Kanad Ray · S. N. Sharan
Sanyog Rawat · S. K. Jain
Sumit Srivastava
Anirban Bandyopadhyay

Editors

Engineering Vibration, Communication and Information Processing

ICoEVCI 2018, India

Springer

Lecture Notes in Electrical Engineering

Volume 478

Lecture Notes in Electrical Engineering (LNEE) is a book series which reports the latest research and developments in Electrical Engineering, namely:

- Communication, Networks, and Information Theory
- Computer Engineering
- Signal, Image, Speech and Information Processing
- Circuits and Systems
- Bioengineering
- Engineering

The audience for the books in LNEE consists of advanced level students, researchers, and industry professionals working at the forefront of their fields. Much like Springer's other Lecture Notes series, LNEE will be distributed through Springer's print and electronic publishing channels.

For general information about this series, comments or suggestions, please use the contact address under "service for this series".

To submit a proposal or request further information, please contact the appropriate Springer Publishing Editors:

More information about this series at http://www.springer.com/series/7818

Kanad Ray · S. N. Sharan
Sanyog Rawat · S. K. Jain
Sumit Srivastava · Anirban Bandyopadhyay
Editors

Engineering Vibration, Communication and Information Processing

ICoEVCI 2018, India

Springer

Editors
Kanad Ray
Amity School of Applied Sciences
Amity University Rajasthan
Jaipur, Rajasthan, India

S. N. Sharan
Department of Electrical Engineering
Manipal University Jaipur
Jaipur, Rajasthan, India

Sanyog Rawat
Department of Electronics and
    Communication Engineering
Manipal University Jaipur
Jaipur, Rajasthan, India

S. K. Jain
Department of Physics
Manipal University Jaipur
Jaipur, Rajasthan, India

Sumit Srivastava
Department of Information Technology
Manipal University Jaipur
Jaipur, Rajasthan, India

Anirban Bandyopadhyay
ANCC
National Institute for Materials Science
Tsukuba, Ibaraki, Japan

Preface

The first International Conference on "Engineering Vibration, Communication and Information Processing" (ICoEVCI 2018) was organized at Manipal University Jaipur, Rajasthan, India, during March 9–10, 2018. The papers included in this book were presented at ICoEVCI 2018.

The purpose of holding such conferences is to bring together, on a common platform, professors, scientists, engineers, medical practitioners, researchers, and students in the field of vibrational research, communication, and information processing, making the conference a perfect platform to share experience, foster collaborations across industry and academia, and evaluate emerging technologies across the globe. Vibrations, oscillations, and rhythms have enriched science, engineering, and medicine, being a very fundamental part of it. Vibration tends to have plenty of advantages and a lot more disadvantages, stirring up researchers worldwide to put a lot of effort to use its advantages and curtail its disadvantages.

This book discusses the revolution of cycles and rhythms that is expected to take place in different branches of science and engineering in the twenty-first century, with a focus on communication and information processing. It presents high-quality papers in vibration sciences, rhythms and oscillations, neurosciences, mathematical sciences, and communication. It includes major topics in engineering and structural mechanics, computer sciences, biophysics and biomathematics, as well as other related fields. Offering valuable insights, it also inspires researchers to work in these fields.

We are thankful to the authors of the research papers for their valuable contribution to the conference and for bringing forth significant research and literature across the field of vibration, communication, and information processing. The editors also express their sincere gratitude to ICoEVCI 2018 patron, plenary speakers, keynote speakers, reviewers, program committee members, international advisory committee, local organizing committee, sponsors, and student volunteers, without whose support the quality of the conference could not be maintained.

We would like to express our sincere gratitude to Dr. J. E. Lugo, University of Montreal; Dr. A. Alphones, NTU, Singapore; Dr. Subrata Ghosh, CSIR-NEIST, Jorhat; Dr. Phool Singh, NorthCap University; Haryana; Dr. Gaurav Saxena,

Jaipur, India	Kanad Ray
Jaipur, India	S. N. Sharan
Jaipur, India	Sanyog Rawat
Jaipur, India	S. K. Jain
Jaipur, India	Sumit Srivastava
Tsukuba, Japan	Anirban Bandyopadhyay

ICoEVCI 2018

Chief Patron

Prof. G. K. Prabhu, President, Manipal University Jaipur

Patrons

Prof. N. N Sharma, Pro-President, Manipal University Jaipur
Prof. G. C. Tikkiwal, Dean FOS, Manipal University Jaipur
Prof. A. K. Sinha, Director—SBS, Manipal University Jaipur
Mr. G. Karthihaivelen, Unit Head, Manipal Hospital, Jaipur

General Chairs

Prof. S. N Sharan, Director–SEEC, Manipal University Jaipur
Prof. Kanad Ray, Professor, Amity University, Rajasthan

Conveners

Dr. Sanyog Rawat, Manipal University Jaipur
Dr. Sushil Kr. Jain, Manipal University Jaipur
Dr. Deepak Yaduvanshi, Manipal Hospital, Jaipur

Co-Conveners

Dr. Rahul Sharma, Narayana Hospital, Jaipur
Dr. Pooja Pathak, Chitrakoot Hospital, Jaipur
Mrs. Parul Pathak, JECRC University, Jaipur

