSF insertion has been investigated in the proposed ultra-thin MoTe₂ PV cell to achieve higher cell conversion efficiency with lower thickness of absorber layer. The performance of MoTe₂ PV cell with As₂Te₃ BSF showed best cell conversion efficiency of 25.08% (FF = 0.843, V_{oc} = 1.07 V and J_{sc} = 27.69 mA/cm²) at 0.9 μ m thickness of MoTe₂ absorber layer. The cell conversion efficiency of baseline case MoTe₂ PV cell is 17.06% (FF = 0.730, V_{oc} = 0.98 V and J_{sc} = 23.74 mA/cm²) without As₂Te₃ BSF layer. With the addition of As₂Te₃ BSF, the conversion efficiency is increased upto 8%. Moreover, the thermal stability is better at higher operating temperature in case of As₂Te₃ BSF placing between the absorber layer and back contact metal.

ACKNOWLEDGMENT

The authors would like to thank the Department of Electrical and Electronic Engineering (EEE) and Renewable Energy Laboratory (REL) of Chittagong University of Engineering and Technology (CUET), Bangladesh.

REFERENCES

- [1] T. Löher, Y. Tömm, C. Pettenkofer, A. Klein and W. Jaegermann, "Structural dipoles at interfaces between polar II-VI semiconductors CdS and CdTe and non-polar layered transition metal dichalcogenide semiconductors MoTe₂ and WSe₂," *Semiconductor Science and Technology*, vol. 15, no. 6, pp. 514–522, 6 March, 2000.
- [2] Mrinmoy Dey, Maitry Dey, M. A. Matin and Nowshad Amin, "Design of High Efficient and Stable Ultra-Thin CdTe Solar Cells with ZnTe as a Potential BSF," in *Proceedings of the 3rd International Conference on Green Energy and Technology*, Dhaka, Bangladesh, September 2015.
- [3] Mrinmoy Dey, M. A. Matin and Maitry Dey, "Arsenic Telluride BSF for High Performance and Stable ultra-thin CdTe PV cell," *International Journal of Research in Computer Engineering and Electronics*, vol. 4, no. 2, July 2015.

- [4] Mrinmoy Dey, M. A. Matin, Nipu Kumar Das and Maitry Dey, "Germanium telluride as a BSF material for high efficiency ultra-thin CdTe solar cell," in *Proceedings of the 9th International Forum on Strategic Technology*, pp. 334-338, Cox's Bazar, Bangladesh, October 2014.
- [5] M. A. Matin and Mrinmoy Dey, "High performance ultra-thin CdTe solar cell with Lead Telluride BSF," in *Proceedings of the 3rd International Conference on Informatics, Electronics & Vision*, pp. 1-5, Dhaka, Bangladesh, May 2014.
- [6] Mrinmoy Dey, Maitry Dey, M. A. Matin and Nowshad Amin, "Design of high performance and ultra-thin CdTe solar cells with SnTe BSF from numerical analysis," in *Proceedings of the 18th International Conference on Computer and Information Technology*, Dhaka, Bangladesh, December 2015.
- [7] H. Zhu, A. K. Kalkan, J. Hou, and S. J. Fonash, "Applications of AMPS-1D for solar cell simulation," *AIP Conference Proceedings*, vol. 462, pp. 309-314, March 1999.
- [8] Mrinmoy Dey, Maitry Dey, M. A. Matin and Nowshad Amin, "High performance and stable molybdenum telluride PV cells with Indium Telluride BSF," in *Proceedings of the 4th International Conference on Developments in Renewable Energy Technology*, Dhaka, Bangladesh, January 2016.

RONI Segmentation for Medical Image Watermarking

Chowdhury M Abid Rahman*

Hussain Nyeem[†]

Department of Electrical, Electronic and Communication Engineering

Military Institute of Science and Technology (MIST), Mirpur Cantonment, Dhaka-1216

*capt.abid.mist@gmail.com

[†]h.nyeem@eece.mist.ac.bd

Abstract—Region of Non-Interest (RONI) based digital watermarking has attracted much attention recently for medical image (MI) applications. The legal and ethical concern of medical professionals grow with watermark embedding as it erratically modifies the MIs. To minimize this concern, RONI segmentation is an essential process of MI watermarking that confines the embedding region. However, finding a suitable RONI segmentation technique in general for multimodality MIs has remained an open challenge. In this paper, we propose a generic model of RONI segmentation for multimodality MIs. We start with investigating the existing RONI segmentation techniques for MI watermarking. We present a comparative summary of popular RONI segmentation techniques and their classification and analysis. We then identify the MI requirements for RONI segmentation. The proposed RONI segmentation model is the first steps toward the development of a suitable RONI segmentation technique for MI watermarking schemes.

Keywords—RONI segmentation; medical image; digital watermarking; data hiding

I. INTRODUCTION

Digital watermarking is a potential data hiding technology for medical image (MI) applications [1]–[12]. Its general model consists of three main processes namely, watermark generation, embedding and detection [13]. In watermark generation, a suitable watermark(s) is generated to attain the watermarking objective such as content-authentication, copy-right protection, integrity establishment and annotation. Watermark embedding embeds the watermark (or payload that includes any side information and the watermark bits). Watermark detection, on the other hand, makes a ‘reliable’ and objective decision, for example to declare the image content is authentic in an image content-authentication scenario. Among these three processes, embedding has a few additional requirements for the MI application scenario.

Watermark embedding incurs an inevitable distortion and thus for MIs, it has to ensure (i) an acceptable level of embedding distortion in the image, and (ii) continuous protection of the image [11], [14]. The first requirement ensures a confined embedding distortion in the watermarked images. Failing to achieve this requirement may arise legal and ethical concern of the medical professionals to use the watermarked images for any medical or clinical purpose. To address this, many reversible embedding techniques have been developed that can restore a watermarked image back to the original [11], [12], [15]–[17]. The later requirement, on the other hand, means that the watermark has to always remain embedded to provide continuous protection of the image. These two

requirements however conflict each other: if the watermarked image is restored to the original version, its security protection discontinues; if not, legal and ethical concerns arise. Region based embedding has showed the potential to resolve this conflict [1]–[5], [7].

The Region Of Non-Interest (RONI) in MIs are of no or little interest to doctors and medical professionals. RONI embedding aims at confining the distortion in the embedding region at a minimum level, while keeping the Region Of Interest (ROI, the complementary region of RONI in MIs that presents the anatomical objects or important features for diagnostic purposes) untouched. Thus, such embedding technique requires segmentation of RONI prior to watermark embedding in MIs.

However, MIs can be of different modalities such as CT—Computed Tomography, MR—Magnetic Resonance, X-ray, DSA—Digital Subtraction Angiography, RF—Radio Fluoroscopy, US—Ultrasound, MG—Mammography, to name a few. Different modality MIs have different perceptual contents and random locations of ROIs. Therefore, finding a suitable RONI segmentation technique, in general, has remained an open challenge for all modality MIs.

A general ROI/RONI segmentation technique would help preserve anatomical objects or important features in MIs irrespective of their modalities during watermark embedding. A RONI based watermark embedding would minimize legal and ethical concerns of medical professionals and improve the practicality of MI watermarking technology in different distant medical services like telemedicines and e-Health.

In this paper, we take the first steps toward developing a general RONI segmentation model for multi-modality MIs. We investigate the state of RONI segmentation in MI watermarking, and then identify their inputs, outputs, and ROI selection processes, types and suitabilities for any specific set of modalities (Sec. II). Thereby, we determine general requirements of RONI segmentation and develop a conceptual model of RONI segmentation for multi-modality MIs (Sec. III). We finally report preliminary results and their analysis to validate the practicality of the proposed model (Sec. IV). We also pointed out new challenges to be addressed in future research for further development of the proposed model (Sec. V).

II. RONI SEGMENTATION TECHNIQUES

This section reviews popular RONI-based watermark embedding schemes for MIs. Particularly, their considerations for ROIs/RONIs, target image modalities, segmentation criteria

TABLE I. SUMMARY OF ROI/RONI CONSIDERATIONS IN MI WATERMARKING

RONI shape	Work	MI Modality	MI Segmentation				Remarks
			Input	Output	Setting	Principle	
Rectangular	Fotopoulos <i>et al.</i> [5]	MR (brain)	Image and manually selected threshold	Location map of ROI/RONI	Semi-automatic	A simple thresholding is used with grouping of horizontal and vertical pixels that covers large no. of pixels bounded by the threshold to specify a ROI	ROI is embedded in RONI Location map of undesired segmented pixels is required Threshold is manually selected
	Fylakis <i>et al.</i> [18]	MR and X-Ray	Image and length payload	ROI as a rectangle containing the imaged tissue	Manual	ROI is defined by simple heuristics to clearly differ it from the background	Both ROI and RONI is used for embedding Least Significant Bits (LSBs) of ROI is used for data embedding Original LSBs of ROI is embedded in RONI
	Navas <i>et al.</i> [19]	X-Ray, MR and CT	Image	Two diagonal vertices of ROI rectangle	Manual	Background is considered as RONI	RONI is used for embedding
	Rahimi <i>et al.</i> [20]	CT and MR	Image	Coordinates of rectangle of ROI	Automatic	RONI is considered as the background of the image and marked as outside region of the ROI rectangle	Both ROI and RONI is used for embedding
	Pandey <i>et al.</i> [2]	Eye image	Image	Segmented ROI	Manual	ROI is defined by forming a rectangular boundary	RONI is used for embedding
Elliptical	Menom <i>et al.</i> [21]	CT, US, MR and X-Ray	Image	Isolated ROI and scrambled NROI	Manual	Logical ellipse is drawn around ROI and ROI is isolated	Only RONI is used for embedding
Elliptical/Rectangular	Al-Dmour and Al-Ani [22]	MRI	Image	Initial coordinates and size of ROI	Manual	Four element vector method specifies ROI with its initial coordinates and size	ROI is preserved and only RONI is used for embedding
Circular	Biswas <i>et al.</i> [23]	MRI and IVUS	Binary image	Sample image with marked ROI and RONI	Semi-automatic	Fuzzy C-means, Sobel edge, and Harris corner based technique specifies a circular ROI, where the center point and diameter are calculated by Harris corner detection algorithm	Only selected RONIs are used for embedding
Polygonal	Al-Haj and Amer [10]	MR, US and X-Ray (Grayscale MI)	Image	Partitioned image of non-overlapped blocks	Manual	Polygon capturing ROI is drawn by a radiologist or computer aided selection tool	Both ROI and RONI are used for embedding Reversible watermarking is used for ROI embedding Wavelet domain singular value decomposition is used for RONI embedding
	Guo and Zhuang [7]	CT, MR and US	Image	Vertex information of selected polygonal ROI	Manual	Polygon capturing ROI is drawn by a radiologist or computer aided selection tool	Outside of ROI is used for embedding
ROI Outline or Silhouette	Coatrieux <i>et al.</i> [3]	MR (head)	Averaged Image	RONI location detected image	Automatic	Morphological operation with a threshold	Both ROI and RONI are used for embedding
	Zhang <i>et al.</i> [9]	Not specified	Image	ROI detected image	Automatic	ROI is selected by block energy based algorithm with an adaptive threshold T to determine ROI	Only ROI is used for embedding using Histogram Shifting based reversible technique
	Al-Dmour and Al-Ani [8]	MR	Image and a defined threshold from Otsu's method	ROI selected image	Semi-automatic	Binarizing and segmenting the image with a threshold obtained from Otsu's method	Only ROI is used for embedding ROI is selected in rectangular shape according to the white pixel intensity

and techniques *etc.* are studied to determine general RONI segmentation requirements for different modality MIs.

Several RONI-based MI watermarking schemes have been developed. A few of them considered an automatic or semi-automatic segmentation of ROIs/RONIs. Coatrieux *et al.* [3] used morphological operation and threshold based RONI segmentation of MR images. Fotopoulos *et al.* [5] used a simple scanning and threshold-based algorithm to automatically find a smallest possible rectangle capturing the whole head shape of a brain-MRI. Zhang *et al.* [9] proposed a block-energy-based algorithm to determine the ROI of the original MIs. Fylakis *et al.* [18] used a ROI pixels' non-consecutive selection criteria based on the JND (Just Noticeable Distortion) for MRI and X-ray images. Al-Dmour and Al-Ani [8] utilized Otsu's method to determine the threshold for partitioning the MR image into ROI and RONI. Biswas *et al.* [23] employed fuzzy C-means, Harris corner and circle for ROI and RONI selection for US and MR images.

Other MI embedding schemes consider manual selection of ROI/RONI by the radiologist or a computer-aided selection tool. Pandey *et al.* [2] employed a manual selection of ROI/RONI of rectangle shape for digital eye image. Al-Haj and Amer [10] used a manual selection of ROI/RONI of a polygon shape for MR, US, and X-Ray gray-scale images. Guo and Zhuang [7] employed similar manual selection of polygon shaped ROI for CT, MR and US images. Memon *et al.* [21] also used manual segmentation of ellipse shaped ROI for CT, US, X-Ray, and MR images.

Different shapes of ROI/RONI have also been used in MI watermarking. Mainly five types of general ROI shapes have been studied; namely, rectangle [2], [5], [18]–[20], polygon [7], [10], ellipse [21], circle [23] or combination of them [22]. Additionally, ROIs have also been defined using their segmented region for example using morphological processing [3], Otsu's threshold [8] and adaptive thresholding [9].

Table I summarizes the findings discussed above, which suggest that no general ROI/RONI segmentation technique for multimodality MIs have been reported so far in the literature. As mentioned in Sec. I, the main possible reason is that the ROIs in multimodality MIs have a variety of visual structures (and properties and locations), to define which using general shapes is a challenging task. In this paper, we attempt to address this challenge by establishing a foundation for developing a general RONI segmentation technique for MI watermarking.

III. PROPOSED MODEL OF RONI SEGMENTATION

This section presents our proposed model of RONI segmentation for multi-modality medical images. This model consists of two main processes: (i) a suitable ROI (or RONI) segmentation process for a specific set of MI modalities, and (ii) a suitable description process that combines some basic shapes to effectively represent the segmented ROIs (or RONIs). In what follows, we briefly explain these processes.

A. ROI/RONI Segmentation

Conventional image segmentation techniques can be classified into three main categories: (i) edge-based, (ii) clustering based, and (iii) Watershed-based.

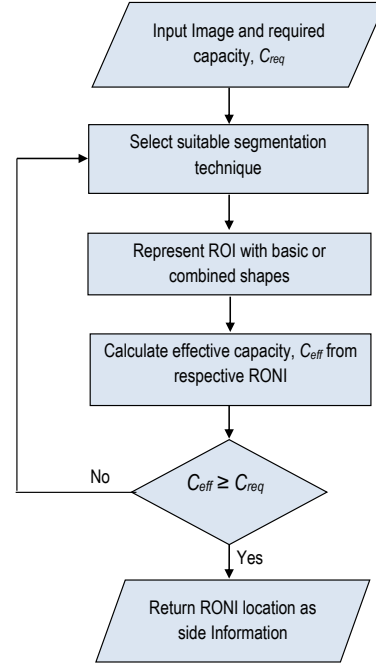


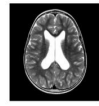
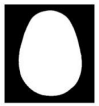
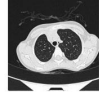
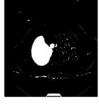
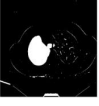
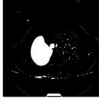
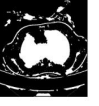
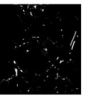
Fig. 1. Flowchart of the proposed general framework of RONI segmentation

Edge-based segmentation mainly employs different edge detection operators, for example, Roberts, Sobel, Laplacian, Prewitt, and Canny. These operators have different features and limitations [24]. For example, classical operators like Prewitt and Robert operators are simple and can detect edges and their orientation. However these techniques are sensitive to noise. Zero crossing operators like Laplacian can detect edges and their orientation with some fixed directional characteristics, but remain sensitive to noises. Gaussian based operators like Canny has separate considerations of error rate, edge localization and response with better signal to noise ratio. These techniques are usually complex and may have false zero crossing.

Clustering based segmentation, on the other hand, is a tool that makes data groups or clusters where the members of clustered data are made as similar as possible or as dissimilar as possible. *K*-means and *C*-means clustering are two popular methods of data clustering. In the first method, a collection of data is divided in to *k* numbers of disjoint clusters. It works in two consecutive phases. At first it investigates the number of centroid termed as *k*, in the second phase it takes data from clusters and compares with nearest centroid from the respective data point. Unlike *K*-means, *C*-means clustering relates the data point with every cluster of data. It deals with the ambiguity of any data point's location in a cluster by replacing its distance from means with probability. Probability here is a function of distance, for example the probability can be related inversely with the distance. It uses a weighted centroid based on those probabilities. Initialization, iteration and termination are same as *k*-means.

The watershed algorithm is another popular tool used for image segmentation. It represents an image topographically. The regions of the lowest intensity levels are compared with low land or lowest elevation and termed as *regional-minima*.

TABLE II. ROI/RONI SEGMENTATION FOR DIFFERENT MODALITY IMAGES

Modality	Original Image	Segmentation technique					
		Canny-based	Prewitt-based	Robert-based	Sobel-based	Log-based	Watershed-based
Xray							
							
MR							
							
CT							
							
Fluoroscopy							
							

The groups of connected pixels with their gray level lower than neighboring pixel are called *minima*. Edges are determined considering the bridges between the highest and lowest intensity levels and regional-minimas, which are calculated as Watersheds.

We use some popular conventional segmentation techniques discussed above to determine their MI modality wise suitability, which will be presented in Sec. IV.

B. ROI/RONI Description

The ROI/RONI segmented region obtained in the previous step discussed above now needs to be represented in terms

of basic shapes. The basic shapes are typically a rectangle, square, circle, ellipse, *etc.* These basic shapes and/or their combinations are used to define the segmented region. The main objective of doing so is to determine the minimum information to represent the segmented RONI/ROI, which will be used in both watermark embedding and extraction for exact localization of the segmented region.

Our proposed model aims to employ an automatic ROI/RONI representation process. It is envisaged that the representation may be done by covering segmented region of an image by minimum number of basic shapes and identifying the co-ordinates of intersected regions. In this paper,

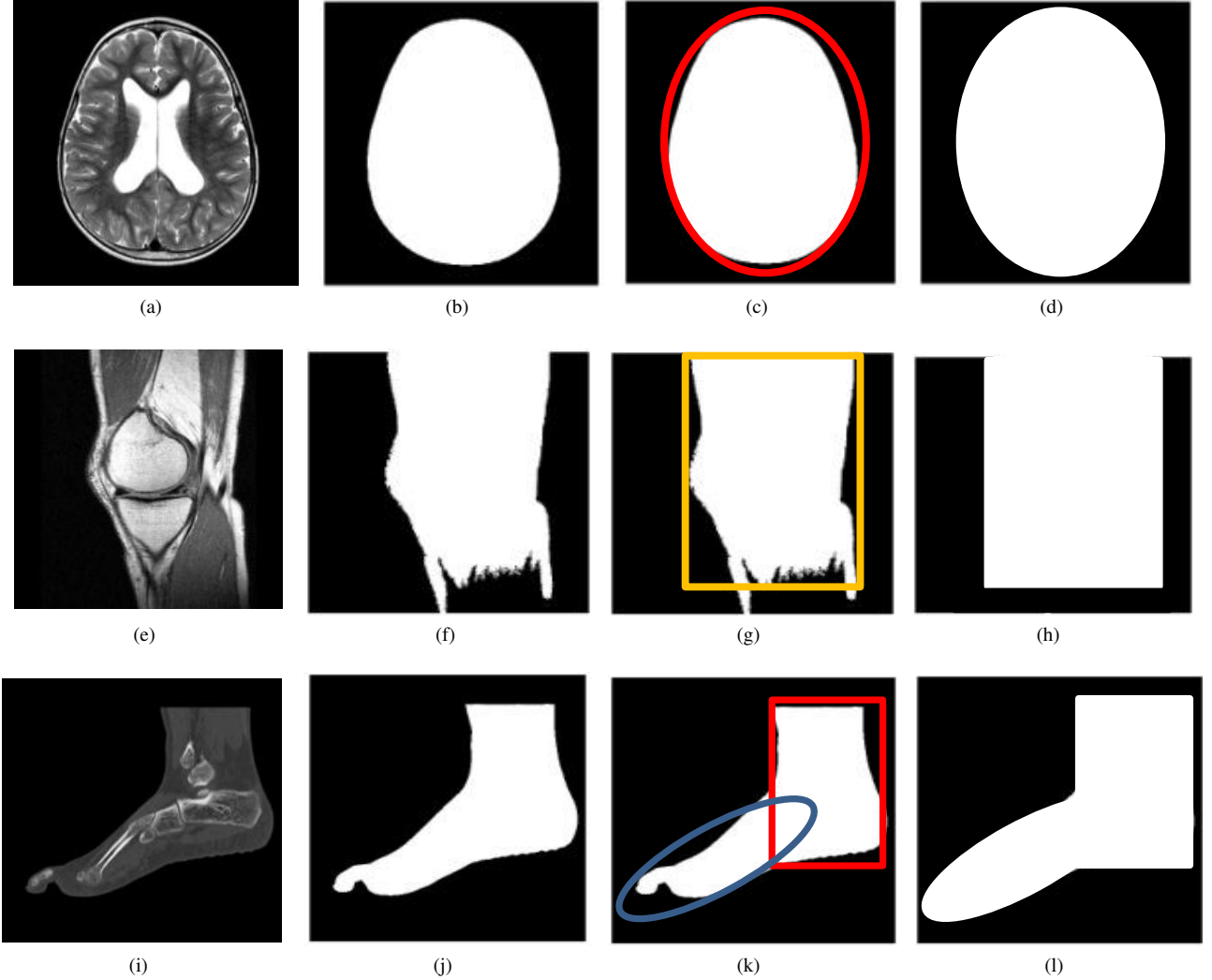


Fig. 2. Representation of segmented ROIs/ROIs using basic shapes for MRI head (1st row), MRI knee (2nd row), and CT leg (3rd row) images: (a,e,i) original images, (b,f,j) segmented image, (c,g,k) represented by basic shapes, and (d,h,l) ROIs/ROIs using basic shapes.

we initially verify this possibility by manually selecting the basic shapes that effectively capture the segmented regions. For example, if a segmented region of an image is captured by an ellipse and a rectangle, the intersected points and coordinates of the ellipse and rectangle are to be computed to exactly identify their covered region. These information are then required to be included in a structured way in the payload as a side information. Our continuing work will focus on the development of this automatic selection process, which is also pointed out in Sec. V.

IV. RESULTS AND ANALYSIS

In this section we present our early experimental results of different processes of our proposed segmentation model. The model has two main processes, one is the segmentation of MIs and the other is the description of segmented region using minimum possible information as discussed in Sec. III. We used all common modality test images for experiments, and a computer having an Intel Core i3 2.30 GHz processor with memory of 4.00 GB and Matlab version R2007b for necessary

simulation.

Table II shows the segmented output of medical images of four modalities using some popular segmentation techniques discussed in Sec. II. Here, different edge-based and Watershed-based segmentation techniques have been considered. We observed that Canny-edge-based and Watershed-based schemes give better results for different modality MIs like MR, CT, X-ray and Fluoroscopy images. Watershed segmentation technique however seems to be relatively more reliable than Canny edge based technique.

Additionally, we have shown the possible outputs of segmented region representation by covering segmented images with basic shapes in Fig. 2 (1st column). Here, the original image is segmented by the best possible segmentation technique (*i.e.*, Watershed in our case) and showed in Fig. 2 (2nd Column). The segmented images were then represented by basic shapes as showed in Fig. 2 (3rd Column). A consistent mask for the segmented region is then created by using the combination of the shapes and their regions were represented in Fig. 2 (4th Column).

The above results in Table II and Fig. 2 suggest that proposed model can be used for multimodality MI segmentation for digital watermarking purpose. Further development of the automatic generation of mask for segmented region representation may also be required, which will be carried out in our future research.

V. CONCLUSIONS

Automatic RONI segmentation has remained a highly challenging task which received a little or no attention yet, although it is desirable in medical image watermarking research. We have taken the first steps toward developing a general RONI segmentation technique for multimodality medical images, by presenting its conceptual model and its simulation. Our preliminary results show that the proposed model can be used to effectively segment the different modality medical images. The Watershed based segmentation seems to be promising to be used in that model for primary segmentation, which is later represented by a mask of combined basic shapes.

Our ongoing research aims to further develop the RONI description to be an automatic process in future.

REFERENCES

- [1] H. Nyeem, W. Boles, and C. Boyd, "Content-independent embedding scheme for multi-modal medical image watermarking," in *Biomedical Engineering Online*. Springer, 2015, vol. 14, no. 1, p. 7.
- [2] R. Pandey, A. K. Singh, B. Kumar, and A. Mohan, "Iris based secure nroi multiple eye image watermarking for teleophthalmology," *Multimedia Tools and Applications*, pp. 1–17, 2016.
- [3] G. Coatrieux, J. Montagner, H. Huang, and C. Roux, "Mixed reversible and roni watermarking for medical image reliability protection," in *Proc. 29th IEEE EMBS*. IEEE, 2007, pp. 5653–5656.
- [4] N. A. Memon and S. Gilani, "Nroi watermarking of medical images for content authentication," in *Proc. 13th IEEE INMIC*. IEEE, 2008, pp. 106–110.
- [5] V. Fotopoulos, M. L. Stavrinou, and A. N. Skodras, "Medical image authentication and self-correction through an adaptive reversible watermarking technique," in *Proc. 8th IEEE Int. Conf. BIBE*. IEEE, 2008, pp. 1–5.
- [6] H. Nyeem, W. Boles, and C. Boyd, "A review of medical image watermarking requirements for teleradiology," in *Journal of Digital Imaging*. Springer, 2013, vol. 26.
- [7] X. Guo and T.-g. Zhuang, "A region-based lossless watermarking scheme for enhancing security of medical data," *Journal of Digital Imaging*, vol. 22, no. 1, pp. 53–64, 2009.
- [8] H. Al-Dmour and A. Al-Ani, "A medical image steganography method based on integer wavelet transform and overlapping edge detection," in *Neural Information Processing*. Springer, 2015, pp. 436–444.
- [9] S. Zhang, T. Gao, and L. Gao, "A novel encryption frame for medical image with watermark based on hyperchaotic system," *Mathematical Problems in Engineering*, vol. 2014, 2014.
- [10] A. Al-Haj and A. Amer, "Secured telemedicine using region-based watermarking with tamper localization," *Journal of Digital Imaging*, vol. 27, no. 6, pp. 737–750, 2014.
- [11] X. Guo and T. . G. Zhuang, "A region-based lossless watermarking scheme for enhancing security of medical data," in *Journal of Digital Imaging*. Springer, 2009, vol. 22.
- [12] P. Tsai, Y.-C. Hu, and H.-L. Yeh, "Reversible image hiding scheme using predictive coding and histogram shifting," *Signal Processing*, vol. 89, no. 6, pp. 1129–1143, 2009.
- [13] H. Nyeem, W. Boles, and C. Boyd, "Digital image watermarking: its formal model, fundamental properties and possible attacks," in *EURASIP Journal on Advances in Signal Processing*. Springer, 2014, vol. 2014, no. 1, pp. 1–22.
- [14] —, "Utilizing least significant bit-planes of RONI pixels for medical image watermarking," in *Proc. DICTA'13*. Piscataway: IEEE, 2013, pp. 1–8.
- [15] J. Fridrich, M. Goljan, and R. Du, "Lossless data embedding for all image formats," in *Electronic Imaging 2002*. SPIE, 2002, pp. 572–583.
- [16] J. Tian, "Reversible data embedding using a difference expansion," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 8, pp. 890–896, 2003.
- [17] Z. Ni, Y.-Q. Shi, N. Ansari, and W. Su, "Reversible data hiding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 3, pp. 354–362, 2006.
- [18] A. Fylakis, A. Keskinarkaus, V. Kiviniemi, and T. Seppanen, "Reversible blind data hiding for verifying integrity and authenticating mri and x-ray images," in *Proc. 9th ISMICT*. IEEE, 2015, pp. 185–189.
- [19] K. Navas, S. A. Thampy, and M. Sasikumar, "Epr hiding in medical images for telemedicine," *International Journal of Biomedical Sciences*, vol. 3, no. 1, 2008.
- [20] F. Rahimi, H. Rabbani *et al.*, "A dual adaptive watermarking scheme in contourlet domain for dicom images," *Biomedical Engineering Online*, vol. 10, no. 1, pp. 1–18, 2011.
- [21] N. A. Memon, S. Gilani, and S. Qayoom, "Multiple watermarking of medical images for content authentication and recovery," in *Proc. 13th IEEE INMIC*. IEEE, 2009, pp. 1–6.
- [22] H. Al-Dmour and A. Al-Ani, "Quality optimized medical image information hiding algorithm that employs edge detection and data coding," *Computer Methods and Programs in Biomedicine*, vol. 127, pp. 24–43, 2016.
- [23] D. Biswas, P. Das, P. Maji, N. Dey, A. Das, and S. S. Chaudhuri, "Visible watermarking within the region of non-interest of medical images based on fuzzy c-means and harris corner detection," *Computer Science & Information Technology*, pp. 161–168, 2013.
- [24] R. Maini and H. Aggarwal, "Study and comparison of various image edge detection techniques," *International journal of image processing (IJIP)*, vol. 3, no. 1, pp. 1–11, 2009.

Performance Analysis of a Power Line Channel for Multipath Propagation

Mubdiul Islam Rizu¹, Jerin Tasnim², Sadat Hasan³ Rakin Irshad Alvi⁴,

Department of Electrical, Electronic and Communication Engineering

Military Institute of Science and Technology, Mirpur Cantonment, Dhaka.

E-mail: ¹mubdiulislam@yahoo.com, ²jerin.misteece@gmail.com, ³sadat.shadow@gmail.com

⁴rakin_irshaad@yahoo.com

Abstract—Power line communication is an evolving technology to support transmission of power and information simultaneously. However, noise, multipath propagation, fading etc. are the main barriers that causes the signal degradation in a power line channel. We have designed a power line coupler for power line channel. In this paper, we have developed and applied a mathematical model of power line and tested the power transfer capability of the power line using the designed coupler. The simulation results show that the couplers are well-suited with the line parameters and provide easy path for high frequency signal. The result also suggest that the performance can be improved by adjusting few transfer function parameters.

Keywords— *Exponent of the attenuation factor, attenuation parameters, weighting factor, PLN.*

I. INTRODUCTION

Power line communication (PLC) has recently attracted much attention to explore the possibility of simultaneously transmitting information & power through power lines [1]. These power lines are also known as power line carrier, power line digital subscriber line (PDSL) [2], mains communication, power line telecommunications, or power line networking [3]. The power line communication technologies capture a wide range of applications from home automation to internet access [4]. The economic and operational viability of using this technology however requires availability of low cost PLC equipment and solutions to be used with the existing power network infrastructure.

Multipath propagation means the propagation of the radio signals to the receiving antenna by two or more paths. This is mainly responsible for the multipath interference and phase shifting of the signal. Broadband transmission characteristics highly depend on multipath propagation considering the channel attenuation, length of the path, path loss etc [5]. Previously different statistical analysis and

simulation based on low voltage channel have been done on the basis of multipath propagation [6].

Transfer function is the ratio of the voltage on the load to the source voltage and the input impedance of the network. It is commonly used in the analysis of systems such as single-input single-output filters, typically within the fields of signal processing, communication theory, and control theory. By changing different parameters we can significantly control the amplitude of transfer function of a power line communication model [7].

II. System Model

Generally, a typical PLC system consists of traditional power line, bidirectional coupling circuits and line trap circuits. The coupling circuits acting as both transmitters and receivers are required for generation and recovery of the high frequency message signals. Here, the Differential Mode Coupling method is mainly used for coupling and decoupling procedure which also involves the live wire as one terminal and the neutral wire as another terminal [8]. The coupler is the heart of the PLC system. The low voltage coupler shown in Fig 1 is suitable for user end distribution network (220 V - 440 V). Live-neutral is connected to power line for the data transmission whereas data input side is connected to low voltage data source. Similar connection goes for receiving section. The transmitted data flows from data input side to live-neutral and for receiving data, it flows in reverse direction.

Transformer coupling circuits are extensively used in power line communication system for providing galvanic isolation from power line and it also act as limiter when saturation occurred by high voltage transient [9]. Here, the ferrite core transformer with turn's ratio 20:1 is used as it has fewer losses at higher frequency than other cores. Furthermore, the combination of series capacitors and parallel inductors forms a high pass filter with cutoff frequency 1 MHz. This filter mainly attenuate 50 Hz and other noise signals present

in power line but allow passing the message signals of higher frequency ranging from 5 MHz to 10 MHz. Moreover, for improving overall bypass effect, parallel resonant circuits can be implemented here.

k = Exponent of attenuation factor (0.5 to 1);
 d_i = Length of i -th path;
 τ_i = delay of i -th path;

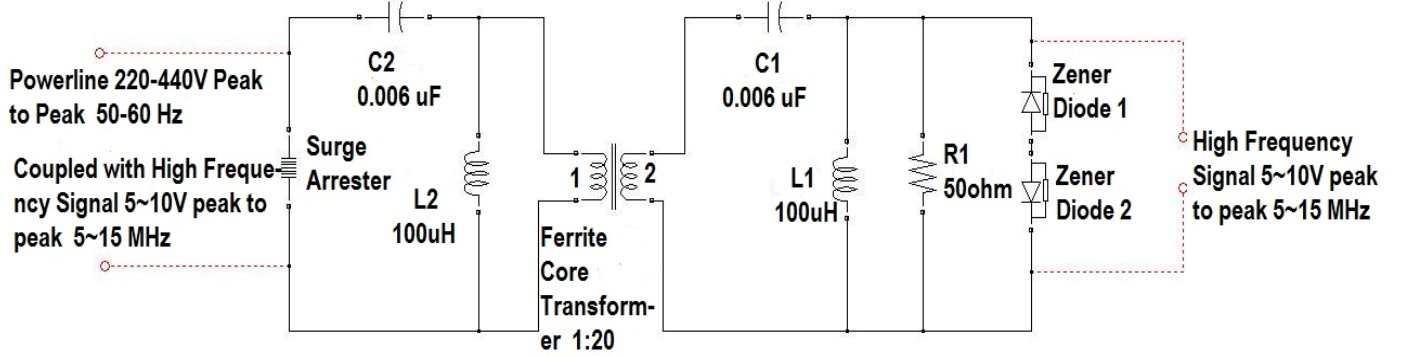


Fig 1: Low Voltage Bidirectional Coupler Circuit

IV. SIMULATION

At the secondary side of the transformer, back to back zener diodes are used to clamp the voltage level and limiting any high voltage that may creep into the data circuit. Also, a resistor is connected in parallel with the capacitor to discharge the capacitor for minimizing the hazard of high voltage peaks caused by the stored charge in capacitor. The metal oxide varistor (MOV) protects the circuit from any incoming surge from the power line providing the protection against very large transients on the power line [10].

The above equation (1) was simulated by MATLAB. Here it is observed that with the changing values of exponent of attenuation factor k , the attenuation parameters, a_0 & a_1 and weighting factor g_i , the corresponding amplitude of transfer function changes significantly.

At first in fig 2, taking the fixed value $a_1=1.0000005$ and $a_0=0.0000001$ we have varied k from 0.6 to 1. We have observed that with the increase of k , the corresponding amplitude of transfer function increases proportionally.

III. POWERLINE CHANNEL MODEL

Power line transfer function helps to signify about the channel behavior, the channel performance, different configurations and load conditions. We consider power line as a multipath radio channel for its transfer function. The multi-path model considers the channel as an unknown function which can be estimated by some analytical results. Based on the Transmission Line principle, it proposes a parametric model of the channel frequency response. This type of modeling approach is useful when the network topology is not available or component connections are desired. The multi-path model was first proposed in [11] and later augmented in [12]. It is mainly based on the superposition of signal reflections at various loads predicted by Transmission Line theory. Its parametric form is given by

$$H(f) = \sum_{i=1}^L g_i e^{-(a_0+a_1 f^k)d_i} e^{-j2\pi f \tau_i} \quad (1)$$

The above transfer function is for broadband power line in the frequency range of 0.3 to 100MHz where

L = No. of path;
 g_i = Weighting factor for path i (less than 1);
 a_0 & a_1 = attenuation parameter;
 f = frequency in Hz;

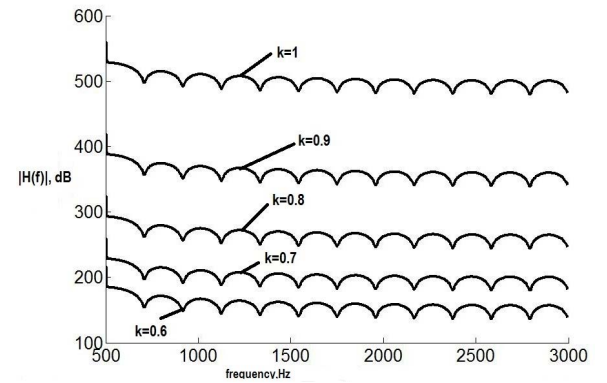


Fig 2: $H(f)$ vs. frequency with varying ' k '

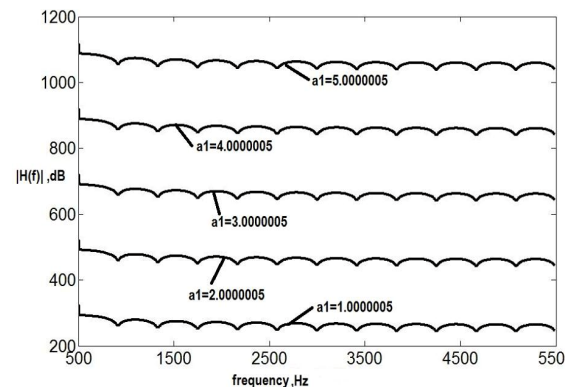


Fig 3: $H(f)$ vs frequency with varying a_1

In Fig 3, a_1 was varied from 1.0000005 to 5.0000005 and kept a_0 and k fixed. Here also the value of $H(f)$ increased proportionally with the increase of a_1 .

Similarly, in Fig 4, the value of a_0 was increased from 0.0000001 to 4.0000001 and kept a_1 and k fixed. Consequently, it is observed that the amplitude of transfer function rises proportionally with the rising value of a_0 . In Fig 5, the value of weighting factor, g_i is varied from 0.1 to 0.9 keeping the other parameters constant and it is observed that the change of $|H(f)|$ is inversely proportional to with the change in g_i .

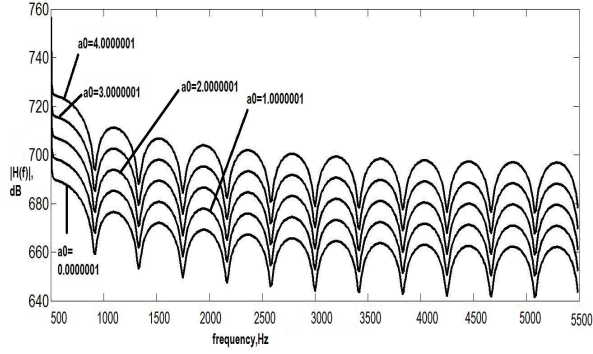


Fig 4: $H(f)$ vs frequency with varying a_0

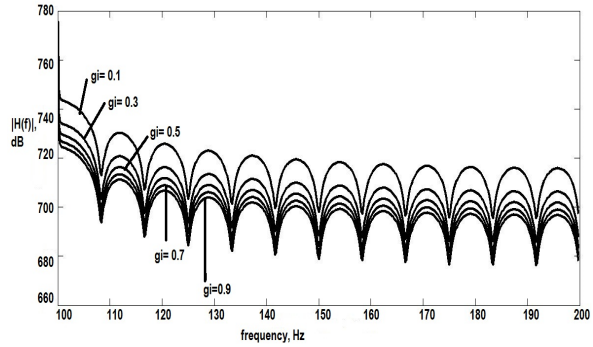


Fig 5: $H(f)$ vs frequency with varying g_i

V. RESULT & DISCUSSION

In this paper, the magnitude of transfer function for different values of exponent of the attenuation factor k , attenuation parameters a_0 and a_1 and different values of weighting factors of i -th path, g_i have been compared. For different values of k , a_0 , a_1 and g_i , the magnitude changes significantly. For example, when $k = 0.6$, $|H(f)| = 170$ and when $k = 1$, $|H(f)| = 540$. Again, for $a_1 = 1.0000005$, $|H(f)| = 290$ and for $a_1 = 5.0000005$, $|H(f)| = 1200$. Similarly, at the value of $a_0 = 0.0000001$, $|H(f)| = 680$ and at $a_0 = 4.0000001$, $|H(f)| = 730$. Again, when $g_i = 0.1$, $|H(f)| = 744$ and when $g_i = 0.9$, $|H(f)| = 725$.

The data clearly indicates that the value of k , a_0 , a_1 and g_i play very important role in the improvement of transfer characteristics of power line communication. With the gradual increase of the attenuation factor and parameters, we get better performance model as the amplitude of transfer function increases proportionally with these parameters. On the contrary, the change of magnitude of transfer function is inversely proportional to the change of weighting factor of i -th path.

In Fig 6, the improvement curve of $H(f)$ with the change of ' k ' has been shown. The curve clearly indicates almost the linear increase in the value of $|H(f)|$ with the rise of ' k '. Similarly, in Fig 7, 8 and 9 the improvement curves of $|H(f)|$ against the value of ' a_1 ', ' a_0 ' and ' g_i ' are shown.

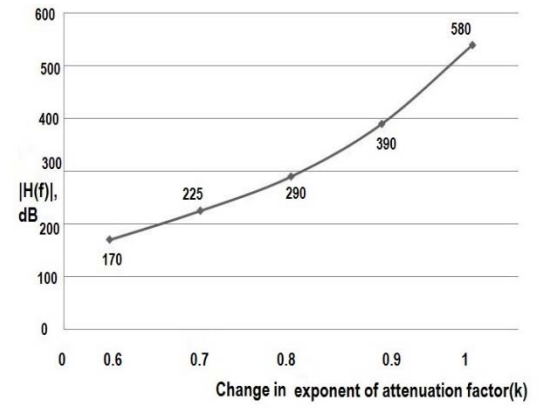


Fig 6: Improvement Curve against ' k '

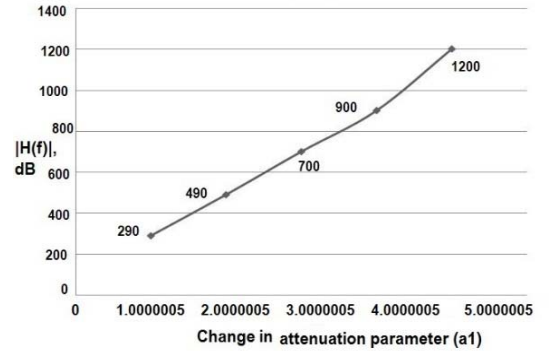


Fig 7: Improvement Curve against ' a_1 '

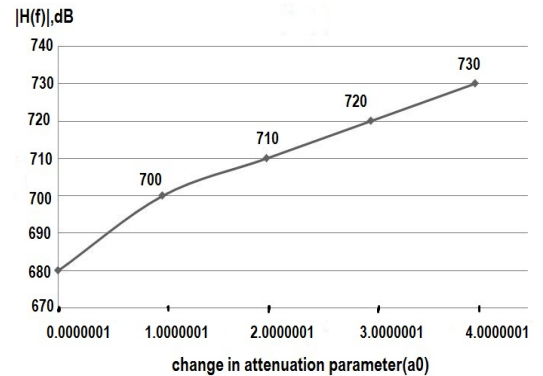


Fig 8: Improvement Curve against ' a_0 '

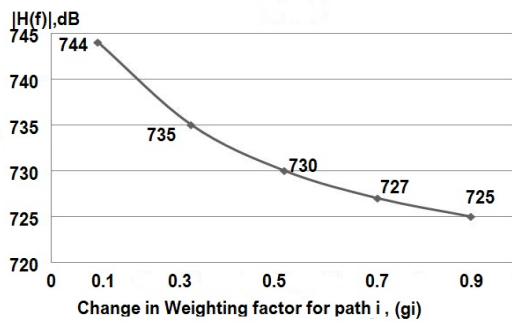


Fig 9: Performance Curve against 'g,'

VI. CONCLUSION

This paper presents a method for improvement of the magnitude of transfer function with the increasing value of exponent of the attenuation factor, attenuation parameters and weighting factor of i -th path. The proposed coupler is well suited for transmitting high frequency communication signals. Impedance mismatching problem between the coupler impedance and power line impedance is also solved using line trap circuit. This coupler can also be effectively used for broad band communication in domestic as well as distribution power networks.

Further improvement and analysis of the transfer function based model may be carried out in future for reducing low noise and attenuation.

VII. REFERENCES

- [1] Dostert, K (1997). "Telecommunications over the Power Distribution Grid- Possibilities and Limitations" Proc 1997 Internat. Symp. on Power Line Comms and its Applications: 1-9.
- [2] Proko, T. ; Vodrazka, J., "Modeling of transmission lines for digital subscriber lines testing"; 15th International Conference on Systems, Signals and Image Processing, (IWSSIP) 2008.; Publication Year: 2008 , Page(s): 53 – 56.
- [3] Perez, A., Sanchez, A.-M., Regue, J.-R., Ribo, M., Aquilue, R., Rodriguez-Cepeda, P. Pajares, F.-J., " Circuital and Modal Characterization of the Power-Line Network in the PLC Band" IEEE Transactions on Power Delivery, Volume: 24 , Issue: 3; Publication Year: 2009, Page(s): 1182 – 1189.
- [4] H. Hrasnica, A. Haidine, and R. Lehnert, Broadband Powerline Communications: Network Design, 2005.
- [5] Zheng Tao; Yang Xiaoxian; Zhang Baohui, "Broadband transmission characteristics for power-line channels" IEEE Transactions on Power Delivery, Year: 2006, Volume: 21, Pages: 1905 – 1911.
- [6] I. C. Papaleonidopoulos; C. N. Capsalis; C. G. Karagiannopoulos; N. J. Theodorou, "Statistical analysis and simulation of indoor single-phase low voltage power-line communication channels on the basis of multipath propagation" IEEE Transactions on Consumer Electronics, Year: 2003, Volume: 49, Pages: 89 – 99.
- [7] T. Banwell and S. Galli, "A new approach to the modeling of the transfer function of the power line channel," Proc. Int. Symp. Power Line Communication. and its Appl., pp. 319-324, 2001.
- [8] Sanya T.O., Hove M., Snyders A.J., Ferreira H.C. "Surge protection of communication equipments for power line communication: Effects on communication signal" AFRICON, 2011; Publication Year: 2011, Page(s): 1 – 4.
- [9] P.A.J van Rensburg, "Design of a Bidirectional Impedance-Adapting Transformer Coupling Circuit for Low- Voltage Power-Line Communications," IEEE Transactions on Power Delivery, Vol. 20, No. 1, 2005, pp. 64-70.
- [10] "IEEE Draft Standard for Test Methods for Surge Protectors Used in Low-Voltage Data, Communications, and Signaling Circuits"; IEEE PC62.36/D16, October 2013; Publication Year: 2013, Page(s): 1 – 130.
- [11] H. Philipps, "Modeling of power line communication channels," Proc. Int. Symp. Power Line Communication. And Its Appl, pp. 14-21, 1999.
- [12] M. Zimmermann and K. Dostert, "A multipath model for the powerline channel," IEEE Trans. Commun., vol. 50, no. 4, pp. 553-559, 2002.

PITCHY: A Pitch Controller to Minimize Frequency Fluctuation in Hybrid Power Systems

Shafinaz Islam, Shabnam Ferdous and MD Rafiqul Islam Sheikh

*Electrical and Electronic Engineering
Rajshahi University of Engineering and Technology
Rajshahi, Bangladesh*

{shafinaz.orthy16, shabnam.shimuleee101067}@gmail.com, ris_ruet@yahoo.com

Abstract— Wind energy is widely considered as a major source of clean energy. As such, wind turbines are becoming popular in the renewable energy world. The output of the wind turbine varies with randomly changing speed of the wind. This is a critical problem for wind turbines in general, and fixed speed wind generators in particular. However, conventional pitch controllers and governor control systems are not capable enough to keep the power system frequency to the expected stage. By using variable reference line power in pitch controller of wind turbine, frequency fluctuation can be minimized significantly. In this paper, we propose PITCHY, a novel pitch controller to minimize output frequency of wind turbines. PITCHY minimizes frequency fluctuation by generating reference power using low pass filter (LPF). We design a simulation model and evaluate PITCHY using PSCAD/EMTDC. We used a number of wind (induction) generators having varying governor control systems in our simulation. The results show that using PITCHY power system frequency fluctuation can be varied from 49.9 to 50.1 Hz.

Keywords— Wind energy; Frequency fluctuation; Hydro Generator; Load Frequency Control Model; Line power reference

I. INTRODUCTION

Consumption of energy has a direct consequence on how people live in a society. Study shows that the higher the consumption of energy in a country per capita, the higher the quality of life of its people [1]. The U.S. Energy Information projects an astonishing 48% increase in world energy consumption by 2040. The world tries to meet up the increasing demand of energy by means of coal, natural gas, oil, biomass, wind and solar. However, not all the energy sources are environment friendly. For example, coal and other fossil fuels have been recognized as sources of global warming. As an alternative, renewable and clean energy resources like wind energy is becoming popular. In 2015, EU alone installed and grid-connected 12,800 MW of wind power [15].

One potential problem related to wind generators is frequency fluctuation. A pitch control system generally works as a controller of the output power of the wind generator at rated value; at the time wind speed is between the cut out speed and the rated speed. Furthermore, when the speed of the wind is within the rated speed and cut speed, the angle of the blade pitch is kept at zero degree as a constant value; and this

enables the wind turbine to produce maximum power [3]. One disadvantage of the wind power generator is that high wind power penetration has an impact on the stability of the power system as the output of the wind turbine is proportional to the third power of wind speed [2]. As the uncertainty rated capacity of the wind turbine rises, the output of the wind generator influences power system frequency [3].

Some existing works proposed techniques to minimize frequency fluctuations in hybrid power systems. One study [10] considered only thermal governor, but it failed to minimize frequency fluctuation to a significant level. Also, some authors [8-9] have used expensive energy storage devices to minimize frequency fluctuation.

In this paper, we propose PITCHY, a novel technique to reduce the frequency fluctuation to a considerable level. In our proposed scheme we have used a new pitch controller in which reference line power is variable. Low pass filters generated this reference power. We considered significant wind farm total capacity and we have considered the thermal governor and the hydro governor. Our results show that by minimizing frequency fluctuation significantly, it is possible to increase the total generation capacity up to 20% of the total capacity.

We organize the paper as follows. Section II introduces methodology. In this section, we discuss the system model for the simulation, models for synchronous generators and our proposed pitch controller PITCHY. Section III presents our simulation results. We review related work in Section IV and conclude the paper in Section V.

II. METHODOLOGY

In this section, we first discuss the model system for the simulation (section A). Then we present our proposed approach PITCHY (section B).

A. Model System For Simulation Analysis

Figure 1 shows the whole system model that is used in the simulation studies. In this novel system, a total number of five Induction Generators are used as wind generators, four synchronous generators (SG) which are used as thermal (TG) and hydro (HG) generators, and two loads are also used. SG1 indicates hydro generator where SG2 and SG3 indicate thermal generators. Two control modes as governor free (GF) for SG1 and SG2 and load frequency control (LFC) for SG3 are also used. To control the combined and individual operations of SGs with operating modes, another switch component is used. In the figure, S1, S2, S3, and S4 denote the switches. In the

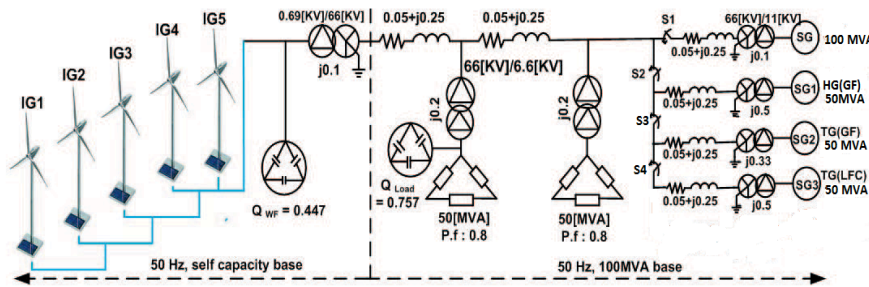


Fig. 1: System Model

following we provide details of the components.

Synchronous Generator Model

Governor:

The basic function of a governor is to automatically adjust turbine's speed of rotation and the output of the generator [2]. For constant load condition of generator, the turbine drives at an unchanged rotating speed. Conversely, for variation in load, there is an imbalance between the generator production and load and as a result, rotating speed varies. And in absence of the load, the rise of the rotating speed is sensed by the governor and an anomalous increase of the speed is restricted because of closing the valve quickly.

Governor for Synchronous Generators:

The governor models [4] shown in Fig. 2 and Fig. 3 are considered in simulation analyses. The values of 65M and 77M for hydro and thermal generators are shown in Table 1 and Table 2 respectively.

In the figures, Sg denotes the revolution speed deviation [pu] and 65M denotes the initial output [pu]; 77M: the load limit ($65\text{M} + \text{rated MW output} \times \text{PLM} (\%)$); PLM: the spare governor operation (%); Pm: the turbine output [pu]; $\text{PLM} > 0$ For governor-free (GF) operation; $77\text{M} = 65\text{M} + \text{rated MW output} \times \text{PLM} (\%)$, 65M and 77M are the initial output [pu]. Sg is kept 0 for SG2 because this generator is governed under LFC to regulate frequency fluctuations with a comparatively longer interval.

Table 1: Standards of Governor (Hydro)

	IG: 10MVA	IG: 20MVA
65M[pu]	0.72	0.68
77M[pu]	0.7733	0.751
PLM[%]	10	20

Table 2: Standards of Governor (Thermal)

	IG: 10MVA	IG: 20MVA
65M[pu]	0.72	0.68
77M[pu]	0.82	0.78
PLM[%]	10	20

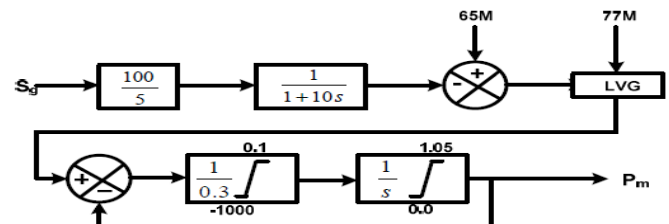


Fig. 2: Hydro governor model

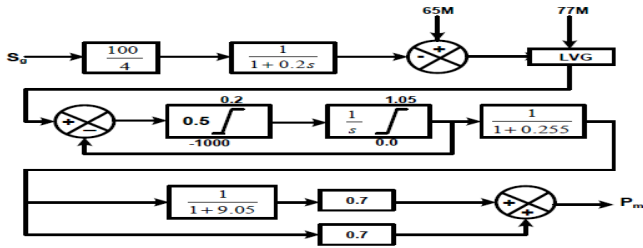


Fig. 3: Thermal governor model

Control Models for Load Frequency:

Load Frequency Control (LFC) primarily controls frequency fluctuations in a longer interval of time. When there is a deviation in the power system frequency, the control output signal of Load Frequency Control (LFC) is sent to the LFC power plant. Then, LFC power plant's output varies along with the signal of LFC. And Low Pass Filter (LPF) is considered to remove short interval fluctuations. Figure 4 shows the LFC model used in this study, where, LFC period, $T_c = 200[s]$; damping ratio is 1; LFC frequency = $1/T_c = 0.005[Hz]$;

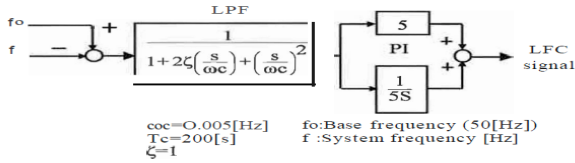


Fig. 4: Load frequency control model

B. PITCHY: A New Pitch Controller

In this paper, simulations have been performed by previously used pitch controllers and newly proposed pitch controller, PITCHY. And the results of these simulations have been compared. Variations in wind speed is not suitable for wind generators. And when the speed of wind is imbalanced with rated speed, a pitch controller is required. At this imbalanced condition of wind speed, pitch controller controls the blade pitch angle of turbine blade. And thus Wind generator output power is maintained at an acceptable level. Again, obtaining adequate output power with a very lower rated wind speed is challengeable by using constant output power reference. So, to minimize the frequency fluctuations as well as to get sufficient power output, in our new pitch controller, variable reference line power is generated by low pass filters. Figure 5 shows the proposed pitch controller.

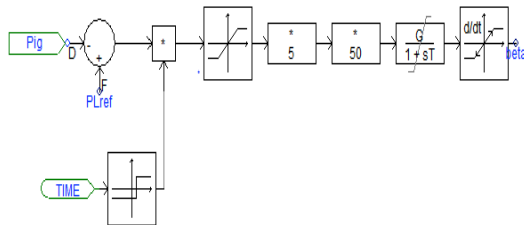


Fig. 5: PITCHY pitch controller

Here,

$$G(\text{Gain})=1, T=5[s]$$

Now we describe the generating process of the reference line power, $P_{\text{Line-ref}}$, using Low Pass Filter (LPF).

Figure 6 shows the generating process of the reference line power, $P_{\text{Line-ref}}$, using Low Pass Filter (LPF). By this process, it can be concluded that fluctuation in wind power is inversely proportional with the parameter, Time constant (T) of the low pass filter. System frequency fluctuation which is influenced by wind generator output fluctuations can be reduced by compensating the transmission line power, P_{IG} along with the reference value.

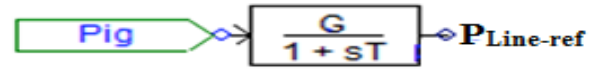


Fig. 6: $P_{\text{Line-ref}}$ generated by LPF

Using mathematical notations the first-order low pass filter can be expressed as follows:

$$P_{\text{IG}} = P_{\text{Lref}} + T * P'_{\text{Lref}} \dots\dots\dots(i)$$

Here, $P_{\text{Line-ref}} = P_{\text{Lref}}$

Where T represents the filtering time constant, P_{Lref} represents the output function of the filter related to the wind turbine, P'_{Lref} is the derivative of P_{Lref} and P_{IG} denotes the filter input function. Now, if we apply a discrete data having Δt time step to a low pass filter and if we expand the derivative of P_{Lref} into a discrete form, then for step k, the equation can be expressed as the following:

$$T \frac{P_{\text{Lref},k} - P'_{\text{Lref},k-1}}{\Delta T} + P_{\text{Lref},k} = P_{\text{IG},k} \dots\dots\dots(ii)$$

Solving for P_{Lref} gives

$$P_{\text{Lref},k} = \frac{T}{T+\Delta T} P_{\text{Lref},k-1} + \frac{T}{T+\Delta T} P_{\text{IG},K} \dots\dots\dots(iii)$$

Defining a constant, $\beta = \frac{T}{T+\Delta T}$, equation (iii) can be written as

$$P_{\text{Lref},k} = \beta P_{\text{Lref},k-1} + (\beta-1) P_{\text{IG},K} \dots\dots\dots(iv)$$

Eventually the equation represents itself as an exponentially weighted moving average filter. For example, we can write as $t_k = t_0 + k\Delta t$, where the subscript k relates to time and Δt is the the starting point and the time step.

III. SIMULATION RESULTS

The system has been simulated and the results have been calculated using both conventional pitch controller and the pitch controller using reference line power generated by low pass filter. Calculations have been done for the cases listed in Table 3 to understand the power system frequency fluctuations.

Table 3: Simulation Analysis

Cases	Governor & Frequency Control	IG Capacity (MVA)	Condition of Switches			
			S1	S2	S3	S4
Case 1	TG-GF	10	ON	OFF	OFF	OFF
Case 2	HG-GF		ON	OFF	OFF	OFF
Case 3	TG-LFC		ON	OFF	OFF	OFF
Case 4	HG-GF + TG-GF		OFF	ON	ON	OFF
Case 5	HG-GF + TG-LFC	20	OFF	ON	OFF	ON
Case 6	TG-GF		ON	OFF	OFF	OFF
Case 7	HG-GF		ON	OFF	OFF	OFF
Case 8	TG-LFC		ON	OFF	OFF	OFF
Case 9	HG-GF + TG-GF		OFF	ON	ON	OFF
Case 10	HG-GF + TG-LFC		OFF	ON	OFF	ON

Figs. 7-11 show the response of power system frequency for total IG capacity 10MVA using conventional pitch controller and pitch controller with reference line power generated by low pass filter. For case 1 and 2, single hydro and single thermal governor with governor free control mode is used respectively with a capacity of 100 MVA. For case 3, single thermal governor with load frequency control mode is used with a capacity of 100 MVA. For case 4, both hydro (50 MVA) and thermal (50 MVA) governor with governor free control mode are used with a total capacity of 100 MVA. For case 5, both hydro governor (50 MVA) with governor free control mode & thermal governor (50 MVA) with load frequency control mode are used with a total capacity of 100 MVA.

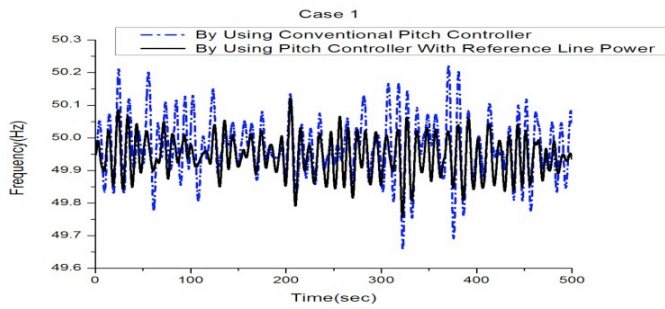


Fig. 7: Frequency of the power system vs. time

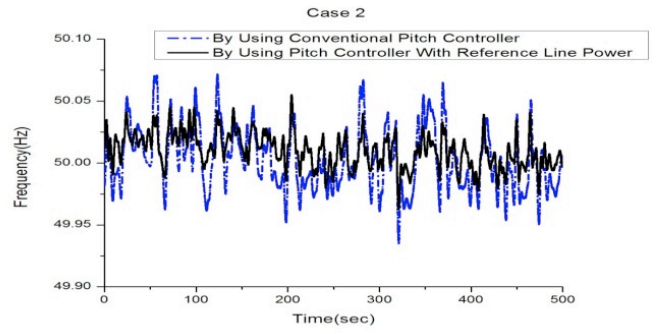


Fig. 8: Frequency of the power system vs. time

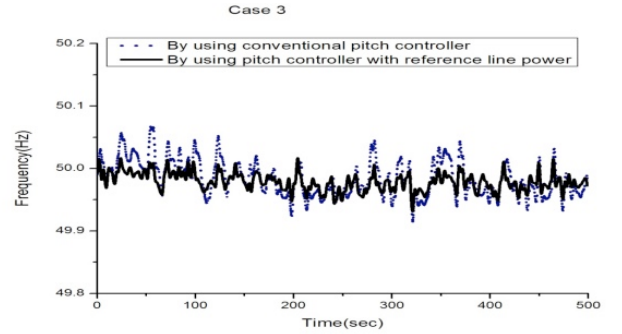


Fig. 9: Frequency of the power system vs. time

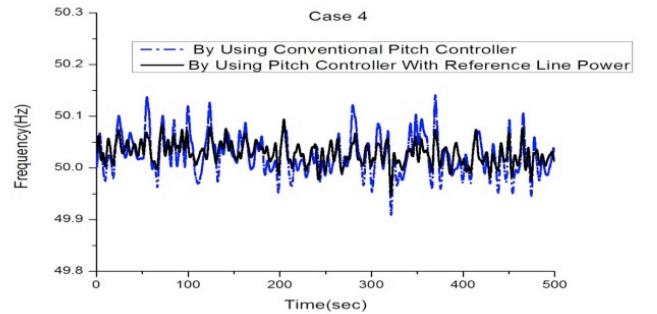


Fig. 10: Frequency of the power system vs. time

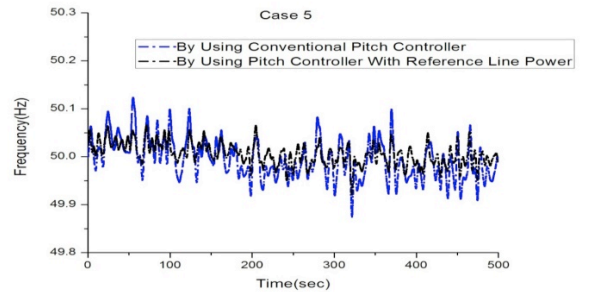


Fig. 11: Frequency of the power system vs. time

Above figures show that for 10% of IG capacity fluctuation of frequency reduced. For fig.7 frequency fluctuates between (49.8-50.2) Hz and for cases 2-7 frequency fluctuates between 49.95-50.01, which is reasonably very low range.

Figures 12-16 show the response of power system frequency for total IG capacity 20MVA with conventional pitch controller and pitch controller with reference line power. For 20 MVA, simulation schemes are similar to 10 MVA as cases 6, 7, 8, 9, 10 are similar to cases 1, 2, 3, 4, 5 respectively.

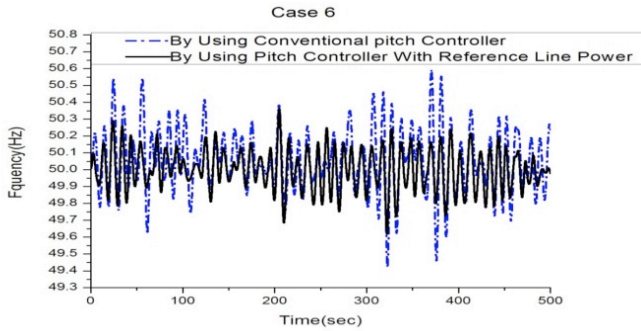


Fig. 12: Frequency of the power system vs. time

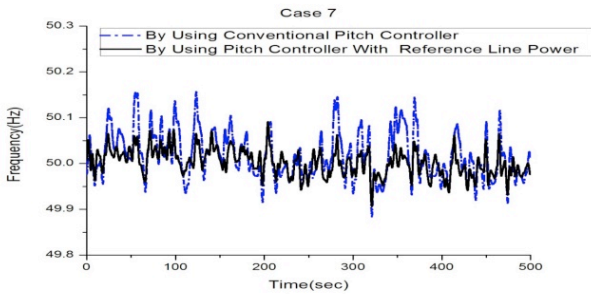


Fig. 13: Frequency of the power system vs. time

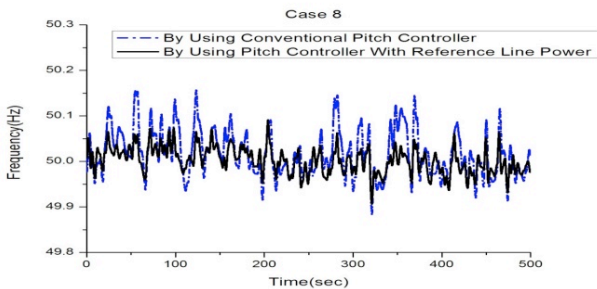


Fig. 14: Frequency of the power system vs. time

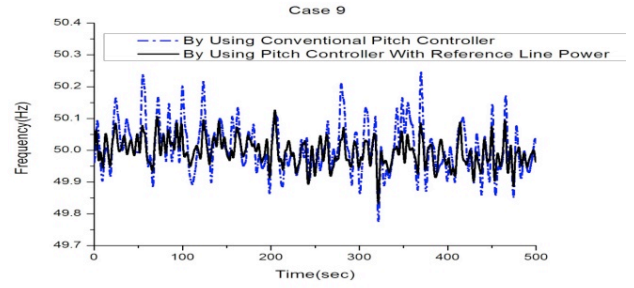


Fig. 15: Frequency of the power system vs. time

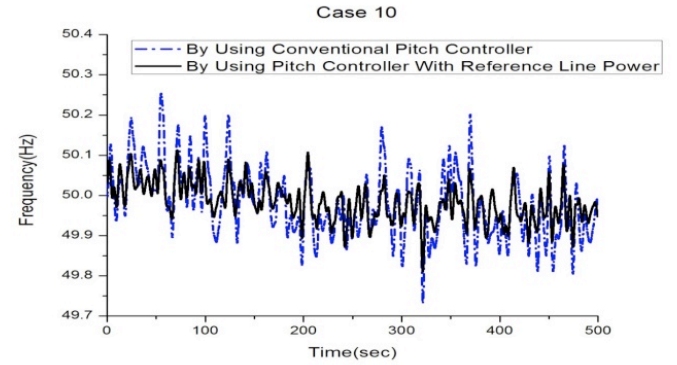


Fig. 16: Frequency of the power system vs. time

Above figures show that for both 10% and 20% of IG capacity fluctuation of frequency reduced significantly. For fig.7 frequency fluctuates between (49.8-50.2) Hz and for cases 2-7 frequency fluctuates between 49.95-50.01, which is reasonably very low range. In fig.12 it can be shown that frequency fluctuates less when our new scheme is used. And for fig 13-16 frequency fluctuates between (49.9-50.1) Hz. So, in most of the cases, the fluctuation range is (49.9-50.1) Hz.

Fig.17 shows the comparison between the output performances of 5 different IGs with and without reference line power generated by low pass filter. It can be shown from fig.17 that the output power fluctuates more when we didn't use reference line power. Hence, by using reference line power, the fluctuation range of the output performances can be minimized.

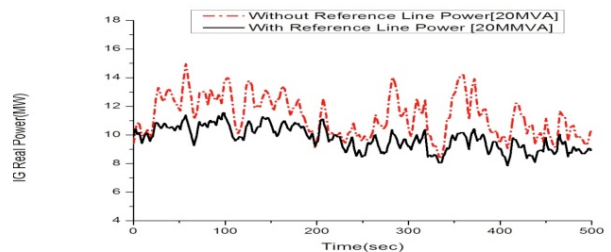


Fig. 17: IG real power vs. time

IV. RELATED WORK

A number studies ([6] - [10]) have investigated frequency fluctuations of wind generators.

