ULAB JOURNAL OF SCIENCE AND ENGINEERING

A RESEARCH PUBLICATION OF ULAB

November 2015 Vol. 6

ISSN 2079-4398 (Print)
ISSN 2414-102X (Online)

CONTENTS

(Contents Continued on Back Cover)

School of Science & Engineering
UNIVERSITY OF LIBERAL ARTS
BANGLADESH

ULAB JOURNAL OF SCIENCE AND ENGINEERING

Vol. 6, November 2015

(Contents Continued from Front Cover)

ULAB
UNIVERSITY OF LIBERAL ARTS BANGLADESH

House 56, Rd 4/A @ Satmasjid Road Dhanmondi, Dhaka-1209, Bangladesh
Phone: 966-1255, 966-1301, 0171-309-1936. Web: www.ulab.edu.bd

Editorial
Vol. 6, 2015

WE are pleased to present the sixth volume of the ULAB's Journal of Science and Engineering (JSE). It's been another successful year for us. Since 2010, ULAB's JSE has maintained its position as the most prestigious national publication in the field of science and engineering. We feel proud that the journal has crossed the national boundary and becomes international. Now we are getting papers from abroad.

In 2010 (first volume), 28 papers were submitted for publication, with 10 being accepted (acceptance rate: 35.71%). In 2011 (second volume), submissions increased to 30 and after peer review 10 papers were being accepted (acceptance rate: 33.33%). In 2012 (third volume) 17 papers (including 1 from UK) were submitted for publication, with 9 being accepted (acceptance rate: 52.94%). In 2013 (fourth volume) 16 papers (including 3 from India and 1 from South Korea) papers were submitted for publication, with 8 being accepted (acceptance rate: 50.00%). In 2014 (fourth volume) 10 papers (including 1 from USA) were received and after peer-review 8 were being accepted (acceptance rate: 80.00%). In the first volume, we found many ordinary papers were being submitted, but the quality of submitted paper is gradually improving in the subsequent volumes. This year we have received 11 submissions and after peer-review 6 are being accepted (acceptance rate: 54.54%).

According to the publication policy all papers submitted to this journal have been subject to a rigorous peer-review. We continuously strive to publish original research that contains elements with technical novelty in a timely manner. The journal's focus is on traditional areas of both theoretical and practical applications of physics, mathematics, statistics, environmental science, electronics, computer science, information and communication engineering. In addition, we shall gladly accept submissions on emerging technologies and other emerging areas related to the above fields.

You are most welcome to read this issue of the ULAB Journal of Science and Engineering. In order to continue publishing a high-quality journal, JSE's editorial board seeks excellent contributions containing original research or reviews. Our editorial board works tirelessly to provide contributors with a prompt and thorough review process.

We would like to extend our heartfelt thanks to every author, reviewer and reader for your support and dedication to JSE. We strongly believe that together, we shall elevate the journal to even higher levels of quality, impact, and reputation.

Mohammad Shorif Uddin
Editor-in-Chief

Rezaul Karim Mazumder
Editor

Sazzad Hossain
Editor

Self Organization Mapping-Based Robust Speaker Identification in Noisy Environment using Linear Discriminant Analysis

Md. Rabiul Islam and Md. Fayzur Rahman
Department of Computer Science & Engineering, Rajshahi University of Engineering & Technology, Bangladesh
Department of Electrical & Electronic Engineering, Daffodil International University, Bangladesh.
E-mail: rabiul_cse@yahoo.com and mfrahman3@yahoo.com

Abstract— The aim of this work is to evaluate the performance of the noise robust speaker identification system using Kohonen self organization mapping based neural network algorithm. Speaker identification system performs well under noiseless environment but its performance degrades when the environments become noisy. In this work, wiener filtering technique has used to reduce the white Gaussian environmental speech noises, standard feature extraction techniques have been applied to extract the effective speech features. Linear Discriminant Analysis based dimensionality reduction technique has been used to reduce the dimension of the extracted speech features. Kohonen self organization learning neural network based algorithm has applied for the learning and recognition model of the proposed system. Finally, NOIZEUS speech dataset has used to measure the efficiency of the proposed system under various environmental noisy conditions with different noise addition rate.

Keywords—noise robust speaker identification, self organization mapping, artificial neural networks, dimensionality reduction linear discriminant analysis.

Manuscript received on 17 March 2015 and accepted for publication on 9 June 2015.

1 Introduction

Speaker identification is a dynamic biometric task where both biological and behavioral biometric modalities are concatenated through speech feature extraction, learning and recognition modeling [1, 2]. Though biological and behavioral biometric modalities are combined through speaker identification task, it can achieve great efficiency than other individual biological and behavioral biometric system such as face, iris, retinal scan, fingerprint, hand geometry etc. Generally speaker identification system performs well under neutral conditions but its performance degrades when the environments become noisy. There are various techniques that have been developed by different researchers for automatic speech and/or speaker identification problem [3, 4, 5, 6, 7, 8, 9]. From the last several years researchers investigates and tries to enhance performance of the speaker identification in noisy environments [10, 11, 12, 13, 14, 15].

Although there are recent studies to handle reverberation and additive noise in feature [15] and model domain [13, 14] for speaker identification systems, the compensation techniques with respect to noise and reverberation for speaker identification systems are still an open question [16]. There are many speech enhancement algorithms proposed for robust automatic speech recognition, most of them relying on the assumption that the additive noise is a stationary process which is not always true for real-world applications [8].

Not only single classification techniques are used for noise robust speech and speaker identification but also hybrid classification techniques have also been applied. Genetically optimized Hidden Markov Model (HMM) based hybrid technique [17] has been used for noise robust speaker identification system. Both audio and visual features are used for noise robust speaker identification purpose. Face and speech features are used with the combination of different types of fusion methods with neutral [18] and noisy environment [19]. Leap reading feature with audio feature have also been used for speaker identification [20].

In this paper, we propose a Kohonen self organizing mapping based approach for noise robust speaker identification which is much more capable than other existing system. To measure the performance, the proposed system has been tested under eight different real world noisy conditions such as airport, babble, car, exhibition, restaurant, street, train and train station with four different noise addition rate i.e., 0dB, 5dB, 10dB and 15 dB.

2 Strategy of the Proposed Speaker Identification System

The block diagram of the proposed speaker identification system is shown in figure 1. After acquisition of the speech signal, it is required to eliminate the noise from the speech signal. Then some speech pre-processing techniques have been applied to pre-process the speech before feature extraction. Various standard speech parameterization techniques are applied to effectively extract the

speech features and finally select best of them. The dimensions of the extracted speech features are very large. As a result, dimension reduction technique is used to reduce the speech feature vector. The reduced feature vector is now feed to the Kohonen self organizing mapping based neural network algorithm for learning and testing. Different parameters of Kohonen self organizing mapping based network has been tested for learning and finally optimum parameters are chosen to find out the final result.

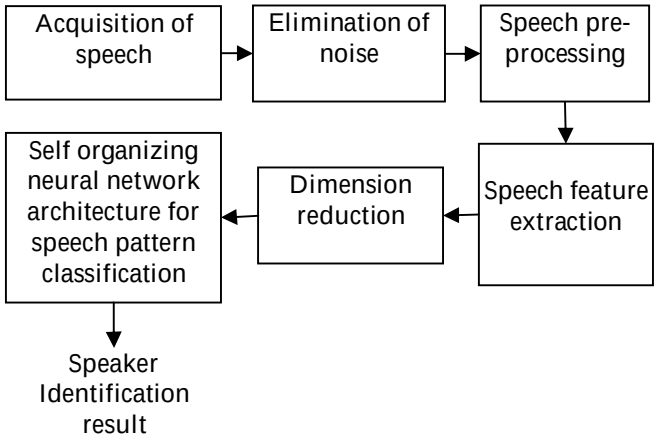


Figure 1: Architecture of the proposed speaker identification system.

3 Noise Elimination and Speech Pre-processing Techniques

Since speech signals generally captures a lot amount of environmental noises, it is necessary to reduce or eliminate noises from the captured speech signal. For this reason, wiener filtering technique has been used in the proposed system. The wiener filter is a noise removing filter based on Fourier iteration. Its main advantage is the short computational time it takes to find a solution [21].

Some pre-processing techniques are necessary to prepare the speech as an input for the system. Speech start and end points detection and silence part removal algorithm [22] have been used to extract the core part of the speech utterances. Pre-emphasis has been used to balance the spectrum of voiced sounds that have a steep roll-off in the high frequency region [23]. Since Sort Time Fourier Transform (STFT) is the appropriate method to analyze the speech features, windowing technique has been used to reduce the effect of the spectral artifacts that results from the framing process. Frame length of 10-30 milliseconds speech has been considered here and 25%-75% frame overlapping has been tested to get the optimum output of the speaker identification result.

4 Speech Feature Extraction and LDA based Dimension Reduction Technique

To extract the features from the speech utterances, various speech feature extraction techniques have been applied and find out the optimum results which will be focused in the experimental result section. Linear Prediction Coefficients (LPCs), Linear Prediction Cepstral Coefficients (LPCCs), Mel Frequency Cepstral Coefficients (MFCCs), Delta MFCCs and Delta Delta MFCCs [24] are tested to find out the best speech feature extraction technique for the proposed speaker identification system. The process of feature extraction and dimensionality reduction technique is shown in figure 2.

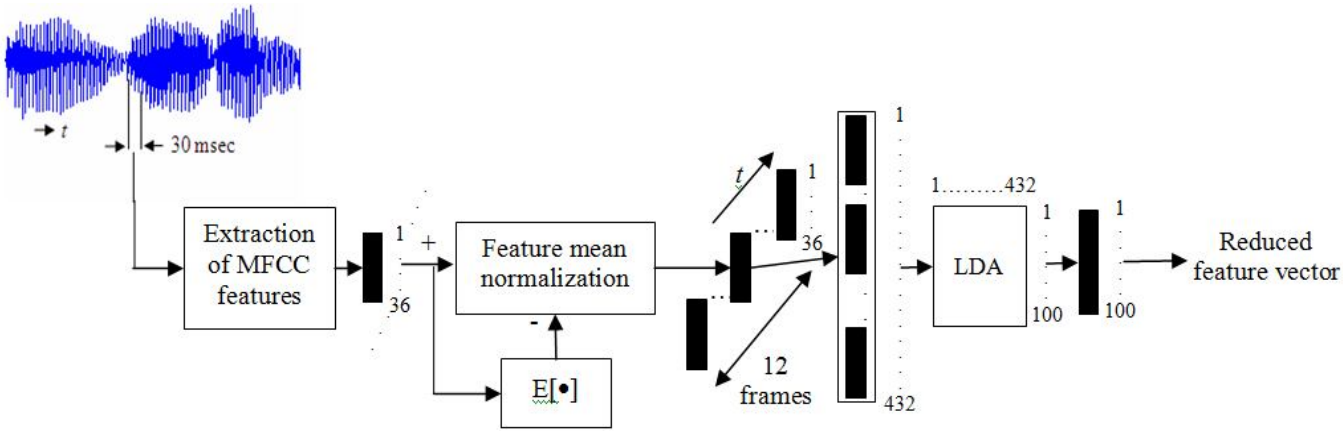


Figure 2: Feature extraction and dimensionality reduction from the speech utterance.

The dimensionality of the extracted speech feature vector is large. As a result, it is necessary to reduce the dimension of the speech features. To reduce the dimension of the speech feature vector, Linear Discriminant Analysis (LDA) based dimension reduction technique has been used.

The first step is to formulate the data sets and test sets, which are to be classified in the original space. The mean of each data set and mean of entire data set is to be calculated. After the mean image has been calculated, mean image is subtracted from each of those images of the training data set. Mean image is subtracted from each original image F_i and stored in the variable Φ_i. Each image in the data set differs from the average face by the vector Φ_i.

$$\Phi_i = F_1 - \Psi \quad (1)$$

Since LDA calculates the difference of features within all images of each person individually. So, one scatter matrix is calculated for each person from its images.

$$S_i = \sum_{j=1}^{L} F_{ij} F_{ij}^T \quad (2)$$

Here,

S_i is the scatter matrix of i^{th} person

L is the number of images of each person

F_{ij} is the j^{th} image of i^{th} person

Summation of all scatter matrices is called within-class scatter matrix which represents variation among images of each persons.

Within-class scatter matrix,

$$S_w = \sum_{i=1}^{M} S_i \quad (3)$$

Here,

M is the number of total persons

S_i is the i^{th} scatter matrix

Between-class scatter matrix represents the variation among persons. For Between-class scatter matrix,

$$S_B = 2\sum_{i=1}^{M} F_{mean,i} F_{mean,i}^T \quad (4)$$

Here,

M is the number of total persons

S_i is the i^{th} scatter matrix

$F_{mean,i}$ represents mean image of i^{th} person

Since LDA maximizes between-class scatter whereas minimizes the within-class scatter. To accomplish this, we must maximize W matrix where,

$$J(W) = |W^T S_B W| / |W^T S_W W| \quad (5)$$

From the matrix W, we will compute eigenvectors (Fishervectors) which will represents linear discriminant features of each person. The steps to compute eigenvectors from W matrix are given below:

1. Columns of W are eigenvectors satisfying the equation given below:

$$S_B W_i = \lambda_i S_W W_i \quad (6)$$

2. Eigenvalues are roots of the equation given below:

$$\left| S_B - \lambda_i S_W \right| = 0 \quad (7)$$

3. Calculation of eigenvectors by solving the equation given below:

$$(S_B - \lambda_i S_W) W_i = 0 \quad (8)$$

Eigenvectors of highest eigenvalues are selected and eigenvectors with lowest eigenvalues of the data set are ignored. Once eigenvectors are found, the next step is to order them by eigenvalue, highest to lowest. This gives the components in order of significance. Now those components having less eigenvalue can be ignored. If the eigenvalues are small, then it contains a less information about the data. To be precise, if original data have n dimensions in data set and so, n eigenvectors and eigenvalues are gained and then only the first p eigenvectors are chosen then the final data set has only p dimensions.

Now the feature vector is to be calculated. Taking the eigenvectors that we want to keep from the list of eigenvectors and forming a matrix with these eigenvectors in the columns construct this. At first eigenvectors are converted in column vector and then each of them are placed on a matrix in each row.

$$Feature\ vector = (eig_1\ eig_2\ eig_3 \ldots\ldots\ldots\ldots\ eig_p) \quad (9)$$

Finally, we get the feature vector in reduced dimension which can be used in classification process. Linear Discriminant Analysis (LDA) searches for those vectors in the underlying space that best discriminate among classes (rather than those that best describe the data). More formally, given a number of independent features relative to which the data is described, LDA creates a linear combination of these which yields the largest mean differences between the desired classes. Thus theoretically, LDA should give better performance than PCA [25, 26, 27].

5 SELF ORGANIZATION MAPPING BASED LEARNING AND TESTING MODEL

Kohonen self organizing mapping based learning and testing model has been developed for the proposed speaker identification system. The SOM is unlike most classification or clustering techniques in that it provides a topological ordering [28] of the classes. The architectural view of Kohonen self-organization map is shown on figure 3. The output grid contains the nodes where each of the node will win for each of the input set. For weight adaptation, the following equation has been used.

$$w_{ij}(t+1) = w_{ij}(t) + \eta(t)(x_i(t) - w_{ij}(t)) \quad (10)$$

In this work, 200 nodes have been used on the output grid and the neighborhood size has been decreased according to the increment of adaptation process which is shown in figure 4. The learning procedure has been stopped when the minimum distance of a grid node is zero from the input nodes by using the equation,

$$d_j = \sum_{i=0}^{n-1} (x_i(t) - w_{ij}(t))^2 \quad (11)$$

Where d_j is the distance between the inputs and each output node j, x_i is the input to node i and w_{ij} the weight from input i to node j.

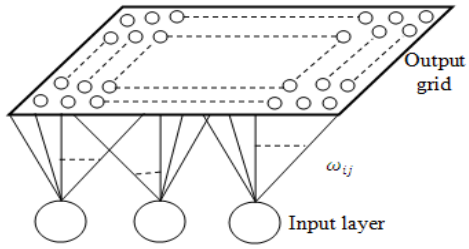


Figure 3: Architecture of Kohonen Self Organizing network.

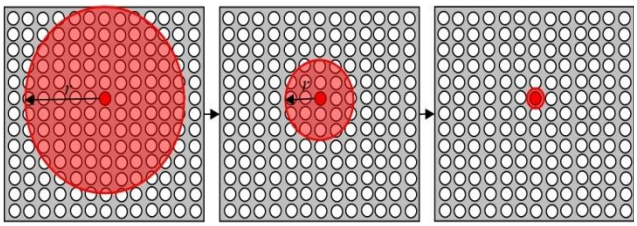


Figure 4: Reduction of topological neighborhood of Kohonen self-organizing network.

6 Experimental Results and Performance Analysis

NOIZEUS speech database has been used to measure the accuracy of the proposed speaker identification system. NOIZEUS speech corpus [29] was developed to facilitate comparison of speech enhancement algorithms among research groups. The noisy database contains 30 IEEE sentences (produced by three male and three female speakers) corrupted by eight different real-world noises at different SNRs. The noise was taken from the AURORA database and includes suburban train noise, babble, car, exhibition hall, restaurant, street, airport and train station noise.

Thirty sentences from the IEEE sentence database were recorded in a sound proof booth. The sentences were produced by three male and three female speakers. The IEEE database (720 sentences) was used as it contains phonetically balanced sentences with relatively low word context predictability. The thirty sentences were selected from the IEEE database so as to include all phonemes in the American English language. The sentences were originally sampled at 25 kHz and downsampled to 8 kHz [30].

One clean speech has been used for training and four others are used for testing purposes where noise addition varies from 0dB to 15dB with 5dB interval. The results of the proposed speaker identification system are shown in the table 1 to 9. The summarization of the results is shown on table 9. In this experiment, eight different real world environmental noises are added to successfully evaluate the performance of the proposed speaker identification system.

Table 1 to table 9 show the experimental results among various standard feature extraction techniques i.e., LPC, LPCC, MFCC, ΔMFCC, ΔΔMFCC. The results are taken with four different variations of SNRs such as 0dB, 5dB, 10dB and 15dB. To measure the overall performance of the proposed speaker identification, average performance is taken which is shown in table 9. From the average identification rate it has been found that MFCC can gives highest identification rate than other feature extraction method.

TABLE 1 Airport Noise Average Identification Rate (%) for NOIZEUS Speech Corpus

Method / SNR	MFCC	ΔMFCC	ΔΔMFCC	LPC	LPCC
15dB	87.00	88.00	64.75	68.00	69.25
10dB	81.25	85.50	57.50	62.25	65.00
5dB	67.75	75.00	48.25	54.50	58.25
0dB	59.00	67.25	50.00	55.50	50.50
Average	73.75	78.94	55.13	60.06	60.75

TABLE 2 Babble Noise Average Identification Rate (%) for NOIZEUS Speech Corpus

Method / SNR	MFCC	ΔMFCC	ΔΔMFCC	LPC	LPCC
15dB	75.00	87.50	57.75	56.25	72.00
10dB	71.25	83.50	47.25	48.00	68.25
5dB	57.50	67.25	36.50	47.75	62.25
0dB	55.00	55.50	36.67	44.25	57.25
Average	64.69	73.44	44.54	49.06	64.94

TABLE 3 Car Noise Average Identification Rate (%) for NOIZEUS Speech Corpus

Method / SNR	MFCC	ΔMFCC	ΔΔMFCC	LPC	LPCC
15dB	68.25	88.25	56.50	67.75	68.50
10dB	65.50	74.50	45.00	55.25	60.00
5dB	55.50	67.25	45.25	55.50	60.25
0dB	55.5-	58.00	39.00	45.00	55.00
Average	61.19	72.00	46.44	55.88	60.94

TABLE 4 Exhibition Hall Noise Average Identification Rate (%) for NOIZEUS Speech Corpus

Method / SNR	MFCC	ΔMFCC	ΔΔMFCC	LPC	LPCC
15dB	73.00	82.25	60.50	65.00	72.25
10dB	70.25	70.00	54.50	63.50	65.00
5dB	65.50	68.25	50.25	62.50	61.50
0dB	57.75	62.25	45.00	53.75	55.50
Average	66.63	70.69	52.56	61.19	63.56

TABLE 5 Restaurant Noise Average Identification Rate (%) for NOIZEUS Speech Corpus

Method / SNR	MFCC	ΔMFCC	ΔΔMFCC	LPC	LPCC
15dB	73.50	80.50	44.25	71.25	72.25
10dB	70.00	70.75	42.25	65.50	61.50
5dB	62.50	65.50	42.25	52.25	60.00
0dB	48.00	52.25	37.75	50.50	53.50
Average	63.50	67.25	41.63	59.88	61.81

TABLE 6 STREET NOISE AVERAGE IDENTIFICATION RATE (%) FOR NOIZEUS SPEECH CORPUS

Method / SNR	MFCC	ΔMFCC	ΔΔMFCC	LPC	LPCC
15dB	75.50	80.25	52.25	67.75	75.00
10dB	67.75	70.00	45.50	53.50	65.25
5dB	65.50	69.25	42.25	67.75	63.25
0dB	53.25	65.50	35.50	55.50	52.25
Average	65.50	71.25	43.88	61.13	63.94

TABLE 7 TRAIN NOISE AVERAGE IDENTIFICATION RATE (%) FOR NOIZEUS SPEECH CORPUS

Method / SNR	MFCC	ΔMFCC	ΔΔMFCC	LPC	LPCC
15dB	73.50	75.50	48.00	57.25	70.00
10dB	65.50	70.00	42.25	53.75	62.25
5dB	50.50	60.25	40.25	50.50	50.50
0dB	45.50	60.50	35.00	48.00	50.00
Average	58.75	66.56	41.38	52.38	58.19

TABLE 8 TRAIN STATION NOISE AVERAGE IDENTIFICATION RATE (%) FOR NOIZEUS SPEECH CORPUS

Method / SNR	MFCC	ΔMFCC	ΔΔMFCC	LPC	LPCC
15dB	65.00	78.50	42.50	55.50	62.50
10dB	60.50	60.25	40.00	50.00	60.00
5dB	49.50	50.50	37.75	43.25	50.50
0dB	45.50	50.00	35.00	38.50	45.25
Average	55.13	59.81	38.81	46.81	54.56

TABLE 9 OVERALL AVERAGE SPEAKER IDENTIFICATION RATE (%) FOR NOIZEUS SPEECH CORPUS

Method / Various Noises	MFCC	Δ MFCC	ΔΔ MFCC	LPC	LPCC
Airport Noise	73.75	66.57	64.60	62.78	61.45
Babble Noise	64.69	64.00	63.13	61.63	60.85
Car Noise	61.19	64.21	62.32	60.95	60.64
Exhibition Hall Noise	66.63	63.60	61.06	60.44	60.63
Restaurant Noise	63.50	60.72	60.01	60.39	60.83
Street Noise	65.50	60.76	60.41	60.78	61.01
Train Noise	58.75	59.17	60.29	60.91	61.09
Train Station Noise	55.13	59.39	60.85	61.22	61.18
Average Identification Rate (%)	63.64	62.30	61.58	61.14	60.96

7 CONCLUSIONS AND OBSERVATIONS

From the experimental results and performance analysis, it can say that the performance of the proposed self-organized mapping based speaker identification system is well enough for the real life applications. In this work, Linear Discriminant Analysis based dimensionality reduction technique has been used which can work well for inter-class variations. Effective noise removing technique and speech preprocessing techniques are used to achieve the highest efficiency. Though the highest identification rate of the proposed system has been found at 63.64%, the performance can be tested by using large speech database. Another feature extraction technique such as wavelet based speech feature extraction technique and other classification techniques i.e., Hidden Markov Model, Gaussian Mixture Model, Genetic Algorithm etc may be used to enhance the efficiency of the system.