Technical Program Committee

Dr. Amit Rathi, Manipal University Jaipur
Prof. V. N. Tiwari, Manipal University Jaipur
Prof. Manish Tiwari, Manipal University Jaipur
Dr. Sunil Joshi, Manipal University Jaipur
Dr. Jyoti Chaudhary, Manipal University Jaipur
Col. Sanjiv Jaswal, Manipal University Jaipur
Dr. Raja Ram Aggarwal, Manipal Hospital, Jaipur
Dr. Ravi Kant Porwal, Manipal Hospital, Jaipur
Dr. Deepak Shukla, Manipal Hospital, Jaipur
Prof. Upendra Mishra, Amity University, Rajasthan
Dr. Vipin Pal, National Institute of Technology Meghalaya, Shillong
Dr. Tarun Kr. Dubey, Manipal University Jaipur
Dr. Pramod Kumar, Manipal University, Manipal
Dr. K. M. Bakwad, Latur, Maharashtra
Dr. Ritu Vijay, Banasthali Vidyapith
Mr. Shiv Narain, Central Electronics Limited, Sahibabad
Dr. Sunil Gupta, Rajasthan University of Health Sciences
Prof. Priyanka Chawla, Jalandhar, Punjab
Dr. Sunita Gupta, SKIT, Jaipur
Dr. Gangaram Mishra, Amity Lucknow
Dr. Sandeep Kumar, Inje University, Gimhae, South Korea
Dr. Surbhi Sharma, Thapar University
Dr. S. K. Gupta, MGNIMS Jaipur
Dr. Chitresh Nayak, Kautilya Institute of Technology and Engineering, Jaipur
Dr. Parvez Ahmad Alvi, Banasthali University, Rajasthan
Dr. Mukesh Arora, SKIT, Jaipur
Dr. Pranay Goswami, Ambedkar University, Delhi
Dr. Rajkumar Chaurasia, ICFAI University, Jaipur
Dr. Rana Mukherjee, ICFAI University, Jaipur
Dr. Manoj Singh Parihar, IIITDM, Jabalpur
Prof. Mukesh Gupta, SGVU, Jaipur
Dr. Pushpendra Singh, JKLU, Jaipur

Advisory Committee

Prof. Sandeep Sancheti, Vice Chancellor, SRM University, Chennai
Prof. Kumkum Garg, Dean Faculty of Informatics, Photonics and Robotics, BSDU, Jaipur
Prof. (Dr.) K. M. Prasad, Manipal University, Manipal
Prof. (Dr.) Bhola Ishwar, BRA Bihar University, Muzaffarpur
Prof. (Dr.) Susmita Sarkar, Department of Applied Maths, University of Calcutta
Dr. Ganesh Naik, University of Technology Sydney
Prof. S. Chakraverty, NIT Rourkela
Prof. Ben H. Jansen, University of Houston, USA
Prof. Gustav Bernroider, University of Salzburg, Austria
Prof. Giorgio Carmignoto, University of Padova, Italy
Prof. Ilija Damjanovic, Russian Academy of Sciences, Russia
Dr. Lipo Wang, Nanyang Technological University, Singapore
Prof. Mark Solms, University of Cape Town, South Africa
Prof. Emilia I. Barakova, Eindhoven University of Technology, Netherlands
Prof. Arun Mishra, McGill University, Canada
Prof. J. Kaplunov, Keele University, New Castle, UK
Dr. C. L. Prajapat, SSPD, BARC Mumbai
Dr. Pankaj Poddar, NCL, Pune
Dr. Gaurav Saxena, GWEC, Ajmer
Dr. S. N. Dolia, University of Rajasthan
Dr. Ghanshyam Singh, MNIT Jaipur
Dr. Ritu Sharma, MNIT Jaipur

Publication Committee

Prof. Sumit Srivastava, Manipal University Jaipur
Dr. K. P. Misra, Manipal University Jaipur
Dr. Saikat Chattopadhyay, Manipal University Jaipur
Mr. Anurag Bhatnagar, Manipal University Jaipur
Mr. Dinesh Yadav, Manipal University Jaipur

Sponsorship and Finance Committee

Prof. D. P. Sharma, Manipal University Jaipur
Mr. Sujiban Ghosh, Manipal University Jaipur
Dr. Jagrati Sahariya, Manipal University Jaipur
Ms. Kuntal Gaur, Manipal University Jaipur

Mr. Govind Vijay, Manipal University Jaipur
Dr. Sanyog Rawat, Manipal University Jaipur
Dr. Sushil Kr. Jain, Manipal University Jaipur

Registration Committee

Dr. Rohit Jain, Manipal University Jaipur
Dr. Jagrati Sahariya, Manipal University Jaipur

Media, Website, Videography, and Photography Committee

Mr. Lokesh Sharma, Manipal University Jaipur
Dr. K. P. Misra, Manipal University Jaipur
Dr. Abhijeet Singh, Manipal University Jaipur
Dr. Ramesh Kumar Rawat, Manipal University Jaipur
Mr. Vikas Verma, Manipal University Jaipur

Publicity Committee

Mr. Ankit Mundra, Manipal University Jaipur
Mr. Dinesh Yadav, Manipal University Jaipur
Ms. Abhilasha, JECRC University
Ms. Himani Bali, JECRC University
Mr. Tejpal, Manipal University Jaipur
Mr. C. P Gupta, Manipal University Jaipur

Accommodation Committee

Dr. Anupam Sharma, Manipal University Jaipur
Mr. Shashank Sharma, Manipal University Jaipur
Dr. Nilanjan Halder, Manipal University Jaipur
Mr. Nitin Gupta, Manipal University Jaipur

Technical Presentation Committee (oral/poster)

Dr. Sumit Srivastava, Manipal University Jaipur
Dr. Nilanjan Halder, Manipal University Jaipur
Dr. Anupam Sharma, Manipal University Jaipur
Dr. Saikat Chattopadhyay, Manipal University Jaipur
Mr. Mahesh Jangid, Manipal University Jaipur
Dr. Saurabh Gupta, Manipal University Jaipur
Mr. Tejpal, Manipal University Jaipur

Transportation and Hospitality Committee

Dr. Madan Mohan Sharma, Manipal University Jaipur
Dr. AbhijeetSingh, Manipal University Jaipur
Dr. Saikat Chattopadhyay, Manipal University Jaipur
Mr. Nitin Gupta, Manipal University Jaipur

Catering Committee

Mr. Dinesh Yadav, Manipal University Jaipur
Mr. Chandra Prakash Gupta, Manipal University Jaipur
Mr. Mohit Kumar Sharma, Manipal University Jaipur

Stage and Certificate/Prize Distribution Committee

Ms. Pallavi Yarde, Manipal University Jaipur
Dr. Rohit Jain, Manipal University Jaipur
Ms. Deepika Bansal, Manipal University Jaipur
Ms. Neha Singh, Manipal University Jaipur

Purchase Committee

Dr. Sushil Kumar Jain, Manipal University Jaipur
Dr. Jagrati Sahariya, Manipal University Jaipur
Dr. Rohit Jain, Manipal University Jaipur
Dr. Sanyog Rawat, Manipal University Jaipur