In ([6] - [7]), the authors proposed the thermal governor as the most efficient frequency controller, but its frequency controlling ability is limited to a lower range of wind generator capacity. Also, authors in ([8] - [9]) suggested some devices which stores energy to reduce voltage, frequency and power fluctuations. Yet, this strategy is costly. Additionally, another technique for controlling is expected for these devices.

The authors in [10] used thermal governor for its better control ability of frequency. They generated reference power for the pitch controller using another method named moving average. But it is unable to maintain frequency to the acceptable limit.

V. CONCLUSIONS

In this paper, we study the problem of power system output frequency fluctuation. To minimize the frequency fluctuation, we propose PITCHY. Using PITCHY we show that frequency deviation can be reduced significantly for both 10% and 20% capacity of IG power when LPF is used for reference power generation with multiple IGs and different types of governor control models with different types of operating modes. As widely varying wind data is used, the fluctuation of output power and terminal voltage were quite noticeable. It is observed from this study that frequency varies between (49.9-50.1) Hz for 20% capacities of IG power. We show that frequency deviation is 0.2 Hz, which is reasonably compatible in wind power applications.

REFERENCES

- [1] Global energy trends to 2040, Global Energy Outlook 2014.
- [2] E. Haque and M.R.I Sheikh, "Grid frequency analysis with the issue of high wind power penetration," International Conf. on Electrical Information and Communication Technology, KUET, Khulna, Bangladesh, 13-15 February 2014.
- [3] M.R.I. Sheikh, M.M. Haque, and Hossain, "Performance of governor system on minimizing frequency fluctuations with wind power generation," International Journal of Electrical and Computer Engineering, vol. 2, no. 1, pp. 46--56, February 2012.
- [4] C. Lou and B.-T. Poi, "Frequency deviation of thermal power plants due to wind farms," IEEE Transl. Enter. Cong, vol. 21, pp. 708-716, 21 August 2006.
- [5] M.R.I. Sheikh, M.S.R. Ashraf, and D. Data, "Co-ordinated Frequency Control of Power System with High Wind Power Penetration", International Conference on Electrical Engineering and Information & Communication Bangladesh, 12-14 Apr. 2014. MIST, Dhaka.
- [6] 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.
- [7] A. Murakami, A. Yokuyama, and Y. Tada, "Basic study on battery capacity evaluation for frequency control in power system with a large penetration of wind power generation," IEEJ Transl. Japan, vol. 126, pp. 236-242, 2006.
- [8] S.M. Mueen, S. Shishido, M.H. Ali, R. Takahashi, T. Murata, and J. Tamura, "Application of energy capacitor system (ECS) to wind power generation," Wind Ener., vol. 11, pp. 335-350, July/August 2008.
- [9] M.R.T. Sheikh, S.M. Mueen, Rion Takahashi, Toshiaki Murata and Junji Tamura, "Minimization of Fluctuations of Output Power and Terminal Voltage of Wind Generator by Using STATCOM/SMES", 2009 IEEE Bucharest PowerTech Conference, June 28th-July 2nd 2009, Bucharest, Romania.
- [10] Takuya Yamazaki, R. Takahashi, Toshiaki Murata, Junji Tamura, Tomoyuki Fukushima, Eiichi Sasano, Koji Shinya, and Takatoshi Matsumoto "Smoothing Control of Wind Generator Output Fluctuations by New Pitch Controller", IEEE Transactions on Power and Energy, Vol. 129, No. 7, pp. 880-888, (2009-7).
- [11] S.M. Mueen, S. Shishido, M.H. Ali, R. Takahashi, T. Murata, and J. Tamura, "Application of energy capacitor system (ECS) to wind power generation," Wind Energy, vol. 11, pp. 335-350, July/August 2008.
- [12] M. S. R. Ashrafi, D. Datta and M. R. I. Sheikh, "Influence of governor control systems on the stability of a wind farm integrated power system," Strategic Technology (IFOST), 2014 9th International Forum on, Cox's Bazar, 2014, pp. 309-312.
- [13] M. R.I. Sheikh, "Stabilization of a Grid-Connected Wind Farm by Using SMES", PhD Thesis, KIT, Japan, Sep 2010.
- [14] PSCAD/EMTDC Manual, Manitoba HVDC Research Center. 1994.
- [15] <https://windeurope.org/about-wind/statistics/>

Designing of a Non-Zero Dispersion Shifted Fiber with Ultra-High Birefringence and High Non-Linearity

Shabbir Chowdhury

Dept. of Electrical and Electronic Engineering
Bangabandhu Sheikh Mujibur Rahman Science and
Technology University
Gopalganj, Bangladesh
shabbirchowdhury01@gmail.com

Japatoosh Mondal

Dept. of Electrical and Electronic Engineering
Bangabandhu Sheikh Mujibur Rahman Science and
Technology University
Gopalganj, Bangladesh
jmondal.eee@bsmrstu.edu.bd

Abstract— This paper proposes a design of a photonic crystal fiber (PCF) with ultra-high birefringence (0.0262 at 1550nm) which is more useful for sensor based on fiber optics. The finite element method (FEM) is used for simulation with $0.6\ \mu\text{m}$ as the perfectly matched boundary layer (PML). This proposed design will offer the characteristics of nonzero-dispersion-shifted fiber (NZ-DSF) for 550nm waveband with very high non-linearity ($50.86\ \text{W}^{-1}\text{km}^{-1}$) at 1550nm.

Keywords— Chromatic Dispersion, Birefringence, NZ-DSF, FEM, Non-linear coefficient.

I. INTRODUCTION

Photonic crystal fiber is the media to convey guided information in the form of light using total internal reflection (TIR) method for communication, medical, astronomical or any other purposes. Generally fused silica is used as core that fenced by a systematic arrangement of air holes over the fiber span [1]. It is very much possible to achieve high birefringence [2], high non-linearity [3], controlled operation [4], large chromatic dispersion [5] and large mode area [6] by tailoring air holes diameter and hole-to-hole distance which is inconvenient in conventional fibers.

High birefringence photonic crystal fiber has become very popular for polarization maintaining properties which is suitable for application of sensor based on optical fiber like as temperature-insensitive polarimetric vibration sensor [7], sensor for lateral pressure monitoring applications [8], seawater temperature measurement [9] and so on. High birefringence can be achieved by growing the difference between the real values of two refractive indices.

In photonic crystal fiber, material dispersion and waveguide dispersion are off opposite sign so proper design can cancel each other out and we can have zero dispersion or near-zero dispersion or NZ-DSF by using this property at fiber operating wavelength. But some factors such as dispersion compensation ratio, fiber bandwidth, fiber mode property should be kept in mind when a fiber is being designed.

In this paper, a design of PCF has been suggested consisting of circular air holes and two ellipses as the fiber cladding. Silica sand has been used as the core. This design will offer ultra-high birefringence, NZ-DSF for large wavelength range with high non-linearity. The high non-linearity can diminish the footprint and power consumption of various optical mechanism. This NZ-DSF can reduce the cross-talk of the wavelength division multiplexing (WDM) of the dispersion shifted fiber.

II. DESIGN METHODOLOGY

Fig.1 illustrates the structure of the proposed design. The fiber is of ring main design consisting of circular air holes of $0.62\ \mu\text{m}$ diameter and two ellipses with semi-axis $0.25\ \mu\text{m}$ and $0.85\ \mu\text{m}$, both are located at $0.6\ \mu\text{m}$ distances from the center of the core. The first ring radius R_1 of the air holes is $1.5\ \mu\text{m}$ and rests of four are as the following manner. $R_{n+1} = (R_n + 0.9)\ \mu\text{m}$, where $n=1, 2, 3, 4$. The first, second, third, fourth and fifth rings have 12, 18, 24, 30, and 36 circular air holes respectively with the same radius.

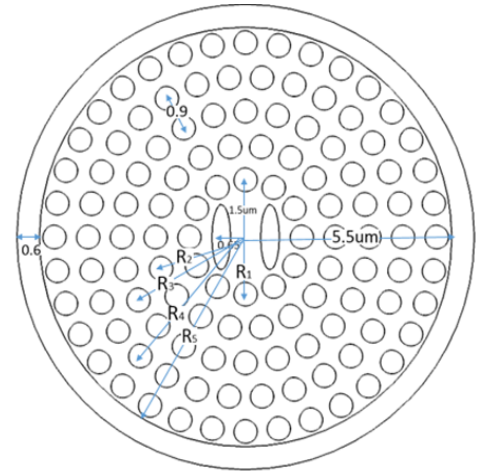


Fig. 1. Cross section of proposed PCF

The air holes are arranged at the same angular distance from one another in the same ring. The last two diameters have

selected for lowering the confinement loss. The difference $0.6 \mu\text{m}$ between the last two diameters is used as PML

III. NUMERICAL ANALYSIS

The proposed photonic crystal fiber is modeled by using finite element method (FEM) software (COMSOL Multiphysics version 4.2). The birefringence of the photonic crystal fiber is calculated from the variance between the effective refractive indices of x and y axis.

$$B = |n_x - n_y| \quad (1)$$

where, n_x and n_y represent the real parts of x and y axis refractive index respectively. We can have the refractive index from wavelength by using the sellmeier equation.

Effective area (A_{eff}), chromatic dispersion (D) and nonlinearity (γ) are calculated by the following equations:

$$A_{\text{eff}} = ((\iint (|E|^2 dx dy)^2) / (\iint (|E|^4 dx dy))) \quad (2)$$

$$D = -\lambda/c \cdot d^2 \text{Re}(n_{\text{eff}}) / d\lambda^2 \quad (3)$$

$$\gamma = 2\pi n_2 / (\lambda A_{\text{eff}}) \quad (4)$$

where c is the light velocity in vacuum, λ is the wavelength and n_2 is the nonlinear refractive index of silica sand.

Fig.2 represents the birefringence changes with respect to wavelength variation. The birefringence at 1550nm wavelength is 0.0262. This high birefringence [10] characteristic makes the fiber useful for sensor application where light is maintained a linear polarization state [11]-[12].

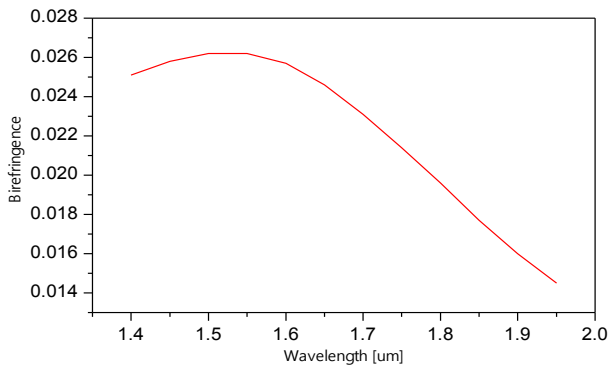


Fig. 2. Wavelength Vs Birefringence

Fig.3 shows the effective area and non-linear coefficients respectively. The effective area and fiber non-linear coefficient at 1550nm wavelength is $1.8331 \mu\text{m}^2$ and $50.86 \text{ W}^{-1} \text{ km}^{-1}$ respectively. The effective area is very small [13]-[14] which

will reduce the bending loss [11], because small mode areas strongly guide the light through the core, where bend losses and other effects of external disturbances are weak. The non-linearity is very high compared to recently published papers which are shown at Table I.

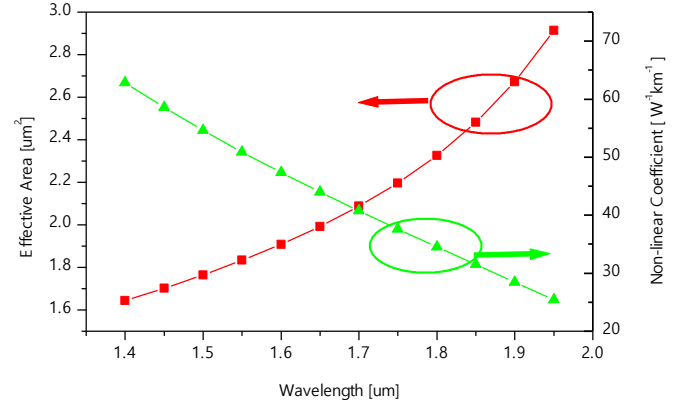


Fig. 3. Wavelength Vs Effective area and Non-linear co-efficient

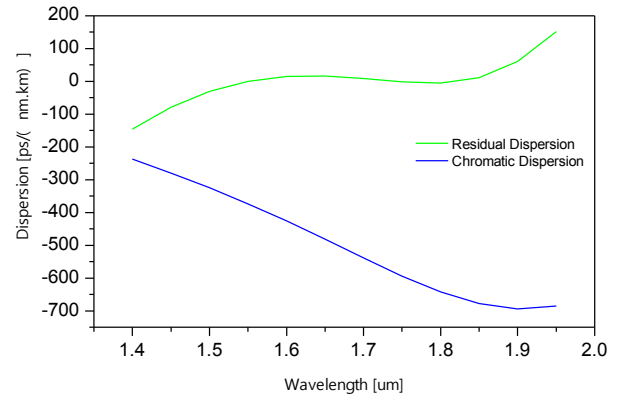


Fig. 4. Wavelength Vs Chromatic Dispersion and Residual Dispersion

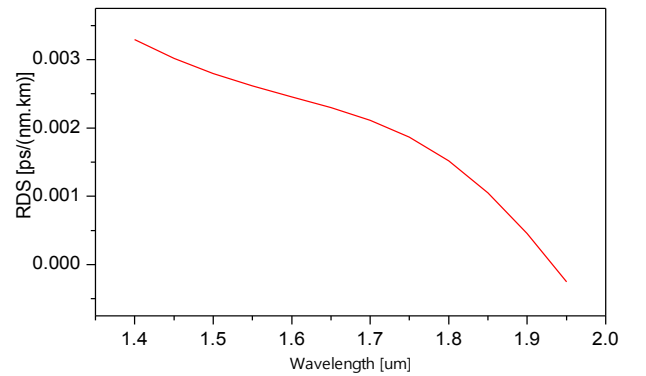


Fig. 5. Wavelength Vs Relative Dispersion Slope

Fig.4 represents the relationship between the wavelength versus residual dispersion and chromatic dispersion.

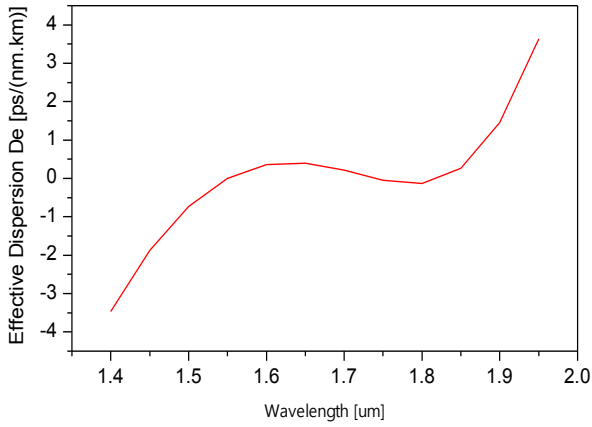


Fig. 6. Wavelength Vs Effective Dispersion

Generally chromatic dispersion affects the bit error rate and also limits the maximum transmission distance [15]. In that sense, our proposed design provides NZ-DSF as well as highly negative chromatic dispersion to enlarge the transmission efficiency [16].

In that order Fig. 5, 6 and Fig.7 show relative dispersion slope, effective dispersion and dispersion compensation ratio. The graph in Fig.6 shows that there is only ± 3.64 ps/nm-km dispersion variation within the wavelength of 1400nm to 1950nm and ± 0.85 ps/nm-km for 1490nm to 1870nm [17]-[18]. So this fiber can be used as NZ-DSF and for WDM. The relative dispersion slope of the proposed PCF is 0.002615 at 1550nm which is shown in Fig. 7.

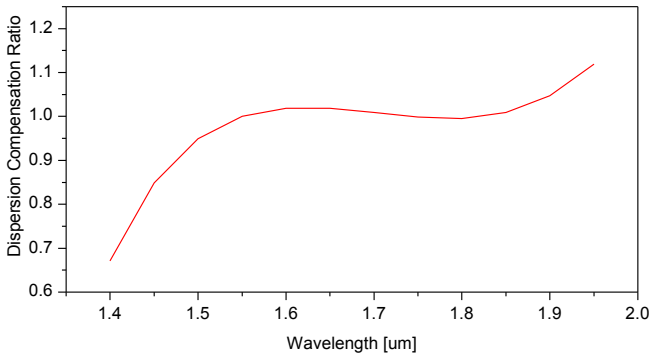


Fig. 7. Wavelength Vs Dispersion Compensation Ratio

TABLE I. COMPARISON BETWEEN THE PROPOSED PCF AND OTHER PCF AT 1550 NM

PCF Design	Birefringence, B	Non-linear Coefficient, γ ($\text{W}^{-1}\text{km}^{-1}$)
Ref [10]	1.75×10^{-2}	39.933
Ref [19]	4.2×10^{-2}	-----
Ref [20]	6×10^{-2}	-----
Proposed PCF	2.62×10^{-2}	50.86

IV. CONCLUSION

The proposed design is of ultra-high birefringence photonic crystal fiber having more advantage as compared to recently proposed design. The designed PCF offers -373.85 ps/(nm-km) chromatic dispersion with $50.86 \text{ W}^{-1}\text{km}^{-1}$ non-linear coefficient at the operating wavelength 1550nm. This design ensue the characteristics of non-zero dispersion shifted fiber of only ± 3.64 ps/nm-km dispersion variation for 1400nm to 1950nm wavebands.

References

- [1] P. Russell, "Photonic-Crystal Fibers", J. Lightwave Technol., vol. 24, pp. 4729-4749, 2006.
- [2] J. F. Liao and J. Q. Sun, "High birefringent rectangular-lattice photonic crystal fibers with low confinement loss employing different sizes of elliptical air holes in the cladding and the core," Opt. Fiber Technol., vol. 18, no. 6, pp. 457-461, Dec. 2012.
- [3] Jianfei Liao, Junqiang Sun, Mingdi Du, and Yi Qin, "Highly nonlinear dispersion-flattened slotted spiral photonic crystal fibers," IEEE Photonics Technology Letters, vol. 26, no. 4, Feb. 2014.
- [4] J. C. Knight, J. Arriaga, T. A. Birks, et.al., "Anomalous dispersion in photonic crystal fiber", IEEE Photonics Technology Lett., vol. 12, pp.807-809, 2000.
- [5] M. A. Islam and M. S. Alam, "Design of a polarization-maintaining equiangular spiral photonic crystal fiber for residual dispersion compensation over E + S + C + L + U wavelength bands," IEEE Photon. Techno. Lett. vol. 24, no. 11, pp. 930-932, Jun. 1, 2012.
- [6] J. C. Knight, T. A. Birks, R. F. Cregan, P. St. J. Russell, and I.-P. de Sandro, "Large mode area photonic crystal fibre," Electron. Lett., vol. 34, no. 13, pp. 1347-1348, Jun. 1998.
- [7] Karima Chah, et.al, "Temperature-insensitive polarimetric vibration sensor based on HiBi microstructured optical fiber", Applied Optics, Vol. 51, No. 25, 1 September 2012.
- [8] Mohammad Karimi, Tong Sun, and Kenneth. T. V. Grattan, "Design evaluation of a high birefringence single mode optical fiber-based sensor for lateral pressure monitoring applications" IEEE Sensors Journal, VOL. 13, NO. 11, November 2013.
- [9] Xin Wang, Hong-Juan Yang, Shan-Shan Wang, Yi-Peng Liao, and Jing Wang, "Seawater temperature measurement based on a high birefringence elliptic fiber sagnac loop" IEEE Photonics Technology Letters, VOL. 27, NO. 16, August 15, 2015.
- [10] Japatoosh Mondal, Mohammad Shaifur Rahman, "Design of highly birefringent dispersion compensating spiral photonic crystal fiber" in 2nd Int'l Conf on Electrical Engineering and Information & Communication Technology (ICEEICT) 2015, Bangladesh, 2015, ©2015 IEEE, DOI: 10.1109/ICEEICT.2015.7307515.
- [11] H. Ademgil, S. Haxha, T. Gorman, and F. AbdelMalek, "Bending effects on highly birefringent photonic crystal fibers with low chromatic dispersion and low confinement losses" Journal of Lightwave Technology, VOL. 27, NO. 5, March 1, 2009.
- [12] Lin An, Zheng Zheng, Zheng Li, Tao Zhou, and Jiangtao Cheng, "Ultrahigh birefringent photonic crystal fiber with ultralow confinement loss using four airholes in the core" Journal Of Lightwave Technology, VOL. 27, NO. 15, August 1, 2009.
- [13] N. A. Mortensen, "Effective area of photonic crystal fibers," Opt. Exp., vol. 10, pp. 341-348, 2002.
- [14] N. A. Mortensen and J. R. Folkenberg, "Low-loss criterion and effective area considerations for photonic crystal fibers," Journal of Optics A: Pure And Applied Optics vol. 5, pp. 163-166, 2003.
- [15] Graham H. Smith, Dalma Novak, and Zaheer Ahmed, "Overcoming chromatic-dispersion effects in fiber-wireless systems incorporating

- external modulators”, IEEE Transactions on Microwave Theory and Techniques, Vol. 45, No. 8, August 1997.
- [16] P. S. Andre I, A. L. J. Teixeira, M. Lima, J. F. da Rocha, J. L. Pinto, “Nonlinear refractive index and chromatic dispersion simultaneous measurement in non-zero dispersion shift optical fibres”, 4th International Conference on Transparent Optical Networks, warsaw poland 2002, vol.1, pp. 111 – 114, 2002.
- [17] Franco, M. A., Serrão, V. A., & Sircilli, F.. Microstructured optical fiber for residual dispersion compensation over wavelength bands. IEEE Photonics Technology Letters, 20(9), pp. 751-753, 2008.
- [18] Habib, M. S., Haque, M. M., Habib, M. S., Hasan, M. I., Rahman, M. S., & Razzak, S. M. A., Polarization maintaining holey fibers for residual dispersion compensation over S+ C+ L wavelength bands. Optik- International Journal for Light and Electron Optics, 125(3), pp. 911-915, 2004.
- [19] T. I. Yang, et.al, “High birefringence and low loss circular air-holes photonic crystal fiber using complex unit cells in cladding”, Opt. Comm., vol. 281, pp. 4334-4338, 2008.
- [20] Russel Reza Mahmud; Muhammad Abdul Goffar Khan; S. M. Abdur Razzak “Design and Analysis of a new photonic Crystal Fiber for Fiber Based Sensor Applications” , 2015 IEEE International Conference on Telecommunications and Photonics (ICTP), ©2015 IEEE, DOI: 10.1109/ICTP.2015.7427933.

Feasibility Study of GSM Network for Tracking Low Altitude Helicopter

A A M Mostafizur Rahman¹, Shartaz Hossain¹, Imtiaz
Jamil Tuku¹, Md Hossam-E-Haider¹,

⁽¹⁾Electrical Electronic & Communication Engineering
Department
Military Institute of Science and Technology
Dhaka, Bangladesh

Md. Syedul Amin²,

⁽²⁾Bangladesh Air Force
Dhaka, Bangladesh
syedul8585@yahoo.com

Abstract— Low altitude helicopters are difficult to track due to the limitation of radar horizon, which may cause severe flight hazard. Thus, tracking of helicopter in real time is vital. There are various ways to track low flying helicopter, such as aircraft to ground data link, satellite phone etc. But the Global System for Mobile Communications (GSM) network has not been reported to be utilized yet to transmit the position information obtained from Global Navigation Satellite System (GNSS) receiver. A study was carried out to check the coverage of GSM network in low altitude for the feasibility of sending location information and the report has been presented in this paper. A GPS and GSM modem was programmed to log the time, location, altitude and signal strength. Two missions were carried out at 1000 feet with the device. Logged data were later on downloaded in a computer to analyze the low altitude coverage of the GSM network. The results show that the signal quality of the GSM was available for more than 75% of the time for sending SMS. Thus, it is evident from the study report that the low altitude helicopter can be tracked using GSM network.

Keywords—AT Commands, GPS, GSM, Helicopter, SMS.

I. INTRODUCTION

Sky is immensely getting crowded day by day due to the increased number of aerial vehicles [1]. Recent development of Unmanned Aerial System (UAS) has escalated it further. For safety of these aerial vehicles, the information of the positions of these aerial vehicles is very important [2]. Radio Detection And Ranging (RADAR) is an excellent tool for tracking aerial vehicles. However, due to the limitation of radar horizon, position of any aerial vehicle cannot be obtained after a certain distance for a particular height [3].

Helicopter normally flies in low altitude which is generally around 1000 feet. Radar can pickup an aircraft maximum up to 75 km if flying at 1000 feet due to the earth curvature [4]. Even the position information cannot also be contacted over the radio due to the same limitation. As such, getting the real time position information is very difficult which poses serious flight safety hazards.

Various researches have been carried out to track low altitude aircraft besides radar. Secondary Surveillance Radar (SSR) is another means for obtaining position information from the transponder. But, the system also suffers from the distance limitation with height. However, GPS can be an excellent tool for obtaining position information [5]. Initially, using the GPS, it was possible to transmit data related to aircrafts in a form of discrete frequency signal, like flight numbers and aircrafts type, etc. to pilots. But that was only one way communication. Ground controls were blind of any information about the altitude and whether the aircraft is turning, climbing or descending. Later on, using Iridium communication system, control units were able to get most of the required information. But, transmitting data using Iridium is very expensive [6]. As such, a low cost system to track low altitude aircraft has an immense importance.

GSM is a digital mobile communication system that is widely used now a days in 213 countries by more than 690 mobile networks and represents 82.4% of all global mobile connections [7]. Besides the voice communication, it also offers Short Message Service (SMS) and General packet radio service (GPRS) to transfer data which is much cheaper than the satellite phone. GSM network can be used to transmit the positional data of GPS from the aircraft. However, GSM works on the principle of line of sight and primarily optimized and targeted for surface communication. Nevertheless, it is possible to communicate at low altitude in GSM network.

Dewdney carried out a study on the connectivity of cell phone at low altitude in Ontario, Canada [8]. The study revealed that the cell phone ceiling is approximately 7000 feet maximum for single-engine light aircraft and approximately 6000 feet maximum for twin engine light aircraft. However, the reliability and steadiness of the connection varied drastically. Another study was carried out by Juan et al. in Spain for the signal radiated by mobile telephony antennas installed in the direction towards the sky, till heights of about 3000 meters [9]. These studies show that GSM works above the ground level at low altitude, although, primarily optimized and targeted for surface communication.

Bangladesh has a well netted GSM network which connects all parts of Bangladesh. However, all the operators' network are not similarly strong, even at times, cannot be used for any sort of transmission. As such, a study was needed to check the availability of GSM network at low altitude. This paper presents the study report of the coverage of GSM network in low altitude for the feasibility of sending location information of helicopter.

II. BACKGROUND

A. GPS Technology

GPS is a satellite based system made up of 24 satellites set into space. GPS satellites transmit accurate signal data to earth by circling the earth twice every day in an extremely exact circle. Data received by the GPS receiver determines the position by comparing the time signals received from the GPS satellites and by trilateration of the known positions of the satellites as shown in Fig. 1 [10, 12]. Thus, to solve for a three dimensional position, four observations are needed since there will always be clock timing differences between the receiver and the satellites. Typical GPS positioning is done by quadrilateration with four or more pseudo range measurements and with no external corrections, is known as stand-alone GPS.

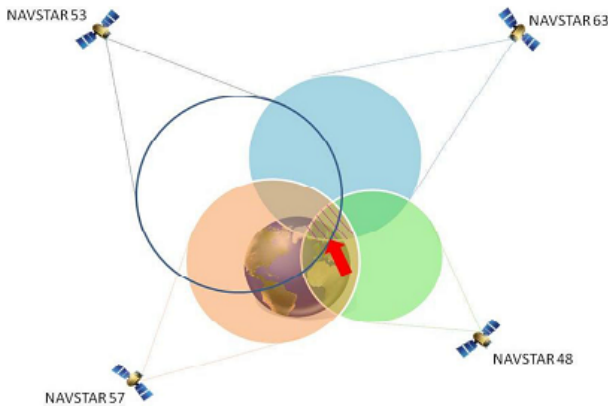


Fig. 1. GPS technology [11]

B. GSM Technology

GSM technology is based on radio signals which pass back and forth between cell sites known as base transceiver station (BTS) and user's cell phone. The voice of the user is converted into digital radio signals which are transmitted in all directions by a built-in antenna. It is then received by more sensitive and larger antennas at the nearest BTS and then to a mobile switching station (MSC). Once a call is picked up by the MSC, it is then transmitted over land lines such as public switched telephone network (PSTN) that connect with the regular telephone network as shown in Fig. 2. It is based on Time Division Multiple Access (TDMA) system. The frequency band for GSM is 900 MHz or 1800 MHz.

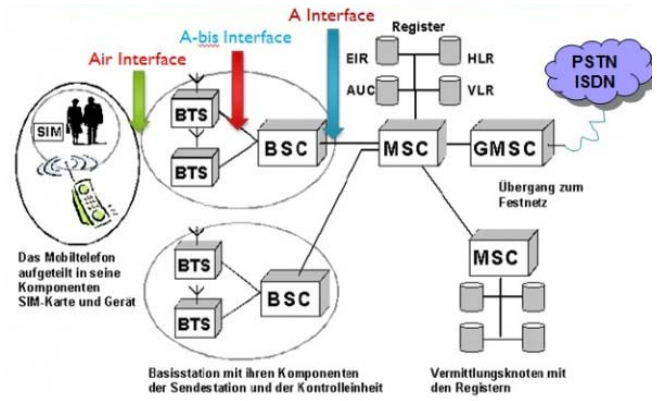


Fig. 2. GSM technology [9]

C. GSM Antenna

The antennas of BTS are linear radio antennas and are normally mounted on rooftops, ground-based masts or other existing structures at the height with a clear view over the existing surrounding terrain or buildings. The patterns of radiation of the antennas normally have relatively narrow main beam which provide rise in the gain from 10 to 20 dB [9]. The side and back lobe levels are normally between -15 and -25 dB. The transmission and reception pattern of the antennas are like an elongated lobe as shown in Fig. 3. Normally most of the energy is broadcasted towards the terrestrial surface. Only a small portion of the energy is transmitted to the space. As such, the vertical signal is as weak above the antenna as below it. However, signals are likely to be obtained from other antennas around.



Fig. 3. GSM antenna technology [8]

The lobes are vertically nested more closely together like a flattened onion. Normally the effective lobe extends to thousand feet vertically. If the cell sites are looked down, it may be imagined as wire-frames as shown in Fig. 4 [8]. Normally there are three antennas in a tower. As such, a cell site has normally three lobes as shown in the figure.



Fig. 4. Lobes of cell sites [8]

III. METHODOLOGY

A device was needed to check the feasibility of the GSM network at the low altitude. The device was composed commercially of the shelf (COTS) hardware. A GPS/GPRS/GSM module comprising of GPS and GSM modem, an Arduino microcontroller unit (MCU), a micro Secure Digital (SD) card and power source were used. The details of the hardware are given below.

A. GPS/GPRS/GSM Module

The GPS/GPRS/GSM Module from DF Robot is used for getting GPS position and querying the signal strength of GSM network. The module houses a SIM908 SMT packaged chip which has a Quad-band GSM/GPRS engine workable in EGSM 900MHz/DCS 1800MHz and GSM850 MHz/PCS 1900MHz frequencies. Both the GPS and GSM modem can be controlled via AT command. The board houses a flexible GSM antenna and a high-gain SMD antenna for GPS.

The GPS receiver is a high performance STE engine with 42 channels and L1 C/A code capable. It has -160dBm tracking sensitivity. In a cold start, it takes 30 sec and in hot start, it takes 1 sec for first fixing. The horizontal position accuracy has less than 2.5m Circular Error Probability (CEP). It provides output in National Marine Electronics Association (NMEA) specification over the serial interface which allows to access the data in various formats.

B. Arduino MCU

The Arduino MCU board is used with the GPS/GPRS/GSM Module for data acquisition from the GPS and querying GSM signal strength. It is based on the ATmega328P microcontroller having 14 digital Input/ Output (IO) pins. Among these 14 pins, 6 can be used as PWM outputs, 6 pins is for analog input. It has 16 MHz quartz crystal, a USB connection, a power jack, an ICSP header and the remaining one is the reset button. It can be connected to computer through USB cable. The ATmega328 has 32 KB memory. The boot loader occupies only 0.5 KB. It contains 2

KB of SRAM and 1 KB of EEPROM. The board can be powered via the USB connection or with an external power supply which can be selected automatically. The compact logger module is used to check the feasibility of GSM signal strength is shown in Fig. 5.



Fig. 5. Logger module

During the missions the device was programmed to acquire the position of the helicopter and signal strength in that location at that time. The information was logged in the SD card. The flow chart of the methodology is shown in Fig. 6. These information were logged every after 30 sec.

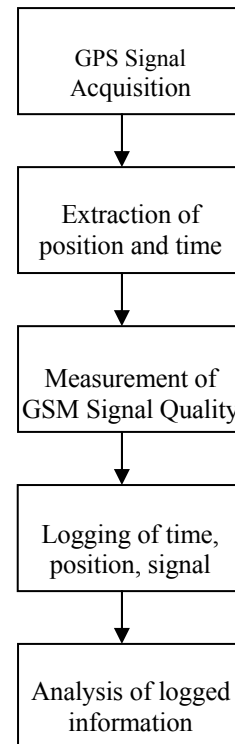


Fig. 6. Flow chart of the methodology

The GPS/GPRS/GSM Module was connected over the Arduino MCU and powered via the USB cable for programming. Arduino language was used for programming. The module was accessed via AT commands. Baud rate of 9600 bps was set by using AT+CGPSIPR=9600 command. AT+CGPSPWR=1 was used to power up the GPS and AT+CGPSRST=1 was used to reset GPS in autonomy mode.

NMEA defines the specification of GPS receiver communication by various self contained sentences. These are independent from each other. Most of the GPS receivers transmit GPRMC sentence which contains time, latitude, longitude, speed over ground, heading etc. Thus, the information from the GPRMC sentence was used to capture the position in terms of latitude, longitude and time.

After capturing position, the signal quality of the GSM network was checked. AT+CSQ command was used to obtain the signal quality report. The CSQ commands return received signal strength indication in Received Signal Strength Indicator (RSSI) and channel Bit Error Rate (BER). RSSI is the measurement of how good the device can receive a signal from cell tower. RSSI and dBm are different measurement units. But both represent the signal strength. However, the difference is that dBm is an absolute number representing power levels in mW (milliwatts), whereas RSSI is a relative index.

The RSSI was considered for the GSM signal quality. The value of RSSI varies from 0 to 99. The relation of RSSI and dBm is given in Table 1 indicating the signal strength. The time, position in terms of latitude and longitude and the signal were stored in the SD Card.

TABLE I. RSSI VS DBM FOR SIGNAL STRENGTH

Value	RSSI dBm	Condition
2	-109	Marginal
3	-107	Marginal
4	-105	Marginal
5	-103	Marginal
6	-101	Marginal
7	-99	Marginal
8	-97	Marginal
9	-95	Marginal
10	-93	OK
11	-91	OK
12	-89	OK
13	-87	OK
14	-85	OK
15	-83	Good
16	-81	Good
17	-79	Good
18	-77	Good
19	-75	Good
20	-73	Excellent
21	-71	Excellent
22	-69	Excellent
23	-67	Excellent
24	-65	Excellent
25	-63	Excellent
26	-61	Excellent
27	-59	Excellent
28	-57	Excellent
29	-55	Excellent
30	-53	Excellent

IV. EXPERIMENTAL RESULTS AND DISCUSSION

Helicopter missions were utilized for checking the signal quality of the GSM network. The device was programmed and assembled with independent power supply following the methodology given in the previous section. The Subscriber Identification Module (SIM) card of Grameen phone was used considered the largest network in Bangladesh. The device was placed inside the helicopter for a clear line of sight for the GPS and the GSM antenna. Two missions were planned from Dhaka. The first mission was from Dhaka to Jessore and the second mission was from Dhaka to Chittagong. The route of the missions are shown in Fig. 7.



Fig. 7. Helicopter routes for signal quality check

In the first mission, the helicopter flew with the device from Dhaka to Chittagong as shown in Fig. 7. The helicopter flew at 200 kph and maintained 1000 feet height. The total flying time was 110 minutes. 220 positions and signal strength could be logged in the mission as the data were logged every after 30 seconds. In the next mission the helicopter flew with the device from Dhaka to Jessore as shown in Fig. 7. The helicopter flew at 200 kph and maintained 1000 feet height like the previous mission. The total flying time was 45 minutes. 90 positions and signal strength could be logged in the mission.

Data stored in the SD card were copied in the computer. The signal strength with time and position were taken from the data. The signal strength were analysed for both the missions. In the first mission, while flying from Dhaka to Chittagong, the helicopter flew over the river Meghna. Before reaching to Chittagong, it flew over the small hills and the sea was towards the right of the track. Signal quality over these areas were poor which was most of the time <2 RSSI. Thus, it was unusable for communication as referenced from Table 1. The unusable or no communication state was checked around for about 20% of the total logged data of the route. Another 10% were also between 2 to 6 RSSI which was considered as unreliable for communication. More than 60% of the signal quality was between 7 to 14 RSSI which were OK for communication. Approximately 10% of the signal was above 15 RSSI which indicated good for communication. However these were mainly during the take off and landing phase as the helicopter was at low height.

In the second mission, while flying from Dhaka to Jessore, the helicopter flew over the river Padma which was wide and over Madhumati river and that was comparatively narrow. Signal quality over the rivers were poor which was most of the time < 2 RSSI. Thus, it was unusable for communication as given in Table 1. The unusable or no communication state was around 17% of the total logged data of the route. Another 8% were also between 2 to 6 RSSI which was considered as unreliable for communication. More than 60% of the signal quality was between 7 to 14 RSSI which were OK for communication. Approximately 15% of the signal was above 15 RSSI which indicated good for communication. However, these were also mainly during the take off and landing phase as the helicopter was at low height.

Following Table 1, it was confirmed from both the missions that approximately 75% of the time workable signal strength (more than 7 RSSI) was received by the logger. As such, it is evident from the study that positional information of a helicopter flying at 1000 feet can be transmitted using the GSM network.

V. CONCLUSION

Presently, managing increased number of flights has become challenging as the sky is getting crowded day by day. Thus, tracking an aircraft in real time is the prime need in ensuring flight safety. Tracking of helicopter flying at low altitude is limited by the radar horizon. However, iridium can be used to mitigate this issue at the expense of huge costing. In consequence, the GPS and GSM based real time tracking system can provide low cost solution. To perform this analysis a logger module was introduced integrating GPS/GPRS/GSM module connected over the Arduino MCU. The GSM signal strength was measured in two separated missions. Finally, the outcome of these missions brought out a conclusion that although GSM network is primarily optimized and targeted for surface communication, it can also be satisfactorily used at low altitude for tracking helicopter.

REFERENCES

- [1] R. G. Fox, "The History of Helicopter Safety," in International Helicopter Safety Symposium, Montréal, Québec, Canada, 2005, pp. 1-17.
- [2] D. McCallie, J. Butts, and R. Mills, "Security analysis of the ADS-B implementation in the next generation air transportation system," International Journal of Critical Infrastructure Protection, vol. 4, pp. 78-87, 2011.
- [3] K.-W. Gurgel, H.-H. Essen, and S. Kingsley, "High-frequency radars: physical limitations and recent developments," Coastal Engineering, vol. 37, pp. 201-218, 1999.
- [4] D. L. Knepp, "Radar measurement of ionospheric scintillation in the polar region," in General Assembly and Scientific Symposium, 2011 XXXth URSI, 2011, pp. 1-4.
- [5] M. S. Amin, M. A. S. Bhuiyan, M. B. I. Reaz, and S. S. Nasir, "GPS and Map Matching Based Vehicle Accident Detection System," in 2013 IEEE Student Conference on Research and Development (SCORED), Putrajaya, Malaysia, 2013, pp. 1-4.
- [6] L. Collins, "Iridium reborn [Comms Iridium]," Engineering & Technology, vol. 5, pp. 64-65, 2010.
- [7] M. S. Amin, M. B. I. Reaz, and S. S. Nasir, "Integrated Vehicle Accident Detection and Location System," TELKOMNIKA Telecommunication, Computing, Electronics and Control, vol. 12, pp. 73-78, 2014.
- [8] A. Dewdney, "Project Achilles Report: Parts One, Two and Three," ed: Physics.
- [9] J. A. Romo, G. Aranguren, J. Bilbao, I. Odriozola, J. Gómez, and L. Serrano, "GSM/GPRS signal strength measurements in aircraft flights under 3,000 meters of altitude," WSEAS Transactions on Signal Processing, vol. 5, pp. 219-228, 2009.
- [10] M. Hossam-E-Haider, A. Tabassum, R. H. Shihab, and C. M. Hasan, "Comparative analysis of GNSS reliability: GPS, GALILEO and combined GPS-GALILEO," in Electrical Information and Communication Technology (EICT), 2013 International Conference on, 2014, pp. 1-6.
- [11] Z. M. Malik, "Design and implementation of temporal filtering and other data fusion algorithms to enhance the accuracy of a real time radio location tracking system," University of Gävle, 2012.
- [12] M. S. Amin, M. B. I. Reaz, M. A. S. Bhuiyan and S. S. Nasir, "Kalman filtered GPS accelerometer-based accident detection and location system: A low-cost approach," Current Science, vol. 106(11), pp. 1548-1554, 2014.

Secure Wireless Multicasting Through Multi-cellular MIMO Networks with linear equalization

R. Sayed¹, M. Z. I. Sarkar¹, S. M. R. Ali² and D. K. Sarker¹

¹ Department of Electrical and Electronic Engineering
Rajshahi University of Engineering and Technology
Rajshahi-6204, Bangladesh.

² Department of Information and Communication Technology
Bangladesh University of Professionals
Mirpur Cantonment, Bangladesh.
Email: rosnisayed@gmail.com

Abstract—In this paper, we consider a confidential communication through multi-cellular co-located multiple-input multiple-output (C-MIMO) and distributed MIMO (D-MIMO) networks in the presence of an eavesdropper. The base station (BS) of each cell sends a common stream of information to a group of receivers within the cell and an eavesdropper observes their communications. We are interested to protect these information from eavesdropping, so that no eavesdropper can decode any information. Moreover, we investigate the performance of maximal ratio combining (MRC), zero-forcing (ZF) and minimum mean-square error (MMSE) detectors in enhancing security in multi-cellular MIMO networks considering their presence at the receivers and eavesdropper. Our results show that although the D-MIMO system is more sensitive to interference than C-MIMO system but in the high interference region, the effect of interference on both the MIMO models are almost same. The increase in transmit antennas enhances security and the MRC and MMSE detectors perform better than the ZF detector when the number of transmit antennas is very small.

Index Terms—Co-located MIMO, distributed MIMO, linear receiver, multi-cellular, secrecy rate.

I. INTRODUCTION

In the last decade, MIMO antenna systems have been shown to be able to substantially increase the data rate and improve the reliability of communication system without extra spectrum and power resources. Two different schemes can be adopted for MIMO systems, one is C-MIMO where antennas are co-located at both transmitting and receiving sides and other is D-MIMO where BS antennas are located at different geographical locations.

Recently, the spectral efficiency of massive MIMO systems for both C-MIMO and D-MIMO systems were investigated showing the effects of path loss, shadowing, multi-path fading and antenna correlation [1]. Here the authors showed that though D-MIMO scheme exhibits higher multiplexing gain, but it does not always outperform the C-MIMO scheme. Electrical efficiency of the D-MIMO and C-MIMO in the uplink cellular systems for composite fading channels were compared in [2], where the authors showed that D-MIMO system is more energy efficient than C-MIMO system. In order to optimize C-MIMO and D-MIMO average capacities, the waveguide multi-mode model and particle-swarm-optimization

method were employed in [3] where the authors showed that D-MIMO system is suitable for those environments where C-MIMO system can not perform properly and desirable performances of the D-MIMO system can be achieved by proper placement of the antennas. In [4], the theoretical performances of D-MIMO and traditional C-MIMO systems were studied and compared for composite Rayleigh fading and log-normal shadowing channel. The analytical approximations of the mean and cumulative distribution function (CDF) of the mutual information and the closed-form expressions for the mean spectral efficiency and mean outage spectral efficiency were derived for C-MIMO and D-MIMO channels. In [5], the performances of MRC, ZF and MMSE detectors were studied for enhancing the ergodic capacity of Nakagami-m fading C-MIMO and D-MIMO channels. The authors showed that MMSE detector is the most efficient detector in enhancing the performance of Nakagami-m fading channels compared to ZF and MRC detectors. However, the study of secure wireless multicasting through multi-cellular MIMO networks with MRC, ZF and MMSE detectors is not available in the literature.

In this paper, we consider a wireless multicasting scenario through Rayleigh fading multi-cellular MIMO networks consisting n cells in the presence of an eavesdropper. Two MIMO models such as C-MIMO and D-MIMO are considered for the formulation of the proposed model with MRC, ZF and MMSE detectors at the receivers. In order to compare the performance of MRC, ZF and MMSE detectors in enhancing security in multi-cellular MIMO networks, at first, we derive the expressions of achievable ergodic secrecy rate separately for MRC, ZF and MMSE detectors considering the effect of the number of cells. Then, we show the numerical results using Monte-Carlo simulation. Finally, we draw the conclusion and identify which detector is efficient of a particular MIMO model.

The rest of the paper is organized as follows. Section II describes the system model. In section III, the achievable ergodic secrecy rate is formulated for C-MIMO system. Section IV formulates the achievable ergodic secrecy rate for D-MIMO system. Section V provides the numerical results.

Finally, Section-VI describes the conclusion of this work.

II. SYSTEM MODEL

The system model is shown in Fig.1 where a multicasting scenario is considered through n number of cells in the presence of an eavesdropper. Each base station (BS) sends a common stream of information to a group of users and the eavesdropper E observes their communications. The channel between BS and users is known as multicast channel and that between BS and eavesdropper is known as eavesdropper channel. We are interested to protect the information of each BS from eavesdropping so that the eavesdropper cannot decode a single bit of information from the multicast channels. The BS-1, BS-2, ..., BS- n communicate with Group-1, Group-2, ..., Group- n , respectively. Note that Group-1, Group-2, ..., Group- n consists of $K_1, K_2, \dots, K_n$ users, respectively. The BS-1, BS-2, ..., BS- n are equipped with $M_1, M_2, \dots, M_n$ antennas and each user of Group-1, Group-2, ..., Group- n are equipped with $N_1, N_2, \dots, N_n$ antennas, respectively. The eavesdropper is equipped with N_e antennas. All the channels are assumed to be Rayleigh fading. Two MIMO models such as C-MIMO and D-MIMO are considered for the formulation of the proposed model.

III. ACHIEVABLE ERGODIC SECRECY RATE FOR C-MIMO

In the case of C-MIMO system, the received signal vector for j th user of Group- n , $j = 1, \dots, K_n$ can be expressed as,

$$\mathbf{y}_{n_j}^{\text{C-MIMO}} = \sqrt{\frac{P_1}{M_1}} \mathbf{G}_{1j} \mathbf{x}_{1j} + \sqrt{\frac{P_2}{M_2}} \mathbf{G}_{2j} \mathbf{x}_{2j} + \dots + \sqrt{\frac{P_{n-1}}{M_{n-1}}} \mathbf{G}_{(n-1)j} \mathbf{x}_{(n-1)j} + \sqrt{\frac{P_n}{M_n}} \mathbf{H}_{nj} \mathbf{x}_{nj} + \mathbf{w}_{n_j}, \quad (1)$$

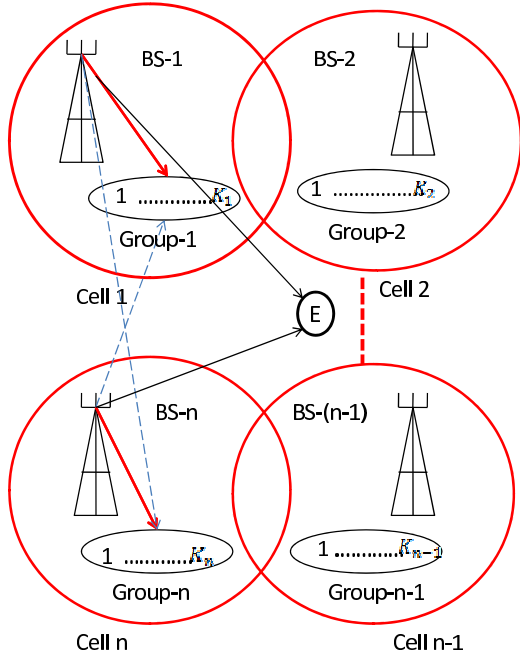


Fig. 1. System model.

where $\mathbf{H}_{nj} \in \mathbb{C}^{N_n \times M_n}$ denotes the direct channel coefficients between BS- n and j th user of group- n and $\mathbf{G}_{1j} \in \mathbb{C}^{N_n \times M_1}$, $\mathbf{G}_{2j} \in \mathbb{C}^{N_n \times M_2}, \dots, \mathbf{G}_{(n-1)j} \in \mathbb{C}^{N_n \times M_{n-1}}$ denote the cross channel coefficients from BS-1, BS-2, ..., BS- $n-1$ to j th user of group- n , respectively. $\mathbf{x}_{1j}, \mathbf{x}_{2j}, \dots, \mathbf{x}_{nj}$ denote the transmitted signal vectors from BS-1, BS-2, ..., BS- n to the j th user of Group- n respectively. The average power transmitted by BS-1, BS-2, ..., BS- n are expressed as $P_1, P_2, \dots, P_n$, respectively. $\mathbf{w}_{n_j}$ denotes the received noise vector at the j th user of Group- n . Similar to (1), the received signal vector at the eavesdropper can be expressed as

$$\mathbf{y}_e^{\text{C-MIMO}} = \sqrt{\frac{P_1}{M_1}} \mathbf{Z}_1 \mathbf{x}_{1j} + \sqrt{\frac{P_2}{M_2}} \mathbf{Z}_2 \mathbf{x}_{2j} + \dots + \sqrt{\frac{P_{n-1}}{M_{n-1}}} \mathbf{Z}_{n-1} \mathbf{x}_{n-1j} + \sqrt{\frac{P_n}{M_n}} \mathbf{Z}_n \mathbf{x}_{nj} + \mathbf{w}_e, \quad (2)$$

where $\mathbf{Z}_1 \in \mathbb{C}^{N_e \times M_1}$, $\mathbf{Z}_2 \in \mathbb{C}^{N_e \times M_2}, \dots, \mathbf{Z}_{(n-1)} \in \mathbb{C}^{N_e \times M_{n-1}}$, $\mathbf{Z}_n \in \mathbb{C}^{N_e \times M_n}$ denote the channel coefficients from BS-1, BS-2, ..., BS- $n-1$, BS- n to eavesdropper, respectively. $\mathbf{w}_e$ denotes the received noise vector at the eavesdropper.

Let us consider a $N_n \times M_n$ linear detector matrix represented by $\mathbf{A}_{n_j}$, which is dependent on $\mathbf{H}_{nj}$. In order to separate the received signal expressed by (1) into streams, we can use the linear detector matrix as follows;

$$\mathbf{r}_n = \mathbf{A}_{n_j}^\dagger \mathbf{y}_{n_j}^{\text{C-MIMO}},$$

where $\mathbf{r}_n$ denotes the received signal vector after linear detection and the linear detector matrix $\mathbf{A}_{n_j}$ is defined as follows for the MRC, ZF and MMSE detectors [6]

$$\mathbf{A}_{n_j} = \begin{cases} \mathbf{H}_{nj} & \text{for MRC,} \\ \mathbf{H}_{nj} (\mathbf{H}_{nj}^\dagger \mathbf{H}_{nj})^{-1} & \text{for ZF,} \\ \mathbf{H}_{nj} (\mathbf{H}_{nj}^\dagger \mathbf{H}_{nj} + \frac{M_n}{P} \mathbf{I}_{M_n})^{-1} & \text{for MMSE.} \end{cases}$$

Hence the received signal vector after using linear detector matrix is given by

$$\mathbf{r}_n = \sqrt{\frac{P_1}{M_1}} \mathbf{A}_{n_j}^\dagger \mathbf{G}_{1j} \mathbf{x}_{1j} + \sqrt{\frac{P_2}{M_2}} \mathbf{A}_{n_j}^\dagger \mathbf{G}_{2j} \mathbf{x}_{2j} + \dots + \sqrt{\frac{P_{n-1}}{M_{n-1}}} \mathbf{A}_{n_j}^\dagger \mathbf{G}_{(n-1)j} \mathbf{x}_{(n-1)j} + \sqrt{\frac{P_n}{M_n}} \mathbf{A}_{n_j}^\dagger \mathbf{H}_{nj} \mathbf{x}_{nj} + \mathbf{A}_{n_j}^\dagger \mathbf{w}_{n_j}$$

Therefore, the received signal at i th antenna, $i = 1, \dots, N_n$ of j th user, $j = 1, \dots, K_n$ of Group- n can be expressed as

$$\mathbf{r}_{n_j,i} = \sqrt{\frac{P_n}{M_n}} \mathbf{a}_{n_j,i}^\dagger \mathbf{h}_{n_j,i} x_{n_j,i} + \sqrt{\frac{P_n}{M_n}} \sum_{t=1, t \neq i}^{N_n} \mathbf{a}_{n_j,i}^\dagger \mathbf{h}_{n_j,t} x_{n_j,t} + \sum_{m=1}^{n-1} \sqrt{\frac{P_m}{M_m}} \mathbf{A}_{n_j}^\dagger \mathbf{G}_{mj} \mathbf{x}_{mj} + \mathbf{A}_{n_j}^\dagger \mathbf{w}_{n_j}, \quad (3)$$

where $\mathbf{a}_{n_j,i}$ denotes the i th column of $\mathbf{A}_{n_j}$ matrix and $\mathbf{h}_{n_j,i}$ denotes the i th column of $\mathbf{H}_{nj}$ matrix. Similarly, the received

signal at the p th antenna, $p = 1, \dots, N_e$ of eavesdropper is given by

$$\mathbf{d}_{n_p} = \sqrt{\frac{P_n}{M_n}} \mathbf{b}_{n_p}^\dagger \mathbf{z}_{n_p} x_{n_{j,i}} + \sqrt{\frac{P_n}{M_n}} \sum_{t=1, t \neq p}^{N_e} \mathbf{b}_{n_p}^\dagger \mathbf{z}_{n_t} x_{n_{j,i}} + \sum_{m=1}^{n-1} \sqrt{\frac{P_m}{M_m}} \mathbf{B}_n^\dagger \mathbf{Z}_m \mathbf{x}_{m_j} + \mathbf{B}_n^\dagger \mathbf{w}_e, \quad (4)$$

where $\mathbf{B}_n$ is a linear detector matrix, dependent on $\mathbf{Z}_n$. Transmitted signal from BS- n to i th antenna of j th user is the desired signal and the transmitted signals from other base stations are the interference signals. $\mathbf{b}_{n_p}$ denotes the p th column of $\mathbf{B}_n$ matrix and $\mathbf{z}_{n_p}$ denotes the p th column of $\mathbf{Z}_n$ matrix. Therefore, the achievable ergodic rate at the i th antenna of j th user of Group- n is given by

$$\langle R_n^{\text{C-MIMO}} \rangle = \mathbb{E} \left\{ \log_2 \left(1 + \frac{\frac{P_n}{M_n} |\mathbf{a}_{n_{j,i}}^\dagger \mathbf{h}_{n_{j,i}}|^2}{\alpha_n} \right) \right\}, \quad (5)$$

where $\alpha_n = \frac{P_n}{M_n} \sum_{t=1, t \neq i}^{N_n} |\mathbf{a}_{n_{j,i}}^\dagger \mathbf{h}_{n_{j,t}}|^2 + \sum_{m=1}^{n-1} \frac{P_m}{M_m} |\mathbf{A}_{n_j}^\dagger \mathbf{G}_{m_j}|^2 + \|\mathbf{A}_{n_j}\|^2$. Similarly, the achievable ergodic rate at the p th antenna of eavesdropper is given by

$$\langle R_{E_n}^{\text{C-MIMO}} \rangle = \mathbb{E} \left\{ \log_2 \left(1 + \frac{\frac{P_n}{M_n} |\mathbf{b}_{n_p}^\dagger \mathbf{z}_{n_p}|^2}{\beta_n} \right) \right\}, \quad (6)$$

where $\beta_n = \frac{P_n}{M_n} \sum_{t=1, t \neq p}^{N_e} |\mathbf{b}_{n_p}^\dagger \mathbf{z}_{n_t}|^2 + \sum_{m=1}^{n-1} \frac{P_m}{M_m} |\mathbf{B}_n^\dagger \mathbf{Z}_m|^2 + \|\mathbf{B}_n\|^2$ and $\mathbf{B}_n = [\mathbf{b}_{n_1}, \mathbf{b}_{n_2}, \dots, \mathbf{b}_{n_{N_e}}]$. Hence, the achievable ergodic secrecy rate is given by

$$\langle R_{S_n}^{\text{C-MIMO}} \rangle = \mathbb{E} \left\{ \log_2 \left(\frac{1 + \frac{\frac{P_n}{M_n} |\mathbf{a}_{n_{j,i}}^\dagger \mathbf{h}_{n_{j,i}}|^2}{\alpha_n}}{1 + \frac{\frac{P_n}{M_n} |\mathbf{b}_{n_p}^\dagger \mathbf{z}_{n_p}|^2}{\beta_n}} \right) \right\} \quad (7)$$

A. MRC Receiver

In the case of MRC detector at the receiver, we have $\mathbf{A}_{n_j} = \mathbf{H}_{n_j}$. Hence $\mathbf{a}_{n_{j,i}} = \mathbf{h}_{n_{j,i}}$. Similarly, for deriving the achievable rate of the transmit signal from BS- n to the i th antenna of j th user of Group- n at the eavesdropper, we have $\mathbf{B}_n = \mathbf{Z}_n$. Hence $\mathbf{b}_{n_p} = \mathbf{z}_{n_p}$. From (7), the achievable ergodic secrecy rate for the transmit signal from BS- n to the i th antenna of j th user of Group- n is given by

$$\langle R_{S_n}^{\text{C-MRC}} \rangle = \mathbb{E} \left\{ \log_2 \left(\frac{1 + \frac{\frac{P_n}{M_n} \|\mathbf{h}_{n_{j,i}}\|^4}{\epsilon_n}}{1 + \frac{\frac{P_n}{M_n} \|\mathbf{z}_{n_p}\|^4}{\zeta_n}} \right) \right\}, \quad (8)$$

where $\epsilon_n = \frac{P_n}{M_n} \sum_{t=1, t \neq i}^{N_n} |\mathbf{h}_{n_{j,i}}^\dagger \mathbf{h}_{n_{j,t}}|^2 + \sum_{m=1}^{n-1} \frac{P_m}{M_m} |\mathbf{H}_{n_j}^\dagger \mathbf{G}_{m_j}|^2 + \|\mathbf{H}_{n_j}\|^2$ and $\zeta_n = \frac{P_n}{M_n} \sum_{t=1, t \neq p}^{N_e} |\mathbf{z}_{n_p}^\dagger \mathbf{z}_{n_t}|^2 + \sum_{m=1}^{n-1} \frac{P_m}{M_m} |\mathbf{Z}_n^\dagger \mathbf{Z}_m|^2 + \|\mathbf{Z}_n\|^2$.

B. ZF Receiver

In the case of ZF detectors at the receiver, we have $\mathbf{A}_{n_j}^\dagger = (\mathbf{H}_{n_j}^\dagger \mathbf{H}_{n_j})^{-1} \mathbf{H}_{n_j}^\dagger$ or $\mathbf{A}_{n_j}^\dagger \mathbf{H}_{n_j} = \mathbf{I}_{M_n}$. Hence $\mathbf{a}_{n_{j,i}}^\dagger \mathbf{h}_{n_{j,t}} = \delta_{i,t}$, where

$$\delta_{i,t} = \begin{cases} 1, & \text{for } i=t \\ 0, & \text{otherwise.} \end{cases}$$

Similarly, at the eavesdropper, we have $\mathbf{B}_n^\dagger = (\mathbf{Z}_n^\dagger \mathbf{Z}_n)^{-1} \mathbf{Z}_n^\dagger$ or $\mathbf{B}_n^\dagger \mathbf{Z}_n = \mathbf{I}_{M_n}$. Hence $\mathbf{b}_{n_p}^\dagger \mathbf{z}_{n_t} = \delta_{p,t}$, where

$$\delta_{p,t} = \begin{cases} 1, & \text{for } p=t \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, using (7), the achievable ergodic secrecy rate for the i th antenna of j th user of Group- n is given by

$$\langle R_{S_n}^{\text{C-ZF}} \rangle = \mathbb{E} \left\{ \log_2 \left(\frac{1 + \frac{\frac{P_n}{M_n}}{\gamma_n}}{1 + \frac{\frac{P_n}{M_n}}{\theta_n}} \right) \right\}, \quad (9)$$

where $\gamma_n = \sum_{m=1, m \neq n}^N \frac{P_m}{M_m} |(\mathbf{H}_{n_j}^\dagger \mathbf{H}_{n_j})^{-1} \mathbf{H}_{n_j}^\dagger \mathbf{G}_{m_j}|^2 + |(\mathbf{H}_n \mathbf{H}_n^\dagger \mathbf{H}_n)^{-1}|^2$ and $\theta_n = \sum_{m=1}^{n-1} \frac{P_m}{M_m} |(\mathbf{Z}_n^\dagger \mathbf{Z}_n)^{-1} \mathbf{Z}_n^\dagger \mathbf{Z}_m|^2 + |\mathbf{Z}_n (\mathbf{Z}_n^\dagger \mathbf{Z}_n)^{-1}|^2$.

C. MMSE Receiver

In the case of MMSE detectors at the receiver, we have $\mathbf{A}_{n_j}^\dagger = (\mathbf{H}_{n_j}^\dagger \mathbf{H}_{n_j} + \frac{M_n}{P_n} \mathbf{I}_{M_n})^{-1} \mathbf{H}_{n_j}^\dagger = \mathbf{H}_{n_j}^\dagger (\mathbf{H}_{n_j} \mathbf{H}_{n_j}^\dagger + \frac{M_n}{P_n} \mathbf{I}_{N_n})^{-1}$. Similarly, in the case of MMSE detectors at the eavesdropper, we have $\mathbf{B}_n^\dagger = (\mathbf{Z}_n^\dagger \mathbf{Z}_n + \frac{M_n}{P_n} \mathbf{I}_{M_n})^{-1} \mathbf{Z}_n^\dagger = \mathbf{Z}_n^\dagger (\mathbf{Z}_n \mathbf{Z}_n^\dagger + \frac{M_n}{P_n} \mathbf{I}_{N_e})^{-1}$. Therefore, the achievable ergodic secrecy rate for the i th antenna of j th user of Group- n is given by

$$\langle R_{S_n}^{\text{C-MMSE}} \rangle = \mathbb{E} \left\{ \log_2 \left(\frac{1 + \frac{\frac{P_n}{M_n} |\mathbf{a}_{n_{j,i}}^\dagger \mathbf{h}_{n_{j,i}}|^2}{\kappa_n}}{1 + \frac{\frac{P_n}{M_n} |\mathbf{b}_{n_p}^\dagger \mathbf{z}_{n_p}|^2}{\mu_n}} \right) \right\}, \quad (10)$$

where $\kappa_n = \frac{P_n}{M_n} \sum_{t=1, t \neq i}^{N_n} |\mathbf{a}_{n_{j,i}}^\dagger \mathbf{h}_{n_{j,t}}|^2 + \sum_{m=1}^{n-1} \frac{P_m}{M_m} |\mathbf{A}_{n_j}^\dagger \mathbf{G}_{m_j}|^2 + \|\mathbf{A}_{n_j}\|^2$ and $\mu_n = \frac{P_n}{M_n} \sum_{t=1, t \neq p}^{N_e} |\mathbf{b}_{n_p}^\dagger \mathbf{z}_{n_t}|^2 + \sum_{m=1}^{n-1} \frac{P_m}{M_m} |\mathbf{B}_n^\dagger \mathbf{Z}_m|^2 + \|\mathbf{B}_n\|^2$.

IV. ACHIEVABLE ERGODIC SECRECY RATE FOR D-MIMO

For a D-MIMO system, the received signal vector $\mathbf{y}_{n_j}^{\text{D-MIMO}} \in \mathbb{C}^{Q_n M_n \times 1}$, at the j th user of Group- n related with the transmitted signal vector $\mathbf{x}_{1_j}, \mathbf{x}_{2_j}, \dots, \mathbf{x}_{n_j}$ is given by

$$\mathbf{y}_{n_j}^{\text{D-MIMO}} = \sqrt{\frac{P_1}{Q_1 M_1}} \mathbf{F}_{1_j} \mathbf{x}_{1_j} + \sqrt{\frac{P_2}{Q_2 M_2}} \mathbf{F}_{2_j} \mathbf{x}_{2_j} + \dots + \sqrt{\frac{P_{n-1}}{Q_{n-1} M_{n-1}}} \mathbf{F}_{(n-1)_j} \mathbf{x}_{(n-1)_j} + \sqrt{\frac{P_n}{Q_n M_n}} \mathbf{L}_{n_j} \mathbf{x}_{n_j} + \varsigma_{n_j}, \quad (11)$$

where $\mathbf{F}_{1_j} = \mathbf{G}_{1_j} \Delta_{1_j}^{\frac{1}{2}}$, $\mathbf{F}_{2_j} = \mathbf{G}_{2_j} \Delta_{2_j}^{\frac{1}{2}}$, $\mathbf{F}_{(n-1)_j} = \mathbf{G}_{(n-1)_j} \Delta_{(n-1)_j}^{\frac{1}{2}}$ and $\mathbf{L}_{n_j} = \mathbf{H}_{n_j} \Delta_{n_j}^{\frac{1}{2}}$. $\mathbf{F}_{1_j}, \mathbf{F}_{2_j}, \dots, \mathbf{F}_{(n-1)_j}$ denote the cross channel coefficients from BS-1, BS-2, ..., BS- $n-1$ to j th user of Group- n , respectively. $\mathbf{L}_{n_j}$ denotes the direct channel matrix between BS- n and the j th user of Group- n . $Q_1, Q_2, \dots, Q_n$ denote the number of radio ports located at the BS-1, BS-2, ..., BS- n respectively. $\Delta_{n_j}^{\frac{1}{2}} = \text{diag} \left(\sqrt{\frac{\sigma_n}{\lambda_n^v}}, \dots, \sqrt{\frac{\sigma_n}{\lambda_n^v}}, \dots, \sqrt{\frac{\sigma_{Q_n}}{\lambda_{Q_n}^v}}, \dots, \sqrt{\frac{\sigma_{Q_n}}{\lambda_{Q_n}^v}}, \dots \right) \in \mathbb{R}^{Q_n M_n \times Q_n M_n}$ represents a diagonal matrix to count the

large-scale fading effect in which the path loss is characterized by $\lambda_{Q_n}^v$ for some exponent v . $\{\sigma_i\}_{i=1}^{Q_n}$ denote the independent random variables. ς_{n_j} denotes the received noise vector at the j th user of Group- n . Similarly as (2), the received signal at the eavesdropper is given by

$$\mathbf{y}_e^{\text{D-MIMO}} = \sqrt{\frac{P_1}{Q_1 M_1}} \mathbf{U}_1 \mathbf{x}_{1j} + \sqrt{\frac{P_2}{Q_2 M_2}} \mathbf{U}_2 \mathbf{x}_{2j} + \dots + \sqrt{\frac{P_{n-1}}{Q_{n-1} M_{n-1}}} \mathbf{U}_{n-1} \mathbf{x}_{n-1j} + \sqrt{\frac{P_n}{Q_n M_n}} \mathbf{U}_n \mathbf{x}_{nj} + \varsigma_e, \quad (12)$$

where $\mathbf{U}_n \in \mathbb{C}^{N_e \times M_n Q_n}$ denotes the direct channel coefficient between BS- n and eavesdropper. $\mathbf{U}_1 \in \mathbb{C}^{N_e \times M_1 Q_1}$, $\mathbf{U}_2 \in \mathbb{C}^{N_e \times M_2 Q_2}, \dots, \mathbf{U}_{(n-1)} \in \mathbb{C}^{N_e \times M_{n-1} Q_{n-1}}$ denote the cross channel coefficients from BS-1, BS-2, ..., BS- $n-1$ to eavesdropper, respectively. ς_e denotes the received noise vector at the eavesdropper. Similar to C-MIMO model, linear detector matrices Υ_{n_j} and Φ_n are dependent on $\mathbf{L}_{n_j}$ and $\mathbf{U}_n$, respectively.

A. MRC Receiver

In the case of MRC detectors at the receiver and eavesdropper, we have $\Upsilon_{n_j} = \mathbf{L}_{n_j}$ and $\Phi_n = \mathbf{U}_n$. Hence $v_{n_j,i} = \mathbf{l}_{n_j,i}$ and $\phi_{n_p} = \mathbf{u}_{n_p}$, where $v_{n_j,i}$ and $\mathbf{l}_{n_j,i}$ are the i th column of Υ_{n_j} and $\mathbf{L}_{n_j}$ matrices, respectively. ϕ_{n_p} and $\mathbf{u}_{n_p}$ are the p th column of Φ_n and $\mathbf{U}_n$ matrices, respectively. From (7), the achievable ergodic secrecy rate for the i th antenna of j th user of Group- n is given by

$$\langle R_{S_n}^{\text{D-MRC}} \rangle = \mathbb{E} \left\{ \log_2 \left(\frac{1 + \frac{\frac{P_n}{Q_n M_n} \|\mathbf{l}_{n_j,i}\|^4}{\xi_n}}{1 + \frac{\frac{P_n}{Q_n M_n} \|\mathbf{u}_{n_p}\|^4}{\rho_n}} \right) \right\}, \quad (13)$$

where $\xi_n = \sum_{m=1}^{n-1} \frac{P_m}{Q_n M_m} |\mathbf{L}_{n_j}^\dagger \mathbf{F}_{m_j}|^2 + \|\mathbf{L}_{n_j}\|^2$ and $\rho_n = \sum_{m=1}^{n-1} \frac{P_m}{Q_n M_m} |\mathbf{U}_n^\dagger \mathbf{U}_m|^2 + \|\mathbf{U}_n\|^2$.