REFERENCES

[1] S. Furui, "40 years of progress in automatic speaker recognition," *Lecture Notes Comput. Sci.,* vol. 5558, pp. 1050–1059, 2009.

[2] J. P. Campbell, "Speaker recognition: A tutorial," *Proc. IEEE*, vol. 85, no. 9, pp. 1437–1462, Sep. 1997.

[3] Jacobsen, J. D., "Probabilistic Speech Detection," *Informatics and Mathematical Modeling,* DTU, 2003.

[4] Jain, A., R.P.W.Duin, and J.Mao., "Statistical pattern recognition: a review," *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 22, pp. 4–37, 2000.

[5] Reynolds, D.A., "Experimental evaluation of features for robust speaker identification," *IEEE Transactions on SAP,* vol. 2, pp. 639-643, 1994.

[6] Wu, D., Morris, A.C. & Koreman, J., "MLP Internal Representation as Disciminant Features for Improved Speaker Recognition," *Proc. NOLISP2005,* Barcelona, Spain, pp. 25-33, 2005.

[7] Picone, J., "Signal modeling techniques in speech recognition," *Proceedings of the IEEE,* vol. 81, no. 9, pp. 1215–1247, 1993.

[8] L.P. Cordella, P. Foggia, C. Sansone, M. Vento., "A Real-Time Text-Independent Speaker Identification System," *Proceedings of 12th International Conference on Image Analysis and Processing,* IEEE Computer Society Press, Mantova, Italy, pp. 632 – 637, 2003.

[9] D. Kewley-Port and Y. Zheng, "Auditory models of formant frequency discrimination for isolated vowels," *Journal of the Acostical Society of America,* vol. 103, no. 3, pp. 1654–1666, 1998.

[10] A. Solomonoff, W.M. Campbell, and I. Boardman, "Advances in channel compensation for SVM speaker recognition," *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing,* pp. 629–632, Philadelphia, USA, March, 2005.

[11] P. Kenny, P. Ouellet, N. Dehak, V. Gupta, and P. Dumouchel, "A study of interspeaker variability in speaker verification," *IEEE Trans. Audio, Speech and Language Processing,* vol. 16, no. 5, pp.980 –988, July 2008.

[12] N. Dehak, P. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-end factor analysis for speaker verification," *IEEE Trans. Audio, Speech and Language Processing,* vol. 19, no. 4, pp.788 –798, May 2011.

[13] J. Ming, T.J. Hazen, J.R. Glass, and D.A. Reynolds, "Robust speaker recognition in noisy conditions," *IEEE Trans. Audio, Speech and Language Processing,* vol. 15, no. 5, pp.1711 –1723, July 2007.

[14] D. Garcia-Romero, X. Zhou, and C. Y. Espy-Wilson, "Multicondition training of gaussian plda models in i-vector space for noise and reverberation robust speaker

recognition," *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing,* pp. 4157-4260, March 2012.

[15] C.M. Vannicola, B.Y. Smolenski, B. Battles, and P.A. Ardis, "Mitigation of reverberation on speaker identification via homomorphic filtering of the linear prediction residual," *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing,*, pp. 5512–5515, May 2011.

[16] R. Saeidi, A. Hurmalainen, T. Virtanen and D. A. van Leeuwen, "Exemplar-based Sparse Representation and Sparse Discrimination for Noise Robust Speaker Identification," *Proc. Odyssey: the speaker and language recognition workshop,* Singapore, June 2012.

[17] Md. Rabiul Islam, Md. Fayzur Rahman and M. Abdus Sobhan, "Genetically Optimized HMM for Robust Speaker Identification in Noise," *International Conference on Advances in Electrical Engineering,* Independent University, Bangladesh, pp. 258-263, Dec 19-20, 2011, Dhaka, Bangladesh.

[18] Gerasimos Potamianos, Chalapathy Neti, and Sabine Deligne, "Joint Audio-Visual Speech Processing for Recognition and Enhancement," *Auditory-Visual Speech Processing Tutorial and Research Workshop (AVSP),* pp. 95-104, St. Jorioz, France, Sep 2003.

[19] Md. Rabiul Islam and Md. Fayzur Rahman, "Hybrid Feature and Decision Fusion based Audio-Visual Speaker Identification in Challenging Environment," *International Journal of Computer Applications (IJCA)*, vol. 9, no. 5, pp. 9-15, Nov 2010.

[20] A. Adjoudant, C. Benoit, "On the integratio of auditory and visual parameters in an HMM-based ASR," *Speechreading by Humans and Machines: Models, Systems, and Speech Recognition, Technologies and Applications*, ed. D.G. Strok and M. E. Hennecke, (Springer, Berlin, Germany, 1996), pp. 461-472.

[21] Rosalind Wang, "Wiener Filtering," *PHYS 3301, Scientific Computing,* Project report for NOISE group, May 2000.

[22] Kitayama, K., Goto, M., Itou, K., and Kobayashi, T., "Speech Starter: Noise-Robust Endpoint Detection by Using Filled Pauses," *Eurospeech 2003*, Geneva, pp. 1237-1240, 2003.

[23] Qi, L., Zheng, J., Tsai, A., and Zhou, Q., "Robust Endpoint Detection and Energy Normalization for Real-Time Speech and Speaker Recognition", *IEEE Transaction on speech and Audion Processing*, vol. 10, no. 3, 2002.

[24] Davis, S. B., and Mermelstein, P., "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-28, no. 4, pp. 357–366, 1980.

[25] Peter N. belhumeur, Joao P. Hespanha, David J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Transaction on Pattern Analysis and Machine Intelligence,* vol. 9, no. 7, pp. 711-720, July 1997.

[26] Kamran Etemad and Rama Chellappa, "Discriminant analysis for recognition of human face images," Department of Electrical Engineering and Center for Automation Research, University of Maryland, College Park, Maryland 20742, A/Vol. 14, No. 8, 1997.

[27] Hongjun Jia and Aleix M. Martinez, "Support Vector Machines in Face Recognition with Occlusions," Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH 43210, USA.

[28] Teuvo Kohonen, "Self-Organizing formation of topologically correct feature maps," *Biol. Cybern.,* vol. 43, pp.59- 69, 1982.

[29] Hu, Y. and Loizou, P., "Subjective comparison of speech enhancement algorithms," *Proc. IEEE Int. Conference on Acoustics, Speech, and Signal Processing (ICASSP),* pp. 153-156, Toulouse, France, May 2006.

[30] Hu, Y. and Loizou, P., "Evaluation of objective quality measures for speech enhancement," *IEEE Transactions on Speech and Audio Processing,* vol. 16, no. 1, pp. 229-238, Jan 2008.

Md. Rabiul Islam was born in Rajshahi, Bangladesh, on December 26, 1981. He received his B.Sc. degree in Computer Science & Engineering and M.Sc. degrees in Electrical & Electronic Engineering in 2004, 2008, respectively from the Rajshahi University of Engineering & Technology, Bangladesh. From 2005 to till now, he is a faculty member in the Department of Computer Science & Engineering at Rajshahi University of Engineering & Technology. He is serving as an associate professor at the same place. He completed his Ph.D. degree programme under the excellent guidance of Dr. Md. Fayzur Rahman, Department of EEE at Rajshahi University of Engineering & Technology, Bangladesh. His research interests include bio-informatics, human-computer interaction, speaker identification and authentication under the neutral and noisy environments.

Md. Fayzur Rahman was born in 1960 in Thakurgaon, Bangladesh. He received the B. Sc. Engineering degree in Electrical & Electronic Engineering from Rajshahi Engineering College, Bangladesh in 1984 and M. Tech degree in Industrial Electronics from S. J. College of Engineering, Mysore, India in 1992. He received the Ph. D. degree in energy and environment electromagnetic from Yeungnam University, South Korea, in 2000. Following his graduation he joined again in his previous job in BIT Rajshahi. Currently he is a Professor in Electrical and Electronics Engineering in Daffodil International University, Bangladesh. He is currently engaged in education in the area of Electronics & Machine Control and Digital signal processing.

Performance of a FSO Link in Presence of Cloud

Md. Mahfuzul Haque [1*], Abu Jahid[2], Md. Mehedi Hasan[3], Pranjan Das[4]
[1-2]Dept. of Electrical and Electronic Engineering
Bangladesh University of Business and Technology (BUBT)
[3]School of Computer Science and Electrical Engineering
University of OTTAWA, Ottawa, Canada
[4]Bangladesh University of Engineering and Technology (BUET)

*mahfuz.ap@gmail.com

Abstract— Over conventional microwave and optical fiber communication systems, Free Space Optical (FSO) communication links have some distinct advantages by virtue of their high carrier frequencies. However, a number of limitations caused by atmospheric phenomena like cloud, fog, aerosols even turbulence makes it difficult to achieve the desired level of performance. In this article, we investigate the bit error rate performance of intensity modulated FSO with direct detection (IM/DD) in single-input single-output (SISO) for the perfect inter symbol interference (ISI) caused by beam broadening at the receiver due to the effect of cloud, assuming that a single information-bearing signal is transmitted.

Keywords— BER, FSO, IM/DD, ISI, SISO

Manuscript received on 12 December 2014 and accepted for publication on 24 June 2015.

1 INTRODUCTION

THE huge amount of data exchange between satellites and ground stations demands enormous capacity that cannot be provided by strictly regulated, scarce resources of the Radio Frequency (RF) spectrum [1]. On the other hand Free Space Optical (FSO) communications has the potential of providing enormous bandwidth. In addition, FSO links are difficult to intercept, immune to interference or jamming from external sources, and because of spatial confinement of laser beams significantly reduces the power loss [2, 3]**.** All these benefits are viable only under ideal channel conditions. But the atmosphere has a detrimental effect on a propagating laser beam which is the main drawback of a FSO link. Various atmospheric phenomena such as clouds, fog, aerosols, and even turbulence can severely degrade the performance by not only scattering the laser beam, giving rise to optical pulse attenuation and broadening in space and times but also disruptions in the amplitude and phase of received signal [4,5]. The broadening of the received optical pulse in space and time causes Inter symbol interference (ISI). The performance of IM/DD FSO systems for different turbulence models has been well studied in the literature. Maximum likelihood sequence detection (MLSD) for IM/DD FSO links was employed by the study of Zhu and Kahn [6]. They continued their study to find out the pairwise error probability of coded FSO links assuming the log-normal distributed turbulence model [7]. A further study of the pairwise error probability for on-off keying (OOK) with temporally correlated *K*-distributed turbulence was carried out by Uysal *et al* in [8].In [3], Ehsan Bayaki et al, find out the pairwise error probability of MIMO FSO link for on-off keying (OOK) with Gamma -Gamma distribution because of its excellent correlation with the measured data for a wide range of turbulence condition (weak to strong) as the log-normal distribution is used for weak turbulence. Pointing errors can affect the FSO system performance since FSO communications requires LOS links. In [9], Farid and Hranilovic presented an FSO channel model considering the effect of fading due to log-normal/Gamma-Gamma atmospheric turbulence and pointing errors by considering beam width, pointing error variance and detector size. The BER of a SISO FSO link impaired by *K*-fading (a special case of Gamma-Gamma fading) and pointing error was expressed in terms of the Meijer's G–function in [10].Later on, this approach has also been extended to MIMO FSO links [11]. Later, Uysal extended their discussion of pair wise error probability for coded OOK FSO links to the cases with independent Gamma-Gamma turbulence [12]. Another new approach, a multiple symbol detection decision metric for OOK in both log-normal and Gamma-Gamma turbulence was investigated by Riediger *et al. in* [13]. Vavoulas et al. demonstrated the design for robust FSO link for various weather conditions, cases like snow, fog and rain [14]. But In this paper, the effect of inter symbol interference caused by cloud is investigated for a SISO free space optical (FSO) communication scheme.

2 SYSTEM MODEL

Basing upon Monte Carlo simulations mathematical models are developed for the temporal characteristics of optical pulse propagation through clouds. These include temporal impulse response, transfer function, bandwidth, received energy and board analysis. The simulation results strongly supports the use of double gamma function model to best describe optical pulse spread through clouds [15]. A FSO link is normally consists of a transmitter, a channel which would be the medium of transmission and it would be atmospheric that includes cloud and finally a receiver to reproduce that transmitted signal. The block diagram is shown in Figure 1.

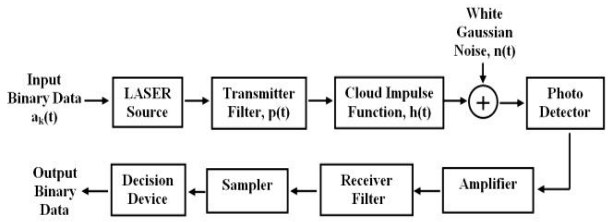


Figure 1: Basic FSO system model

Most practical wireless optical channels use light emitting diodes or laser diodes as transmitters and photodiodes as detectors as shown above. These device modulate and detect solely the intensity of carriers not its phase which implies that all transmitted signal intensities are non-negative.

The input binary data is used to modulate the laser using intensity modulation and thus pass through the transmitted filter and then it passes through the atmosphere where the impulse response has a greater impact. The optical signal is detected by a photo detector and received by the receiver circuit. The sampler and decision device is used to determine the output binary data.

2.1 Intensity Modulated Direct Detection (IM/DD)

When the optical power output of a source is varied in accordance with some characteristics of the modulating signal then this modulation from is named as intensity modulation and in the receiver side direct detection technique is used for the regeneration of the modulating signal.

2.2 Cumulus Cloud Model and Gamma Constant

Cumulus clouds are generally located at the height of 200 and 20000 feet over the ground. This elevation range is very relevant for space communications involving planes. Low lever stratus clouds are usually between 1000 and 2000 feet elevation and as such are unsuitable for aircraft communications. Since particulate scatter close to receiver is of more serious consequences than scatter far away from it. Therefore we will concentrate our study on cumulus clouds.

3 PERFORMANCE ANALYSIS

In this section we present mathematical analysis of BER of SISO (Single-Input Single-output) FSO link for Intensity Modulation – Direct Detection (IM / DD). For the perfor mance analysis of the FSO link signal we have considered some parameters that are listed below-

TABLE 1
DOUBLE GAMMA FUNCTION CONSTANTS: CLOUD THICKNESS OF 200 M [15]

Gamma Function Constant	Wavelengths		
	0.532 μm	0.8 μm	1.3 μm
k_1	120.1	62.4	16.5
k_2	1.9×10^7	1.8×10^7	1.1×10^7
k_3	1.55	2.9	0.67
k_4	3×10^6	3.5×10^6	2.13×10^6

Optical radiation propagating through clouds experiences temporal distortions. A function that describes well the temporal impulse response is the double gamma function [15]-

TABLE 2
DOUBLE GAMMA FUNCTION CONSTANTS: CLOUD THICKNESS OF 200 M [15]

Gamma Function Constant	Wavelengths		
	0.532 μm	0.8 μm	1.3 μm
k_1	12.4	5.2	2
k_2	1.1×10^7	0.8×10^7	0.71×10^7
k_3	0.66	0.41	0.3
k_4	2.4×10^6	1.9×10^6	1.8×10^6

$$h(t) = \{k_1(c_1)te^{-[k_2(c_1)]t} + k_2(c_1)te^{-[k_{42}(c_1)]t}\}U(t) \quad (1)$$

Where, h (t) is in m^{-2}, c_1 is a parameter defining the physical characteristics of the optical channel such as particulate size distribution, particulate refractive index, geometrical cloud thickness and radiation wavelengths, and k_1-k_4 are the gamma function constants depending on c_1 and U (t) is a unit step function. The temporal frequency transfer function can be evaluated by Fourier transforming the temporal impulse response [15]

$$H(f) = \int_{-\infty}^{\infty} h(t)e^{-(j2\pi ft)}d(t) \quad (2)$$

Where, f is the temporal frequency (Hz). Substituting (1) into (2) yields-

$$H(f) = \{\frac{k_1(c_1)}{[k_2(c_1 + j2\pi f)]^2} + \frac{k_3(c_1)}{[k_4(c_1 + j2\pi f)]^2}\} \quad (3)$$

$$H(f)=G\left[1+j\left(\frac{f-b}{f_3}\right)\right]\left[1+j\left(\frac{f+b}{f_3}\right)\right]\Big/\left[\left[1+j\left(\frac{f}{f_3}\right)\right]^2\left[1+j\left(\frac{f}{f_2}\right)\right]^2\right] \quad (4)$$

For a given set of parameters [15]- The transmitted optical signal is given by

$f_1=\frac{k_2}{2\pi}$	$f_2=\frac{k_4}{2\pi}$	$f_2=\frac{(k_1k_4+k_3k_2)}{2\pi(k_1+k_3)}$
$G=\frac{4\pi^2(k_1+k_3)}{(k_2k_4)^2}k_3^2$	$b=\frac{4\pi^2(k_1+k_3)}{(k_2k_4)^2}f_3^2$	

Where, p_T is the transmitted optical power, a_k is the k-th

$$s(t)=\sqrt{2pr}\sum_{k=-\alpha}^{\alpha}a_k\,p(t-kT_b)e^{(jw_ct)} \quad (5)$$

information bit whose value is 1 and 0, p (t) is the optical pulse shape of bit duration T_b and carrier frequency of fc.

The received optical signal is given by

$$r(t)=\sqrt{2p_s}\sum_{k=-\alpha}^{\alpha}a_k\,p(t-kT_b)e^{(jw_ct)} \quad (6)$$

where, P_s is the received optical power and g (t) = h (t) $\otimes$ p(t) is the received optical pulse shape which overlaps over a number of bits and produce Inter Symbol Interference (ISI) .

$$\begin{aligned} i(t)&=|r(t)|^2R_d \\ &=2R_dP_S\left|\sum_{k=-\infty}^{\infty}a_kg(t-kT_b)\right|^2+i_n(t) \end{aligned} \quad (7)$$

The photo current can be expressed as –
where, R_d is the responsivity of the detector and $i_n(t)$ is the noise current due to photo diode and receiver noise current which can be expressed as [17]

$$i_n(t)=i_{sh}(t)+i_{th}(t) \quad (8)$$

SINR can be defined as the ratio of signal power to noise power plus interference power, i.e.

$$SINR=\frac{SignalPower}{NoisePower+InterferencePower}=\left[\frac{i_s^2}{\sigma_n^2+\sigma isi^2}\right] \quad (9)$$

$$i_s(t)=2R_dP_S|a_0|^2.|g(t)|^2$$

Mean Signal Current,

$$I_s(t)=\overline{i_s(t)^2}=2R_dp_s\frac{1}{T_b}\int_0^{T_s}|g(t)|^2dt$$

Mean ISI Current,

$$\sigma_{isi}(t)=\overline{\sigma_{isi}(t)^2}=2R_dp_s\frac{1}{T_b}\int_0^{T_s}|a_kg(t-kT_b)|^2dt$$

Where,
Noise Power, $\sigma_n^2=\sigma_{shot}^2+\sigma_{th}^2$
Shot Noise Power, $\sigma_{shot}^2=2eBIs$
Thermal Noise Power, $\sigma_{th}^2=4KTB/R_L$

B= Signal Bandwidth (Hz), R_L= Load Resistance
K= Boltzmann Constant, T= Ambient Temperature

The expression of BER for Intensity Modulation Direct Detection (IM / DD) can be expressed as [16]

$$BER=0.5erfc\left(\frac{\sqrt{SINR}}{2\sqrt{2}}\right) \quad (10)$$

3 RESULTS AND DISCUSSIONS

In this section we compare the BER performance of SISO optical communication system in presence of cloud. We present the numerical results of BER of the FSO link considering the effect of cloud and without could. We make the theoretical analysis of the system for the given set of Gama function constant for different wavelengths mentioned in Table 1 and Table 2. We used MATLAB tools for this simulation. Figure 2, demonstrates the received optical pulse shape for different carrier wavelength for a transmission bandwidth of 1GHz. It shows that the broadening of the optical pulses arriving at the receiver due to the scattering cloud. This broadening of pulse causes inter-symbol interference (ISI) at the receiver.