Contents

About the Editors

Dr. Kanad Ray is Professor and Head of Physics at the Amity School of Applied Sciences, Amity University Rajasthan (AUR), Jaipur, India. He has obtained M.Sc. and Ph.D. in physics from Calcutta University and Jadavpur University, West Bengal, India. In an academic career spanning over 22 years, he has published and presented research papers in several national and international journals and conferences in India and abroad. He has authored a book on the electromagnetic field theory. His current research areas of interest include cognition, communication, electromagnetic field theory, antenna and wave propagation, microwave, computational biology, and applied physics. He has served as Editor of a Springer Books Series. Presently, he is Associate Editor of the Journal of Integrative Neuroscience published by IOS Press, the Netherlands. He has established an MOU between his university and the University of Montreal, Canada, for various joint research activities. He has also established collaboration with the National Institute for Materials Science (NIMS), Japan, for joint research activities and visits NIMS as a visiting scientist. He organizes international conference series such as SoCTA and ICoEVCI as General Chair. He is Executive Committee Member of IEEE Rajasthan Chapter.

Dr. S. N. Sharan is responsible for academic and administrative monitoring of the Department of Electrical, Electronics and Communication Engineering at Manipal University, Jaipur, India. In addition to this, he is also involved in teaching and research in the field of electronics and communication engineering/energy. His areas of expertise are energy and digital communication. He has 2 books, 20 research papers, and 15 conference papers published to his credit. He has evaluated over ten Ph.D. theses pertaining to the areas of energy and signal processing.

Dr. Sanyog Rawat is presently associated with the Department of Electronics and Communication Engineering, SEEC, Manipal University, Jaipur, India. He holds a B.E. in electronics and communication, an M.Tech. in microwave engineering, and a Ph.D. in planar antennas. He has been involved in organizing various workshops on "LabVIEW" and antenna designs and simulations using FEKO. He has taught

various subjects, including electrical science, circuits and system, communication system, microprocessor systems, microwave devices, antenna theory and design, advanced microwave engineering, and digital circuits.

Dr. S. K. Jain is Head of the Department of Physics, Manipal University, Jaipur, India. His areas of interests are condensed matter physics, materials development and their characterization, and development of magnetic materials. He completed his B.Sc., M.Sc., and Ph.D. (condensed matter physics) from the University of Rajasthan, India. He has qualified NET, SLET, and GATE examinations. He is Executive Committee Member of Indian Physics Association, Jaipur, India, and Life Member of Material Research Society of India. He is currently teaching subjects like engineering physics, nanoscience and nanotechnology-II, oscillation and wave optics, and classical mechanics. He holds several publications to his credit.

Dr. Sumit Srivastava is Professor and Head of the Department of Information Technology, Manipal University, Jaipur, India. He has expertise in the domain of data analytics and image processing. He is also Senior Member of IEEE with more than 14 years of teaching experience. He has organized many national and international conferences under IEEE and Springer banner in Rajasthan. He has also served as Publication Chair of two consecutive editions of IEEE International Conference on MOOC, Innovation and Technology in Education (MITE). He has coordinated and mentored the bilateral exchange between the HEIG-VD University, Switzerland, and Manipal University, Jaipur, India, for two consecutive years. He has published more than 50 research papers in peer-reviewed international journals, book series, and IEEE/Springer/Elsevier conference proceedings.

Dr. Anirban Bandyopadhyay is Senior Scientist at the National Institute for Materials Science (NIMS), Tsukuba, Japan. He completed his Ph.D. in supramolecular electronics from the Indian Association for the Cultivation of Science (IACS), Kolkata, 2005. From 2005 to 2008, he was Independent Researcher, called ICYS Research Fellow at the International Center for Young Scientists (ICYS), NIMS, Japan, where he worked on the brain-like bio-processor building. In 2008, he joined as Permanent Scientist at NIMS, working on the cavity resonator model of human brain and design synthesis of brain-like organic jelly. From 2013 to 2014, he was Visiting Scientist at the Massachusetts Institute of Technology (MIT), USA. He has received several honors, such as the Hitachi Science and Technology Award 2010, Inamori Foundation Award 2011–2012, Kurata Foundation Award, Inamori Foundation Fellow (2011–), and Sewa Society International Member, Japan. He has patented ten inventions: (i) a time crystal model for building an artificial human brain, (ii) geometric musical language to operate a Fractal tape to replace the Turing tape, (iii) fourth circuit element that is not memristor, (iv) drugs for cancer and Alzheimer's diseases, (v) nanosubmarine as a working factory and nanosurgeon, (vi) Fractal condensation-based synthesis, (vii) a thermal noise harvesting chip, (viii) a new generation of molecular rotor,

(ix) spontaneous self-programmable synthesis (programmable matter), and (x) Fractal grid scanner for dielectric imaging. He has also designed and built multiple machines and technologies—(i) THz-magnetic nanosensor, (ii) a new class of fusion resonator antenna, etc. Currently, he is building time crystal-based artificial brain using three ways—(i) knots of darkness made of fourth circuit element, (ii) integrated circuit design, and (iii) organic supramolecular structure.

Temporomandibular Joint Syndrome Prediction Using Neural Network

Navodit Sharma, Ishfaq Gaffar Dar, Jayesh Kumar, Azzan Khan and Anita Thakur

Abstract The temporomandibular joints (TMJs) consist of complex formation of bones, muscles and tendons. Pain in jaw area is a result of disorders and injury of these structures. This TMJ disorder causes tinnitus headaches, vertigo, migraines and several TMJ arthritides. Prediction of TMJ syndrome is complex because this joint is different from the load joint of knee or hip. For diagnosis and prediction of TMJ syndrome, artificial intelligent techniques are indeed worth exploring. In this paper, the proposed model is based on neural network theory which will be helpful to sense whether a patient is suffering from TMJ disorder or not. This model automatically predicts the TMJ on the basis of risk factors and symptoms given by the patients.