B. ZF Receiver

In the case of ZF detectors at the receiver, we have $\Upsilon_{n_j}^\dagger = (\mathbf{L}_{n_j}^\dagger \mathbf{L}_{n_j})^{-1} \mathbf{L}_{n_j}^\dagger$ or $\Upsilon_{n_j}^\dagger \mathbf{L}_{n_j} = \mathbf{I}_{M_n}$. Hence $v_{n_j,i} \mathbf{l}_{n_j,t} = \delta_{i,t}$, where

$$\delta_{i,t} = \begin{cases} 1, & \text{for } i=t \\ 0, & \text{otherwise.} \end{cases}$$

Similarly, at the eavesdropper, we have $\Phi_n^\dagger = (\mathbf{U}_n^\dagger \mathbf{U}_n)^{-1} \mathbf{U}_n^\dagger$ or $\Phi_n^\dagger \mathbf{U}_n = \mathbf{I}_{M_n}$. Hence $\phi_{n_p}^\dagger \mathbf{u}_{n_t} = \delta_{p,t}$, where

$$\delta_{p,t} = \begin{cases} 1, & \text{for } p=t \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, using (7), the achievable ergodic secrecy rate for the i th antenna of j th user of Group- n is given by

$$\langle R_{S_n}^{\text{D-ZF}} \rangle = \mathbb{E} \left\{ \log_2 \left(\frac{1 + \frac{\frac{P_n}{Q_n M_n}}{\tau_n}}{1 + \frac{\frac{P_n}{Q_n M_n}}{\omega_n}} \right) \right\}, \quad (14)$$

where $\tau_n = \sum_{m=1}^{n-1} \frac{P_m}{Q_n M_m} |(\mathbf{L}_{n_j}^\dagger \mathbf{L}_{n_j})^{-1} \mathbf{L}_{n_j}^\dagger \mathbf{F}_{m_j}|^2 + |\mathbf{L}_{n_j} (\mathbf{L}_{n_j}^\dagger \mathbf{L}_{n_j})^{-1}|^2$ and $\omega_n = \sum_{m=1}^{n-1} \frac{P_m}{Q_n M_m} |(\mathbf{U}_n^\dagger \mathbf{U}_n)^{-1} \mathbf{U}_n^\dagger \mathbf{U}_m|^2 + |\mathbf{U}_n (\mathbf{U}_n^\dagger \mathbf{U}_n)^{-1}|^2$.

C. MMSE Receiver

In the case of MMSE detectors at the receiver, we have $\Upsilon_{n_j}^\dagger = (\mathbf{L}_{n_j}^\dagger \mathbf{L}_{n_j} + \frac{M_n}{P_n} \mathbf{I}_{M_n})^{-1} \mathbf{L}_{n_j}^\dagger = \mathbf{L}_{n_j}^\dagger (\mathbf{L}_{n_j} \mathbf{L}_{n_j}^\dagger + \frac{M_n}{P_n} \mathbf{I}_{N_n})^{-1}$. Similarly, in the case of MMSE detectors at the eavesdropper, we have $\Phi_n^\dagger = (\mathbf{U}_n^\dagger \mathbf{U}_n + \frac{M_n}{P_n} \mathbf{I}_{M_n})^{-1} \mathbf{U}_n^\dagger = \mathbf{U}_n^\dagger (\mathbf{U}_n \mathbf{U}_n^\dagger + \frac{M_n}{P_n} \mathbf{I}_{N_e})^{-1}$. The achievable ergodic secrecy rate for the i th antenna of j th user of Group- n is given by

$$\langle R_{S_n}^{\text{D-MMSE}} \rangle = \mathbb{E} \left\{ \log_2 \left(\frac{1 + \frac{\frac{P_n}{Q_n M_n} |v_{n_j,i}^\dagger \mathbf{l}_{n_j,i}|^2}{\varphi_n}}{1 + \frac{\frac{P_n}{Q_n M_n} |\phi_{n_j,i}^\dagger \mathbf{u}_{n_j,i}|^2}{\chi_n}} \right) \right\},$$

where $\varphi_n = \sum_{m=1}^{n-1} \frac{P_m}{Q_n M_m} |\Upsilon_{n_j}^\dagger \mathbf{F}_{m_j}|^2 + \|\Upsilon_{n_j}\|^2$ and $\chi_n = \sum_{m=1}^{n-1} \frac{P_m}{Q_n M_m} |\Phi_{n_j}^\dagger \mathbf{U}_{m_j}|^2 + \|\Phi_{n_j}\|^2$.

V. NUMERICAL RESULTS

Fig.2 shows the achievable ergodic secrecy rate as a function of the number of antennas at BS- n . Here we consider two types of MIMO configurations such as C-MIMO and D-MIMO. For each configuration, three types of detectors such as MRC, ZF and MMSE are considered. We see that the performance of D-MIMO is better than that of C-MIMO in the case of MRC, ZF and MMSE detectors. If we compare the performance of ZF and MMSE detectors at the receiver of D-MIMO system, we see that ZF shows better performance than MMSE.

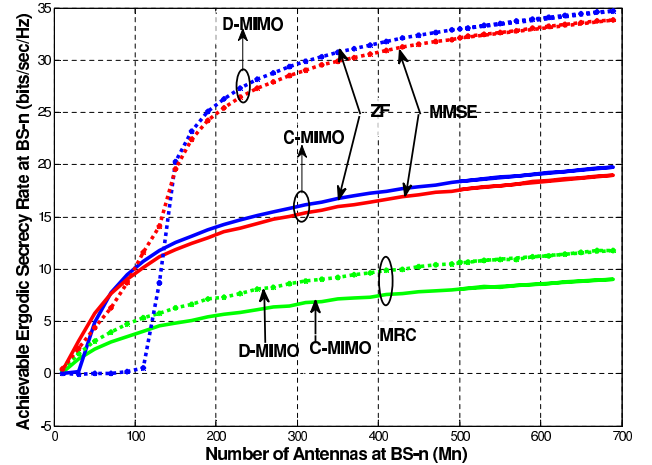


Fig. 2. The achievable ergodic secrecy rate as a function of the number of antennas at BS- n with $K_n = 4$, $N_n = 4$, $N_e = 3$, $Q_n = 4$.

The effects of interferences on the performance of MRC detector are shown Figs. 3 and 4 for two and four consecutive cells, respectively. Note that for all the figures, the dotted

curves and solid curves show the achievable ergodic secrecy rate for D-MIMO and C-MIMO systems, respectively. We see that the effect of interferences on the D-MIMO system is greater than that of C-MIMO system. But when the number of cells increases, the performance of D-MIMO and C-MIMO systems are almost same. It means that in the high interference region, both the D-MIMO and C-MIMO systems show the similar performance. Figs. 5 and 6 show the effects of

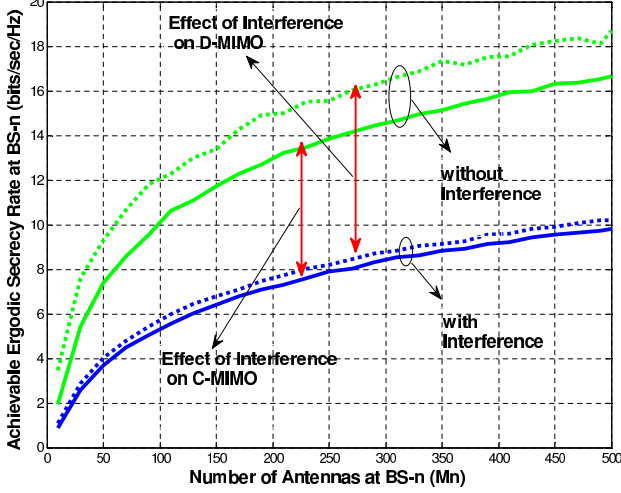


Fig. 3. The effect of interference on the performance of MRC for two consecutive cells with $K_n = 4$, $N_n = 4$, $N_e = 3$, $Q_n = 4$.

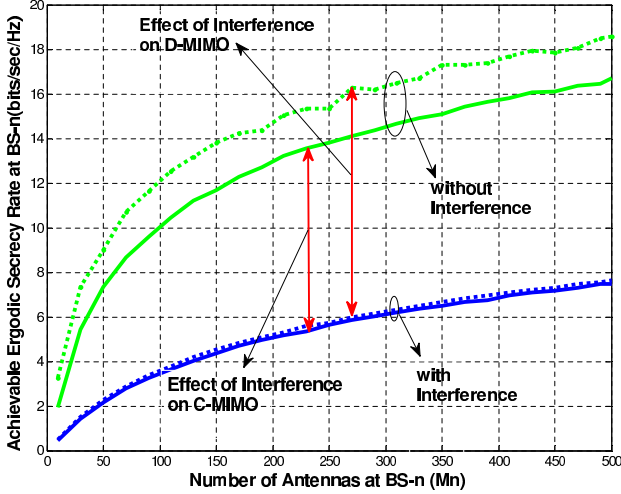


Fig. 4. The effect of interference on the performance of MRC for four consecutive cells with $K_n = 4$, $N_n = 4$, $N_e = 3$, $Q_n = 4$.

interferences on the performance of ZF detector for two and four consecutive cells, respectively. It is observed that the effect of interferences on the D-MIMO system is very significant than that of C-MIMO system. Moreover, in the case of C-MIMO system, the effect of interference is greater for MRC detector compared to ZF detector at the receiver.

Similar to the phenomena of Figs. 3, 4, 5 and 6, the effects of interferences on the performance of MMSE detector are

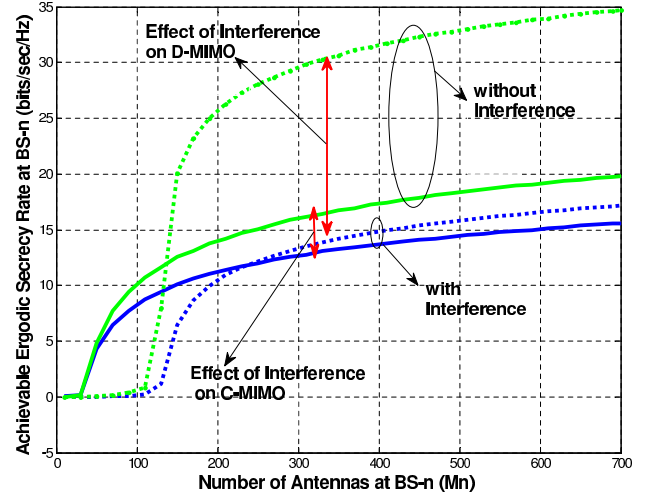


Fig. 5. The effect of interference on the performance of ZF for two consecutive cells with $K_n = 4$, $N_n = 4$, $N_e = 3$, $Q_n = 4$.

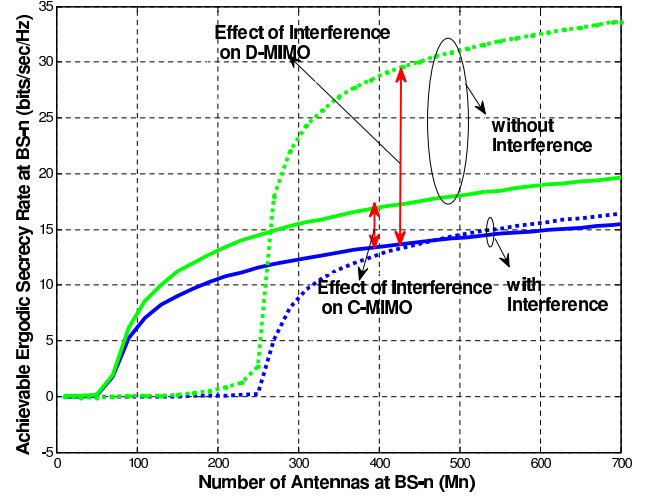


Fig. 6. The effect of interference on the performance of ZF for four consecutive cells with $K_n = 4$, $N_n = 4$, $N_e = 3$, $Q_n = 4$.

depicted in Figs. 7 and 8 for two and four consecutive cells, respectively. Comparing the performance of MRC, ZF and MMSE, we can conclude that when the number of transmit antenna is very small, the MRC and MMSE detectors perform better than the ZF detector.

VI. CONCLUSION

We investigate the performance of MRC, ZF and MMSE detectors at the receiver in enhancing wireless security of multi-cellular C-MIMO and D-MIMO interference networks. Based on our observations and numerical results, we can conclude that the D-MIMO system is more interference sensitive than C-MIMO system. But, in the high interference region, the effect of interference on the D-MIMO and C-MIMO systems are almost same. In the case of C-MIMO system, the effect of interference is greater for MRC detector compared to ZF detector at the receiver. When the number of transmit antenna

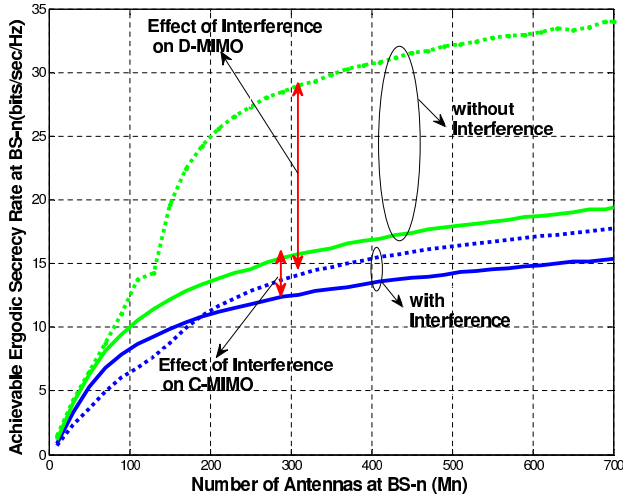


Fig. 7. The effect of interference on the performance of MMSE for two consecutive cells with $K_n = 4$, $N_n = 4$, $N_e = 3$, $Q_n = 4$.

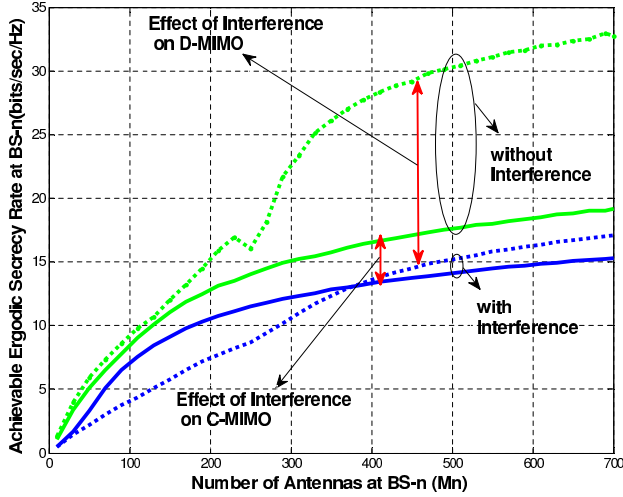


Fig. 8. The effect of interference on the performance of MMSE for four consecutive cells with $K_n = 4$, $N_n = 4$, $N_e = 3$, $Q_n = 4$.

is very small, the MRC and MMSE detectors perform better than the ZF detector.

REFERENCES

- [1] M. X. Gervais N. Kamga and S. Aissa, "Spectral-efficiency analysis of massive MIMO systems in centralized and distributed schemes," *IEEE Transactions on Communications*, vol. in press, January 2016.
- [2] P. Z. X. T. Chunlong He, Nanjing, "Energy efficient comparison between distributed MIMO and co-located MIMO in uplink cellular system," in *Proc. IEEE Vehicular Technology Conference (VTC Fall)*, 03-06 Sept. 2012, pp. 1–5.
- [3] D. G. Arghavan Emami Forooshani, Abbas Ali Lotfi-Nayestanak, "Optimization of antenna placement in distributed MIMO systems for underground mines," *IEEE Transactions on Wireless Communications*, vol. 13, no. 9, pp. 4685–4692, July 2014.
- [4] X. Y. Dongming Wang, Jiangzhou Wang and Y. Wang, "Spectral efficiency of distributed MIMO systems," *Journal on Selected Areas in Communications*, vol. 31, no. 10, pp. 2112–2127, 17th September 2013.
- [5] M. M. A. M. Z. I. Sarkar, K. N. Faisal, "Performance analysis of nakagami-m fading massive MIMO channels with linear receivers," in *International Conference on Computer and Information Technology (ICCIT)*. IEEE, 22-23 Dec. 2014, pp. 510–513.

- [6] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*. New York: Cambridge University Press, 2005.

A Packet Level Simulation Study of Adhoc network with Network Simulator-2 (NS-2)

Iftekharul Mobin¹, Sifat Momen², and Nabeel Mohammed³

Department of Computer Science and Engineering
University of Liberal Arts Bangladesh, 4/A Dhanmondi, Dhaka-1209
E-mail: {iftekharul.mobin¹, sifat.momen², nabeel.mohammed³}@ulab.edu.bd

Abstract—Researchers in the field of computer networking often relies on appropriate simulators to diagnose network and to analyze the impact of various parameters on the network performance. There now exists several simulators (in the areas of wireless ad-hoc network)that offer various features to measure network performance. One such simulator that recently gained high popularity among researchers in the area of networking is Network Simulator 2 (NS-2). In this paper, we present a short survey on NS-2 for packet level simulation. The paper demonstrates how packet level simulation is conducted using NS-2. The paper furthermore explores the advantages and the drawbacks of NS-2 in the light of packet loss, interval period, traffic type and pattern. The techniques used in this work to measure network performance can be replicated in other related work. Simulated results with the NS-2 simulator show remarkable correlation with that of the theoretical estimation. The most remarkable contribution of this paper is this paper shows simulation for packet level analysis with NS-2 with details parameter setting, network scenarios and other necessary configurations which will lead the researchers to investigate network efficiency and performance measurements.

Index Terms—Simulation, Noisy Channel, NS-2, Multi-hop, Network Performance, Packet Efficiency

I. INTRODUCTION

Simulation is the imitation of how a model of a system or a real-world system operates. Simulations allow us to analyze or visualize the impact of experiments in a controlled and repeatable environment without real-time implementation. Most of the researches pertaining to the performance analysis of wireless network is conducted through simulation tools [1], [2]. Some of the simulation tools that are heavily used in the wireless research network community include NS-2 [3]–[5], OPNET [6], GloMoSim [7], J-sim [8], MATLAB [9] and OMNeT++ [10]. A comparison of different network simulator is given in table I [11].

Although these simulators are popular for simulating network, they have their own strengths and weaknesses in terms of protocols that they support, simulation methodologies, evaluation techniques, run-time performance etc [1]. Prior to the starting a simulation, one has to decide upon the simulation criteria, parameter settings and configuration. However, these features vary from one tool to another. Hence this is extremely crucial to know of the features offered by simulation tools used by the researchers.

The open source simulator, NS-2, is found to be a popular choice among research communities in studying packet level simulation for adhoc networks [4], [12]–[14]. Therefore we will constrict our study using the NS-2 simulator only. In this paper we systematically study the pros and cons of the packet level simulation of adhoc network. This paper presents a detailed analysis of packet loss, packet interval rate, traffic pattern, packet size and multi-hop network in adhoc networks. Most of the analysis and research have been done before was not analyzed considering the specific network scenario and network setting [4], [12]–[14]. Only the network facilities and which parameters setting are available that indication are given. But how the experiments can be conducted and how to analyze performance, investigation for efficiency are revealed in this paper.

A. NS-2 as Adhoc Network Packet Level Simulator

Network Simulator (version 2), widely known as NS-2, [4] is one of the most popular choices for performance analysis of wired and wireless networks [12]–[14]. NS-2 simulator [4], [5] is written in C++ which is the core engine and uses OTcl, an object oriented version of TCL language for command, configuration and simulation scripts. The C++ code describes the parameters and methods that are to be made available for OTcl scripting. By using OTcl scripts, one can easily set up network topology, configure nodes, traffic pattern etc. The use of two languages provide two major advantages:

- 1) C++ is slow to change but fast to run.
- 2) OTcl is slow to run but easy to change within the simulation script

In 2004, NS-2 introduced the Wireless Local Area Network (WLAN) functionality. [4], [5]. Furthermore, the mobility of nodes and realistic OSI layer are incorporated within the NS-2 core engine. After this, the NS-2 simulator emerged as a well known simulator for adhoc network for the packet level simulation. NS-2 has many advantages compared to other simulators for packet level simulation owe to some unique features that the simulator offers. It can measure the mobility of nodes, co-ordinates of the nodes, energy, packet behavior, traffic pattern, channel condition etc [15]. Most of the model of wireless networks in NS-2 follows theoretical formulas which are well established in network performance measurements [16]–[19].

TABLE I
COMPARISON OF DIFFERENT NETWORK SIMULATOR IN GENERAL

Attributes	NS-2	J-Sim	Omnet++	Qualnet	OPnet
Coding Style	C++	Java	C++/JAVA	Parse C++	C++/Java
Simulation Time	Moderate	High	Average	High	High
User Interface	Minimal	Average	Average	High Quality	High Quality
Analysis Tools	Custom built	Custom Built	Built in	Built in	Built In
License	Freeware	Freeware	Freeware	Commercial	Commercial
User Adaptability	Long	Moderate	Moderate	Very Easy	Long
Documentation	Poor	Moderate	Moderate	Good	Excellent

B. Major Advantages for simulation with NS-2

NS-2 offers a number of advantages. Some of the major advantages offered by NS-2 are discussed below:

- Parameters such as packet size, traffic patterns(e.g. exponential, pareto, uniform etc...), traffic rate, co-ordinates of the nodes) are straightforward and easy to maintain. The size of the packet can also be changed in run-time which is sometimes crucial for estimating the optimal packet size [16]–[19]. NS-2 can provide extensive information about packet level simulation analysis (see Tab. II).
- The error model of NS-2 can characterize the bursty packet loss behavior of noisy channel. It follows the theoretical equations of the Bit Error Rate (BER) and Packet Error Rate (PER).
- It has built in Forward Error Correction (FEC) module that maintains a formula similar to that used in the theoretical understanding of error correction [20], [21].
- The built in energy model of NS-2 can be used to estimate the energy consumption of particular nodes as well entire simulated network during packet transmission [15].
- NS-2 is a freeware, has large community and lots of built in examples of simulation scripts.

C. Disadvantages for Simulation with NS-2

NS-2 also has some disadvantages in terms of a network simulator. Following are some of the disadvantages of NS-2:

- To extract the simulation results, the trace file (see section II) needs to be parsed with either other post processing tools or scripting languages (e.g. awk, perl, excel etc) or even manually.
- Installation process is quite complex and time consuming. Ideally, NS-2 needs to be installed in Unix environments. In windows environment it can be installed by installing CYGWIN [22].
- Source code customization is complex and time consuming.

Comparing the pros and cons, NS-2 is generally accepted as one of the most beneficial tools for doing simulations for network research. A brief description of NS-2 focusing packet level simulation is given in the following sections.

D. Simulation Process of NS-2

The simulation is processed mainly in four steps [5]. At first, the simulation and parameter settings are described in a TCL script. The simulation TCL script is run and outputs

are generated in the trace file. After that the trace file is analyzed with other programming languages. The results of the simulation are written in a trace file after every millisecond intervals containing detailed information about the network attributes (e.g. node id, coordinates, packet size, type etc). Table II shows the different parameter specifications of a trace file.

By processing this trace file with post processing scripting language (by using awk, perl etc) delay, throughput, packet loss, energy level, routing path etc can be analyzed and monitored.

II. CHANNEL NOISE AND BIT ERROR RATE (BER)

In a digital transmission, BER is defined as the percentage of bits with errors per unit total number of bits that have been transmitted, received or processed over a given time period. BER is usually equivalent to Signal to Noise Ratio (SNR) but not necessarily the same as SNR [24], [25]. SNR measures the quality of a transmission channel or an audio signal over a network channel. Within a network, many kinds of noise can be present including thermal noise, platform noise, noise from environment and interference caused by other packets received by the receiver simultaneously. If only one packet is received by the receiver the SNR can be calculated by the following formula:

$$SNR = 10 \log \frac{P_{receive}}{P_{noise}} \quad (1)$$

Where, P_{noise} is the noise power and $P_{receive}$ is the Receiver power that can be calculated from the propagation model. SNR of zero indicates that the received signal power is same as Carrier sense noise [3] and is virtually indistinguishable from the unwanted noise.

A. Packet Error Rate and Bit Error Rate Estimation

There is a direct relationship between the PER and the BER (see Eq. 2). NS-2 simulator converts PER to BER by using the following formula [24], [25]

$$\begin{aligned} P_p &= 1 - (1 - P_e)^N \\ P_e &= 1 - (1 - P_p)^{\frac{1}{N}} \end{aligned} \quad (2)$$

Where, P_p is packet error rate (PER), P_e is the bit error rate (BER) and N is the packet size in bits. For a small bit error the Eq. 2 can be approximated as $P_p \approx P_e N$. That is why when the packet size increases the error probability increases which consequently leads to more packet error and hence packet drop.

TABLE II
NS-2 TRACE FILE FORMAT AND DIFFERENT FIELDS' SPECIFICATIONS [3]–[5]

Event Information	Specifications	Tag
Event Type	Send, Receive, drop, forward	S, R, D,F
Time	Packet transmission time in milliseconds	t
Source and destination Id	Source and Destination node Id	Hs, Hd
Current Node ID	Intermediate node Id	Ni
Nodes Position	X, Y, Z coordinate	Nx, Ny, Nz
Nodes Energy	Nodes Energy level	Ne
Trace Level	Router, Agent or MAC trace	Nl
Reasons of packet drop	Queue full, TTL, MAC Error etc	Nw
Packet Information IP Level	Packet type, size, TTL, flow Id etc	Is, It, If, Iv
Packet info MAC level	Ethernet address, type etc	Ma, Mt
Packet Info level	TCP, UDP, DSR, DSDV, AODV etc	Po, Ps, Pp etc

B. Channel Noise in A Multi-hop Topology

In a multi-hop network [26], [27] the nodes at the edges do not send data directly. Instead intermediary nodes are used as routers for data forwarding. Two nodes remain at the edges while the intermediary nodes between the source and the destination nodes participate in forwarding data. Since there is a probability of the inclusion of noise in each hop, multi-hop routing results in the noise to increase significantly. Consequently the PER for multi-hop routing deviates from that in Eq.2. To find out the efficient packet size for a multi-hop environment with a constant BER, PER for multi-hop routing follows the equation given below:

$$P_p = 1 - (1 - PER)^N \quad (3)$$

In this Eq. 3, P_p is the new packet error rate (PER) for multiple hops.

C. NS-2 Error Model

NS-2 imposes error on packet transmission by using its error module [3]–[5]. This module simulates the number of error packets after receiving a packet. If packet is simulated as an error packet, either the packet is dropped or marked as corrupted. Error model can be used for wired and wireless networks. Errors can be generated randomly in a controlled way by defining the error rate [3], [25] and error unit in the simulation (TCL) script. The user only needs to describe the error procedure in TCL code in the simulation scripts.

According to the packet error rate (PER) defined in the simulation script, simulator creates a Bernoulli distribution with uniform random variable from 0 to 1. NS-2 uses the following formula:

$$var[m] = \binom{size}{i} b^i (1 - b)^{size-i} \quad (4)$$

where, b is the bit error rate and m is counter variable. The value of m can be changed to number of maximum bits user wants to be corrupted during the simulation. No bit error will take place if ($var[0] \geq u$) since error function will return false. The variable $size$ is the total packet size, m changes with the iteration loop value. When $i = 0$ there is no error. Otherwise for every cases ($var[1]$, $var[2]$ etc) packet will

be corrupted. Here i specifies the number of bits that are corrupted. So, when $i = 1$, it means that only one bit in a packet is corrupted whereas when $i = 2$ two bits in a packet are corrupted and so on. However, this entire procedure has some limitations:

- 1) Simulation outputs are written in the trace file in packets and not in bits. It only specifies how many packets are corrupted and dropped but does not specify how many error bits the dropped packet contained.
- 2) BER (Bit error rate) is defined in the simulation scripts but the output is coming in PER (Packet error rate) which can be obtained from the trace file (by calculating $\frac{drop-packet}{sent-packet}$). Hence, the existing source codes need modified to see the bit error in the trace file.

Bit Error Rate is used in the simulation script and then measure the packet error rate by estimating ($\frac{drop-packet}{sent-packet}$). The packet error rate is calculated theoretically from the defined BER and compared with the trace file results.

III. SIMULATION STRATEGY

To verify the performance of various packet sizes for the adhoc network, a small network with two nodes is created in NS-2. After that the experiments are extended for complex multi-hop network topology. Simulation scenario files, communication pattern and configuring the adhoc node within the TCL script is needed to be done before starting the simulation. As the simulation is run for a particular scenario, one trace file is created for each run. The trace files are processed through a variety of scripts (e.g. perl, awk, shell etc). Further analysis can be made using Matlab or Microsoft Excel. They can be used to produce the graphs such as that depicted in the latter sections of the paper.

A. Simulator Parameter Settings

Most of the parameters used in simulation are default parameter settings of NS-2. Some of the default parameter settings are shown in Tab. III. The packet size and network topology are changed to see the packet drop or packet loss of different packet size for different topology.

TABLE III
PARAMETERS SETTINGS FOR THE SIMULATION

Simulator	NS-2 Version 2.34	
	Network Size	1000m x 1000m
	Number of Nodes	Variable
	Duration	100,000s
Physical Layer	Signal Propagation model	Two Ray Ground
	Max Transition Range	250m
	Antenna Model	Omni Antenna
Mac Layer	MAC protocol	802.11
	Link Bandwidth (data Rate, RTS/CTS)	1Mb, 1Mb (by default)
Traffic Model	Date Rate	20 Kb
	Traffic Type	CBR, UDP
	Traffic Interval Rate	Variable
Mobility Speed	MTU size	1000 byte (default)
	Random Waypoint Model	
	Nodes Average Speed	Random ($0 \sim 20ms^{-1}$)
	Pause Time	Variable
Queue	Type	Drop Tail/PriQueue
	Size	50 (default)
TTL	32	

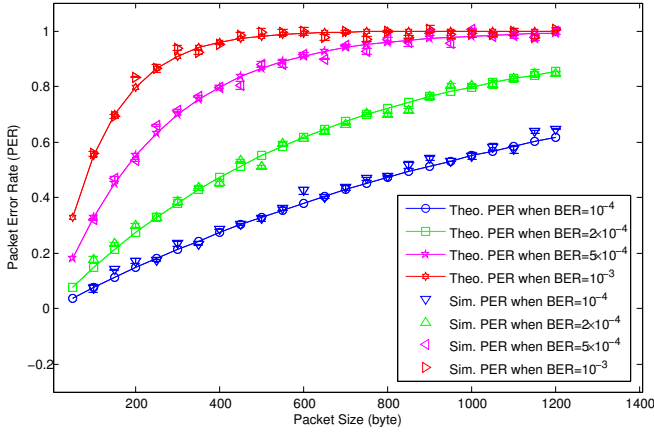


Fig. 1. For different BER theoretical PER and simulated PER is compared with error bar for two hops

B. Simulation with various packet size and fixed data rate

Simulation is run with a BER of 10^{-4} and packet sizes varying (randomly) from 50 bytes to 1200 bytes. The simulation is run for 10,000 sec long and is repeated 10 times for each packet size with a constant data rate of 20 Kb/sec. The number of propagated Packets PER Second (PPS) is different. To achieve the 20 Kb data rate with 50 bytes packet sizes, the simulator needs to send almost 50 packets per second as $20Kb = 20 \times 10^3bits$, thus $n \times 50 \times 8$ need to be 20×10^3bits resulting $n = 50$ packets. Thereby, for the 1200 bytes packet size simulator will send 2.08 packet per second approximately. Hence, the numbers of packets are different for other packet sizes accordingly. However, the number of packets doesn't have effect on the error model because NS-2 follows the Eq. 2 to calculate the bit error probability and corrupt packet according to the error probability. When the numbers of packets are more for a fixed data rate it actually affects the other parameters of the network such as: queuing

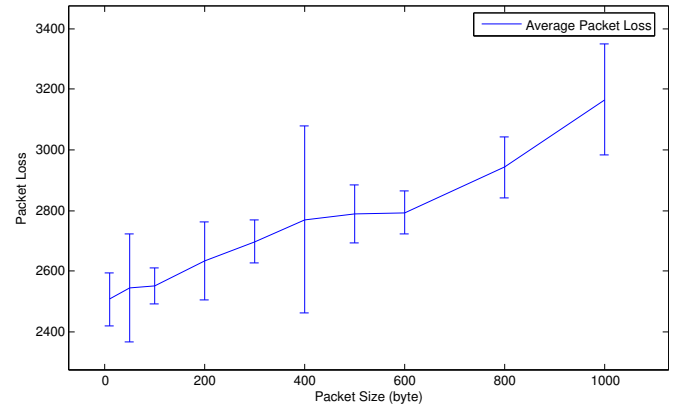


Fig. 2. Average packet drop of various packet size from 10 bytes to 1000 bytes with fixed packet rate (5% packet error rate)

capacity of the nodes, collision probability etc.

C. Simulation with fixed packet size and different data rate

We also simulated network with different data rate and fixed packet size (which is varied from 10 bytes to 3000 bytes for experimentation purposes). Simulation is run for 10,000 sec and each simulation is repeated 10 times for a fixed packet size. The packet error rate (PER) is kept at 5% and interval rate i.e. Packet Per Second (PPS) is fixed at 100 packet per second. Different packet sizes result in different data rates. For the 10 bytes packet size the data rate is $100sec \times 8 \times 10 = 8000bits/sec$ whereas for a 3000 bytes packet size data rate becomes $100sec \times 8 \times 3000 = 2.4Mbits/sec$. In this simulation the same number of packets is sent per second - resulting the error model to receive the same number of packet per second. Thus the error model becomes active for every packet size equally. Figure 2 shows how the same number of different packet sizes interact with erroneous channel.

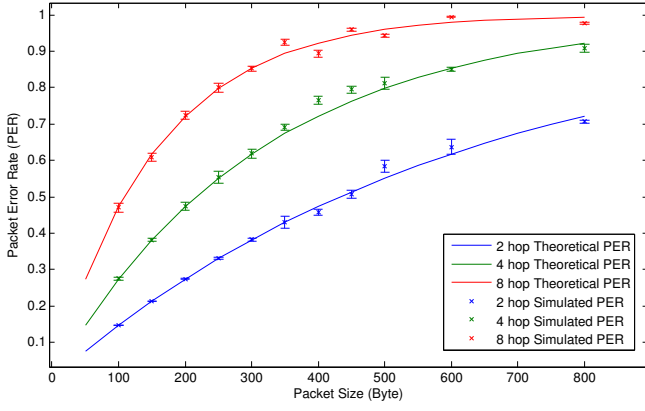


Fig. 3. Theoretical PER and simulated PER comparison in the multi-hop linear chain topology

D. Simulation with Multi-hop Topology Network

To do experiment for a multi-hop network, a multi-hop topology is created with three nodes. There is an intermediate node that works as a router. The edge nodes send data, other nodes act as a receiver and receive data. The edge nodes cannot send data directly so they need to use the intermediate nodes as a router for forwarding data. As there is a data error rate included in each link for every hop, the PER will also be different. The PER should maintain Eq. 3. From the output it is found that the packet error rate does not vary for the three nodes topology. It maintains the equation shown in Eq. 3. To investigate the variations in PER, the hop number is increased linearly and simulation is repeated 10 times. When there are three hops (i.e. four nodes), the packet error rate remains at 5%. But for the five nodes topology (i.e. for four hop distances), the packet drops significantly and is found different in comparison to the theoretical PER (Fig. 3).

IV. RESULTS AND DISCUSSION

The experiments are carried out for various packet sizes with UDP packet (20 Kbps data rate) to compare the theoretical error rate with the simulated error rates. The simulation is repeated 10 times each running for 10,000 seconds each for different packet sizes with constant data rate. Figure 1 shows how the simulated PER varies with the packet size (at different BER). The figure reveals that when the data rate is kept constant, the simulated results show close resemblance with the theoretical predictions.

From Fig. 2 it can be further observed that with a constant BER of 10^{-4} , the PER increases linearly with the packet size (Eq. 2). The simulation is conducted for 3000 sec 10 times each for each experiment and took the average values for the sent packets, received packets and PER. The numbers of sent and dropped packets are calculated from the NS-2 trace file where every packet's status is recorded automatically. After that, the packet error rate is calculated by $\frac{\text{drop} - \text{packet}}{\text{sent} - \text{packet}}$ and the average packet error rate is measured for the simulation results for every packet size. Figure 1 shows that the theoretical packet error rate matches our simulated packet error rate for

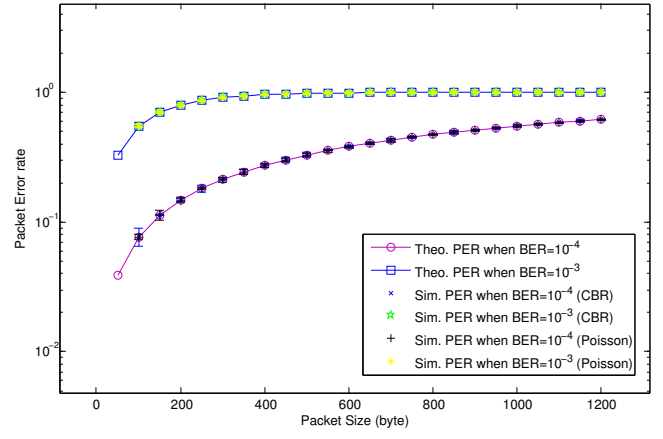


Fig. 4. Theoretical PER and simulated PER comparison with different traffic flow

10^{-4} BER. So, our theoretical estimation is correct for a constant bit error rate of a two nodes topology where the packet size is small. However, as the packet size increases the error probability increases with it resulting in a higher deviation from the theoretical estimation (for 1000 and 1200 byte packet size). The BER between the nodes is independent. Therefore, in a multi-hop network the PER follows Eq. 3. In order to validate our assumption, an experiment is conducted between 5 and 9 nodes in a multi-hop network where the hop size is gradually increased from 2 to 8. The simulated result shows that our assumption is correct. When the sender and receiver nodes use the same channel to send data at the same time there will be competition between them to acquire the channel resources. When one node sends data to another node, it needs to wait otherwise collision may occur. When the number of packets is more than the long list of queue, some packets may get dropped because if the channel is not free for a long time the packet in the source node will make the queue full and causing packets to be dropped. Also the packet cannot be transmitted directly until it gets the channel free or wins in the competition randomly to allocate channel. Thus there is more packet drop and more error occurs than our theoretical error rate.

Packet error is estimated for different types of traffic such as: POISSON, Constant Bit Rate (CBR), with average data rate 20 Kbps. From Fig. 4 it is observed that using different traffic patterns does not vary the average packet error rate theoretically with simulation results. From the experiments, we find that theoretical error rates match with the simulation results with low error bars. It is also revealed that for the low average data rate theoretical PER and simulated PER remain almost the same for various traffic flows.

A. Result Analysis

From the simulation it is further observed that packets of packet sizes from 10 bytes to 1000 bytes, the average packet loss trend is almost linear and is a function of the packet size. But after packet size reaches the MTU size (1000 bytes), the number of packet loss suddenly goes high. Analyzing the

simulation trace file it is observed that the number of packets is more than our defined packet rate. This is because after 1000 bytes when the packet size is more than MTU the packets are split into two (1000 bytes and 200 bytes) packets; hence increases the number of packets within the channel that increases the probability of corrupting more packet. For the 1200, 1500, 2000, 2500 and 3000 bytes packet size it is shown that the number of packet drops are 11729, 11809, 12666, 37378 and 51541. The number of sent packets during the simulation for the 1200, 1500 and 2000 bytes packet size is 1999980 whereas from 10 to 1000 bytes the number of sent packet is 999990 which is exactly the half number. Similarly for the 2500 and 3000 bytes packet size the number of sent packet is three times greater 2999970 packets in the simulation. Therefore, more than MTU size packets are error prone and inefficient. Due to large number of packets smaller packets are affected by the channel congestion comparatively more than bigger packet sizes. But bigger packet sizes increased the error probability. When the packet sizes is more than MTU it splits into two and increase the number of packet and data rate.

V. CONCLUSION

This paper discusses the suitability of NS-2 as a simulator for packet level simulation. We present a range of packet level simulation experiment with NS-2. Simulation results are then compared with theoretical estimation for single-hop, multi-hop as well as different kinds of traffic (including that of CBR, Poisson and uniform). The results indicate a strong match (with low standard deviation) between the theoretical estimation and the simulated results. Thus the simulation results using NS-2 is found to be well validated when compared with the real data. The analysis and experiments conducted in this paper will provide a remarkable guideline for the network engineers and other researchers to evaluate packet level performance of adhoc network. The further research plan is to investigate the reliability of NS-2 for complex adhoc network scenarios having large number of nodes, congestion control and dynamic packet scheduling.

REFERENCES

- [1] M. Korkalainen, M. Sallinen, N. Karkinen, and P. Tukeya, "Survey of wireless sensor networks simulation tools for demanding applications," in *Networking and Services, 2009. ICNS'09. Fifth International Conference on*. IEEE, 2009, pp. 102–106.
- [2] V. Rajendran, K. Obraczka, and J. J. Garcia-Luna-Aceves, "Energy-efficient, collision-free medium access control for wireless sensor networks," *Wireless Networks*, vol. 12, no. 1, pp. 63–78, 2006.
- [3] K. Fall and K. Varadhan, "The Network Simulator NS-2 online documentation," <http://www.isi.edu/nsnam/ns/doc/index.html>, MONARCH Group, 2007.
- [4] T. Issariyakul and E. Hossain, *Introduction to Network Simulator NS2*, 1st ed. Springer Publishing Company, Incorporated, 2008.
- [5] C. Jae and C. Mark, "NS-2 by example," <http://nile.wpi.edu/NS/>, WPI Computer Science, 2007.
- [6] X. Chang, "Network simulations with OPNET," *Proceedings of the 31st conference on Winter simulation: Simulation—a bridge to the future*, vol. 1, pp. 307–314, 1999.
- [7] X. Zeng, R. Bagrodia, and M. Gerla, "GloMoSim: a library for parallel simulation of large-scale wireless networks," in *Parallel and Distributed Simulation, 1998. PADS 98. Proceedings. Twelfth Workshop on*, May 1998, pp. 154–161.
- [8] A. Sobeih, W.-P. Chen, J. C. Hou, L.-C. Kung, N. Li, H. Lim, H.-Y. Tyan, and H. Zhang, "J-sim: A simulation environment for wireless sensor networks," in *Proceedings of the 38th annual Symposium on Simulation*. IEEE Computer Society, 2005, pp. 175–187.
- [9] J. H. Mathews and K. D. Fink, *Numerical methods using MATLAB*. Prentice hall Upper Saddle River, NJ, 1999, vol. 31.
- [10] A. Varga, "The OMNeT++ discrete event simulation system," *Proceedings of the European Simulation Multiconference (ESM'2001)*, June 2001.
- [11] V. Mishra and S. Jangale, "Analysis and comparison of different network simulators," *International Journal of Application or Innovation in Engineering & Management*, 2014.
- [12] A. Chaudhary, S. Sachdev, T. Pradhan, and S. Kamath, "Improving performance of manet via cooperative communication with selective cooperation method considering dynamic traffic pattern using ns2," in *High Performance Computing and Applications (ICHPCA), 2014 International Conference on*. IEEE, 2014, pp. 1–5.
- [13] D. Kliazovich, P. Bouvry, and S. U. Khan, "Greencloud: a packet-level simulator of energy-aware cloud computing data centers," *The Journal of Supercomputing*, vol. 62, no. 3, pp. 1263–1283, 2012.
- [14] V. Kanakaris, D. Ndzi, and G. Papakostas, "Sensitivity analysis of aodv protocol regarding forwarding probability," *Optik-International Journal for Light and Electron Optics*, vol. 127, no. 3, pp. 1016–1021, 2016.
- [15] R. Rachman, E. P. Laksana, D. S. Putra, and R. F. Sari, "Energy consumption at the node in underwater wireless sensor network (uwsns)," in *Computer Modeling and Simulation (EMS), 2012 Sixth UKSim/AMSS European Symposium on*. IEEE, 2012, pp. 418–423.
- [16] M. Choi, M. Kim, T. Lee, J.-M. Chung, and R. C. Garcia, "Frequency hopping mobile ad hoc network and adjacent channel interference implementation in ns-2," in *Consumer Electronics (ISCE 2014), The 18th IEEE International Symposium on*. IEEE, 2014, pp. 1–2.
- [17] Q. Chen, F. Schmidt-Eisenlohr, D. Jiang, M. Torrent-Moreno, L. Delgrossi, and H. Hartenstein, "Overhaul of ieee 802.11 modeling and simulation in ns-2," in *Proceedings of the 10th ACM Symposium on Modeling, analysis, and simulation of wireless and mobile systems*. ACM, 2007, pp. 159–168.
- [18] W. Junfeng, W. Xiaoqing, and W. Ke, "Aero-sim: An ns-2 based simulator for aeronautical ad hoc networks," *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 9, no. 7, pp. 2548–2567, 2015.
- [19] A. Venkataramana, J. V. Rao, and S. P. Setty, "Impact of propagation models on qos issues of routing protocols in mobile ad hoc networks," in *Emerging ICT for Bridging the Future-Proceedings of the 49th Annual Convention of the Computer Society of India CSI Volume 2*. Springer, 2015, pp. 267–274.
- [20] C. Boutremans and J.-Y. Le Boudec, "Adaptive joint playout buffer and fec adjustment for internet telephony," in *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies*. IEEE Societies, vol. 1. IEEE, 2003, pp. 652–662.
- [21] C.-H. Lin, C.-H. Ke, C.-K. Shieh, and N. K. Chilamkurti, "An enhanced adaptive fec mechanism for video delivery over wireless networks," in *Networking and Services, 2006. ICNS'06. International conference on*. IEEE, 2006, pp. 106–106.
- [22] J. Racine, "The cygwin tools: a gnu toolkit for windows," *Journal of Applied Econometrics*, vol. 15, no. 3, pp. 331–341, 2000.
- [23] W. Dargie, X. Chao, and M. K. Denko, "Modelling the energy cost of a fully operational wireless sensor network," *Telecommunication Systems*, vol. 44, no. 1-2, pp. 3–15, 2010.
- [24] R. A. Shafik, M. S. Rahman, and A. Islam, "On the extended relationships among evm, ber and snr as performance metrics," in *Electrical and Computer Engineering, 2006. ICECE'06. International Conference on*. IEEE, 2006, pp. 408–411.
- [25] J. McDougall and S. Miller, "Sensitivity of wireless network simulations to a two-state markov model channel approximation," in *Global Telecommunications Conference, 2003. GLOBECOM'03. IEEE*, vol. 2. IEEE, 2003, pp. 697–701.
- [26] P. C. Ng and S. C. Liew, "Throughput analysis of ieee802.11 multi-hop ad hoc networks," *IEEE/ACM Transactions on Networking (TON)*, vol. 15, no. 2, pp. 309–322, 2007.
- [27] H. Zhai, Y. Fang et al., "Physical carrier sensing and spatial reuse in multirate and multihop wireless ad hoc networks," in *INFOCOM*, vol. 6, 2006, pp. 1–12.

Performance Analysis of a DS-CDMA Wireless Communication System with Multi Carrier(MC) Modulation in Presence of Carrier Frequency Offset (CFO)

Md.Al-Mamun Sabuz^{1*}, K.M.Niaz Reza¹, Md.Rashedul Islam¹, Md.Jubayed Hasan¹ and Satya Prasad Majumder²

¹Department of Electrical Electronic and Communication Engineering

¹Military Institute of Science and Technology, Dhaka-1216, Bangladesh

²Bangladesh University of Engineering and Technology, Dhaka-1205, Bangladesh

*sabuz.mamun@gmail.com

Abstract—Wireless communication with multicarrier modulation is very attractive for present and future communication network due to immunity to delay spread and fading. However, such system are sensitive to carrier frequency offset (CFO) due to Doppler effect in a mobile communication system. Analysis is carried out for a wireless communication system with multi carrier (MC) direct sequence (DS) code division multiple access (CDMA) with CFO to find the expression of signal to interference and noise ratio (SINR) and the bit error rate (BER) with BPSK coherent detection receiver. The results are evaluated in term of SINR and BER for various values of CFO variance, considering Gaussian distribution of CFO. The result shows that there occurs BER floor due to CFO for SC and MC DS-CDMA and there is power penalty at a given BER which is high for MC DS-CDMA compared to SC DS-CDMA. It is also observed that increasing the number of subcarriers does not affect much in MC DS-CDMA which will definitely allow to minimize the delay spread which is one of the vital limitations of wireless communication system

Keywords—Multicarrier, CFO, CFO variance, DS-CDMA, power penalty.

I. INTRODUCTION

Every communication system can be bounded by its own frequency range, application implementation cost, capacity, system [1]. Communication system can be classified into two types basing on the transmission system.

- Wired communication system
- Wireless communication system

Wireless communication means [2] transferring of information without the use of electrical conductors or "wires" over a distance. The distances may vary from few meters to thousands or millions of kilometers for radio communication. In wireless communication, electromagnetic waves or RF (instead of wire) carries the desired signal over a part or the entire communication path.

Since 1990s, the demand for new wireless capacity has been growing at a very rapid pace. But, the traditional resources used for adding capacity [3] to wireless systems are radio bandwidth and transmitter power. Unfortunately, these

two resources are mostly limited in the deployment of modern wireless networks. Radio bandwidth is limited due to the tight situation of useful radio spectrum and transmitter power [4] is limited due to the use of portable devices like mobile phone having limited battery power. These two resources are not improving at rates that may fulfill the anticipated demands for wireless capacity. Under these circumstances, there has been considerable efforts [5] in recent years aiming at the development of new wireless capacity [6] through greater intelligence in wireless networks.

A wireless system needs to accommodate as many users as possible by effective sharing of the limited bandwidth [7]. As a result, the term 'multiple access' plays an important role in the field of communications which can be defined as a mean of allowing multiple users simultaneously [8] to share the finite bandwidth with least possible degradation in the performance of the system. Keeping the limitations in mind, this paper focuses on the use of limited bandwidth [9] effectively with different techniques.

II. THEORETICAL BACKGROUND

The MC-DS-CDMA transmitter spreads the serial-to-parallel converted data streams [10] using a given spreading code in the time domain, so that the resulting spectrum of each sub-carrier can satisfy the orthogonality condition with the minimum required frequency separation [11]. In this scheme the initial data stream is serial-to-parallel converted to a number of lower-rate streams and each of these lower-rate streams is spread by the spreading code $C_k(t)$ with the same rate in the time domain, so that it is easier to synchronize the spreading sequences by a large chip time [12]. Furthermore, each of the lower-rate parallel streams is bit-interleaved and these streams modulate the orthogonal sub-carriers with a

successively overlapping bandwidth [11]. The receiver block diagram of a MC-DS-CDMA system is shown in Fig: 1.

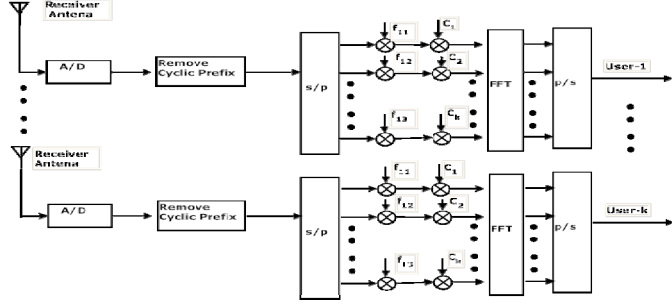


Fig.1. Block diagram of the receiver of a MC-DS- CDMA system.

The receiver contains a multiple correlator for each sub-carrier and each correlator is synchronized to a different resolvable path in the received composite signal. Then the serial data stream is recovered by a parallel-to-serial converter.

The MC-DS-CDMA has the following advantages. First, the spreading processing gain is increased compare to the corresponding single-carrier DS-CDMA scheme. Second, the effect of multipath interference is mitigated by implementing DS spreading. Finally, frequency/time diversity can be achieved [7].

In addition, each sub-carrier signal is subject to frequency selective fading in the MC-DS-CDMA scheme system. Consequently, a high complexity RAKE receiver can be implemented and forward error control (FEC) techniques can be utilized in order to enhance its associated performance [11].

System Block diagram:

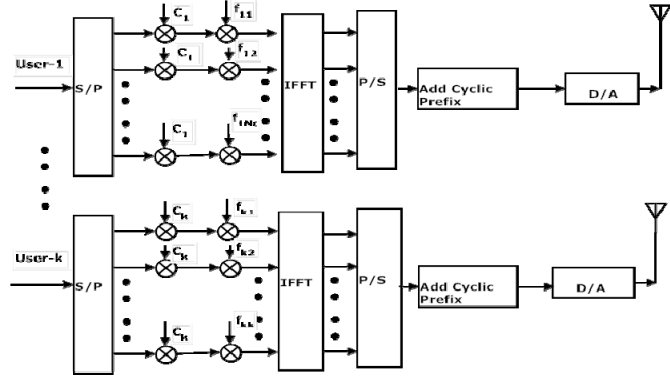


Fig.2(a).Transmitter Block diagram of a MC-DS-CDMA System.

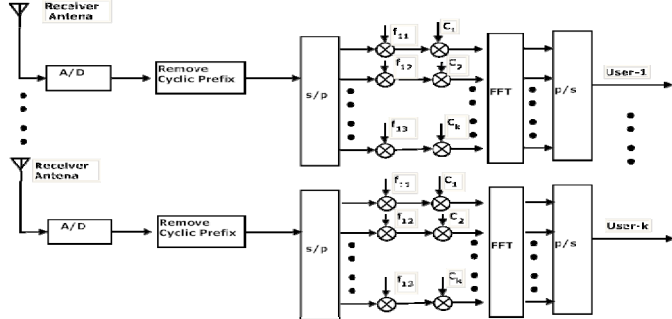


Fig. 2(b). Receiver block diagram of a MC-DS-CDMA System.

Analysis of SINR and BER of MC-DS-CDMA with CFO:

Figure-2(a) and 2(b) are the block diagram of a MC-DS-CDMA system.

The transmitted signal $S(t)$ can be expressed as,

$$S(t) = \sum_{j=1}^k x_j(t) c_j(t) \sum_{i=1}^M \cos[w_0 + (i-1)\Delta w]t;$$

Where

$x_j(t)$ = Message Signal;

$c_j(t)$ = Coded Signal;

K = No of User;

M = No of Subcarrier;

w_0 = Carrier Frequency (rad/sec);

Δw = Frequency offset(rad/sec);

We will first examine the OFDM portion of the system and then we gradually develop the necessary equation for effect of CFO on MC-DS-CDMA system. The OFDM system model over multipath slowly time varying fading channel with CFO can be described as [8]:

$$y_1(k) = e^{j2\pi \frac{k}{M} \sum_{l=0}^{L-1} h_l x_1(k-l) + n_1(k);}$$

$$0 \leq k \leq M-1 \quad (1)$$

Where,

h_l 's ($0 \leq l \leq L-1$) are i.i.d (independent and identically distributed) complex valued Rayleigh fading random variables and $n(k)$'s ($0 \leq k \leq M-1$) are independent complex-valued Gaussian random variables with zero mean and variance σ_n^2 for both real and imaginary components $x_1(0), \dots, x_1(M-1)$ are time domain transmitted symbols.

Where,

L =Length of the time domain channel impulse response (CIR)

It should be noted that this model can be employed in AWGN channel by setting:

$$h_0 = 1;$$

$$h_i = 0; \text{ for } i \geq 1;$$

ϵ = the normalized carrier frequency offset;

$\epsilon = \frac{\Delta f}{f_{ofdm}} \times M$; where f_{ofdm} = Total occupied bandwidth by OFDM

M = total subcarrier,

Now considering the CDMA and OFDM together i.e MC-DS-CDMA, multiplying each serial to parallel converted symbol with PN sequence code we have:

$$y_1(k) = e^{j2\pi \frac{k\epsilon}{M}} \sum_{l=0}^{L-1} h_1 x_1(k-l) C_1 + n_1(k);$$

$$0 \leq k \leq M-1 \quad (2)$$

Now if we add cyclic prefix in each data frame, with its length chosen to be longer than L, there will be no inter frame Interference (IFT) among OFDM frames.

After removing cyclic prefix and performing FFT at the receiver, we obtain the received data frame in the frequency domain [8]:

$$Y_1(m) = \frac{\sin \pi \epsilon}{M \sin \frac{\pi \epsilon}{M}} X_1(m) H(m) e^{j\pi \frac{(M-1)\epsilon}{M}} + ICI_1(m) + MAI(m) + N_1(m),$$

$$(3)$$

Where,

$H(m)$ = Frequency response of the channel at subcarrier m.

$N(m)$ = Frequency domain Gaussian noise variable at subcarrier

$$ICI_1(m) = \frac{1}{M} \sum_{k=0}^{M-1} \sum_{n \neq m} X_1(n) C_1(n) H(n) e^{j2\pi \frac{k(n-m+\epsilon)}{M}}; n \neq m,$$

$$ICI_1(m) \neq 0 \text{ for } \epsilon \neq 0;$$

For single user MAI=0;

Equation (3) shows that the frequency offset degrades the amplitude of the received signal in each subcarrier and introduce Inter carrier Interference in addition to the common phase shift to received signal.

Since CFO is not known at the receiver, the common phase $\frac{\pi(M-1)\epsilon}{M}$ in all subcarriers is unknown and causes degradation to BER.

From equation (3) we have

$$Y_1(m) = \frac{\sin \pi \epsilon}{M \sin \frac{\pi \epsilon}{M}} X_1(m) H(m) e^{j\pi \frac{(M-1)\epsilon}{M}} + ICI_1(m) + MAI(m) + N_1(m),$$

$$\text{Let, } I(\epsilon) = \frac{\sin \pi \epsilon}{M \sin \frac{\pi \epsilon}{M}} e^{j\pi \frac{(M-1)\epsilon}{M}}$$

hence equation (3) can be written as,

$$Y(m) = I(\epsilon) X(m) + ICI(m) + MAI(m) + N(m),$$

$$(4)$$

Where

$$ICI_1(m) = \frac{1}{M} \sum_{k=0}^{M-1} \sum_{n \neq m} X_1(n) C_1(n) H(n) e^{j2\pi \frac{k(n-m+\epsilon)}{M}}; n \neq m,$$

$$ICI_1(m) \neq 0 \text{ for } \epsilon \neq 0;$$

For single user MAI=0;

For multiple user MAI can be expressed as:

$$MAI = \frac{1}{T_b} \int_0^{T_b} \left[\sum_{j=2}^K x_j(t) \times c_j(t) \times c_1(t) \times \sigma_x \right] dt$$

$$= \sigma_x \sum_{j=2}^K x_j(t) \times c_{j1}(t)$$

$$(5)$$

$$\text{Where } C_{j1} = \frac{1}{T_b} \int_0^{T_b} [c_j(t) \times c_1(t)] dt$$

Cross correlation between user code-1 & jth user code.

Desired signal power

$$E[|I(\epsilon) X(m)|^2] = |I(\epsilon)|^2 \sigma_x^2$$

$$= \frac{\sin^2 \pi \epsilon}{M^2 \sin^2 \frac{\pi \epsilon}{M}} \sigma_x^2$$

$$(6)$$

It is clear that signal power is attenuated since

$$|I(\epsilon)|^2 = \frac{\sin^2 \pi \epsilon}{M^2 \sin^2 \frac{\pi \epsilon}{M}} \sigma_x^2 < 1$$

Now,

$$E[|ICI(m)|^2] = \sigma_x^2 (1 - E[|I(\epsilon)|^2])$$

$$= \sigma_x^2 \left(1 - \frac{\sin^2 \pi \epsilon}{M^2 \sin^2 \frac{\pi \epsilon}{M}}\right)$$

$$= \frac{\sin^2 \pi \epsilon}{M^2 \sin^2 \frac{\pi \epsilon}{M}} \sigma_x^2$$

$$(7)$$

Now considering multipath fading and multiuser, the Gaussian approximation the analytical SINR with BPSK can be expressed as [8][6]:

$$SINR(\epsilon) = \frac{\frac{\sigma_x^2 \sin^2 \pi \epsilon}{M^2 \sin^2 \frac{\pi \epsilon}{M}} \cdot \cos^2 \frac{\pi \epsilon (M-1)}{M}}{[\sigma_x^2 + MS\{MAI\}](1 - \frac{\sin^2 \pi \epsilon}{M^2 \sin^2 \frac{\pi \epsilon}{M}}) + \sigma_n^2}$$

$$(8)$$

Now For multiple user (no of user= K), MS {MAI}

$$\approx \sigma_x^2 \frac{K-1}{L} \otimes C_{rx}$$

;C_{rx}= Cross correlation factor

Hence we can write,

$$SINR(\varepsilon) = \frac{\frac{\sigma_x^2 \sin^2 \pi \varepsilon}{M^2 \sin^2 \frac{\pi \varepsilon}{M}} \cdot \cos^2 \frac{\pi \varepsilon (M-1)}{M}}{[\sigma_x^2 + \sigma_n^2 \frac{K-1}{L} \otimes C_{rx}](1 - \frac{\sin^2 \pi \varepsilon}{M^2 \sin^2 \frac{\pi \varepsilon}{M}}) + \sigma_n^2}$$

$$SINR(\varepsilon) = \frac{\frac{\sigma_x^2 \sin^2 \pi \varepsilon}{M^2 \sin^2 \frac{\pi \varepsilon}{M}} \cdot \cos^2 \frac{\pi \varepsilon (M-1)}{M}}{\sigma_n^2 (1 + \frac{K-1}{L} \otimes C_{rx})(1 - \frac{\sin^2 \pi \varepsilon}{M^2 \sin^2 \frac{\pi \varepsilon}{M}}) + \sigma_n^2}$$

$$SINR(\varepsilon) = \frac{\frac{\sigma_x^2 \sin^2 \pi \varepsilon}{M^2 \sin^2 \frac{\pi \varepsilon}{M}} \cdot \cos^2 \frac{\pi \varepsilon (M-1)}{M}}{\sigma_n^2 (1 + \frac{K-1}{L} \otimes C_{rx})(1 - \frac{\sin^2 \pi \varepsilon}{M^2 \sin^2 \frac{\pi \varepsilon}{M}}) + \sigma_n^2}$$

$$SINR(\varepsilon) = \frac{\frac{\sigma_x^2}{\sigma_n^2} \cdot \frac{\sin^2 \pi \varepsilon}{M^2 \sin^2 \frac{\pi \varepsilon}{M}} \cdot \cos^2 \frac{\pi \varepsilon (M-1)}{M}}{1 + \frac{\sigma_x^2}{\sigma_n^2} (1 + \frac{K-1}{L} \otimes C_{rx})(1 - \frac{\sin^2 \pi \varepsilon}{M^2 \sin^2 \frac{\pi \varepsilon}{M}})}$$

$$(9)$$

Finally Bit Error Rate (BER) of MC-DS-CDMA can be expressed as

$$BER(\varepsilon) = Q(\sqrt{SINR(\varepsilon)})$$

$$BER(\varepsilon) = Q\left(\sqrt{\frac{\frac{\sigma_x^2}{\sigma_n^2} \cdot \frac{\sin^2 \pi \varepsilon}{M^2 \sin^2 \frac{\pi \varepsilon}{M}} \cdot \cos^2 \frac{\pi \varepsilon (M-1)}{M}}{1 + \frac{\sigma_x^2}{\sigma_n^2} (1 + \frac{K-1}{L} \otimes C_{rx})(1 - \frac{\sin^2 \pi \varepsilon}{M^2 \sin^2 \frac{\pi \varepsilon}{M}})}}\right)$$

$$BER(\varepsilon) = Q\left(\sqrt{\frac{SNR_0 \cdot \frac{\sin^2 \pi \varepsilon}{M^2 \sin^2 \frac{\pi \varepsilon}{M}} \cdot \cos^2 \frac{\pi \varepsilon (M-1)}{M}}{1 + SNR_0 (1 + \frac{K-1}{L} \otimes C_{rx})(1 - \frac{\sin^2 \pi \varepsilon}{M^2 \sin^2 \frac{\pi \varepsilon}{M}})}}\right)$$

$$(10)$$

$$SNR_0 = \frac{\sigma_x^2}{\sigma_n^2} = \frac{P_{sig}}{P_{noise}}$$

Where,

III. RESULT AND DISCUSSION

The plots of BER as a function of SNR are shown in fig 3(a) and fig 3(b) for different values of the variance of the carrier frequency offset $\sigma^2 = 0.001, 0.002, 0.003, 0.004, \text{ and } 0.005$. It is noticed that the BER decreases with SNR at a given value of CFO and results in BER floor which is due to

CFO. The system thus suffers from CFO in term of power penalty (PP), which is depicted in fig 4 (a), 4 (b) and Table-1.

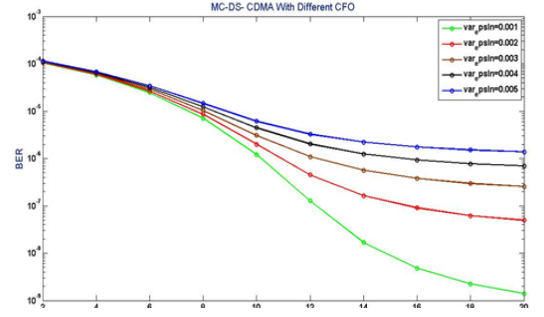


Fig. 3(a). Plot of BER versus SNR (dB) for MC-DS-CDMA system with having variance $\sigma^2 = 0.001, 0.002, 0.003, 0.004, \text{ and } 0.005$.

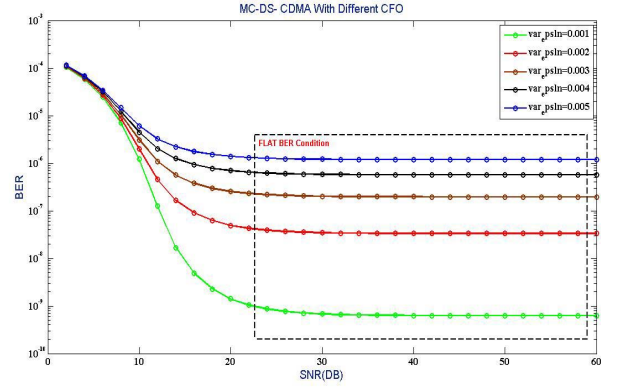


Fig. 3(b). Plot of BER versus SNR (dB) for MC-DS-CDMA system with CFO having variance 0.001, 0.002, 0.003, 0.004, and 0.005 showing BER Floor.

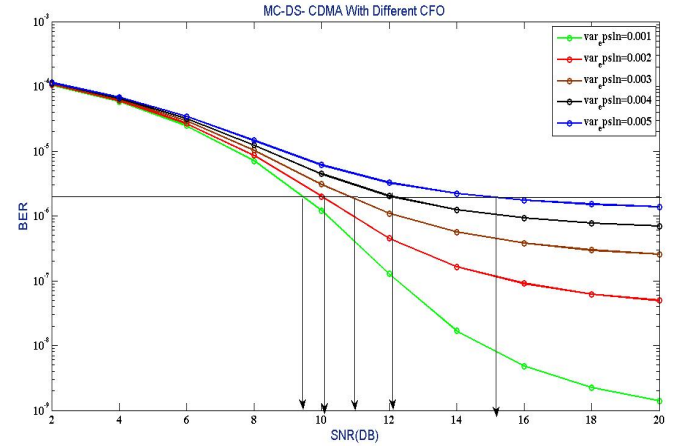


Fig.4(a). Plot of BER versus SNR (dB) for MC-DS-CDMA system with CFO having variance $\sigma^2 = 0.001, 0.002, 0.003, 0.004, \text{ and } 0.005$ for power penalty (PP) calculation.

TABLE-1: POWER PENALTY FOR DIFFERENT CFO (MC-DS-CDMA)

Considering,
Bit Error Rate (BER) = 10^{-7} ;
User = 255;
Code Length = 16;

IV. CONCLUSION

In this paper, we have developed an analytical approach to find out the effect of CFO on the SINR and BER of a SC-DS-CDMA and MC-DS-CDMA system in order to compare the performance of those two systems under BPSK modulation scheme and AWGN channel. It was observed that both the system suffers significantly due to the presence of CFO which results BER floor. However a MC-DS-CDMA system suffers comparatively more than SC-DS-CDMA. For example, when CFO variance $\sigma^2 = 0.003$, and SNR=20 (dB), BER is about 10^{-8} for a SC-DS-CDMA system, whereas for MC-DS-CDMA, with same parameter, BER is $10^{-6.5}$. It is also observed that for a given BER, user and code length, a MC-DS-CDMA system having variable CFO variance has more power penalty than that of a SC-DS-CDMA system with same parameter. For example, when CFO variance is 0.003 a SC-DS-CDMA system has a power penalty about 1 dB, whereas a MC-DS-CDMA system has 1.8 dB. For a given BER, CFO variance and code length, a MC-DS-CDMA system having user as variable has more power penalty than that of a SC-DS-CDMA system with same parameter. For example, for 300 users a SC-DS-CDMA system has a power penalty about 1.4 dB, whereas a MC-DS-CDMA system has 1.5 dB. The effect of code length on the system. When code length is increased both the system performance improved. It was also observed that for a MC-DS-CDMA system increasing the no of subcarriers does not affect much on BER performance. Hence we can minimize delay spread using more no of subcarriers which will definitely help to minimize on of the major drawbacks of wireless communication system.

V. FUTURE WORK

Further research can be carried out on DS-CDMA system considering the effect of carrier frequency offset (CFO) as follows: Work can be carried out to analyze BER performance in a Multiple input multiple output (MIMO) Orthogonal Frequency Division Multiplexing (OFDM) system considering Space Time Block Code (STBC) and Space Frequency Block Code (SFBC). Work can be carried out to find the BER performance considering a fading environment such as Nakagami and Rician fading.

Different Coding can be tested such as Convolutional Code, Turbo Code etc to evaluate the performance of BER.