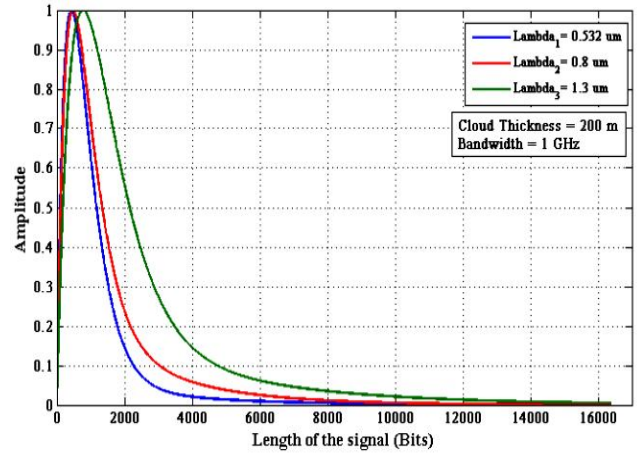


Figure 2: Received optical pulse shape

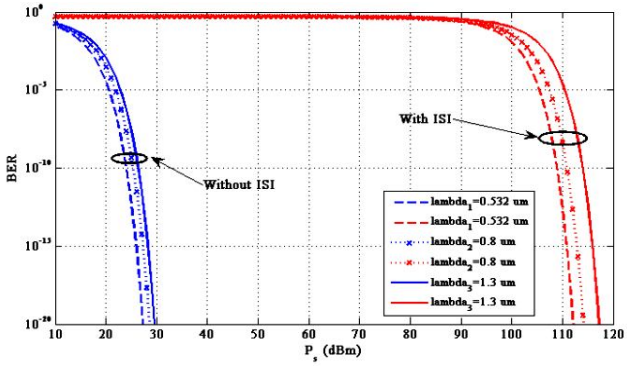


Figure 3: BER performance of a FSO link with ISI and without ISI for different wavelengths by taking the cloud thickness as 200 m (for the parameters listed in Table 1)

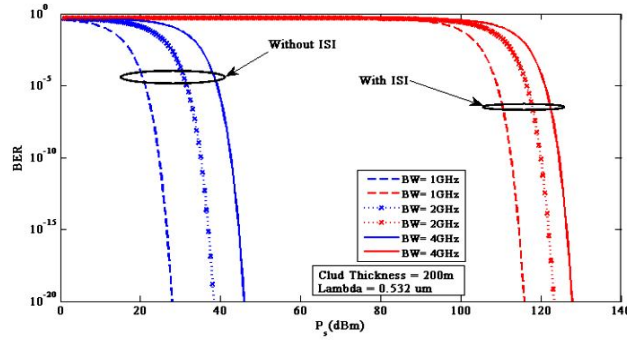


Figure 5: BER performance of a FSO link with ISI and without ISI for different Bandwidths for λ=0.532μm (for the parameters listed in Table 1)

In Figure 3, we present the graphical representation of BER vs. received power in dBm considering the effect of ISI and without ISI. It shows that there exists a significant variation in received power due to the effect of cloud for all optical carrier wavelengths. The received power without the inter symbol interference is around 25-30 dBm, whereas the required received power to overcome the inter symbol interference for maintaining a specific bit error 10^{-10} is 105-110 dBm. This value is unrealistic. It shows unrealistic values of power required for transmission in the presence of cloud.

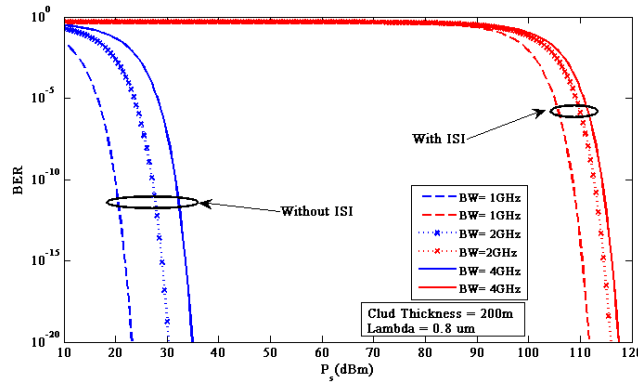


Figure 6: BER performance of a FSO link with ISI and without ISI for different Bandwidths for λ=0.8μm (for the parameters listed in Table 1)

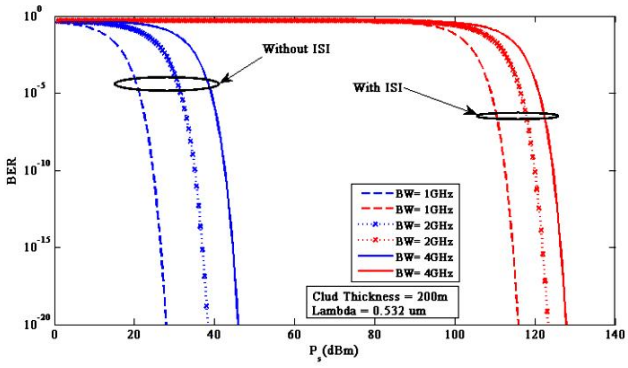


Figure 4: BER performance of a FSO link with ISI and without ISI for different wavelengths by taking the cloud thickness as 250 m (for the parameters listed in Table 2)

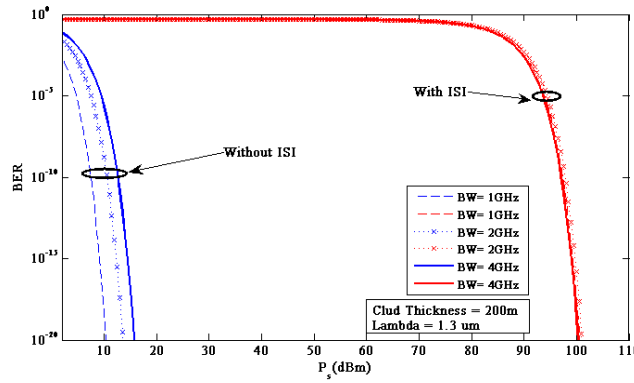


Figure 7: BER performance of a FSO link with ISI and without ISI for different Bandwidths for λ=1.3μm (for the parameters listed in Table 1)

In Fig. 4, we demonstrate the effect of cloud for all carrier wavelengths considering the transmission bandwidth of 1GHz for the parameters listed in Table II. For further investigation we varied the transmission bandwidth for different carrier wavelength communication. Fig. 5, 6, and 7 illustrates the BER performance of a SISO link for various bandwidths.

The whole performance of the optical system can be quantified in terms of power penalty for better understanding. The power penalty is defined as the increase of optical power required to overcome a given effects with respect to an ideal system.

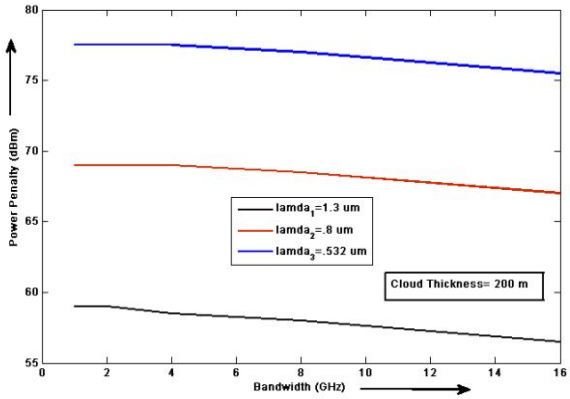


Figure 8: Relation between Power Penalty and Bandwidths for different wavelengths by taking the cloud thickness as 200 m (for the parameters listed in Table 1)

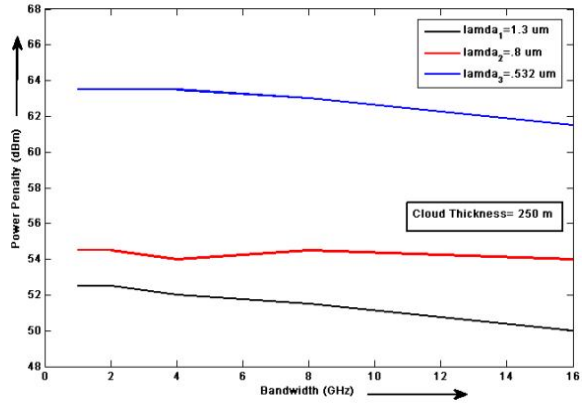


Figure 9: Relation between Power Penalty and Bandwidths for different wavelengths by taking the cloud thickness as 250 m (for the parameters listed in Table 2)

Fig. 8 and Fig. 9 shows the required optical power in dBm for various transmission bandwidths for maintaining a bit error rate (BER) of 10^{-10}.From Fig. 8, the required optical power (for the cloud thickness of 200 m) to overcome the inter symbol interference varies about 55 dBm to 59 dBm for a wide range of transmission bit rate (1Gb/s-16 Gb/s) for the carrier wavelength of λ=1.3 μm. While the required optical power for the wavelength of λ=0.8 μm is about 67 dBm to 69 dBm. Whereas for λ=0.532 μm the value is about 74 dBm to 78 dBm for the same bit rate. In Figure 9, the required optical power (for the cloud thickness of 250 m) to overcome the inter symbol interference varies about 50 dBm to 53 dBm for a wide range of transmission bit rate (1Gb/s-16 Gb/s) for the carrier wavelength of λ=1.3 μm. While the required optical power for the wavelength of λ=0.8 μm is about 54 dBm to 55 dBm. Whereas for λ=0.532 μm the value is about 60 dBm to 65 dBm for the same bit rate.

4. CONCLUSION

In this paper, we have presented a very basic approach to performance analysis of SISO free space optical systems (FSO) considering the effect of cloud. The proposed technique is based on finding the noise power caused by inter symbol interference (ISI) from the received pulse shape by using MATLAB tools for a single bit transmission. Then we investigated the BER performances for various optical wavelength communications by varying transmission bit rate. Then we also quantified the performance of the link in terms power penalty. It is found for SISO free space optical link, for any transmission bit rate under the effect of cloud thickness of 200 m and 250 m, the required received power value is unrealistic (more than 100 dBm) which suggest no communication is possible using this modulation scheme. This area badly needs an experimental research demonstration for possibilities of communication as FSO link is very important issue for future optical networking. Further research can be carried to overcome this problem. There is ample scope to conduct more detail research in this field for a concrete decision and subsequent remedy.

References

[1] Ivan B. Djordjevic, Bane Vasic and Mark A. Neifield, "Multilevel Coding in Free-Space Optical MIMO Transmission with Q-ary PPM Over the Atmospheric Turbulence Channel, IEEE Photonics Technology Letter, Vol. 18, No. 14, (2006).

[2] Zeinab Hajjarian and Jarir Fadlullah, Mohsen Kavehrad, MIMO Free Space Optical Communications in Turbid and Turbulent Atmosphere(Invited Paper),The Pennsylvania State University Department of Electrical Engineering, University Park, PA. 16802, USA.

[3] Ehsan Bayaki, Robert Schober, and Ranjan K.Mallick, Performance Analysis of Free–Space Optical Systems in Gamma–Gamma Fading, IEEE Global Communications Conference (Globecom), New Orleans, Dec. (2008).

[4] Hanling Wu, Haixing Yan, Xinyang Li, "Performance analysis of bit error rate for free space optical communication with tip-tilt compensation based on gamma–gamma distribution," Optica Applicata, Vol. 39, No. 3, (2009).

[5] S. Mohammad Navidpour and Mohsen Kaverhrad, "BER Performance of Free Space optical Transmission with Spatial Diversity," IEEE Transactions on Wireless Communications. Vol. 6, No. 08, (2007).

[6] Zhu X, Kahn J M, "Free-space optical communication through atmospheric turbulence channels," IEEE Trans. Commun. 50(8): (2002)1293-1300.

[7] Zhu X, Kahn J M, "Performance bounds for coded free-space optical communications through atmospheric turbulence channels," IEEE Trans. Commun. 51(8): (2003)1233-1239.

[8] Uysal M, Navidpour S M, Li J, "Error rate performance of coded free-space optical links over strong turbulence channels," IEEE Commun. Lett. 8(10): (2004) 635-637.

[9] Farid A A, Hranilovic S "Outage capacity optimization for free-space optical links with pointing errors, IEEE/OSA J. Light wave Technol. 25(7): (2007)1702-1710.

[10] Sandalidis H G, Tsiftsis T A, Karagiannidis G K, Uysal M, "BER performance of FSO links over strong atmospheric turbulence channels with pointing errors," IEEE Commun.Lett. 12(1): (2008) 44-46.

[11] T. A. Tsiftsis, H. G. Sandalidis, G. K. Karagiannidis, and M. Uysal, "FSO links with spatial diversity over strong atmos-

pheric turbulence channels," in Proc. IEEE Intern. Conf. Commun. (ICC), (2008) pp. 5379–5384.

[12] Uysal M,Li J, Yu M, "Error rate performance analysis of coded free-space optical links over Gamma-Gamma atmospheric turbulence channels," IEEE Trans. Wireless Commun. 5(6): (2006) 1229-1233.

[13] Riediger ML B, Schober R, Lampe L, "Fast multiple-symbol detection for free-space optical communications," IEEE Trans. Wireless Commun. 57(4) (2009) 1119-1128.

[14] Alexander Vavoulas, Harilaos G. Sandalidis, and Dimitris Varoutas, "Weather Effects on FSO Network Connectivity" J. OPT. COMMUN. NETW. / VOL. 4, NO. 10 / OCTOBER 2012

[15] S. Arnon, D. Sadot and N.S. Kopeika, "Analysis of Optical Pulse Distortion through Clouds for Satellite to Earth Adaptive Optical Communication, Journal of Modern Optics," Vol. 41, No. 8, (1994), 1591-1605.

[16] John. M. Senior, "Optical Fibre Communications"- Principles and Practice, 3rd edition, Pearson Education.

[17] Pranjan Das, "Analysis of a Free Space Optical (FSO) Link Under The Influence of Spatial Widening Effect From Propagation Through Clouds", 2011.

Md. Mahfuzul Haque is currently working as Lecturer in the Department of Electrical & Electronic Engineering at Bangladesh University of Business &Technology (BUBT). Previously he worked as a Lecturer in the Department of Information & Communication Technology at Metropolitan University, Sylhet. He received B.Sc. & M.S. degrees in Applied Physics, Electronic & Communication Engineering from the University of Dhaka. His research interest includes wireless communication, Semiconductor Technology and Opto Electronics.

Abu Jahid is currently working as Lecturer in the Department of Electrical & Electronic Engineering at Bangladesh University of Business &Technology (BUBT). Previously he worked as a BSS Engineer in Huawei Technology BD. He received Bachelor degree in 2009 from MIST. And currently he is continuing his Master degree from MIST. His research interest includes wireless communication and free space optical communication.

Md. Mehedi Hasan is currently doing M. Sc. degree in the School of Computer Science and Electrical Engineering at the University of Ottawa, Canada. Previously he was working as a Lecturer in MIST and Assistant Professor in the Department of Electrical & Electronic Engineering at Bangladesh University of Business &Technology (BUBT). He received Bachelor degree in 2008 from MIST. His research interest includes wireless sensor networking.

Pranjan Das achieved B.Sc. and M.Sc. from Bangladesh University of Engineering and Technology (BUET) in 2008 and 2011 respectively. His research interest is fiber optic communication and MC CDMA.

Impact of Temperature on Threshold Voltage of Gate-All-Around Junctionless Nanowire Field-Effect Transistor

A. Arefin
Department of Electrical & Electronic Engineering
Northern University Bangladesh
arefin.mist@gmail.com

Abstract—Commonly used transistors are based on the use of semiconductor junctions formed by introducing doping atoms into the semiconductor material. As the distance between junctions in modern devices drops below 10 nm, extraordinarily high doping concentration gradients become necessary. For this reason, a new device was proposed which has full CMOS functionality and is made by using junctionless nanowires. They have near-ideal sub-threshold slope, extremely low leakage currents and less degradation of mobility with gate voltage and temperature than classical transistors. Among several types of field effect transistors, gate-all-around junctionless nanowire FET (GAA-JL-NW-FET) is the recently invented one. In this article, temperature dependency of threshold voltage of GAA-JL-NW-FET has been analyzed for different channel materials such as Si, GaAs, InAs and InP. From the simulation result, it is observed that the threshold voltage is minimum for InAs and it decreases when the temperature is increased for all the above mentioned channel materials.

Keywords—Gate-All-Around, Junctionless, Nanowire, Temperature dependency, Threshold voltage.

Manuscript received on 10 August 2015 and accepted for publication on 18 October 2015.

1 Introduction

NANOELECTRONIC devices have been used in very extreme environments like low temperature ranges as well as high temperature ranges. So their behavior in response to temperature has attractted the attention of researchers. Different studies and investigations are going on to understand the performance of all types of transitors in different temperatures. Threshold voltage, one of the major parameters that can be influenced very easily by temperature change, has been taken under consideration in this article [1].

The existing transistors are based on the formation of junctions. Junctions are capable of both blocking current and allowing it to flow, depending on an applied bias. The most common junction is the p-n junction. Other types of junctions include the metal-semiconductor 'Schottky' junction and the heterojunction, which is a p–n junction comprising two different semiconductor materials. The bipolar junction transistor (BJT) contains two p-n junctions, and so do the metal-oxide–semiconductor field-effect transistor (MOSFET) and most of the modern transistors. The junction field-effect transistor (JFET) has only one p–n junction and the metal–semiconductor field-effect transistor (MESFET) contains a Schottky junction.

Conventional MOSFETs are usually fabricated on a semiconductor bulk substrate. With the continuous scaling of the device dimension, these bulk MOSFETs are facing serious challenges, such as the increasing gate leakage current and more serious short channel effects [2]. Gate-all-around (GAA) nanowire (NW) transistors are regarded as a promising device structure to extend the scaling limit due to their superior gate controllability. On the other hand, the formation of the abrupt source and drain junctions in the conventional NW devices imposes severe challenges on doping techniques and thermal budget [3]. As an alternative, junctionless (JL) devices with a uniform doping concentration and type throughout the channel and source/drain extensions are fabricated to overcome these challenges.

In recent times transistors have become of nanometer sizes, due to the aggressive scaling. At this size it is very hard to control the sharp source/drain-channel junctions from the device fabrication point of view. Also many other unwanted effects such as gate leakage, short channel effects, hot carrier effects etc. have been seen to be increasing. For reducing the short channel effects the GAA-FETs are the best since they provide the best control over the channel from all around [4]. But if channel with corners, i.e. rectangular shape, is used for this purpose then it leads another effect known as corner effect [5]. To avoid this, cylindrical structure looks to be promising and has been widely used for getting rid of corner effect and also to improve other short channel effect performance parameters.

The first JL transistor was introduced back in 1920s [6], [7], but it has attracted researches recently. JL transistors are inherently depletion mode devices and needs a gate voltage to be applied to make them OFF [8]. One advan-

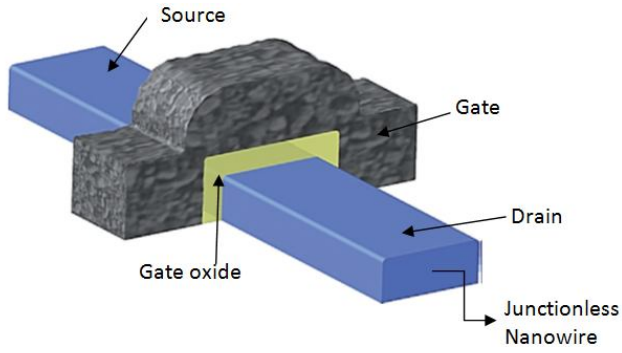


Figure 1: Schemetic view of a JL-NW-FET. Here souce, channel and drain are consists of a single JL NW. The underlying insulator (buried oxide) is not shown [10].

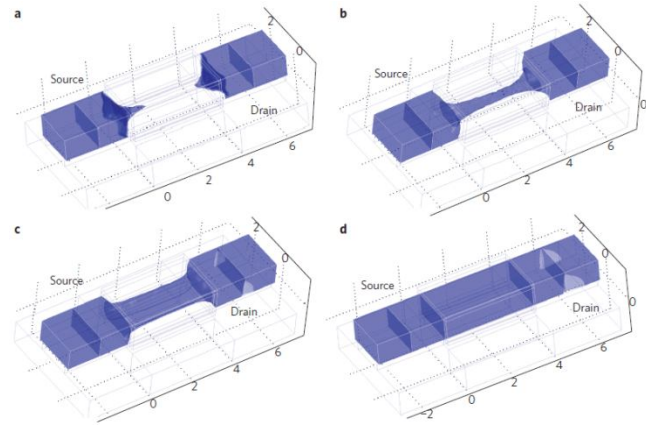


Figure 2: Electron concentration contour plots in an n-type junctionless gated resistor. a–d, Plots result from simulations carried out for a drain voltage of 50 mV and for different gate voltage (V_G) values: below threshold ($V_G<V_{TH}$) the channel region is depleted of electrons (a); at threshold ($V_G=V_{TH}$) a string-shaped channel of neutral n-type Si connects source and drain (b); above threshold ($V_G>V_{TH}$) the channel neutral n-type Si expands in width and thickness (c); when a flat energy bands situation is reached ($V_G=V_{FB}>>V_{TH}$) the channel region has become a simple resistor (d). The plots were generated by solving the Poisson equation and the drift-diffusion and continuity equations self-consistently. The device has a channel width, height and length of 20, 10 and 40 nm, respectively. The n-type doping concentration is 1×10^{19} cm^{-3} [10].

tage of these are that it operates due to the conduction of carriers in bulk as opposed to the conventional MOSFETs (inversion mode devices), in which the carriers conducts through a thin layer of inversion charge layer created at the oxide-semiconductor interface. Hence, the current driving capability is improved a lot. However, it is also degraded slightly due to the high doping concentrations used in the channel.

2 GAA-JL-NW-FET

2.1 NW-FET

Semiconductor NWs have attracted significant interest because of their potential for a variety of different applications, including logic and memory circuitry, photonic devices, and chemical and biomolecular sensors [9].

2.2 JL-NW-FET

Because of the laws of diffusion and the statistical nature of the distribution of the doping atoms, conventional junctions represent an increasingly difficult fabrication challenge for the semiconductor industry. A new type of transistor has been proposed and demonstrated in which there are no junctions and no doping concentration gradients [10]. These devices have full CMOS functionality and are made using NWs. They have near-ideal subthreshold slope, extremely low leakage currents, and less degradation of mobility with gate voltage and temperature than classical transistors.

Figure 1 presents a schematic view of a JL-NW-FET. Having no junctions presents a great advantage. Modern transistors have reached such small dimensions that ultra sharp doping concentration gradients are required in junctions. Typically the doping must switch from n-type with a concentration of 1×10^{19} cm^{-3} to p-type with a concentration of 1×10^{18} cm^{-3} within a couple of nanometers. This imposes severe limitations on the processing thermal budget and necessitates the development of costly millisecond annealing techniques. In a JL gated resistor, on the other hand, the doping concentration in the channel is identical to that in the source and drain. Because the gradient of the doping concentration between source and channel or drain and channel is zero, no diffusion can take place, which eliminates the need for costly ultrafast annealing techniques and allows one to fabricate devices with shorter channels. The key of fabricating a JL gated resistor is the formation of a semiconductor layer that is thin and narrow enough to allow full depletion of carriers when the device is turned off. The semiconductor also needs to be heavily doped to allow for a reasonable amount of current flow when the device is turned on. Putting these two constraints together imposes the use of nanoscale dimensions and high doping concentrations.