1 Introduction

Today, around 30–40% of adults have oral problems and the second most common cause of oral problems is TMJ disorder [1, 2]. TMJ disorder is a condition that causes pain and dysfunction of jaw joint and the muscles which help in jaw movement. The

N. Sharma (✉) · I. G. Dar · J. Kumar · A. Khan · A. Thakur
Department of Electronics and Communication Engineering, Amity University Noida, Noida
201303, Uttar Pradesh, India
e-mail: navoditd@gmail.com

I. G. Dar
e-mail: ishfaqgaffar6@gmail.com

J. Kumar
e-mail: jayeshkr11@gmail.com

A. Khan
e-mail: azzanali_khan@yahoo.com

A. Thakur
e-mail: athakur@amity.edu

temporomandibular joint connects the lower jaw, also known as the mandible to the bone situated at the side of head called the temporal bone. Because the joint is flexible in nature, the jaw has the freedom to move freely up and down and from side to side, which allows us to perform various functions like talk, chew and yawn. Muscles which are attached to and around the joint control the position as well as movement of joint. When we are opening our mouth, the round end of the lower jaw, known as condyles, slide along the socket of the joint of temporal bone [3]. The condyles return to their initial position as the mouth is closed. To ensure that this motion remains smooth, a soft disc is present between condyles and the temporal bone. The function of this disc is to absorb shocks generated from chewing and other movements. The temporomandibular joint is slightly different from other joints in the body. This joint is not only a hinge but also provides sliding motions which make this joint very complicated. Also, this joint is made up of tissues unlike any other load bearing joint tissues in the body. Due to the complex motion and unique build, the muscles and jaw joint pose an enormous challenge to both healthcare provider and patient when any problem occurs.

Disorders related to the jaw joint and chewing muscles and how people respond to them vary widely. Researchers are generally in agreement that the condition mainly falls into three categories [4, 5]:

1. Myofascial pain involving distress or ache in the muscles which operates the jaw.
2. Internal derangement or improper fitting of the jaw which includes dislocated jaw, injury to the condyle or displaced disc.
3. Arthritis concerns the group of inflammatory/degenerative joint disorders which affects the temporomandibular joint.

A person can have any one or more of the above conditions at the same time. It is possible that someone has different health issues which coexist with TMJ disorders, such as sleep disturbance, chronic fatigue syndrome or fibromyalgia, a painful situation which has effects on muscles and some other soft tissues in the entire body. Rheumatic diseases like osteoarthritis or rheumatoid arthritis can also have an effect on the temporomandibular joint. Rheumatic diseases concerns with a group of disorders that result in ache, stiffness in the joints and inflammation in muscles and bone [6]. TMJ disorder has resembling symptoms. How disorders of muscles and jaw joint progress over time is not clear and varies from patient to patient. It has been observed that symptoms have eased and worsened with time, but the reason behind these changes cannot be predicted. Most patients have a comparatively mild form of disorder where significant improvement in symptoms is seen with time but there are cases where the pain is consistent and unbearable.

Artificial Neural Network (ANN) has become very popular recently in medical research and studies [7, 8]. Several studies have been proposed based on ANN models for prediction of different diseases like hypertension, cardiac arrhythmias and tuberculosis. It is more popular because it predicts the solution with fuzzy and incomplete data set. Neural network provides a flexible model for the prediction-based application. It does not require prior hypothesis for input and output data set, also not impose any functional form in data set. Neural network theory has much strength to

handle the big data set and missing data relation between input and output data set. It has hidden property to handle the fuzziness in data set. That why it is used in many fields like biology, avionics, communication, medical and so many [9, 10].

This paper proposes a solution for TMJ syndromes prediction using neural network model which can be prediction based on risk factor and symptoms of TMJ. We have collected more than 2300 patient's data of TMJ syndrome. Following organization method is used in this paper. TMJ prediction method is described in Sect. 2. Model simulation and results are in Sect. 3. Section 4 shows the conclusion of the TMJ syndrome prediction.

2 TMJ Prediction Method

2.1 TMJ Disorder Risk and Symptoms

As we know, the jaw is working as a joint which regulates our actions like yawn, eat, talk, etc. Jaw joint known as temporomandibular joint is connected to skull, teeth and spine with various parts of the body. They are very delicate in nature. This joint controls the various musculoskeletal systems; in case of any type of misalignment or imbalance, it can lead to other bodily problems. Like in plethora occurred due to TMJ disorder which is due to incorrect bite habits. For many people, it starts with bad bite; for others, it starts with clicking or popping sound in movement, although only clicking sound is not sufficient factor to call the condition TMJ disorder and warrant treatment.

The different risk factors and symptoms that are associated with TMJ are locking of jaw or limited movement, painful clicking or grating sound, change in fitting of upper and lower teeth, radiating pain in face, jaw or neck, aching pain in and around ear, jaw muscle stiffness, etc. [11]. There are some other factors which have a significant effect on TMJ, such as diabetes, deficiency of vitamin D, parathyroid and genetic.

2.2 TMJ Prediction Using Neural Network Model

Neural network is a robust method to predict the solution of problem with fuzzy data or missing data [12]. According to neural network theory, every system is defined in layer model. In this paper, TMJ neural network model consists of three layers which have hidden layer in between input and output layers. As first input layer is consisting of the information on risk factors and symptoms of TMJ syndromes, 2300 patient's information is obtained in terms of their symptoms of TMJ and other risks which increase the TMJ syndromes. On the basis of input data correlation, output data is formed in three levels decision. So, the input–output matrix of neural network model

Table 1 Performance comparison matrix

Algorithm for training	Epochs	Correlation coefficient (R)	Mean square error (MSE)
BFGS Quasi-Newton	36	0.98899	0.0048621
Gradient descent backpropagation	1000	0.98429	0.0060292
Resilient backpropagation	27	0.98888	0.0047029
Scaled conjugate gradient	33	0.90792	0.070927
Levenberg–Marquardt	17	0.98919	0.003533
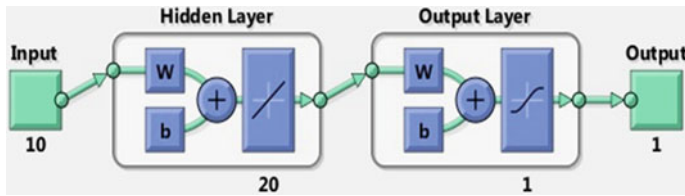
Fig. 1 The TMJ prediction neural network model

is 2300×10 and 2300×3 for the proposed model. In between the input and output layers, one hidden layer is in the model. For interaction of the layer, the activation function is used. Pure line transfer function is used in hidden to output layers, and log sigmoid function is used between hidden and input layers in the proposed TMJ model. To learn and train the supervised error, backpropagation (BP) algorithm is used in the proposed neural model. For the rapid convergence of the system, different training algorithms are implemented. Their comparison performance is shown in Table 1 (Fig. 1).