REFERENCES

- [1] Ling, A.; Milstein, L.B., "The effects of spatial diversity and imperfect channel estimation on wideband MC-DS-CDMA and MC-CDMA," Communications, IEEE Transactions on , vol.57, no.10, pp.2988,3000, October 2009
- [2] Mukesh Kumar Mishra, Neetu Sood and Ajay K Sharma, "BER Performance of OFDM-BPSK over Nakagami Fading Channels", International Journal of Computer Applications (0975 – 8887) Volume 18– No.1, March 2011.
- [3] S.Mallick and S.P.Majumder "Performance analysis of an OFDM system in the presence of Carrier Frequency Offset, phase noise and timing jitter over rayleigh fading channels" in Proc.ICECE'08,pp-205-2010,Dhaka,Dec.2008.
- [4] K.A. Taher, M.R. Minhaz, S.P. Majumder "Performance analysis of multi-carrier direct-sequence (MC DS) CDMA with fading" in Proc.ICECE'08,pp-810-813,PyeongChang,Jan.2013.
- [5] R.U. Mahesh, A.K. Chaturvedi "Closed form BER expressions for BPSK OFDM systems with fractional timing offset and carrier frequency offset" IEEE Communications Letters (Impact Factor: 1.16). 09/2010; DOI: 10.1109/LCOMM.2010.08.100088.
- [6] K. Sathananthan, C. Tellambura "Probability of error calculation of OFDM systems with frequency offset" IEEE Transactions on Communications,2/2001; DOI: 10.1109/26.966051.
- [7] EA Sourour, M Nakagawa "Performance of orthogonal multicarrier CDMA in a multipath fading channel" *IEEE Trans. Commun.*, vol 44, no. 3, pp.356-367, March 1996.
- [8] Xiaoqiang Ma, Hisashi Kobayashi and Stuart C. Schwartz, "Effect of Frequency Offset on BER of OFDM and Single Carrier System", 2003 the 14th IEEE International Symposium on Personal, Indoor and Mobile Radio Communication.
- [9] Pravin W. Raut, Dr. S.L. Badjate, "MIMO Technology for Wireless Communication and Wireless Computer Networking", International Journal of Innovative Technology and Research, Volume No. 1, Issue No. 2, February - March 2013, 194 – 200.
- [10] Pravin W. Raut, S.L. Badjate, "MIMO-Future Wireless Communication", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-2, Issue-5, April 2013.
- [11] L.Hanzo,L-L,Yang,E-L.Kuan, and K. Yen, "Single- and multi-carrier CDMA: multi-user detection, space-time spreading ,synchronization, networking, and standards" IEEE press, IEEE communications society, 2003.
- [12] S. Hara, and R. Prasad. "Overview of Multicarrier CDMA", IEEE Communication Magazine, Volume 35, pp. 126-133, December 1997.

Pre-Detection PMD Compensation Based on Feedback of the First and Second RF Harmonic Tones in an Optical Fiber Transmission System

Kazi Abu Taher^{1,*} and Satya Prasad Majumder¹

¹Department of EEE, Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh

*kataher@yahoo.com

Abstract—We propose and demonstrate a noble technique for compensation of polarization mode dispersion (PMD) induced impairments in polarization division multiplex (PDM) quadrature phase shift keying (QPSK) systems. The technique uses first and second radio frequency (RF) harmonic tone powers to generate the control signal for the compensation. Optical notch filter is used to filter out one of the sidebands of the optical spectrum to remove the effects of chromatic dispersion (CD) from the detected RF signal. Then power levels of the RF harmonics vary with the PMD induced differential group delay (DGD) alone. The compensation technique takes control signal from the electrical domain by comparing the slopes of the power of the RF harmonics. The technique is found to be very effective covering the window of a symbol period. After compensation, the maximum electrical power improvements are 0.76 dB and 0.85 dB for data rate of 40 and 160 Gbps respectively.

Index Terms—polarization mode dispersion, compensation, differential group delay, polarization-division multiplex.

I. INTRODUCTION

PMD arises in a long single mode fiber (SMF) transmission system due to random anisotropy in caused by the non-uniform cores and variations in physical parameters such as the stresses, bending and cabling [1]. The very small amount of birefringence generated due to the variations in environmental and manufacturing parameters randomly varies along the length of the fiber [2,3]. This random and stochastic birefringence causes the optical pulses to divide between the two orthogonal polarization modes. As the modes experience different refractive indices, they travel at slightly different speeds and reach the receiving end at different times. It gives rise of a DGD that results in pulse spreading and inter symbol crosstalk [4,5]. The dominant effects of PMD come from its first order which is frequency independent. The frequency dependent higher order PMD effects are not much significant comparing to the first order PMD induced impairments [6,7].

When the Kerr nonlinearity induced effects of the fiber couples with other transmission impairments such as PMD, it can lead to complex dynamics. The adverse effects of PMD worsen with the increase of bit rate and distance [8,9]. Moreover, the instantaneous value of the DGD follows a Maxwell probability distribution, having possibility of network outage at any time [10,11]. Stochastic processes of PMD poses the main difficulty in controlling and compensating its effects in long SMF transmission links. Thus a system can randomly wander in and out of high-penalty states with Maxwellian distribution and need measures to reduce the probability that the penalty will exceed a certain unacceptable level [12]. This leads to the need for dynamic monitoring and compensation techniques.

A number of monitoring and compensation techniques for PMD in optical and electrical domain have been studied. Optical domain techniques include: 1) calculating the degree of polarization (DOP) of polarized lights [12,13]; 2) measuring the phase difference between two frequency components of orthogonal principal state of polarizations (PSPs) [14] and 3) calculating the arrival time of polarization-scrambled light [15]. The electrical domain techniques include: 1) analysing the spectral frequency components of received signals [5,16]; 2) measuring the eye-opening penalty [17-19] and 3) adding a subcarrier tone to the transmitting data [8]. But all of these techniques suffers from one or more of the following disadvantages [4]: (1) limited monitoring sensitivity, (2) dependence on pulse width, (3) single-channel operation and (4) the necessity for major system modification. NRZ signals suffer greater degradation due to PMD than RZ signals [17,20]. Thus this report focuses on the PMD compensation techniques for NRZ PDM-QPSK signals.

This paper is organized as follows; system model is described in Section II, Section III covers the spectral analysis for PMD compensation, Section IV covers results and discussion and finally conclusion is given in Section V.

II. SYSTEM MODEL

The system model for the compensation of PMD induced DGD in a single channel optical PDM-QPSK transmission link is shown in Fig. 1. OptSim simulator is used for the simulation of the model. Matlab is used to process the offline data. For 40 Gbps PDM-QPSK system, the data rate of each channel is 10 Gbps. Two IQ modulators are used to modulate the input data. Thus the symbol period is 100 ps and the simulation is carried out for $0 \leq \text{DGD} \leq 100$ ps. We generated 10 Gbps data using PRBS generators for the four inputs as shown in Fig.1. The data were modulated onto a 193.14 THz continuous wave (CW) light. The transmitter laser power was set at -5 dBm to keep the nonlinear effects to minimum. One of the two IQ modulators is rotated by 90 degree before combining the polarization modes to ensure the orthogonality of the two polarization modes.

We implemented the dispersion compensated fiber (DCF) by post compensation method. The link length was set to 100 km having two segments each of 50 km. Dispersions for the first and second segments were set at 16 ps-nm/km and -16 ps-nm/km respectively. The loss of the fiber is neglected setting the EDFA gain to a value to compensate the signal attenuation. The fiber length can be increased by setting the iteration value to required number. The PSP of the signal $E_i(t)$ is so adjusted so that it matches to that of the fiber to ensure equal power among both the polarizations.

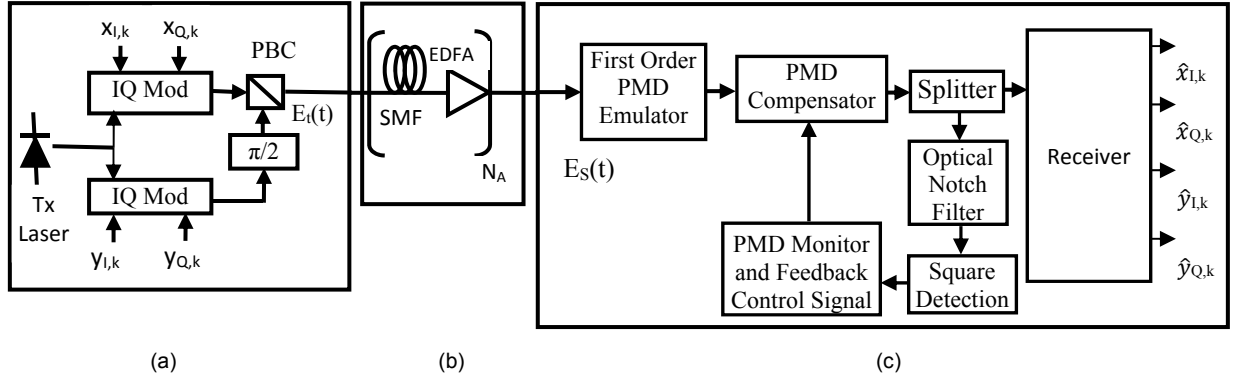


Fig. 1: Diagram of an optical PDM-QPSK transmission system (a) PDM-QPSK transmitter (b) Optical link and (c) PDM-QPSK receiver with PMD compensator

At the receiver, a first order PMD emulator introduces required amount of DGD to the received signal $E_s(t)$. The PMD compensator has two inputs. One input is the received signal with PMD induced DGD and another input is the feedback control signal. Basing on the control signal, the PMD compensator introduces corrective delay to the fast polarization mode. Then a splitter splits the signal to tap out small portion to the optical notch filter and the main portion is sent to the receiver. The notch filter removes the lower side band (LSB) of the optical spectrum and passes the rest of the signal for square detection. Raised cosine notch filter of order 1 with roll-off factor of 0.5 of 20 GHz bandwidth is set at 10 GHz below the carrier frequency. The notch filter is used here because it can remove one of the sideband terms efficiently while leaving the rest of the spectrum and the carrier unaffected. Detected electrical signal is analysed to monitor the amount of DGD and a control signal is generated basing on the analysis. The control signal is passed to the PMD compensator for required compensation.

III. SPECTRAL ANALYSIS FOR PMD COMPENSATION

In PDM-QPSK transmission system, the optical spectrum has the same property as that of QPSK signals. When double sideband (DSB) optical NRZ-QPSK data is detected, the RF

power is given by [4]:

$$P_{DSB} = P_0 [1 - 4\gamma(1 - \gamma)\sin^2(\pi f_{RF}\Delta\tau)] \times \cos^2\left(\frac{\pi c D f_{RF}^2}{f_c^2}\right) \quad (1)$$

where P_0 is RF power without considering the effects of CD and PMD, γ is the power splitting ratio between the orthogonal modes, f_{RF} is the data rate, $\Delta\tau$ is the average DGD of the link, f_c is the optical carrier frequency, D is the collective residual dispersion of fiber link and c is the speed of light. Here $\Delta\tau$ and γ are the PMD related terms. We neglected the chirp parameter because the amplitude of QPSK modulation may be considered constant. To separate the effects of CD and PMD, we exploit the behavior of the optical and electrical frequency spectrum. Since the effect of CD arises from the beating of the two side bands, the effect of CD may be removed from equation (1) by removing one of the side bands. If we filter out the lower side band (LSB), the detected RF spectrum can be given as [4]:

$$P_{SSB} = 0.5P_0 [1 - 4\gamma(1 - \gamma)\sin^2(\pi f_{RF}\Delta\tau)] \times |H(f_{RF})|^2 \quad (2)$$

where H is electrical field transfer function of the filter. The factor 0.5 is introduced due to the removal of one sideband.

Fig. 2 shows the electrical spectrum after detecting the optical signal with both sidebands. During the square

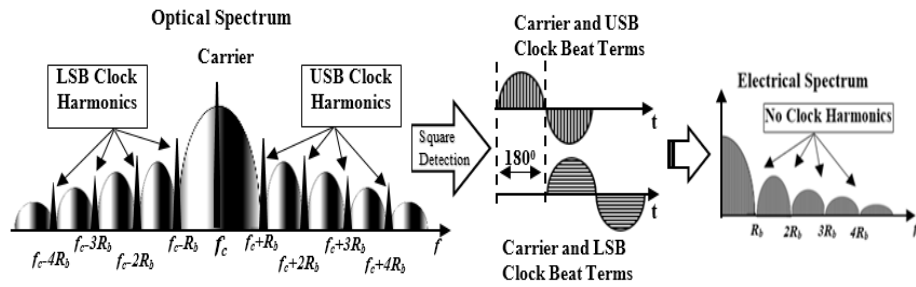


Fig. 2. Optical and electrical spectrum with carrier, USB and LSB terms

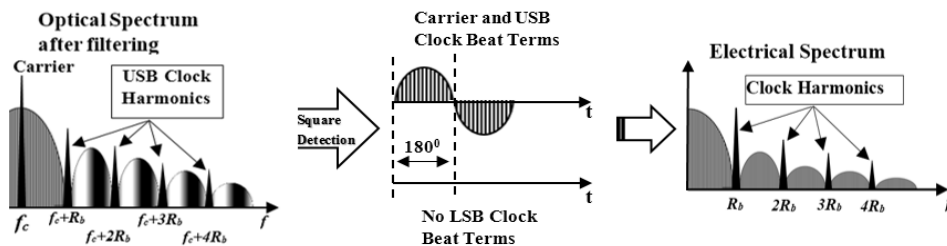


Fig. 3. Optical spectrum with one sideband and corresponding electrical spectrum

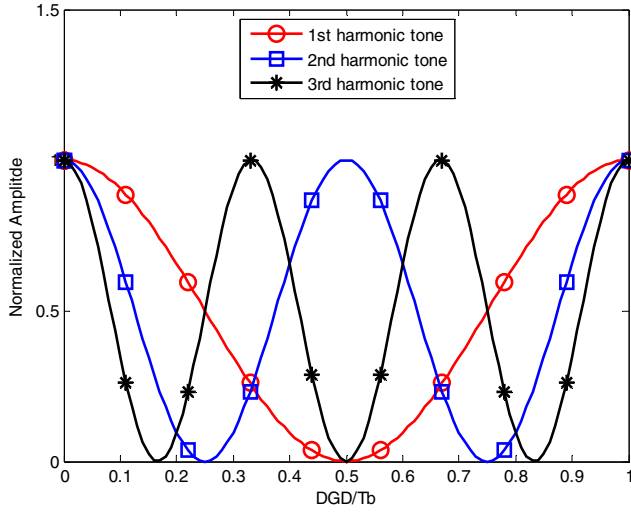


Fig. 4. Normalized RF power of 1st, 2nd and 3rd harmonics using Eq. (2)

detection, LSB and USB terms are 180 degree out of phase. The carrier and the USB terms beat with LSB terms resulting in minimum or no RF power at tone harmonics as shown in the electrical spectrum of Fig. 2. Fig. 3 shows the spectrum of the detected electrical signal after passing the optical signal with LSB removed. It is to be noted that Equation (2) now have only PMD related parameters γ and $\Delta\tau$. The RF power of SSB signal changes periodically and the period is related to the data rate. It is noted that we can get the RF power at tone harmonics by setting $f_{RF} = n f_{RF}$ in Equation (2) where $n = 1, 2$ and 3 for 1st, 2nd and 3rd harmonics respectively. Fig. 4 shows the normalized RF power at first three harmonics where DGD is varied from 0 to T_b . Actual RF power at higher order harmonics will be very less. As seen from Fig. 4 that the slope of RF power at different harmonics for the same value of DGD are different. By comparing the slope of RF power levels of different harmonics help to generate the control signal for PMD compensator. It is found that the more the number of harmonics, the more accurate will be the control signal.

The PSP of the signal are adjusted so that both the polarization modes have equal power. When the modulated light enters the fiber, the polarization components have power depending on power splitting factor. We set the power splitting factor as 0.5 because it causes the worst power fading condition of the RF harmonic tones. We used the slope of the power pattern of the regenerated RF clock tone harmonics to generate the control signal. The slope of the power of each harmonic is different at a particular DGD. The more number of harmonics are used to compare the power slope, the more accurate precise value of the control signal is generated. Again the number of harmonics is limited by ultralow power level. Higher the harmonic order, lower is the power level. Minimum two harmonics must be used to generate the control signal. We used RF power of first and second harmonics to generate the control signal.

The trend and pattern of previous four samples are taken as the reference to generate the control signal. Care is also taken so that local maxima or minima does not provide inaccurate control signal. The control signal determines how much delay needs to be introduced to the fast mode through the delay element. In fact the monitoring signal is processed and converted to control signal. The control signal is passed to the compensator where the required delay is introduced to the

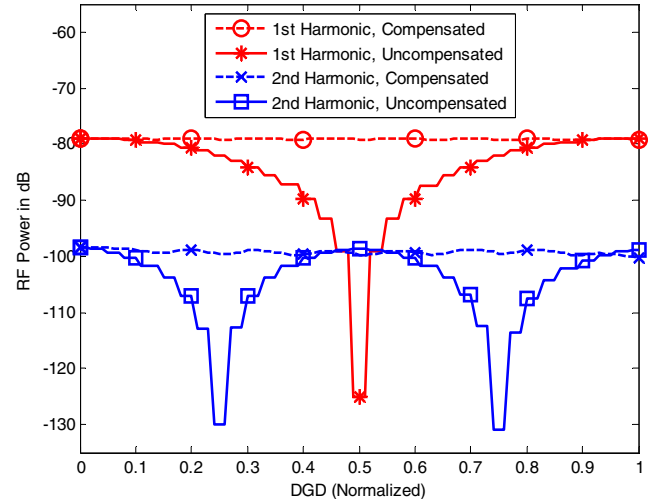


Fig. 5. Compensated and uncompensated RF power at 1st and 2nd harmonics for data rate of 40 Gbps.

fast mode. This compensator also works as a polarizer and improves the received electrical power level because the output of the polarizer is linearly polarized.

IV. RESULTS AND DISCUSSION

For the emulation of the first order PMD compensation, variable amount of DGD is introduced between the modes as shown in Fig 1. We simulated 512 independent bits with sampling rate of 32 samples/bit. Link length is set at 100 km with equal segments for implementing post dispersion compensation fiber (DCF). Fig. 5 shows the graph of RF power of the 1st and 2nd harmonics as a function of normalized DGD. Uncompensated RF tones of 1st and 2nd harmonics have the similar frequency similar to the waveform shown in Fig. 4. For $0 \leq \text{DGD} \leq 0.5T_b$, the 1st harmonic power slope is negative but the 2nd harmonics power slope goes negative and then positive having minimum value at $\text{DGD} = 0.25T_b$. The power swings of the RF power are 46.12 and 32.36 dB for the 1st and 2nd harmonics respectively. The slope changes of the harmonics help generating the control signal using the appropriate algorithm. The graph also shows the RF harmonic power level after compensation.

Fig. 6 show the electrical power level as a function of normalized DGD for compensated and uncompensated optical signals at data rates of 40 and 160 Gbps. After compensation,

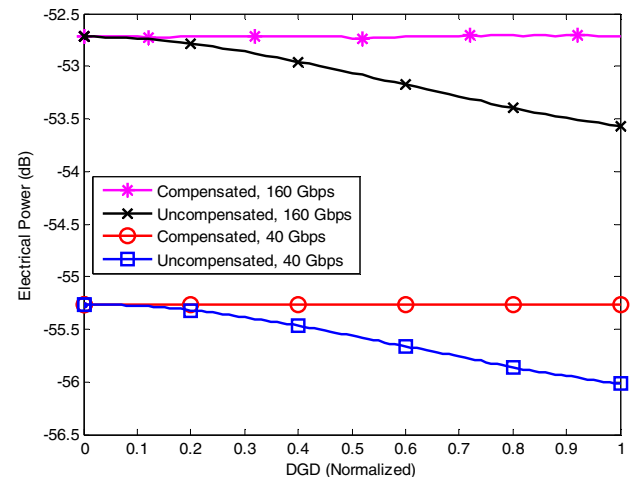


Fig. 6. Compensated and uncompensated electrical power for data rate of 40 and 160 Gbps

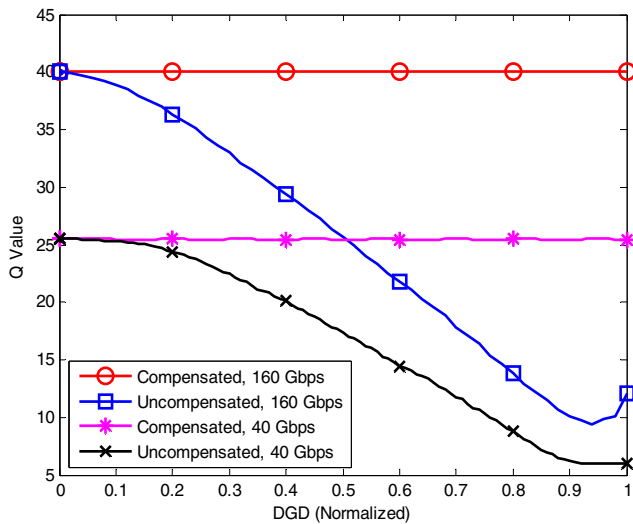


Fig.7. Compensated and uncompensated Q-value for data rate of 40 and 160 Gbps

there is no change in the electrical power level. But there is a power loss for uncompensated signal. This loss increases with the increase of DGD. The maximum power losses are 0.76 dB and 0.85 dB at $DGD = T_b$ for data rate of 40 and 160 Gbps respectively comparing with the compensated power level. The average power loss of the RF tone is 0.32 dB and 0.37 dB for 40 and 160 Gbps respectively for $0 \leq DGD \leq 0.5T_b$.

Fig. 7 shows the Q-value improvement of the compensated and uncompensated signals at 40 and 160 Gbps. For uncompensated case, the Q-value decreases with the increase of DGD. The Q-value goes below the acceptable level at higher DGD. But after PMD compensation, the Q value improves to the desired level for reliable data transmission.

V. CONCLUSIONS

The proposed technique of PMD compensation of PDM-QPSK transmission system by regenerated first and second RF harmonic tones shows satisfactory performance. The monitoring window for generating the control signal increases to cover the whole of the bit period. The electrical power level improves in an average value of 0.32 dB and 0.37 dB for data rates of 40 and 160 Gbps. The Q-value is also improves as a result of the compensation. The data may be detected by direct detection, self-coherent or coherent method. The proposed technique has the better ability in compensating the first order PMD at higher data rates. There should be further study to find out its performance at even higher data rates and in multi-channel systems.

ACKNOWLEDGMENT

The work is carried out as a part of PhD dissertation in the Department of Electrical and Electronic Engineering of Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh.

REFERENCES

- L. L. Minkov, I. B. Djordjevic, L. Xu, and T. Wang, "PMD Compensation in Polarization-Multiplexed Multilevel Modulations by Turbo Equalization", *IEEE Photonics Technol. Lett.*, vol. 21(23), pp. 1773-5, Dec 2009.
- T. Ilavarasan and M. Meenakshi, "An overview of fiber dispersion and nonlinearity compensation techniques in optical orthogonal frequency division multiplexing systems", *J. of Optics*, pp. 255-270, Sep 2015.
- K. A. Taher, S. P. Majumder and B. M. A. Rahman, "Performance of multilevel modulation formats in 92 Gb/s systems in the presence of PMD and nonlinear effects", in *proc. of IEEE Conference on 1st Int. Conf. on Electrical Engineering and Information & Commun Techno (ICEEICT)*, pp. 1-4, May 2014.
- J. Yang, C. Yu, L. Cheng, Z. Li, C. Lu, A. P. T. Lau, H. Tam and P. K. A. Wai, "CD-insensitive PMD monitoring based on RF power measurement", *Opt. Express*, **19**(2), pp. 1354-1359, 2011.
- C. Yu and J. Yang, "CD and PMD monitoring based on RF spectrum analysis with optical filtering," in *proc. of SPIE-OSA-IEEE Asia Communications and Photonics*, SPIE vol. 798604, pp. 1-8, 2011.
- M. W. Chbat, J. P. Soigne, T. Fuerst, J. T. Anthony, S. Lanne, H. Fevrier, A. H. Bush and D. Penninckx, "Long term field demonstration of optical PMD compensation on an installed OC-192 link," in *proc. of Opt. Fiber Comm. Conf. and the Int. Conf. on Integrated Opt. and Opt. Fiber Comm.(OFC/IOOC)*, PD12/1-PD12/3, 1999.
- S. B. Yanyushev, I. H. Galeev, I. L. Vinogradova, A. N. Garifinov, A. I. Salikhov, and A. H. Sultanov, "Chromatic dispersion and PMD measurements of optical fiber", in *proc. of SPIE vol. 7992 79920G*, pp. 1-8, 2011.
- S. M. Nezam, Y. Song, C. Yu, J. McGeehan, A. Sahin and A. E. Willner, "First-order PMD monitoring for NRZ data using RF clock regeneration techniques," *J. of Light. Techno.*, **22**(4), pp. 1086-1093, 2004.
- A. Bononi, A. Vannucci, A. Orlandini, E. Corbel, S. Lanne, and S. Bigo, "Degree of polarization degradation due to cross-phase modulation and its impact on polarization-mode dispersion compensators," *J. of Light. Techno.*, **21**(9), pp. 1903-1913, 2003.
- F. Roy, C. Francia, F. Bruyere, and D. Penninckx, "Simple dynamic polarization mode dispersion compensator," in *proc. of Conf. Opt. Fiber Comm. and Int. Conf. Integrated Opt. (OFC/IOOC)*, **1**, pp. 275-278, 1999.
- J. Zhao, C. Lu and X. R. Ma, "Simultaneously monitoring of chromatic dispersion and polarization mode dispersion for DPSK signal based on RF spectrum analysis," *J. of Applied Mechanics and Materials*, **130-134**, pp. 3851-3854, 2012.
- H. Rosenfeldt, C. Knothe, R. Ulrich, E. Brinkmeyer, U. Feiste, C. Schubert, J. Berger, R. Ludwig, H. G. Weber and A. Ehrhardt, "Automatic PMD compensation at 40 Gbit/s and 80 Gbit/s using a 3-dimensional DOP evaluation for feedback", in *proc. of Opt. Fiber Comm. Conf. and Exhibit (OFC)*, vol. **4**, pp. PD27-1-PD27-3, 2001.
- S. M. R. M. Nezam, L. S. Yan, Y. Q. Shi, A. E. Willner and S. Yao, "Wide-dynamic-range DGD monitoring by partial optical signal spectrum DOP measurement," in *proc. of Opt. Fiber Comm. Conf.* pp. 924-926, 2002.
- B. W. Hakki, "Polarization mode dispersion compensation by phase diversity detection," in *IEEE Photon. Techno. Lett.*, **9**(1), pp. 121-123, 1997.
- H. Y. Pua, K. Peddanarappagari, B. Zhu, C. Allen, K. Demarest, and R. Hui, "An adaptive first-order polarization-mode dispersion compensation system aided by polarization scrambling: theory and demonstration," *J. of Light. Techno.*, **18**(6), pp. 832-841, Jun 2000.
- K. A. Taher, S. P. Majumder, and B. M. A. Rahman, "PMD monitoring of polarization division multiplex transmission system using RF tone harmonics", in *proc. of the 18th International Conference on Computer and Information Technology (ICIT)*, pp. 527-530, Dec 2015.
- K. A. Taher, S. P. Majumder, B. M. A. Rahman, Y. Yu, and C. Yu, "Simultaneous monitoring of CD and PMD using RF tone power," *Elsevier Procedia Engineering*, pp. 209-216, Jul 2016.
- H. Sun and K. Jhang, "Adaptive PMD compensation using FBMC in optical transmission system", in *proc. of International Conference on Cyberspace Technol.*, 14 May 2015
- A. Djupsjöbacka, A. Berntson, and J. Mårtensson, "A method to calculate PMD-induced eye-opening penalty and signal outage for RZ-modulated signal formats," *J. of Light. Techno.*, **26**(17), pp. 3186-9, Sep 2008.
- X. Tang, X. Zhang, L. Xi, Q. Xiong, X. Li, "PMD mitigation using combination of PMD compensator and a phase regenerator in optical 40 Gb/s DPSK system" in *proc. of Conf. on Lasers and Electro-Optics*, pp 1-2, Sep 2009.

On the Selection of Passive Elements for Low Phase Noise LC Tank VCO in 65 nm Process

Md. Tawfiq Amin, *Member, IEEE*

Department of EECE,
Military Institute of Science and Technology (MIST),
Dhaka-1216, Bangladesh

Abstract—This paper describes the LC tank design for low phase noise LC VCO on the selection of passive elements. LC tank based VCO is commonly used for the low phase noise performance in most RF applications. Since LC tank quality factor is mainly dominated by the quality factor of the passive elements (inductor & capacitor) selection, this work presents the passive elements optimization based on the inductor and varactor optimization. Simulated in 65 nm process, higher Q is obtained for large width and minimum number of turns for the inductor and for the varactor design; the PMOS varactor operating in the depletion and accumulation regions allows large tuning range and less parasitic resistance.

Keywords— Voltage-controlled oscillator (VCO), phase noise, quality factor, LC tank, integrated spiral inductor, varactor.

I. INTRODUCTION

The rapid growth of commercial wireless communication market has brought an increasing demand for high performance, low cost, low power consumption radio frequency integrated circuits (RFICs). A critical building block in almost all wireless communication systems is the local oscillator (LO), where it is used with a mixer to allow frequency translation and channel selection of RF signals. LC VCO topologies are commonly used in RF circuits because of their good phase noise characteristics and their ease of implementation [1] - [4]. The most critical specification for any VCO is its spectral purity, usually characterized by its phase noise. Phase noise in oscillators is a non-ideal behavior that appears as noise in the side bands of the carrier frequency. This is undesirable because it distorts not only the carrier frequency but also the adjacent channels and if the oscillator is used for frequency conversion purposes, the phase noise causes reciprocal mixing. Phase noise analysis for LC VCO shows that, total phase noise in LC VCO design is contributed by three main sources namely; from the tank resistance, the differential pair MOSFETs and from the current source. This paper presents how the phase noise contribution from the tank resistance can be optimized by proper selection of passive elements in LC VCO design.

The output amplitude of the VCO depends on the parallel resistance of the LC tank,

$$R_P = Q \cdot 2\pi f_{osc} \cdot L \quad (1)$$

R_P is the combination of oscillation frequency, f_{osc} , quality factor, Q and the size of the inductor, L . The phase noise at an offset frequency f_{offset} from the oscillation frequency f_{osc} can be defined as [5]

$$\mathcal{L}(f_{offset}) = \frac{F \cdot K T}{A^2} \cdot \frac{R_P}{Q^2} \cdot \left(\frac{f_{osc}}{f_{offset}} \right)^2 \quad (2)$$

For a constant I_{bias} , and f_{osc} , the design tradeoff is that the phase noise of the VCOs is inversely proportional to the square of the tank's Q factor, and for constant Q , the phase noise is proportional to the size of L .

II. LC VCO TOPOLOGY SELECTION

The VCO is an oscillator whose output signal is a function of tuning voltage [6]. Although ring- or relaxation type oscillators can be found in some applications, their poor

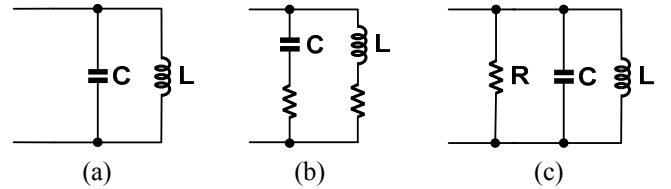


Fig. 1. Models for the LC tank (a) model for the ideal, (b) model with series resistance, and (c) model with parallel resistance.

phase noise performance prohibited them in most RF applications. LC tank based VCOs are widely used because of their outstanding phase noise performance.

Inductor and capacitor combined together with active circuit to compensate the passive elements losses build the LC VCOs. The parallel combination of these passive elements is the LC tank where the resonance frequency is defined as $\omega_0 = 1/\sqrt{LC}$ where inductor's impedance and capacitor's impedance are equal but opposite in their sign as shown in Fig. 1(a). Practically, these passive elements suffer from a series resistance which is shown in Fig. 1(b) and an

equivalent parallel resistance can be formed in place of the series resistance [Fig. 1(c)]. If the quality factors of those passive elements are expressed by Q_L and Q_C respectively, they can be given by

$$Q_L = \frac{\omega L}{R_{SL}} = \frac{R_{PL}}{\omega L} \quad (3)$$

$$Q_C = \frac{1}{\omega C R_{SC}} = \omega C R_{PC} \quad (4)$$

Since the capacitor's Q is higher than that of the inductor, the inductor's Q determines the tank Q in LC VCO tank design.

The single transistor VCO topology can be realized by using the LC tank which could be connected at the drain of the MOS transistor and the gate or source can be connected to the feedback signal. Single transistor oscillator topology (e.g. Colpitts oscillator) requires a simple, yet considered as the reliable topology where harmonic distortion is considered to be not critical [7].

Cross coupled LC tank VCO topology shown in Fig. 2 is the most common and convenient topology in modern CMOS technology where transistors M_1 and M_2 are connected in differential configuration. The inductor is an on-chip spiral integrated inductor and the varactors are the MOS capacitors as the frequency tuning elements. The integrated on-chip spiral inductors has relatively low quality factor compare to the MOS varactors. For the NMOS cross-coupled topology as shown in Fig. 2(a), the DC bias point is defined by setting V_{GS} and V_{DS} equal to V_{DD} , and the NMOS transistors operate in saturation; the drain current can be expressed as [6]:

$$I_{DS} = \frac{\mu_n C_{ox}}{2} \left(\frac{W}{L} \right) (V_{GS} - V_{TH})^2 \quad (5)$$

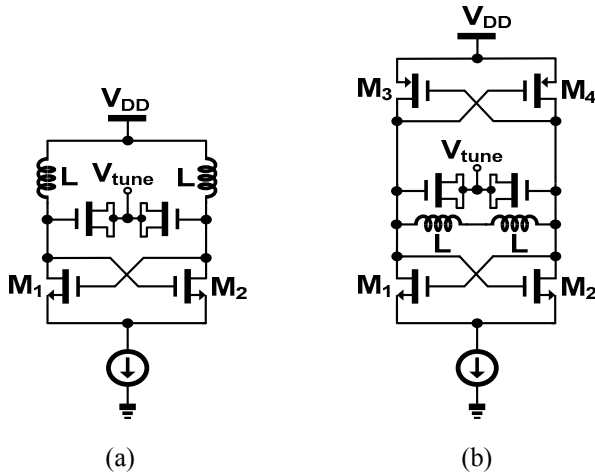


Fig. 2. Cross-coupled LC -tank VCO topology (a) NMOS only, and (b) CMOS cross-coupled

Where μ_n defines surface mobility of the electrons (NMOS transistor), C_{ox} represents gate oxide capacitance, and V_{TH} defines the threshold voltage. The transconductance can be calculated for the low frequency model:

$$g_m = \frac{\partial I_{DS}}{\partial V_{GS}} = \mu_n C_{ox} \left(\frac{W}{L} \right) (V_{GS} - V_{TH}) \quad (6)$$

The CMOS cross coupled LC VCO topology shown in Fig. 2(b) adds an additional cross coupled pair of PMOS placed on top of the NMOS cross coupled topology. CMOS LC VCO topology has the advantage that the peak output amplitude is twice compared to the NMOS only topology for the same bias current. Additionally, the CMOS topology generates a negative conductance of $-2/g_{M,N} - 2/g_{M,P}$ for a given bias current, where, for the NMOS only topology, the negative conductance is only $-2/g_{M,N}$.

III. DEVICE SELECTION

A. Integrated Spiral Inductor

Inductors are widely used as the frequency determining LC tank in VCO design. The lack of high quality on-chip inductors are by far the most conspicuous of standard CMOS process for RF applications. The reason behind that, the magnetic flux is created through loops which are small and relatively lossy conductors are used. On the other hand, compared with other passive elements, on-chip inductors show poor area efficiencies [8]. The most common on-chip spiral inductor structure forming its coils with at least two metal layers to form the basic spiral coil and the underpass contact to return the inner terminal of the coil to the outside.

In case of differential LC VCOs, the differential output signals appear at the inductor nodes. The swing of this signal is large and the inductor is subject to large signal AC currents. A simple lumped-element model (π -model) is shown in Fig. 3 to describe the electrical device characteristics and its loss. Where L is the series inductance, R is the series resistance, C_{OX} is the oxide capacitance, C_{SUB} is the substrate capacitance, R_{SUB} is the resistance of conductive substrate, and C_P is the shunt capacitance. The DC resistive losses are exacerbated by the skin effect, which causes a non-uniform current distribution in a conductor at RF. The consequence is a reduction in the effective cross-section, increasing the series resistance. Together with the series resistive loss, capacitance to the substrate is the other conspicuous problem of on-chip spiral inductors [9].

For the low phase noise and low power performance of the LC VCO, the LC tank should have high Q . Generally, the Q of inductors are lower than the Q of conventional diodes or varactors and therefore, improve the performance of integrated inductors have been the active research topics. More turns can be filled to increase the inductance for a given inductor area.

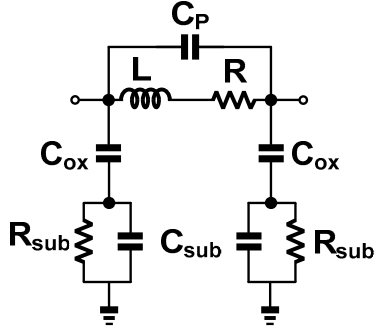


Fig. 3. Lumped model of square planar spiral inductor.

This approach is normally prohibited as small inductance and significant loss can be contributed by the inner turns.

Connecting multiple metal layers in parallel is a common practice for improving Q which will reduce the series resistance for the coil resulting low self-resonance. Inductors driven differentially have a higher quality factor and common mode or even mode capacitive parasitics along with the associated losses can be suppressed by using differential structure, thus improves the self-resonance frequency and Q .

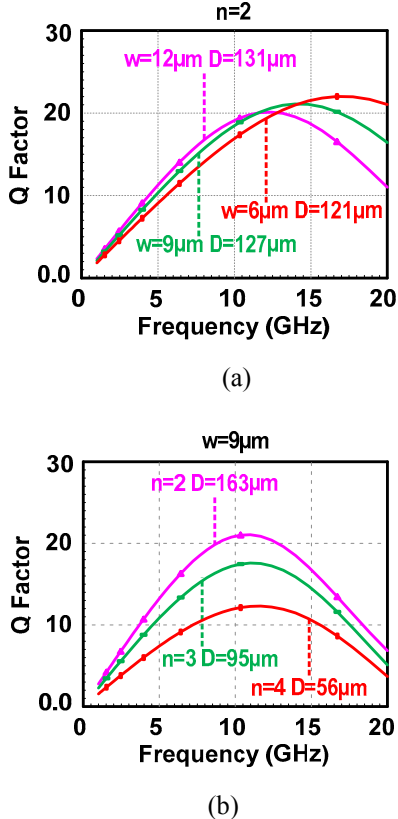


Fig. 5 (a) Q versus frequency of a 1-nH inductor for different widths (w), $n = 2$. (b) Q versus frequency of a 1.5-nH inductor for different turns (w), and $w = 9\mu m$.

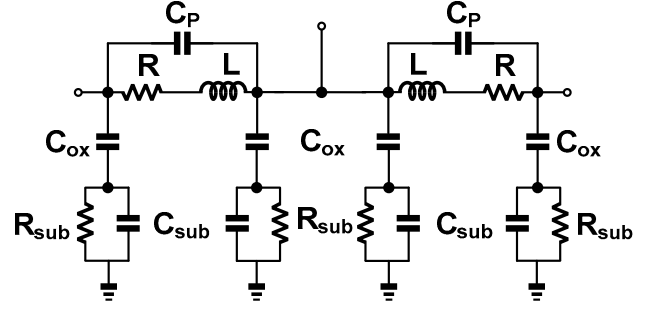


Fig. 4. Centre-tapped model of square planar spiral inductor.

The series resistance itself is frequency dependent and due to proximity and skin effects, it can be increased since the frequency increases. Initially, the Q of the inductor rises linearly with frequency as the coil's dc series resistance dominating the loss. Therefore, Q reaches to its maximum value, thus it experiences a fast decline as frequency approaches the self-resonance frequency of the coil. As shown in Fig. 5(a), for a 1-nH inductance in 65 nm CMOS (number of turns is kept constant to $n = 2$), higher Q can be achieved for large width, while Fig. 5(b) shows the Q variations for different number of turns. For the value of 1.5-nH inductance, higher Q is achieved for minimum number of turns but with increased area.

B. MOS Varactor

Integrated voltage controlled capacitor (varactor) is widely used as frequency tuning element in LC VCO for RF circuits. Together with the integrated inductor, the varactors build the core of the VCOs and determine the VCO tuning frequency. As the inductance is constant, the varactors should provide a large capacitance ratio C_{max}/C_{min} to allow for a large tuning range for the VCO.

In CMOS technology, a number of options available for the realization of varactor tuning elements. Reverse-biased p-n junctions can be realized as varactors, where this can be accomplished using the available p+/n- diffusions and N-or P-wells. However, by large voltage swings, this kind of junction varactors can be forward-biased. This can be avoided with proper layout techniques and the achievable quality factor can be higher.

MOS capacitor can be realized with a MOS transistor by connecting together the drain, source, and body (D, S, B) terminals. The voltage between body and gate (V_{BG}) will set the capacitance [10]. An inversion channel is formed with mobile holes for $V_{BG} > |V_{TH}|$ for a PMOS capacitor, where V_{TH} is defined as the threshold voltage of the transistor. MOS capacitor works in the strong inversion region for $V_{BG} \gg |V_{TH}|$ where MOS device acts as a transistor. An accumulation region builds up for the MOS device when $V_B > V_G$, and for the free movement of electrons, the voltage is a positive value

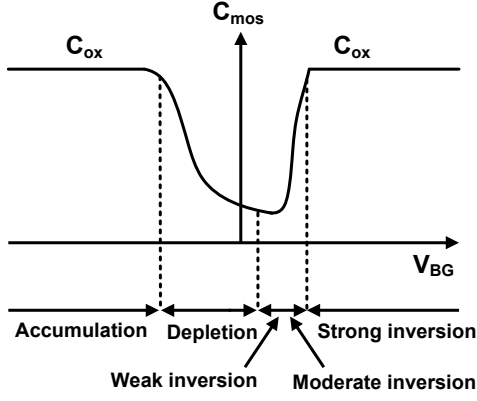


Fig. 6. Tuning characteristics for the PMOS capacitor with $B = D = S$.

at the interface between gate oxide and semiconductor. The value of the MOS capacitance C_{mos} is equal to $C_{ox} = \epsilon_{ox}S/t_{ox}$ in both strong inversion and accumulation region where S and t_{ox} are the transistor channel area and the oxide thickness, respectively.

For intermediate values of V_{BG} , three more regions can be distinguished (moderate inversion, weak inversion, and depletion) as shown in Fig. 6. Now C_{mos} can be modeled as C_{ox} in series with the parallel of capacitances C_b and C_i , where C_b accounts for the modulation of the depletion region below the oxide, and C_i represents the holes variation at the gate oxide interface. If C_b dominates, the MOS device is working in the depletion region, while, if C_i dominates, the MOS device is working in the moderate inversion region.

Dc tuning characteristics for the MOS varactor can be different based on the terminal connections. Drain, source, and body terminals can be connected together to form a capacitance between the gate and the $D = S = B$ terminal and a non-monotonic tuning characteristic is observed by this kind of varactor which is difficult to handle by a frequency synthesizer. Simulated capacitance versus tuning voltage (C-V) curve for a PMOS varactor of this type is shown in fig. 7(a). An almost monotonic function for C_{mos} can be obtained if the transistor does not enter the accumulation region for a very wide range of V_g . Connecting B to one of the supplies (highest dc-voltage for PMOS and ground for NMOS) rather than connected with D and S this can be accomplished. This configuration is known as inversion mode varactor. Fig. 7(b) shows the simulated C-V curve for a PMOS varactor of this kind, where $V_g = 0.7$ V and $V_{DD} = 1.2$ V is used. Fig. 7(c) shows the variations of dc tuning characteristics for an inversion mode PMOS varactor with different V_g .

The parasitic resistance associated with the MOS channel that mainly determines the quality of a MOS varactor can be approximated for a PMOS channel working in inversion mode is given by:

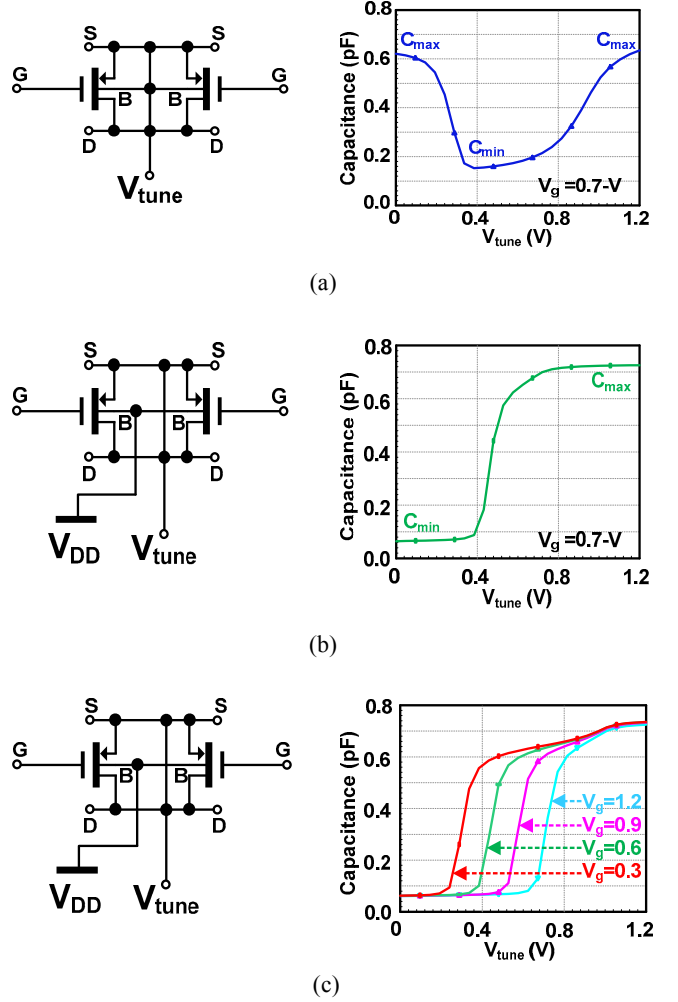


Fig. 7. (a) Terminal connection and simulated C-V curve for PMOS varactor $D = S = B$, (b) terminal connection and simulated C-V curve for the inversion type PMOS varactor, and (c) terminal connection and simulated C-V curve for different V_g for inversion type PMOS varactor (bottom). All varactor transistors have the width of 20- μ m and length of 0.5- μ m.

$$R_{mos} = \frac{L}{12K_p W(V_{BG} - |V_{TH}|)} \quad (7)$$

Where W , L , and K_p represents the width, length and the gain factor for PMOS transistor. According to eqn. (7), when V_{BG} approaches $|V_{TH}|$, R_{mos} increases, and infinite when V_{BG} is equal to $|V_{TH}|$.

IV. CONCLUSIONS

An optimization of LC VCO for low phase noise performance based on passive elements selection is presented in this paper. Low phase noise of LC VCO mainly depends on the Q of the LC tank which is determined by the Q of the inductor. This optimization process is confirmed by the

simulation of Q for inductor and varactor in 65 nm CMOS process. Based on the simulation results, selection of inductors and varactors will find the solution for high performance LC VCO design for RF applications.

REFERENCES

- [1] M. Zargari, K. D. Su, C. P. Yue, S. Rabii, D. Weber, B. J. Kaczynski, S. S. Mehta, K. Singh, S. Mendis, B. A. Wooley, "5-GHz CMOS transceiver for IEEE 802.11a wireless LAN systems," *IEEE J. Solid-State Circuits*, vol. 37, no. 12, pp. 1688-1694, Dec. 2002.
- [2] W. Deng, K. Okada, and A. Matsuzawa, "Class-C VCO with amplitude feedback loop for robust start-up and enhanced oscillation swing," *IEEE J. Solid-State Circuits*, vol. 48, no. 2, pp. 429-440, Feb. 2013.
- [3] J. J. Kim and B. Kim, "A low-phase-noise CMOS LC oscillator with a ring structure," in *IEEE ISSCC Dig. Tech. Papers*, Feb. 2000, pp. 430-431.
- [4] J. Chen, F. Jonsson, M. Carlsson, C. Hedenas, and L. R. Zheng, "A low power, start-up ensured and constant amplitude class-C VCO in 0.18 μm CMOS," *IEEE Microw. Wireless Compon. Lett.*, vol. 21, no. 8, pp. 427-429, Aug. 2011.
- [5] S. A. Yu and P. R. Kinget, "Scaling LC oscillators in nanometer CMOS technologies to a smaller area but with constant performances," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 56, no. 5, pp. 354-358, May. 2009.
- [6] B. Razavi, Design of analog CMOS integrated circuits, New York: McGraw-Hill, 2000.
- [7] B. Razavi, RF Microelectronics. Upper Saddle River, NJ: Prentice Hall, 1998.
- [8] J. R. Long and M. A. Copeland, "The modeling, characterization and design of monolithic inductors for silicon RF IC's," *IEEE Journal of Solid-State Circuits*, vol. 32, no. 3, pp. 357-369, Mar. 1997.
- [9] A. M. Niknejad and R. G. Meyer, "Analysis, design, and optimization of spiral inductors and transformers for Si RF IC's," *IEEE Journal of Solid-State Circuits*, vol. 33, no. 10, pp. 1470-1481, Oct 1998.
- [10] P. Andreani and S. Mattisson, "On the use of MOS varactors in RF VCOs," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 6, pp. 905-910, Jun. 2000.

A New LSTM Model by Introducing Biological Cell State

Lamia Rahman, Nabeel Mohammed, Abul Kalam Al Azad,

Department of Computer Science and Engineering

University of Liberal Arts Bangladesh (ULAB)

Satmasjid Road, Dhaka, Bangladesh

Email:llamiarahaman5@gmail.com, nabeel.mohammed@ulab.edu.bd, abul.azad@ulab.edu.bd

Abstract—Long Short Term Memory (LSTM) has been a very successful augmented recurrent neural network model employed to learn sequential information with long term dependencies where LSTM can store and compute information for a long period of time. In this study, a biologically inspired variation has been incorporated in LSTM by introducing additive cell state into the functionally computational system. The novel biological variant of LSTM model has been employed to conduct sentiment analysis of textual data. As the learning dataset, fifty thousand movie reviews have been used from IMDB where equal number of review data has been used for training and testing purposes. The comparative performance of the new variant is found to be promisingly better and show more stability than that of the traditional LSTM.

I. INTRODUCTION

In the machine learning computational tool repertoire, Long Short Term Memory (LSTM) is widely employed in learning varied long term sequential dataset. LSTM typically learn a sequence with long term dependencies by virtue of a memory cell infrastructure that uses the interplay of three gates, namely, an input, a forget and a output gate. Since its inception, this model has undergone a handful of variations mostly retaining the recurrent architecture and multiplicative gates in processing the input sequences. Though the interactive structure of multiplicative gates in the model has been very effective, it may be noted that the neurons building blocks of biological nervous system embed an additive interplay of the different channel gates to computational purposes. In this study, a biologically motivated additive inner structure of the interacting gates are considered aligned with the traditional chemically driven different gates or channels elucidated by the seminal works of Hodgkin and Huxley. The Biologically motivated variant of the model proposed here is employed to learn standard IMDB dataset of fifty thousand entries. The performance of the variant model is compared with that of the traditional LSTM with interesting outcome.

The paper is arranged as following: after a brief interlude to the traditional LSTM different variations of the model is discussed. After that the biological motivation of the proposed model is explored followed by elucidation of the proposed biologically motivated model. A brief discussion of the IMDB dataset is presented and the model is employed for sentimental analysis of the dataset. The result and different metrics of

performance are presented thereafter followed by concluding remarks.

II. BACKGROUND

Neural Network based different methods have made great progress in the recent past. Applications like handwriting recognition, speech recognition, language modeling, image captioning and many more have enjoyed renewed success through the use of Neural Networks and deep architectures.

A. Artificial Neural Network (ANN)

Artificial neural network is a computational model which is inspired by natural neurons[3]. Natural neurons have four basic parts which are cell body or soma, axon, dendrites or nerve ending and synapses. In the biological brain dendrites act like input channels; they receive their inputs through the synapses of other neurons. The soma processes the information and turns the processed information as an output value which is sent to the other neurons. In 1943, the first conceptual model of a neural network was developed by Warren S. McCulloch, a neuroscientist, and Walter Pitts, a logician [9]. Like natural neurons, artificial neurons were designed to have input layers, where the input is given, then some hidden layers which process the information and forwards to the output layer. This simple model, known as a Perceptron, is still popular today and used in many classification/learning tasks.

B. Recurrent Neural Network (RNN)

Most of us do not process large amounts of information from start to finish in one go. Instead we process small bits of information, build mental models, and process some more information and so on until we reach target. Traditional neural network models do not have the capacity to process information in steps or sequences. Recurrent Neural Networks were developed specifically to address the sequence learning problem.

RNNs incorporate a loop in its network model for information persistence. The basic feature of a RNN is that - it has at least one feed-back-connection, so the activation can flow round in a loop which enables the network to do temporal processing and learn sequences. It consists of a layer of inputs connected to a set of hidden memory cells, a connected set of

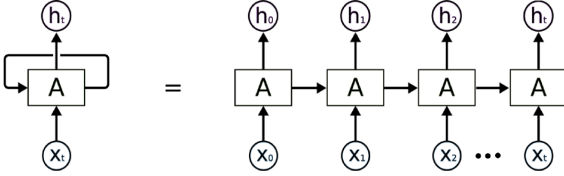


Fig. 1. An unrolled recurrent neural network[12]

recurrent connections amongst the hidden memory cells, and a set of output nodes.

These models can be visualized as multiple copies of the same network, each passing a message to a successor (Figure 1). One of the appeals of RNNs is that they might be able to connect previous information to the present task. Traditional RNNs were indeed able to achieve this, but for short sequences. Unfortunately, as that gap grows, RNNs are unable to successfully model long term dependencies. Long Short Term Memory (LSTM) was developed to address this short coming of traditional RNNs.

C. Long Short Term Memory (LSTM)

Long Short Term Memory networks are a special kind of RNN, capable of learning long-term dependencies. It is widely used because of its feature of remembering information for long periods of time. This is done by creating special modules that is designed to allow information to be gated-in and gated-out when needed. A basic difference between LSTM and RNN is that LSTM uses a memory cell with linear activation function to store information. LSTM was first proposed almost 20 years ago [4]. The design incorporates a memory cell using logistic and linear units with multiple interactions as shown in Equations 1 – 5. Where σ is the logistic sigmoid function and i , f , o and c are respectively the input gate, forget gate, output gate, and memory cell activation vectors and have exactly the same equations with different weights W s.

$$i_t = \sigma(W^{(xi)}x_t + W^{(hi)}h_{t-1} + W^{(ci)}c_{t-1} + b^{(i)}) \quad (1)$$

$$f_t = \sigma(W^{(xf)}x_t + W^{(hf)}h_{t-1} + W^{(cf)}c_{t-1} + b^{(f)}) \quad (2)$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tanh(W^{(xc)}x_t + W^{(hc)}h_{t-1} + b^{(c)}) \quad (3)$$

$$o_t = \sigma(W^{(xo)}x_t + W^{(ho)}h_{t-1} + W^{(co)}c_{t-1} + b^{(o)}) \quad (4)$$

$$h_t = o_t \cdot \tanh(c_t) \quad (5)$$

Each memory cell c_t has its net input modulated by the activity of an input gate, and has its output modulated by the activity of an output gate. These input and output gates provide a context-sensitive way to update the contents of a memory cell. The forget gate modulates amount of activation of memory cell kept from the previous time step, providing a method to quickly erase the contents of memory cells.

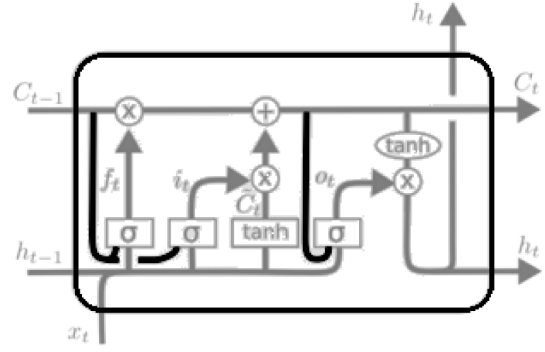


Fig. 2. LSTM variant: With peephole connection

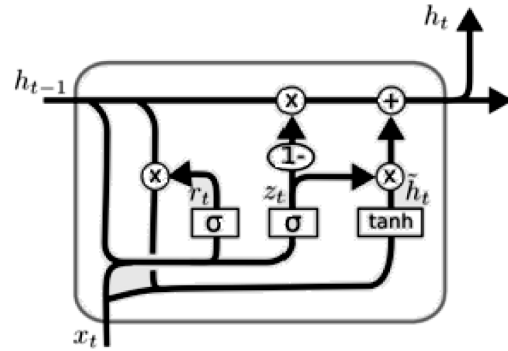


Fig. 3. LSTM variant: Gated Recurrent Unit

Alternate versions of LSTM have been proposed. Some of these are discussed below.

D. Variations of LSTM

A popular variant of LSTM, proposed by [2], adds peephole connections (figure 2), which enable the gates to incorporate the value of c_{t-1} in their updates.

$$f_t = \sigma(W_f \cdot [C_{t-1}, h_{t-1}, x_t] + b_f) \quad (6)$$

$$i_t = \sigma(W_i \cdot [C_{t-1}, h_{t-1}, x_t] + b_i) \quad (7)$$

$$o_t = \sigma(W_o \cdot [C_{t-1}, h_{t-1}, x_t] + b_o) \quad (8)$$

A slightly more interesting variation on the LSTM is the Gated Recurrent Unit (GRU) which was introduced by Cho, et al. [1]. It combines forget and input gates into a single gate which is known as an update gate (z_t in Equations 9 – 12). It also merges the cell state and hidden state. The resulting model, shown in figure 3, is simpler than standard LSTM models, and has been growing increasingly popular.

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \quad (9)$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \quad (10)$$

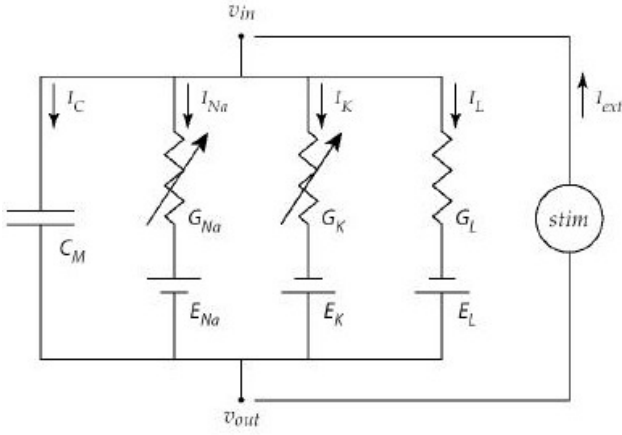


Fig. 4. Electrical equivalent circuit for a short segment of squid giant axon[6]

$$\tilde{h}_t = \tanh(W_r \cdot [r_t * h_{t-1}, x_t]) \quad (11)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (12)$$

III. BIOLOGICAL MOTIVATION FOR ADDITIVE MODEL

At the heart of the neuro-biological computation lies the physiological building block known as neuron. Neuron, anatomically, is primarily a cellular structure consists of two anatomical processes known as dendrite and axon which function as the input and output apparatus to and from neuron, respectively. The bio-chemical properties of the neuron are well elucidated by the works of Alan Hodgkin and Andrew Huxley [5]. In explain the firing or the action potential of neuron, sodium and potassium ion gates or channels with complex voltage-gated properties were posited into an elegant mathematical model.

Hodgkin and Huxley neuron model uses a series of voltage clamp experiments and varying extracellular sodium and potassium concentrations. In this celebrated model the properties of an excitable cell are described by a set of four ordinary differential equations which are given in Equations 13-16. Where I is the current per unit area, and α_i and β_i are rate constants for the i -th ion channel, which depend on voltage but not time. $\bar{g}_n$ is the maximal value of the conductance. n , m and h are dimensionless quantities between 0 and 1 that are associated with potassium gate activation, sodium gate activation, and sodium gate inactivation. It is interesting to note that the neuro-computations are governed by the interplay of the stochastic behavior relating the opening and closing rates of ion gates which appear mathematically additive as seen in Equation 14. The motivation of developing a biologically inspired LSTM came from the mathematical model of Hodgkin and Huxley where neurons, the neuro-computational unit in the living organism, process information controlling the traffic of different ions through gates. The relation between Hodgkin Huxley model and LSTM is that the latter model adopts a very similar mathematical structure of the input, output and forget gates as the related biological counterpart.

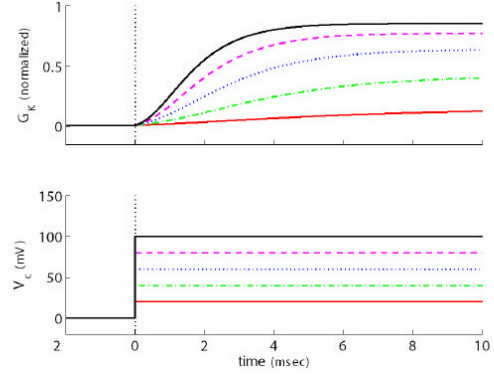


Fig. 5. Mathematical model of the voltage dependent and time dependent properties of the Na+ and K+[6]

Hodgkin and Huxley created a model by using a series of voltage clamp experiments and varying extracellular sodium and potassium concentrations. In that model the properties of an excitable cell are described by a set of four ordinary differential equations which are given below in Equations 13 – 16.

$$I = C_m \frac{dV_m}{dt} + \bar{g}_k n^4 (V_m - V_k) + \bar{g}_{Na} m^3 h (V_m - V_{Na}) + \bar{g}_l (V_m - V_l) \quad (13)$$

$$\frac{dn}{dt} = \alpha_n(V_m)(1 - n) - \beta_n(V_m)n \quad (14)$$

$$\frac{dm}{dt} = \alpha_m(V_m)(1 - m) - \beta_m(V_m)m \quad (15)$$

$$\frac{dh}{dt} = \alpha_h(V_m)(1 - h) - \beta_h(V_m)h \quad (16)$$

Where I is the current per unit area, and α_i and β_i are rate constants for the i -th ion channel, which depend on voltage but not time. $\bar{g}_n$ is the maximal value of the conductance. n , m , and h are dimensionless quantities between 0 and 1 that are associated with potassium channel activation, sodium channel activation, and sodium channel inactivation. The main motivation behind creating a biological LSTM came from the biological model of Hodgkin and Huxley where they created an additive model of giant squids axons. The gating functions in an LSTM model resemble the functionality of the three ion channels (leakage, Sodium, Potassium channels). However, a study of the traditional LSTM model, in Equations 1-5, shows that it is multiplicative model. In this work we modify the LSTM model to have an additive cell state, thus rendering it closer to the biological model of Hodgkin and Huxley.

IV. PROPOSED MODEL: BIOLOGICALLY VARIANT LSTM

We propose a biologically inspired LSTM model as outlined in the previous section. Our proposed model is shown in Equations 17 – 21, with the particular changes to the cell state shown in Equation 19.

Here the value is determined for is 0.001 for ensuring computational precision. Our only change was to the way in

which we update c , which in this new model is an incremental update from the older cell state, c_{t-1} , where the increment depends on a fraction of the summation of i and f . This new model is much simpler and biologically plausible.

$$i_t = \sigma(W^{(xi)}x_t + W^{(hi)}h_{t-1} + W^{(ci)}c_{t-1} + b^{(i)}) \quad (17)$$

$$f_t = \sigma(W^{(xf)}x_t + W^{(hf)}h_{t-1} + W^{(cf)}c_{t-1} + b^{(f)}) \quad (18)$$

$$c_t = c_{t-1} + \delta * (i + f + b_c) \quad (19)$$

$$o_t = \sigma(W^{(xo)}x_t + W^{(ho)}h_{t-1} + W^{(co)}c_{t-1} + b^{(o)}) \quad (20)$$

$$h_t = o_t \cdot \tanh(c_t) \quad (21)$$

V. METHOD AND DATASET

In our proposed model of LSTM, we are introducing a change to how the model updates/calculates the cell state c_t , to more closely resemble a biological process (more on this in Section IV). The performance of proposed model will be compared with that of LSTM on a sentiment analysis task on the popular IMDB movie review dataset.

A. Sentiment Analysis

Sentiment analysis is a computational process to automatically estimate the sentiment expressed by people, (most commonly) in written material. The sentiment could have been expressed over anything worth having an opinion on - a product we want to use; book we want to read or can be a movie we want to watch. Sentiments have been categorized into two basic categories: positive or negative and also an n-point scale eg. good, very good, satisfactory[10]. In this age of globalization, e-commerce sites have achieved great popularity and success. This has also meant that it is becoming increasingly difficult to manually sift through all the product reviews, as the volume of comments and reviews generated by platforms such as Twitter, Facebook, blogs is simply too high for manual inspection [11]. Sentiment analysis tools can play an important role here, as they can automatically gauge user/review sentiment. Since early 2000, sentiment analysis has become one of the most popular areas in natural language processing [7]. Plainly, the challenge for a sentiment analysis model is to understand the sentiment, to detect which words are using as a positive review or which words are for negative. Normally “good”, “amazing”, “wonderful” are thought as a positive words and “bad”, “terrible”, “poor” are considered as a negative sentiment words, and also there are other phrases which express sentiment. A list of such words or phrases are known as a sentiment lexicon[7]. However, a sentence containing such earmarked words may not actually express the assumed sentiment. This is especially true for very human actions such as sarcasm. Essentially systems which are more robust than just word-based solutions are required. Neural Networks, in particular Recurrent Neural Networks and its variants have enjoyed resurgence in this area over the last few years.

B. Dataset

To test our model on a sentiment analysis task we decided to test against a standard dataset. We are using the IMDB movie review dataset [8] to do the sentimental analysis on our model. The dataset was constructed by collecting 5000 reviews from movies from IMDB, which doesn't allow more than 30 reviews per movie. The dataset has an even number of positive and negative reviews, so randomly guessing yields 50% accuracy. Based on previous work on polarity classification, only highly polarized classification are considered. The score is given for negative review is $\bar{i}=4$ out of 10, and $\bar{i}=7$ out of 10 for positive reviews. Neutral reviews were not included in the dataset. The dataset was evenly divided for the training and test sets.

C. Implementation Detail

Our models were implemented using the open source, Python-based, Keras deep learning library. We utilized Keras sequential model, which allowed us to create a multi-layer, deep networks. The first layer of the sequential model is an embedded layer, which is used to implement the word to vector representation [8]. The second layer is an LSTM/ layer which has 128 dimensional internal states. Max-feature is set to 2000 in the code which means it will take 2000 unique word from the dataset at the training time. The maximum length considered was set to 100 characters. Both models were trained on this data set for 15 epochs.

For comparison, we wrote a new Keras layer which implements our proposed version. A new model was created which mimics the one described above, except that the second layer (LSTM) layer was replaced with the one we developed.

VI. RESULTS

As with other neural network-based studies, we present our results in terms of the loss and accuracy during training and validation. The loss function used is categorical cross entropy and accuracy is expressed as a fraction with 1.0 being completely accurate. Before discussing the results, it is worthwhile to review that fact that the loss values are in fact used to train the network (through back propagation) and are ideally meant to be reflective of the generalised performance. Practically however we find that reduction in loss values are not always reflected through improvement in accuracy, however a level of mirroring is expected.

Figures 6 and 7 shows how the loss and accuracy measures for the training and validation sets. Two things are immediately apparent - first, the behaviour of the loss function is inconsistent between the training and validation sets. The inconsistency is highlighted even more when we take into account the accuracy measures, as the loss values do not accurately reflect the effectiveness of the model.