In a MOSFET, carriers are confined in an inversion channel in which scattering events rapidly increase in frequency with gate voltage, thereby decreasing transconductance and current drive [11]. In the heavily doped gated resistor, the drain current essentially flows through the entire section of the nano-ribbon, instead of being confined in a surface channel. Figure 2 shows the electron concentration in an n-type JL gated resistor for different values of gate voltage ranging from device pinch-off (Figure 2a) to flat-band conditions (Figure 2d). The conduction path is clearly located near the centre of the NW, not at the semiconductor-oxide interface. This allows the electrons to move through the core of the channel with bulk mobility, which is influenced much less by scattering than the surface mobility experienced by regular FETs. It is possible to create surface accumulation channels by increasing the gate voltage beyond the flat-band voltage, if a further increase of drain current is desired [10]. Because it operates under bulk conduction rather than channel conduction, the gated resistor sees its transconductance degrade much more slowly when gate voltage is increased. As a result, higher current and, therefore, higher-speed performance, can be expected from the gated resistor.

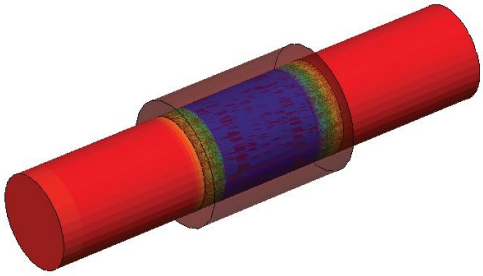


Figure 3: Structure of cylindrical GAA-JL-NW-FET [12].

2.3 GAA-JL-NW-FET

The JL-NW-FET, as illustrated in figure 3, has become widely recognized as one of the most promising candidates for future nanoscale CMOS technology due to its excellent subthreshold slope (SS), low drain-induced barrier lowering (DIBL) and low leakage current [12]. But it has a limitation which is related to the carrier transportation or mobility. In general, carrier transport can be divided into three regimes: diffusive transport, ballistic transport, and quantum transport [13]. When the gate length is much greater than the carrier mean free path, the carriers transit within the diffusive transport regime where scattering dominates. Carriers can reach their equilibrium states through sufficient scattering events and thus Drift-Diffusion (DD) model can capture well this equilibrium carrier transport property. When the gate length is shorter than de Broglie wavelength [14], the carriers transit within the quantum transport regime, where the full quantum transport approach such as Non Equilibrium Green Function (NEGF) is suitable to simulate strong quantum confinement and significant source to drain tunneling. In the range between these two regimes, the carriers transit from quasi-ballistic to ballistic transport regime. The semi-classical approach Monte Carlo simulator incorporated with full band structure is able to capture the non-equilibrium transport effects [15].

Figure 4 depicts the GAA-JL-NW-FET with its circular cross-sectional and top views [16].

3 Temperature Effect on Threshold Voltage

For an enhancement mode, n-channel MOSFET body effect upon threshold voltage is computed according to the Shichman-Hodges model. This model was accurate for the devices having channel lengths in the range of microns [17].

$$V_{TH} = V_{T,0} + \gamma\left(\sqrt{|V_{SB} + 2\varphi_F|} - \sqrt{|2\varphi_F|}\right)$$

where V_{TH} is the threshold voltage when substrate bias is present, V_{SB} is the source-to-body substrate bias, $2\varphi_F$ is the surface potential, and $V_{T,0}$ is threshold voltage for zero substrate bias, $\gamma = (t_{ox}/\varepsilon_{ox})\sqrt{2q\varepsilon_S N_A}$ is the body effect parameter, t_{ox} is oxide thickness, ε_{ox} is the relative permittivity of oxide, ε_S is the relative permittivity of semiconductor, N_A is a doping concentration, q is the charge of an electron.

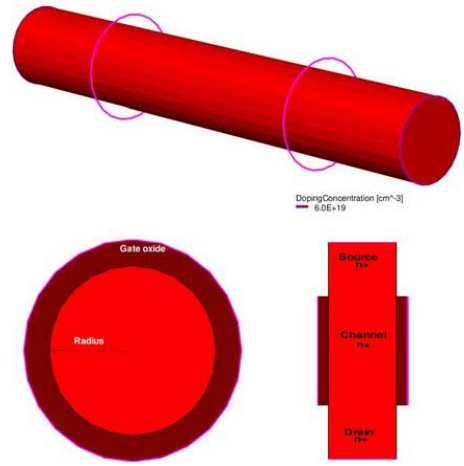


Figure 4: 3D structure of GAA-JL-NW with their cross-sectional views (both side and top views) [16].

As with the case of oxide thickness affecting threshold voltage, temperature has an effect on the threshold voltage of a CMOS device. Expanding on part of the equation in the body effect section,

$$\varphi_F = \frac{kT}{q}\ln\left(\frac{N_A}{n_i}\right)$$

where φ_F is half the contact potential, k is Boltzmann's constant, T is Temperature and n_i is the intrinsic doping parameter for the substrate. And it is known that [18],

$$n_i = 5.2\times10^{15}\times T^{3/2}\times\exp\left(-\frac{E_g}{2kT}\right)$$

Here E_g is the bandgap energy of the channel material.

It can be seen that the surface potential has a direct relationship with the temperature. Looking above, that while the threshold voltage does not have a direct relationship but is not independent of the effects.

The variation of the threshold voltage of a gated resistor with temperature is similar to that of a bulk MOSFET, with values of approximately –1.5 mV°C^{-1} measured in these devices. Interestingly, the decrease of mobility with temperature is much smaller in the gated resistors than in tri-gate FETs. In a lightly doped FET, the mobility is little affected by impurity scattering and tends to be phonon limited, so it shows a strong temperature dependence. In the highly doped gated resistor, on the other hand, mobility is limited by impurity scattering rather by phonon scattering, and its variation with temperature is much smaller. For instance, the electron mobility measured at room temperature in tri-gate FETs and gated resistors is 300 and 100 cm^{2}V^{-1}s^{-1}, respectively. When heated to 200 °C, the tri-gate FETs show a 36% loss of mobility, whereas the gated resistor has a reduction in mobility of only 6% [10].

For GAA-JL-NW-FET, the equation stands for the threshold voltage depends upon the work function, radius of device, thickness of oxide and the permittivity of oxide and material that is used [19].

$$V_{TH} = \Delta\varphi + \frac{kT}{q}\ln\left(\frac{8\mathrm{kT}\varepsilon_S}{q^2 n_i}\right) - \frac{2kT}{q}\ln\left[R\left(\frac{1+t_{ox}}{R}\right)^{\left(\frac{2\varepsilon_S}{\varepsilon_{ox}}\right)}\right]$$

Where, $\Delta\varphi$ is the work function difference and R is the radius of the device.

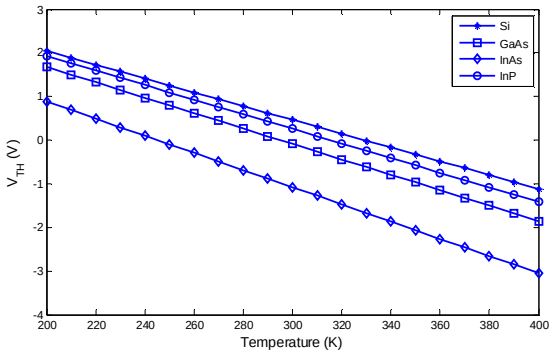


Figure 5: V_{TH} vs. temperature of GAA-JL-NW-FET. Four different matrials, Si, GaAs, InAs and InP, have been considered for the analysis.

4 Simulation and Result

In this article, the change of threshold voltage of GAA-JL-NW-FET having channel of different materials as a function of temerature has been simulated using MATLAB and then analyzed. Before the simulation, careful approximations should be made regarding the device dimentions of the GAA-JL-NW-FET. The parameters have already been approximated in different research works. From those, suitable estimated have been selected for this article [16].

The radius of the NW and the oxide-thickness are considered to be 2.5 nm and 0.77 nm, respectivey. The relative permittivity of the gate-oxide, considering SiO_2, has been taken as 3.9. And the workfunction difference has been approximated as 4.63 V.

For the analysis, four different channel materials have been considered, Si, GaAs, InAs and InP. The bandgap energies of these materials are 1.12 eV, 1.42 eV, 0.36 eV and 1.27 eV, respectively, and their relative permittivities are 11.7, 13.1, 14.6 and 12.4, respectively [20].

Considering the above mentioned values along with the general constants the simulation has been run over the temperature range of 200 K to 400 K. Threshold voltage versus temperature curve has been shown in figure 5.

From this figure it can be seen that for all the above metioned materials the threshold voltage reduces if the temperature is increased. And over the 200 K range, the change in threshold voltage is significant, around 3 to 4 V.

It means, if the temperature of the environment is as low as 200 K, they have a threshold voltage of 1 to 2 V. Whereas, at around 400 K temperature their threshold voltage become around –1 to –3 V. This result indicates that, if the temperature rises, the devices need less gate-voltage to be turned ON. It can be an important finding regarding in which applications they can be utilized.

Furthermore, at room temperature, i.e., 300 K, the threshold voltage of the experimented Si-channel GAA-JL-NW-FET is 0.4639 V. Device having GaAs-channel has the threshold voltage of –0.0865 V at room temperature. For InAs-channel transistor at 300 K, the threshold voltage is –1.0804 V. And if the channel is of InP, the threshold voltage is 0.2580 V at room temperature.

These results are clear indications of the material dependancy of threshold voltage of GAA-JL-NW-FET. The lowest threshold voltage can be achieved from InAs and the highest from Si. It proves the fact that, using a channel material with higher bandgap energy ensures a higher threshold voltage and vice versa. Keeping these in mind, the channel materials should be chosen so that they can accomplish the necessity.

5 Conclusion

GAA-JL-NW-FET is one of the most promising transistors for the near future for its simplicity and other electronic properties like near-ideal sub-threshold slope, extremely low leakage currents and less degradation of mobility with gate voltage and temperature than classical transistors. However, the high doping concentration in the channel reduces carrier mobility, which hurts drive current and transconductance of JL MOSFETs.

The temperature effect on the threshold voltage of this device has been deeply investigated over a range of 200 K to 400 K. And from this analysis, a conclusion may be drawn that the threshold voltage of GAA-JL-NW-FET is temperature dependant and it decreases with the increament of temperature.

The study has been made for Si, GaAs, InAs and InP NW based devices. From this study, it is clear that the minimum and maximum threshold voltages are achieved using InAs and Si, respectively.

It has already been mentioned earlier that GAA-JL-NW-FET has a great potential to replace the existing transistors. It is very important for the researchers to study its different characteristics for proper utilization. In the near future, the properties of this device along with the use of different channel materials may be studied.

References

[1] D. A. Neamen, *Semiconductor Physics and Devices – Basic Principles*. Tata McGraw-Hill Publishing Co. Ltd., pp. 537-542, 2006.

[2] H. Lou, H. Zhang, Y. Zhu, X. Lin, S. Yang, J, He and M. Chan, "A Junctionless Nanowire Transistor With a Dual-Material Gate," *IEEE Trans. Electronic Devices*, vol. 59, no. 7, pp. 1829-1836, Jul. 2012, doi: 10.1109/TED.2012.2192499.

[3] C. Lee, A. Borne, I. Ferain, A. Afzalian, R. Yan, N. Dehdashti-Akhavan, P. Razavi and J. Colinge, "High-Temperature Performance of Silicon Junctionless MOSFETs," *IEEE Trans. Electron Devices*, vol. 57, no. 3, pp. 620-625, Mar. 2010.

[4] C. J. Su, T. I. Tsai, H. C. Lin, T. Y. Huang and T. S. Chao, "Low-Temperature Ply-Si Nanowire Junctionless Devices with Gate-All-Around TiN/Al_2O_3 Stack Structure using an Implant-Free Technique," *SpringerOpen Journal:Nanoscale Research Letters,* 2012 7:339, doi: 10.1186/1556-276X-7-339.

[5] G. S. Kumar and B. Srimanta, "Novel Characteristics of Junction less Dual Metal Cylindrical Surround Gate (JLDM CSG) MOSFETs," *Research Journal of Recent Scences,* vol. 2, no. 1, pp. 44-52, Jan. 2013.

[6] J. E. Lilienfield, "Method and apparatus for controlling electric current," *US Patent* 1, 745,175 (1925).

[7] J. E. Lilienfield, "Device for controlling electric current,"*US Patent* 1,900,018 (1928).

[8] C. J. Su, T. I. Tsai, Y. L. Liou, Z. M. Lin, H. C. Lin, T. S. Chao, "Gate-All-Around Junctionless Transistors With Heavily Doped Polysilicon Nanowire Channels," *IEEE Electron Device Letters*, vol. 32, no. 4, pp. 521-523 Apr. 2011.

[9] Y. Cui, Z. Zhong, D. Wang, W. U. Wang and C. M. Lieber, "High Performance Silicon Nanowire Field Effect Transistor," *Nano Letters*, vol. 3, no. 2, pp. 149-152, 2003.

[10] J. P. Colinge, C. W. Lee, A. Afzalian, N. Dehdashti-Akhavan, R. Yan, I. Ferain, P. Razavi, B. O'Neill, A. Blake, M. White, A. M. Kelleher, B. McCarthy and R. Murphy, "Nanowire Transistors without Junctions," *Nature Nanotechnology*, Feb. 2010, doi: 10.1038/NNANO.2010.15.

[11] S. R. Sahu, R. S. Agrawal and S. M. Balwani, "Review of Junctionless Transistor using CMOS Technology and MOSFETs," *International Journal of Computer Applications*, pp. 8-11, 2012.

[12] T.-H. Yu, E. Hsu, C.-W. Liu, J.-P. Colinge, Y.-M. Sheu, Jeff Wu, and C.H. Diaz, "Electrostatics and Ballistic Transport Studies in Junctionless Nanowire Transistors," *IEEE Trans. Electronic Devices*, 2013, doi: 978-1-4673-5736-4/13.

[13] D. Vasileska, S.M. Goodnick, and G. Klimeck, Computational Electronics Semiclassical and Quantum Device Modeling and Simulation, *CRC press*, 2010**.**

[14] H. H. Lau, I. H. Hii, Aaron C. E. Lee, M. TaghiAhmadi, R. Ismail, and V. K. Arora, "The High-Field Drift Velocity in Degenerately-Silicon Nanowires," *Proc, INEC*, 2008.

[15] R. Ravishankar, G. Kathawala, and U. Ravaioli, S. Hasan, and M. Lundstrom, "Comparison of Monte Carlo and NEGF Simulations of Double Gate MOSFETs," *J. Computational Electronics*, 2005.

[16] B. Lakshmi and R. Srinivasan, "Investigation of f_t and Non-Quasi-Static Delay in Conventional and Junctionless multigate transistors using TCAD Simulations,"*ARPN Journal of Engineering and Applied Sciences*, vol. 7, no. 7, pp. 847- 852, Jul. 2012.

[17] S. M. Sze and M. K. Lee, *Semiconductor Devices – Physics and Technology*. John Wiley & Sons, Inc., pp. 166-167, 2012.

[18] B. Razavi, *Fundamentals of Microelectronics.* John Wiley & Sons, Inc., pp. 23-24, 2006.

[19] D. Jiménez, B. Iñíguez, J. Suñé, L. F. Marsal, J. Pallarès, J. Roig, and D. Flores, "Continuous Analytic *I–V* Model for Surrounding-Gate MOSFETs,"*IEEE Electron Device Letters*, vol. 25, no. 8, pp. 571-573 Aug. 2004.

[20] The Semiconductors Information website. [Online]. Availabe: http://www.semiconductors.co.uk/propiiiv5653.htm

A. Arefin has completed his B.Sc. in Electrical Electronic & Communication Engineering from Military Institute of Science & Technology, Dhaka. Now he is pursuing his M.Sc. Engineering from the same department. Earlier he attended Ideal School & College, Motijheel, Dhaka and Notre Dame College, Dhaka. Currently he is serving as an Assistant Professor in the Department of Electrical & Electronic Engineering, Northern University Bangladesh. Before joining here, he was a Lecturer in the Department of Electrical Electronic & Telecommunication Engineering, Dhaka International University. During the time as a faculty member, he has been associated with a number of research works on nanoelectronic devices, VLSI technology and renewable energy. He has also an experience of working as a Technical Manager in two projects of Bangladesh Election Commission. Mr. Arefin is a member of a number of associations like The American Center and International Association of Computer Science & Information Technology.

Testing Correlation and Homoscedasticity in a Bivariate T-Population

Anwar H. Joarder[1] and M. Hafidz Omar[2]
[1]School of Business, University of Liberal Arts Bangladesh, Dhanmondi, Dhaka 1209, Bangladesh
[2]Department of Mathematics and Statistics, King Fahd University of Petroleum and Minerals, Dhahran 31261, Kingdom of Saudi Arabia
Emails: anwar.joarder@ulab.edu.bd and omarmh@kfupm.edu.sa

Abstract—Testing uncorrelation or independence of the two components of a bivariate normal distribution is known for independent and identically distributed bivariate normal observations. What happens if the some of three strong assumptions do not hold good? A recent trend among business experts and econometricians is to use bivariate *t*-distribution whose components are correlated and which has thicker sides. It has been pointed out that the tests for uncorrelation developed in the bivariate normal case remains the same for observations following identical bivariate *t*-distributions. The implication, in this case, is that the failure of rejection of hypothesis of uncorrelation of the two components on the basis of a test would not necessarily mean independence. Similarly, testing the equality of true variances or homoscedasticity by sample variance ratio is also well known for independent and identically distributed observations from two normal populations. What is less known is that even if the sample obervations follow independent and identical bivariate normal distributions, the test remains the same. In this paper, we prove that the test of equality of variances remains the same even for observations following identical bivariate *t*-distributions.

Keywords— correlation coefficient; variance ratio; bivariate t-distribution, homoscedasticity, test of hypotheses.

Manuscript received on 15 September 2015 and accepted for publication on 18 October 2015.

1 INTRODUCTION

LET each of the sample observation $\underset{\sim}{X}_j$, $(j=1,2,\cdots,N)$, follow a bivariate *t*-distribution with mean $\underset{\sim}{\theta}$ (column vector order 2 components) and scale matrix Σ (a 2×2 matrix) where $\underset{\sim}{\theta}'=(\theta_1,\theta_2)$ and $\Sigma=(\sigma_{ik})$, $i=1,2;\ k=1,2$. We denote $\sigma_{11}=\sigma_1^2$, $\sigma_{22}=\sigma_2^2$, $\sigma_{12}=\rho\sigma_1\sigma_2$ with $\sigma_1>0,\ \sigma_2>0$ and the quantity ρ $(-1<\rho<1)$ as the product moment correlation coefficient between X_1 and X_2. The sample mean vector is $\underset{\sim}{\bar{X}}$, where $\underset{\sim}{\bar{X}}'=(\bar{X}_1,\bar{X}_2)$, so that the sums of squares and cross product matrix is given by $A=(a_{ik})$, where $a_{ik}=\sum_{j=1}^{N}(X_{ij}-\bar{X}_i)(X_{kj}-\bar{X}_k),\quad i=1,2;k=1,2.$ Obviously, $a_{ii}=\sum_{j=1}^{N}(X_{ij}-\bar{X}_i)^2,\quad (i=1,2),$ and $a_{12}=\sum_{j=1}^{N}(X_{1j}-\bar{X}_1)(X_{2j}-\bar{X}_2).$ The sample correlation coefficient is then given by $r=a_{12}/(ms_1s_2)$, where $s_i^2=a_{ii}/m$ $(i=1,2)$ and $m=N-1$.

In 1915, Fisher derived the distribution of the bivariate matrix A to study the distribution of correlation coefficient based on independent observations from a bivariate normal population [1]. The distribution of A is also known for uncorrelated observations following identical bivariate *t*-distributions. See, for example, in [2] page 157 and [3].

A recent interest among the applied scientists is the use of fat tailed distributions for modeling business data such as stock returns. Since the bivariate *t*-distribution has fatter tails, it has been increasingly applied for modeling business data. Interested readers may go through [4],[5] and [6].

If sample observations follow bivariate normal distributions, the test statistic $(m-1)^{1/2}(R-\rho)(1-R^2)^{-1/2}$ is known to have a *t*-distribution with $m-1$ degrees of freedom under the null hypothesis $H_0:\rho=0$ against the alternative $H_1:\rho\neq 0.$ In this paper we will expound that the assumption of bivariate normality can be relaxed to bivariate *t*-distribution.

Consider the scaled variances $U=mS_1^2/\sigma_1^2$ and $V=mS_2^2/\sigma_2^2.$ Assuming that the observations are from a bivariate normal population, Bose in 1935 and Finny in 1938 derived the density function of the variance ratio $H=U/V$ (See equation 4.1) [7],[8]. The distribution specializes to usual *F*-distribution $F(m,m)$ if $\rho=0.$ The random variables U and V have a bivariate chi-square distribution [9] with correlation coefficient ρ^2 and found application in signal processing [10].

Can we relax the assumption of bivariate normality to fat tailed distributions, say, to bivariate *t*-distribution and study the behaviour of the variance ratio? In this paper, we prove that if the sample observations are uncorrelated *t*-distributions, the distribution of the variance ratio remains the same.

In Section 3, we derive the joint distribution of scaled variances (U,V) and correlation coefficient (R) whenever the sample observations are uncorrelated bivariate *t*-distributions. We pinpoint that the uncorrelation does not necessarily imply independence in this case. In Section 4, we prove that even if the observations follow identical bivariate *t*-distributions, the distribution of variance ratio $(H = U/V)$ is the same as that developed by Bose orFinney [7], [8].

2 Bivariate Normal and T-Distributions

Let $\underset{\sim}{X}$ be bivariate normal random vector with density function

$$f_{2.1}(\underset{\sim}{x}) = (2\pi)^{-1}|\Sigma|^{-1/2}\exp\left[\frac{-1}{2}\left((\underset{\sim}{x}-\underset{\sim}{\theta})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\theta})\right)\right], \quad (2.1)$$

which will be denoted by $N_2(\underset{\sim}{\theta},\Sigma)$. Note that $E(\underset{\sim}{X}) = \underset{\sim}{\theta}$ and the variance-covariance matrix $Cov(\underset{\sim}{X}) = E(\underset{\sim}{X}-\underset{\sim}{\theta})(\underset{\sim}{X}-\underset{\sim}{\theta})'$ is given by $Cov(\underset{\sim}{X}) = \Sigma$. The bivariate normal distribution is also called elliptical normal distribution as the density function in (2.1) is constant on the ellipsoid $(\underset{\sim}{x}-\underset{\sim}{\theta})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\theta}) = c^2$ for any constant c.