Input Layer
In neural network model, input layer is the layer which consists of the characteristics data of application. In TMJ syndrome prediction, we used 10 data as input, which include risk factors and symptoms. This symptoms and risk factor included in sample data are prepared by medical experts and dentists [4, 5]. Following risk factors and symptoms are used as input to the neural network model.
S1 Locking of Jaw or limited movement.
S2 Painful clicking, popping or grating sound.
S3 Change in fitting of upper and lower teeth.
S4 Radiating pain in face, jaw or neck.
S5 Aching pain in and around ear.
S6 Jaw muscle stiffness.
R7 Genetic.
R8 Diabetes.
R9 Parathyroid.
R10 Deficiency of vitamin D.

Hidden Layer

In between the input and output hidden layers lies in neural network which extracts the features of the input automatically to minimize the dimensionality. The number of hidden layers in the model that was chosen according to the mean square error is to be reduced. In the proposed model, one hidden layer is used which gives less error in actual output to targeted output.

Output Layer

The last layer is the output layer of the network; it is designed according to the nature of the problem which needs to be predicted. In the proposed TMJ prediction, three stages are needed to predict. They are severe chances, mild chance and no chance of TMJ disorders. Here if the output is 1, it is the case of severe TMJ disorder and the patient needs immediate treatment. If the output is 0.5, then it is the case of moderate disorder and patient should consult a doctor, and if the output is 0, it means no need to concern the doctor.

3 Results and Discussion

The proposed model of TMJ disorder prediction is simulated in neural network MATLAB tool. In which, nftool fitting tool is used to develop the neural network architecture. Where first decide the input, output layer with changeable hidden layer, then, the activation function has to be decided between the layer to get the desired output. Then, network has to be trained with supervised training algorithm by changing the number of neuron in the hidden layer. The best performance of the network
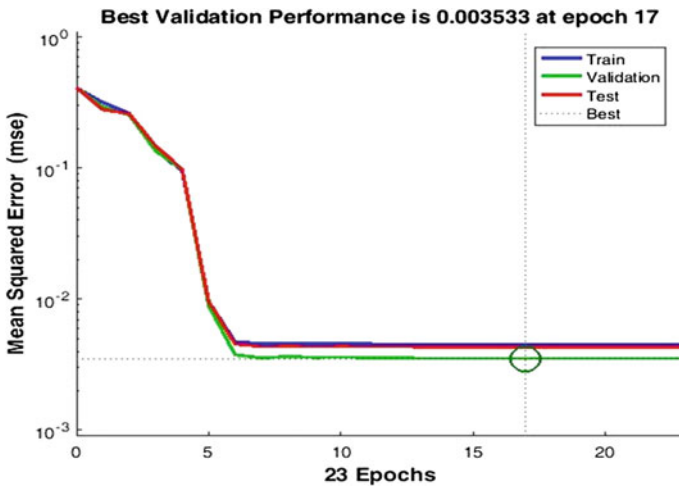
Fig. 2 Performance curve of TMJ neural network model

chooses according to minimum mean square error. The proposed work performance comparison matrix is shown in Table 1, where the mean square error is found out with respect to the number of correlation coefficients, iteration and with varying the training algorithm.

In the proposed neural network model, 2300 patient's data is used in training the network. By performance curve shown in Fig. 2, we get to know that best training performance was meet at epoch 17, and the mean square error is 0.003533 with 20 number of neuron in the hidden layer. We chose the minimum mean square error in the network by changing the number of neuron in the hidden layer.

In neural network model, the output is also known according to input data. So best fitting between the actual data and targeted data can be find out with the help of correlation coefficient. If percentage of correlation coefficient is high, it means less error in model which implies targeted data is near to actual data. Figure 3 shows the correlation coefficient in terms of regression curve, which is 98.9% for the proposed model of TMJ disorder prediction.
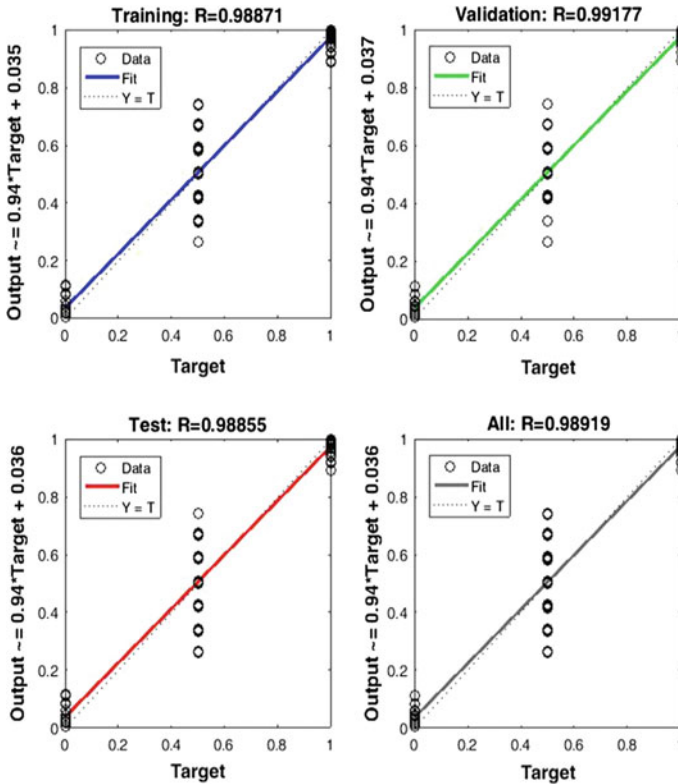
Fig. 3 Regression curve of TMJ prediction model

4 Conclusion

Nowadays, early prediction of chronic disease helps in treatment of patients. In that new era of soft computing techniques helped to the medical practicing and for common people. It is also saved the time, cost and overall valuable life of human being. In the proposed TMJ neural network model which work on the basis of risk factor and symptoms, the neural TMJ network is trained with various training algorithms, and best results are listed in the paper. The correlation coefficient of output data and target data is 98.9% of the proposed model. So the early prediction of TMJ disorder using neural network technique is value added to the treatment of dentistry.