In comparison, Figures 8 and 9 shows how the loss and accuracy measurements change over the 15 epochs for the proposed model. The accuracy measures are comparable to the LSTM model. In both cases, the accuracy measures plateau out around the 80% mark. However, the divergence between the validation loss and training loss values is not as large

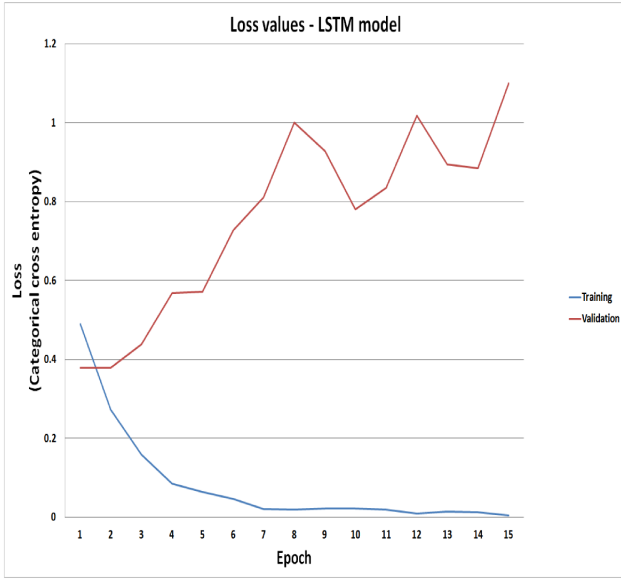


Fig. 6. Training and validation loss of LSTM

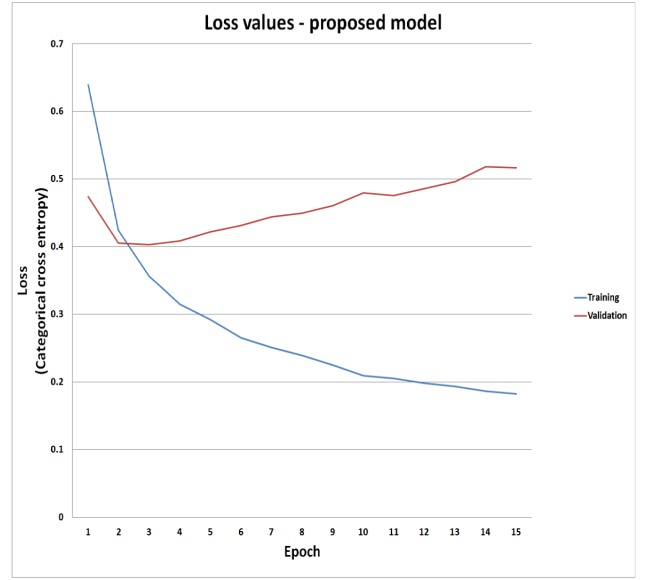


Fig. 8. Training and validation loss of the proposed model

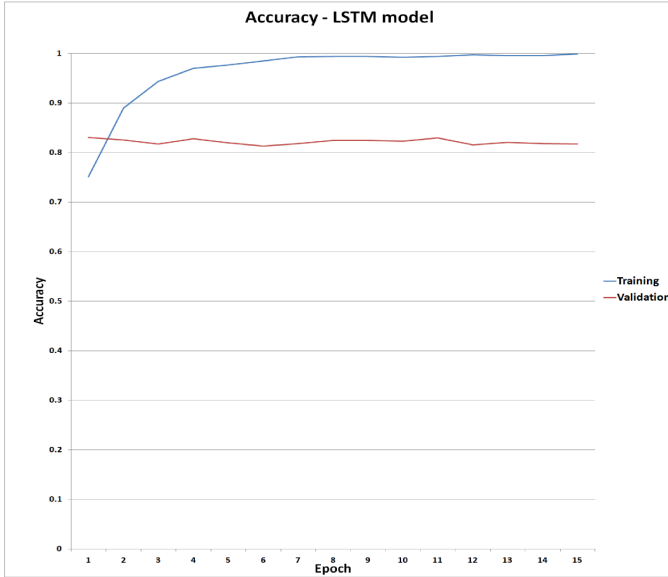


Fig. 7. Training and validation accuracy of LSTM

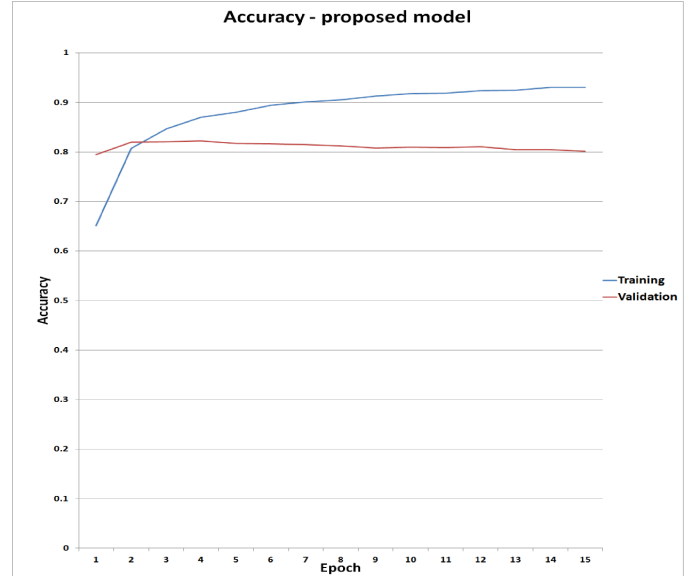


Fig. 9. Training and validation accuracy of the proposed model

as the LSTM model. Also, the proposed model is not as prone to over-training as the LSTM model, which can be seen by comparing the training accuracy for the LSTM and proposed models. It should be noted, that these differences arise purely through the introduction of an additive cell state model with no other changes done in the training pipeline. These two things, lower divergence of the loss values and better resistance to over training leads us to conclude the the proposed additive model displays more stability compared the LSTM's multiplicative model, while maintaining comparable accuracy.

Finally for a direct comparison, figure 10 shows the validation accuracy of the LSTM and proposed models. Both models

stagnate at around 80% accuracy, so the proposed model does not result in an improvement in accuracy. However in terms of consistency and stability, the proposed additive model shows superior behaviour over the training and validation sets while being less susceptible to over-training.

VII. CONCLUSION

In this study, the proposed model incurs the motivation from the traditional neuronal cell model structure where the interactive gates are additive rather than multiplicative as in the traditional LSTM. This rather simple modification to the model when applied in a sentimental analysis task yielded a better comparative stability in training and testing. The comparative analysis demonstrates the training loss and validation

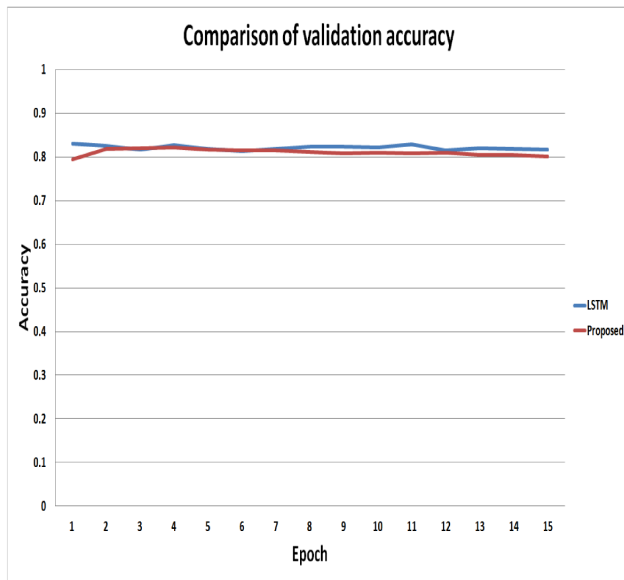


Fig. 10. Comparison of validation accuracy

loss of traditional LSTM, showing excellent results at the time of training period as the loss function is almost null after some time from the start. But when it comes to testing, the loss increases continuously, where the loss and validation loss of the proposed variant shows more stable result. The loss of the proposed model is not as good as LSTM at the time of training but when it comes to testing, the growth of the the loss function is not nearly as acute as that of the LSTM model.

The discrepancy between loss and validation loss of the biologically motivated variant is found to be significantly lower compared to that of the LSTM. Furthermore, the LSTM model achieves almost 100% training accuracy within the 15 epochs, but the validation accuracy showed no significant improvement; which leads us to conclude that the LSTM model was over training.

We have demonstrated that the proposed additive model is not as prone to over training, at least for the sentiment analysis task shown in this paper, while achieving a comparable validation accuracy, in the same number of epochs. As the current model shows promise, the further study is planned to accommodate noise as per the biological reality and a wider scope of the learning dataset is deemed for implementation.

REFERENCES

- [1] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.
- [2] F. A. Gers and J. Schmidhuber, "Recurrent nets that time and count," in *Neural Networks, 2000. IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference on*, vol. 3. IEEE, 2000, pp. 189–194.
- [3] C. Gershenson, "Artificial neural networks for beginners," *arXiv preprint cs/0308031*, 2003.
- [4] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [5] A. L. Hodgkin and A. F. Huxley, "A quantitative description of membrane current and its application to conduction and excitation in nerve," *The Journal of physiology*, vol. 117, no. 4, p. 500, 1952.
- [6] S. Koslow and S. Subramaniam, "Electrophysiological models in: Databasing the brain: From data to knowledge," 2004.
- [7] B. Liu, "Sentiment analysis and opinion mining," *Synthesis lectures on human language technologies*, vol. 5, no. 1, pp. 1–167, 2012.
- [8] A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, "Learning word vectors for sentiment analysis," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. Association for Computational Linguistics, 2011, pp. 142–150.
- [9] D. Shiffman, S. Fry, and Z. Marsh, *The nature of code*. D. Shiffman, 2012.
- [10] G. Vinodhini and R. Chandrasekaran, "Sentiment analysis and opinion mining: a survey," *International Journal*, vol. 2, no. 6, 2012.
- [11] X. Wang and Z. Min, "Parallel algorithm for hill cipher on mapreduce," in *Progress in Informatics and Computing (PIC), 2014 International Conference on*. IEEE, 2014, pp. 493–497.
- [12] K. Yao, T. Cohn, K. Vylomova, K. Duh, and C. Dyer, "Depth-gated recurrent neural networks," *arXiv preprint arXiv:1508.03790*, 2015.

BER Performance Limitation of an Optical Link with Multicore Fiber (MCF) Due to Crosstalk Caused by Mutual Coupling Among the Cores

Md Al Amin Sajeeb, Member, *IEEE*, Farhat Tasnim Farin, Member, *IEEE*, Abdur Rahim¹
and S.P.Majumder²

¹Department of EECE, MIST
Dhaka, Bangladesh

²Department of EEE, BUET
Dhaka, Bangladesh

Abstract— Analysis is presented for an optical link by using a multicore fiber (MCF) considering the effect of mutual coupling among the fiber cores which results in optical crosstalk. Analysis is carried out to find the optical crosstalk power at the output of the multicore fiber (MCF) and the photocurrents due to optical signal and crosstalk at the output of the receiver low pass filter. The signal to crosstalk plus noise ratio (SCNR) and the bit error rate (BER) are then determined for a given data rate. The results show that there is significant amount of optical crosstalk at the output of the multicore fiber (MCF) and thus limits the maximum allowable transmission length at a given data rate. Further, crosstalk is found to be dependent on optical carrier wavelength and crosstalk is found to be minimum at certain wavelengths for a given number of cores of multicore fiber (MCF).

Keywords— Multicore Fiber (MCF), Cross-Talk, bit error rate (BER), Coupling Co-efficient, signal to crosstalk noise ratio (SCNR), penalty.

I. INTRODUCTION

Optical fiber is one of the most authentic transmission medium for expeditious outgiving of information. Multicore fiber is the latest surprise of optical technology. Implicating with multiplexing technology like space division multiplexing and mode division multiplexing different multicore fibers are introduced [1-3].

Generally MCF has mode to mode coupling features between neighboring cores and hence the transmission of optical signal is tethered away [4-5]. In multimode transmission fibers, unintended mode coupling can arise from several sources. In MCF, there are two types of crosstalk; intra-core and inter-core crosstalk. The Intra-core crosstalk is caused by the coupling between the fiber modes within each core due to the fiber imperfections such as index perturbations, bends, and twists. The inter-core crosstalk arises from the mode coupling between the neighboring cores [6-7]

In this paper analysis is presented for evaluating the bit error rate (BER) performance of an optical transmission with intensity modulation and direct detection using multicore fiber. The amount of optical crosstalk and crosstalk photocurrent at the output of the receiver is evaluated analytically. The SCNR and BER are evaluated for different fiber and system parameters and power penalty suffered by the system due to crosstalk are determined at a given BER of 10^{-9} .

II. SYSTEM MODEL

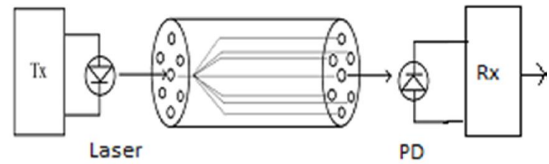


Fig.1 Block diagram of 7 core homogenous MCF.

A 7 core homogenous multicore transmission link is shown in Fig. 1. (The cores are identical and labeled as 1, 2, 3,.....7). Each core has core index 1.51 and cladding index is 1.44. As the system is homogeneous so all cores have same diameter of 4.5 micrometer and core to core distance is considered 30 micrometer [4]. This core to core distance is called pitch (Λ). Later on index diameter and pitch are varied to get desired result. Input is given into only one core but inside the fiber link it will be cross coupled. So the output signal will be found each core at the terminal end. Though the received signal is drawn from one core other cores also contain some parts of incoming signal due to cross coupling. Laser is considered as the source of input signal and photodiode (PD) is used for receiving desired core output. The wavelengths are varied from 770 nm to 1600 nm.

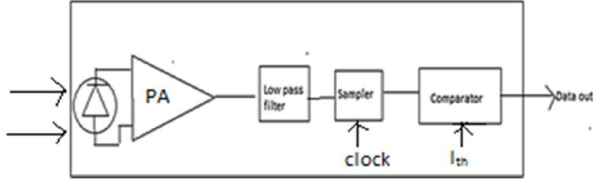


Fig.2 Block diagram of receiver model.

Block diagram of direct detection receiver is shown in Fig. 2. Incoming signals noise are reduced by LPF. Then final message signal is obtained from precision circuit (sampler and comparator).

III. SYSTEM ANALYSIS

The transmitted signal is given by[4]

$$E_p(t, 0) = E_0 e^{j\omega_c t} \quad (1)$$

when $z=0$.

The signal at a distance z is given by[4]

$$E_p(t, z) = E_0 e^{j\omega_c t} e^{-j\beta_p z} \quad (2)$$

The coupling coefficient C_{pq} is a measure of the spatial overlapping of the mode fields of core p and q over the cross-sectional area of core q . [6]

$$C_{pq} = \sqrt{2\Delta_q} W_q U_p K_0 \left(\frac{W_p d_{pq}}{a_p} \right) \left[a_p U_q J_1(U_q) I_0 \left(\frac{W_p a_q}{a_p} \right) + a_q W_p J_0(U_q) I_1 \left(\frac{W_p a_q}{a_p} \right) \right] [V_p J_1(U_q) K_1(W_p) (a_p^2 U_q^2 + a_q^2 W_p^2)]^{-1} \quad (3)$$

where J_1, I_1 and K_1 are the Bessel function of the first kind and the modified Bessel functions of the first and second kinds of order 1 respectively and d_{pq} is the distance between the centers of core p and q .

When light is launched into core1 (center core), i.e. $A_1(0)=1$ and $A_p(0)=0$ for $p \neq 1$, analytical solution for the mode amplitude at distance z are obtained as[6].

$$A_1(z) = \left[\cos(\sqrt{7}C_{12}z) + \frac{j}{\sqrt{7}}(\sin(\sqrt{7}C_{12}z)) \right] e^{-(jC_{12}z)} \quad (4)$$

And

$$A_p(z) = \left[-\frac{j}{\sqrt{7}}(\sin(\sqrt{7}C_{1p}z)) \right] e^{-(jC_{1p}z)} \quad (5)$$

This is the signal which is found at the end of the multicore fiber. The positive exponential indicates the signal propagation in a specific direction. Due to crosstalk, signal of p -th core will also enter the other remaining cores. So the signal in q -th core is given by

$$E_q(t, z) = E_0 e^{j\omega_c t} e^{-j\beta_q z}; \quad q \neq p \quad (6)$$

The total crosstalk signal is formed by the signal passing through all q -th cores except p -th core.

$$E_{CT}(t, z) = \sum_{q=1}^M E_q(t, z); \quad q \neq p \quad (7)$$

Now signal at 1st core at z distance is given by [5]

$$E_1(z) = E_1(0) \{ \cos(\beta_1 z) - j \sin(\beta_1 z) \} + \{ E_2(0) \cos(C_{21}z) + E_3(0) \cos(C_{31}z) + \dots + E_7(0) \cos(C_{71}z) \} - j \{ E_2(0) \sin(C_{21}z) + E_3(0) \sin(C_{31}z) + \dots + E_7(0) \sin(C_{71}z) \} \quad (8)$$

So, when input light signal is given to core 1, i.e. $A_1(0)=1$ and $A_p(0), p \neq 1$, then from equation (8) signal power (P_r) and crosstalk power (P_{CT}) for core 1 can be obtained as under

$$P_r = |E_1(0)|^2 \sqrt{\cos^2(\beta_1 z) + \sin^2(\beta_1 z)}$$

$$\text{or,} \quad P_r = |E_1(0)|^2 \quad (9)$$

and

$$P_{CT} = |E_2(0) \cos(C_{21}z) + E_3(0) \cos(C_{31}z) + \dots + E_7(0) \cos(C_{71}z) + E_2(0) \sin(C_{21}z) + \dots + E_7(0) \sin(C_{71}z)|^2 \quad (10)$$

So from equation (9) and (10) signal to crosstalk ratio can be calculated. So

$$(SCR)_{optical} = \frac{P_r}{P_{CT}} \quad (11)$$

Eventually signal current is obtained and that is given by

$$I_r(t) = R_d |E_q|^2 \quad (12)$$

where R_d is the responsivity of the photodiode.

And the crosstalk current can be found as following

$$I_{CT}(t) = R_d |E_{CT}|^2$$

$$\text{or,} \quad I_{CT}(t) = R_d \left| \sum_{q=1}^M E_q(t, z) \right|^2 \quad (13)$$

The received optical signal with crosstalk at the input of the photodetector at a given distance z from the transmitter is given by[5]

$$e_r(t, z) = \sqrt{2GP_r(z)} e^{j\omega_c t} + \sqrt{2P_{CT}(z)} e^{j(\omega_c t - \beta z)} \quad (14)$$

where G is the gain of the optical pre-amplifier.

The photo current at the output of the photodetector can be expressed as:

$$i_d(t, z) = R_d \cdot E |Re\{e_r(t, z)\}|^2$$

$$\begin{aligned}
&= R_d \cdot E |Re\{\sqrt{2GP_r(z)} e^{j\omega_c t} + \sqrt{2P_{CT}(z)} e^{j(\omega_c t - \beta z)}\}|^2 \\
&= R_d \cdot E |\sqrt{2GP_r(z)} \cos \omega_c t + \sqrt{2P_{CT}(z)} \cos(\omega_c t - \beta z)|^2 \\
&= R_d \cdot P_r(z) \cdot G + R_d \cdot P_{CT}(z) + 2R_d \sqrt{P_r(z) \cdot P_{CT}(z)} \cos \beta z \\
&= I_r(z) \cdot G + I_{CT}(t) + I_{shot}(t) + I_{r-CT}(t) \quad (15)
\end{aligned}$$

The pre-amplifier is followed by photodetector and the current at the output of pre amplifier is given by

$$I_o(t) = I_r(z) \cdot G + I_{CT}(T) + I_{shot}(t) + I_{r-CT}(t) + I_{th}(t) \quad (16)$$

where $I_{th}(t)$ is the thermal noise due to preamplifier.

The variance of the noise components can be expressed as:

$$\sigma_{shot}^2 = 2eB(R_d \cdot G \cdot P_r + R_d \cdot P_{CT}) \quad (17)$$

$$\sigma_{CT}^2 = (R_d \cdot P_{CT})^2 \quad (18)$$

$$\sigma_{r-CT}^2 = 4R_d^2 P_r \cdot P_{CT} \cdot \cos^2 \beta z \quad (19)$$

$$\sigma_{th}^2 = \frac{4KT}{R_L} B \quad (20)$$

where σ_{shot} , σ_{CT} , σ_{ASE} , σ_{th} are the shot noise, crosstalk noise, amplifier spontaneous noise and thermal noises respectively. R_d , n_{sp} , h , ν are the receiver responsivity, spontaneous emission factor, Planck constant and frequency of light wave respectively. B is the bandwidth of optical amplifier and B is the electrical bandwidth of the receiver. Then the electrical signal to crosstalk plus noise ratio at the output of low pass filter is given by

$$SCNR = \frac{I_r^2}{\sigma_{shot}^2 + \sigma_{CT}^2 + \sigma_{r-CT}^2 + \sigma_{th}^2} G^2$$

where $I_{CT} = R_d P_{CT}$ and $I_r = R_d P_r$. Now, for the direct detection system the BER can be evaluated from

$$BER = 0.5 \operatorname{erfc} \frac{\sqrt{SCNR}}{2\sqrt{2}}$$

where erfc is the complementary error function.

IV. RESULTS AND DISCUSSION

Following the theoretical analysis presented in sec. III we evaluate numerically the coupling coefficients and crosstalk due to coupling at the output of a multicore fiber (MCF). The plot of coupling coefficient versus wavelength (nm) is shown in Fig.3 and comparison between Fig. 5 and Fig. 6 has been made to ascertain the minimum coupling coefficient for a selected range of wavelength, varying the values of core diameter and the pitch, severally.

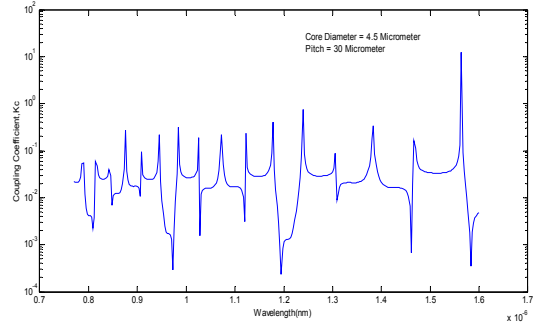


Fig. 3 Plot of Coupling Coefficient as a function of wavelength for core diameter=4.5 Micrometer and pitch=30 Micrometer.

The plots of coupling coefficient versus wavelength are demonstrated in Fig.4. Here, the coupling coefficients for different diameter of cores incorporated in an optical fiber are found. It is noticed that the minimum of all the coupling coefficient values is obtained for core diameter=4.5 micrometer.

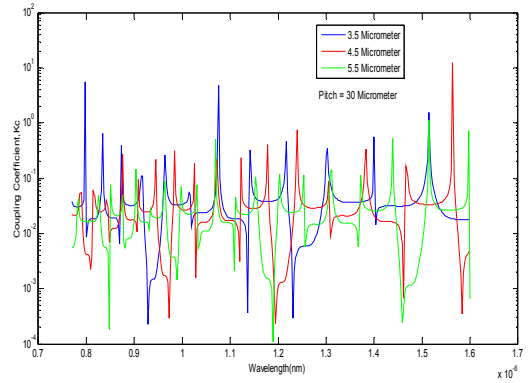


Fig. 4 Plots of coupling coefficient as a function of wavelength for different core diameter and pitch=30 Micrometer.

The plots of coupling coefficient versus wavelength for different values of pitch are demonstrated in Fig. 5. It is resolved that there is no substantial shift in values of coupling coefficient, due to the fact that coupling coefficient is a function of core diameter rather than core to core distance. The positions of coupling coefficients remain unchanged.

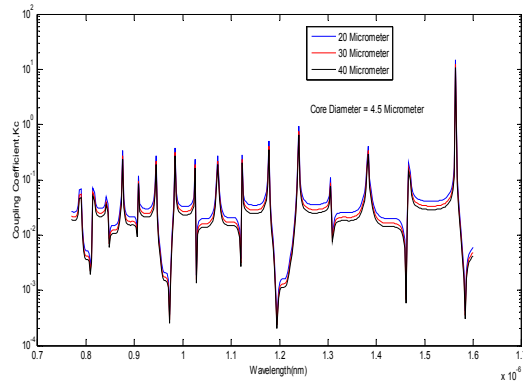


Fig. 5 Plots of coupling coefficient as a function of wavelength for different pitch and core diameter=4.5 Micrometer.

From Fig. 6 to Fig. 9 $\log_{10}$ BER vs. signal power for different core diameter is plotted for different coupling coefficient to compare with the state of zero coupling coefficient. So BER varies as a function of coupling coefficient and also as a function of core diameter. The variation in BER is less with core having greater diameter. Again, it can be reasoned that with less coupling,

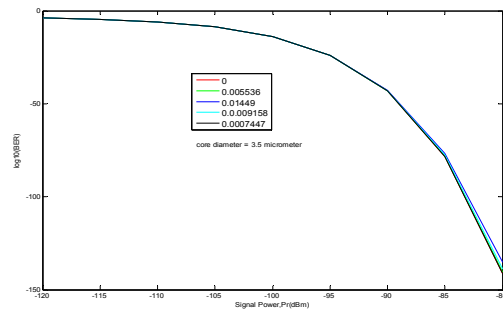


Fig. 6 Plots of as $\log_{10}$ BER a function of signal power for different coupling coefficient and core diameter=3.5 Micrometer.

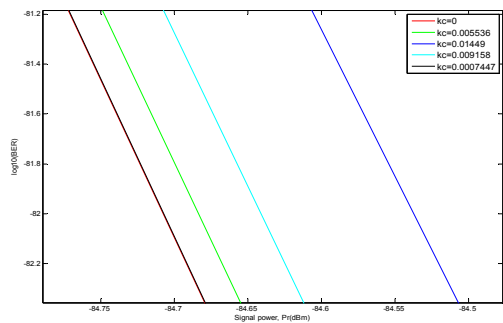


Fig. 7 Plots of $\log_{10}$ BER as a function of signal power for different coupling coefficient and core diameter=3.5 Micrometer (after zoom).

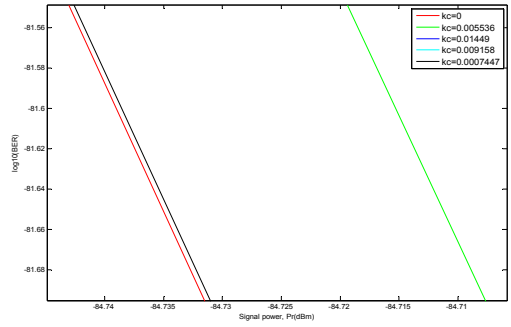


Fig. 8 Plots of $\log_{10}$ BER as a function of signal power for different coupling coefficient and core diameter=3.5 Micrometer (after zoom).

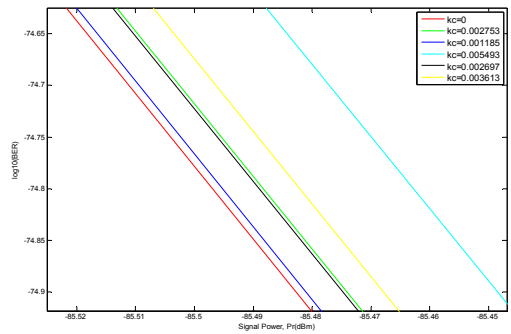


Fig. 9 Plots of $\log_{10}$ BER as a function of signal power for different coupling coefficient and core diameter=5.5 Micrometer (after zoom).

Fig.10 demonstrates the plot of penalty versus coupling coefficient for different core diameters. We get to see that for a specific amount of coupling, $K_c=0.006$, penalty is 0.1736 dB, 0.1348 dB and 0.08919dB for 5.5 micrometer, 3.5 micrometer and 4.5 micrometer, severally. So, for 4.5 micrometer the penalty is lowest in this case.

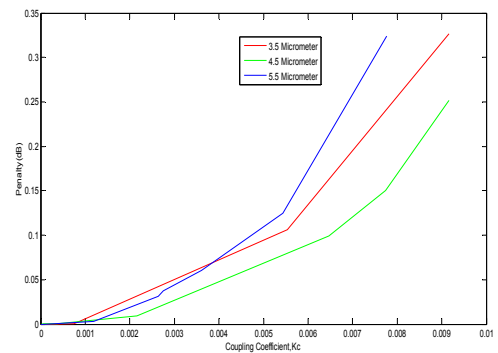


Fig. 10 Plots of penalty (dB) as a function of coupling coefficient for different core diameter.

V. CONCLUSIONS

An analytical approach is presented for a MCF optical link taking into consideration the crosstalk due to coupling among the cores. The results are presented in terms of BER and power penalty due to crosstalk. It is noticed that there are significant amount of crosstalk at certain wavelengths and system suffers power penalty at a given BER. The outcome of this research will find applications in design of MCF optical fiber transmission systems.

ACKNOWLEDGMENT

This work is a part of B.Sc. thesis work in the department of Electrical, Electronics and Communication Engineering in Military Institute of Science and Technology (MIST) Mirpur, Dhaka, Bangladesh.

REFERENCES

- [1] M. Koshiba, K. Saitoh, and Y. Kokubun, "Heterogeneous multi-core Fibers: proposal and design principle," IEICE Electronics Express, vol. 6, no. 2, pp. 98-103, Jan.25, 2009
- [2] Y. Kokubun and M. Koshiba, "Novel multi-core fibers for mode division multiplexing: proposal and design principle," IEICE Electronics Express, vol 6, no 8, pp. 522-528, Apr. 25, 2009.
- [3] T. Morioka, Y. Awaji, R. Ryf, P. J. Winzer, D. Richardson, and F. Poletti, "Enhancing optical communications with brand new fibers," IEEE Commun. Mag. vol 50, no 2, pp. s31-s42, Feb. 9, 2012.
- [4] J. H. Chang, H. G. Choi, S. H. Bae, D. H. Sim, Hoon Kim, and Y. C. Chung, "Crosstalk analysis in homogeneous multi-core two-mode fiber under bent condition" Opt. Express, vol. 23, issue 8, pp. 9649-9657 Apr. 20, 2015
- [5] Md Jahangir Hossain, S.P. Majumder, "Performance limitations due to crosstalk in an optical transmission link over multi-core fiber" 2015 IEEE International Conference on Telecommunications and Photonics (ICTP) (ieeexplore.org), Dhaka 26-28 Dec., 2015.
- [6] F.Y.M. Chan, A.P. T. Lau, and H.Y. Tam, "Mode coupling dynamics and communication strategies for multi-core fiber systems," Optics Express, vol. 20, no. 4, pp. 4548-4563, Feb. 13, 2012.
- [7] J. M. Kahn, K. Ho, and M. B. Shemirani, "Mode Coupling Effects in Multimode Fibers," Optical Fiber Communication Conference and Exposition (OFC/NFOEC), 2012 and the National Fiber Optic Engineers Conference, 4-8 Mar. 2012, Los Angeles, California, USA

Optimization of Interval Type-2 Fuzzy Logic Controllers with Rule Base Size Reduction using Genetic Algorithms

Sonia Yesmin, Animesh Kumar Paul,¹ Pintu Chandra Shill²

Department of Computer Science and Engineering

Khulna University of Engineering & Technology

Khulna, Bangladesh

¹animesh10kuet@gmail.com, ²pintucskuet@yahoo.com

Abstract—In this article, we propose optimization technique to develop type-2 fuzzy inference system (FISs) through hybrid genetic algorithms (HGAs). The proposed optimization technique works as follows: (i) Optimize the type-2 membership functions (ii) Learn the rule base through genetic algorithms (iv) Apply the reducing technique to reduce the rule base. (v) Build the FISs based on type-2 membership functions and the reduced rule base. For concurrently works step (i) and (ii), we used real and binary coded coupled genetic algorithms, where real coded genetic algorithms are used to tune the type-2 membership functions and binary coded genetic algorithms is used to learn and reducing the fuzzy rules. For intelligent control of a double inverted pendulum system, a control algorithm is used. Finally, the simulation studies demonstrate that the proposed type-2 fuzzy controller performance is better and comparable to the existing methods under normal conditions.

Keywords—Type-2 fuzzy logic controller; Binary and real coded couple genetic algorithms; Optimization; Two degrees of freedom inverted pendulum system.

I. INTRODUCTION

Type-1 Fuzzy systems (FSs) are usually generated using type-1 fuzzy sets (FSs). Real Application areas [1] of these FLCs are in the area of control [2-3], decision making [4-7, 28-29], prediction [8], classification [9], and inference [10]. Despite their popularity, difficulties in modeling and reducing the effect of uncertainties [11, 12] are present in type-1 FLCs for the sustainable number of researches. One restriction being that for each input, a type-1 fuzzy set is consolidating in the sense that a crisp value is the membership grade.

Recently, membership functions (MFs) that are themselves fuzzy have been exhibited in type-2 FSs and received the most attention. The key concept of such type-2 FSs used to develop type-2 FLCs (T2-FLCs) is the footprint of uncertainty (FOU). Type-1 fuzzy sets shape and position uncertainties are modeled with it using extra mathematical dimension.

The idea behind the type-2 fuzzy set (T2-FS) was first introduced by Zadeh [13] and it is a development and generalization of the type-1 fuzzy sets (T1-FSs). FLCs that are described using at least one T2-FS are called T2-FLCs [14]. It has been shown in [15] that uncertainties and nonlinearities are directly handled powerfully by T2-FSs.

Now a day's, many real applications such as forecasting time-series data [16], communication and networks [17], decision making [18], data and survey processing [19], modeling word [20], and recognition model [21] are modeled with T2-FLCs. The literature view further revealed

that only a few type-2 FLCs are employed in the controlling field although fuzzy control is an important area and widely used in the field of control. Interval type-2 FLCs have been employed for controlling mobile robot [12], sound speakers' quality control [22], and ATM networks' connection control [23].

Moreover, the number of rules on the rule base is an important factor while designing T2-FLCs. Even though it is understandable that for getting the better generalizing ability of fuzzy model, we need to remove the unnecessary, inappropriate, incorrect and conflicting fuzzy rules from the rule base. Finding the effective rules are not an easy exercise. For the real time application, fuzzy systems need rule base reduction techniques.

For the reduction of the size of rule base, a remarkable number of techniques have been developed [24, 25, 26, 27]. Some articles works on eliminating the redundant rules, merging the rules that contains some common selecting the important rules for the rule base.

Wang et al. introduced a multiple knowledge bases integration process based on genetic selection. Using GAs, [24] for the inference mechanism rules are extracted. Rule base is reduced by applying pruning and merging method [25] on rules.

For building the initial fuzzy inference system, a preset number of fuzzy rules and their respective type-1 fuzzy sets are used for achieving good performance and it contains a high computational complexity. Existing rule base reduction techniques are failed to preserve the inference. We need to get identical reduced rule bases and maintain outputs of the original system. Therefore, a technique is needed to be developed for reducing the rule base size and optimizing the type-2 MFs concurrently and constructing the fuzzy systems that resolve the existing problems and preserves the inference.

In this study, we propose optimization methods with reducing the rule base size to develop type-2 fuzzy inference system (FISs) through hybrid genetic algorithms (HGAs). The main scheme of optimization method is divided into two steps: (i) Optimize type-2 MFs (ii) Learning and reducing rule base. Type-2 MFs optimization is concerned with optimizing the parameter of type-2 fuzzy sets and rule learning-reducing phase used for automatically generating as well as reducing fuzzy rules. A new double coding schema, first binary coded chromosome part for rule learning as well as reducing and the second real coded chromosome part for tuning the type-2 fuzzy sets is developed in this paper. An intelligent control of double inverted pendulum is used for applying the control

algorithm. Finally, ensuring the performance of the controller and comparing to the existing methods under normal conditions, a simulation results are used.

Removing the erroneous rules and rules that reduce the system performance by setting zero to all consequent weight factors, combining the rules which are conflicting during the learning process are the main amenities of the proposed method. Fewer rules are used as rule base than the initial rule sets and the fuzzy models get improved computational efficiency and interpretability.

A significant difference exists in the proposed method compared with the described exiting method that makes a type-2 FLCs using hybrid GA. In the existing methods, for building a fuzzy system a preset number of rules is used. But in the article, to build the fuzzy model we proposed a model for selecting a number of rules automatically by hybrid GA. In [27], defining MFs parameters linguistic representation of 2-tuple is used and for learning the linguistic rules it used same representation as a post processing. On the other hand, for defining only MFs, we used two tuple real numbered representation and for learning and reduce the rule base, binary valued GAs is used. Generally the fuzzy model used random partition and width into initial fuzzy sets and center of MFs are obtained by using GAs, uniform partitioned are used for initial fuzzy sets and GA is used for obtaining the displacement of lateral.

The basic association of the article is given as follows: Section 2 discuss about the type-2 fuzzy logic controllers and fuzzy sets. In section 3, we give key concepts of fuzzy rules learning and reducing technique on the basis of double coding schema genetic algorithms and advanced the procedure of genetic solution. We describe the way how the linguistic rules are encoded and corresponding type-2 fuzzy sets with optimization technique also. In Section 4, an experimental result of control application is shown for the double inverted pendulum systems and finally the conclusion is described in Section 5.

II. TYPE-2 FUZZY SETS AND FLCs

A. Interval Type-2 Fuzzy sets

The ideas of a type-2 fuzzy system were introduced by Prof. Zadeh [6] in 1975. Membership function is used for constructing the type-2 fuzzy logic controller. The fuzzy grade value of the type-2 fuzzy set in the closed interval between [0, 1] rather than a point between [0, 1]. A type-2 fuzzy logic controller operates on type-2 fuzzy sets as shown in Fig.1, which are used to map the inputs and outputs of the fuzzy logic controller.

B. Type-2 Fuzzy Logic Controller (FLC)

Type-2 fuzzy logic controllers (FLCs) have five components: fuzzifier, fuzzy knowledge base (KB) that is combining rule base (RB) and database (DB) which is consists of shape and type of membership function, fuzzy inference engine, type- reducer and defuzzifier as shown in Fig.1. The structure of type-2 FLCs is same as type-1 FLCs with the only difference being the introduction of type-reducer components.

A fuzzifier is used to fuzzify the crisp input and transform it from crisp normalized values to fuzzified values usually T2 fuzzy sets to be used by the inference engine. In

this way activate the process of inference engine, used the rule base to generate the T2 fuzzy sets. Then, use the type-reducer that combines the output sets and performs a centroid calculation, leading to type-reduced sets known as type-1 fuzzy sets. The defuzzifier takes the type-1 (type reduced fuzzy sets) obtained at the output of type reducer then defuzzify the type-reduced sets to generate the crisp outputs.

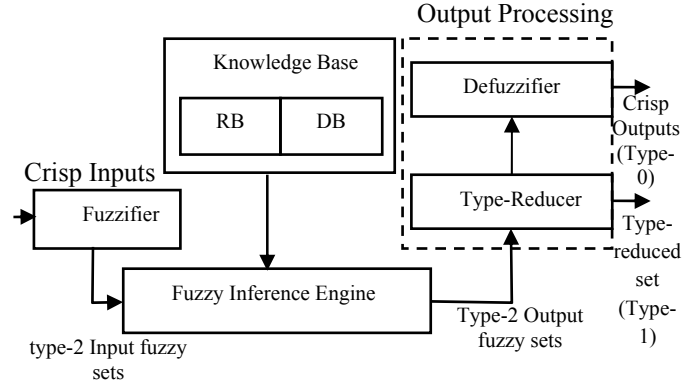


Fig.1. A type-2 Fuzzy Logic Controller

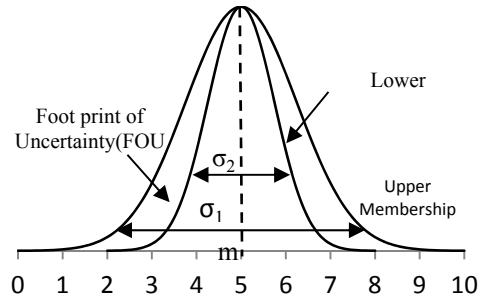


Fig.2. An Interval type-2 fuzzy set

III. GENETIC TYPE-2 FUZZY SYSTEM (GT2FS): OPTIMIZATION AND RULE REDUCING

On the basis of genetic algorithm (GA), by learning and reducing process a genetic type-2 fuzzy inference system (GT2FIS) is basically a fuzzy logic system augmented. Different complexity levels are covered by genetic learning and tuning processes according to the structural changes produced by the genetic algorithm, from the simplest case of optimize the parameters of fuzzy system to the highest complexity levels of learning and tuning the rule set of a fuzzy inference system. The fuzzy system parameter optimization has been proceeding utilized to adapt a wide range of different fuzzy systems, as in genetic-fuzzy or genetic neuro-fuzzy systems.

From the review of the existing fuzzy systems, most of the genetic T2 fuzzy logic controller used genetic algorithms for learning or tuning different parts of a fuzzy system (FSs). In this paper, we focus on genetic learning and reducing of fuzzy rule base and optimizing the parameters of MFs using genetic tuning for T2 fuzzy logic controllers.

A. Fuzzy Rule Learning and Reducing Technique

In this paper, the structure of fuzzy rules used as follows:

$R^i = \text{IF } x_1 \text{ is } \tilde{A}_1^i \text{ and } x_2 \text{ is } \tilde{A}_2^i \dots \dots \text{ Then } y \text{ is } w_i \tilde{C}^i$
where $x_1, x_2, \dots, x_n$ and y are problem domain

variables, $\tilde{A}_1^r$, $\tilde{A}_2^r$ and C^r are type-2 fuzzy sets defined over the domain of their corresponding variables, and w_r is the weight factor associated to rule r . The value of the weight factor, w_r is defined to be 0 or 1.

Let us assume that through the genetic algorithm, it learnt the N candidate fuzzy rules. Design an accurate and interpretable fuzzy logic controllers from the N candidate fuzzy rules are the task of genetic fuzzy rule selection technique.

To integrate genetic algorithms into our scheme, the rule base weight factors are encoded into a binary string. The number of weight factors on each rule depends on the number of linguistic variable of the output variable and this number is equal to the number of linguistic variable of the output variable. All the weight factor representing bits is used to indicate whether the rule is added in the rule base or not. Any subset S of the N candidate fuzzy rules can be defined by a binary string of length N as $w = w_1, w_2, w_3, \dots, w_n$ where $w_i = 1$ and $w_i = 0$ meaning that the i th generated fuzzy rule is include and exclude in the rule base S , respectively.

In this study, independent rules with a good performance are not found for the proposed methodology, but, given a set of candidate rules, an optimal set of rules is searched by means of a combinatorial optimization algorithm.

B. Genetic Tuning Process to Obtain MFs Parameters

We employed a real and binary coded coupled genetic algorithm to generate the type-2 membership functions parameters and to build the best type-2 fuzzy logic controller.

B1. MFs coding

The center or the standard deviation is associated with uncertainties in Gaussian type-2 fuzzy sets. Gaussian type-2 fuzzy sets presents the MFs with uncertain standard deviation and uncertain center are shown in Fig.3, respectively. The mathematical function for the symmetrical gaussian MFs is given as follows:

$$\bar{\mu}(x) = \exp \left(-\frac{1}{2} \frac{(1-c)^2}{\sigma^2} \right)$$

Where center are c and σ is the width of the MFs and x is the input vector. In this study, we consider uncertain mean MFs where the center is defined by the range as follows: $c \in [c_1, c_2]$. Therefore, each type-2 fuzzy membership function is defined by 3 parameters as demonstrated in the following figure.

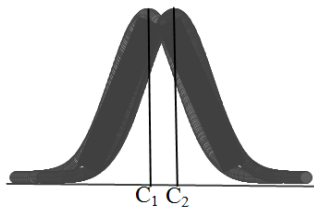


Fig. 3. T2 fuzzy set with uncertain center and certain width

In this case, for tuning uncertain mean C_1 , C_2 and certain width σ , a real coded genetic technique is employed. Therefore, each potential sub-chromosome consists of some real numbers and the sub-chromosome length is calculated

by the following equation: $n \times ((\sigma + C_1 + C_2) \times m)$, where n is the number of model variables with m linguistic terms. In the type-2 fuzzy set, a linguistic can be uniquely defined by its three parameters, C_1 , C_2 and σ as shown in Fig. 4. The Gaussian MFs has been used to define the type-2 fuzzy set in the fuzzy model. Here, because of simplicity gaussian shape is chosen; membership functions of any other form, for example, sigmoid or triangular, may be used as well. Here, l_m^i be the i th linguistic term of the input variable m . Fig.4 shows the coding of a fuzzy rule base system with two input variables and one output variable.

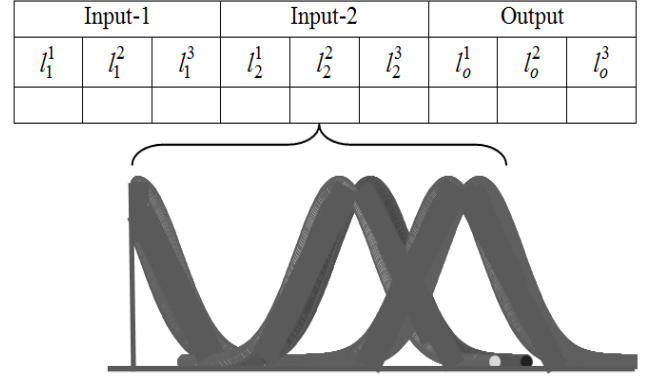


Fig.4. Type-2 fuzzy set encoding technique for input/output variable

IV. SIMULATION RESULTS AND DISCUSSIONS

To interpret the effectiveness of the proposed T2 fuzzy controller, the experimental results are presented.

A. Controllability theory: Double inverted pendulum

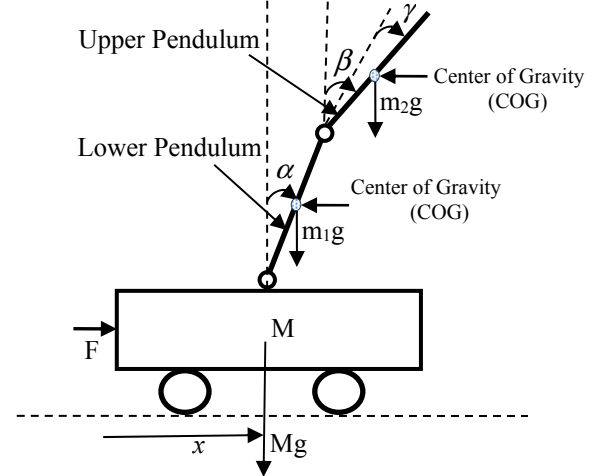


Fig.5. Double inverted pendulum system.

Fig.5 shows the two degrees of inverted pendulum system. The pendulum has a cart which is moving on the rail, a lower pendulum, a straight line rail, and an upper pendulum. x is the cart's position from the rail origin. If cart locates on the right side of the rail origin then cart's position is defined as positive. The lower pendulum's angle, α and the upper pendulum's angle, β both from upright position where clockwise direction is defined as the positive direction. The force defined by F and measuring unit is newton (N), is applying to horizontally to the cart and is consider as positive while moving the cart toward right direction.

There is no collision in the two degrees of pendulum system, the system equation of two degrees inverted pendulum system is acquired from Lagrange's equation of motion as follows:

$$\begin{aligned} a_{11}\ddot{x} + a_{12}\ddot{\alpha} + a_{13}\ddot{\beta} &= b_1 \\ a_{21}\ddot{x} + a_{22}\ddot{\alpha} + a_{23}\ddot{\beta} &= b_2 \\ a_{31}\ddot{x} + a_{32}\ddot{\alpha} + a_{33}\ddot{\beta} &= b_3 \end{aligned} \quad (1)$$

Where the coefficients are given by:

$$\begin{aligned} a_{11} &= M + m_1 + m_2 \\ a_{22} &= 4m_1l_1^2/3 + m_2L_1^2 \\ a_{33} &= 4m_2l_2^2/3 \\ a_{12} &= a_{21} = (m_1l_1 + m_2L_1)\cos\alpha \\ a_{13} &= a_{31} = m_2l_2\cos\beta \\ a_{23} &= a_{32} = m_2L_1l_2\cos(\alpha - \beta) \end{aligned} \quad (2)$$

and

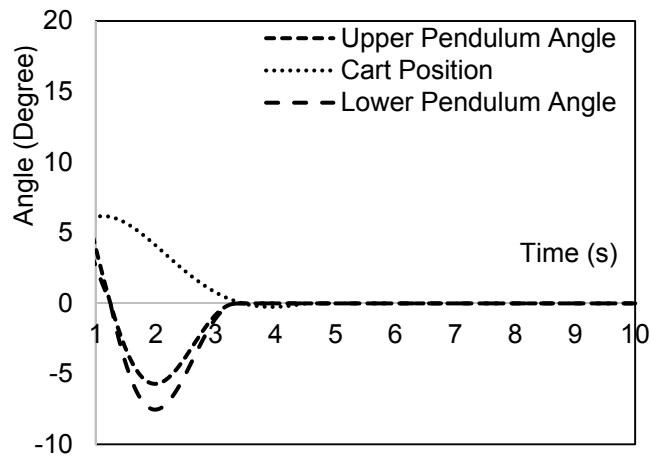
$$\begin{aligned} b_1 &= F + (m_1l_1 + m_2L_1)\dot{\alpha}^2\sin\alpha + m_2l_2\dot{\beta}^2\sin\beta \\ b_2 &= (m_1l_1 + m_2L_1)g\sin\alpha - m_2L_1l_2\dot{\beta}^2\sin(\alpha - \beta) \\ b_3 &= m_2l_2g\sin\beta + m_2L_1l_2\dot{\alpha}^2\sin(\alpha - \beta) \end{aligned} \quad (3)$$

The state variables are as follows: x = cart's position, $\dot{x}$ = cart's angular velocity, α = Lower pendulum angle, $\dot{\alpha}$ = lower pendulum's angular velocity of β = Upper pendulum angle, $\dot{\beta}$ = Upper pendulums' angular velocity.

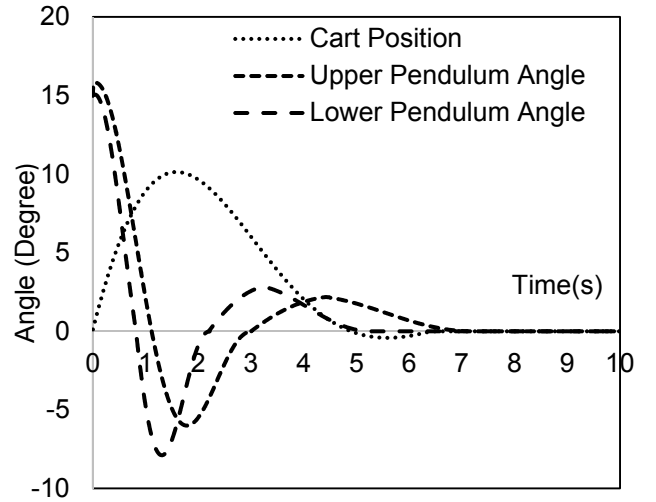
Here, the parameters M , m_1 , m_2 are the cart mass, the lower pendulum mass and upper pendulum mass respectively and their unit in [kg]. The gravity acceleration is $g=9.8$ (m/s²). Suppose for each pendulum uniform property exists on the distribution of the mass. The lower pendulum's full length and the lower pendulum's full length and the lower pendulum's length from the gravity center of to the pivot are denoted as L_1 , l_1 in the unit (m), and $L_1=2l_1$. For the upper pendulum, the length of gravity center to the linkage is denoted as l_2 in the unit (m). Unit radian is used for each angle and radian per second is used for the angular velocity of each angular velocity.

B. Stabilization Control Simulation

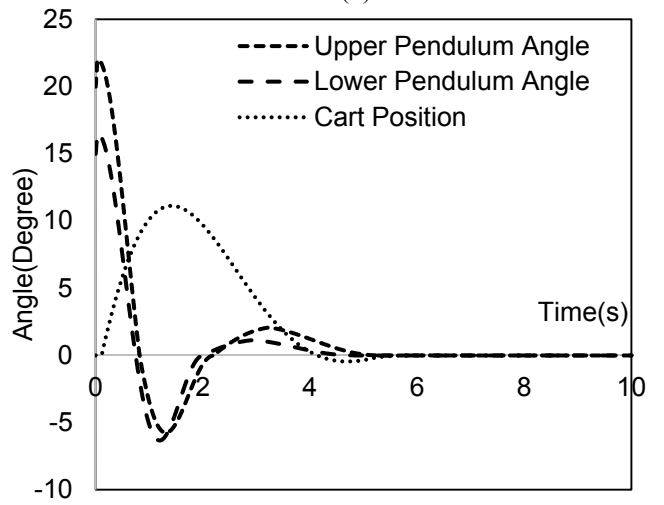
Several experiments have been done with different configuration of double inverted pendulum to prove the feasibility of the proposed stabilization fuzzy controller.



(a)



(b)

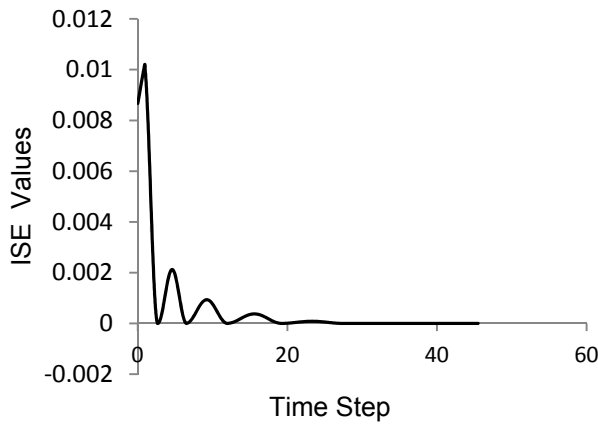


(c)

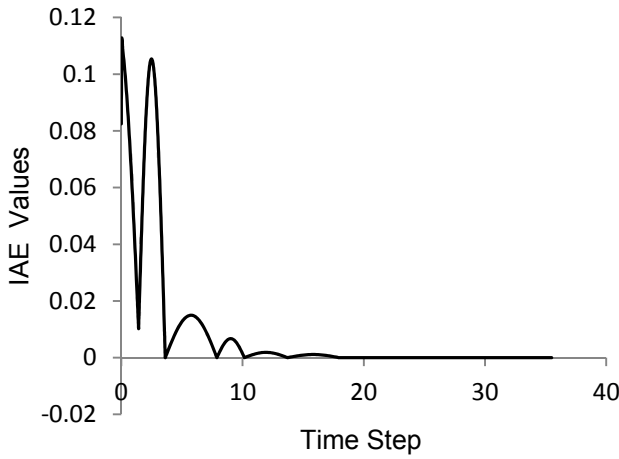
Fig. 6 shows the control results in order when the initial angles (a) $\alpha = 15.0$ and $\beta = 10.0$ (b) $\alpha = 15.0$ and $\beta = 15.0$ (c) $\alpha = 15.0$ and $\beta = 20.0$. In this study, the other state variables are reset to zero for the initial case.

Error rate is used for evaluating the proposed control algorithms. Three different performance criteria's such as rise time, the oscillation behavior and the behavior at the end of transition period are analyzed to quantify the errors. These three performance criteria's are as follows: Integral of Square Error ($ISE = \int_0^{t_f} [et]^2 dt$) Integral of the Absolute value of the Error ($IAE = \int_0^{t_f} |e(t)| dt$) and Integral of the Time multiplied by Square Error ($ITSE = \int_0^{t_f} t[et]^2 dt$).

Here $t_f = 20$ second subject to control input constraint and minimizing the integral by picking the controller parameters. Fig. 7 shows the oscillation at steady state, a rise time and adoption errors at the end of transition period that the pendulum by our controller.



(a)



(b)

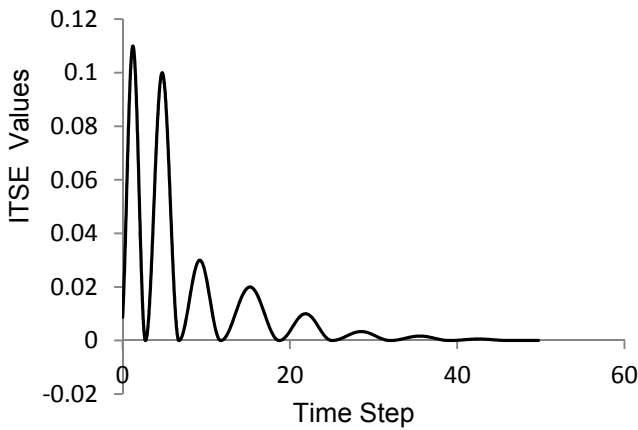


Fig. 7. Error values of the tuned controller while $x(0)$ is fixed to 0 and pendulum angle $\theta = 0.1$ rad. (a) Integral of Square Error (ISE), in this case steady state rise time of the controller response is very short. (b) Integral of the absolute Error (IAE): Before becoming stable the pendulum has less oscillation while using the generated controller. (c) Integral of the time multiplied by square error (ITSE): The adoption error at the end of transition period is less important for the generated fuzzy controller.

C. Generalization Ability of Proposed Controller

To validate the proposed controller generalization ability, varies the pendulum length and at the same time the other parameters remain fixed.

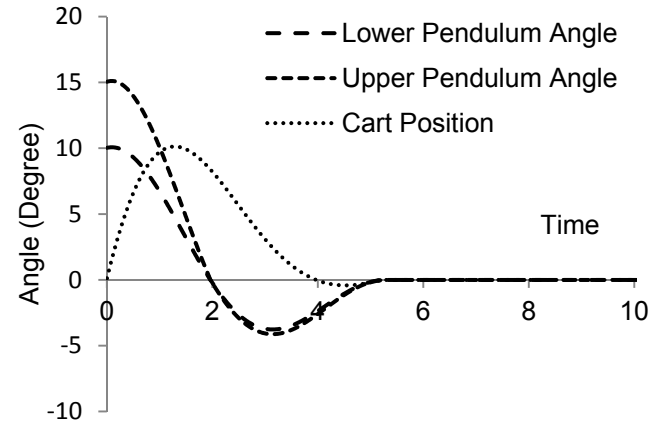


Fig. 8 shows for $m_1=0.1$ kg, $l_1=0.1$ m, $m_2=0.1$ kg, $l_2=0.1$ m, $M=1.0$ kg for the initial angles 15.00 and 10.00 of the two pendulums, the simulation results stabilizing a shorter series-type two degrees inverted pendulum system. The complete stabilization time is required about 6.18 and the maximum driving force is about 9.0N.

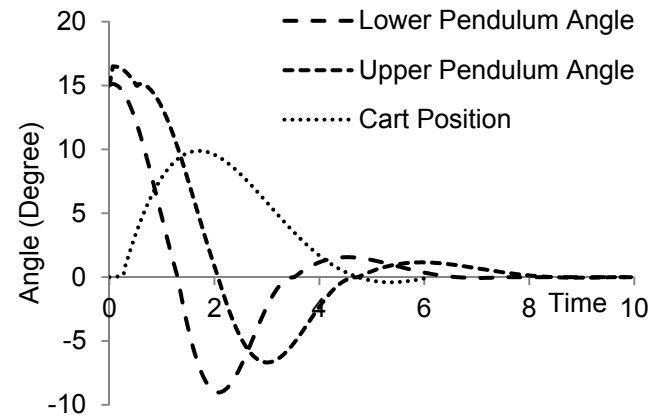


Fig. 9 shows for $m_1=0.1$ kg, $l_1=0.5$ m, $m_2=0.1$ kg, $l_2=0.5$ m, $M=1.0$ Kg, the stabilization control results of a longer series-type two degrees of freedom inverted pendulum system. The initial angles of upper and lower pendulums are set to 15.0°. Complete stabilization time is about 8.21 s and the maximum driving force is about 11.0 N without vibration.

The two pendulums stabilization is found in short time interval for a wide range of the initial angles when applying our proposed type-2 fuzzy controller, so high generalization ability is established for the controller.

V. CONCLUSION

In this article, a computationally-efficient systematic procedure was proposed for designing an Optimal Type-2 Fuzzy Inference System (OT2-FIS) with rule base size reductions using the hybrid genetic algorithm (HGAs). The proposed OT2-FIS was performed to control the double inverted pendulum system. The Integral Absolute Error (IAE), Integral Square Error (ISE) and Integral Time Square Error (ITSE) as well as the computational cost is considered as performance evaluation criteria. Finally, a proper validation came for the success and practicality of the proposed systematic design procedure for type-2 fuzzy inference system.

REFERENCES

- [1] F. J. Cabrerizo, J. Lopez-gijon and A. A. Ruiz, "A model based on fuzzy linguistic information to evaluate the quality of digital libraries," *International Journal of Information Technology & Decision Making*, vol. 9, no.3, pp. 455-472, 2010.

- [2] L. Ying and L. Yuanchun, "Neural-fuzzy Control of Truck Backer-Upper System using a Clustering Method," *Neurocomputing*, vol. 70, pp. 680-688, 2007.
- [3] F. Herrera and M. Lozano, "Adaptive Genetic Operators Based on Coevolution with Fuzzy Behaviors," *IEEE Trans. Evolutionary Computation*, vol. 5, no. 2, pp. 149-165, 2001.
- [4] F. J. Cabrerizo, S. Alonso, and E. Herrera Viedma, "A consensus model for group decision making problems with unbalanced fuzzy linguistic information," *Int. Journal of Information Tech. & Decision Making*, vol. 8, no. 1, pp. 109-131, 2009.
- [5] Z. S. Xu and H. Hu, "Projection models for intuitionistic fuzzy multiple attribute decision making," *Int. Journal of Information Tech. and Decision Making*, vol. 9, no. 2, pp. 267-280, 2010.
- [6] Y. Shi, "The research trend of information technology and decision making in 2009," *Int. Journal of Information Tech. and Decision Making*, vol. 9, no. 1, pp. 1-8, 2010.
- [7] Z. W. Gong, L. S. Li *et al.*, "On additive consistent properties of the intuitionistic fuzzy preference relation," *International Journal of Information Technology and Decision Making*, vol. 9, no. 6, pp. 1009-1025, 2010.
- [8] F. Herrera and M. Lozano, "Adaptive Genetic Operators Based on Coevolution with Fuzzy Behaviors," *IEEE Trans. Evolutionary Computation*, vol. 5, no. 2, pp. 149-165, 2001.
- [9] A. Li, Y. Shi, J. He and Y. Zhang, "A fuzzy linear programming based classification method," *Int. Journal of Information Tech. & Decision Making*, vol. 10, no. 6, pp. 1161-1174, 2011.
- [10] B. Kosko, *Neural Networks and Fuzzy Systems* (Englewood Cliffs: Prentice Hall Inc., 1992).
- [11] Mendel, J. M., *Rule-Based Fuzzy Logic Systems: Introduction and New Directions*. Prentice-Hall, Englewood Cliffs, NJ, 2001.
- [12] H. A. Hagras, "A hierarchical type-2 fuzzy logic control architecture for autonomous mobile robots," *IEEE Trans. Fuzzy Syst.*, vol. 12, pp. 524-539, 2004.
- [13] L.A. Zadeh "The concept of a linguistic variable and its application to approximate reasoning," *Information Science*, vol. 8, pp. 199-249, 1975.
- [14] J. M. Mendel, *Uncertain rule-based fuzzy logic systems: introduction and new directions*, Upper Saddle River, NJ: Prentice-Hall; 2001.
- [15] J. M. Mendel, R. I. John, and L. Feilong "Interval type-2 fuzzy logic systems made simple," *IEEE Trans Fuzzy Syst.*, vol. 14, pp. 808-21, 2006.
- [16] J. M. Mendel, *Rule-Based Fuzzy Logic Systems: Introduction and New Directions*, Prentice Hall, Englewood Cliffs, NJ.
- [17] Q. L. Liang and J. M. Mendel, "MPEG VBR video traffic modelling and classification using fuzzy technique," *IEEE Trans. on Fuzzy Systems*, vol. 9, no. 1, pp. 183-193, 2001.
- [18] T. Ozen, J. Garibaldi, and S. Musikasuwana, "Modelling the variation in human decision making," *In Proc. of FUZZ-IEEE*, vol. 2, 2004, pp. 617-622.
- [19] R. I. John, P. R. Innocent, and M. R. Barnes, "Neuro-fuzzy clustering of radiographic tibia images using type 2 fuzzy sets," *Information Science*, vol. 125, pp. 65-82, 2000.
- [20] H. W. Wu and J. M. Mendel, "Antecedent connector word models for interval type-2 fuzzy logic systems," *In Proc. of FUZZ-IEEE*, 2004, vol. 2, pp. 1099-1104.
- [21] J. Zeng and Z. Q. Liu, "Type-2 fuzzy hidden markov models to phoneme recognition," *In Proc. of the 17th Int. Conf. on Pattern Recognition*, 2004, vol. 1, pp. 192-195.
- [22] P. Melin and O. Castillo, "A new approach for quality control of sound speakers combining type-2 fuzzy logic and fractal theory," *In Proc. of the IEEE Int. Conf. on Fuzzy Systems*, vol. 2, pp. 825-830, 2002.
- [23] Q. Liang, N. Karnik and J. Mendel, "Connection Admission Control in ATM Networks using Survey-Based Type-2 Fuzzy Logic Systems," *IEEE Int. Conf. on SMC*, 2000, .
- [24] H. Roubos and M. Setnes, "Compact and transparent fuzzy models and classifiers through iterative complexity reduction," *IEEE Trans. Fuzzy Systems*, vol. 9, no. 4, pp. 516-524, 2001.
- [25] T. Sudkamp, A. Knapp and J. Knapp, "Model generation by domain refinement and rule reduction," *IEEE Trans. Systems Man Cybernet. B*, vol. 33, no. 1, pp. 45-55, 2003.
- [26] L. Zhang, X. Chang and H. R. Karimi, "Fuzzy Modeling and Control For a Class of Inverted Pendulum System," *Abstract and Applied Analysis*, vol. 2014, Article ID 936868, 6 pages, doi: 10.1155/2014/936868.
- [27] Alcalá, J. Alcalá-Fdez and F. Herrera, "A Proposal for the Genetic Lateral Tuning of Linguistic Fuzzy Systems and its Interaction with Rule Selection," *IEEE Trans. on Fuzzy Systems*, vol. 15, no. 4, pp. 616-635, 2007.
- [28] Zhen Zhang, Chonghui Guo, "Consistency and consensus models for group decision-making with uncertain 2-tuple linguistic preference relations". *International Journal of Systems Science* 47:11, pp. 2572-2587, 2016
- [29] Francisco Javier Cabrerizo, Francisco Chiclana, Rami Al-Hmouz, Ali Morfeq, Abdullah Saeed Balamash, Enrique Herrera-Viedma, Reza Langari, "Fuzzy decision making and consensus: Challenges". *Journal of Intelligent & Fuzzy Systems* 29:3, pp. 1109-1118, 2015

Analytical SISO Bit Error Rate Evaluation of a Satellite to Ground Link Under the Influence of Log-Normal Atmospheric Turbulence

Sharif,^{1,*} S. P. Majumder,²

¹EECE, MIST

Mipur-12, Dhaka-1216

²EEE, BUET

*¹shikdersharif@gmail.com

Abstract—Analysis is carried out for a satellite to ground link to evaluate the bit error rate performance limitation imposed by atmospheric turbulence. Considering log-normal distribution for the turbulence induced fading, the SNR and BER are determined. The average BER Performance results shows that the satellite to ground link at 30 GHz channel suffers significantly due to atmospheric turbulence. Power penalty due to turbulence induced fading are evaluated at a given BER for several atmospheric channel parameters and turbulence length.

Index Terms—Turbulence, Power penalty, satellite.

I. INTRODUCTION

Satellite communication is very much important for next generation communication systems and networks due to its high reliability and matured technology [1-2]. However, performance of satellite to ground links are very much deteriorated due to atmospheric turbulence effects specially at carrier frequency higher than 30 GHz. Significant research works are reported on the effect of atmosphere turbulence on broadband communications operating near 10 GHz [3]. Analytical performance of satellite to earth link are investigated considering the effect of atmospheric turbulence for various modulation formats [4]. Some simulation results are also available on the effect of turbulence on satellite to earth link [5]. Performance analysis with diversity in transmit and receive are also recently respected in presence of turbulence effects for satellite to earth and earth satellite link [6-7]. MIMO satellite to earth link are developed using Gaussian [8-9] and log-Normal [10-11] channel models for turbulence induced fading. Simulation results are recently respected on satellite to earth SISO and MIMO links user Log-Normal [12-13]. Besides, channel capacity [14] is more important in satellite communication. In this paper, we present an analytical approach to evaluate the bit error rate (BER) performance of a satellite to earth link at 30 GHz taking into account the effect of turbulence induced fading which is considered to have log-normal distribution results are evaluated numerically for several atmospheric turbulence condition and compared to simulation results respected easier. Power penalty due to turbulence induced fading of a given BER of 10^{-5} are also evaluated.

II. ANALYSIS

We consider a satellite to ground link as 30 GHz channel. The transmitted signal for satellite is given as

$$x(t) = \sqrt{2EIRP_{sat}} \cos[\omega_c t + \theta] \quad (1)$$

Where, $EIRP = P_t G_t$; P_t, G_t represent the transmitted power and antenna gain of satellite transponder, ω_c is the angular frequency of the downlink carrier, θ is the random phase of the carrier.

An overall 1×1 input-output relation between the transmit signal vector $\mathbf{x}(t)$ and the receive signal vector $\mathbf{y}(t)$ is formulated by(1).

$$y(t) = \int_{\tau} \mathbf{H}(t, \tau) x(t - \tau) d\tau + \mathbf{n}(t) \quad (2)$$

where, $\mathbf{H}(t, \tau)$ represents time-varying channel matrix, whose element $h_{i,j}(t, \tau)$ ($i = 1, 2, \dots, m_R$; $j = 1, 2, \dots, m_T$) denotes the time-varying channel parameter from the j^{th} transmitting antenna to the i^{th} receiving antenna, τ is excess delay, and $\mathbf{n}(t)$ is a noise vector, the element $n_i(t)$ is the noise into the receiver via the i^{th} receiving antenna, whose mean is 0 and variance is $\sigma_{n_i}^2(t)$. If the channel is frequency-flat fading and time-invariant, (1) is simplified as (2)

$$y(t) = \mathbf{H}\mathbf{x}(t) + \mathbf{n}(t) \quad (3)$$

In equation (2), $\mathbf{H}$ expresses channel matrix, whose element $h_{i,j}$ ($i = 1, 2, \dots, m_R$; $j = 1, 2, \dots, m_T$) represent the random channel parameter from the j^{th} the transmitting antenna to i^{th} receiving antenna. A frequency selective fading channel is represented as a combination of several frequency-flat channel is shifted in delay domain. For instance, $h_{i,j}$ at a given time t is a random value decided by Rician model with constant mean and variance. For a time-variant case, the model (2) can be applied by predicting the stochastic fading model and its statistical property parameters. By sampling the received signal as (2) at the symbol rate T_s , which means $y(k) = y(kT_s)$, the sampled received signal is written as

$$y(k) = \mathbf{H}\mathbf{x}(k) + \mathbf{n}(k) \quad (4)$$

In fact, a complete model for $h_{i,j}$ is also written as (4).

$$h_{i,j} = h_{1,i,j} h_{2,i,j} \quad (5)$$

where, $h_{1,i,j}$ represents path loss, which is a constant for a given specific link and specific propagation environment parameters. $h_{2,i,j}$ is a complex random variable, which indicates the randomly fading coefficient include by random multipath propagation of the incoherent scatterings in the troposphere.

The rapid fluctuations of the amplitude and phase of the received signal above 10 GHz are caused by the atmospheric turbulence in clear air.