The j-th observation with i-th characteristic is denoted by x_{ij} so that the random sample is denoted by $\{X_{ij}; i=1,2; j=1,2,\cdots,N\}$. Then the sample $\underset{\sim}{X}_1, \underset{\sim}{X}_2, \cdots \underset{\sim}{X}_N, (N>2)$ has the joint probability density function

$$f_{2.2}(\underset{\sim}{x}_1, \underset{\sim}{x}_2, \cdots, \underset{\sim}{x}_N) = (2\pi)^{-N}|\Sigma|^{-N/2}\exp\left(\frac{-1}{2}\sum_{j=1}^{N}(\underset{\sim}{x}_j-\underset{\sim}{\theta})'\Sigma^{-1}(\underset{\sim}{x}_j-\underset{\sim}{\theta})\right). \quad (2.2)$$

Each observation $\underset{\sim}{X}_j (j=1,2,\cdots,N)$ in (2.2) follows (2.1). Since the observations in (2.2) are uncorrelated, by virtue of normality, they are also independent and we call it Independent and Identical Bivariate Normal (IIBN) model for sample. Note that sufficient statistics for $\underset{\sim}{\theta}$ and Σ exist in model (2.2).

In [1], Fisher derived the distribution of A in order to study the distribution of correlation coefficient from a normal sample. The distribution of A is given by

$$f_{2.3}(A) = 4\pi\Gamma(m-1)|\Sigma|^{-m/2}|A|^{(m-3)/2}\exp\left(-\frac{1}{2}tr\Sigma^{-1}A\right), A>0, m>2. \quad (2.3)$$

The joint density function of the elements of A can be written as

$$f_{24}(a_{11},a_{22},a_{12})\alpha(\sigma_1\sigma_2)^{-m}(a_{11}a_{22}-a_{12}^2)^{(m-3)/2}\times \exp\left(\frac{a_{11}}{2(1-p^2)\sigma_1^2}-\frac{a_{22}}{2(1-p^2)\sigma_2^2}+\frac{pa_{12}}{(1-p^2)\sigma_1\sigma_2}\right) \quad (2.4)$$

where,

$a_{11}>0,\ a_{22}>0,\ -\sqrt{a_{11}a_{22}}<a_{12}<\sqrt{a_{11}a_{22}},\ m>2,\ -1<\rho<1.$

Let $\underset{\sim}{X}$ be bivariate t-random vector with probability density function

$$f_{2.5}(\underset{\sim}{x}) \propto |\Sigma|^{-1/2}\left(1+(\underset{\sim}{x}-\underset{\sim}{\theta})'(\nu\Sigma)^{-1}(\underset{\sim}{x}-\underset{\sim}{\theta})\right)^{-(\nu/2)-1}, \quad (2.5)$$

where, the scalar ν is assumed to be a known positive constant, see page 48 in [11]. The variable will be denoted by $T_2(\underset{\sim}{\theta},\Sigma;\nu)$. Note that $E(\underset{\sim}{X}_j) = \underset{\sim}{\theta}$ and $Cov(\underset{\sim}{X}) = E(\underset{\sim}{X}-\underset{\sim}{\theta})(\underset{\sim}{X}-\underset{\sim}{\theta})' = (1-(2/\nu))^{-1}\Sigma,\ \nu>2.$

Bivariate t-distribution can be generated by Conditionality Principle, Conditional Independence or by Stochastic Decomposition [12]. If sample observations $\underset{\sim}{X}_j (j=1,2,\cdots,N)$ are independent $T_2(\underset{\sim}{\theta},\Sigma;\nu)$, the joint density for the sample is given by

$$f_{2.6}(\underset{\sim}{x}_1,\underset{\sim}{x}_2,\cdots,\underset{\sim}{x}_N) \propto |\Sigma|^{-N/2}\prod_{j=1}^{N}\left(1+(\underset{\sim}{x}_j-\underset{\sim}{\theta})'(\nu\Sigma)^{-1}(\underset{\sim}{x}_j-\underset{\sim}{\theta})\right)^{-(\nu/2)-1}, \quad (2.6)$$

which will be called an Independent and Identical Bivariate T (IIBT) model for the sample. Note that sufficient statistics for $\underset{\sim}{\theta}$ and Σ do not exist in model (2.6). Is there any other alternative model for sample that shares intrinsic features, namely, marginality, conditionality, symmetry, equiprobable contour of (2.2)?

Now consider a sample $\underset{\sim}{X}_1, \underset{\sim}{X}_2, \cdots \underset{\sim}{X}_N, (N>2)$ $X_1, X_2, \cdots X_N (N>2)$ having the joint probability density function

$$f_{2.7}(x_1,x_2,\ldots,x_N)\alpha\left|\Sigma\right|^{-N/2}\left(1+\sum_{j=1}^{N}(x_j-\mu)'(\nu\Sigma)^{-1}(x_j-\mu)\right)^{-(\nu/2)-1} \quad (2.7)$$

Each observation $X_j (j=1,2,\cdots,N)$ in (2.7) follows (2.5). Since the observations in (2.7) are uncorrelated but not necessarily independent, (2.7) is called Uncorrelated and Identical Bivariate T (UIBT) model for the sample. Note that the sample observations in (2.7) are independent if $\nu\to\infty$ in which case (2.7) converges to (2.2) which is the joint density function of N uncorrelated observations from bivariate normal distribution. Samples can be generated from (2.7) by conditionality principle, conditional independence or much easily by stochastic decomposition [12]. In [13], Kelejian and Prucha proved that the Uncorrelated and Identical Bivariate model (2.7) captures

fat tailed behaviour better than the independent model in (2.6).

The density function of A based on UIBT model (2.7) is given by

$$f_{2.8}(A) \propto |\Sigma|^{-m/2} \ |A|^{(m-3)/2} \left(1+tr(\nu\Sigma)^{-1}A\right)^{-(\nu/2)-m}, \quad A>0, m>2 \tag{2.8}$$

where, $m>2,$ see page 160 in [3]. The above can also be written in terms of the elements as

$$f_{2.9}(a_{11},a_{22},a_{12}) \propto (\sigma_1\sigma_2)^{-m}\left(a_{11}a_{22}-a_{12}^2\right)^{(m-3)/2} \times\left(1+\frac{1}{\nu(1-\rho^2)}\left(\frac{a_{11}}{\sigma_1^2}+\frac{a_{22}}{\sigma_2^2}-\frac{2\rho a_{12}}{\sigma_1\sigma_2}\right)\right)^{-(\nu/2)-m}, \tag{2.9}$$

where, $a_{11}>0,\ a_{22}>0,\ -\sqrt{a_{11}a_{22}}<a_{12}<\sqrt{a_{11}a_{22}},\ m>2,\ -1<\rho<1.$

3 TESTS ON CORRELATION COEFFICIENT

3.1 Testing the Significance of Correlation Coefficient under Bivariate Normality

If each of the sample observations follow a bivariate normal distribution, it is well known that under the null hypothesis $H_0:\rho=0,$ the test statistic $R^2 \sim Beta(1/2,(m-1)/2),$ and $\sqrt{m-1}\,R\left(1-R^2\right)^{-1/2}$ has a Student *t*-distribution with $(m-1)$ degrees of freedom.The likelihood ratio test of the null hypothesis $H_0:\rho=0$ against the alternative $H_1:\rho\neq 0$ is done by the above statistic.

3.2 Testing the Significance of Correlation Coefficient under Bivariate t-Distribution

If sample observations follow UIBT model (2.7), then we need the distribution of R for testing the null hypothesis $H_0:\rho=0$ against the alternative $H_1:\rho\neq 0.$ It has been proved by Fang and Anderson; and Ali and Joarder in [2] and [14] respectively that the distribution of sample correlation coefficient remains the same as that for bivariate normal distribution. The proofs were done for a general class of distributions. For wide spectrum of readers, we sketch the proof in Theorem 3.2. The joint density function of scaled variances and correlation coefficient is derived below.

Theorem 3.1. *Let S_1^2, S_2^2 and R be sample variances and correlation coefficient based on a sample following bivariate Uncorrelated T-model (2.7). Then the joint density function of $U=mS_1^2/\sigma_1^2,\ V=mS_2^2/\sigma_2^2,$ R is given by*

$$f_{U,V,R}(u,v,r)\alpha(uv)^{(m/2)-1} (1-r^2)^{(m-3)/2}\left(1+\frac{1}{v(1-p^2)}\left(u+v-2pr\sqrt{uv}\right)\right)^{-(\nu/2)-m} \tag{3.1}$$

Where, $m>2$ and $-1<\rho<1.$

Proof. The density function of the elements of A based on UIBT model (2.7) is given by (2.9), where $a_{11}>0, a_{22}>0, -\infty<a_{12}<\infty, -1<\rho<1,$ $m>2, \sigma_1>0, \sigma_2>0.$ Under the transformation $a_{11}=ms_1^2, a_{22}=ms_2^2, a_{12}=mrs_1s_2$ (i.e., $a_{11}a_{22}=m^2s_1^2s_2^2,\ a_{12}^2=m^2r^2s_1^2s_2^2$) with Jacobian $J(a_{11},a_{22},a_{12}\to r,s_1^2,s_2^2)=m^3s_1s_2,$ we have,

$$f_{S_1,S_2,R}{}^{m-2}(1-r^2)^{(m-3)/2} \left(1+\frac{1}{v(1-p^2)}\left(\frac{m^2s_1{}^2}{\sigma^2{}_1}+\frac{m^2s_2{}^2}{\sigma^2{}_2}\frac{2pms_1s_2}{\sigma_1\sigma_2}\right)\right)^{-(\nu/2)-m}$$

The transformation $ms_1^2=\sigma_1^2u,\quad ms_2^2=\sigma_2^2v$ with Jacobian $J(s_1^2,s_2^2\to u,v)=m^{-2}\left(\sigma_1\sigma_2\right)^2,$ leads to (3.1).

As $\nu\to\infty,$ the joint density of U,V and R converges to the joint density function of U,V and R if the sample is drawn from a bivariate normal population [15].

Theorem 3.2. *Let S_1^2, S_2^2 and R be sample variances and correlation coefficient based on uncorrelated bivariate t-model (2.7). Then the density function of R is given by*

$$f_R(r)=\frac{2^{m-2}\left(1-\rho^2\right)^{m/2}}{\pi\,\Gamma(m-1)}\left(1-r^2\right)^{(m-3)/2}\sum_{k=0}^{\infty}\frac{(2\rho r)^k}{k!}\Gamma^2\left(\frac{m+k}{2}\right), \quad -1<r<1, \tag{3.2}$$

where, $m>2$ and $-1<\rho<1.$

Proof. Since $u+v-2\rho r\sqrt{uv}\le v(1-\rho^2),$ by expanding the last term of (3.1), the probability density function of R can be written as

$$f_R(r)\alpha(1-r^2)^{(m-3)/2} \int_0^{\alpha}\int_0^{\alpha}(uv)^{m/2-1}\left[1+\frac{1}{v(1-p^2)}\left(u+v-2pr\sqrt{uv}\right)\right]^{-(\nu/2)-m} dudv$$

Then the transformation $u=y_1\left(1-\rho^2\right),\ v=y_2\left(1-\rho^2\right)$ with Jacobian $J(u,v\to y_1,y_2)=\left(1-\rho^2\right)^2$ yields

$$h_R(r) \propto \left(1-r^2\right)^{(m-3)/2} J(\rho r, m, \nu/2) \quad (3.3)$$

where,

$J(\rho,m,\nu) = \int_0^\infty\int_0^\infty (uv)^{m/2-1}\left(1+u+v-2\rho\sqrt{uv}\right)^{-\nu-m} dudv$. It can be evaluated that

$J(\rho,m,\nu) = \frac{\Gamma(\nu)}{\Gamma(m+\nu)}\sum_{k=0}^{\infty}\frac{(2\rho)^k}{k!}\Gamma^2\left(\frac{m+k}{2}\right)$, so that from (3.3), we have

$h(r) \propto \left(1-r^2\right)^{(m-3)/2}\sum_{k=0}^{\infty}\frac{(2\rho)^k}{k!}\Gamma^2\left(\frac{m+k}{2}\right)$, which, apart from normalizing constant, is known to be the density function of R.

The density function of R was derived originally by Fisher in 1915 for independent sample observations following identical bivariate normal distributions [1]. Theorem 3.2 indicates that the assumption of bivariate normality under which tests on correlation coefficient are developed can be relaxed to bivariate *t*-distribution.

Acceptance of the null hypothesis does not mean independence unless the sample is from bivariate normal distribution. In view of Theorem 3.2, the test is true for bivariate *t*-distribution in which case acceptance of $H_0: \rho = 0$ implies uncorrelation but not necessarily independence.

We warn that the distribution of R is not necessarily the same if we had independent model (2.6) for the sample. The approximate distribution of R for independent observations from bivariate *t*-population can be obtained from page 157 in [11]. For the distributions of R in nonelliptical populations, the reader is referred to [16] and the references therein.

4 TESTING EQUALITY OF VARIANCES

Consider testing the equality of variances under three different situations. Suppose that we want to test the hypothesis $H_0: \sigma_1^2 = \sigma_2^2$, against the alternative hypotheses $H_0: \sigma_1^2 \neq \sigma_2^2$.

4.1 Independently and Identically Distributed Observations from $N(\theta_1,\sigma_1^2)$ and $N(\theta_2,\sigma_2^2)$

Consider two independent samples $X_{1j}\ (j=1,2,\cdots,N)$ and $X_{2j}\ (j=1,2,\cdots,N)$ from $N(\theta_1,\sigma_1^2)$ and $N(\theta_2,\sigma_2^2)$ respectively. Then the likelihood ratio is given by

$\lambda = \frac{2^m H^{m/2}}{(1+H)^m}$, Where, $H = U/V$. Since, λ is a monotonic function of H, the test can be carried out by H. The critical region $\{\lambda: 0<\lambda<\lambda_0\}$ is equivalent to $\{H: \{H \le H_l\} \cup \{H \ge H_u\}\}$ where H_u and H_l can be determined so that under null hypothesis $P(H \ge H_u) = \alpha/2$ and $P(H \ge H_l) = 1-(\alpha/2)$.

4.2 Independently and Identically Distributed Observations from $N_2(\underset{\sim}{\theta},\Sigma)$

Consider observations $X_1, X_2, \cdots X_N\ (N>2)$ each having $N_2(\underset{\sim}{\theta},\Sigma)$. Then it follows from Bose (1935) or Finney (1938) that H has a density function

$$f_H(h) = \frac{(1-\rho^2)^{m/2}}{B\left(\frac{m}{2},\frac{m}{2}\right)}\frac{h^{(m-2)/2}}{(1+h)^m}\left(1-\frac{4\rho^2 h}{(1+h)^2}\right)^{-(m+1)/2},\ h>0,$$

Which, will be denoted by $F(m,m;\rho)$. In [8], Finney compared the variability of the measurements of standing height and stem length for different age group of school boys by his method with the help of Hirschfeld [17]. In [18], Wilks developed the likelihood ratio test for testing the equality of variances in presence of correlation if the parent population is bivariate normal. An excellent review is available in [19] by Modarres, who also performed Monte Carlo simulation to determine the behavior of the likelihood ratio test.

4.3 Uncorrelated and Identically Distributed Observations from Bivariate T-Distribution

In this section we will prove that the even if each of the sample observations $\underset{\sim}{X}_1, \underset{\sim}{X}_2, \cdots \underset{\sim}{X}_N\ (N>2)$ follow identical bivariate *t*-distribution and the sample has the model (2.7), the distribution of $H = U/V$ remains the same as $F(m,m;\rho)$. Note that the observations are uncorrelated and not necessarily independent though the correlation between the components X_{1j} and X_{2j} of $X_j\ (j=1,2,\cdots,N)$ is ρ [14].

Theorem 4.1 Let S_1^2, S_2^2 and R be sample variances and correlation coefficient based on uncorrelated bivariate *t*-model (2.7). Also let $U = mS_1^2/\sigma_1^2$ and $V = mS_2^2/\sigma_2^2$ be scaled sample variances. Then the density function of $H = U/V$ is given by

$$f_H(h) = \frac{(1-\rho^2)^{m/2}}{B\left(\frac{m}{2},\frac{m}{2}\right)}\frac{h^{(m-2)/2}}{(1+h)^m}\left(1-\frac{4\rho^2 h}{(1+h)^2}\right)^{-(m+1)/2},\ h>0, \quad (4.1)$$

Where, $m>2$ and $-1<\rho<1$, and $B(a,b)$ is the usual beta function.

Proof. It follows from (3.1) that the joint density function of U and V is given by

$$f_{U,V}(u,v)\alpha(uv)^{(m/2)-1}$$

$$\int_{-1}^{1}(1-r^2)^{(m-3)/2}\left(1+\frac{1}{v(1-p^2)}\left(u+v-2pr\sqrt{uv}\right)\right)^{-(v/2)-m}dr \quad (4.2)$$

It follows that the density function of $H=U/V$ is given by

$$f_H(h)\alpha h^{(m-2)/2}\int_{r=-1}^{1}\left(1-r^2\right)^{(m-3)/2}$$

$$\int_{v=0}^{\alpha}v^{m-1}\left(1+\frac{1}{v(1-p^2)}\left(vh+v-2pr\sqrt{vhv}\right)\right)^{-(v/2)-m}dvdr$$

Substituting $v(1+h)-2\rho rv\sqrt{h}=y(1-\rho^2),$ with the Jacobian $J(v\to y)=\dfrac{1-\rho^2}{(1+h)-2\rho r\sqrt{h}}$, we have

$$f_H(h)\alpha\frac{h^{(m-2)/2}}{(1+h)^m}\int_{r=-1}^{1}\left(1-\frac{2pr\sqrt{h}}{1+h}\right)^{-m}$$

$$\int_{v=0}^{\infty}v^{m-1}(1+(y/v))^{-(v/2)-m}dydr \quad (4.3)$$

Since the integral in y can be converted to a beta type integral and it gets absorbed into the normalizing constant,

$$f_H(h)\propto\frac{h^{(m-2)/2}}{(1+h)^m}\int_{r=-1}^{1}\left(1-\frac{2\rho r\sqrt{h}}{1+h}\right)^{-m}dr. \quad (4.4)$$

which is equivalent to what was obtained by Bose in 1935 [7] or Finney in 1938 [8], and is equivalent to (4.1).

Equation (4.1) is well known for bivariate normal distribution (Bose, 1935). This proves the robustness of variance ratio in the class of bivariate elliptical *t*-distributions. The distribution of test statistic $H=U/V$ given by (4.1) will be denoted by $F(m,m;\rho)$.

Example 4.1 A chemical engineering is investigating the effect of process operating temperature (x) on product yield (y). The study results in the following data:

x	y
100	45
110	51
120	54
130	61
140	66
150	70
160	74
170	78
180	85
190	89

See page 457 in [20]. We assume that X and Y have an elliptical *t*-distribution with correlation coefficient ρ given by (2.7).

a. We want to test $H_0:\rho=0$ against $H_1:\rho\neq 0.$ The statistic $T=\dfrac{R\sqrt{m-1}}{\sqrt{1-R^2}}$ has a *t*-distribution with a degrees of freedom of $m-2$ so that the Rejection Region is $\{t:-2.306<t<2.306\}$. Since the sample produces

$$r=\frac{a_{12}}{\sqrt{a_{11}a_{22}}}=\frac{3985}{\sqrt{(8250)(1932.10)}}=0.998128718$$

and

$t=\dfrac{r\sqrt{m-1}}{\sqrt{1-r^2}}\approx 48.9696,$ we reject the null hypothesis and accept the alternative hypothesis. In the classical method, the assumption of bivariate normality of X and Y and the independence of bivariate observations were required.

b. We want to test $H_0:\sigma_1^2=\sigma_2^2$ against $H_1:\sigma_1^2<\sigma_2^2$ where σ_1^2 and σ_2^2 are the true variations of temperature and that of product yield. Let s_1^2 and s_2^2 are the sample variations of temperature and that of product yield.

The statistic $F=\dfrac{s_1^2/m}{s_2^2/m},$ has a variance ratio distribution with the same degrees of freedom of freedom of 9 so that the Rejection Region is $\{F:0.3145<F<3.18\}$. Since the sample yields $F=\dfrac{1932.10/9}{8250/9}\approx 0.2342,$ we do not reject null hypotheses. In the classicalmethod, the assumption of normality of X and Y and their independence are required.

5 CONCLUSION

The testing of equality of variances in presence of correlation with a bivariate normal population has a long history. Under the null hypothesis, the test statistic $H=U/V$ has a $F(m,m;\rho)$ distribution and can be used for testing $H_0:\sigma_1^2=\sigma_2^2,\ \rho\neq 0$ against $H_0:\sigma_1^2\neq\sigma_2^2,\ \rho\neq 0$ if ρ is known.

In this paper, we have proved that the assumption of normality can be relaxed to bivariate *t*- distribution for testing equality of variances in presence of correlation. However, the acceptance of the null hypothesis in this case would mean uncorrelation; It would mean independence in the special case of bivariate normality. The robustness of the distribution of the variance ratio or of the test will stimulate statisticians, econometricians and business experts to embark on further investigation in the area, let alone the use of classical results with confidence.

ACKNOWLEDGMENT

The authors acknowledge the excellent research support provided by King Fahd University of Petroleum and

Minerals, Saudi Arabia through the project IN131015. The first author also acknowledges the excellent research facilities at University of Liberal Arts Bangladesh.

REFERENCES

[1] Fisher, R.A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population, Biometrika, 10, 507-521.

[2] Fang, K.T. and Anderson, T.W. (1990). Statistical Inference in Elliptically Contoured and Related Distributins. Allerton Press.

[3] Sutradhar, B.C. and Ali, M.M. (1989). A generalization of the Wishart distribution for the elliptical model and its moments for the multivatiate t model. Journal of Multivariate Analysis, 29(1), 155-162.

[4] Sutradhar, B.C. and Ali, M.M (1986). Estimation of the parameters of regression model with a multivariate t error variable. Communications in Statistics: Theory and Methods, 15, 429-450.