References

1. Stephen, V.E., Kamakshi, V.: Occlusal splint treatment in Temporomandibular disorder. J. Pain Manag. **6**(4), 291–295 (2013)
2. http://www.who.int/mediacentre/factsheets/fs318/en/
3. Sidebottom, A.J.: Current thinking in Temporomandibular joint management. Br. J. Oral Maxillo Facial Surg. **47**(2), 91–94 (2009
4. Guarda, L., Manfredini, D., Ferronato, G.: Temporomandibular joint total replacement prosthesis: current knowledge and considerations for the future. Int. J. Oral Maxillo Facial Surg. **37**, 103–110 (2008)
5. Fernandez, C.E., Amiri, A., Jaime, J., Delaney, P.: The relationship of whiplash injury and temporomandibular disorders: a narrative literature review. J. Chiropr. Med. **8**(4), 171–186 (2009)
6. Manfredini, D., Winocur, E., Guarda-Nardini, L., Paesani, D., Lobbezoo, F.: Epidemiology of bruxism in adults: a systematic review of the literature. J. Orofacial Pain **27**(2), 99–110 (2013)
7. Wang, A., An, N., Xia, Y., Li, L., Chen, G.: A logistic regression and artificial neural network-based approach for chronic disease prediction: a case study of hypertension
8. Lisboa, P.J.G., Etchells, T.A., Jarman, I.H., Aung, M.S.H., Chabaud, S., Bachelor, T., Perol, D., Gargi, T., Bourdes, V., Bonnevay, S., Negrier, S.: Time-to-event analysis with artificial neural networks: an integrated analytical and rule-based study for breast cancer. In: International Joint Conference on Neural Networks, pp. 2533–2538 (2007)
9. Shanthi, D., Sahoo, G., Saravanan, N.: Designing an artificial neural network for the prediction of thromboembolic stroke. Int. J. Biom. Bioinform. **3**(1), 10–18 (2009)
10. Thakur, A., Bhanot, S., Mishra, S.N.: Early diagnosis of ischemia stroke using neural network. In: Proceedings of the International Conference on Man-Machine Systems (ICoMMS), Malaysia (2009)
11. Tanaka, E., Detamore, M.S., Mercuri, C.G.: Degenerative disorder to the Temporomandibular joint: etiology, diagnosis, and treatment. J. Dent. Res. **87**(4), 296–307 (2008)
12. Haykin, S.: Neural Networks: A Comprehensive Foundation, 2nd edn. Prentice Hall (1999)

Optimization of AlGaN QW Heterostructure for UV Applications

Richa Dolia, M. Abu-Samak and P. A. Alvi

Abstract This paper reports the existence of ultraviolet (UV) optical gain in AlN/AlGaN/AlN nanoscale heterostructure consisting of 50 nm quantum well (QW) of AlGaN material sandwiched between layers of AlN material. The designed heterostructure is of type-I and assumed to be grown on GaN substrate. The optical gain of the heterostructure has been optimized utilizing k.p method. The simulation result shows that the peak optical gain in the AlGaN QW heterostructure lies at ~2400 Å (UV region) of the order of ~760 cm, which proves the potentiality of the AlN/AlGaN/AlN heterostructure as a source of UV radiations.

Keywords Optical gain · Type-I heterostructure · AlN · AlGaN

1 Introduction

Recently, nitride-based heterostructures have attracted a great attention of entire scientific community due to its potential applications in field effect transistors (FETs), laser diodes (LDs), light emitting diodes (LEDs), and tandem solar cells (TSC) [1–5]. In the present decade, Alvi et al. [6–9] have analyzed the AlGaN/GaN-, GaInN/InN-, and AlInN/InN-based multilayer heterostructures and reported their band structures and strain profiles. In order to realize these structures, both the growth quality of epilayers and fabrication of heterojunctions (HJs) are required. Apart from this, the ternary material AlInN has also been much popular due to its potential characteristics, such as greater index of refraction, and therefore, it proves its potentiality in the

R. Dolia
Department of Electronics, Banasthali Vidyapith, Banasthali 304022, Rajasthan, India

M. Abu-Samak
Department of Physics, Al-Hussein Bin Talal University, P.O. Box 20, Ma'an, Jordan

P. A. Alvi (✉)
Department of Physics, Banasthali Vidyapith, Banasthali 304022, Rajasthan, India
e-mail: drpaalvi@gmail.com

distributed Bragg reflectors (DBRs), high mobility electron transistors (HEMTs), and ultraviolet optical confinement laser diodes [10–12].

So far reported heterostructures are suitable for visible and infrared regime, and most of them are based on antimonide and arsenic materials such as InGaAlAs, InGaAsP, InGaAs, AlGaAs, and GaAsSb [13–16]. But keeping in views the functionality in UV region, the III-nitride-based nanoscale heterostructure is designed and found that it is suitable heterostructure as a light source for the UV wavelength region. Hence, interestingly the objective of this paper is to focus on the numerical simulation and analysis of the energy band diagram along with the associated wave functions, dispersion curve, and the gain spectrum of AlN/AlGaN/AlN nano–heterostructure.

2   Simulation Results and Discussion

Numerical simulations were performed by solving the 6×6 Luttinger–Kohn Hamiltonian along with effective mass approximation and description of the conduction band, and valence sub-bands (i.e., light and heavy hole sub-bands) were made in order to study the behavior of conduction band electrons and valence band holes (light and heavy holes). By knowing their localization or probability density, the type of designed heterostructure (i.e., type-I or type-II) can be confirmed. In the calculations of gain spectra, the valence split-off sub-bands were not taken into account because these were found to lie much deeper in the valence band. The necessary parameters required during calculations were taken from literature [17]. In our calculations, the expansion of wave functions associated with holes was made in terms of basis functions of Luttinger–Kohn representation with $J = 3/2$, where J is the total angular momentum. Their projections were corresponding to light holes and heavy holes for $m_J = \pm 1/2$ and $m_J = \pm 3/2$, respectively [18].