The amplitude fading can be described by $\chi = \ln(A/A_0)$ in clear environment, where A and A_0 denote the instantaneous value and the mean of the receive signal amplitude in V. χ follows the Gaussian distribution with zero mean and variance σ_χ^2 which are presented as the following equation.

$$P(\chi) = \frac{1}{\sqrt{2\pi}\sigma_\chi} \exp\left[-\frac{\chi^2}{2\sigma_\chi^2}\right] \quad (6)$$

where, σ_χ^2 is the variance of χ . It is deduced as(6).

$$\sigma_\chi^2 = 0.307C_n^2 \left(\frac{2\pi}{\lambda}\right)^{7/6} L^{11/6} - 0.742C_n^2 \left(\frac{2\pi}{\lambda}\right)^{1/6} L^{11/6} L_0^{-2} \quad (7)$$

where, C_n^2 is the structure constant of refractive index in $m^{-2/3}$. λ is wavelength in m, L_0 is the average outer scale in m, L is the equivalent path length through the turbulence area in m, L for a satellite earth-link can be expressed in terms of turbulence mean height H_T in m and link elevation angle θ .

$$L = \frac{2H_T}{\sqrt{\sin^2\theta + \frac{2H_T}{8500000} + \sin\theta}} \quad (8)$$

The received signal at the earth station can be expressed as,

$$e_r(t) = A \cos(\omega_c t + \theta) \quad (9)$$

where, $A = A_0 e^\chi$ and $\chi = \ln(A/A_0)$ represents the amplitude fading and A, A_0 represent the instantaneous value and mean of the received signal amplitude. The SNR can be expressed as

$$SNR(\chi) = \frac{E_s/T_b}{P_n} = \frac{E_s R_b}{P_n} \quad (10)$$

where, $E_s = E_{s0} |\chi|^2$ The conditional Bit- Error Rate (BER) can be expressed as

$$BER(\chi) = \frac{1}{2} \operatorname{erfc}\left[\frac{\sqrt{SNR(\chi)}}{\sqrt{2}}\right] \quad (11)$$

The average BER can be evaluated by averaging the conditional BER are the pdf of χ .

$$BER = \int BER(\chi) P(\chi) d\chi \quad (12)$$

III. RESULT AND DISCUSSION

Following the theoretical analysis presented in section III, we evaluate the BER performance results of a satellite to ground link in presence of fading due to atmospheric turbulence. The parameters used are shown in table I.

TABLE I
SOME IMPORTANT PARAMETERS

Parameters Name	Value
Average outer scale of turbulence (L_0)	60 m
Refractive index (C_n)	$5-11 \times 10^{-12} m^{-2/3}$
Receiving frequency	30.0 GHz
Equivalent path length through turbulence area	70-110 m
Orbital altitude	1500 km

Fig.1, shows the pdf curve for different value of amplitude fading χ . Probability Density Function (PDF) describes the

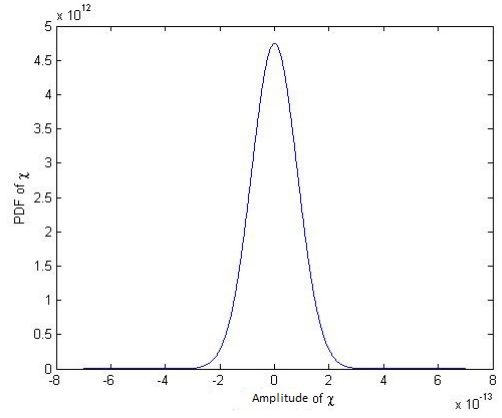


Fig. 1. PDF of different value of χ

approximate value that can be taken from a certain limit or random variable.

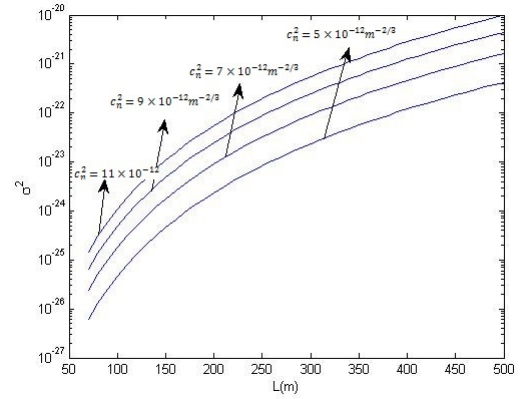


Fig. 2. Variance for different path length(L)

From Fig.2, we get the different variance curve for the various refracting index which depends on the turbulence length.

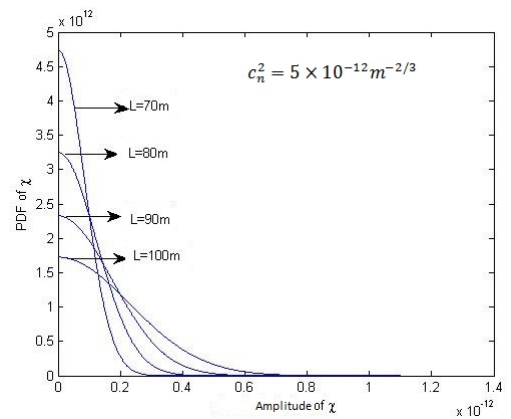


Fig. 3. Different PDF for different value of turbulence length

The Fig. 3 provides the information us that pdf curve is faded with the rise of turbulence length. It means that when the pdf peak maximum then it contain maximum density of data and bit error rate is minimum.

TABLE II
ACCORDING TO THE TURBULENCE LENGTH SNR AND BER AT 10^{-5}

BER at 10^{-5}		
Path length through turbulence area L(m)	SNR(dB)	Power penalty(dB)
70	48	0
80	55	7
90	59	11
100	65	17
110	70	22

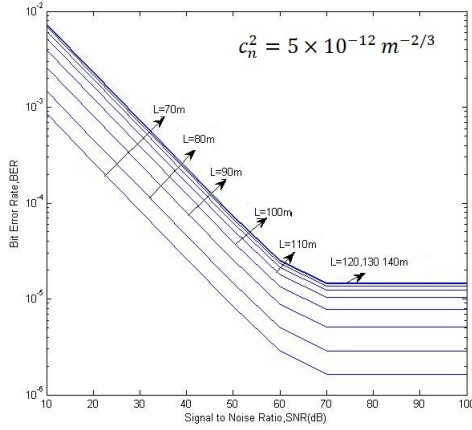


Fig. 4. BER for different value of turbulence length

Fig.4 shows the BER vs SNR curve, we get the acceptable result according to the turbulence length. It is noticeable that satellite link suffers significantly to the turbulence length which results in power penalty. The plots of power penalty is increased with the turbulence length but its become constant after overcoming the 110m which we can observe from. As a result penalty about 5dB for every 5m from the fig. 6.

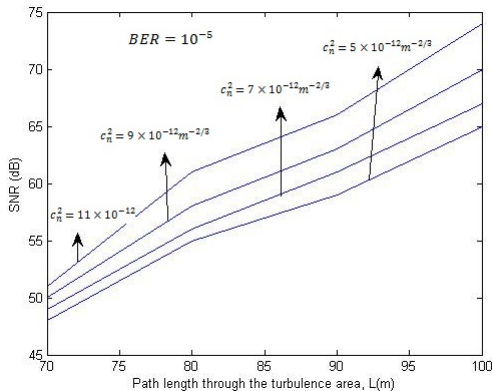


Fig. 5. SNR (dB) for difference turbulence length (m)

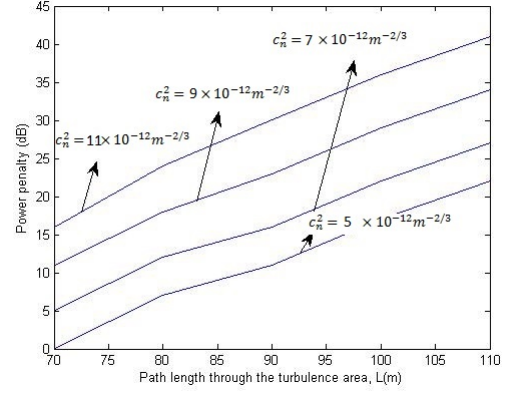


Fig. 6. Power penalty (dB) for difference turbulence (m)

Finally, after analysis we get the power penalty for various turbulence length and refractive index. It is 5dB in every 5m turbulence length.

IV. CONCLUSION

A theoretical approach is presented to evaluate the BER performance degradation of a satellite to ground link due to atmospheric turbulence. It is noticed that at 30 GHz channel, system suffers significant power penalty depending on the refractive index and turbulence length.

ACKNOWLEDGEMENT

The work is carried out as a part of MSc Engineering thesis in the department of EECE, MIST.

REFERENCES

- [1] D. Vanhoenacker; H. Vasseur; A. V. Vorst, "The Effect of Atmospheric Turbulence on Broadband Communication Channels Above 10 GHz", IEEE International Communication on Discovering a New World of Communication, vol. 2, pp. 1064-1068, 1992
- [2] S. Frank, Marzano, and D. A. Giovanni "Model-based prediction of amplitude scintillation variance due to clear-air tropospheric turbulence on Earth-satellite microwave links", IEEE Transactions on Antennas and Propagation, vol. 46, no. 10, pp. 1506-1518, 1998
- [3] Hanada, Tatsuyuki, Kiyotaka Fujisaki, and Mitsuo Tateiba. "Theoretical Analysis of Effects of Atmospheric Turbulence on Bit Error Rate for Satellite Communications in Ka-band", INTECH Open Access Publisher, 2011
- [4] K. Fujisaki, M. Tateiba, Hanada and Tatsuyuki, "Average Bit Error Rate for Satellite Uplink Communication in Ka-Band Under Atmospheric Turbulence Given by Gaussian Model", The 15th Asia-Pacific Conference on Communications, pp. 438-441, IEEE, 2009
- [5] W. liu; David G. Michelson, "Effect of Turbulence Layer Hight and Satellite Altitude On Tropospheric Scintillation on Ka-Band Earth-Band Earth-LEO Satellite Links ", IEEE Transactions on Vehicular Technology, vol. 59, no. 7, pp: 3181-3192, 2010
- [6] J. Shouda, W. Chang Yang and Wu "An Improved Atmospheric Turbulence Simulation Method", First International Conference on Pervasive Computing Signal Processing and Applications, pp. 1236-1239. IEEE, 2010
- [7] Anantrasirichi, Nantheera, Alin-Achim, Nicholas G. Kingsbury and David R. Bull, "Atmospheric Turbulence Mitigation Using Complex Wavelet-Based Fusion", IEEE Transaction on Image Processing, vol. 22, no. 6, pp. 2398-2408, 2013
- [8] Levant, Arie, "Universal Single-Input-Single-Output (SISO) Sliding-Mode Controllers with Finite-Time Convergence", IEEE Transactions on Automatic Control, vol. 46, no. 9, pp. 1447-1451, 2001
- [9] A. Piemontese, Colavolpe and Giulio, "Novel SISO Detection Algorithms for Nonlinear Satellite Channels", Global Telecommunications Conference, pp. 1-5, IEEE, 2011
- [10] L. Zhu, Zhang, Lei and Cheng Ju, "Generalized MIMO Channel Model and Its Capacity Analysis in Formation Flying Satellite Communication Systems", The 6th International ICST Conference on Communications and Networking in China, pp. 1079-1082, IEEE, 2011

- [11] S. Gong, D. Wei, X. Xue and M. Y. Chen, "*Study on the Channel Model and BER Performance of Single-Polarization Satellite-Earth MIMO Communication Systems at Ka Band*", IEEE Transaction on Antennas and Propagation, vol. 62, no. 10, pp. 5282-5297, 2013
- [12] B. M. Villaceros, P. Cerdeira, Lacoste, Frederic, Roberto and Joel Lemorton, "*SISO and MIMO Enhanced 2-State Modelling on the Land Mobile Satellite Channel for Various Frequencies Environments and Elevation Angles*", The 8th European Conference on Antennas and Propagation, pp. 2277-2281, IEEE, 2014
- [13] S. P. Winter, T. Robert, Schwarz and Andreas Knopp, "*MIMO application for reduced adjacent satellite interference in SATCOM downlinks*", International Conference on Communications (ICC), pp. 3570-3575. IEEE, 2014.
- [14] A. Hoffmann, Andreas Knopp, "*Impact of Atmosphere on The Signal Phase and the Channel Capacity in EHF MIMO Satellite Link*", IEEE Global Communication Conference, pp. 1-7, 2015

A Single Phase Input Switched SEPIC PFC AC-DC Converter

Kazi Khurshidi Haque Dia

Department of EEE
Ahsanullah University of Science and Technology
Dhaka, Bangladesh

kkh.dia@gmail.com

Ahammad

abdh.jf@gmail.com

Samia Islam

Department of EEE
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
mac@eee.buet.ac.bd

Mohammad Ali Choudhury

Abstract—A single phase new type AC-DC converter based on SEPIC topology has been proposed in this paper which has power factor correction and input current shaping facilities. The proposed rectifier is input switched and utilizes a bidirectional switch at the input. The rectifier gives buck-boost voltage gain without polarity reversal. The design has been made separately to operate the rectifier as a buck or a boost ac-dc input power factor corrected (PFC) converter with near sinusoidal input current having very low total harmonic distortion (THD). A simple feedback control circuitry has been used for simulated result generation to show that the proposed converter achieve the above at good conversion efficiency.

Keywords— Single-ended primary inductance converter (SEPIC); Total harmonic distortion (THD); Proportional Integral (PI) Controller; Power factor correction (PFC); Power factor (PF).

I. INTRODUCTION

Single phase AC-DC converters (rectifiers) are used in power supplies to provide dc to electronic equipment, computers, uninterruptible power supplies (UPS), battery charges, dc and inverter fed ac drives etc. Uncontrolled single phase diode bridge rectifiers with output capacitive filter draw non-sinusoidal currents which has high THD and causes low input power factor. Both are detrimental to utility power supply. Power factor and input current THD of diode bridge rectifiers can be improved by passive and active methods. The boost PFC circuit is the common active methods of power factor correction (PFC). Buck, Buck-boost, Ćuk and SEPIC topologies have also been used for the same purpose [1-16]. In an active PFC technique either the output DC or the input AC is switched at high frequency to make the input current switched sine wave in phase with the input sinusoidal voltage [10, 13, 17-18]. Feedback control of PFC circuit with feedbacks from output and input makes a rectifier power factor corrected and output voltage regulated with small input current THD at good conversion efficiency [1, 2].

This paper reports an input ac switch based single phase SEPIC ac-dc converter topology. The converter has two separate SEPIC ac to dc topologies each working for positive and negative supply cycles through a combination of inductor and capacitor keeping the input current smooth and in-phase with the supply voltage. The SEPIC topology of the converter has advantages of input boost topology stage and overall boost-buck gain [1-3]. The input boost stage allow the same controller philosophy be used as used in conventional boost pfc rectifiers.

II. THE PRINCIPLE OF OPERATION OF THE PROPOSED 1-PHASE INPUT SWITCHED SEPIC RECTIFIER

The proposed input switched SEPIC AC-DC converter configuration shown in Figure 1. It has four states as shown from Figures 2 - 5 where, when switch S is turned ON and OFF continuously during positive and negative ac supply cycles. As shown in Figures 2 and 3, during positive half cycle of the input, when S is ON, the input inductor L1 gets charged and the capacitor C1 discharges and energizes the inductor L2. When S is OFF, the inductor L1 and L2 transfer energy to the upper capacitor Co1 and load Ro. Figures 4 and 5 represents the operation of the circuit during the negative half cycle of the ac supply as switch turns ON and OFF respectively. When S is ON during negative half cycle, the input inductor L1 is energized from opposite direction and C2 discharges to energize L3. When S is OFF, L1 and L3 transfers energy to Co2 and the load. The voltage of the load (Ro in the circuit) is the combined voltage of the capacitors Co1 and Co2.

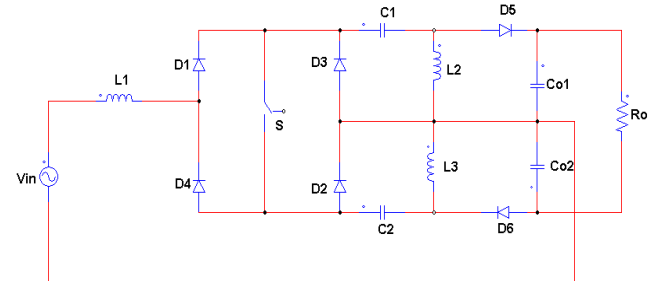


Figure 1: Proposed SEPIC AC-DC Converter.

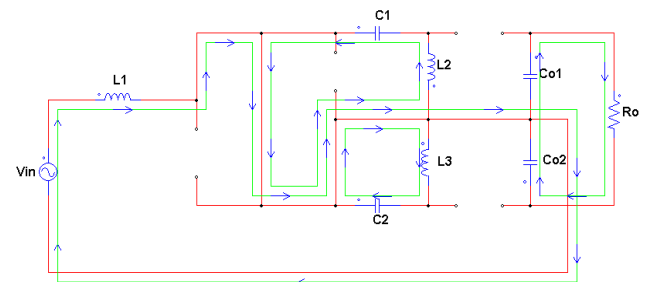


Figure 2: Equivalent circuit of the circuit of Figure 1 for positive cycle switch ON

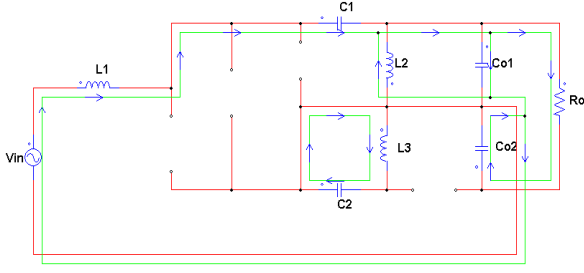


Figure 3: Equivalent circuit of the circuit of Figure 1 for positive cycle switch OFF

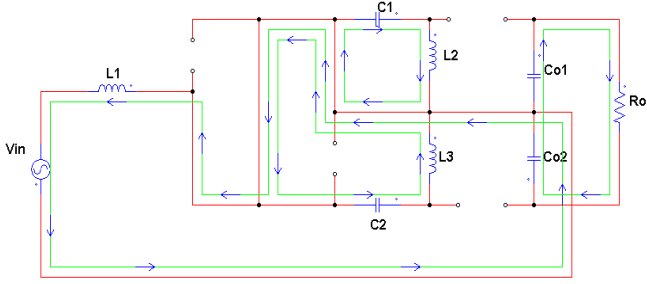


Figure 4: Equivalent circuit of the circuit of Figure 1 for negative cycle switch ON

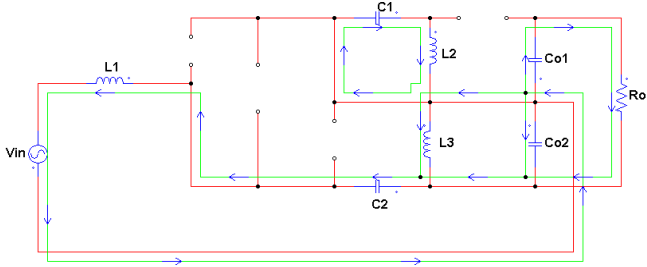


Figure 5: Equivalent circuit of the circuit of Figure 1 for negative cycle switch OFF

III. THE 1-PHASE OUTPUT SWITCHED SEPIC RECTIFIER

The circuit configuration of a conventional output switched SEPIC rectifier is shown in Figure 6. In the circuit input AC is converted to DC by a bridge rectifier and then this DC voltage is fed to a SEPIC DC-DC Converter. To understand the difference of operation and the performance of the proposed configuration over the conventional one, principle of operation of the conventional circuit is briefly described. The working principle of the circuit has two states, during positive and negative half cycle when switch S is ON and OFF. During positive and negative half cycle of AC input, when S is ON, the input inductor is energized and the capacitor C discharges to energize L2. When S is OFF the inductors L1 and L2 deliver energy to the load Ro. The single-ended primary inductance converter (SEPIC) can produce an output voltage that is either greater or less than the input but with no polarity reversal.

The relationship between input and output voltage of the DC-DC SEPIC converter is,

$$V_o = \frac{D}{1-D} V_{in} \quad (1)$$

$$I_o = \frac{1-D}{D} I_{in} \quad (2)$$

Where, V_o , I_o and V_{in} , I_{in} are output voltage, current and input voltage, current respectively and D is the duty cycle of the gate pulses.

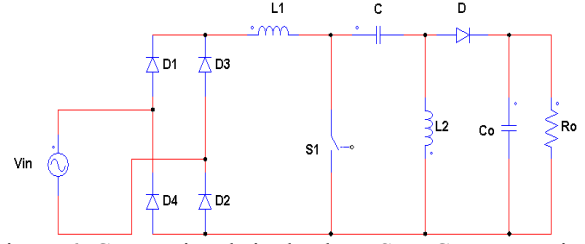


Figure 6: Conventional single phase SEPIC output switched AC-DC converter

III. PFC FEEDBACK CONTROLLER

In the open loop control the power factor and efficiency of the proposed circuit are reasonably good for a wide range of duty cycle variation. However, the input current total harmonic distortions are not in acceptable range for proposed and conventional circuit. It is therefore necessary that appropriate feedbacks are adopted in the circuit for all performance parameters to be in the acceptable range and is also necessary for output voltage regulation around a constant value.

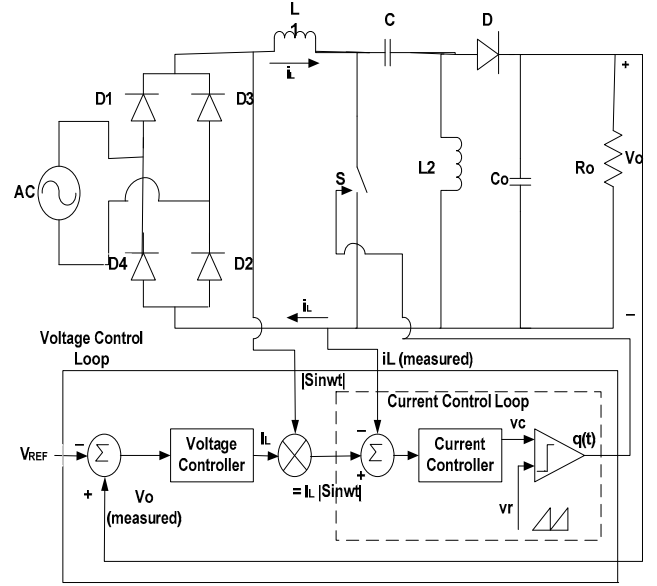


Figure 7: PFC feedback controller circuit of conventional output switched single phase SEPIC Rectifier

In controlling a PFC rectifier, the main objective is to draw a sinusoidal current, in-phase with the utility voltage. The reference inductor current is of the full wave rectified form. The requirements on the form and the amplitude of the inductor current lead to two control loops: voltage control loop and inner current control loop [1]. The transfer function of the current controller needed to be used as follows:

$$G_i(s) = k_{pc} \frac{s + \frac{k_{ic}}{k_{pc}}}{s} \quad (3)$$

where, k_{ic} is the integrator gain and k_{pc} is the proportional gain of the current controller.

The following transfer function is used for the voltage controller as below:

$$G_v(s) = k_{pv} \frac{s + \frac{k_{iv}}{k_{pv}}}{s} \quad (4)$$

where, k_{iv} is the integrator gain and k_{pv} is the proportional gain of the voltage controller.

The two control loops can be designed separately and can be realized by operational amplifier circuits working as PI controllers, each for two different loops. The inner current controller loop is faster than the outer voltage controller loop.

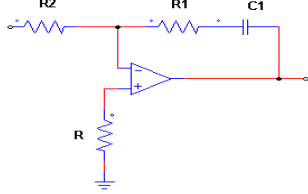


Figure 8: PI controller

The PI current controller parameters are:

$R_1 = 220\text{k}\Omega$ $R_2 = 100\Omega$ $R = 10\text{k}\Omega$ $C_1 = 0.022\mu\text{F}$

The PI voltage controller parameters are:

$R_1 = 10\text{k}\Omega$ $R_2 = 15\text{k}\Omega$ $R = 10\text{k}\Omega$ $C_1 = 0.1\mu\text{F}$

Switching frequency of the Pulse Width Modulator generator is taken 10KHz. With properly designed PFC feedback controller, increase/decrease of the reference voltage (V_{REF}) of the feedback controller can increase/decrease the output voltage of the power stage respectively.

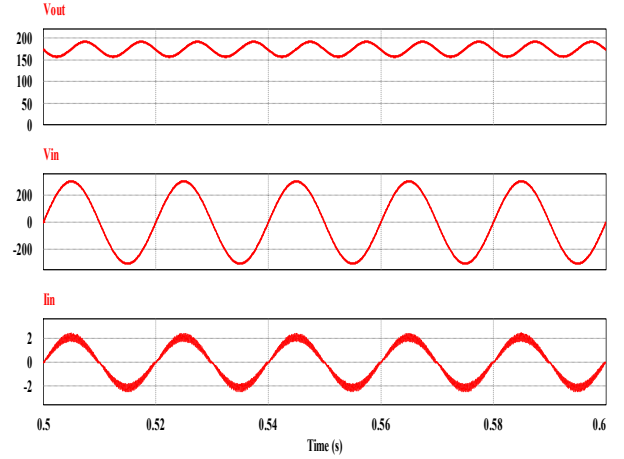
IV. SIMULATION AND RESULTS

To attain high conversion efficiency, low input current THD and high input power factor under both buck and boost operation mode of the proposed and conventional circuit, is not possible keeping the circuit parameters same for both mode of operation. So in this paper circuit design has been made separately for buck and boost mode. The circuit parameters used to investigate and evaluate the performance of the proposed and conventional topology is given in Table I. All the simulated results are obtained for proposed and conventional circuits using same PFC feedback controller.

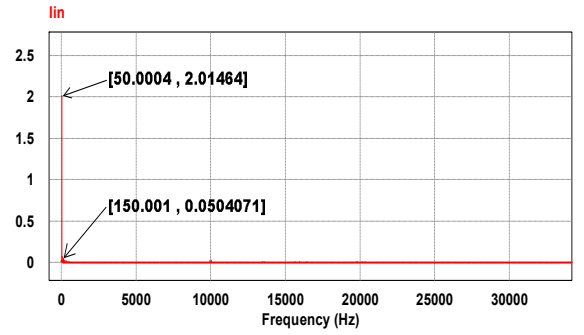
TABLE I: CIRCUIT PARAMETER SPECIFICATIONS OF PROPOSED AND CONVENTIONAL TOPOLOGY

Parameter name	Parameter value	Used to operate as
L1	10 mH	Buck converter
L2 = L3	12 mH	
L1	6 mH	Boost converter
L2 = L3	5 mH	
Co1, Co2	330 μF	Both buck and boost converter
Ro	100 Ω	
C1 = C2	1.5 μF	

Simulated results of the typical input-output voltages, input current and spectrum of input current waveform of proposed circuit in Figure 10 operated in buck and boost region are shown in Figures 9 and 10 respectively.



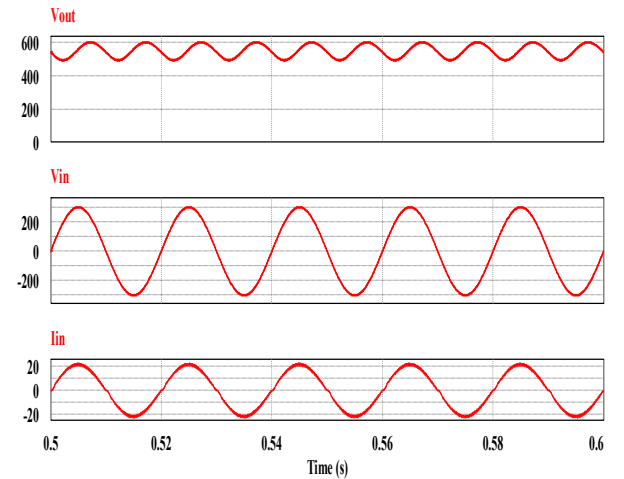
(a)



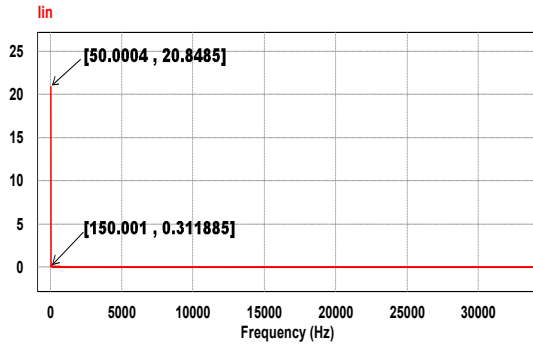
(b)

Figure 9: Input switched single phase SEPIC rectifier operating in buck mode

(a) Output voltage (Upper picture), input voltage (middle picture) and input current (Lower picture), (b) Spectrum of input current in (a)



(a)



(b)

Figure 10: Input switched single phase SEPIC rectifier operating in boost mode.

(a) Output voltage (Upper picture), input voltage (middle picture) and input current (Lower picture), (b) Spectrum of input current in (a)

In the typical waveforms of Figures 9 and 10 of the proposed single phase SEPIC rectifier, it is clear that the AC-DC converter gives buck/boost voltage gain. The output voltages of Figures 9 and 10 are 180V DC and 550V DC for an input voltage of 212.12 V R.M.S. value and of 50 Hz frequency. The proposed circuit gives near unity power factor and the spectrum of the input currents indicate that the harmonic components are lower than the fundamental frequency component producing very low input current THD when operating at in buck-boost region of operation.

Dynamic response of the proposed circuit to regulate the output voltage and input current with load variation is shown in Figures 11 and 12 for buck and boost operation respectively. $\pm 20\%$ variation of 100 Ω load is considered for simulation that is from 100 Ω to 120 Ω to 80 Ω to 100 Ω to 120 Ω to 80 Ω at 0.50, 1, 1.50, 2.0 and 2.50 seconds. The proposed circuit of Figure 1 gives well regulated output voltage of 180V DC (Figure 11-upper picture) and 550V DC (Figure 12-upper picture).

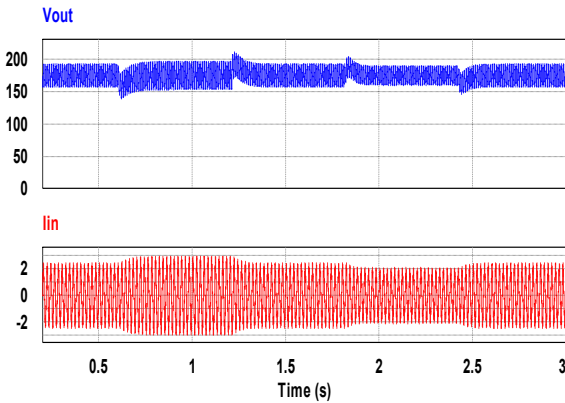


Figure 11: Dynamic response of output voltage (upper picture) and input current (lower picture) with load variation when the proposed circuit is under buck operation

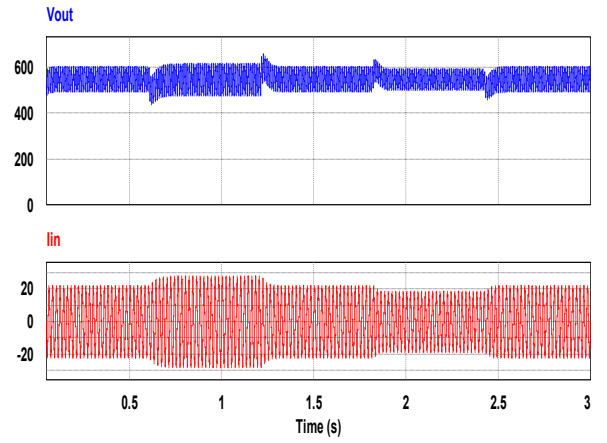


Figure 12: Dynamic response of output voltage (upper picture) and input current (lower picture) with load variation when the proposed circuit is under boost operation

With an increase upto 500V V_p and decrease upto 260V V_p of input voltage with respect to the 300V V_p (212.12V_{rms}) the output voltage of the proposed converter remains constant around 180V DC and 550V DC which is shown in Figures 13 and 14 respectively.

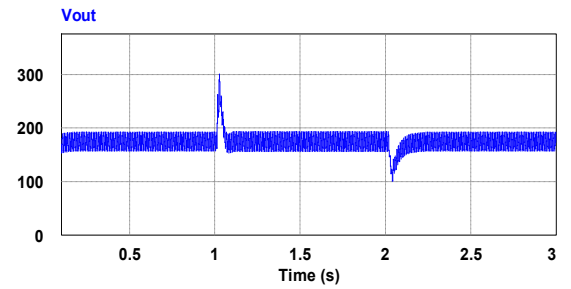


Figure 13: Dynamic response of the output voltage with input voltage variation of the proposed circuit under buck operation

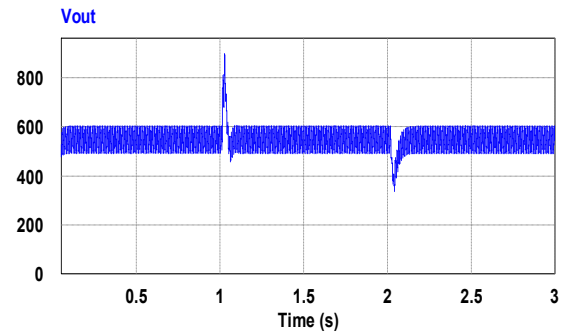


Figure 14: Dynamic response of the output voltage with input voltage variation of the proposed circuit under boost operation

Table II and III and Figures 15 to 20 show the performance comparison of proposed configuration of input switched single phase SEPIC topology based rectifier with conventional output switched SEPIC rectifier in terms of efficiency, input current

THD and input power factor with reference voltage variation V_{REF} of 0.15V (corresponds to 70V o/p DC of converter) to 0.42V (Corresponds to 193.2V o/p DC of converter) and V_{REF} of 0.5V (corresponds to 245.3V o/p DC of converter) to 1.5V (corresponds to 670V o/p DC of converter) of the feedback PFC controller for buck and boost region operation respectively.

TABLE II: COMPARISON OF CONVENTIONAL AND PROPOSED TOPOLOGIES OF 1-PHASE AC-DC SEPIC CONVERTER OPERATING IN BUCK MODE

V_{REF} (V)	EFFICIENCY (%)		INPUT CURRENT THD (%)		INPUT POWER FACTOR	
	Propo- sed Conf.	Conven- tional Conf.	Propo- sed Conf.	Conven- tional Conf.	Propo- sed Conf.	Conven- tional Conf.
0.15	97.65	91.51	38.2	70.61	0.9322	0.819
0.2	97.94	92.63	23.43	48.39	0.9733	0.899
0.25	97.73	92.90	16.31	34.11	0.9869	0.945
0.3	97.59	92.96	11.48	24.85	0.9932	0.968
0.35	97.40	93.17	8.92	20.22	0.9957	0.977
0.4	97.19	93.27	7.13	16.91	0.9968	0.981
0.42	97.14	93.26	6.59	15.68	0.9969	0.983

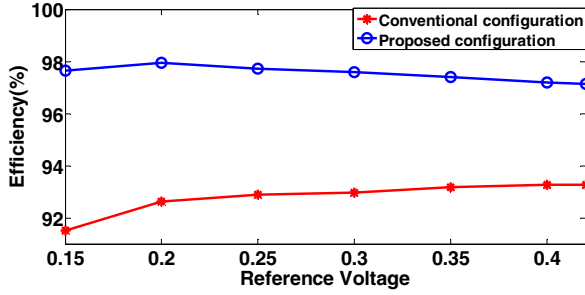


Figure 15: Comparison of efficiency between conventional and proposed topologies of single phase AC-DC SEPIC converter in buck mode.

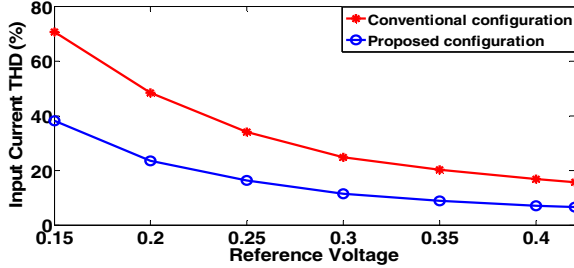


Figure 16: Comparison of input current THD between conventional and proposed topologies of single phase AC-DC SEPIC converter operating in buck mode.

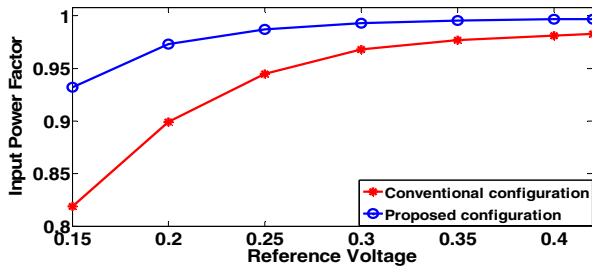


Figure 17: Comparison of input P.F. between conventional and proposed topologies of single phase AC-DC converter operating in buck mode.

TABLE II: COMPARISON BETWEEN CONVENTIONAL AND PROPOSED 1-PHASE AC-DC SEPIC CONVERTERS OPERATING IN BOOST MODE

V_{REF} (V)	EFFICIENCY (%)		INPUT CURRENT THD (%)		INPUT POWER FACTOR	
	Propo- sed Conf.	Conven- tional Conf.	Propo- sed Conf.	Conven- tional Conf.	Propo- sed Conf.	Conven- tional Conf.
0.5	94.80	92.51	8.16	19.82	0.9951	0.981
0.7	94.48	92.61	5.04	12.58	0.9964	0.992
0.9	94.04	92.14	3.75	9.23	0.9957	0.995
1.1	93.49	91.45	3.44	7.45	0.9942	0.997
1.3	92.83	90.58	3.93	6.58	0.9922	0.997
1.5	92.18	90.76	5.09	6.65	0.9897	0.997

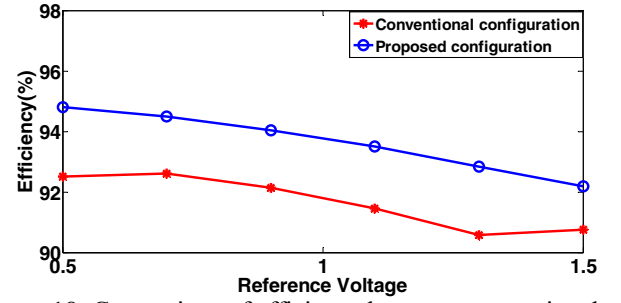


Figure 18: Comparison of efficiency between conventional and proposed topologies of single phase AC-DC SEPIC converter operating in boost mode.

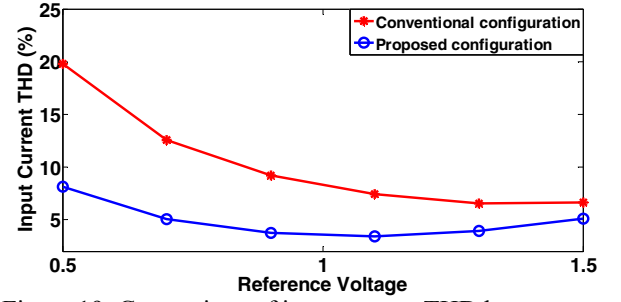


Figure 19: Comparison of input current THD between conventional and proposed topologies of single phase AC-DC SEPIC converter operating in boost mode.

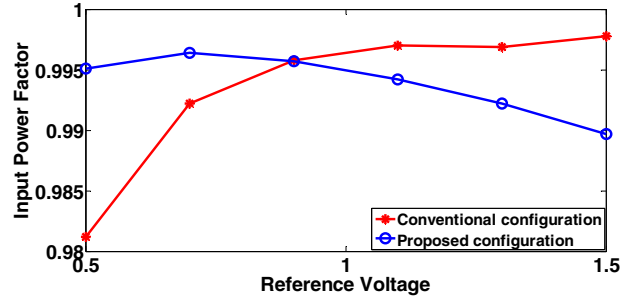


Figure 20: Comparison of input P.F. between conventional and proposed topologies of single phase AC-DC converter operating in boost mode.

Table II and Figures 15 to 17 demonstrate the simulation results of the proposed circuit and conventional circuit under buck operation. The efficiency of the proposed circuit is above 97 percent and gives greater efficiency than the conventional output switched SEPIC rectifier for a wide range of buck operation controlled by the PFC feedback controller with V_{REF}

from .15V to .42V. The proposed circuit is able to give very low input current THD as low as below 10% from .35V to .42V and gives lower THD throughout the buck range of operation whereas using the same feedback controller and same circuit parameter the conventional circuit gives very high input current THD which is not in acceptable range. In terms of input power factor the proposed circuit gives very high p.f. above 0.93 and gives improved p.f. than the conventional circuit for buck mode of operation.

Table III and Figures 18 to 20 represent the simulation results of the proposed circuit and conventional circuit under boost operation. The efficiency of the proposed circuit is above 92 percent and gives greater efficiency than the conventional output switched SEPIC rectifier for a wide range of boost operation controlled by the PFC feedback controller with V_{REF} from .5V to 1.5V. The proposed circuit is able to give very low input current THD as low as below 10% for entire boost range whereas, using the same feedback controller and same circuit parameter the conventional circuit gives high input current THD than the proposed circuit. The input power factor of the proposed circuit is very high giving p.f. above 0.99.

V. CONCLUSION

The proposed input switched new single phase SEPIC PFC rectifier gives buck-boost voltage gain capability, very high conversion efficiency above 92%, high power factor above 0.93 and very low input current THD below 10% under buck and boost region of operation than the conventional single phase output switched SEPIC rectifier.

ACKNOWLEDGMENT

This research work is supported by Bangladesh University of Engineering and Technology (BUET) and Ahsanullah University of Science and Technology (AUST). The authors acknowledge the facilities provided by BUET and AUST.

REFERENCES

- [1] N. Mohan, Power Electronics. John Wiley and Sons, Inc, USA, 2012.
- [2] D. W. Hart, Power Electronics. McGraw-Hill Companies, Inc, New York, 2011.
- [3] O.Garcia, J. A.Cobos, "Single Phase Power Factor Correction: A Survey." IEEE Transactions on Power Electronics, VOL. 18, NO. 3, pp. 749-755, May 2003.
- [4] A. De Bustiani Lange,T. B. Soeiro,M. Silveria Ortmann,M. Lobo Heldwein, " Three Level Single Phase Bridgeless PFC Rectifiers.", IEEE Transactions on Power Electronics, VOL. 30, issue-6, pp. 2935-2949, June 2015.
- [5] L. Huber, Jang Yungtaek, M. Jonanovic Milan, "Performance Evaluation of Bridgeless PFC Boost Rectifiers." IEEE Transactions on Power Electronics, VOL. 23, NO. 3, May 2008.
- [6] I. Siyana,V. A. Manjusha, "Single Phase Single Stage High Power Factor Ćuk Rectifier.", International Journal of Advanced Research in Electrical Electronics and Instrumentation Engineering, VOL. 3, issue-9, pp. 2320-3765, September, 2014.
- [7] F. Musavi, "A High-Performance Single Phase Bridgeless Interleaved PFC Converter for Plug-in Hybrid Electric Vehicle Battery Chargers", IEEE Transactions on Industrial Applications, VOL. 47, NO. 4, pp. 1833-1843, July/August. 2011.
- [8] E. H. Ismail, "Bridgeless SEPIC Rectifier With Unity Power Factor and Reduced Conduction Losses", IEEE Transactions on Industrial Electronics, VOL. 56, NO. 4, pp. 1147-1157, April. 2009.

- [9] R. Balamurugan, Dr. G. Gurusamy, " Harmonic Optimization for Single Phase Improved Power Quality AC-DC Power Factor Corrected Converters" International Journal of Computer Applications, VOL. 1, NO. 5, pp. 46-53, 2010.
- [10] B. Singh,B. N. Singh, A. Chandra,K. Al-Haddad,A. Pandey, D. P. Kothari,"A Review of Single-Phase Improved Power Quality AC-DC Converter", IEEE Transaction on Industrial Electronics, VOL. 50, NO. 5, pp. 962-981, October, 2003.
- [11] Y. Dagur, Dr. M. K. Gupta, "Single Phase AC-DC Converter employing Power Factor Correction with High Frequency Isolation Using Buck-Boost PWM Converter." International Journal of Innovative Research in Science, Engineering and Technology, VOL. 4, issue – 4, pp. 2607-2617, April 2015.
- [12] R. Ramesh, U. Subathra, M. Ananthi, " Single phase AC-DC power factor corrected converter with high frequency isolation using buck converter." International Journal of Engineering Research and Applications, VOL.4, issue-3, pp.79-82, March 2014.
- [13] R.V. Krishnendu, P. R. Sunil Kumar, "Bridgeless CUK Rectifier with Output Voltage Regulation using Fuzzy controller." ISOR Journal of Electrical and Electronic Engineering, VOL. 8, issue-1, pp. 93-100, Nov.-Dec., 2013.
- [14] K. Umamaheswari, V. Venkatachalam, "Single Phase Converters for Power Factor Correction With Tight Output Voltage Regulation", International Journal of Emerging Technology and Advanced Engineering, VOL. 3, issue-2, pp. 516-521, February 2013.
- [15] S. Singh, B. Singh, "Single-Phase SEPIC Based PFC Converter for PMBLDCM Drive in Air-Conditioning System", Asian Power Electronics Journal, VOL. 4, NO. 1, pp. 16-21, April 2010.
- [16] Z. Ye,F. Greenfeld,Z. Liang, "Single-Stage Offline SEPIC Converter with Power Factor Correction to Drive High Brightness LEDs", Applied Power Electronics Conference and Exposition, Twenty Fourth Annual IEEE, pp.546-553.
- [17] M. M. S. Khan, M. S. Arifin, M. R. T. Hossain, M. A. Kabir, A. H. Abedin, M. A. Choudhury, "Input Switched Single Phase Buck and Buck Boost AC-DC Converter with Improved Power Quality". ICECE, Dhaka, Bangladesh, pp. 189-192, December, 2012.
- [18] A. H. Abedin, Rahaman, M. H. Khan, M. M. S. Choudhury, M. A. Rubaiyat, M. Hossain, T. Uddin, M. N. PEDES, Bengaluru, India, pp. 1-5, December 2012.

Numerical Modeling and Analysis of Ultra Thin Film Cu(In,Ga)Se₂ Solar Cell using SCAPS-1D

Mohammad Sijanur Rahaman Robin¹, Mohamed Mansoor Mohamed Rasmi¹

MD. Shakowat Zaman Sarker², A. S. M. Rabbi Al Mamun¹,

Electrical and Electronic Engineering, International Islamic University Chittagong¹, Chittagong-4314, Bangladesh

Electrical and Electronic Engineering, Uttara University², Dhaka-1230, Bangladesh

Email: Sijan.robin24@gmail.com

Abstract—Numerical analysis of ultra thin film Cu(In,Ga)Se₂ (CIGS) solar cell has been performed using SCAPS-1D simulator. In this paper, the performances of CIGS solar cell has been observed by varying absorber layer, buffer layer and window layer thickness. Also the effect of temperature has been carried out. The aim of this work is optimization of CIGS solar cell thickness as well as the efficiency. Results represent that solar cell performances have enhanced with rise in thickness of CIGS absorber layer. Meanwhile opposite scenario is found with the rise in buffer layer and window layer thickness. The highest cell efficiency achieved is 22.67% with 0.7862 V open circuit voltage (Voc), 34.130 mA/cm² short circuit current density (Jsc) and 84.49% Fill Factor (FF) on 2000 nm CIGS absorber layer.

Keywords—CIGS; SCAPS-1D; thickness; efficiency.

I. INTRODUCTION

Thin-film solar cells (TFSCs) are the latest development in the field of photovoltaic cells, and there is considerable hope for wide spread use in the production of electricity. It is a microstructure created by the individual layers of a material on the surface of a solid material. TFSCs are more preferable due to its low manufacturing cost, flexibility, reliability, availability and being lightweight. Amorphous-Si, Cadmium Telluride (CdTe), Copper Indium Gallium Di-Selenide (CIGS) are the most promising TFSCs. Among them CIGS based solar cells attracted a lot of interest as a marketable PV technology [1] and now leading the thin film market due to its high efficiency. Most attractive feature of this solar cell is adjustable band gap [2] and it has high absorption coefficient (10⁵cm⁻¹) [3]. In spite of its high efficiency, the loss mechanism and inaccessibility of some materials makes it a matter of concern. With the increasing possibility of TFSCs have to keep in mind that materials used for CIGS solar cell should be lessen in a way that it doesn't affect the efficiency. Though it has some limits, it has non-toxic absorber in contrast with CdTe absorber.

The highest efficiency in laboratory is 22.3% achieved by “Solar frontier” [4] (also confirmed by NREL) and 21.7% by “Solar Energy and Hydrogen Research (ZSW)” [5] whereas crystalline Si efficiency is 25% [6]. CIGS solar have caught attention of researchers due its high efficiency. Although it has reached the maximum efficiency, still its parameter needs to

be optimized to get better efficiency in industrial production. Mostefaoui et al. showed in their comparative buffer layer study that optimal thickness of CIGS is 2500 nm. Whereas other reports depicted that the thickness of CIGS absorber layer would be optimal at around 3000 nm [8][9]. Besides, Sharmin et al. found optimum value of the CIGS absorber layer at 2000 nm [10]. Due to the scarcity of indium, its thickness needs to be thinner in order to reduce the material cost [11]. However, maximum numerical study based only on optimization of absorber and buffer layer. Window layer optimization needs to be taken into consideration for further study.

The purpose of this study is to observe performance of CIGS solar cells and optimize its results with best possible way and minimize different layer thickness to reduce its manufacturing cost. In this study the effect of the absorber and buffer layer thickness as well as window layer thickness and temperature effect on the electrical output parameters of CIGS solar cell has been carried out.

II. METHODOLOGY

A. CIGS solar cell structure

The structure of the CIGS solar cell is as follows ZnO:Al/i-ZnO/CdS/CIGS/Mo/substrate. Fig. 1 illustrates the schematic structure of simulated CIGS solar cell.

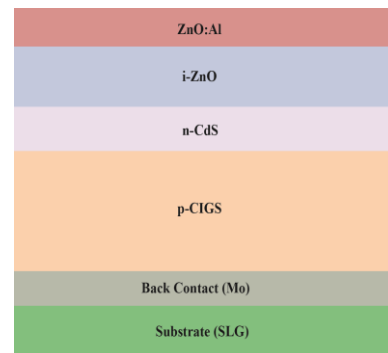


Fig. 1. Model of simulated CIGS solar cell.

In this solar cell, Transparent Conductive Oxide (TCO) which is a bi-layer of ZnO having a bandgap of 3.3 eV is used as window layer. The first layer is intrinsic Zinc-Oxide (i-ZnO) and the second layer is Aluminum (Al) doped ZnO

(ZnO:Al). Its transmission is 90% [12] which deposited through various processing techniques such as CVD, sputtering etc. CIGS absorber layer band gap can be tuned with variation of In:Ga ratio[2]. This results maximum incident sunlight absorption near to the p-n hetero-junction and allows the CIGS absorber layer to have lower thickness compared to traditional Si cells. Cadmium Sulfide (CdS) is mostly used buffer layer in CIGS solar cell to establish p-n junction with absorber layer and yielded high- efficiency cells [13]. It is fabricated between TCO window layer and CIGS absorber layer. It is a direct band gap semiconductor has a band gap of 2.4 eV that utilizes most energy carrying photons to pass through. Molybdenum (Mo) showed constancy with processing temperature, hence consider as a most suitable back contact [13]. It has a metal work function of 4.95 eV [14]. The purpose of this layer is to collect the carriers from the absorber layer. Thickness of back contact varies with the requirement of resistance of the solar cell Solar cell is fabricated on substrate and it is the starting point of the solar cell which determines the flexibility and rigidity of the cell. Glass as substrate is widely used in production as because of its expense is low and it is able to withstand corrosion. Solar cells are mostly fabricated on soda-lime glass (SLG) substrate [4].

B. Numerical modelling and cell parameters

Numerical method helps to examine the results and optimize the process parameters of the cell without fabrication. The semiconductor parameters used for CIGS were relative permittivity ϵ_r , band gap E_g , electron affinity χ_e , density of states in the conduction band N_c , density states in valence band N_v , electron band mobility μ_e , hole band mobility μ_h , donor density N_D and acceptor density N_A . Gaussian defects are also used in this simulation as it has a role in efficiency change. All those parameters of the cell are given in TABLE I. ZnO which is used as front contact has a long range of photon transmission capacity compared to others and Mo is used as back contact has a very high metal work function and those parameters are given in TABLE II. The Interface properties between CIGS/CdS and CdS/i-ZnO are also presented in TABLE III.

In our study, we numerically modeled and simulated CIGS solar cells by using SCAPS (Solar Cell Capacitance Simulator) version of 3.0.0.2 which is a 1D simulator, programmed by the University of Gent [15]. Its motive is to simulate especially the CIGS and CdTe solar cells to observe its AC and DC characteristics before fabrication [16] and observe the numerical data through this. The numerical CIGS solar cell performed on two conditions, i.e. under the dark and under the light. While solar spectrum is AM 1.5G and power density is 100mW/cm². Temperature was kept 300 K in order to observe theoretical structure CIGS cell. The function of current density versus voltage (J-V) characteristic curve is used for extraction of electrical parameters such as V_{oc} , J_{sc} , FF and efficiency. The function of current density versus voltage (J-V) characteristic curve is used for extraction of electrical parameters such as V_{oc} , J_{sc} , FF and efficiency. In this analysis, CIGS layer thickness varied from 500 nm to 2000 nm, CdS

and i-ZnO layer thickness varied from 50 nm to 200 nm and ZnO:Al layer thickness was kept constant at 100 nm and all other semiconductor parameters were fixed value as shown in TABLE I, TABLE II and TABLE III aiming to high efficient and thinner CdS/CIGS solar cell.

TABLE I
Material properties for basic CIGS materials

Parameter	ZnO:Al	i-ZnO	CdS	CIGS
Layer (nm)	100	50-200	50-200	500-2000
E_g (eV)	3.3	3.3	2.4	1.15
χ_e (eV)	4.45	4.45	4.2	4.1
ϵ_r	9	9	10	13.6
N_c (cm ⁻³)	2.20×10^{18}	2.20×10^{18}	2.20×10^{18}	2.20×10^{18}
N_v (cm ⁻³)	1.80×10^{19}	1.80×10^{19}	1.80×10^{19}	1.80×10^{19}
v_e (cm/s)	1.00×10^7	1.00×10^7	1.00×10^7	1.00×10^7
v_h (cm/s)	1.00×10^7	1.00×10^7	1.00×10^7	1.00×10^7
μ_e (cm ² /Vs)	100	100	100	100
μ_h (cm ² /Vs)	31	31	25	25
N_D (cm ⁻³)	1.00×10^{20}	1.00×10^{18}	5.00×10^{17}	0
N_A (cm ⁻³)	0	0	0	2.00×10^{18}

Defects properties

Defect type	Donor	Donor	Acceptor	Donor
Energetic distribution	Gaussian	Gaussian	Gaussian	Gaussian
Energy level (eV)	1.65	1.65	1.2	0.6
σ_n	1.00×10^{-15}	1.00×10^{-15}	1.00×10^{-15}	5.00×10^{-17}
σ_p	5.00×10^{-13}	5.00×10^{-13}	5.00×10^{-13}	1.00×10^{-13}
N_t	1.00×10^{17}	1.00×10^{17}	1.00×10^{18}	1.00×10^{14}

TABLE II
Contact properties

Parameter	Back contact (Mo)	Front contact (ZnO:Al)
SN_0 (cm/s)	1.00×10^{17}	1.00×10^{17}
SP_0 (cm/s)	1.00×10^{17}	1.00×10^{17}
ϕ_m (eV)	4.95	4.40
Reflectivity/Transmittivity	0.80 (Reflectivity)	0.90 (Transmittivity)

TABLE III
Interface properties

Parameter	CIGS/CdS	CdS/i-ZnO
Defect type	Acceptor	Acceptor
σ_n	1.00×10^{-14}	1.00×10^{-15}
σ_p	1.00×10^{-14}	1.00×10^{-15}

III. RESULTS AND DISCUSSIONS

A. Optimization of CIGS Absorber Layer Thickness

The performances of solar cell depend on the thickness of the cell. In this study, CIGS absorber layer has varied between 500 nm and 2000 nm with CdS 50 nm and i-ZnO 50 nm. In Fig. 2. the performances of simulated cells are shown in respect to V_{oc} , J_{sc} , FF and efficiency. All those solar cell output parameters in Fig. 2 show an upward trend with the absorber layer is increment. This is because with rise in CIGS absorber layer thickness, to be absorbed the wavelengths of illumination are also increased. The long wavelength photons will go deep within the CIGS layer and which will result rise in electron hole pair (EHP) generation, so that the V_{oc} , J_{sc} , FF

and efficiency increased with rise in CIGS layer thickness and the best value of the efficiency found at the thickness 2000 nm with highest efficiency 22.67%. On the other hand for a very thin absorber layer, back contact is very close to it, so that capturing electrons by back contact will grow up. Thus reduces the V_{oc} and J_{sc} . However, further increase in CIGS layer can increase the efficiency but it will ultimately increase the material cost.

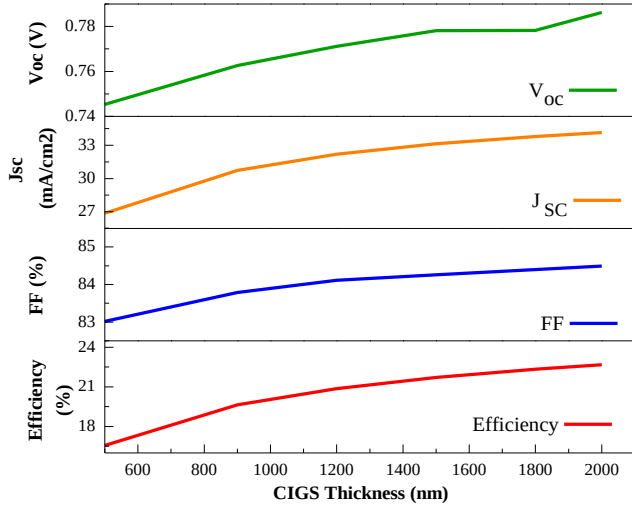


Fig. 2. CIGS absorber layer thickness variation on solar cell performance

Fig. 3 illustrates the quantum efficiency (QE) of variable CIGS absorber layer thickness. Thicker CIGS layer will lead to rise in number of longer wavelength photons and it will contribute to more EHP generation and collection. The effect can clearly be seen in the QE of the solar cell.

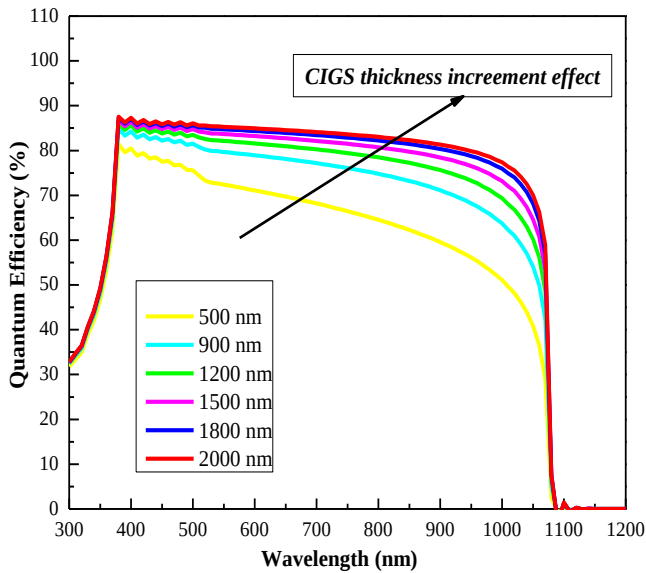


Fig. 3. CIGS absorber layer thickness variation on Quantum Efficiency

QE in thicker absorber layer is much higher than the thinner absorber layer as illustrated in Fig.3. In case of thick absorber layer generation process has occurred far from the back contact region in the long wavelength portion.

B. Optimization of CdS Buffer Layer Thickness

The optimum thickness of CdS layer found is (50-60) nm. Here CdS layer is changed from 50 nm to 200 nm with CIGS 2000 nm and i-ZnO 50 nm to observe the optimum performance. Fig. 4 illustrates the effects in details of with increase in thickness of CdS layer.

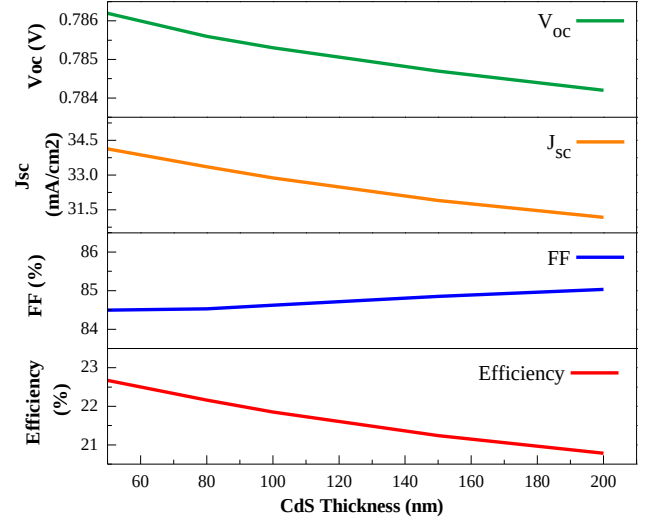


Fig. 4. CdS buffer layer thickness variation on solar cell performance

It depicts that J_{sc} reduces greatly with increased CdS layer and V_{oc} affected a little. Meanwhile FF has increased slightly as it is inversely proportional to V_{oc} and J_{sc} . Thus, from the overall effect finally the efficiency has decreased with the rise in CdS layer thickness. This is because with increment in buffer layer, more energy carrying photons will be absorbed here. Therefore it lessens the number of photons which could have the possibility to go near the absorber layer and hence the efficiency has been decreased from 22.67% to 20.78%.

C. Optimization of i-ZnO Layer Thickness

The effect of i-ZnO thickness on the output parameter is similar to the effect of CdS buffer layer but is varied in magnitude.

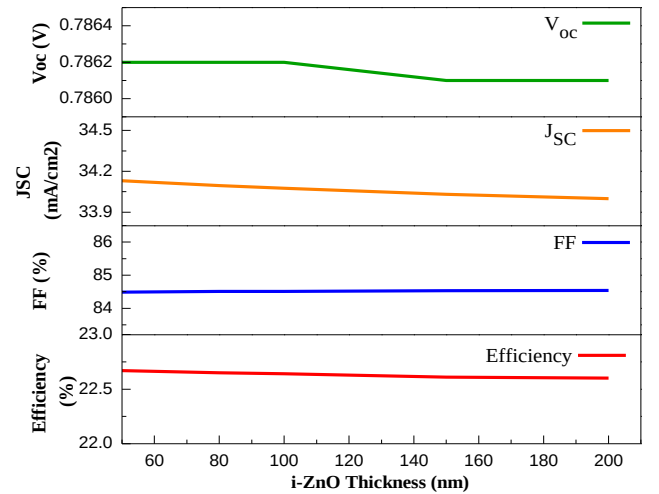


Fig. 5. i-ZnO window layer thickness variation on solar cell performance

An increase in the thickness of i-ZnO layer from 50 nm to 200 nm Fig. 5 shows that efficiency reduced from 22.67% to 22.60% where other layer is as follows CIGS 2000 nm and CdS 50 nm. Besides, V_{oc} is also dropped by a little margin and J_{sc} decreased from 34.130 mA/cm² to 33.999 mA/cm² whereas FF is almost remained same.

These effects have also reduced QE as it depends on the thickness of window layer. Light crossed through window layer first to arrive near the p-n junction. QE will be reduced when light is absorbed by the layer most as because carriers number is decreased. The maximum QE is found with 50 nm i-ZnO as shown in Fig. 5.

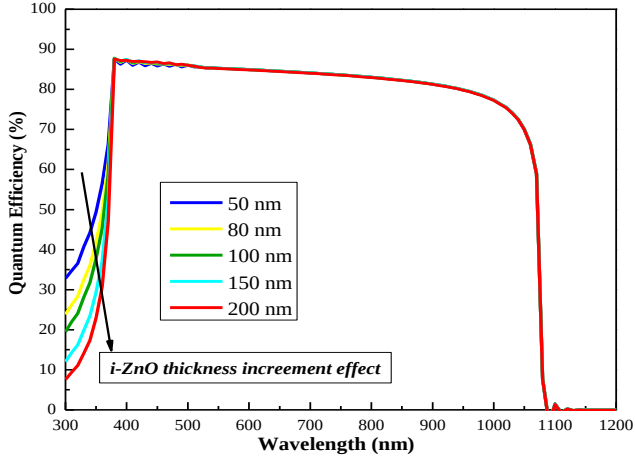


Fig. 6. I-ZnO window layer thickness variation on Quantum Efficiency

D. Optimization of Operating Temperature

Operating temperature plays a critical role in solar cell electrical parameters. The operating temperature optimal at 300 K or 27 °C. In a solar cell, V_{oc} and efficiency is mostly affected by an increase in temperature.

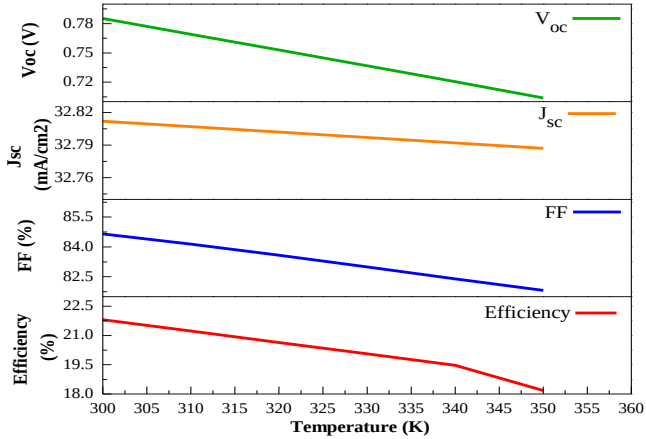


Fig. 7. Temperature variation on solar cell performance

Fig. 7 represents that solar cell performances have degraded with increase in temperature. With temperature increment band gap energy becomes unstable [9] resulting electrons and holes recombination in the regions. This results reduction in J_{sc} , V_{oc} and FF. Hence, efficiency has also reduced.

E. J-V Characteristic Curve of High Efficiency CIGS Cell

Fig. 10 shows the J-V curve of high efficient CIGS solar cell under light and dark condition on AM 1.5G where $V_{oc} = 0.7862$ V, $J_{sc} = 34.130$ mA/cm², FF = 84.49% with the layers as follows 2000 nm CIGS, 50 nm CdS, 50 nm i-ZnO and 100 nm ZnO:Al

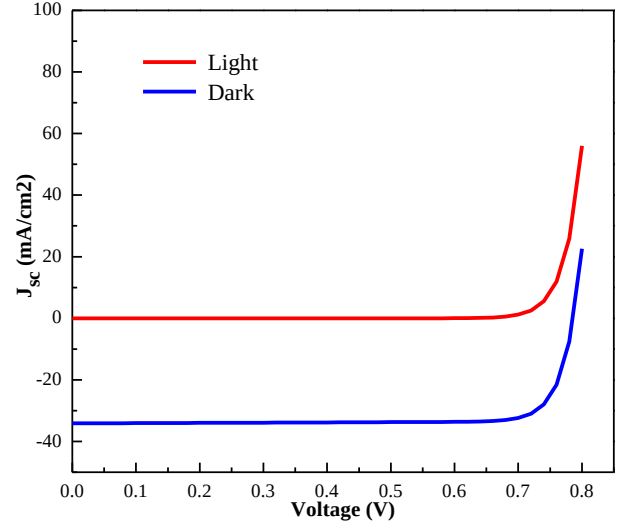


Fig. 8. J-V Curve of high efficiency CIGS cell

In our simulation, CIGS solar cell has crossed over the highest efficiency of the laboratory based cell as shown in TABLE IV. Though current density of simulated CIGS cell is less compared to laboratory produced CIGS cell, open circuit voltage and fill factor is much higher than that cell.

TABLE IV
Comparison of simulated and lab based CIGS solar cell output parameters

	Efficiency (%)	V_{oc} (V)	FF (%)	J_{sc} (mA/cm ²)
Simulated CIGS	22.67	0.7862	84.49	34.13
Solar frontier [4]	22.30	0.4219	78.2	39.38
ZSW [5]	21.70	0.746	79.3	36.6
Solibro [4]	21.0	0.757	77.6	35.70

IV. CONCLUSION

In our work, numerical modeling and analysis of CIGS solar cell has carried out using SCAPS-1D simulator by varying the layer thickness of the CIGS solar cell. The output parameters depicts that optimum efficiency is obtained with increment in the thickness of CIGS absorber layer from 500 nm to 2000 nm. Besides V_{oc} , J_{sc} and FF are also enhanced with increased absorber layer. As well as QE also improved with widen CIGS absorber. Meanwhile, a rise in buffer layer and window layer reduces the performances of CIGS solar cell. The highest efficiency of CIGS cell obtained is 22.67% whereas at present best efficiency is 22.3%. These results can be useful for a desired CIGS solar cell fabrication. But some other parameters need be investigated in further studies.