[5] Lange, K.L.; Little, R.J.A. and Taylor, J.M.G. (1989) Robust statistical modeling using the t distribution. Journal of American Statistical Association, 84, 881-896.

[6] Billah, M.B. and Saleh, A.K.M.E. (2000). Performance of the large sample tests in the actual forecasting of the pretest estimators for regression model with t-error. Journal of Applied Statistical Science, 9(3), 237-251.

[7] Bose, S.S. (1935). On the distribution of the ratio of variances of two samples drawn from a given normal bivariate correlated population, Sankhya, 2, 65-72.

[8] Finney, D.J. (1938). The distribution of the ratio of the estimates of the two variances in a sample from a bivariate normal population, Biometrika, 30, 190-192.

[9] Gunst, Richard F. And Webster, John T. (1973). Density functions of the bivariate chi-square distribution. Journal of Statistical Computation and Simulation, 2, 275-288.

[10] Gerkmann, T. And Martin, R. (2009). On the statistics of spectral amplitudes after variance reduction by temporal cepstrum smoothing and cepstral nulling. IEEE Transactions on Signal Processing, 57(11), 4165-4174.

[11] Muirhead, R.J. (1982). Aspects of Multivariate Statistical Theory. John Wiley and Sons, New York.

[12] Johnson, M.E. (1984). Multivariate Statistical Simulation. John Wiley and Sons, New York.

[13] Kelejian, H.H. and Prucha, I. (1985). Independent and uncorrelated disturbances in linear regression: An illustration of the difference. Economic Letters, 19, 35-38.

[14] Ali, M.M. and Joarder, A.H. (1991). Distribution of the correlation coefficient for the class of bivariate elliptical models, Canadian Journal of Statistics, 19, 447-452.

[15] Joarder, A.H. (2009). Moments of the product and ratio of two correlated chi-square random variables. Statistical Papers, 50(3), 581-592.

[16] Johnson, N.L.; Kotz, S. and Balakrishnan, N. (1995). Continuous Univariate Distributions (volume 2). John Wiley and Sons, New York.

[17] Hirschfeld, H.O. (1937). The distribution of the ratio of covariance estimates in two samples drawn from normal bivariate populations, Biometrika, 29, 65-79.

[18] Wilks, S.S. (1946). Sample Criteria for Testing Equality of Means, Equality of Variances, and Equality of Covariances in a Normal Multivariate Distribution. The Annals of Mathematical Statistics, 17, No. 3 , 257-281.

[19] Modarres, R. (1993) Testing the Equality of Dependent Variances. Biometrical Journal 35(7), 785-790.

[20] Hines, William W. and Montgomery, Douglas C. (1990). Probability and Statistics in Engineering and Management Science. John Wiley and Sons.

Anwar H. Joarder obtained PhD and MS degrees in Statistics from the University of Western Ontario in 1992 and 1988, respectively. Currently he is a professor at University of Liberal Arts Bangladesh (ULAB). Before joining ULAB, he worked at many institutions including the University of Western Ontario, Monash University and University of Sydney. He serves on the editorial board of about 10 international research journals. He has authored or co-authored about 80 methodical papers published in internationally reputed research journals. His research area includes robustness of statistical methods, correlated statistical analysis, multivariate parametric estimation, and survey sampling and pedagogical statistics. He has been a lifelong Fellow of Royal Statistical Society, England, since 2005.

M. Hafidz Omar obtained PhD degree in Educational Measurement and Statistics in 1995 from the University of Iowa, USA. He obtained MS degree in Statistics in 1992 and BS in Actuarial Science in 1990 from the same university. He worked in companies and taught statistics at several institutions in the USA including the University of Iowa, the University of Kansas, and Riverside Publishing Company. Currently he is an Associate Professor of Statistics and Actuarial Science at King Fahd University of Petroleum and Minerals. He has refereed many research papers from reputable international journals. He has authored or co-authored more than 50 papers published in internationally reputed research journals. His research area includes applied multivariate statistics, correlated statistical analysis, probability modeling, mathematical statistics, educational statistics, actuarial statistics and statistics education.

Group Performance in a Swarm of Simulated Mobile Robots

Sifat Momen[1] and Kazi Tanjila Tabassum[2]
[1]Department of Computer Science and Engineering
University of Liberal Arts Bangladesh
Email: sifat.momen@ulab.edu.bd
[2]Department of Computer Science and Engineering
East West University

Abstract— Division of labour (or task allocation) is considered to be one of the most fundamental fields of research within the context of swarm intelligence and swarm robotics. In this paper, we look at how swarms of simulated mobile robots (i.e. mobile agents) carry out the decision of doing a particular task in an artificial world. In this paper, we present an agent based model wherein groups of mobile agents make decisions based on some simple rules. We, furthermore present an in-depth analysis on the efficiency of the swarm – typically how factors such as the number of robots and the amount of tasks affect the average energy of the swarm (energy efficiency) and how well they are utilized (robot utilization). Our simulation results reveal a number of interesting findings including: (1) the robot utilization decreases with the increase in the number of robots and vice versa, (2) the robot utilization increases with the increase in the amount of task and vice versa, and (3) there exists a swarm size, *S*, which results in an optimal energy for the swarm.

Keywords— Agent based model, division of labor, energy efficiency, mobile agents, robot utilization.

Manuscript received on 25 October 2015and accepted for publication on 11 November 2015.

1 Introduction

Division of labor (or task allocation) is often referred to as one of the most salient features for the organization and success of a team [1, 2]. Many multi-agent systems, such as that of the swarm systems (both biological and artificial), comprise of agents that carry out tasks. Typically there are more than one task in the environment and therefore there needs to be a mechanism through which the tasks to be done are to be assigned to agents for execution. This ability to allocate tasks among the agents is referred to as task allocation or division of labor. Often, the demand for the completion of a particular task may change (i.e. increase / decrease) due to perturbations caused by internal or external factors. Consequently, agents may require to re-assign tasks on the fly to meet the changing environment.

This paper addresses the issues of division of labor within the context of a group of mobile agents that are assumed to be simple (i.e. they have simple behavioral rules) and adaptive (i.e. they can switch tasks as required). Agent based modeling (ABM) approach has been chosen to model the behavior of agents since this approach can handle more complex behavior of agents (as opposed to that of the mathematical modeling). A simple stimulus response algorithm has been developed to provide the ability for the agents to re-assign tasks on the fly without the need of any centralized controller. The paper carefully analyzes how the collective behavior of agents affect the efficiency of the system at the macro-level. We introduced and defined robot utilization and energy efficiency to understand how the efficiency at the macro-level is affected by the number of robots in the environment.

The rest of the paper is organized as follows: In section II, a background of task allocation and swarm systems is given followed by our proposed model in section III. Experimental details and their results are discussed in section IV. Finally the paper concludes in section V with a summary of our findings and potential implications in our future research endeavors.

2 Background

Division of labor is highly noticeable in biological swarm systems such as ants, bees, termites etc. where individual agents respond to changing environment to serve the colonial demand. Division of labor in ants, in particular, are well-studied behavior within the context of social insects.

2.1 Division of labor in ants

There are more than 12000 species of ants and they show different kinds of division of labor [2, 3, 13]. The most fundamental form of division of labor in ants is the reproductive division of labor where a few individuals (often limited to one – called the queen) is solely responsible for the reproductive tasks while the other females of the colony (called workers) work as non-

reproductive agents. Apart from the basic division of labor, there are three prominent kinds of division of labor (not necessarily mutually exclusive of each other) found within different species of worker ants [3]: (1) Worker polymorphism, (2) Age polyethism and (3) Individual variability.

The ants that exhibit worker polymorphism tend to specialize in particular tasks due to their morphological advantages. For example, in case of the Pheidole ants [3], there exists two kinds of workers: major workers and minor workers. Major workers have disproportionately large heads and strong mandibles. Therefore, they tend to specialize in tasks that are laborious such as guarding a nest from intruders and/or transporting heavy objects from one place to another. Minor workers, on the flip side, tend to be smaller in size and therefore they tend to specialize in softer tasks i.e. tasks that do not require much expense of energy (such as cleaning the nests or feeding the young brood members)

Some species of ants (such as the Pogonomyrmex barbatus and Catalyphis bicolor) [4] are found to display age polyethism in which the task preferred by workers tend to change with their age. The young workers tend to be closer to the queen specializing in tasks such as feeding brood members and cleaning nests while the older workers tend to specialize in foraging and carrying food items to the nest.

However, most species of ants tend to demonstrate individual variability wherein they select tasks depending on their learning experience based on different factors. Consequently such colonies of ants tend to adapt to environment quickly and respond to the change in demand of the colony. For instance, in case of the red harvester ants, the worker class tends to carry out a number of tasks including foraging, patrolling, feeding younger brood members and cleaning nests. However, if the system is perturbed by external agencies, the situation changes. For instance, rain often results in the damage to the nest and this increases the demand to clean up the nests. When this happens, some of the ants doing other tasks are found to switch to nest cleaning job in response to the increased colonial demand for cleaning the messes [5]. This behavioral plasticity is the key factor behind the effective organization and the success of a colony.

2.2 Division of labor in swarm robotics

Animal behavior and in particular the behavior of social insects have strongly inspired many engineering disciplines. One such discipline that emerged over the past fifteen years is the field of swarm robotics where groups of robots (physical robots or simulated robots) or agents interact with each other to perform some tasks. These agents are autonomous in the sense that they carry out tasks without the need of any intervention of a centralized controller. The agents have simple behavioral rules and interact only locally i.e. they do not have the knowledge of the global template of the environment. One of the early papers written in this area was that by Krieger and Billeter [6] where they used up to twelve mobile robots to make the decision of whether to forage or rest. Unfortunately, the robots were not fully autonomous in the sense that periodically radio messages were sent to the robots from a control station briefing out when to forage or not. This is in violation with the swarm intelligence paradigm. However, the paper remains an important one as this was the first time, ant behavior was simulated in real robots. Subsequent studies by Labella [7] and Wenguo Liu and colleagues [8] worked towards building a fully decentralized system for allocating tasks. Yongming and colleagues [9] used a fixed response threshold model to develop a system where simulated robots can automatically decide whether to forage or rest. Ducatelle and colleagues [10], inspired from the work of Momen and colleagues [11] in mixed species flocking, presented task allocation strategies in the realms of heterogeneous groups of mobile robots. In their model, they had two kinds of simulated robots (wheeled robots and flying robots) working together to perform a common task. Zahadat and colleagues [14, 16] developed distributed adaptive partitioning algorithm for swarms of underwater robots. Their work was inspired by the age polyethism behavior demonstrated by honeybees. Brutschy and colleagues [15] on the other hand developed a method of allocating sequentially interdependent tasks to swarms of robots. In 2013, Momen [17] proposed how swarms of heterogeneous robots allocate tasks in a stochastic environment.

All the studies in the literature looks into developing systems that allocate tasks in a dynamic fashion. However, the factors affecting the efficiency of a multi-agent system is not well addressed in the literature. In this paper we present a simulated group of robots that can divide their labor to carry out complex tasks. We also carefully inspect the issue of efficiency in the context of swarm systems.

3 PROPOSED MODEL

Agent based modeling approach has been taken to model a swarm system in which agents make decisions of what tasks to do at any given moment. The model is developed using Netlogo [12], a cross-platform multi-agent modeling environment.

The specification of the proposed model is given below:

3.1 Specification of the model

The model consists of an environment (which is a 2D grid world) populated with mobile robots and objects (called boxes) placed randomly in the world (see figure 1). The topology of the world is referred to as 'Box' as the world is bounded in all dimensions. The agents (robots) can carry out three major tasks. Within each major tasks, there are a number of sub-tasks. The major tasks carried out by the simulated robots in our model are outlined below:

1) Charging- In this case, the robots go to a designated charging area (in the top left hand corner of the environment) in order to charge themselves up.

2) Box searching- In this case, robots search for boxes in the environment.

3) Box transport- Once a robot finds a box, it grabs the box using its gripper, moves towards a designated dumping area (on the top right hand corner of the environment) and leaves the box there.

The space is treated in discrete patches (81x41). However, the movement of the robots is modelled in continuous space so that at each time step t, each robot's floating point coordinates is mapped to an integer type coordinate in the local patch.

3.2 Behavioral rules

Each robot possesses orientation and follows simple rules as described below (see figure 5 for further details).

1. Each robot maintains an energy value [initialized randomly between 500 and 1000].
2. A robot moves randomly within the environment.
3. After every step, the energy value of a robot decre ments by a fixed small value.
4. Every robot has to maintain a minimum energy in order to operate.
5. If the energy of the robot falls below the minimum energy, the robot activates the charging task. (see figure 2) and moves towards the charging area.
6. Once it goes to the charging area, the energy value of the robot increases and it gets charged up.
7. If a robot finds a box within its visible range, it approaches the box, grip it using its gripper and starts moving towards the dumping area.
8. Once it reaches the dumping area, it releases the box and switch to box-searching mode for picking up new boxes.
9. The robots know the locations of the dumping area and the charging area and therefore they can simply face towards the area they need to move and starts travelling in that direction.

3.3 Parameters used in this model

The following table summarizes the parameters used in this model.

TABLE I. Parameters in the Model.

Parameter Name	Value	Remarks
Number of robots	Variable	This parameter specifies how many robots are used in a particular simulation.
Number of boxes	Variable	This parameter specifies the number of boxes placed randomly in the environment.
Minimum energy	185	This is the minimum energy a robot must have in order to operate. A robot having energy below this value needs to go to the charging area and charge itself up.
Energy decrement	0.12	This is the amount by which the energy of a robot decreases after every move.
Vision range	8.5	This is the depth of patch a robot can see.
Vision angle	440	This is the angular vision of the robot. The robot can see 220 to its left and 220 to its right.
Grab range	3.7	If a box is within its grab range, the robot can grip the box and start transporting.

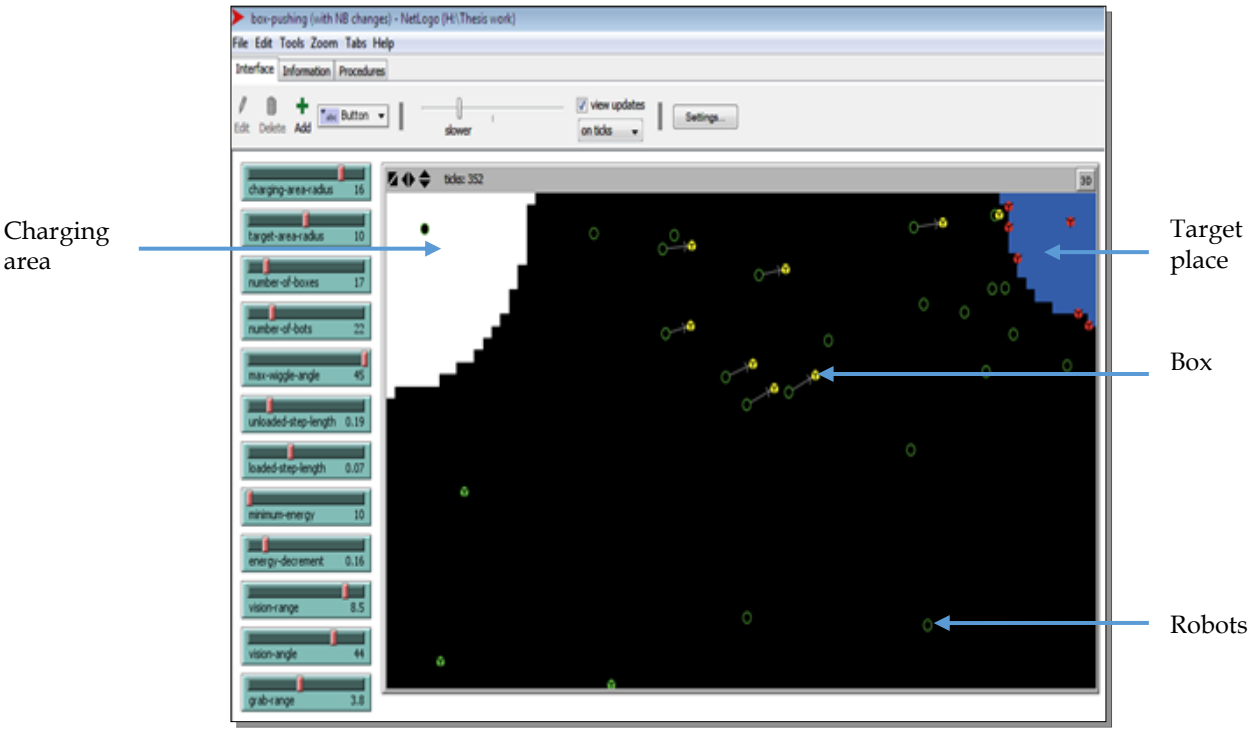


Figure 1: Snapshot of the model (robots are marked as green circles, the boxes are boxed shape and colored green if not yet being picked up, yellow if it is currently being transported by a robot and red if it is left in the target area.

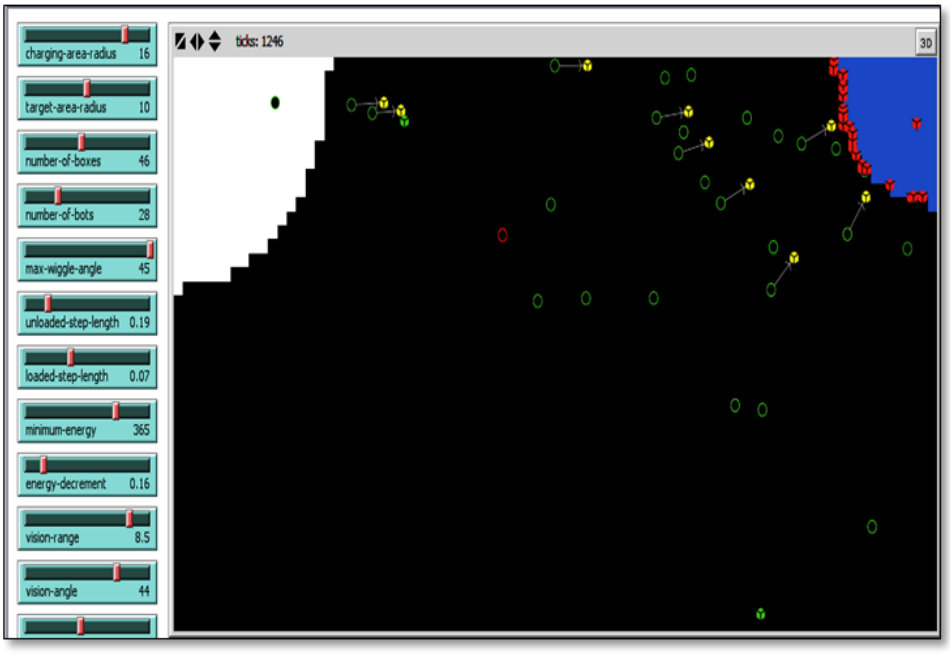


Figure 2: The robot colored red (shape circle) has low energy and has activated the charging task. Hence it is moving towards the charging area.

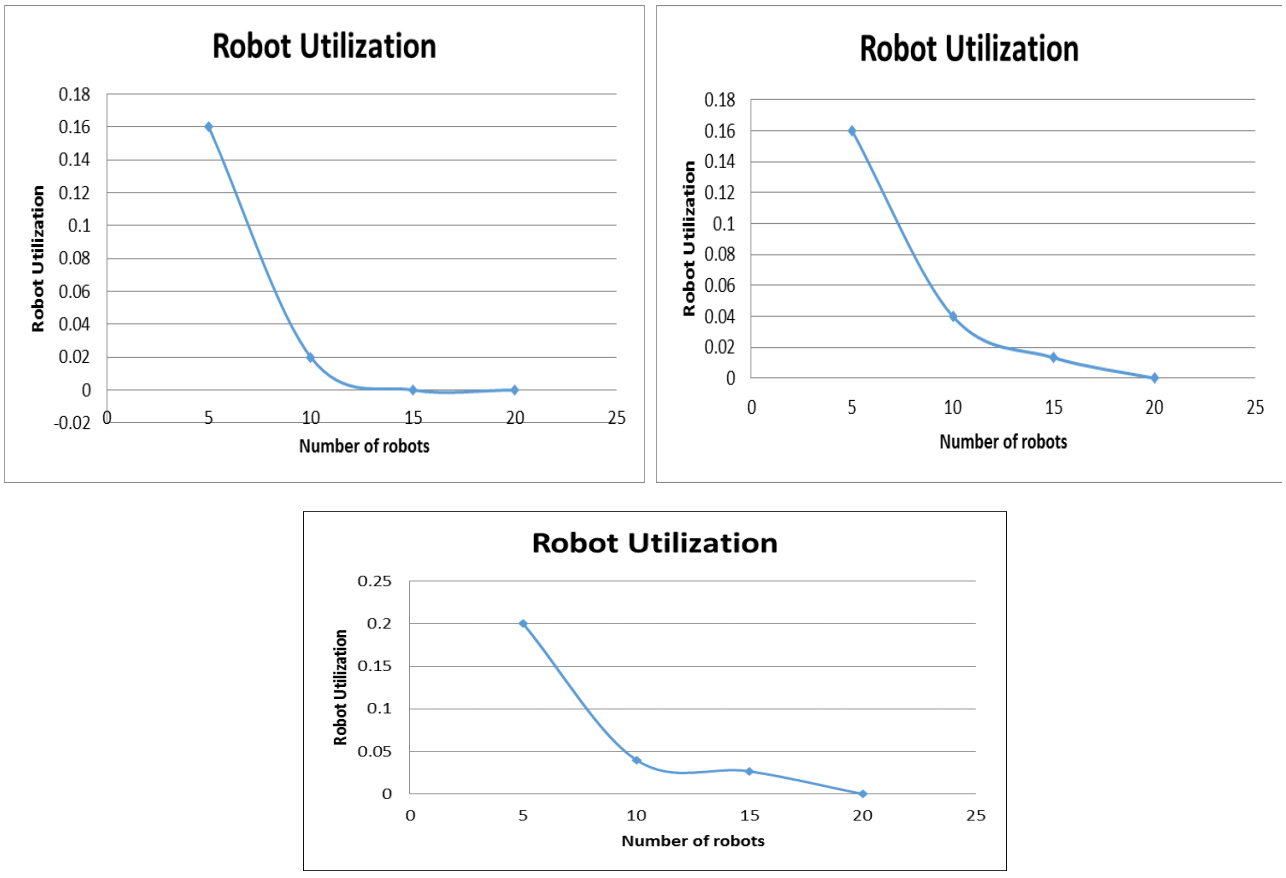


Figure 3: Robot Utilization as a function of the number of robots. In a) there are 5 boxes in the environment whereas in b) and c) there are 15 and 20 robots in the environment.