The resulting wave functions associated with the conduction and valence sub-bands (light and heavy hole bands) of the designed heterostructure are shown in Fig. 1. According to the calculations, this is strained heterostructure. It can be confirmed from Fig. 1, in which light holes are found to lie above the heavy holes. From Fig. 1, it is also clear that the wave function associated with the electron of the first energy state is high in the central part of the QW, and hence, it plays important role in the optical transitions, while the electrons of second and third energy states do not contribute in the transitions responsible for optical gain because the wave functions associated with these electrons are almost negligible (i.e., flat wave functions). Also focusing on valence sub-bands, the wave functions of heavy hole are high in magnitude rather than light holes, and hence, the contribution of heavy holes is greater than the light holes in the optical gain spectra of the designed heterostructure. The plot of dispersion relation (E-K curve) for the heterostructure is shown in Fig. 2. In Fig. 2, it can be predicted that the energy curves of conduction band electrons are almost parabolic, while energy curves of valence sub-bands are non-parabolic. Again, the location of heavy holes can be seen above the light holes which confirm the existence of strain (tensile strain) in the heterostructure. Again from Fig. 2, it can

be observed that for the strained AlGaN QW under investigation, the first valence sub-band (i.e., first heavy hole-hh1 state) corresponds to the ground state while the second valence sub-band (i.e., first light hole-lh1 state) corresponds to the just next one on the energy scale (see Fig. 2). The optical gap is actually determined by the optical transitions made between the lowermost electron state of the conduction band (i.e., e1) and the uppermost hole state of the valence band (i.e., hh1). Hence, the transitions between e1 state and hh1 state decide the regime of the optical gain of the designed heterostructure. Refer Fig. 2, the interesting feature is that the split-off holes (SOHs) are also the next one (i.e., valence sub-bands), but these holes do not contribute in the optical transitions due to their low-lying locations. In the E-K curve, this gap is of the order of ~5 eV, and hence, the radiations produced resulting from the transitions between e1 and hh1 will lie within the UV wavelength region. This is also proved from the gain spectra shown in Fig. 3a, b. The optical gain is the basic characteristic of the quantum well heterostructures or lasing heterostructures by virtue of which their application (in terms of wavelength regime) is decided in optoelectronic devices. In fact, the gain occurs due to recombination of electrons and holes, and its direct measurement gives the information on the stimulated recombination process. The optical gain characteristic is plotted in terms of photonic energy in Fig. 3a and in terms of wavelength in Fig. 3b. In Fig. 3, it can be observed that the optical gain spectrum has a single broadening peak which indicates that the major contribution in optical transitions is from the e1 and hh1 and a little bit from e1 and lh1, while no contribution is given from e1 and SO holes. From Fig. 3, it is clear that the peak optical gain is achieved of the order of ~700 cm at the energy ~5.165 eV and at wavelength ~2400 Å (wavelength of UV region). This shows that the emitted radiations from the heterostructure can have possible energy of the order of ~5.165 eV and corresponding wavelength ~2400 Å (UV wavelength). Here, it is important to note that the optical gain calculated for this heterostructure can be enhanced by increasing the concentration of 2D carrier injection. Thus, the designed AlN/AlGaN/AlN nano–heterostructure can be suitable as a source of UV radiations. For example, recently, GaN-based LEDs (that include multiple quantum wells) have been designed, and their properties, such as localized carriers density, normalized optical gain, power spectral density, and luminous power, have been simulated by using 6×6 Kohn–Luttinger Hamiltonian. The designed LEDs have been reported to show their spectral emission peaks in UV range [19]. The development of GaN-based LEDs as a UV radiation source has replaced UV lamps. Moreover, the UV-LEDs have been reported to show their increased power up to 12 W (14 A) [20]. Apart from this, UV-LEDs have shown significant advantages over the mercury lamps such as high efficiency, low power consumption, robustness, and non-toxicity. Due to these all factors, LEDs are more preferred rather than the other conventional sources.