REFERENCES

- [1] Elizabeth M. Tennyson, Jesse A. Frantz, Jason D. Myers, Jas S. Sanghera, et al., "Imaging EQE in CIGS solar cells with high spatial resolution", IEEE 42nd Photovoltaic Specialist Conference (PVSC), pp. 1-4, June 2015.
- [2] Clas Persson, "Thin-film ZnO/CdS/CuIn_{1-x}Ga_xSe₂ solar cells: anomalous physical properties of the CuIn_{1-x}Ga_xSe₂ absorber," Brazilian Journal of Physics, vol. 36, no. 3B, pp. 948-951, 2006.
- [3] S.S. Viswanathan, C Ik-Ho, L. Chi-Woo, "Progress in electrodeposited absorber layer for CuIn(1-x)Ga_xSe₂(CIGS) solar cells," Solar Energy vol.85, pp. 2666-2678, 2011.
- [4] Martin A. Green, Keith Emery, Yoshihiro Hishikawa, Wilhelm Warta and Ewan D. Dunlop "Solar cell efficiency tables (version 48)" Progress Photovoltaics journal, vol. 24, pp. 905-913, 2016.
- [5] Theresa Magorian Friedlmeier, Philip Jackson, Andreas Bauer, et al., "Improved Photocurrent in Cu(In,Ga)Se₂ Solar Cells: from 20.8% to 21.7% Efficiency," IEEE Journal of Photovoltaics, vol. 5, pp. 1487-1491, 2015.
- [6] Masuko K, Shigematsu M, Hashiguchi T, Fujishima D, et al., "Achievement of more than 25% conversion efficiency with crystalline silicon heterojunction solar cell", IEEE Journal of Photovoltaics, vol. 4, pp. 1433-1435, 2014.
- [7] M. Mostefaoui, H. Mazari, S.Khelifi, A.Bouraiou, R.Dabou, "Simulation of High Efficiency CIGS solar cells with SCAPS-1D software", International Conference on Technologies and Materials for Renewable Energy, Environment and Sustainability, TMREES15, Energy Procedia 74, vol. pp. 736 – 744, 2015.
- [8] A. Benmira, M. S. Aidab, "Analytical modeling and simulation of CIGS solar cells", TerraGreen13 International Conference, Advancements in Renewable Energy and Clean Environment, Vol. 36, pp.618-627, 2013.
- [9] Puvaneswaran Chelvanathan, Mohammad Istiaque Hossain, Nowshad Amin, "Performance analysis of copper-indium-gallium-diselenide (CIGS) solar cells with various buffer layers by SCAPS", 19th International Photovoltaic Science and Engineering Conference and Exhibition (PVSEC-19), vol. 10, Issue 3, pp. S387-S39, 2010.
- [10] Sharmin Aktar Chowdhury, Md. Shah Alam, "Enhancement of Device Performance with Optimized Absorber Layer in CIGS Solar Cell", Proceedings of International Conference on Electrical Information and Communication Technology (EICT 2015), pp. 475 – 479, December 2015.
- [11] Chiara Candelise, Jamie F. Spiers, Robert J.K. Gross, "Materials availability for thin film (TF) PV technologies development: A real concern? " Renewable and Sustainable Energy Reviews, vol. 15, pp. 4972-4981, 2011.
- [12] Fengyan Zhang, Chivin Sun, Cyril Bajracharya, Rene G. Rodriguez, and Joshua J. Pak, "Fabrication and Characterization of Thin Film Solar Cell Made from CuIn_{0.75}Ga_{0.25}Se₂ Wurtzite Nanoparticles", Hindawi Publishing Corporation Journal of Nanomaterials, vol. 2013, pp. 1-5, 2013.
- [13] Udai P. Singh and Surya P. Patra, "Progress in Polycrystalline Thin-Film Cu(In,Ga)Se₂ Solar Cells", International Journal of Photoenergy, vol. 2010, pp. 1-19, 2010.
- [14] P. Ranade, H. Takeuchi, T.-J. King, and C. Hu, "Work function engineering of molybdenum gate electrodes by nitrogen implantation," Electrochemical and Solid-State Letters, vol.4 (11), pp.G85-G87, 2001.
- [15] M. Burgelman, P. Nollet and S. Degrave, "Modelling polycrystalline semiconductor solar cells", Thin Solid Films, vol. 361-362, pp. 527-532, 2000.
- [16] M. Burgelman, K. Decock, S. Khelifi and A. Abass, "Advanced electrical simulation of thin film solar cells", Thin Solid Films, vol 535, pp. 296-301, 2013.

BER performance analysis of a MC-DS-CDMA wireless communication system with rake receiver employing MRC under Nakagami – m fading

Anupama Tasneem ^{# 1}

Department of EEE
Ahsanullah University of Science and Technology
Dhaka, Bangladesh
anupamatasneem.eee@aust.edu

Satya Prashad Majumder ^{# 2}

Department of EEE
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
spmajumder@eee.buet.ac.bd

Abstract— A novel theoretical analysis is presented for a wireless communication system using MC-DS-CDMA with OFDM subcarriers and Rake receivers under Nakagami - m fading environment considering maximal ratio combining (MRC) for the rake combiner. Here the probability density function (PDF) expression of the signal to noise ratio at the output of rake combiner is developed for individual sub-channels of the OFDM sub carriers. The bit error rate (BER) is then determined for each individual OFDM sub-channels and the average BER is then evaluated numerically. The outcomes show that the performance with respect to BER improves significantly with the increase in the number of rake fingers.

Keywords—MC-CDMA; Nakagami – m fading; Bit Error Rate (BER) performance evolution; Maximal Ratio Combining (MRC); Rake Receiver; Multi Access Interference (MAI).

I. INTRODUCTION

Multicarrier code-division multiple access (MC-CDMA) is a promising spread spectrum transmission method where each input data stream is partitioned into parallel sub streams that are modulated at different subcarriers [1-11]. Through BPSK modulated signals subcarriers having narrowband are generated with frequencies multiples of a fundamental frequency [12]. A combination of narrowband subcarrier signals results in a MC CDMA signal [13, 14]. In MC CDMA use of orthogonal subcarriers by OFDM technique provides higher data rate which leads to improved bandwidth efficiency compared to DS CDMA. Moreover MC CDMA has many other advantages like high spectrum efficiency, strong resistivity to noise, flexibility to handle multiple data rates and capable of spreading the original signal bandwidth by mitigating the inimical effect of Inter Symbol Interference [15].

In Wireless Communication systems fading and co-channel interference degrade signal quality. Several techniques are used in order to improve the intensity of signal such as adaptive arrays, antenna diversity, equalization etc. Among all the diversity techniques used in wireless communication, Equal Gain Combining (EGC) and Maximal Ratio Combining (MRC)

are the most effective ones. MRC is considered as optimal combining with careful estimation of channel induced fading. The Nakagami - m distribution is a statistical distribution used to model different fading parameters [16]. Usually this distribution is more flexible and accurate to match different experimental data compared to other distributions such as Rayleigh, Lognormal and Rician distribution [17] particularly in some urban radio multipath channels [18].

Here we evaluate the characteristics of MRC diversity reception in the Nakagami fading environment. In spite of the complexity of the MRC compared to other diversity techniques, we've considered this since it has the highest improvement possibility which can be attained through by a diversity system through a fading channel [19]. Signals received in MRC are calculate in such a way that the weighted sum at the combiner output has the highest SNR (signal to noise ratio) [19]. In this paper MC-CDMA system model is first described. The analysis is carried out to derive the expressions of signal to noise ratio (SINR), multi access interference (MAI), bit rate (BER) and finally average bit error rate of MC-DS-CDMA system having RAKE receiver under Nakagami-m fading environment. Performance is evaluated numerically for different system parameters as improvement due to rake receiver.

II.

III. SYSTEM MODEL

The MC-DS-CDMA system model with OFDM modulator is described in this section where the Transmitter model is shown in fig. 1(a) and the receiver model is shown in fig. 1(b). In transmitter block, it shows MC-DS-CDMA with OFDM modulator where input data stream is transformed into N_c sub stream having bit duration, T_b . Parallel data symbols of each data stream of j^{th} user are coded by the pseudo random noise (PN) sequence having chip duration (T_c) and processing gain, G_p . Thus each data symbol is spread in time domain. In each OFDM modulator there is sub carrier of respective channel which modulates the coded symbol of each data stream. MC-DS-

CDMA signal is then transmitted through a multipath Nakagami-m fading wireless channel.

Receiver block mainly consists of OFDM demodulator, Rake Receiver with L_r branches, Integrators and Maximal ratio combiner (MRC). During reception, the j -th user of the receiver receives all the signals transmitted by K no of users. Each receiver is equipped with Rake fingers and these Rake fingers outputs are combined by Maximal ratio combiner.

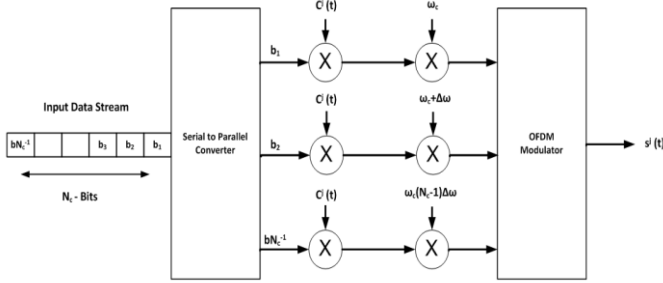


Fig. 1(a): Transmitter Model

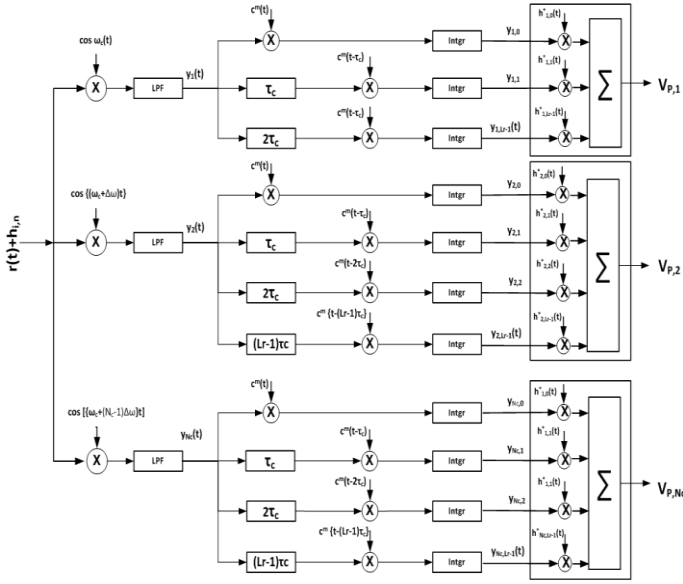


Fig. 1(b): Receiver Model

IV. ANALYSIS OF MULTICARRIER CDMA WITH RAKE RECEIVER

For the j -th user with code $c^j(t)$, the general expressions of the signal transmitted is,

$$s^j(t) = \sum_{n=1}^{N_c} \sqrt{2P_T} \cdot b_n^j(t) \cdot c^j(t) \cdot \cos[\{\omega_c + (n-1)\Delta\omega\}t + \varphi_n(t)] \quad (1)$$

[20]

Where,

$$c^j(t) = \sum_{l=0}^{L-1} c_l^j p(t - lT_c)$$

Here, P_T is the transmitted power per symbol, T_c is the chip period, b_n^j is the transmitted symbol of j -th user in n -th sub channel which is being modulated by the n -th sub-carrier of frequency $\{\omega_c + (n-1)\Delta\omega\}$, $c^j(t)$ is the PN sequence code for the j -th user and $c_l^j(t)$ represents the l -th chip of the j -th code word, N is the code length, ω_c is the carrier frequency of the reference channel, $\Delta\omega$ is the frequency spacing between two successive channels, $\varphi_n(t)$ is the instantaneous /random phase angle of each sub-carrier, considered to have uniform distribution over $[0, 2\pi]$.

Now the expression of the received signal corresponding to the i -th path when there are j active users, is

$$r_{in}(t) = \int_{j=1}^K \sqrt{2P_r} \cdot b_n^j(t - i\tau_c) \cdot c^j(t - i\tau_c) \cdot h_{in}(t) \cdot \cos[\{\omega_c + (n-1)\Delta\omega\}(t - i\tau_c) + \varphi_n] + n(t) \quad (2)$$

Where, P_r is the received signal power, K is the number of user, τ_c is the chip delay and $n(t)$ is additive white Gaussian noise (AWGN) with two sided power spectral density $N_0/2$. Here $h_{i,n}$ is the fading parameter.

Now the composite signal received for the whole sequence is,

$$r(t) = \sum_{i=0}^{L_r-1} r_{i,n}(t) \quad (3)$$

Where, L_r is the no of resolvable paths and $n=1:N$

Now, signal at the output of the n -th coherent demodulator is,

$$y_n(t) = \sum_{i=0}^{L_r-1} \sum_{j=1}^K \sqrt{\frac{P_r}{2}} b_n^j(t - i\tau_c) \cdot c^j(t - i\tau_c) \cdot h_{in}(t) \cdot \cos\{\varphi_n(t)\} + n(t) \quad (4)$$

To determine the signals of the i -th branch of the Rake receiver and N_c -th sub-carrier $y_n(t)$ is multiplied with $c^m(t - i\tau_c)$ where $i=0 : L_r-1$. So, the output of the i -th correlator corresponding to the m -th user is

$$y_{n,i}(t) = \sum_{i=1}^{L_r-1} \sum_{j=1}^K \sqrt{\frac{P_r}{2}} b_n^j(t - i\tau_c) \cdot h_{in}(t) \cdot \cos\{\varphi_n(t)\} \cdot \frac{1}{T_b} \int_0^{T_b} c^j(t - i\tau_c) \cdot c^m(t - i\tau_c) dt \quad (5)$$

Now the maximal ratio combiner (MRC) output of is,

$$v_{p,n} = \sum_{i=1}^{L_r-1} y_{n,i}(t) \cdot h_{in}^*(t) \quad (6)$$

Here, $j=m$ gives signal component and is given by the expression,

$$v_{n,sig} = \sum_{i=0}^{L_r-1} \sqrt{\frac{p_r}{2}} \cdot [h_{in}(t) \cdot h_{in}^*(t)] \cdot b_n^m(t - i\tau_c) \cdot C_{o,m} \quad (7)$$

Where, $C_{o,m} = G_P T_C$

Now the Mean Square Signal Power of n-th subcarrier is,

$$P_r = L_r \times \frac{p_r}{2} \times |h_{in}(t) \cdot h_{in}^*(t)|^2 \quad (8)$$

Where,

$$[h_{in}(t) \cdot h_{in}^*(t)]^2 = |h|^4 \text{ and } |C_{o,m}|^2 = 1$$

Again, the condition $j \neq m$ gives Multi Access Interference (MAI) and the expression is,

$$v_{n,MAI} = \sum_{i=0}^{L_r-1} \sum_{j=0}^k \sqrt{\frac{p_r}{2}} \times [h_{in}(t) \cdot h_{in}^*(t)] \times b_n^j(t - i\tau_c) \times \cos\{\varphi_n(t)\} \times r_{j,m} \quad (9)$$

Where,

$$r_{j,m} = \frac{1}{T_b} \int_0^{T_b} c^j(t - i\tau_c) \cdot c^m(t - i\tau_c) dt \text{ and } r_{j,m} = \frac{1}{G_P} \sum_{l=1}^{G_P} C_{j,l} \cdot C_{m,l}$$

Now the MAI power for the n-th subcarrier,

$$P_{MAI,n} = MS\{v_{n,MAI}\}$$

$$P_{MAI,n} = (k-1) \cdot L_r \cdot \frac{p_r}{2} \cdot [h_{in}(t) \cdot h_{in}^*(t)]^2 \cdot \frac{1}{G_P} \quad (10)$$

Where, $MS(r_{j,m}) = \frac{1}{G_P}$ and the variance of noise power,

$$\sigma_n^2 = \frac{N_0}{4T_b} \quad (11)$$

Now,

The Conditional BER for BPSK modulated signal is given by,

$$BER(n, h) = \frac{1}{2} \operatorname{erfc} \left[\sqrt{\frac{\gamma(h)}{2}} \right] \quad (12)$$

Where,

Signal to interference + noise ratio,

$$SINR = \gamma_p(n, h) = \frac{L_r \cdot \frac{p_r}{2} \cdot [h_{in}(t) \cdot h_{in}^*(t)]^2}{(k-1) \cdot L_r \cdot \frac{p_r}{2} \cdot [h_{in}(t) \cdot h_{in}^*(t)]^2 + \frac{1}{G_P} + \frac{N_0}{4T_b}}$$

Here, received signal power is denoted as P_r and the bit duration is denoted as T_b . This is obtained by combining equation (8), (10), (11). Hence unconditional BER can be derived as below –

$$BER(n) = \int_0^\infty BER(n, h) \cdot p(h) \cdot dh \quad (13)$$

Now, $h_{l,k}$ is the probability density function (PDF) of Nakagami-m distributed random variable which is given by [21],

$$p(h_{l,k}) = 2 \cdot \left(\frac{m}{\Omega_{l,k}}\right)^m \cdot \frac{h_{l,k}^{2m-1}}{\Gamma m} \cdot e^{-\left(m \cdot \frac{h_{l,k}^2}{\Omega_{l,k}}\right)} \quad (14)$$

Where, m is the Nakagami-m fading parameter and parameter m of amplitude distribution characterizes the severity of fading over the l-th resolvable path [22].

Average BER is then obtained as,

$$BER = \frac{1}{N} \sum_n BER(n)$$

V. RESULTS AND DISCUSSIONS

Following the theoretical analysis shown in section 3, we've figured out the bit error rate performance of a wireless communication system using MC-DS-CDMA with rake receiver employing nakagami-m fading environment. The plots of BER v number of users for a MC-DS-CDMA system with nakagami fading of order, m=4 is shown in figure 3(a) for different values of processing gain and number of users without rake receiver. It is observed that bit error rate increases with increased number of users due to the effect of MAI. The level of the BER floor depends on the value of processing gain, G_p and number of users, K. BER reduces for higher values of processing gain for a given number of users and vice versa.

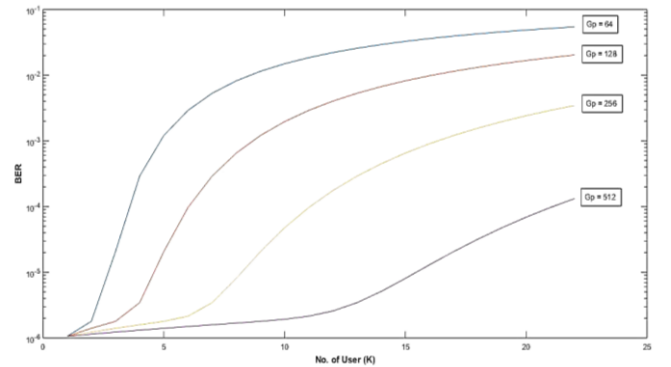


Fig.2(a): BER Vs number of simultaneous users in nakagami fading channel with MRC equalization for different processing gain, $G_p=64, 128, 256, 512$ with $L_r=1$.

The BER curves with rake fingers $L_r=2$, $L_r=4$ and $L_r=6$ are indicated in figure 2(b), 2(c) and 2(d) respectively. Compared to the performance without rake receiver of figure 2(a) it is found that there is a significant improvement in BER with rake fingers. It is noticed that BER reduces by order of magnitude as L_r is increased from 2 to 6.

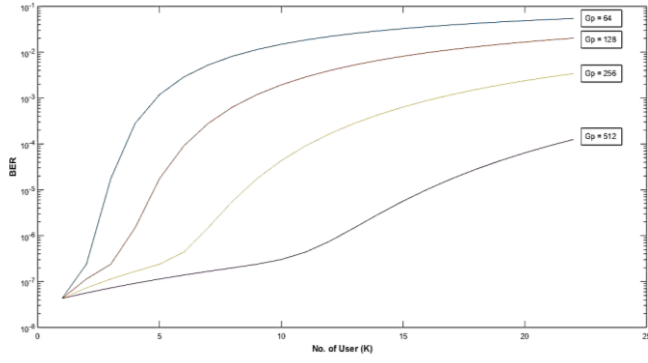


Fig.2(b): BER Vs number of simultaneous users in nakagami fading channel with MRC equalization for different processing gain ,Gp=64,128,256,512 with Lr=2.

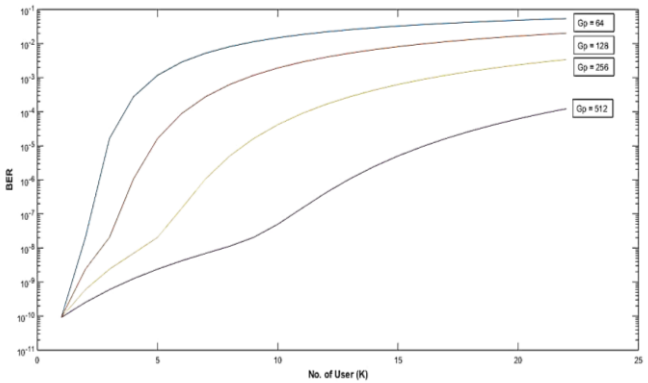


Fig.2(c): BER Vs number of simultaneous users in nakagami fading channel with MRC equalization for different processing gain ,Gp=64,128,256,512 with Lr=4.

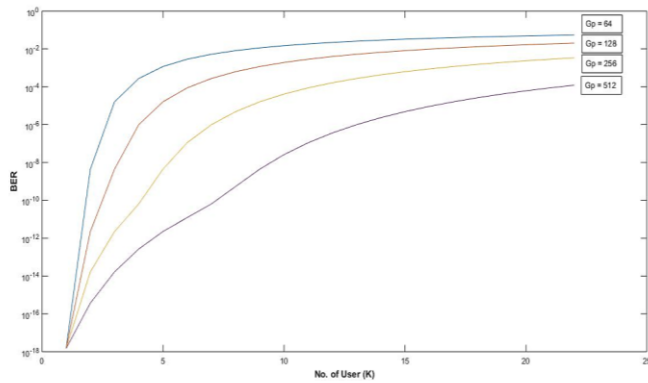


Fig.2(d): BER Vs number of simultaneous users in nakagami fading channel with MRC equalization for different processing gain ,Gp=64,128,256,512 with Lr=6.

The plots of number of user Vs processing gain with Nakagami fading of order $m = 4$ at a fixed BER = 10^{-6} are shown in fig. 2(e) for different values of rake fingers.

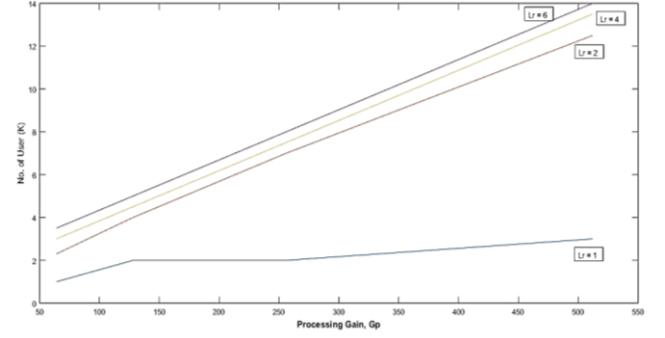


Fig.2(e): Number of users Vs different number of processing gain in nakagami fading channel with MRC equalization with Lr=1,Lr=2,Lr=4andLr=6.

The results are evaluated at number of rake fingers, for example $L_r = 1, 2, 4, 6$ and it indicates that number of user increases with increased number of rake fingers along with increased processing gain.

VI. CONCLUTIONS

Analysis is presented for a wireless communication system which employs MC-DS-CDMA with rake receiver for each OFDM sub channels. Bit error rate is numerically evaluated for each OFDM sub band considering MRC for the rake combiner. The results related to average BER show that there is a significant improvement in BER performance with increase in Rake fingers.

ACKNOLEDGEMENT

This work is carried out in the department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh for the partial fulfilment of the degree of MSC Engineering.

REFERENCES

- [1] K. Fazel, S. Kaiser, and M. Schnell, "A flexible and high performance cellular mobile communications system based on orthogonal multi-carrier SSMA," *Wireless Per. Commun.*, vol.2, no.1&2, pp.121–144, Mar.1995.
- [2] S.Hara and R.Prasad, "Overview of multi carrier CDMA," *IEEE Commun. Mag.*, vol.35, no.12, pp.126–133, Dec.1997.
- [3] S. Hara and R. Prasad, "Design and performance of multicarrier CDMA system in frequency-selective Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol.48, no.5, pp.1584–1595, Sept.1999.
- [4] T. Mueller, K. Brueninghaus, and H. Rohling, "Performance of coherent OFDM-CDMA for broadband mobile communications," *Wireless Per. Commun.*, vol.2, no.4, pp.295–305, 1996.

- [5] E. A. Sourour and M. Nakagawa, "Performance of orthogonal multicarrier CDMA in a multipath fading channel," *IEEE Trans. Commun.*, vol. 44, no. 3, pp. 356–367, Mar. 1996.
- [6] L. Vandendorpe, "Multitone spread-spectrum multiple-access communications system in a multipath Rician fading channel," *IEEE Trans. Veh. Technol.*, vol. 44, no. 2, pp. 327–337, May 1995.
- [7] T. T. Liu and C. Y. Yang, "Equivalent signal-alignment-based frequency domain equalization for MC-CDMA two-way relay systems," *IEEE Trans. Veh. Technol.*, vol. 61, no. 1, pp. 237–248, Jan. 2012.
- [8] W. Yang, J. Y. Liu, C. L. Xu, and S. X. Cheng, "Joint carrier frequency offset estimation and compensation for downlink group-orthogonal multicarrier CDMA," *Wireless Pers. Commun.*, vol. 62, no. 1, pp. 107–116, Jan. 2012.
- [9] C. C. Lin, W. C. Chen, and C. D. Chung, "Spectral sidelobe decaying property of Walsh-Hadamard code in MC-CDMA systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 10, pp. 3151–3157, Oct. 2011.
- [10] Y. Meng, M. L. You, H. W. Luo, and G. Liu, "The subspace-based linear conjugate CMA in BPSK-modulated MC-CDMA systems," *Wireless Pers. Commun.*, vol. 56, no. 4, pp. 761–777, Feb. 2011.
- [11] F. Zabini, B. M. Masini, A. Conti, and L. Hanzo, "Partial equalization for MC-CDMA systems in non-ideally estimated correlated fading," *IEEE Trans. Veh. Technol.*, vol. 59, no. 8, pp. 3818–3830, Oct. 2010.
- [12] A. B. Djebbar, A. Djebbari, M. Bouziani, and J. M. Rouvaen, "Derivation of new expressions of bit error rate for MC-CDMA system in Nakagami fading channel," *AEÜ - Int. J. Electron. Commun.*, vol. 57, no. 6, pp. 395–402, 2003.
- [13] Webber, C. L.; Huth, G. K.; Baston, B.: Performance consideration of code division multiple access. *IEEE Trans on vehicular technology* 30(1981), 3-9.
- [14] Yee, N.; Linnartz, J.: Multi-carrier code division multiple access (MC-CDMA): A new spreading technique for communication over multipath channels. Final report for Micro project pp. 93-101. University of California, Oakland, 1994.
- [15] L. L. Yang and L. Hanzo, "Multicarrier DS-SS-CDMA: A multiple access scheme for ubiquitous broadband wireless communications," *IEEE Commun. Mag.*, vol. 41, no. 10, pp. 116–124, Oct. 2003.
- [16] M. Nakagami, "The m-distribution: A general formula of intensity distribution of rapid fading," in *Statistical Methods in Radio Wave Propagation*, W. C. Hoffman, Ed. Oxford, U.K.: Pergamon, 1960.
- [17] Tanda, M.: Bit error rate of DQPSK signals in slow nakagami fading. *Electronics Letters* 29 (1993), 134-136.
- [18] Sujiki, H.: A statistical model for urban multipath propagation. *IEEE Trans. on Communications* 25 (1977), 673-680.
- [19] Emad K. and Abdel Aziz M.: Performance of MRC Diversity Systems for the detection of signals with nakagami fading. *IEEE Trans. on Communications* 33 (1985), 1315-1319.
- [20] L. L. Yang and L. Hanzo, "Slow Frequency – Hopping Multicarrier DS-SS-CDMA for Transmission over Nakagami Multipath Fading Channels," *IEEE Journal on Selected areas in Communications*, vol. 19, no. 7, pp. 1211–1221, July 2001.
- [21] Lie-Liang Yang and Lajos Hanzo, "Performance of Generalized Multicarrier DS-SS-CDMA over Nakagami-m Fading Channel". *IEEE Transactions on Vehicular Technology*, vol. 47, no. 4, pp. 1234 – 1244, November, 1998.
- [22] M. – S. Alouini and A. J. Goldsmith, "A unified approach for calculating error rates of linearly modulated signals over generalized fading channels," *IEEE Transactions on Communications*, vol. 47, pp. 1324 – 1334, September, 1999.

Numerical Simulation of CdS/CdTe Tandem Solar Cell with Addition of ET-HBL

Al Amin

Department of Electrical & Electronic Engineering
Rajshahi University of Engineering & Technology
Rajshahi-6204, Bangladesh
alamin.3eee@gmail.com

Abstract—Heterojunction structure CdS/CdTe solar cell has long been regarded as one promising choice for the development of cost-effective and reliable solar cells. CdS/CdTe tandem solar cell with addition of Electron Transport-Hole Blocking Layer (ET-HBL) on the Cadmium Sulfide (CdS) window layer has been simulated using One-Dimensional Device Simulation Program for Analysis of Microelectronic and Photonic Structures (AMPS-1D) in this simulation model. The band diagram of the solar cell has been obtained at the thermodynamic equilibrium condition, & electron current & hole current density are also observed at short circuit & under AM 1.5G light illumination. The hole density of CdTe has tremendous effect on the efficiency of solar cell. The maximum efficiency of 36.908% has been achieved by varying the hole density of CdTe. The performance parameters of the CdS/CdTe solar cell with ET-HBL are compared with the without ET-HBL.

Keywords—Tandem Solar Cell; ET-HBL; Hole Density; Heterojunction; Band Diagram; Electron & Hole Current Density; AMPS-1D; Efficiency

I. INTRODUCTION

The world is suffering from impending death of fossil fuels and serious environment pollution resulted from the burning of fossil fuels. The burning of fossil fuels causes greenhouse effect, environment gets warm and becomes polluted. So it is more and more widely recognized that renewable energy sources, especially solar energy is now regarded as one promising solution to the global energy crisis [1]. So the solar cell efficiency should be enhanced to subside the energy crisis problem. In Different times, many scientist have worked to enhance the sunlight energy to electric energy conversion efficiency by using different techniques and diverse materials, semiconductor-metal alloys. $\text{In}_{0.52}\text{Ga}_{0.48}\text{N}$, Copper-Indium Selenide (CIS), copper indium gallium diselenide (CIGS) are the examples of different semiconductor alloy solar cell [2-4]. Ohl first traced the Si solar cell in 1941 [5]. Professor Stephen J. Fonash showed 12.45% efficiency of Si solar cell by using front Hole Transport-Electron Blocking Layer (HT-EBL) and back ET-HBL in his book [6]. After that the efficiency of Si solar cell with front HT-EBL was investigated 26.285% at 2.10 eV band gap energy of HT-EBL in 2015 [7]. From this works the idea has been developed in this simulation model.

Cadmium telluride (CdTe) is now regarded as one leading material for the development of cost-effective tandem solar cell and it is also the first PV technology with the price for Wp below \$1 [8]. The crystalline structure of CdTe is cubic zinc blend and has direct optical band gap of 1.50 eV [9]. Efficiency of 16.5% has been achieved for CdS/CdTe heterojunction structure in laboratory in 2001 [10] and the efficiency reached at 17.3% with solar cell modules efficiency of 11.7% in 2011 [11]. Under global AM 1.5G illumination ultrathin CdS/CdTe solar cell using back surface field and backside distributed Bragg reflector has an open-circuit voltage of 1.062 V, a short-circuit current density of 24.64 mA/cm², and a fill factor of 81.3%, and efficiency of 21.02% [12]. The magnetron sputtered CdS/CdTe solar cell shows efficiency of 14% with sputtered ZnO:Al [13]. Many researchers have tried to enhance the solar cell performance by using ZnTe, $\text{Cd}_{1-x}\text{Zn}_x\text{Te}$ materials at back surface of CdS/CdTe heterojunction solar cell [14-16].

The main target of this work is how to enhance the CdS/CdTe tandem solar cell efficiency to solve the enormous energy crisis problems. In this simulation model, the effect of hole concentration in p type CdTe at 5000 nm layer thickness and ET-HBL at 40 nm has been investigated. The carrier lifetime of minority carrier of CdTe has been assumed 10 ns [17] and the designed cell “Fig. 1” has been simulated. The short circuit current density, open circuit voltage, Fill Factor and Efficiency are 27.103 mA/cm², -1.637 V, 91.5% and 36.908% respectively at hole density of $1.0\text{e}+20\text{ cm}^{-3}$. By using the ET-HBL, the cell efficiency increases from 34.381% to 36.908%.

II. SOLAR CELL MODELLING AND SIMULATION

AMPS-1D is a very general computer programming code for analyzing and designing two terminal microelectronic and photonic devices. It can simulate p-n and p-i-n homo- and hetero-junctions semiconductor devices. These devices may have poly-crystalline, amorphous or single crystal layers or their combinations. Device operation can be characterized in dark or under light illumination. Hence the behavior of solar cells and photodiodes can be observed by using AMPS-1D code [18]. The physics of device transport can be captured in three governing equations: Poisson’s equation, the continuity equation for free holes, and the continuity equation for free electrons [19]. The three equations are given bellow.

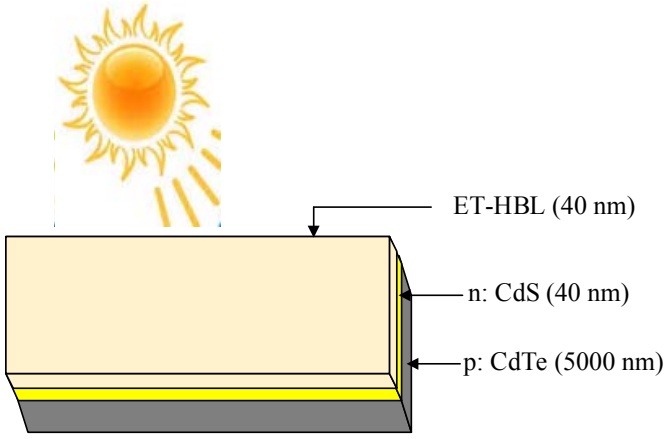


Fig. 1. Schematic diagram of the CdS/CdTe tandem solar cell with addition of ET-HBL.

$$\frac{d}{dx} \left(-\epsilon(x) \frac{d\psi}{dx} \right) = q [p(x) - n(x) + N_D^+(x) - N_A^-(x) + p_t(x) - n_t(x)] \quad (1)$$

$$\frac{\partial p_n}{\partial t} = G_p - \frac{p_n - p_{n0}}{\tau_p} - p_n \mu_p \frac{\partial \xi}{\partial x} - \mu_p \xi \frac{\partial p_n}{\partial x} + D_p \frac{\partial^2 p_n}{\partial x^2} \quad (2)$$

$$\frac{\partial n_p}{\partial t} = G_n - \frac{n_p - n_{p0}}{\tau_n} + n_p \mu_n \frac{\partial \xi}{\partial x} + \mu_n \xi \frac{\partial n_p}{\partial x} + D_n \frac{\partial^2 n_p}{\partial x^2} \quad (3)$$

where the electrostatic potential ψ , and the free electron n , free hole p , trapped electron n_t , and trapped hole p_t , as well as the ionized donor-like doping N_D^+ and ionized acceptor-like doping N_A^- concentrations are all functions of the position coordinate x . Here, ϵ is the permittivity and q is the magnitude of the charge of an electron. G_n and G_p are the electron and hole generation rate and ξ is electrostatic field.

After determining transport characteristics from above three equations then becomes a task of solving these three coupled non-linear differential equations, each of which has two associated boundary conditions. In AMPS, these three coupled equations, along with the appropriate boundary conditions, are solved simultaneously to obtain a set of three unknown state variables at each point in the device: the electrostatic potential, the hole quasi-Fermi level, and the electron quasi-Fermi level. From these three state variables, the carrier concentrations, fields, currents, etc. can then be computed. To determine these state variables, the method of finite differences and the Newton-Raphson technique are incorporated by the computer.

The device operating temperature is used as 300k and all performance parameters of cell have been taken under AM1.5G illumination. The material parameters are listed in "TABLE I".

TABLE I. MATERIAL PARAMETERS

Parameter	ET-HBL	CdS	CdTe
Layer Thickness (nm)	40	40	5000
Relative permittivity	9.00	9.00	9.40
Electron mobility, μ_n (cm ² /Vs)	350	350	500
Hole mobility, μ_p (cm ² /Vs)	50	50	60
Electron & Hole Density (cm ⁻³)	N: 1.0e+17	N: 1.0e+18	Variable
Bandgap (eV)	3.20	2.42	1.50
Effective density of states in conduction band (1/cm ³)	1.80e+19	1.80e+19	7.50e+17
Effective density of states in valence band (1/cm ³)	2.40e+18	2.40e+18	1.80e+18
Electron Affinity (eV)	4.50	4.50	4.28
Carrier Lifetime (ns)	P: 0.2	P: 0.2	N: 10

TABLE II. FRONT AND BACK CONTACT PARAMETERS

Front contact	Back contact
PHIBO = 0.00 eV	PHIBL = 1.05 eV
SNO = 1.0e+7 cm/s	SNL = 1.0e+7 cm/s
SPO = 1.0e+7 cm/s	SPL = 1.0e+7 cm/s
RF = 0.07	RB = 0.30

III. RESULTS AND DISCUSSION

CdTe is a p type material. The carrier concentration strongly depends on the deposition method [20]. If the hole density of CdTe is varied from 1.0e+15 cm⁻³ to 1.0e+21 cm⁻³, then a great variation of solar cell performance parameters is noticed due to increased carrier concentration according to physics of semiconductor. The short circuit current density (J_{sc}) and open circuit voltage (V_{oc}) changes with change in hole density. The

J_{sc} decreases and V_{oc} increases with increasing in hole density “Fig. 2”. The solar cell efficiency increases up to hole density of $1.0 \times 10^{20} \text{ cm}^{-3}$, and be maximum of 36.908% at $1.0 \times 10^{20} \text{ cm}^{-3}$ hole density, after this density the efficiency decreases with increasing in hole density of CdTe and efficiency of 11.662% is at $1.0 \times 10^{21} \text{ cm}^{-3}$ “Fig. 3”.

The thermodynamic equilibrium band diagram of the simulated solar cell is shown in “Fig. 4”. It has been reported that the energy level changes with position from 0 to 250 nm. The electron current density decreases and whole current density increases with increasing in position “Fig. 5”.

The Electron transport-hole blocking later (ET-HBL) is n type semiconductor material which transports electron from CdTe to CdS easily. The use of ET-HBL on the window layer of CdS enhances the short circuit current, open circuit voltage, Fill Factor and efficiency “Fig. 6”. The efficiency develops from 34.381% to 36.908% for the same cell thickness and surface area.

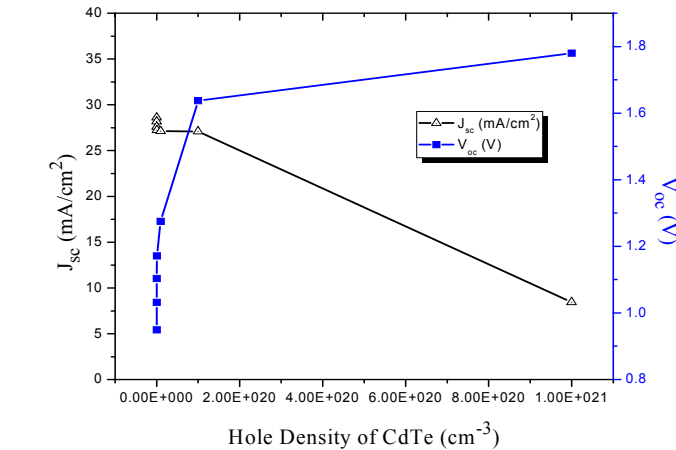


Fig. 2. Effect of hole density of CdTe on the short circuit current density (J_{sc}) & open circuit voltage (V_{oc}) of CdS/CdTe tandem solar cell with ET-HBL.

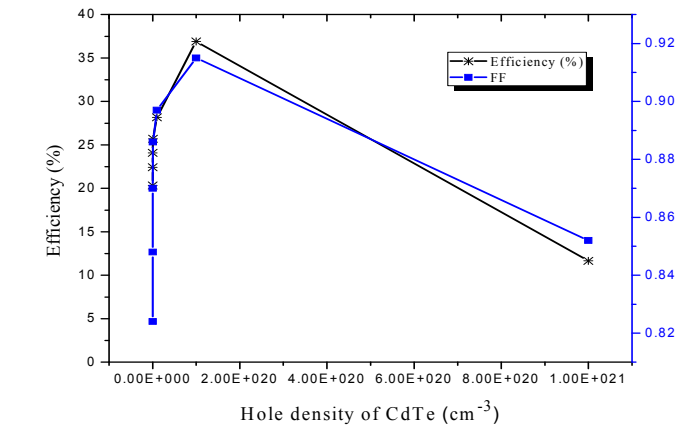


Fig. 3. Effect of hole density of CdTe on the Efficiency & Fill Factor (FF) of CdS/CdTe tandem solar cell with ET-HBL.

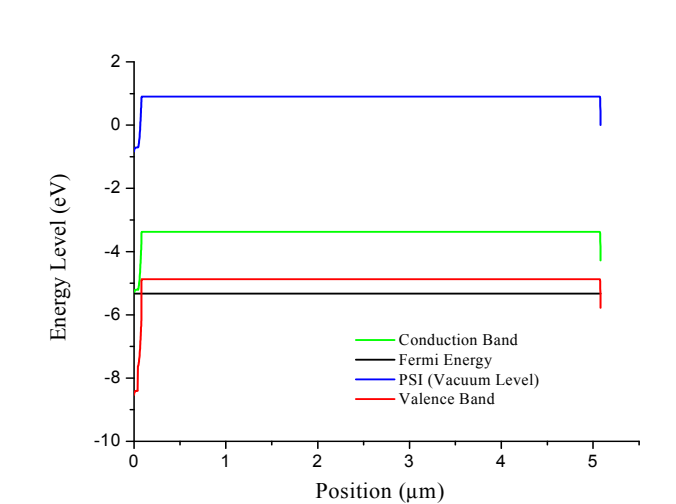


Fig. 4. The computed Thermodynamic Equilibrium (TE) band diagram of the CdS/CdTe with the addition of a 40-nm ET-HBL on the window layer of CdS at the hole density of CdTe of $1.0 \times 10^{20} \text{ cm}^{-3}$.

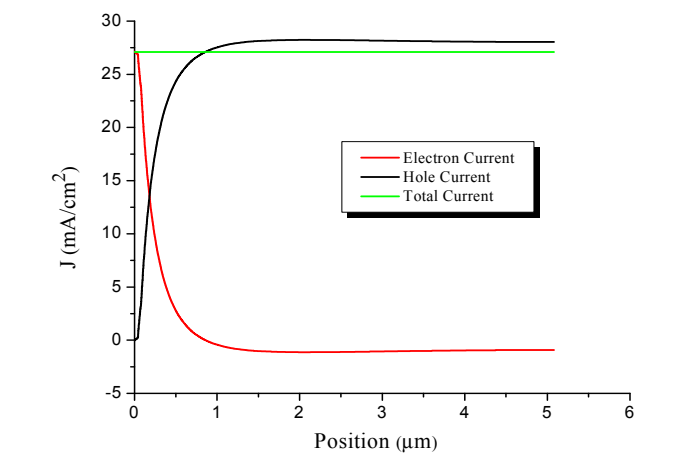


Fig. 5. The computed electron and hole current density components at short circuit for the device at the hole density of CdTe of $1.0 \times 10^{20} \text{ cm}^{-3}$.

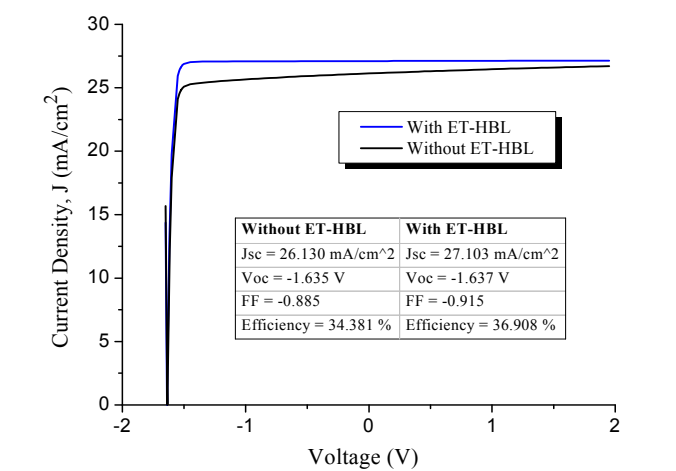


Fig. 6. Current density vs. Voltage curve of CdS/CdTe tandem solar cell and comparison between with and without ET-HBL.

IV. CONCLUSION

The light energy to electric energy conversion efficiency is an important factor to design a solar cell in practical. In this simulation model, the efficiency of CdS/CdTe tandem solar cell with addition of ET-HBL on the front CdS layer for hole density of CdTe of $1.0 \times 10^{20} \text{ cm}^{-3}$ is more than CdS/CdTe solar cell. The efficiency and Fill Factor have been enhanced respectively 2.527% and 3% due to the use of ET-HBL. So this type of solar cell can be fabricated and implemented in the laboratory.

ACKNOWLEDGMENT

The author would like to thank to Professor Stephen J. Fonash, Pennsylvania State University, USA, for providing us the Analysis of Microelectronic and Photonic Structures of one dimensional device (AMPS-1D) simulation software.

REFERENCES

- [1] Islam MR, Islam MR, Beg MRA, "Renewable Energy Resources and Technologies Practice in Bangladesh," *Renewable and Sustainable Energy Reviews*, 2008, 12(2):299–343.
- [2] Dennai Benmoussa, Benslimane Hassane, Helmaoui Abderrachid, "Simulation of In_{0.52}Ga_{0.48}N Solar Cell Using AMPS-1D," in *Proceedings of the IEEE 2013 International Renewable and Sustainable Energy Conference (IRSEC)*, pp. 23–26, 7–9 march 2013.
- [3] Nowshad Amin, Michael Tang, Kamaruzzaman Sopian, "Numerical Modeling of the Copper-Indium-Selenium (CIS) Based Solar Cell Performance by AMPS-1D," in *Proceedings of the IEEE 5th Student Conference on Research and Development, 2007. SCORED 2007*, pp. 1–6, 11–12 Dec. 2007.
- [4] J. R. Ray, C. J. Panchal, M. S. Desai, U. B. Trivedi, "Simulation of CIGS Thin Film Solar Cell Using AMPS-1D," *J. Nano- Electron. Phys.*, 3 (2011) No1, P. 747–754.
- [5] R. S. Ohl, "Light Sensitive Electric Device," U. S. Patent 2, 402, 662 (1941).
- [6] Stephen J. Fonash, "Solar Cell Device Physics," Elsevier, 2010.
- [7] Md. Feroz Ali, Md. Faruk Hossain, "Effect of Bandgap of EBL on Efficiency of the p-n Homojunction Si Solar Cell from Numerical Analysis," in *proceeding of IEEE 2015 International Conference on Electrical & Electronic Engineering (ICEEE)*, pp. 245–248, 4–6 Nov. 2015.
- [8] V. Avrutin, N. Izyumskaya, and H. Morko, "Semiconductor solar cells: recent progress in terrestrial applications," *Superlattices and Microstructures*, vol. 49, no. 4, pp. 337–364, 2011.
- [9] A. Luque, S. Hegedus(Eds.), "Handbook of Photovoltaic Science and Engineering," Wiley, UK, 2003.
- [10] X. Wu, J. C. Keane, R. G. Dhere, et al., "16.5%-efficient CdS/CdTe polycrystalline thin-film solar cells," in *Proceedings of the 17th European Photovoltaic Solar Energy Conference and Exhibition (PVSEC '01)*, pp. 995–1000, 2001.
- [11] First Solar, "First solar sets world record for CdTe solar PV efficiency," 2011.
<http://investor.firstsolar.com/releasedetail.cfm?ReleaseID=593994>.
- [12] Saeed Khosroabadi, Seyyed Hossein Keshmiri, "Design of a high efficiency ultrathin CdS/CdTe solar cell using back surface field and backside distributed Bragg reflector," (310.6845) *Thin film devices and applications*; (040.5350) *Photovoltaic*; (350.6050) *Solar energy*; (310.6805) *Theory and design*, Optical Society of America, 2014.
- [13] Alvin D. Compaan, Akhlesh Gupta, Sunghyun Lee, Shanli Wang, Jennifer Drayton, "High efficiency, magnetron sputtered CdS/CdTe solar cells," *Solar Energy* 77 (2004) 815–822.
- [14] Nowshad Amin, Kamaruzzaman Sopian, Makoto Konagai, "Numerical modeling of CdS/CdTe and CdS/CdTe/ZnTe solar cells as a function of CdTe thickness," *Solar Energy Materials & Solar Cells* 91 (2007) 1202 – 1208.
- [15] B.L. Williams, J. D. Major, L. Bowen, L. Phillips, G. Zoppi, I. Forbes, K. Durose, "Challenges and prospects for developing CdS/CdTe substrate solar cells on Mo foils," *Solar Energy Materials & Solar Cells* 124 (2014) 31 –38.
- [16] Mohammadnoor Imamzai, Mohammadreza Aghaei, Nowshad Amin, "Numerical Modeling and Analysis of CdS/Cd_{1-x}Zn_xTe Solar Cells as a Function of CdZnTe Doping, Lifetime and Thickness," in *proceeding of 2013 IEEE 4th International Conference on Photonics (ICP)*, pp. 87–89, 28–30 Oct. 2013.
- [17] S. Khosroabadi, S. H. Keshmiri, S. Marjani, "Design of a high efficiency CdS/CdTe solar cell with optimized step doping, film thickness, and carrier lifetime of the absorption layer," *J. Europ. Opt. Soc. Rap. Public.* 9, 14052 (2014).
- [18] S. Fonash, J. Arch., J. Cuiffi, J. Hou, W. Howland, P. McElheny, A. Moquin, M. Rogosky, T. Tran, H. Zhu, F. Rubinelli, "A Manual for AMPS-1D for Windows 95/ NT a One-Dimensional Device simulation program for the Analysis of Microelectronic and Photonic Structures," The Pennsylvania State University, USA, 1997.
- [19] S.M. Sze, "Physics of Semiconductor Devices," 2nd ed., John Wiley & Sons, New York, 1981, p. 51.
- [20] Arturo Morales-Acevedo, "Can we improve the record efficiency of CdS/CdTe solar cells?," *Solar Energy Materials & Solar Cells* 90 (2006) 2213 – 2220.

BER and CCDF Performance Analysis of DCT-SCFDMA System using Pulse Shaping Filter

Md. Rabiul Hossain

Department of Applied Physics, Electronics &
Communication Engineering
University of Chittagong
Chittagong-4331, Bangladesh
robi_apece@yahoo.com

Kazi Tanvir Ahmmed

Department of Applied Physics, Electronics &
Communication Engineering
University of Chittagong
Chittagong-4331, Bangladesh
tanvir@cu.ac.bd

Abstract—Single carrier frequency division multiple access based on discrete cosine transform (DCT-SCFDMA) is a salient technique for allowing high Quality of Service (QoS) to the users by mitigating the fading of signals. It also provides high data rates in multimedia services. But High peak-to-average power ratio (PAPR) is a technical challenge which cut down the efficiency of RF power amplifiers. In this article, we analyze the bit error rate (BER) and complementary cumulative distribution function (CCDF) using two efficacious filter- raised cosine (RC) and root raised cosine (RRC) filter for different subcarrier mapping techniques e.g. interleaved frequency division multiple access (DCT-IFDMA) and localized frequency division multiple access (DCT-LFDMA). Simulation experiments present that RRC pulse shaping will be more advantageous as far as the lower PAPR performance is concerned and among the subcarrier schemes, performance of DCT-IFDMA with root-raised cosine pulse shaping scheme is superior to that of the other scheme.

Keywords—Pulse shaping, peak-to-average power ratio (PAPR), complementary cumulative distribution function (CCDF), discrete cosine transform, single carrier frequency division multiple access (SC-FDMA).

I. INTRODUCTION

Modulated signals of single-carrier frequency division multiple access (SCFDMA) scheme compared to other multicarrier modulations have a much smaller PAPR. Though pulse shaping enlarges the PAPR of SCFDMA signal, to band-limit the signal available at the transmitter it is needed [1]. As a result, a trade-off is needed between PAPR reduction and spectrum efficiency in conventional single carrier systems. Therefore, over the last several years research has been ongoing trying to design a pulse shaping filter which limits the PAPR without worsening the system's performance keeping on mind that, compared to existing PAPR reduction process, pulse shaping is an effective and simple method. In case of SCFDMA, several novel methods have been introduced to better handle the PAPR problem.

A novel pulse shaping filter with a new parameter β independent from the roll-off factor α is presented in [2] for improved PAPR reduction. This proposed filter increases the

complexity of the system though it maintains the zero inter-symbol interference condition and the same expansion of bandwidth. In [3], a few bit errors are introduced to modify the modulated symbols of each data symbol in a sub-frame to improve the PAPR reduction while the BER performance is slightly trimmed down. Jinwei et al. proposed a technique [4] to use a set of weighting windows with zero excess bandwidth. In this technique a candidate signal block having lowest PAPR is picked out to be transmitted. Side information is not required for data detection at the receiver and this system has potential to be applicable in SCFDMA systems with high order modulations. Above all this advantage, the computational complexity of this system has been increased for the calculation of the candidate signal block.

In this paper, the parametric approach to construction of RC an RRC filter for PAPR reduction on DCT-SCFDMA is presented. At first, DCT-SCFDMA system model is analyzed. The mathematical analysis related to PAPR of SCFDM signal is discussed in section III. Section IV demonstrates the brief background regarding pulse shaping. Simulation parameters and results are shown in section V where BER and CCDF performance of DCT-SCFDMA system has been analyzed.

II. SYSTEM MODEL

A. SC-FDMA System

SC-FDMA can be termed as “linearly pre-coded” orthogonal frequency division multiple access (OFDMA). It is the combination of SC-FDE and FDMA system where the envelope fluctuations are less pronounced than that of the OFDMA system. SCFDMA has relatively lower PAPR than OFDMA due to the pre-coding [5]-[7], which introduces a low cost and effective of power amplifier at transmit section. For downlink, PAPR having value to a great extent may not be a problem but in case of uplink it means a lot. That's why SCFDMA is employed in the LTE uplink. Only in base station complicated frequency domain equalizers are needed whereas it is not required in mobile terminals [8]. At mobile terminals employing SCFDMA, the power amplifiers can be simpler and more power-efficacious than they would be with OFDMA

transmission [9], though equalizer of SCFDMA system is more perplexed than OFDMA.

B. DCT-SC-FDMA System

Real arithmetic only used in Discrete Cosine Transform that reduces the signal processing complexity at great extent [10]. It shows large energy compaction property [11] which leads to a reduction in the effect of ISI while in DFT process it uses complex arithmetic. That's why it suffers from I/Q imbalance problem. A schematic diagram of DCT-SC-FDMA system is depicted in Fig. 1.

The spectral efficiency and length of the CP are same in both DCT-SCFDMA and DFT-SCFDMA system. In DCT, a single set of co-sinusoidal functions $\cos(2\pi m f_\Delta t)$, where $n = 0, 1, \dots, N-1$ and $0 < t < T$ is used. The minimum subcarrier spacing, f_Δ required to satisfy the orthogonality condition

$$\int_0^T \sqrt{\frac{2}{T}} \cos(2\pi k f_\Delta t) \sqrt{\frac{2}{T}} \cos(2\pi n f_\Delta t) dt = \begin{cases} 1, & k = n \\ 0, & k \neq n \end{cases} \quad (1)$$

is $1/T$.

At time instants $T(2n+1)/2N$, sampling the continuous-time signal gives a discrete-time sequence after the discrete cosine transform [10] as follows:

$$X(k) = \sqrt{\frac{2}{N}} \beta(k) \sum_{n=0}^{N-1} x(n) \cos\left(\frac{\pi k(2n+1)}{2N}\right), \quad k = 0, 1, \dots, N-1 \quad (2)$$

Where,

$x(n)$ denotes the n^{th} sample of the input signal and

$$\beta(k) = \begin{cases} \frac{1}{\sqrt{2}}, & k = 0 \\ 1, & k = 1, 2, \dots, N-1 \end{cases} \quad (3)$$

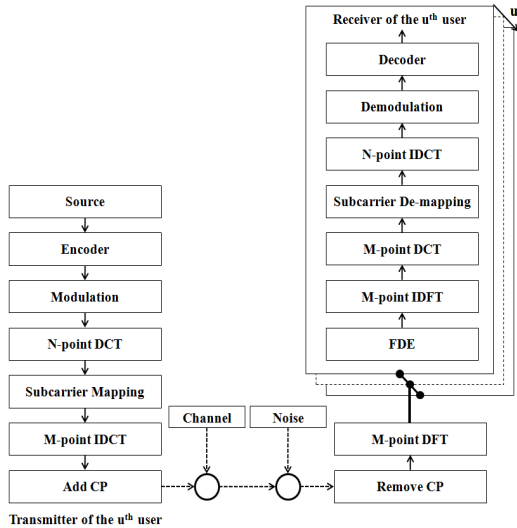


Figure 1. Block Diagram of DCT-SCFDMA transceiver

After the inverse DCT (IDCT), the discrete-time signal can be written as [11]:

$$x(n) = \sqrt{\frac{2}{N}} \sum_{k=0}^{N-1} X(k) \beta(k) \cos\left(\frac{\pi k(2n+1)}{2N}\right), \quad n = 0, 1, \dots, N-1 \quad (4)$$

M frequency domain symbols is produced via M-point DCT where it modulates M out of N orthogonal subcarriers over a bandwidth

$$W_{\text{Channel}} = N \cdot f_\Delta \quad (5)$$

If Q denotes the bandwidth spreading factor, i.e., $Q = N/M$ then, the SCFDMA system can handle up to Q orthogonal source signals without the co-channel interference with each source occupying a different set of M orthogonal subcarriers.

III. PAPR OF SC-FDMA SIGNALS

PAPR is the prime indicant of the transmitted power efficiency. For a particular transmission block, PAPR of transmitted signal is termed as the ratio of peak power to average power [8], [12], [13]. The pass-band transmit signal, $x(t)$, is given as [8]-

$$x(t) = e^{j\omega_c t} \sum_{m=0}^{M-1} \tilde{x}_m r(t - m\tilde{T}) \quad (6)$$

Where,

$r(t)$ = baseband pulse

ω_c = carrier frequency

$\tilde{T}$ = symbol duration

$\tilde{x}_m$ = time domain symbols after IDFT

PAPR for the transmit signal $x(t)$, is written as follows-

$$PAPR = \frac{\max_{0 \leq t \leq M\tilde{T}} |x(t)|^2}{\frac{1}{M\tilde{T}} \int_0^{M\tilde{T}} |x(t)|^2 dt} \quad (7)$$

The symbol will bring out the exact PAPR value in both sampling and continuous case when pulse shaping is not implied in the system. The reason behind this is the signal is modulated over a single carrier in SC-FDMA system. Considering no pulse shaping technique, PAPR with symbol's sampling rate is given as follows [8]-

$$PAPR = \frac{\max_{m=0,1,\dots,M-1} |\tilde{x}_m|^2}{\frac{1}{M} \sum_{m=0}^{M-1} |\tilde{x}_m|^2} \quad (8)$$

When pulse shaping process is induced at transmit section, the totality of PAPR is computed by aggregating the process of modulation with pulse shaping. PAPR of filter relies on the excess bandwidth and the supported symbol's number in a filter [14]. In SCFDMA system, PAPR maintains a proportional behavior with the number of symbols for a pulse shaping filter while maintain an inverse relationship with the excess bandwidth.

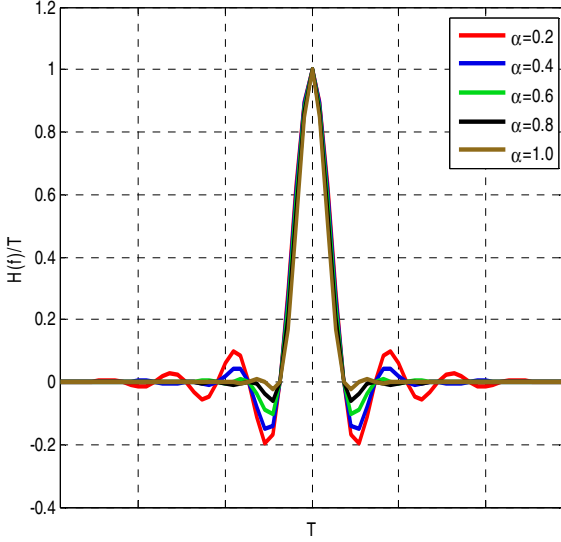


Figure 2. Impulse response of the RC filter.

IV. PULSE SHAPING FILTER

In a band-limited channel an impulse response $h(t)$ must fulfill the Nyquist-I criterion for proper transmission to ensure that there is no inter-symbol interference [15]. A pulse $h(t)$ having zero ISI is [16]-

$$h(kT) = \begin{cases} 1, & k = 0 \\ 0, & k = \pm 1, \pm 2, \dots \end{cases} \quad (11)$$

Where,

k = symbol time index

T = symbol period

Frequency characteristic of Eq. (11) is written as follows-

$$\frac{1}{T} \sum_{m=-\infty}^{+\infty} H\left(f + \frac{m}{T}\right) = 1 \quad (12)$$

Where,

$H(f)$ = Fourier transform of $h(t)$.

Standard raised-cosine (RC) filter has a linear phase and it justifies the Eq. (11) and Eq. (12). The spectral magnitude of this filter is [16]-

$$H_{RC}(f) = \begin{cases} T, & 0 \leq |f| \leq \frac{1-\alpha}{2T} \\ \frac{T}{2} \left[1 + \cos \left\{ \frac{\pi T}{\alpha} \left(|f| - \frac{1-\alpha}{2T} \right) \right\} \right], & \frac{1-\alpha}{2T} \leq |f| \leq \frac{1+\alpha}{2T} \\ 0, & |f| \geq \frac{1+\alpha}{2T} \end{cases} \quad (13)$$

Here α is the roll-off factor [$0 < \alpha < 1$], which define the bandwidth of filter, $B = (1 + \alpha)/(2T)$.

TABLE I. PARAMETERS USED IN SIMULATION

Parameter	Value
Transmission Bandwidth	5MHz
Modulation	QPSK/16QAM
Number of Subcarriers	512
Input Data Block Size	128
Spreading Factor, Q	4
Subcarrier Allocation	Interleaved/Localized
Equalization	MMSE
Subcarriers Spacing	9.765625KHz

Inverse Fourier transform of Eq. (13) gives the corresponding impulse response which is shown in Fig.2 and it is expressed as follows-

$$h(\tau) = \sin c(\tau) \frac{\cos(\pi\alpha\tau)}{1 - 4\alpha^2\tau^2} \quad (14)$$

where τ is the normalized time, equal to $\tau = t/T$.

The root raised-cosine (RRC) filter has vestigial symmetry property. It is an implementation of a low-pass Nyquist filter, whose spectral magnitude is,

$$H_{RRC}(f) = \sqrt{H_{RC}(f)} \quad (15)$$

The impulse response of this filter is given by [17]-

$$h(\tau) = \frac{\sin[\pi\tau(1-\alpha)] + 4\alpha\tau \cos[\pi\tau(1+\alpha)]}{\pi\tau(1-16\alpha^2\tau^2)} \quad (16)$$

V. SIMULATION PARAMETERS AND RESULTS

In our simulation, we analyzes the PAPR and BER of the DCT-SCFDMA signals with the standard raised-cosine filter and root raised-cosine filter using Minimum Mean Square Error (MMSE) equalizer to verify the effectiveness of this method. In Table-I the simulation parameters implemented in the DCT-SCFDMA system using localized and interleaved allocation are shown.

A. BER Performance

The performance gain is better for the DCT-SCFDMA system over the DFT-SCFDMA and the OFDMA systems at a certain BER. Desired symbol inherits more energy through the distribution of DCT operation while less energy is distributed to the interference [11]. Therefore, the desired symbol suffers

less interference coming from neighboring symbols in the DCT-SCFDMA system. Thus, the ISI is reduced in the DCT-SCFDMA system, which reduces the BER.

Fig. 3 picturizes the BER performance of the DCT-SCFDMA with different subcarriers mapping schemes and modulation formats using a vehicular-A outdoor channel [18]. It is clear from figure a significant improvement in BER performance for interleaved mode of DCT-SCFDMA using QPSK has been achieved. The gain is almost 4.5dB which is much better when compared to that of DCT-LFDMA system using QPSK. In case of 16QAM modulation format DCT-IFDMA also shows its deftness in improving the BER performance over the localized subcarrier mapping by about 4.4dB.

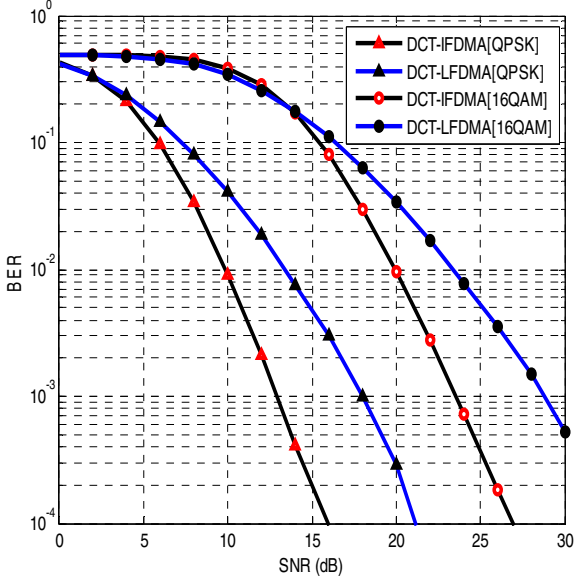


Figure 3. Comparison of BER performance between different subcarrier mapping of DCT-SCFDMA using QPSK and 16QAM.

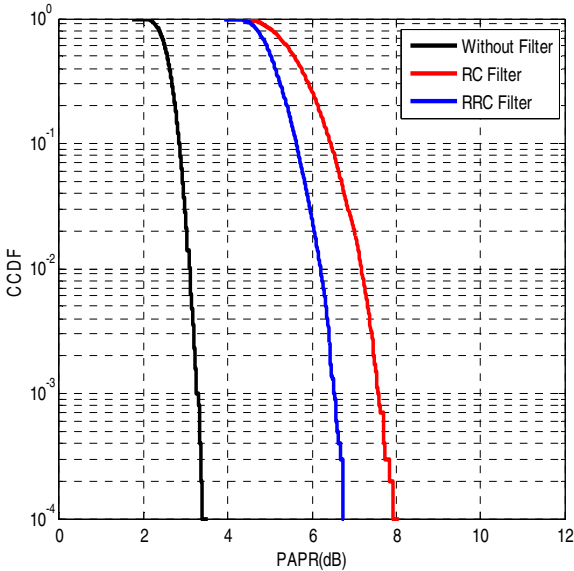


Figure 4. CCDF of PAPR for DCT-IFDMA system for $\alpha=0.35$ using 16QAM.

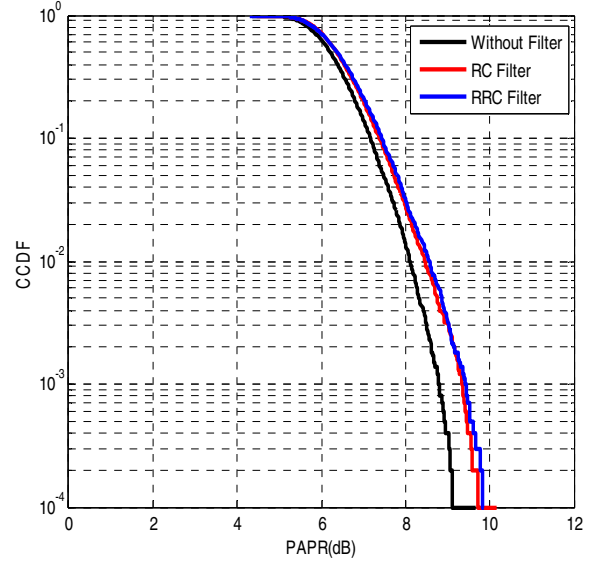


Figure 5. CCDF of PAPR for DCT-LFDMA system for $\alpha=0.35$ using 16QAM.

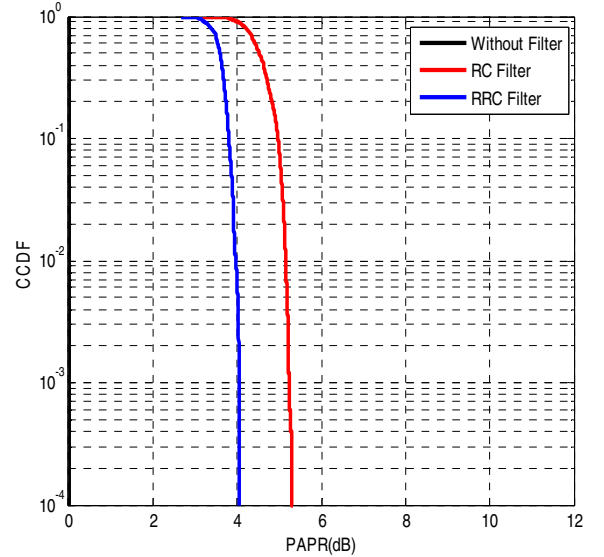


Figure 6. CCDF of PAPR for DCT-IFDMA system for $\alpha=0.35$ using QPSK.

B. CCDF Performance

To show the effectuation of pulse shaping process, each transmitted symbol waveform is convoluted with the respective truncated pulse shaping filter (raised-cosine and root raised-cosine) [20]. The PAPR value of every block is represented in terms of CCDF. Additionally, the probability for which the signal's PAPR surpass a certain threshold value $PAPR_0(P_r\{PAPR > PAPR_0\})$ is computed by CCDF [13].

QPSK and 16-QAM are the applied modulation format in the 3GPP LTE uplink and a bandwidth allocation of 5MHz is implemented [19]. In wireless communication systems pulse shaping filters are widely used for avoiding ISI and the implemented roll-off factors α is in the range of $0.22 < \alpha < 0.5$ [21]. In all simulation, we used the typical roll-off factors of 0.35.

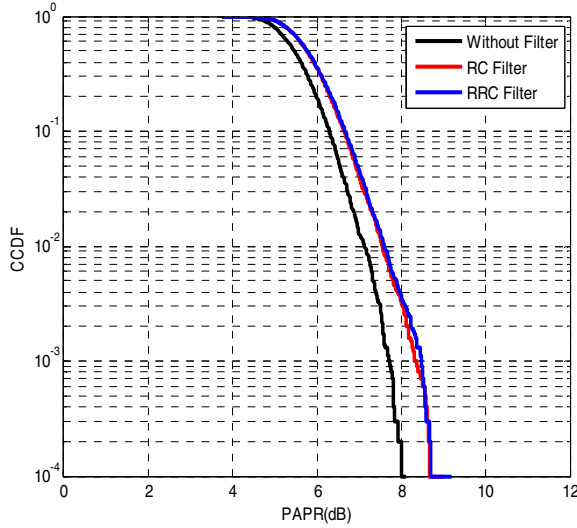


Figure 7. CCDF of PAPR for DCT-LFDMA system for $\alpha=0.35$ using QPSK.