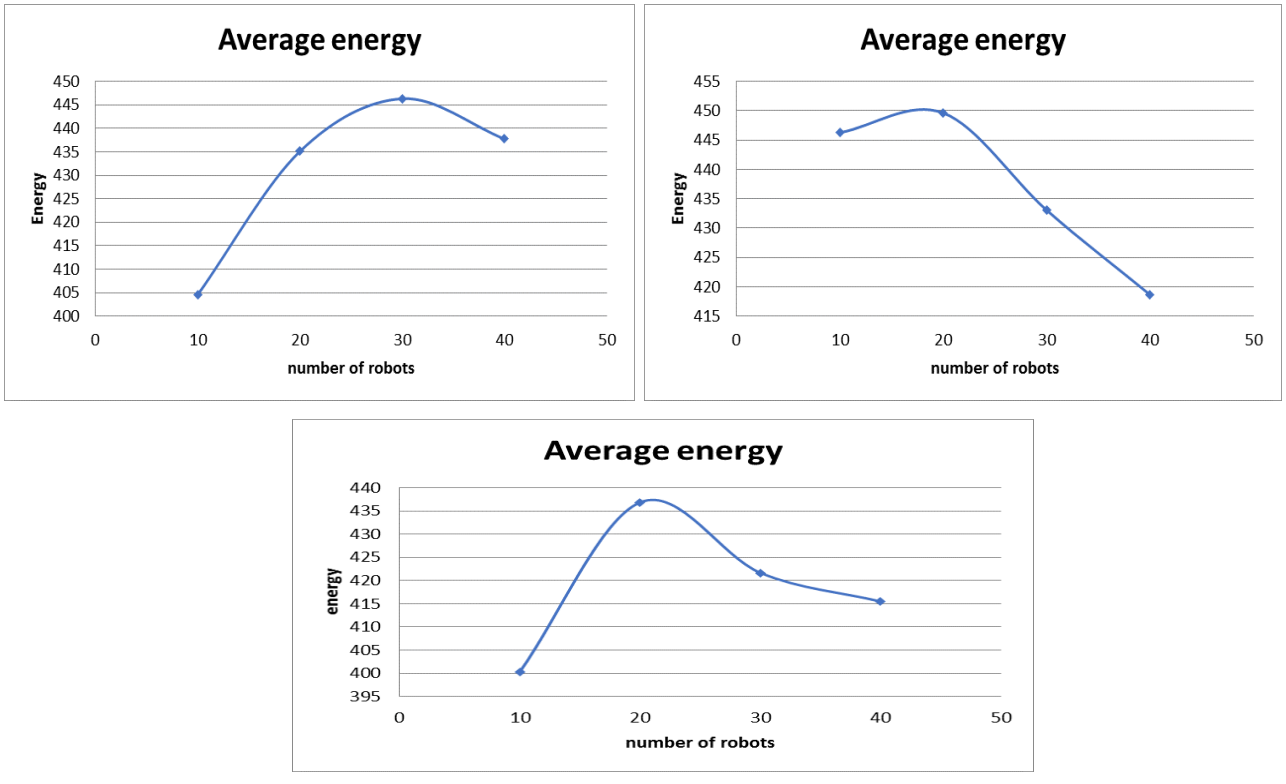


Figure 4: Average Energy of robots. There are 10 boxes in (a), 20 boxes in (b) and 30 boxes in (c).

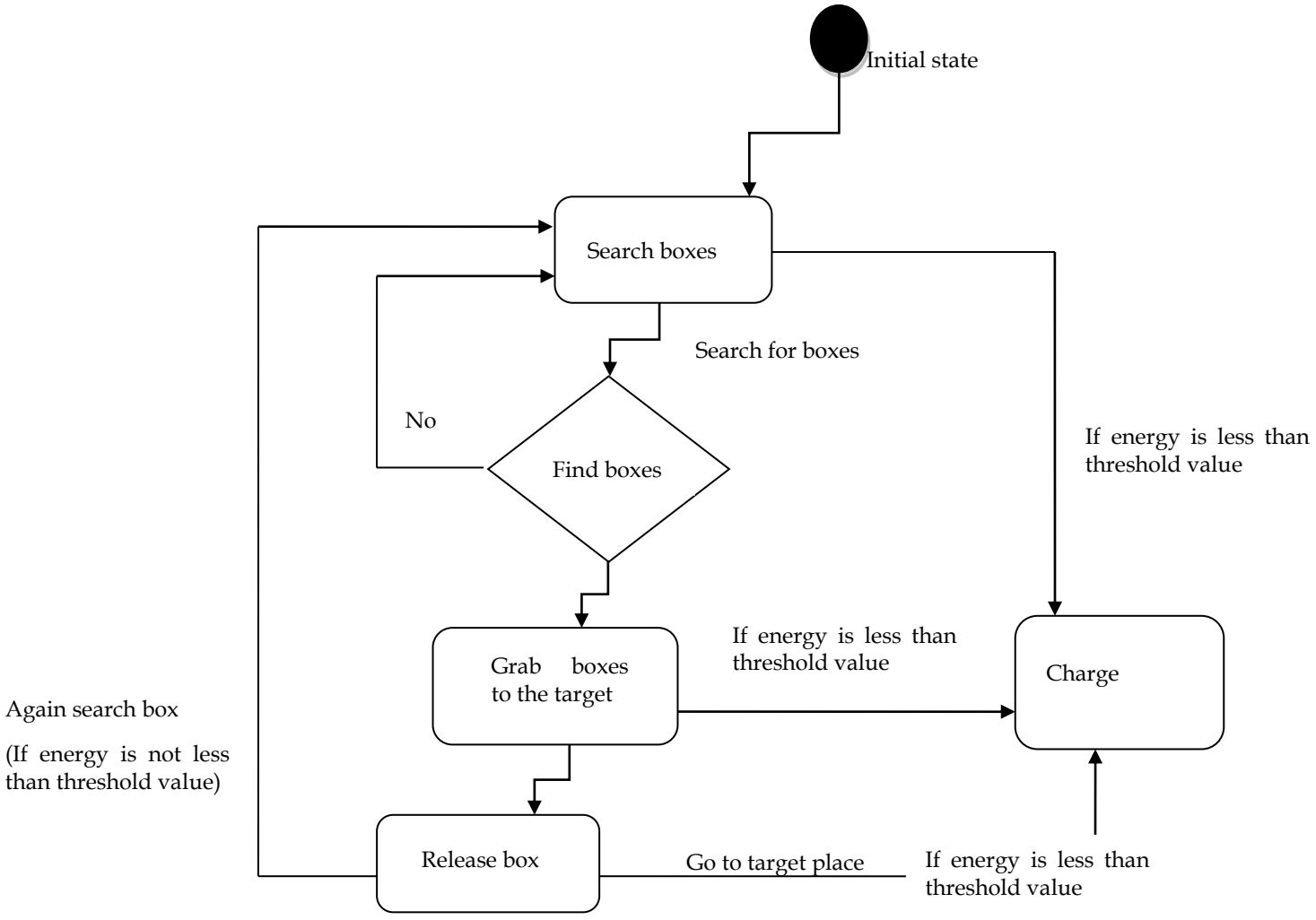


Figure 5: State Transition diagram

4 EXPERIMENTS AND RESULTS

In this section, we discuss how the number of robots participating affects the efficiency of the swarm system. This is particularly important since there is a notion of understanding in the literature that the increase in number of robots will improve the overall efficiency of the system. Swarm sytems are characterized by large number of redundant agents to improve the efficiency of the swarm. However, in reality, it can be different. If there are too less robots participating compared to the number of boxes in the environment (i.e. the amount of task in the world), then it will take long time to complete tasks and hence the efficiency would be low. On the other hand, if there are too many robots compared to the number of boxes, then too few robots will be effectively used and consequently it will result in poor resource utilization. In order to therefore understand efficiency, we define and inspect robot utilization and energy efficiency. Each experiment is repeated 20 times (runs) for 4000 time steps and the average reading is taken into account.

4.1 Robot Utilization

Robot utilization refers to the average number of robots been used in transporting boxes It is the ratio of the average number of robots carrying boxes compared to the total number of robots present in the environment – thus this is the measure of the fraction of robots that are engaged in effective tasks. Figure 3 shows how the robot utilization varies with the number of robots in the environment. The graphs show some interesting trends: (1) as the number of robots increases, the robot utilization decreases. So, it is not a good idea to always use a really high number of robots. Higher number of robots in a swarm system will allow tasks to be done in a shorter time but at the same time there are many robots that do not have any effective task to do, (2) as the number of boxes in the environment increases, robot utilization increases. This indicates that with greater number of tasks (if there are more boxes in the environment, there will be more transporting activity – so the number of tasks also increases), there is an increase in robot utilization. However, if the robots are over utilized, the efficiency of the system will decline. So, when there are more tasks in the environment, the increase in the number of robots to participate is justified.

4.2 ENERGY EFFICIENCY

In this section, we investigate how the average energy of a swarm at the end of experiments varies with the

number of robots participating in the experiments. Figure 4 shows the results obtained. Experiments were repeated with different number of boxes in the environment. All graphs indicate that there is an optimal number of robots for which the average energy of the swarm is maximum which agrees to our hypothesis. If the number of robots is less than the optimal value, the average energy of the swarm will be less. Similarly, if the number of robots participating is more than the optimal value, the average energy of the swarm again declines. This indicates that if there are too few robots in the world, they need to carry a lot of tasks and consequently the average energy of the swarm stays low. On the other hand, if there are too many robots, many of them will roam around in the environment trying to look for a box that they can pick up. Meanwhile, they continue to lose their energy and consequently again the average energy of the swarm stays low.

5 Conclusion and Future Work

This paper presents an agent based model in which swarms of simulated mobile robots carry out more than one task. The robots make dynamic decisions of when to carry out a particular task without the need of any centralized controller. The issue of efficiency of a swarm is investigated in details and some interesting trends have been found as illustrated in Section 4. In future, we plan to explore this model further and analyze how different parameters affect robot utilization and energy efficiency in a swarm. We also plan to adopt genetic algorithm to investigate the range of parameters that result in optimal energy efficiency and robot utilization.

References

[1] S. Momen, "Design and modelling of decentralised task allocation mechanisms in groups of mobile agents", PhD thesis, University of Sheffield, 2011.

[2] S.N. Beshers and J.H. Fewell, "Models of division of labour in social incects" Annu. Rev. Entomol. 2001, vol. 46, pp: 413 – 440, 2001.

[3] E. Bonabeau, M.Dorigo and G. Theraulaz, "Swarm Intelligence: From natural to artificial systems, Oxford University Press, 1999.

[4] K.K. Ingram, P. Oefner, D.M. Gordon, "Task specific expression of the foraging gene in harvester ants", Molecular Biology, vol. 14, pp: 813 – 818, 2005.

[5] D.M. Gordon, "Ant encounters: Interaction networks and colony behavior, Princeton University Press, 2010

[6] M.J.B. Krieger and J.B. Billeter, "The call of duty: Self organised task allocation in a population of up to twelve mobile robots", Robotics and Autonomous Systems, vol. 30, pp: 65 – 84, 2000

[7] T.H. Labella, "Division of labour in groups of robots". PhD thesis, Universite Libre De Bruxelles, 2007.

[8] W. Liu and A.F.T. Winfield, "Modelling and optimization of adaptive foraging in swarm robotic systems, The International Journal of Robotics Research, pp: 1 – 18, 2010.

[9] Y.Yongming, C. Xihui, L. Qingjun, T.Yantao, "Swarm robots task allocation based on local communication", International Conference on Computer, Mechatronics, Control and Electronic Engineering, pp: 415 - 418, 2010.

[10] F. Ducattelle, A.Foerster, G.A. Di Caro and L.M. Gambardella, "New task allocation methods for robotic swarms" , 9th IEEE/RAS Conferrence on Autonomous Robot Systems and Competitions, 2009.

[11] S. Momen, B.P. Amavasai, N.H. Siddique, "Mixed species flocking for heterogeneous robotic swarms, IEEE Eurocon 2007: The International Conference on Computer as a Tool, pp: 2329 – 2336, 2007.

[12] U.Wilensky, Netlogo, https://ccl.northwestern.edu/netlogo/, Center for Connected Learning and Computer-Based Modeling, Northwestern University, 1999.

[13] S.Momen and A.J.C. Sharkey, "Strategies of division of labor for improving task efficiency in multi-robot systems", IEEE World Congress on Nature and Biologically Inspired Computing, pp: 672 – 677, 2009

[14] P. Zahadat, S.Hahshold, R.Thenius, K.Crailsheim and T.Schmickl, "From hoonebees to robots and back: Division of labor based on partitioninng social inhibition", Bioinspiration & Biomimetics, vol. 10, no. 6, 2015

[15] A. Brutschy., G. Pini, C. Pinciroli, M. Birattari and M.Dorigo, "Self-organized task allocation to sequentially interdependent tasks in swarm robotics", Autonomous Agents and Multi-Agent Systems, vol. 28, no. 1, pp: 101-125, 2014.

[16] P. Zahadat, K. Crailsheim, and T. Schmickl. "Social inhibition manages division of labour in artificial swarm systems." In Advances in Artificial Life, ECAL, vol. 12, pp: 609-616, 2013.

[17] S.Momen, "Ant-Inspired decentralised task allocation strategy in groups of mobile agents", Complex Adaptive Systems, Procedia Computer Science, vol. 20, pp: 169 – 176, 2013

Sifat Momen is currently working as an Assistant Professor within the Department of Computer Science and Engineering at the University of Liberal Arts Bangladesh. He received his PhD in Computer Science from the University of Sheffield, UK. He was also involved as research associate in the School of Biological Sciences at the University of Bristol, UK. Prior to his PhD, he completed his MSc with distinction in Electronics and Information Technology from Sheffield Hallam University, UK. Dr. Momen received his undergraduate degree in Computer Science from North South University, Bangladesh. His current research interests include that of biomimetics, swarm intelligence, swarm robotics and artificial life.

Kazi Tanjila Tabassum is currently pursuing her MSc degree in Computer Science and Engineering from East West University, Bangladesh. She completed her BSc (Honors) in Computer Science and Engineering from the University of Liberal Arts Bangladesh. Her research interests are in the areas of artificial intelligence – particularly that in multi-agent systems, swarm intelligence and machine learning.

An Explicit Finite Difference Method

T M A K Azad[1] and L S Andallah[2]
[1]Department of Computer Science & Engineering,
University of Liberal Arts Bangladesh, Dhaka-1209,
[2]Department of Mathematics,
Jahangirnagar University, Savar, Dhaka-1342, Bangladesh
E-mail: abul.kalam@ulab.edu.bd

Abstract—This paper presents a one dimensional advection diffusion equation (ADE) as a simple mathematical model for the estimation of river water pollution. The numerical solution of ADE is obtained by using explicit centered difference scheme with FTBSCS techniques for prescribed initial and boundary data which may be used to predict the contaminant concentration levels in a river. Numerical results for the scheme are compared in terms of accuracy by error estimation with an exact solution of the ADE, and also the numerical features of the rate of convergence are presented graphically. Computational result verifies the qualitative behavior of the solution of ADE for various considerations of the parameters.

Keywords—Advection Diffusion Equation, Finite Difference Scheme, Exact Solution, Stability Condition, Water pollution, Rate of Convergence.

Manuscript received on 25 October 2015 and accepted for publication on 10 November 2015.

1. INTRODUCTION

River water pollution can be modeled using one dimensional advection–diffusion equation (ADE). This equation reflects physical phenomena where in the diffusion process particles are moving with certain velocity from higher concentration to lower concentration. It is a partial differential equation of parabolic type, derived on the principle of conservation of mass using Fick's 1st law. Due to the growing surface and subsurface hydro-environment degradation and the air pollution, the advection–diffusion equation has drawn significant attention of hydrologists, civil engineers and mathematical modelers. The analytical / numerical solutions along with an initial condition and two boundary conditions help to understand the contaminant or pollutant concentration distribution behavior through an open medium like air, rivers, lakes and porous medium like aquifer. It has wide applications in other disciplines too, like soil physics, petroleum engineering, chemical engineering and biosciences.

Many researchers have already been worked on it. Ogata and Banks [7] obtained an analytical solution of the one dimensional ADE by reducing the original ADE into a diffusion equation by applying moving coordinates. Banks and Ali[12] obtained an analytical solution of the one dimensional ADE by reducing the original ADE into a diffusion equation by introducing another dependent variable. Atul Kumar, Dilip Kumar Jaiswal and Naveen Kumar [6] presented an analytical solution of the one dimensional ADE by reducing the original ADE into a diffusion equation by using Laplace transformation technique. Agusta and Bamingbola [8] studied on the numerical treatment of the mathematical model for water pollution. This study was examined by various mathematical models invoving water pollution. The authors used the implicit centered difference scheme in space and a forward difference scheme in ti
mefor the evaluation of the generalized transport equation.ChangiunZhu, LipingWa and Sha Li [20] presented a numerical simulation of hybrid finite analytic methods for ground water pollution. Changjun and Shuwen [21] made a numerical simulation on river water pollution by using grey differential model. They corrected the model in finding the truncation error and found that the obtained results from the grey model are excellent and reasonable. Thongmoon and Mckibbin [14] compared some numerical methods for the advection-diffusion equation. They reported that the finite difference methods (FTCS, Crank Nicolson) give better point-wise solutions than the spline methods. M. M. Rahman, L.S. Andallah [22] presented a simulation of water pollution by finite difference method. They estimated and analyzed the extent of water pollution at different time and points.

In the present paper our intention is to investigate mathematical models and subsequent numerical methods to predict the contaminant concentration levels in a river at different time and different points of water bodies.

2.GOVERNING equation

The one-dimensional advection-diffusion equation (1) is considered as water pollution model is given as

$$\frac{\partial c}{\partial t} + u\frac{\partial c}{\partial x} = D\frac{\partial^2 c}{\partial x^2} \dots\dots\dots (1)$$

where c represents the solute concentration [ML^{-3}] at the point x, along longitudinal direction at time t, D is the solute dispersion, if it is independent of position and time, is called dispersion coefficient [L^2T^{-1}], t = time [T]; x = distance [L] and, u is the mean flow velocity [LT^{-1}] assumed to be constant.

Appended with initial condition

$$c(x,0) = f(x) 0 \le x < l$$

and boundary conditions

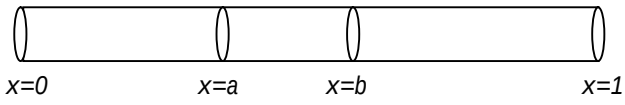
$$c(x = 0, t) = g_0(x) \qquad 0 < t \le T$$

$$c(x = l, t) = g_l(x) \qquad 0 < t \le T$$

the ADE formulates an initial boundary value problem (IBVP).

2.1 Advection diffusion equation and its derivation

For our model we assumed that the advection-diffusion equation may be a good first approximation to model the river pollution levels. It was also assumed that the river had a uniform cross-sectional area. Therefore, the river was assumed to be linear or one-dimensional (a pipe with a uniform cross-sectional area).



A one-dimensional river cross-section with arbitrary interior and end points at $x=0$ and $x=1$.

The advection-diffusion equation in one dimension is given by

$$c_t = Dc_{xx} - uc_x + f \dots\dots\dots (2)$$

where, the parameters are defined as follows- c is the concentration, x is the position on the river, t is the amount of time that passes, D is the diffusion factor, u is the velocity of the river, and f is the source or sink. The dimensions of the terms and coefficients in the equation are as follows-

$$c_t = \left[\frac{concentration}{T}\right],\ c_x = \left[\frac{concentration}{L}\right],\ c_{xx} = \left[\frac{concentration}{L^2}\right], D = \left[\frac{L^2}{T}\right], u = \left[\frac{L}{T}\right] \text{and } f = \left[\frac{concentration}{T}\right].$$

The derivation of equation (1) is as follows-

The time rate of change in concentration amount of contaminants due to a tributary in the interval from point a to point b in the river is represented by

$$A\int_a^b f_0(x,t)dx \dots\dots\dots (3)$$

where A is the cross-sectional area of the river and f0 is the source / sink.The total amount of contaminant in the river on the interval (a, b), is given by

$$A\int_a^b c(x,t)dx \dots\dots\dots (4)$$

The flux in concentration across a plane due to diffusion is the amount of concentration that passes through this plane due to the diffusion process. According to Fick's Law of Diffusion, the concentration flux due to diffusion across any cross-section at a point a is proportional to the product of the cross-sectional area and the concentration gradient c_x. The flux in concentration due to diffusion at the points a and b with c_x being the concentration gradient is given by respectively.

$$-AD\big(a, c(a,t)\big)c_x(a,t) \dots\dots\dots (5)$$

and

$$-AD\big(b, c(b,t)\big)c_x(b,t) \dots\dots\dots (6)$$

The concentration flux due to advection across any cross-section at point x is proportional to the product of the velocity, cross-sectional area, and concentration. By conservation of mass, we have

time rate of change **= diffusion flux at ($x=a$) - diffusion flux at ($x=b$) + advection flux at($x=a$) - advection flux at ($x=b$) + sources or sinks,**

or equivalently,

$$\frac{d}{dt}A\int_a^b cdx = -AD\big(ac(a,t)\big)c_x(a,t) + AD\big(b, c(b,t)\big)c_x(b,t) + uAc(a,t) - uAc(b,t) + \int_a^b f_0(x,t)dx \dots\dots\dots (7)$$

Now, divide through by A and let (f_0/A) = f). Using the fundamental theorem of calculus,

$$\int_a^b \frac{d}{dx}f(x)dx = f(b) - f(a) \dots\dots\dots (8)$$

the result is

$$\int_a^b [(c)_t - (Dc_x)_x + (uc)_x - f]dx = 0 \dots\dots\dots (9)$$

The choices of points a and b are arbitrary, so the integral equation is written as the partial differential equation

$$c_t - (Dc_x)_x + (uc)_x = f \dots\dots\dots (10)$$

Assume that there is no source or sinks holds ($f = 0$) and let D and u be constants. The final equation appears as

$$c_t + uc_x - Dc_{xx} = 0$$

Implies that

$$\frac{\partial c}{\partial t} + u\frac{\partial c}{\partial x} = D\frac{\partial^2 c}{\partial x^2} \dots\dots\dots (11)$$

2.2 Analytical Solution

By coordinate transformation, the exact solution of the advection-diffusion equation in unbounded is given by [1]

$$C(x,t) = \frac{M}{A\sqrt{4\pi Dt}} exp\left(-\frac{(x - (x_0 + ut))^2}{4Dt}\right) \dots\dots\dots (12)$$

where, M = mass of the pollutant

A = cross sectional area perpendicular to x
with the initial condition $c(x,0)=(M/A)\delta(x)$, where $\delta(x)$ is the Dirac delta function.