TABLE II. PAPR VALUES AT $\text{CCDF}=10^{-4}$ FOR DIFFERENT SUBCARRIER MAPPING TECHNIQUE IN DCT-SCFDMA SYSTEM USING 16QAM, QPSK AND ROLL-OFF FACTOR=0.35

Pulse shaping Filter	PAPR of DCT-SCFDMA (dB)			
	16QAM		QPSK	
	IFDMA	LFDMA	IFDMA	LFDMA
None	3.5	9.5	0	8.1
RC	8	9.7	5.2	8.7
SRRC	6.7	9.8	4	8.7

The effect of the pulse-shaping filter (RC and RRC) on the PAPR performance in the DCT-SC-FDMA system has been studied for the different subcarriers mapping techniques and modulation formats and the results are shown in Fig. 4, 5, 6 and 7. From these figures, it is obvious that the DCT-IFDMA system has a lower PAPR than that of the DCT-LFDMA and higher order modulation schemes have an impact on the performance of PAPR of SC-FDMA system. Simulations results showed that the effectiveness of the RRC pulse shaping filter outperforms and has a much simpler impulse response expression, therefore, adding straightforwardness and flexibility into the system. In Table-II, a full summary of comparison regarding PAPR for 16QAM and QPSK modulation format is shown.

It is affirmed that pulse shaping is must in a band-limited channel for proper transmission. Following this note and the BER analysis, among all the process DCT-IFDMA using SRRC filter is the efficient one which shows better PAPR reduction performance as well as good BER.

VI. CONCLUSION

From the analysis results above, it shows that the root raised cosine (RRC) filter provides better performance in PAPR reduction compared to that of the raised cosine filter for DCT-IFDMA whereas PAPR of the DCT-LFDMA system is nearly the same for both filter. Moreover, it is noted that the higher order modulation schemes have an impact on the PAPR of SC-FDMA. From the simulation results, it can also be asserted that the modulation formats have significant impact on the PAPR of the DCT-SC-FDMA system. Compared to MU-MIMO OFDMA better diversity can be achieved by combining SCFDMA with MU-MIMO (multi user multiple input and output), due to the properties of SCFDMA which is the prime motivating scope for our future work.

ACKNOWLEDGEMENT

The authors would like to acknowledge support for this work provided by ICT Fellowship, 2014-2015 under the Information and Communication Technology Division, Ministry of Post, Telecommunications and IT, Government of the People's Republic of Bangladesh.

REFERENCES

- [1] Huang, G.; Nix, A.; and Armour, S., "Impact of radio resource allocation and pulse shaping on PAPR of SC-FDMA signals," *IEEE 18th International Symposium on PIMRC*, pp. 1–5, September 2007.
- [2] Azurdia-Meza, Cesar A.; Lee, Kyujin and Lee, Kyesan; "PAPR Reduction in Single Carrier FDMA Uplink by Pulse Shaping Using a β - α Filter," *Wireless Personal Communications*, vol. 71, no. 1, pp. 23–44, July 2013.
- [3] J. Ji, G. Ren and H. Zhang, "PAPR reduction in coded SC-FDMA systems via introducing few bit errors," *IEEE Commun. Lett.*, vol. 18, no. 7, pp. 1258–1261, July 2014.
- [4] Ji, J.; Ren, G. and Zhang, H., "PAPR Reduction of SC-FDMA Signals via Probabilistic Pulse Shaping," *IEEE Trans. on Vehicular Technology*, vol. PP, no. 99, pp. 1, October 2014.
- [5] G. Berardinelli, L.; A. Ruiz de Temino; S. Frattasi; M. Rahman, and P. Mogensen, "OFDMA vs. SC-FDMA: performance comparison in local area IMT-A scenarios," *IEEE Wireless Commun.*, vol. 15, no. 5, pp. 64–72, October 2008.
- [6] Cruz-Roldan, F.; Dominguez-Jimenez, M.E.; Vidal, G.S.; Amo-Lopez, P.; Blanco-Velasco, M.; Bravo-Santos, A., "On the use of discrete cosine transforms for multicarrier communications," *IEEE Trans. on Signal Processing*, vol. 60, no. 11, pp. 6085–6090, July 2012.
- [7] M. Rinne; M. Kuusela; E. Tuomaala; P. Kinnunen; I. Kovacs; K. Pajukoski and J. Ojala, "A performance summary of the Evolved 3G (E-UTRA) for voice over Internet and best effort traffic," *IEEE Trans. Veh. Technol.*, vol. 58, no. 7, pp. 3661–3673, Sep. 2009.
- [8] Myung, H. G., Lim, J., & Goodman, D. J., "Peak-to-average power ratio of single carrier FDMA signals with pulse shaping," *IEEE 17th International Symposium on PIMRC*, pp. 1–5, Sept. 2006.
- [9] J. Ji and G. Ren, "A new modified SLM scheme for wireless OFDM systems without side information," *IEEE Signal Process. Lett.*, vol. 20, no. 11, pp. 1090–1093, Nov. 2013.
- [10] P. Tan and N. C. Beaulieu, "A comparison of DCT-based OFDM and DFT-based OFDM in frequency offset and fading channels," *IEEE Trans. Commun.*, vol. 54, no. 11, pp. 2113–2125, November 2006.
- [11] A. V. Oppenheim; R. W. Schaffer and J. R. Buck, *Discrete-Time Signal Processing*, 2nd ed., Prentice Hall, Upper Saddle River, NJ, 1999.

- [12] Cesar A. Azurdia-Meza, "PAPR reduction in SC-IFDMA uplink system using parametric pulses," *IEEE Latin-American conf. on commun.*, pp. 1-5, Nov. 2013.
- [13] Nee, R. V., and Prasad, R., "OFDM for wireless multimedia communications. Boston," Massachusetts: Artech House, 2000.
- [14] Chatelain, B. and Gagnon, F., "Peak-to-average power ratio and intersymbol intersymbol interference reduction by Nyquist pulse optimization," *IEEE 60th Vehicular Technology Conference*, vol. 2, pp. 954–958, Sept. 2004.
- [15] Scanlan, J. O. (2007). Pulses satisfying the Nyquist criterion. *Electronics Letters*, 28(1), 50–52.
- [16] Mohapatra, S. and Das, S., "Peak-to-average power reduction (PAPR) by pulse shaping using a modified raised cosine filters," *Annual IEEE India Conference*, pp. 1–4. Dec. 2009.
- [17] T. S. Rappaport, *Wireless Communications Principles and Practice*, 2nd ed., Pearson Education, Indianapolis, IN, 2002.
- [18] 3rd Generation Partnership Project and 3GPP TS 25.101, "Technical Specification Group Radio Access Network; User. Equipment (UE) Radio Transmission and Reception (FDD). (Release 7), Section B.2.2," September 2007.
- [19] Myung, H. G., "Introduction to single carrier FDMA," *15th European Signal Processing Conference*, pp. 2144–2148, Sep. 2007.
- [20] Myung, H. G., Lim, J., & Goodman, D. J., "Single carrier FDMA for uplink wireless transmission," *IEEE Vehicular Technology Magazine*, vol. 3, no. 1, pp. 30-38, Sep. 2006.
- [21] Du, K. L., & Swamy, M. N. S., *Wireless communication systems*. Cambridge, UK: Cambridge University Press, 2010.

Adaptive Neuro-Belief Rule Based Diabetes Diagnosis System

Saifur Rahaman

Dept. of Computer Science & Engineering

International Islamic University Chittagong, Chittagong, Bangladesh

University of Chittagong, Chittagong, Bangladesh

saifurcubd@gmail.com, saifurcse@iiuc.ac.bd,

Abstract—Diabetes is the most regular disease in medical science. It can affect the organs of the human body. Due to the handling both uncertain diabetic medical and clinical data to diabetes diagnosis system is a complex problem. However, a computer-base diagnosis system for diabetes would help to enhance the accuracy of the diagnosis and reduce the time and cost. This paper describes an effective new approach “Adaptive Neuro-Belief Rule Based System (ANBRBS)” with Evidential Reasoning (ER) to diagnose diabetes, which can reduce the errors and medical uncertainties. This paper used the medical and clinical real data to implement and test of this proposed system. It has been observed that, this new adaptive methodology provides more reliable diabetes diagnosis result in percentage and recommendations.

Keywords— *Diabetes, Neural Network; Belief Rule Base; ER Algorithm, Neuro-BRB, Diagnosis*

I. INTRODUCTION

Diabetes is a metabolic disease. 100 million people in the world are affected by diabetes [1] and it's the cause of 1.1 million deaths every year, 80% of which found in lower and middle income countries [2]. Now 32.7 million diabetic patients are estimated in India and this amount is expected to twice by 2025. Prevalence is expected to double worldwide in the next 20 years [3] and enlarged by 150% in lower and middle income countries [2][3][4].

A same situation having in other Asian countries [4][5]. The diabetic patient rate is enlarging from year to year. Diabetes disease is not diagnose and treated properly at an early stage of patients, it would affect people and lead to various diseases like cardiovascular disease, visual impairment, leg amputation and renal failure etc. However diabetes diagnosis at right time is a difficult to Physicians due to lack of subjective specialists or inexperience with the previous diabetic cases. The manual diabetes diagnosis system is does not sound due to the various types of uncertainty come from the medical and clinical diabetic data and knowledge. In general, the manual diabetes diagnosis system is time delay and expensive. An automated medical diagnosis system can enhance the accuracy of diagnosis and reduce the expenses of diagnosis.

There are numerous researches have been found in various methodologies such as Rule Based, ANN, Decision Tree and BRB for diabetes diagnosis. Artificial neural networks based Diabetes mellitus forecast [13] having the biggest challenge was the missing values in the data set. Rule base classification system for Diabetic Patient using Cascaded K-Means and Decision Tree cannot handle the

various types of uncertainty from diabetic data. A Decision Support System to analyze elder diabetic patients using Decision Tree. This system does not consider younger patients [14]. A diabetes expert system [16] does not consider diabetes type 1. Belief Rule based (BRB) CDSS for Heart Failure Suspicion [17]. Diabetes diagnosis system by using OLAP and data mining integration can't handle the all types uncertainties [15]. Pima Indian Diabetes diagnosis system [18] is used the general regression neural networks to classify the patient as diabetic or not diabetic. It can't recommend any reliable value like a patient having the percentage of diabetes. The above survey highlights that the development of a diabetes diagnosis system, which aims to better diagnosis result than other existing systems by handling all types of uncertainties..

Therefore, it can be inferred that to handle the uncertainties of existing diabetes diagnosis and generate reliable results is needed to processed by using refined knowledge base schemata and inference system[6][7]. Hence, this research paper proposes to adopt a new adaptive methodology namely Adaptive Neuro-Belief Rule-Base (BRB) System with ER to the design and development of the diabetes diagnosis system. This methodology consists of two parts are neural network algorithm and BRB inference system with ER. BRB is the expanded form of the traditional IF-Then rule-base, which belief degree can handle. The ER approach [7] deals with MADA problems having both numeric and subjective attributes in the inference process.

This research paper is presents to develop an adaptive neuro-belief rule based system for diabetes diagnosis.. Pima Indian Diabetes dataset [9] has been used to development and test of this paper proposed system. It would find that the results of this proposed system are more accuracy than the other diabetes diagnosis systems.

This paper consists of five sections, including the preceding introduction. The second section presents an overview of the neural network and BRB system. The third section presents the design and implementation of the adaptive neuro-BRB system for diabetes diagnosis. Results and discussion are described in section four. Section five concludes with some final comments and suggestions for future researches.

II. NEURAL NETWORK AND BELIEF RULE BASE

Artificial neural networks used the back propagation algorithm using multi layer perceptions (MLA). This

section presents the description of ANN and belief rule based (BRB) inference system.

A. Artificial Neural Network (ANN)

The construction of the neural network involves three different layers with feed forward architecture. This is the most popular network architecture in use today. The input layer of this network is a set of input units, which accept the elements of input feature vectors. The input units (neurons) are fully connected to the hidden layer with the hidden units. The hidden units (neurons) are also fully connected to the output layer. The output layer supplies the response of neural network to the activation pattern applied to the input layer. The information given to a neural net is propagated layer-by-layer from the input layer to output layer through (none) one or more hidden layers.

Important issues in Multi Layer Perceptions (MLP) design include specifications of the number of hidden layers and the number of units in these layers. The number of input and output units is defined by the problem the number of hidden units of use is far from clear. That is the amount of hidden layers and their neurons is more difficult to determine. A network with one hidden layer is sufficient to solve most tasks. In this paper, neural network used seven inputs in the input layer, one hidden layer and one node in the output layer.

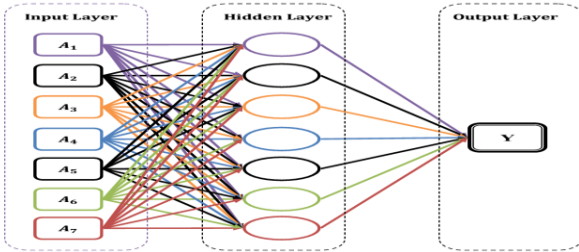


Figure 1. Artificial Neural network with Seven inputs, One hidden layer and one output

B. Belief Rule Base (BRB) inference system

The Belief Rules are the main components of a BRB that consists of three extra features such as belief degrees, rule weight and attributes weight. These belief degrees are the expanded form of general IF-THEN rules. The antecedent attributes of BRB takes referential values and consequent with belief degrees structure [20]. The Belief Rule Base (BRB) is used to represent the knowledge base schema of an expert system that allows handling of uncertain data. A belief rule can be elucidated as following way [7].

$$R_k : \begin{cases} \text{IF } (P_1 \text{ is } A_1^k) \cap (P_2 \text{ is } A_2^k) \cap \dots \cap (P_{T_k} \text{ is } A_{T_k}^k) \\ \text{THEN } \{(C_1, \beta_{1k}), (C_2, \beta_{2k}), \dots, (C_N, \beta_{Nk})\} \end{cases} \quad (1)$$

$$R_k : \left(\beta_{jk} \geq 0, \sum_{j=1}^N \beta_{jk} \leq 1 \right) \text{ with a rule weight } \theta_k, \text{ attribute weights } \delta_{k1}, \delta_{k2}, \delta_{k3}, \dots, \delta_{kT_k} \quad k \in \{1, \dots, L\}$$

Where the antecedent attributes of k th rule are represent by $P_1, P_2, P_3 \dots P_{T_k}$. $A_i^k (i = 1, \dots, T_k, k = 1, \dots, L)$ represents one of the referential values of the i th antecedent attribute P_i in the k th rule. C_j represents one of the consequent reference values of above belief rule. $\beta_{jk} (j = 1, \dots, N, k = 1, \dots, L)$ is the belief degree of the consequent reference value C_j of BRB that believed to be true. If $\sum_{j=1}^N \beta_{jk} = 1$ the k th rule is said to be complete,

otherwise, it is incomplete. T_k is presents the total number of antecedent attributes used in k th rule. L is considered the number of all belief rules in the BRB. N is the number of all possible referential values of consequents in a belief rule. An example of belief rule in the diabetes diagnosis system can be defined as follows-

$$R_k : \begin{cases} \text{IF } A_1 \text{ is High AND } A_2 \text{ is High AND } A_3 \text{ is High AND } A_4 \text{ is High AND} \\ A_5 \text{ is High AND } A_6 \text{ is Medium} \\ \text{THEN Diabetes is } \{(High, (0.860)), (Medium, (0.140)), (Low, (0.000))\} \end{cases} \quad (2)$$

Where $\{(High, 0.860), (Medium, 0.140), (Low, 0.000)\}$ is a belief degrees distribution with "Diabetes" consequent of belief rule as present as in Eq. (2). "High", "Medium" and "Low" are the referential values of consequent "Diabetes" in this research. The belief distribution present that the degree of belief of "High" is 84%, 16% degree of belief for "Medium", while 0% degree of belief is considered with "Low". In this belief rule, the total degree of belief is $(0.860+0.140+0.000) = 1.000$; hence, the rule is complete

The BRB inference system of this research methodology consists of input transformation, rule activation weight calculation, rule belief degree update, and aggregation by using ER algorithm [22]. The inference engine of this system architecture works in the following ways [17] – i) First take the input data from the user by the interface layer, this input data are transformed in its attributes referential values of BRB. ii) Calculate the activation weights of all BRB rules by the using (A.2) equation. iv) Update the belief degree of consequence in belief rules by using (A.3) and v) Finally aggregates all rules by using analytical ER algorithm (A.4). After all, system inference engine generates the overall inference result by using utility equation (A.5).

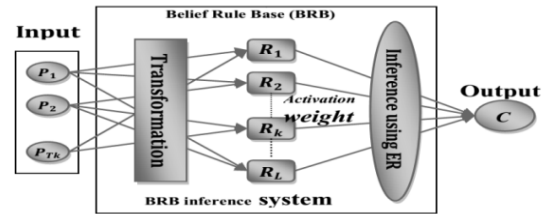


Figure 2. BRB Inference System
Figure 2 illustrates the BRB inference system, which consists of mention BRB, input transformation, activation weight calculation; update the belief rules and aggregation by using ER.

III. ADAPTIVE NEURO-BELIEF RULE BASE SYSTEM FOR DIABETES DIAGNOSIS

Adaptive Neuro-BRB system is a combination of neural network and belief rule based system. This expert system adapts the layered based system architecture that represents how its components consisting of inputs, process, and outputs are organized by using Adaptive Neuro-BRB and ER. This proposed adaptive neuro-Belief Rule Base system architecture consists of three basic layers namely interface layer, application processing layer and data management layer [30]. This research considers a numerical multi-input single-output system, which can always be separated into a set of single-output systems [24]. The adaptive neuro-belief rule-based system with ER methodology is well adapted to develop a multi-input single-output system. The diabetes diagnosis system is a multi-input single-output

system. Hence this approach employs to develop the diabetes diagnosis system.

The adaptive Neuro-BRB diabetes diagnosis system with ER consists of patient information and three layers system architecture. This section presents the diabetic information, factors, and system architecture along with design and implementation.

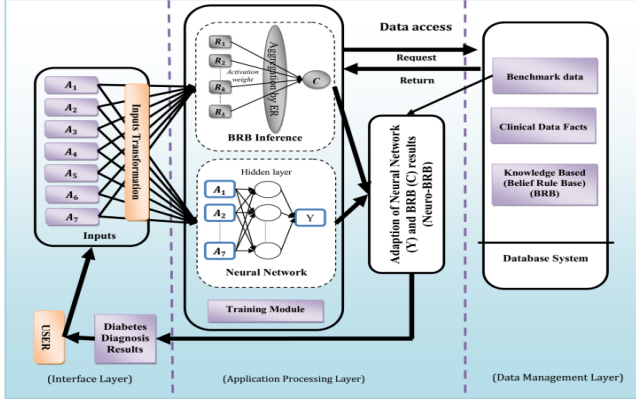


Figure 3. System Architecture of Neuro-BRB Diabetes Diagnosis System

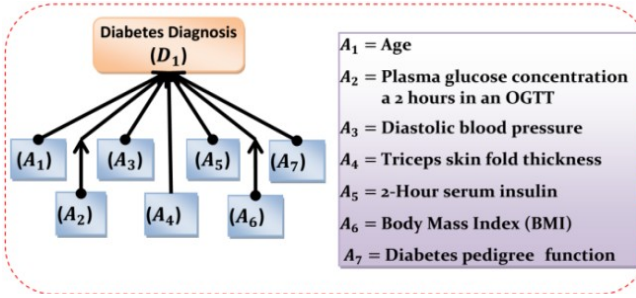


Figure 4. Hierarchical BRB Framework and antecedent attributes for diabetes Diagnosis System

Figure 3 illustrates the main components of the three-layered architecture of the adaptive Neuro-BRB diabetes diagnosis system along with the components as mentioned above.

A. Diabetic Information and factors

The Age (A_1), plasma glucose concentration a 2 hours in an oral glucose tolerance test (OGTT) (A_2), diastolic blood pressure (A_3), Triceps skin fold thickness (A_4), 2Hour serum insulin (A_5), Body mass index (BMI) (A_6), and Diabetes pedigree function (A_7) that were taken from the patients are considered the input facts and input antecedent attributes for this proposed adaptive neuro-BRB diabetes diagnosis system. The input antecedent names along with their referential values present in Table A.I.

B. System Analysis Design, Architecture and its Implementation

The system architecture of adaptive Neuro-BRB diabetes diagnosis system consists of three layers mention in above. Figure 3 and 4 illustrates the BRB framework with input antecedent attributes and system architecture of the adaptive Neuro--BRB diabetes diagnosis System.

Interface layer - used to get the seven diabetes referential values from the patient or from the physician as inputs and to show the diabetes diagnosis outputs in percentage. It facilitates the acquiring of input (antecedent attributes of BRB) as diabetes clinical data, signs, symptoms and risk factors (BRB framework in Figure 1) This data are distributed over the referential values associated with each antecedent attribute by taking account of diabetes expert belief degree [26] and these seven diabetic data also input into the ANN component of application layer. The system interface also enables the displaying of the system results as the patient is diabetic or not with the recommendation in percentage.

Application layer- this layer having main components are BRB knowledge base and ER inference system, neural network and adoption process of BRB ER with ANN. it's also process to transform input, rule activation, rule update and rule aggregation by ER and generate neural network outcomes. Finally, adopt the neural network outcomes with BRB inference system output. The processing of this layer work as following way-

First, the knowledge base of the BRB inference system construct by using belief degree structured BRB which store and organized by data management layer. The BRB framework along with antecedents attributes of this research (by taking account of "Pima Indian Diabetes dataset") [9] have been developed and present in Figure 4. From this framework, it can be stated that input data of diabetes diagnosis include seven antecedent attribute (A_1 - A_7) are mentioned previously. A BRB can be designed in many ways [28]. The initial BRB of this system have been constructed by the domain expert and benchmark dataset. This system BRB consists of only one consequent as "diabetes (D_1)". All antecedent attributes consist of three referential values as high, medium and Low. Hence, this entire BRB system consists of 2187 (L) belief rules as illustrated in Table I. It is considered that all belief rules have equal rule weight and all antecedent attributes have equal weight that is 1.00.

An example belief rule of this BRB diabetes diagnosis system is given below-

R2: IFA_1 is High $AND A_2$ is High $AND A_3$ is High $AND A_4$ is High $AND A_5$ is High $AND A_6$ is High $AND A_7$ is Medium **THEN** Diabetes (D_1) is {High (0.86), Medium (0.14), Low (0.00)}

TABLE I. INITIAL BELIEF RULE BASE FOR NERO-BRB DIABETES DIAGNOSIS SYSTEM

Rule ID	Rule Weight	IF Antecedent attributes (A_1 - A_7)							Then Consequent Diabetes (D_1)		
		A_1	A_2	A_3	A_4	A_5	A_6	A_7	High (H)	Medium (M)	Low (L)
R1	1.00	High	High	High	High	High	High	High	1.000	0.000	0.000
R2	1.00	High	High	High	High	High	High	Medium	0.860	0.140	0.000
R3	1.00	High	High	High	High	High	High	Low	0.860	0.000	0.140
.....											
R2186	1.00	Low	Low	Low	Low	Low	Low	Medium	0.000	0.140	0.860
R2187	1.00	Low	Low	Low	Low	Low	Low	Low	0.000	0.000	1.000

Above belief rule distributes the degrees of belief into the Diabetes (D1) consequent's referential values. Consequently, it can assign the initial degrees of belief to the diabetes consequent in the rule as "0.860" to "High", "0.140" to "Medium", "0.00" to "low" as shown in rule R2.

Second, The BRB inference system of this architecture of consists of a four step process such as take the input and its transformation in referential values of consequent, rule activation weight calculation by using (A.2), rule update by using (A.3), and aggregation of the rules of a BRB by using the analytical

IV. RESULT AND DECISION

This research proposed system input antecedent's data in the BRB framework (Figures 2 and 3) and ANN have been collected from the female patients of Pima Indian heritage [14]. There are seven input attributes of diabetes data have been used in this adaptive Neuro-BRB diabetes diagnosis system to diagnose the diabetes in percentage. The input data form same data set are used to generate the diabetes diagnosis results in manual system and BRB expert system (BRBES) are also shown in Table II.

TABLE II. DIABETES DIAGNOSIS RESULT BY BRB, EXPERT AND NEURO-BRB SYSTEM

ID	Age	Plasma glucose concentration a 2 hours in an OGTT	diastolic blood pressure (mm Hg)	Triceps skin fold thickness (mm)	2 Hour serum insulin (mu U/ml)	Body mass index (BMI)	Diabetes pedigree function	Actual Diabetes Diagnosis Outcome (Benchmark)	Diabetes Diagnosis using BRBES	Expert Opinion	Diabetes Diagnosis using Neuro-BRB (This system)
P1	50	148	72	35	0	33.60	0.627	1	76.41%	70%	78.10%
P2	31	85	66	29	0	26.60	0.351	0	32.12%	30%	35.04%
P3	32	183	64	0	0	23.30	0.672	1	85.40%	80%	90.00%
P4	21	89	66	23	94	28.10	0.167	0	43.31%	45%	42.45%
P5	33	137	40	35	168	43.10	2.288	1	91.85%	90%	94.00%

ER algorithm (A.4). [27]. After that BRB inference system generate the BRB inference output (C_m) using utility equation (A.5).

Third, the seven inputs antecedent attributes used as inputs of the input layer of neural networks and it generates the one output (Y_m) in the outputs layer. This neural network of this system illustrate at Figure 1 and Figure 3. This network use only one hidden layer and one output layer. The diabetes diagnosis benchmark data (B_m) are used as target data on this network. The output of this network is also converted in to percentage value.

Fourth (optional), The second part of this layer is to build the BRB training module to learn the initial belief rule base (BRB) of the data management layer. Its aim to find optimal BRB parameters set $(\theta_k, \delta_i, \beta_{jk})$ by minimizing the discrepancy between the system results and the sampled data by using (A.6). [30].

Finally, the output of the BRB inference system and neural network are adopted in a single value output by using following algorithm 1.

Algorithm 1: Special Rule Based Algorithm
Input : BRB inference output(C_m) and Neural network output (Y_m)
Output: Final Neuro-BRB system result $O((A_{1...7})_m)_m$
BEGIN
BRB inference output := (C_m);
Neural Network := (Y_m)
Diabetes diagnosis Benchmark value $B_m := \{0,1\}$
IF $C_m > Y_m$
THEN $O((A_{1...7})_m)_m = 1/2 (Y_m + C_m) + \{1/2 (C_m - Y_m)\}B_m$
ELSE
$O((A_{1...7})_m)_m = 1/2 (Y_m + C_m) + \{1/2 (Y_m - C_m)\}B_m$
END
RETURN Neuro-BRB system result $O((A_{1...7})_m)_m$
END

Data management layer – manage the initial Belief Rule-Based (BRB) and diabetes diagnosis facts including clinical data and actual benchmark data of diabetes diagnosis.

We have considered the actual diabetes diagnosis results as the benchmark data to compare the adaptive Neuro-BRB diabetes diagnosis system results with other diabetes diagnosis tools and train this proposed system. This benchmark data are one (1) or zero (0). The data one represents the patient having diabetes and zero data represent the patient non-diabetic as shown in Table II. The data set consists of 768 samples. For simplicity only five patients data set are presented in Table II.

The column 12 of Table II illustrates this adaptive Neuro-BRB diabetes diagnosis system results in percentage and column 10 presents the BRBES' generated output in percentages. The column 11 gives the expert opinion (known as manual system) on diabetes diagnosis also in percentage. It is observed that the Adaptive BRB diabetes diagnosis system results are more realisable than other system tools like manual system and BRBES.

V. CONCLUSION

This paper has been proposed new adaptive methodology as adaptive neuro-BRB system with ER approach and explained the design, development and application of proposed system methodology for diabetes diagnosis to account of signs, symptoms, risk factors and data of diabetes. This diabetes diagnosis system can handle the various types of uncertainties found in diabetes medical and clinical data. This system is the adaption of BRB inference system and artificial neural network. Hence, adaptive Neuro-BRB with ER can be considered as a robust method, which can be utilized in the diabetes diagnosis system. The system will facilitate the diabetic patient to diagnose their improvement of diabetes level as well. In this research, the system provides the results as a percentage of diabetes diagnosis of patient (Diabetic) that it's more informative and reliable than from the traditional expert's opinion. This system result is cost effective and time consumes to diagnose diabetes. The outputs of this adaptive Neuro-BRB diabetes diagnosis system have been compared to other diabetes diagnosis tools like traditional expert opinion and BRBES etc, this system results founded as more reliable than outputs on patient having diabetes. Further, this system methodology will train and use to implement for other disease diagnosis system.

APPENDIX

TABLE A.1 THE INPUTS FACTORS WITH REFERENTIAL VALUES

Sl. No.	Input Antecedent	Input	Referential Values
A1	Age	20-34	Low
		35-46	Medium
		>46	High
A2	plasma glucose concentration a 2 hours in an OGTT	< 95,	Low
		95-150	Medium
		>150	High
A3	diastolic blood pressure	>100	High
		70-100	Medium
		<70	Low
A4	Triceps skin fold thickness (mm)	> 40	High
		21-40	Medium
		<21	Low
A5	2-Hour serum insulin (mu U/ml)	>200	High
		140-200	Medium
		140>	Low
A6	Body mass index (BMI)	>29	High
		23-29	Medium
		23>	Low
A7	-Diabetes pedigree function	>0.8	High
		0.4-0.8	Medium
		0.4>	Low

$$\alpha_k = \prod_{i=1}^{T_k} (\alpha_i^k)^{\bar{\delta}_{ki}} \quad (\text{A.1})$$

$$\omega_k = \frac{\theta_k \alpha_k}{\sum_{j=1}^L \theta_j \alpha_j} = \frac{\theta_k \prod_{i=1}^{T_k} (\alpha_i^k)^{\bar{\delta}_{ki}}}{\sum_{j=1}^L \theta_j \left[\prod_{i=1}^{T_k} (\alpha_i^j)^{\bar{\delta}_{ji}} \right]} \text{ And } \bar{\delta}_{ki} = \frac{\delta_{ki}}{\max_{i=1, \dots, T_k} \{\delta_{ki}\}} \quad (\text{A.2})$$

$$\beta_{ik} = \frac{\sum_{t=1}^{T_k} (\tau(t, k) \sum_{j=1}^{J_t} \alpha_{ij})}{\sum_{t=1}^{T_k} \tau(t, k)} \quad (\text{A.3})$$

Where, $(t, k) = \begin{cases} 1, & \text{if } P_i \text{ is used in defining } R_k(t = 1, \dots, T_k) \\ 0, & \text{otherwise} \end{cases}$

Here $\bar{\beta}_{ik}$ is the original belief degree and β_{ik} is the updated belief degree

$$\beta_j = \frac{\mu \times \left[\prod_{k=1}^L ((\omega_k \beta_{jk} + 1 - \omega_k \sum_{j=1}^N \beta_{jk})) - \prod_{k=1}^L (1 - \omega_k \sum_{j=1}^N \beta_{jk}) \right]}{1 - \mu \times \left[\prod_{k=1}^L (1 - \omega_k) \right]} \quad (\text{A.4})$$

With

$$\mu = \left[\sum_{j=1}^N \prod_{k=1}^L \left(\omega_k \beta_{jk} + 1 - \omega_k \sum_{j=1}^N \beta_{jk} \right) - (N - 1) \times \prod_{k=1}^L \left(1 - \omega_k \sum_{j=1}^N \beta_{jk} \right) \right]^{-1}$$

$$C_m = H(A^*) = \sum_{j=1}^N u(C_j) B_j \quad (\text{A.5})$$

$$\zeta(P) = 1/M \sum_{m=1}^M (y_m - \hat{y}_m)^2 \quad (\text{A.6})$$

REFERENCES

- [1] O. Hejlesen, S. Plogmann, and D. Cavan. DiasNet, An Internet Tool for Communication and Education in Diabetes. International Symposium on Computer and Diabetes Care. Rochester MN, September 8-10, 2000
- [2] Zimmet P: The burden of type 2 diabetes: are we doing enough? Diabetes Metab 2003;29:6S9
- [3] Fagot-Campagna A, Bourdel-Marchasson I, Simon D: Burden of diabetes in an aging population: prevalence, incidence, mortality, characteristics and quality of care. Diabetes Metab 2005;25:S35-S52
- [4] Van Dieren S, Beulens JW, van der Schouw YT, Grobbee DE, Neal B: The global burden of diabetes and its complications: an emerging pandemic. Eur J Cardiovasc Prev Rehabil 2010;17:S3-S8
- [5] Deshpande AD, Harris-Hayes M, Schootman: Epidemiology of diabetes and diabetes-related complications. MPhys Ther 2008;88:1254-1264.
- [6] Lin, L., Hu, P.J.-H., Liu Sheng, O.R., () A decision support system for lower back pain diagnosis: uncertainty management and clinical evaluations. Decis. Support Syst. 42, 1152-1169, 2006
- [7] Rahaman, Saifur, and Mohammad Shahadat Hossain. "A belief rule based (BRB) system to assess asthma suspicion." *Computer and Information Technology (ICCIT), 2013 16th International Conference on*. IEEE, 2014.
- [8] Wang, Y.M., Yang, J.B., Xu, D.L., Environmental impact assessment using the evidential reasoning approach. Eur. J. Oper. Res. 174, 1885-1913, 2006
- [9] Smith, J.W., Everhart, J.E., Dickson, W.C., Knowler, W.C., & Johannes, R.S. (1988). Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. In Proceedings of the Symposium on Computer Applications and Medical Care (pp. 261-265). IEEE Computer Society Press
- [10] Nicki R. Colledge, Brian R. Walker and Stuart H. Ralston, *Davidson's Principles and Practice of Medicine 21st Edition* (Elsevier Limited, 2010) 795-818.
- [11] S. Rahaman, "Diabetes diagnosis decision support system based on symptoms, signs and risk factor using special computational algorithm by rule base", *Computer and Information Technology (ICCIT), 2012 15th International Conference on*, pp. 65-71, IEEE, 2012
- [12] D. C. Serlin and R. W. Lash, "Diagnosis and management of gestational diabetes mellitus," *Am Fam Physician*, vol. 80, pp. 57-62, Jul 2009
- [13] Siti Farhanah Bt Jaafar and Dannawaty Mohd Ali, "Diabetes mellitus forecast using artificial neural networks", Asian conference of paramedical research proceedings, 5-7, September, 2005, Kuala Lumpur, MALAYSIA
- [14] Sudajai Lowanichchai, Saisunee Jabjone, Tidanut Puthasimma, Knowledge-based DSS for an Analysis Diabetes of Elder using Decision Tree, Faculty of Science and Technology Nakhon Ratchasima Rajabhat University, Nakhonratchasima,. 30000, Thailand
- [15] Rupa Bagdi et al, Diagnosis of Diabetes Using OLAP and Data Mining Integration (2003) International Journal of Computer Science & Communication Networks, Vol 2(3), 314-322
- [16] Texas University-ESDIABETES (An Expert system In diabetes)- Consortium for Computing Sciences in Colleges ,Proceedings of the twelfth annual CCSC South Central conference on The journal of computing in small colleges April 2001, Journal of Computing Sciences in Colleges(JCSC) March 2001.
- [17] Rahaman, Saifur, and Mohammad Shahadat Hossain. "A belief rule based clinical decision support system to assess suspicion of heart failure from signs, symptoms and risk factors." In *Informatics, Electronics & Vision (ICIEV), 2013 International Conference on*, pp. 1-6. IEEE, 2013.
- [18] Kamer Kayaer, Tulay Yildirim (2003), Medical Diagnosis on Pima Indian Diabetes Using General Regression Neural Networks, Yildiz Technical University, Department of Electronics and Comm. Eng. Besiktas, Istanbul 34349 TURKEY
- [19] Rupa Bagdi et al, Diagnosis of Diabetes Using OLAP and Data Mining Integration (2003) International Journal of Computer Science & Communication Networks, Vol 2(3), 314-322
- [20] S. Issac Niwas, P. Palanisamy. Rajni Chibbar & W. J. Zhang, (2012), An Expert Support System for Breast Cancer Diagnosis using Color Wavelet Features, J Med Syst () 36:3091-3102 DOI 10.1007/s10916-011-9788-9, 2012
- [21] J.B. Yang, Rule and utility based evidential reasoning approach for multiple attribute decision analysis under uncertainty, European Journal of Operational Research 131 (1) (2001) 31-61.
- [22] J.B. Yang, J. Liu, J. Wang, H.S. Sii, H.W. Wang, A belief rule-based inference methodology using the evidential reasoning approach - RIMER, IEEE Transactions on Systems, Man, and Cybernetics - Part A 36 (2) (2006) 266-285
- [23] Y.M. Wang, J.B. Yang, D.L. Xu, Environmental impact assessment using the evidential reasoning approach, European Journal of Operational Research 174 (3) (2006) 1885-1913
- [24] A. Gegov, Distributed Fuzzy Control of Multivariable Systems, Kluwer Academic Publishers, The Netherlands, Dordrecht, 1996
- [25] Hui-Ling Chen, Bo Yang, Gang Wang, Jie Liu, Yi-Dong Chen, Da-You Liu (), A Three-Stage Expert System Based on Support Vector Machines for Thyroid Disease Diagnosis, J Med Syst 36:1953-1963

- [26] Huang, M.-J. & Chen, M.-Y. () Integrated design of the intelligent web-based Chinese Medical Diagnostic System (CMDS) - Systematic development for digestive health. *Expert Systems with Applications*, 32, 658-673, 2007
- [27] S. Issac Niwas, P. Palanisamy. Rajni Chibbar & W. J. Zhang, (2012), An Expert Support System for Breast Cancer Diagnosis using Color Wavelet Features, *J Med Syst* () 36:3091–3102 DOI 10.1007/s10916-011-9788-9, 2012
- [28] Lin, C., Lin, C. M., Lin, B. & Yang, M.-C. A decision support system for improving doctors' prescribing behavior. *Expert Systems with Applications*, 36, 7975 -7984, 2009
- [29] Rahaman, Saifur, M.M. Islam, MS Hossain, "A Belief Rule Based Clinical Decision Support System Framrwork", *Computer and Information Technology (ICCIT), 2014 17th International Conference on*. IEEE, 2014
- [30] Thilo, B., Vasileios, A., Gregor, H., Friederike, L., Jens, K., Simon, P. W. S., Hamid, G., Angelika Zabel-Du, B., Stefanie, M.-Z., Jgen, D., Hubert, J. B. & Walter, E. H. () Multidisciplinary pain management based on a computerized clinical decision support system in cancer pain patients. *Pain*, 147, 20-28, 2009
- [31] Roger S. Pressman, *Software engineering: a practitioners approach*. 5th ed. (McGraw-Hill series in computer science) Includes index. ISBN 0 -07-365578-3, p373- 37
- [32] Rahaman, S. "Diabetes diagnosis expert system by using Belief Rule Base with evidential reasoning", *Electrical Engineering and Information Communication Technology (ICEEICT), 2015 International Conference on*, On page(s): 1 – 6, IEEE, 2015

Suppression of White and Colored Noise in Bangla Speech Using Kalman Filter

Muhammad Navid Anjum Aadit¹, Sharadindu Gopal Kirtania², and Mehnaz Tabassum Mahin³

^{1,2}Department of Electrical and Electronic Engineering

³Department of Computer Science and Engineering
Bangladesh University of Engineering and Technology

Email: ¹navidanjumaadit@gmail.com, ²sharadindugopal@gmail.com, and ³mehnaztabassummahin@gmail.com

Abstract—We introduce adaptive Kalman filter approach for suppression of white and colored noise that corrupts Bangla speech. It is difficult to retrieve the desired information from noise corrupted speech signals. Bangla speech is frequently corrupted by white and colored noise in chaotic environment and so noise suppression is extremely necessary for smooth communication purpose. Background noise in Bangla speech may be stationary or dynamic in nature. Kalman filter is a well-known recursive filter that enhances speech signals corrupted by both static and dynamic noises. In this paper, we apply adaptive Kalman filter and show that it effectively reduces white and colored noise from corrupted Bangla speech and retrieves the information carried by the original speech. We compare the retrieved speech to the corresponding noise-free speech in terms of pitch values and analyze error to evaluate performance of the proposed filter. We take oral speech as voice record by native Bangla speakers and work with both vowels and consonants to show the feasibility of Kalman filter in both cases. Both male and female voices are analyzed to prove the effectiveness of the filter for Bangla speech irrespective of gender. Elimination of noise would provide non-disruptive communication in Bangla enhancing the efficacy of Bangla speech processing.

Keywords—Bangla Speech, Colored Noise, Kalman Filter, Pitch, White Noise

I. INTRODUCTION

We always face the problem in hearing speech and sounds due to unwanted background disturbance or noise that degrades the quality of speech and corrupts the information to reduce its intelligibility. We cannot eliminate those background noise completely from any system but we can reduce them to improve the renditions. In public and industrial places, it is a very acute problem and highly intense noise may mask the speech recognition system [1]. There are two main types of noise known as white noise and colored noise. White noise is the noise signal whose integrated power is constant in a large bandwidth. It consists of almost all the frequencies and will have constant power at all these frequencies [2]. Colored noise has different integrated power at different frequency bands and classified accordingly as pink noise, red noise, violet noise etc. [3]. Clear speech with precise information is required in many fields including cellular network [4], car navigation system [5] and audio recording [6].

Noise suppression is necessary for speech enhancement and so many researches have been in progress on this area over the years. A common noise suppression method that

was previously used is Wiener filtering [7]. Talmon *et al.* presented a novel approach for transient noise reduction that relies on non-local (NL) filtering [8]. Mendiratta and Jha presented Adaptive Noise Cancelling (ANC) method by least mean square algorithm [9]. Kawamura *et al.* proposed a noise reduction method using linear predictor with variable step size which reduces the error of the linear predictor [10]. These methods have some limitations and do not ensure high quality speech enhancement or noise suppression. Weiner algorithm is for specific time span of noise signal and does not have good performance for speech signals corrupted by dynamic noise [11]. ANC method reduces noise signal using the periodicity of speech signal and may sacrifice speech quality if multiple channels are not available [12]. Kalman filter overcomes these drawbacks and ensures high speech quality [11]. It is very useful in a stochastic environment where the noise characteristics is known [13]. Speech signal is modeled as autoregressive (AR) process and defined in the state-space domain in this filtering process [14]. Kalman filter is the most widely used technique to remove noise from speech signals in different languages. Bangla (or Bengali) is spoken by about 300 million people in the world and it has become one of the prominent languages. Though research on bangla speech signal processing is a new concept, some promising research works are found in literature regarding Bangla speech signal processing. Hossain *et al.* described detailed analysis of Bangla dental consonants [15]. Kamal *et al.* proposed a fuzzy based Bangla vowel sign recognition system [16]. Chowdhury *et al.* presented a new word separation algorithm for real time speech i.e. Continuous Bangla Speech Recognition (CBSR) [17]. To make communication in Bangla over phones, internet and other media efficiently, noise reduction in audio signals is a mandatory task. But no experimental method or simulation approach has been reported so far to improve noisy Bangla speech signal particularly. In this paper, for the first time, we analyze effect of white noise and colored noise on Bangla speech signal and present a way to suppress them from speech.

We record voice signals from native Bangla speakers in chaotic environment and sampled the signal using MATLAB. We implement adaptive Kalman filtering approach and suppress white and colored noise from the corrupted speech. We also record the same speech by the same speakers in a silent environment and compare it with the filtered speech in terms

of pitch. We calculate error and find that the discrepancy is very insignificant and Kalman filter reproduces almost noise-free speech signal. We analyze both vowels and consonants and also consider both male and female voices to show the effectiveness of Kalman filter in all possible conditions. Enhancement of Bangla speech using Kalman filter would pave the way for many new researches on Bangla speech processing, recognition and communication.

II. OVERVIEW OF BANGLA SPEECH

Bangla, as a widely used language is ranked as seventh based on the total number of speakers [18]. Like most other languages, Bangla alphabet is classified into vowels and consonants. There are 11 vowels and 39 consonants in Bangla alphabet [19]. There are two forms of Bangla speech: literary form and colloquial form. For general speaking and writing, people use colloquial form. Even in standard colloquial form, Bangla is differed according to region or religion of speakers. There are also some words in Bangla speech which are borrowed from Sangshkrit, Arabic, Persian, English and other foreign languages. Despite being a popular language with enriched vocabulary and variety [19], speech processing in this language has just recently started and so not that much progress is found in literature. Specially, noise elimination from Bangla speech in chaotic environment is yet to be done for further speech processing works.

III. SIMULATION TECHNIQUE

Kalman filter is the most popular filtering approach used for noise suppression in the recent years [11]. There are many types of Kalman filters found in literature. Among them, adaptive Kalman filter is found to be the most effective in case where noise is non-stationary [20]. The advantage of using adaptive Kalman filter is that it can detect dynamic time-varying noise signals. For suppression of white and colored noise from Bangla speech which is corrupted in rigorous chaotic environment, it is necessary to achieve fast tracking and estimation of time-varying noise. Hence, adaptive Kalman filter may be the best choice for filtering corrupted Bangla speech and is introduced in this paper.

The Kalman filter estimates the state of a system as a variable quantity taking the system's dynamic model as input. Primitive Kalman filters only included whiteness characteristics of noise assuming that noise is not correlated. However, advanced Kalman filters are developed to include colored noise in Kalman filtering algorithm [20]. We introduce the following state vectors to fit $n \times 1$ one-dimensional speech signal into Kalman filter:

$$u(m) = [u(m-n+1) u(m-n+2) \dots u(m)]^T, \quad (1)$$

where, $u(m)$ refers to noise-free speech signal at time m . Mixing white noise $w(m)$, the corrupted signal is given by:

$$v(m) = u(m) + w(m), \quad (2)$$

Invoking AR process [20] of order n ,

$$u(m) = \sum p_i u(m-i) + x(m) \quad i = 1, \dots, n, \quad (3)$$

$$w(m) = \sum q_i w(m-i) + y(m) \quad i = 1, \dots, r, \quad (4)$$

where, p, q are AR coefficients, x, y are assumed errors and r is order of noise signal. Substituting (1) and (2) into (3) and (4) respectively, we find,

$$u(m) = A_u u(m-1) + B_u x(m), \quad (5)$$

$$w(m) = A_w w(m-1) + B_w y(m), \quad (6)$$

where,

$$B = [0 \ 0 \ \dots \ 0 \ 1]^T$$

$$A_u = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 1 \\ p_n & p_{n-1} & p_{n-2} & \dots & \dots & \dots & p_1 \end{bmatrix}$$

and

$$A_w = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 1 \\ q_n & q_{n-1} & q_{n-2} & \dots & \dots & \dots & q_1 \end{bmatrix}$$

The observation equation then becomes,

$$v(m) = \mathbf{H}u(m) + w(m), \quad (7)$$

where,

$$\mathbf{H} = B^T$$

Equations to describe state model for speech signal $u(m)$ and observation signal $v(m)$ are,

$$u(m) = A u(m-1) + z(m-1), \quad (8)$$

$$v(m) = \mathbf{H}u(m) + w(m), \quad (9)$$

where, $z(m)$ is uncorrelated process noise and A is an $n \times n$ matrix of state transition.

The pre-estimation and post-estimation of errors e^- and e can be calculated respectively as the followings:

$$e^-(m) = u(m) - u(m|m-1), \quad (10)$$

$$e(m) = u(m) - u(m|m), \quad (11)$$

Finally, the pre-estimation and post-estimation of error covariances c^- and c can be calculated respectively as the followings:

$$c^-(m) = E\{e^-(m) e^{-T}(m)\}, \quad (12)$$

$$c(m) = E\{e(m) e^T(m)\}, \quad (13)$$

An equation that can relate pre-estimation and post-estimation of error to measure weighted difference between actual and predicted measurement is given by,

$$u(m|m) = u(m|m-1) + M(m)(v(m) - \mathbf{H}u(m|m-1)), \quad (14)$$

where, $(v(m) - \mathbf{H}_u(m|m-1))$ is known as innovation of measurement and $M(m)$ is given by:

$$M(m) = c^-(m)H^T(Hc^-(m)H^T + N)^{-1}, \quad (15)$$

where, N is measurement noise co-variance and tends to maximize measurement innovation as following,

$$\lim_{N \rightarrow 0} M(m) = H^{-1}, \quad (16)$$

Oppositely, pre-estimation of error co-variance c^- tends to minimize measurement innovation as following,

$$\lim_{c^-(m) \rightarrow 0} M(m) = 0, \quad (17)$$

Thus, in adaptive Kalman filter algorithm, the pre-estimation of error co-variance is minimized, and it may be expected that the predicted measurement is more accurate.

We have taken similar Bangla speech signal as audio record in both chaotic and noise-free environments. Both native male and female voices are analyzed and both vowels and consonants have been considered. As speech signal varies over time, only a small portion of speech signal has been sampled which may be called a window signal. The window signal is then downsampled and read by built-in 'audioread' function in MATLAB. The adaptive Kalman filter algorithm is also implemented in MATLAB by writing custom codes. The pitches of both noise-free and corrupted signals have been calculated using cepstrum algorithm [21]. The pitch of the retrieved signal after Kalman filtering is also calculated using similar technique and compared with the noise-free version. The closeness of the pitch values shows the accuracy of adaptive Kalman filter in suppression of white and colored noise from Bangla speech. Finally, we also calculate Mean Square Error (MSE) to evaluate performance of the filter.

IV. RESULTS AND DISCUSSIONS

At first, we take record of sample male speech for Bangla vowel 'অ' in a noise-free environment. Downsampling a window portion of the speech signal and reading it by MATLAB as described in Section III, we calculate pitch of the signal using cepstrum algorithm. In Fig. 1, the noise-free signal spectrum for vowel 'অ' is shown on top and the corresponding cepstrum is shown at the bottom. From the cepstrum, pitch of this signal is found to be 156.86 Hz.

Now, the same vowel 'অ' is recorded by the same male, but in a chaotic environment where uniform white noise is generated using pre-recorded noisy audio. Fig. 2 shows the spectrum and cepstrum of its downsampled window. From the cepstrum, the pitch of white noise corrupted speech is found to be 123.25 Hz. From the spectrum, it is also found that the corrupted signal has very much higher dB value compared to the noise-free signal. Thus, white noise not only deteriorates the pitch of the original signal but also corrupts the carried information by deviating the amplitude.

The white noise corrupted Bangla speech for vowel 'অ' is then passed through the adaptive Kalman filter designed in MATLAB. The output spectrum and its cepstrum are shown in

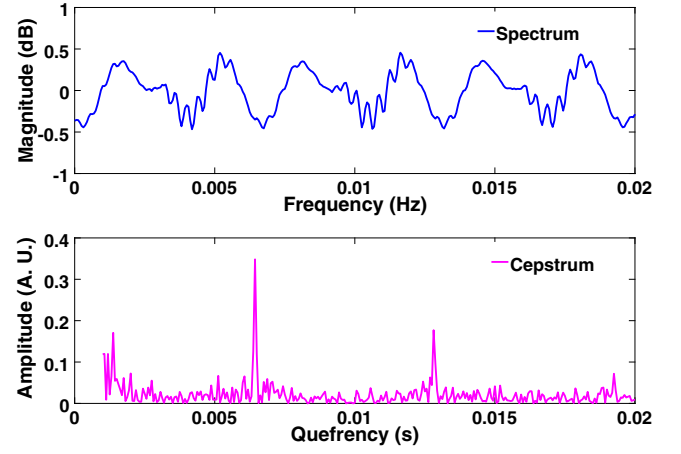


Fig. 1. Spectrum and cepstrum of vowel 'অ' uttered by a male in a noise-free environment

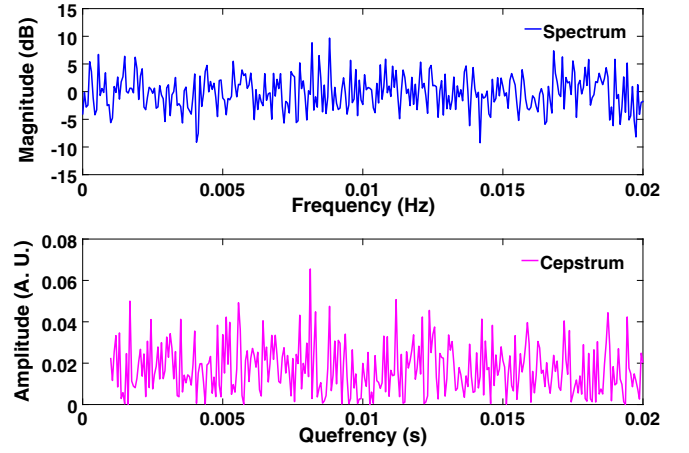


Fig. 2. Spectrum and cepstrum of vowel 'অ' uttered by a male in a chaotic (white noise) environment

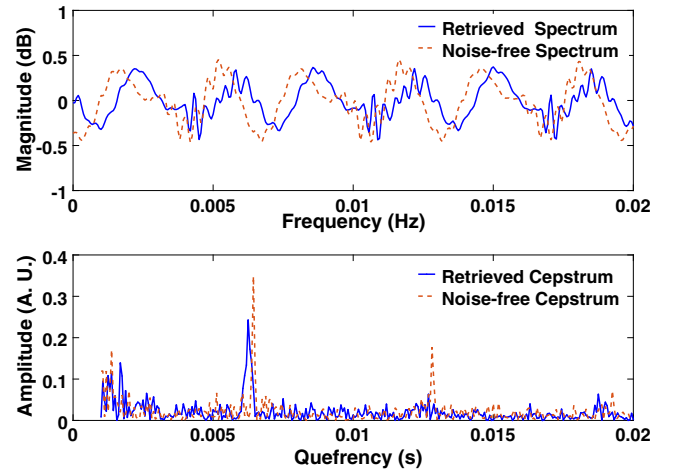


Fig. 3. Spectrum and cepstrum of vowel 'অ' uttered by a male after Kalman filtering for white noise

Fig. 3 where these are compared with the corresponding noise-free versions. The pitch value of the retrieved speech is found to be 160.01 Hz which is very close to that of the noise-free speech (156.86 Hz). The slight deviation is not unexpected as the recorded speech cannot be exactly equal each time even if it is uttered by the same human. Also, the figure shows that the adaptive Kalman filter not only recovers pitch but also provides correction for amplitude and successfully retrieves the information carried by the original signal.

Fig. 4 shows the spectrum and cepstrum of vowel 'অ' uttered by the same human in another chaotic environment generated by a pre-recorded colored noise corrupted audio. From the cepstrum, the pitch value is found to be 231.84 Hz. The spectrum shows that colored noise affects the original signal more than white noise in terms of dB values. Thus, it is more challenging to recover a speech signal successfully that is corrupted by colored noise. However, adaptive Kalman filter is very effective in this purpose and the retrieved speech signal is shown in Fig. 5. The pitch of the signal retrieved from colored noise corrupted speech of vowel 'অ' is found to be 155.59 Hz which is close to the noise-free pitch value (156.86 Hz). Also, the dB value of amplitude is successfully recovered by Kalman filter. Therefore, adaptive Kalman filter can suppress both white and colored noises and recover the desired Bangla speech effectively.

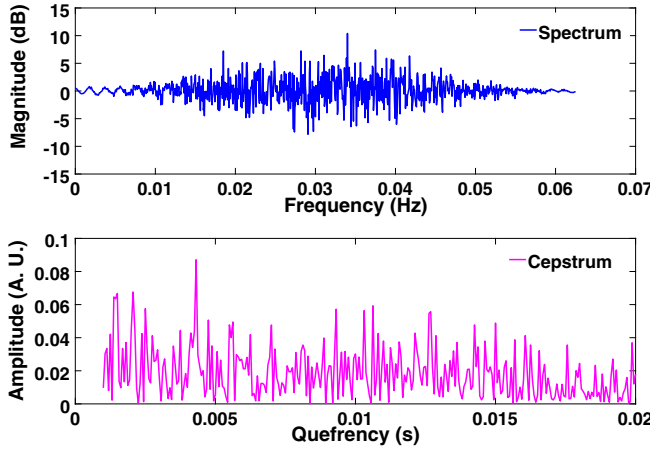


Fig. 4. Spectrum and cepstrum of vowel 'অ' uttered by a male in a chaotic (colored noise) environment

Now, we consider female voice as well as consonant instead of vowel to show the efficacy of adaptive Kalman filter irrespective of type of alphabet or gender. Fig. 6 shows the spectrum and cepstrum of white noise corrupted female voice of consonant 'ক'. The pitch frequency is found to be 276.63 Hz. After passing the corrupted speech into Kalman filter, we retrieve the original spectrum and cepstrum as shown in Fig. 7. The pitch frequency is found to be 21.1 Hz for the retrieved speech. From the figure, we can also see that pitch is found to be 27.33 Hz for the noise-free case. The magnitude spectra for both the cases are also quite similar. Hence, the original information is successfully retrieved from the white noise corrupted female Bangla consonant speech 'ক'.

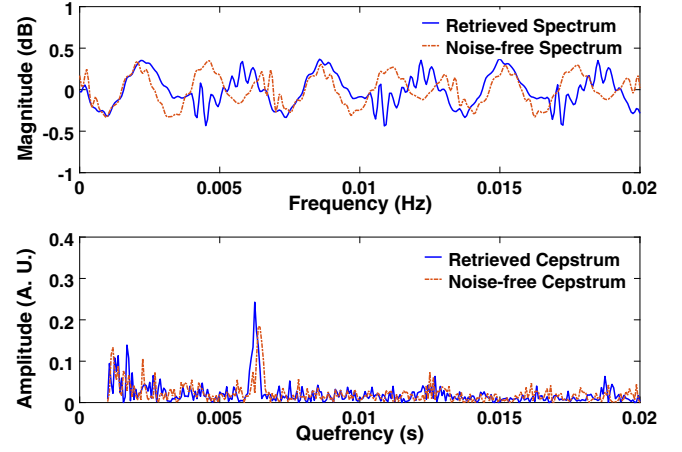


Fig. 5. Spectrum and cepstrum of vowel 'অ' uttered by a male after Kalman filtering for colored noise

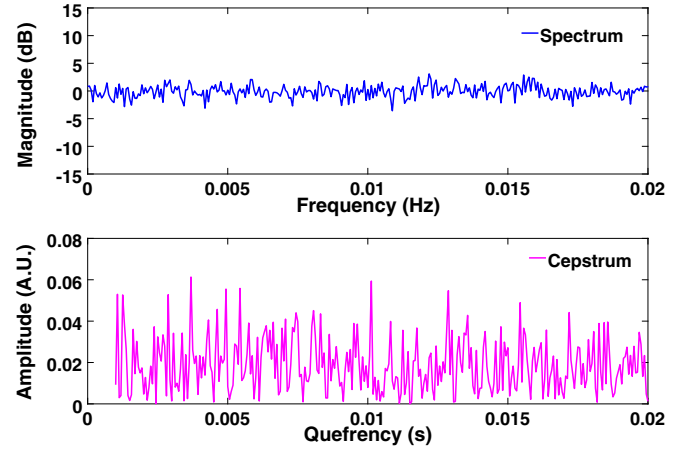


Fig. 6. Spectrum and cepstrum of consonant 'ক' uttered by a female in a chaotic (white noise) environment

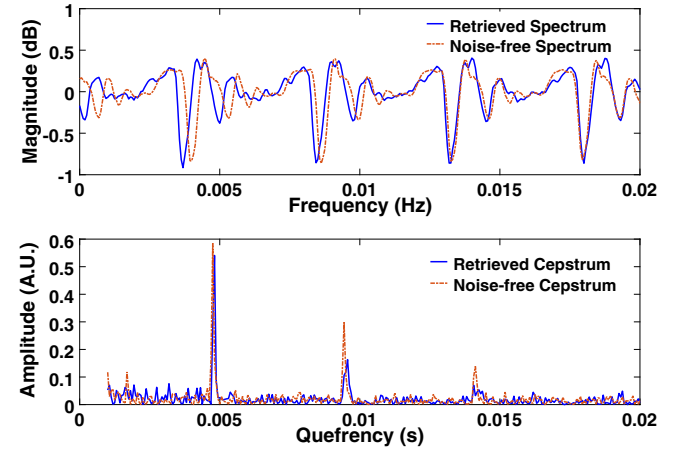


Fig. 7. Spectrum and cepstrum of consonant 'ক' uttered by a female after Kalman filtering for white noise

Fig. 8 shows the spectrum and cepstrum of the same consonant 'ক' by the same female in a colored noise corrupted chaotic environment. The pitch frequency for this case is 551.72 Hz. White and colored noises may be so random in nature that the pitch values differ extremely between these two corrupted cases. The magnitudes are also disturbed for these cases which is also same as in the case for male or vowel speeches. Fig. 9 shows the recovered speech from colored noise corrupted speech compared to the noise-free version. The two spectra are very similar and thus show that adaptive Kalman filter recovers lost information from successfully in colored female Bangla consonant too. So, it is clear from the discussed cases that adaptive Kalman filter is effective in all possible cases: either vowels or consonants, either female or male and either white or colored noise corruption for Bangla speech.

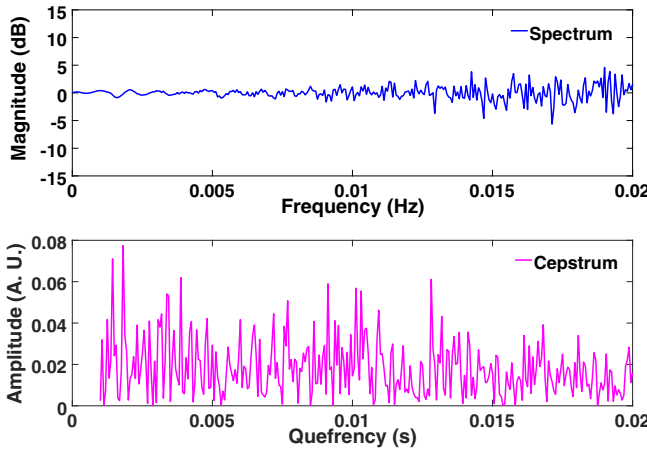


Fig. 8. Spectrum and cepstrum of consonant 'ক' uttered by a female in a chaotic (colored noise) environment

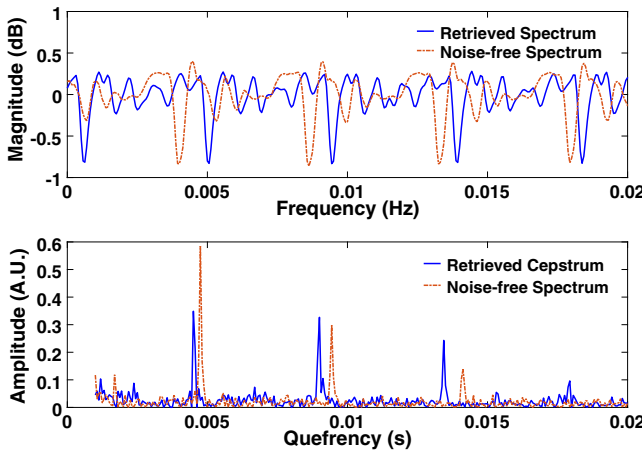


Fig. 9. Spectrum and cepstrum of consonant 'ক' uttered by a female after Kalman filtering for colored noise

Finally, to justify our findings and evaluate the performance of Kalman filter, we calculate Mean Square Error (MSE) between the noise-free signal and the retrieved signal under

different chaotic conditions with different Signal to Noise Ratios (SNRs). The plot of MSE vs. SNR for both white and colored noises is shown in Fig. 10. It is found that as SNR increases, error also increases. Colored noise exhibits a slightly higher error than white noise, which is obvious as it is more difficult to predict and eliminate colored noise. However, the maximum MSE is found to be only 0.2 dB, which verifies the efficiency of Kalman filter to suppress both white and colored noises in Bangla speech.

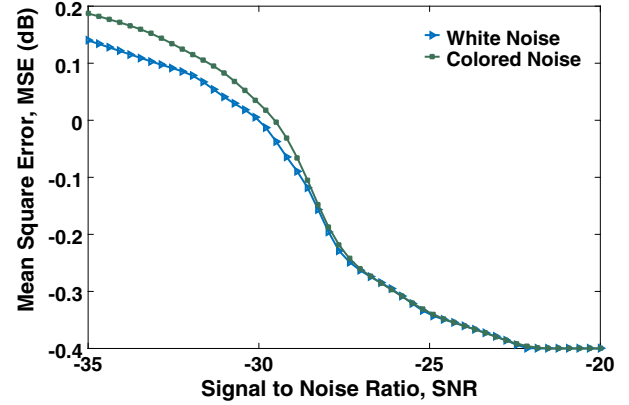


Fig. 10. Mean Square Error vs. Signal to Noise Ratio for white and colored noises in Bangla speech

V. CONCLUSION

We have presented noise suppression technique for Bangla speech considering both white and colored noises. We have implemented adaptive Kalman filter algorithm to retrieve the information carried by the original signal. Both vowels and consonants have been analyzed and gender effect has also been included. The retrieved speech is compared with noise-free recordings to verify the accuracy and applicability of the proposed algorithm for Bangla speech. It is found that adaptive Kalman filter is an effective and precise tool in all possible conditions to suppress white and colored noises from Bangla speech. Clear and precise speech provided by this technique would extensively aid future speech processing researches on Bangla speech.

REFERENCES

- [1] D. C. Popescu and I. Zeljkovic, "Kalman filtering of colored noise for speech enhancement," in *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, vol. 2, May 1998, pp. 997–1000 vol.2.
- [2] C. E. Shannon, "Communication in the Presence of Noise," *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, Jan 1949.
- [3] S. Kamath and P. Loizou, "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise," in *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, vol. 4, May 2002, pp. IV-4164–IV-4164.
- [4] Z. Tong, G. D. Li, and W. X. Xu, "The Impact of Noise or Noiseless on Cellular Neural Network to Detect Image Edge," in *Intelligent Systems Design and Engineering Applications (ISDEA), 2014 Fifth International Conference on*, June 2014, pp. 1056–1059.

- [5] Y. Hiratani, A. Hirano, and M. Kanazawa, "A noise-robust echo canceller on V830 multimedia RISC processor integrated into a car navigation system," in *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, vol. 3, May 1998, pp. 1753–1756 vol.3.
- [6] N. Fechner and M. Kirchner, "The Humming Hum: Background Noise as a Carrier of ENF Artifacts in Mobile Device Audio Recordings," in *IT Security Incident Management IT Forensics (IMF), 2014 Eighth International Conference on*, May 2014, pp. 3–13.
- [7] B. Widrow, J. R. Glover, J. M. McCool, J. Kaunitz, C. S. Williams, R. H. Hearn, J. R. Zeidler, J. E. Dong, and R. C. Goodlin, "Adaptive noise cancelling: Principles and applications," *Proceedings of the IEEE*, vol. 63, no. 12, pp. 1692–1716, Dec 1975.
- [8] R. Talmon, I. Cohen, and S. Gannot, "Clustering and suppression of transient noise in speech signals using diffusion maps," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2011, pp. 5084–5087.
- [9] A. Mendiratta and D. Jha, "Adaptive Noise Cancelling for audio signals using Least Mean Square algorithm," in *Electronics, Communication and Instrumentation (ICECI), 2014 International Conference on*, Jan 2014, pp. 1–4.
- [10] A. Kawamura, Y. Iiguni, and Y. Itoh, "A noise reduction method using linear predictor with variable step-size," in *Communications and Information Technology, 2004. ISCIT 2004. IEEE International Symposium on*, vol. 1, Oct 2004, pp. 563–566 vol.1.
- [11] N. Tanabe, T. Furukawa, H. Matsue, and S. Tsujii, "Noise Suppression with High Speech Quality Based on Kalman Filter," in *2006 International Symposium on Intelligent Signal Processing and Communications*, Dec 2006, pp. 315–318.
- [12] K. Wooll, "Noise variance estimation for Kalman filtering of noisy speech," *IEICE TRANSACTIONS on Information and Systems*, vol. 84, no. 1, pp. 155–160, 2001.
- [13] S.-S. Xiong and Z.-Y. Zhou, "Neural filtering of colored noise based on Kalman filter structure," *IEEE Transactions on Instrumentation and Measurement*, vol. 52, no. 3, pp. 742–747, June 2003.
- [14] M. Gabrea, "An adaptive kalman filter for the enhancement of speech signals in colored noise," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2005.*, Oct 2005, pp. 45–48.
- [15] S. A. Hossain, M. L. Rahman, and F. Ahmed, "Analysis and synthesis of Bangla Dental consonants," in *Computer and Information Technology, 2008. ICCIT 2008. 11th International Conference on*, Dec 2008, pp. 166–170.
- [16] M. S. Kamal, M. M. Hoque, M. M. U. Hasan, and M. S. Arefin, "Bangla vowel sign recognition by extracting the fuzzy features," in *Computer and Information Technology, 2008. ICCIT 2008. 11th International Conference on*, Dec 2008, pp. 306–311.
- [17] N. Chowdhury, M. A. Sattar, and A. K. Bishwas, "An Efficient Algorithm Bangla Speech Separation," in *Computer Sciences and Convergence Information Technology, 2009. ICCIT '09. Fourth International Conference on*, Nov 2009, pp. 663–666.
- [18] Lewis, M. Paul, G. F. Simons, and C. D. Fennig. (2016) Summary by Language Size. [Online]. Available: <http://www.ethnologue.com/statistics/size>
- [19] Wikipedia. (2016) Bengali Language. [Online]. Available: https://en.wikipedia.org/wiki/Bengali_language
- [20] M. Mathe, S. P. Nandyala, and T. K. Kumar, "Speech enhancement using Kalman Filter for white, random and color noise," in *Devices, Circuits and Systems (ICDCS), 2012 International Conference on*, March 2012, pp. 195–198.
- [21] Wikipedia. (2016) Cepstrum. [Online]. Available: <https://en.wikipedia.org/wiki/Cepstrum>