3. NUMERICAL METHOD FOR GOVERNING EQUATION

We consider the one–dimensional water pollution model problem as an initial and boundary value problem.

$$\frac{\partial C}{\partial t} + u\frac{\partial C}{\partial x} = D\frac{\partial^2 C}{\partial x^2} \dots\dots\dots (13)$$

With initial condition

$$c(t_0, x) = c_0(x); \qquad a \le x \le b$$

and boundary conditions

$$c(t, a) = c_a(x); \qquad t_0 \le t \le T$$

$$c(t, b) = c_b(x); \qquad t_0 \le t \le T$$

Finite difference techniques for solving the one dimensional advection diffusion equation can be considered according to the number of spatial grid points involved, the number of time levels used, whether they are explicit or implicit nature.

In Mathematics, the finite difference methods (FDM) are numerical methods for solving differential equations by approximating them with difference equations, in which finite differences approximate the derivatives. FDMs are thus discretization methods.

Today, FDMs are the dominant approach to numerical solutions of partial differential equations. Our goal is to approximate solutions to differential equations. i.e. to find a function (or some discrete approximation to this functions) which satisfies a given relationship between various of its derivatives on some given region of space / and or time, along with some boundary conditions along the edges of this domain. In general, this is a difficult problem and rarely an analytic formula can be found for the solution. A finite difference method proceeds by replacing the derivatives in the differential equation by the finite difference approximations. This gives a large algebraic system of equation to be solved in place of the differential equation, something that is easily solved on a computer.

3.1 EXPLICIT CENTERED DIFFERENCE SCHEME BY FTBSCS TECHNIQUES

Consider the model equation $\frac{\partial C}{\partial t} + u\frac{\partial C}{\partial x} = D\frac{\partial^2 C}{\partial x^2} \dots\dots\dots (14)$
In order to develop the scheme, we discretize the *x-t* plane by choosing a mesh width $h \equiv \Delta x$ space size and a time step size $k \equiv \Delta t$. The finite difference methods, we will develop, produce approximations $c_i^n \in R^n$ to the solution $c(x_i, t_n)$ in the discrete points by

$$x_i = ih, \quad i = 0, 1, 2, 3, \dots\dots\dots$$

$$t_n = nk, \quad n = 0, 1, 2, 3, \dots\dots\dots$$

Let the solution $c(x_i, t_n)$ be denoted by C_i^n and its approximate value by c_i^n.
By Explicit upwind time difference formula

$$\frac{\partial C}{\partial t} = \frac{C_i^{n+1} - C_i^n}{\Delta t} \dots\dots\dots (15)$$

Next use the backward space difference formula

$$\frac{\partial C}{\partial x} = \frac{C_i^n - C_{i-1}^n}{\Delta x} \dots\dots\dots (16)$$

And centered space difference formula

$$\frac{\partial^2 C}{\partial x^2} = \frac{C_{i+1}^n - 2C_i^n + C_{i-1}^n}{\Delta x^2} \dots\dots\dots (17)$$

Substituting equations (15 - 17) into equation (14) and rearrange according the time level, lead to

$$\frac{C_i^{n+1} - C_i^n}{\Delta t} + u\frac{C_i^n - C_{i-1}^n}{\Delta x} = D\frac{C_{i+1}^n - 2C_i^n + C_{i-1}^n}{\Delta x^2}$$

Which leads to

$$C_i^{n+1} = C_i^n - \frac{u\Delta t}{\Delta x}(C_i^n - C_{i-1}^n) + \frac{D\Delta t}{\Delta x^2}(C_{i+1}^n - 2C_i^n + C_{i-1}^n)$$

$$\therefore C_i^{n+1} = \left(\frac{u\Delta t}{\Delta x} + \frac{D\Delta t}{\Delta x^2}\right)C_{i-1}^n + \left(1 - \frac{u\Delta t}{\Delta x} - 2\frac{D\Delta t}{\Delta x^2}\right)C_i^n + \frac{D\Delta t}{\Delta x^2}C_{i+1}^n$$

Implies to

$$C_i^{n+1} = (\gamma + \lambda) * C_{i-1}^n + (1 - \gamma - 2 * \lambda) * C_i^n + \lambda * C_{i+1}^n \dots\dots\dots (18)$$

In which

$$\gamma = \frac{u\Delta t}{\Delta x}, \qquad \lambda = \frac{D\Delta t}{\Delta x^2}$$

3.2 STABILITY CONDITIONS FOR THE SCHEME BY FTBSCS TECHNIQUES

The explicit centered difference scheme for (14) is given by

$$C_i^{n+1} = (\gamma + \lambda) * C_{i-1}^n + (1 - \gamma - 2 * \lambda) * C_i^n + \lambda * C_{i+1}^n \dots\dots\dots (19)$$

In which

$$\gamma = \frac{u\Delta t}{\Delta x}, \qquad \lambda = \frac{D\Delta t}{\Delta x^2}$$

The equation (19) implies that if

$$0 \le \gamma + \lambda \le 1 \dots\dots\dots \text{(i)}$$

$$0 \le 1 - \gamma - 2\lambda \le 1 \dots\dots\dots \text{(ii)}$$

$$0 \le \lambda \le 1 \dots\dots\dots \text{(iii)}$$

then the new solution is a convex combination of the two previous solutions. That is, the solution at new time-step (*n+1*) at a spatial node *i* is an average of the solutions at the previous time-step at the spatial-nodes *i-1*, *i* and *i+1*. This means that the extreme value of the new solution is the average of the extreme values of the previous two solutions at the three consecutive nodes. Therefore, the new solution continuously depends on the initial value $c_i^0,\ i = 1, 2, 3, \dots\dots\dots M$.
(ii) implies $\gamma \le 1 - 2\lambda \le 1 + \gamma \dots\dots\dots$ (iv)

(i) implies $-\lambda \le \gamma \le 1 - \lambda$
$\therefore -\lambda \le \gamma \le 1 - 2\lambda$ by (iv)

Therefore,
the conditions are $0 \le \lambda \le 1$ and $-\lambda \le \gamma \le 1 - 2\lambda$
That is, $0 \le \frac{D\Delta t}{\Delta x^2} \le 1$ and $-\frac{D\Delta t}{\Delta x^2} \le \frac{u\Delta t}{\Delta x} \le 1 - 2\frac{D\Delta t}{\Delta x^2}$ are the st ability conditions of (19).

4. ALGORITHM FOR THE NUMERICAL SOLUTION

To find the numerical solution of the model, we have to accumulate some variables which are offered in the following algorithm.

Input: *nx* and *nt* are the number of spatial and temporal mesh points respectively.

t_f, the right end of (0, T)
x_d, the right end point of (0, b)
C_0, the initial concentration density, apply as initial condition
C_a, left hand boundary condition
C_b, right hand boundary condition
D, diffusion rate
u, velocity

Output: *C(x, t)*, the solution matrix

Initialization: $dt = \frac{T-0}{nt}$, the temporal grid size

$dx = \frac{b-0}{nx}$, the spatial grid size

$gm = \frac{u*dt}{dx}$, the courant number

$ld = \frac{D*dt}{(dx)^2}$ (peclet number)

Step 1. Calculation for concentration profile of explicit centered difference scheme
forn=1 to *nt*

for*i*=2 to *nx*

$$C(n+1,i) = (\gamma+\lambda)*C(n,i-1) + (1-\gamma-2*\lambda)*C(n,i) + \gamma*C(n,i+1)$$

end
end

Step 2: output *C*(*x*, *t*)
Step 3: Figure Presentation
Step 4: Stop

5. COMPUTATIONAL RESULTS

5.1. ERROR ESTIMATION AND CONVERGENCE

We have discussed two types of explicit finite difference schemes in the previous sections.
Now, we compute the relative error of the explicit difference scheme by using FTBSCS technique which is defined by the relative error in $L_1 -$ norm as

$$err = \frac{\| C_e - C_n \|_1}{\| C_n \|} \ldots\ldots\ldots (20)$$

where, C_e is the exact solution and C_n is the numerical solution computed by the finite difference scheme by FTBSCS techniques for time $t \in [0, 6]$.

The following figure 5.1 shows the convergence of relative error by the scheme FTBSCS techniques.

Numerical computation of ADE is presented by using explicit finite difference methods by FTBSCS techniques and compared with an exact solution of the ADE. A good agreement between the numerical solutions and the analytical solutions are obtained and the error becomes clear when using large size step for time. The choice of smaller discretization parameters (δtand δx) produce less errors. However we can say that FTBSCS techniques show the stable and accurate solutions for the advection diffusion equation.

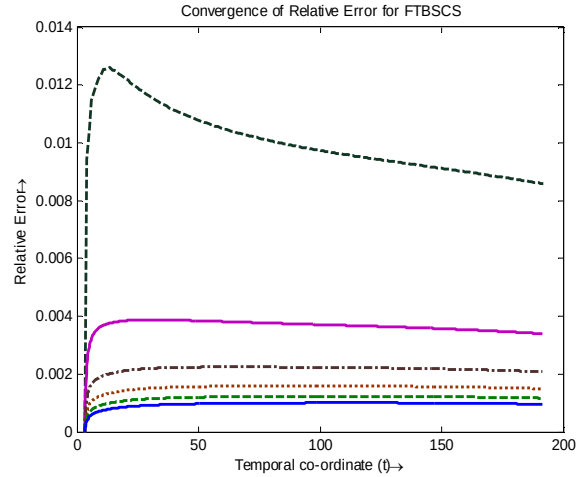


Figure 5.1 Rate of Numerical feature of Convergence.

5.2 NUMERICAL SIMULATION AND RESULTS DISCUSSIONS

This section presents the numerical simulation results for pollutant transportation in river water with increasing water flow velocity and increasing diffusion coefficient. To test the accuracy of the numerical scheme by FTBSCS technique for the ADE, we implement the model for some artificial data for the transport of the pollutant in the river water. Our aim is to show that for the water pollution, any substance with bigger diffusion results a wider pollutant front or a bigger diffusion distance. For different coefficients ranging from $9 m^2/s$ to $25\ m^2/s$, as shown in figures 5.2 to 5.4.

Problem description: Estimation of pollutant in a river of length=600 *m*=0.6 *km* at all time *t* = 1*min* to *t* = 6 *min*.

If *u* =1 m/s = 3.6 km/h and *D* = 25 m^2/s at time from 1 *min* to 6 *min*, for the numerical scheme FTBSCS technique is shown in Figure 5.2. Which shows that the pollutant distribution within the described domain.

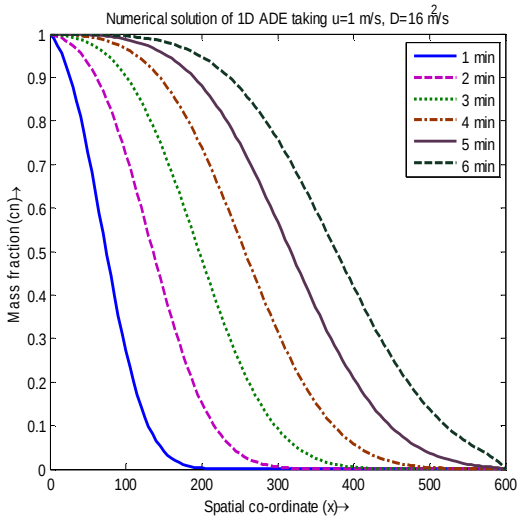


Figure 5.2:1D ADE with u = 1 m/s = 3.6 km/h and D = 9 m^2/s.

The curve marked by "solid line" shows the concentration profile for 1 minute (left), the curve visible by "dash line" represents the concentration profile for 2 minutes(left). The curve "dot line" shows the concentration profile for 3 minutes, the curve visible by "dash-dot line" represents the concentration profile for 4 minutes, The "solid line" curve shows the concentration profile for 5 minutes (right) and the curve visible by "dash line" represents the concentration profile for 6 minutes(right). We have seen that the pollutant concentration is increasing with respect to time increasing.

Now, we observe the following different figures for different diffusion coefficients and for different velocities. If $u = 1\ m/s = 3.6\ km/h$ and varying the diffusion rate a time $t = 6\ min$, the solution appeared is given below:

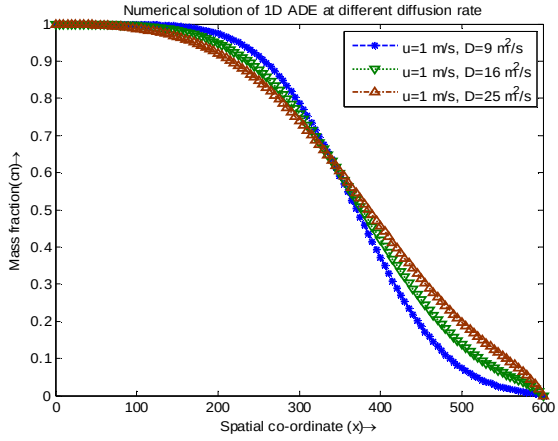


Figure 5.3:1D ADE with u=1 m/s=3.6 km/h and varying the diffusion rate at time 6 min.

If $D = 9\ m^2/s$ and varying the velocity at time 6 *min*, the solution appeared is given below:

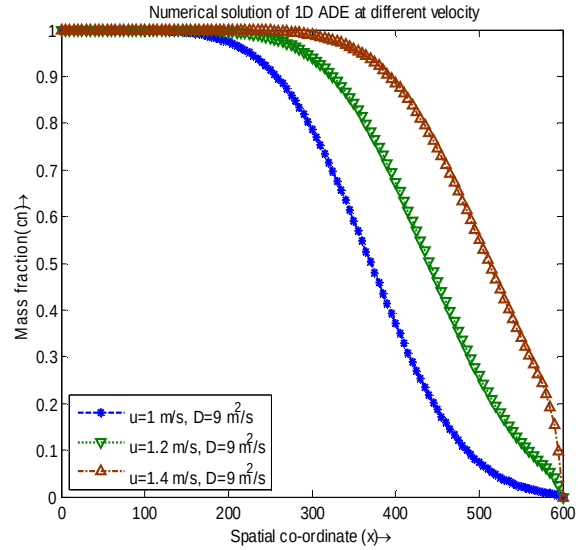


Figure 5.3:1D ADE with diffusion D = 9 m^2/s = 3.6 km/h and varying the velocity at time 6 min.

If varying both velocity and diffusion coefficient at time 6 *min*, the solution appeared is given below:

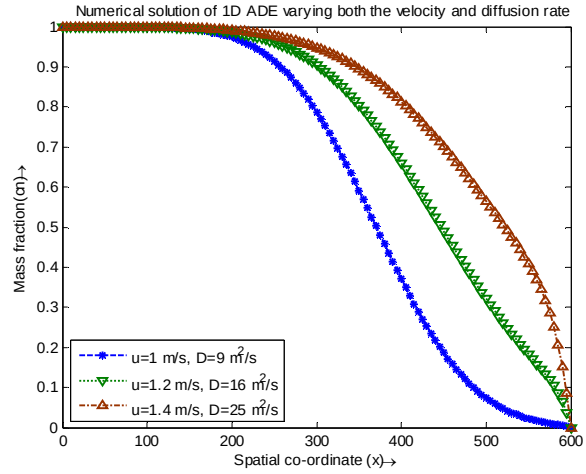


Figure 5.4:1D ADE with varying both the velocity and diffusion at time 6 min.

As we know the stability conditions of the scheme by FTBSCS technique are $0 \leq \frac{D\Delta t}{\Delta x^2} \leq 1$and $-\frac{D\Delta t}{\Delta x^2} \leq \frac{u\Delta t}{\Delta x} \leq 1 - 2\frac{D\Delta t}{\Delta x^2}$, this course of action will be continued until this stability conditions are satisfied. From this conditions, we have the diffusivity coefficient ranges from $D = 9\ m^2/s$ to $25\ m^2/s$ and the velocity ranges from $1\ m/s$ to $1.4\ m/s$.

6. Conclusion

In this paper, analytical solutions and numerical solutions for 1D advection diffusion equation, with an initial condition and two boundary conditions, have been presented. The ADE has been considered as model equation for estimation of water pollution by using explicit finite difference scheme (FTBSCS techniques). We have computed

relative errors for the scheme which shows a good rate of convergence of the numerical scheme. So, the scheme (FTBSCS techniques) for ADE is stable and consistent.

Also, we have presented the numerical solutions graphically by varying the value of velocity and value of diffusion coefficient. The graphical presentations are verifying the qualitative behavior of the solutions of ADE for various considerations of the parameters. The results show that the water pollutions are spreading with varying the diffusion term and advection term with respect to time and space.

This method can be extended for higher dimensional ADE as a water pollution model which demands the further study.

REFERENCES

[1] Scott A. Socolofsky& Gerhard H. Jirka, “Environmental Fluid Mechanics”, part 1, 2nd Edition, 2002.

[2] John D. Anderson JR, “Computational Fluid Dynamics”.

[3] Yijun Liu, “Introduction to the Finite Element Method”, lecture Notes; University of Cincinnati, U.S.A.

[4] Anderson, Mary P., and William W. Woessner. “Applied Groundwater Modeling, Simulation of Flow and Advective-Transport”. Academic Press Inc., San Diego, 1992. pp: 6-68.

[5] Bender, Edward A. “An Introduction to Mathematical Modeling”. John Wiley &Sons Inc., New York, 1978. pp: 1-140.

[6] Atul Kumar, Dilip Kumar Jaiswal and Naveen Kumar,“Analytical solution of one dimensional Advection diffusion equation with variable coefficients in a finite domain”, J.Earth Syst. Sci.118, No.5, pp: 539-549, October 2009.

[7] Ogata A, Banks RB.,”A Solution of the differential equation of longitudinal dispersion in porous media”, US Geological Survey, Paper 1961; 411-A; 1961.

[8] F.B. Agusto and O. M. Bamingbola, “Numerical Treatment of the Mathematical Models for Water Pollution”, Research Journal of Applied Sciences 2(5), pp: 548-556, 2007.

[9] Young-San Park, Jong-JinBaik,”Analytical solution of the advection diffusion equation for a ground level finite area source”, Atmospheric Environment 42, 9063-9069, 2008.

[10] Al-Niami A N S and Ruston K R 1977, “Analysis of flow against dispersion in porous media”, J.Hydrol. 3387-97.

[11] Aral M M and Liao B 1996, “Analytical solutions for two-dimensional transport equation with time-dependent dispersion coefficients; J. Hydrol. Engg. 1(1), pp: 20-32.

[12] Banks R B and Ali J 1964, “Dispersion and adsorption in porous media flow”, J. 90, pp: 13- 31.

[13] Nicholas J.Higham, “Accuracy and stability of Numerical Algorithms”, Society of Industrial and Applied Mathematics, Philadelphia, 1996.

[14] M. Thongmoon and R. Mckibbin, “A comparison of some numerical methods for the advection-diffusion equation”, Int. Math. Sci. 2006, Vol. 10, pp: 49-52.

[15] L.F. Leon, P.M. Austria, “Stability Criterion for Explicit Scheme on the solution of Advection-Diffusion Equation”, Maxican Institute of Water Technology.

[16] Randall J. Leveque, “Numerical methods for conservation laws”, second edition, 1992, Springer.

[17] Donea J, Guilinai S, Laval H. Quartpelle, “Time accurate solution of advection-diffusion problems by finite elements”. Comput Methods ApplMechEng 1984; 45; pp: 123-45.

[18] S G Ahmed, “A Numerical Algorithm for solving Advection-Diffusion Equation with Constant and Variable Coefficients”. Department of Engineering Physics and Mathematics, Faculty of Engineering,Zagazig University, Egypt.

[19] Romao, Silva and Moura, “Error analysis in the numerical solution of 3D convection diffusion equation by finite difference methods”. Thermal technology, vol-08, pp: 12-17, 2009.

[20] Changiun Zhu, LipingWa and Sha Li, “A numerical Simulation of Hybrid Finite Analytic Methods For Ground Water Pollution”. Advanced Materials Research Vol. 121-122(2010),pp: 48-51.

[21] Changjun Zhu and Shuwen Li, “A numerical Simulation of River Water Pollution using Grey Differential Model”. Journal of Computer, No.9, September 2010.

[22] M. M. Rahman, L.S. Andallah, “Simulation of Water Pollution by Finite Difference Method”, IJRT International Journal of Research in Information Technology, Volume 2, Issue 1, January 2014, pp: 17-24.

T. M. Abul Kalam Azad was born on February 15, 1966 in Sirajgonj, Bangladesh. He received his M.Phil degree on Ring Theory, M.Sc and B.Sc (Hons) degrees from the Department of Mathematics, Rajshahi University. He has been working in the Department of Computer Science and Engineering, University of Liberal Arts Bangladesh (ULAB), Dhaka from 2010 as a full time faculty. Previously, he worked at home, ‘Rajshahi University of Engineering & Technology (RUET)’, and abroad. Now, he is doing his PhD degree on Computational fluid dynamics in the Department of Mathematics, Jahangirnagar University. He has published a number of papers in peer-reviewed international and national journals. He is a life member of Bangladesh Mathematical Society.

Laek Sazzad Andallah was born on July 12, 1967 in Nilphamari, Bangladesh. He received his PhD degree on Applied Mathematics from TU Ilmenau Institute, Germany and M.Sc degree on industrial Mathematics from University of Kaiserslautern, Germany. He received his M.Sc and B.Sc (hons) degrees from the Department of Mathematics, Jahangirnagar University. He joined at the Department of Mathematics, Jahangirnagar University on 1992 as teaching associate. Now, he is working in the same department as a professor. He has published a number of papers in peer-reviewed international and national journals and presented original research articles in numerous international conferences. He received various scholarships, research and travel grants as recognition of his research works. He is a member of Bangladesh Mathematical Society.