Fig. 1 Representations of
wave functions associated
with conduction and valence
band of AlN/AlGaN/AlN
nano–heterostructure
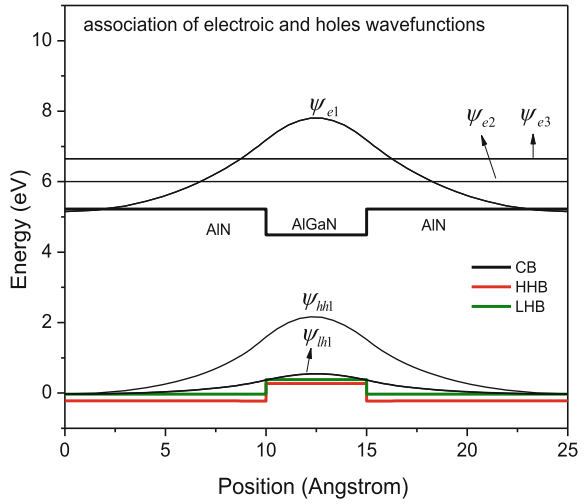
Fig. 2 Representation of
dispersive curve of
conduction band electron
and valence sub-bands of
AlN/AlGaN/AlN
nano–heterostructure
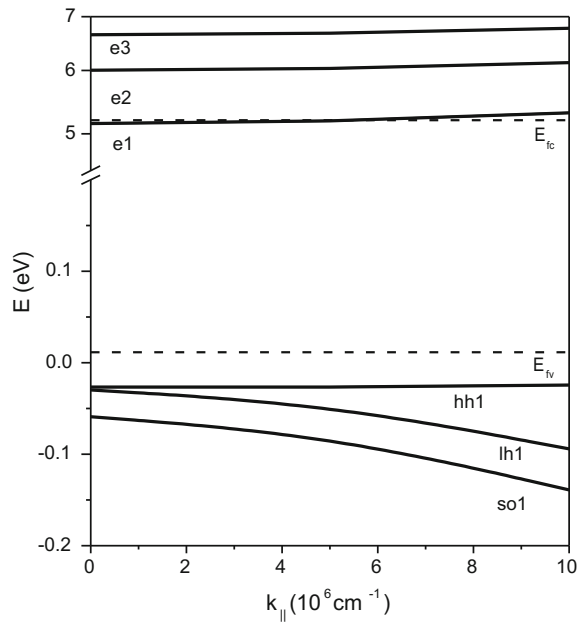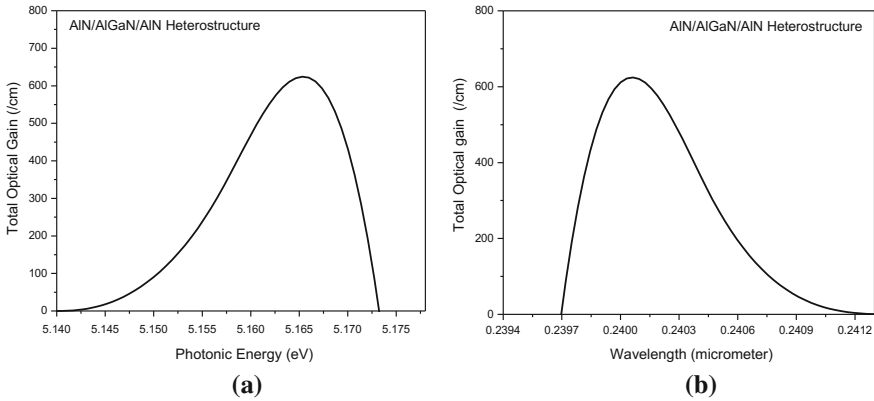
Fig. 3 Gain spectra in terms of **a** energy and **b** wavelength of AlN/AlGaN/AlN nano–heterostructure

3 Conclusion

The optical gain of AlN/AlGaN/AlN nanoscale heterostructure has been optimized with the help of 6×6 Kohn–Luttinger Hamiltonian. According to the results achieved, it has been found that the peak optical gain has the order of ~700 cm at the energy ~5.165 eV and at wavelength ~2400 Å. Thus, the designed AlN/AlGaN/AlN nano–heterostructure can be suitable as a source of UV radiations.

References

1. Nakamura, S., Pearton, S., Fasol, G.: The Blue Laser Diode. Springer, Berlin (2000)
2. Kuramata, A., Domen, K., Soejima, R., Horino, K., Kubota, S., Tanahashi, T.: Japan. J. Appl. Phys. **36**, L1130 (1997)
3. Sheu, J.K., Huang, F.W., Lee, C.H., Lee, M.L., Yeh, Y.H., Chen, P.C., Lai, W.C.: Appl. Phys. Lett. **103**, 063906 (2013)
4. Wu, F.Y., Keller, P.B., Kapolnek, D., Denbaars, P.S., Mishra, U.K.: IEEE Electron. Dev. Lett. **17**, 455–457 (1995)
5. Wu, F.Y., Keller, P.B., Parikh, P., Kapolnek, D., Denbaars, S.P., Mishra, U.K.: IEEE Electron Dev. Lett. **18**, 290–292 (1997)
6. Alvi, P.A., Gupta, S., Siddiqui, M.J., Sharma, G., Dalela, S.: Modeling and simulation of GaN/Al$_{0.3}$Ga$_{0.7}$N new multilayer nano-heterostructure. Phys. B: Condens. Matter **405**(10), 2431–2435 (2010)
7. Alvi, P.A., Gupta, S., Vijay, P., Sharma, G., Siddiqui, M.J.: Affects of Al concentration on GaN/Al$_x$Ga$_{1-x}$N new modeled multilayer nano-heterostructure. Phys. B: Condens. Matter **405**(17), 3624–3629 (2010)

8. Gupta, S., Rahman, F., Siddiqui, M.J., Alvi, P.A.: Strain profile in III-Nitride based multi-layer nano-heterostructures. Phys. B: Condens. Matter **411**, 40–47 (2013)

9. Alvi, P.A., Gupta, S., Sharma, M., Jha, S., Rahman, F.: Computational modeling of novel $InN/Al_{0.30}In_{0.70}N$ multilayer nano-heterostructures. Phys. E: Low-dimens. Syst. Nanostruct. **44**, 49–55 (2011)

10. Wang, J., et al.: Evidence of type-II band alignment in III-nitride semiconductors: experimental and theoretical investigation for $In_{0.17}Al_{0.83}N/GaN$ heterostructures. Sci. Rep. **4**, 6521 (2014)

11. Carlin, J.-F., Ilegems, M.: High-quality AlInN for high index contrast Bragg mirrors lattice matched to GaN. Appl. Phys. Lett. **83**, 668–670 (2003)

12. Butte, R., et al.: Current status of AlInN layers lattice-matched to GaN for photonics and electronics. J. Phys. D Appl. Phys. **40**, 6328–6344 (2007)

13. Nirmal, H.K., Yadav, N., Rahman, F., Alvi, P.A.: Optimization of high optical gain in type-II $In_{0.70}Ga_{0.30}As/GaAs_{0.40}Sb_{0.60}$ lasing nano-heterostructure for SWIR applications. Superlattices Microstruct. **88**, 154–160 (2015)

14. Nirmal, H.K., Yadav, N., Dalela, S., Rathi, A., Siddiqui, M.J., Alvi, P.A.: Tunability of optical gain (SWIR region) in type-II $In_{0.70}Ga_{0.30}As/GaAs_{0.40}Sb_{0.60}$ nano-heterostructure under high pressure. Phys. E: Low-dimens. Syst. Nanostruct. **80**, 36–42 (2016)

15. Nirmal, H.K., Anjum, S.G., Lal, P., Rathi, A., Dalela, S., Siddiqui, M.J., Alvi, P.A.: Field effective band alignment and optical gain in type-I $Al_{0.45}Ga_{0.55}As/GaAs_{0.84}P_{0.16}$ nano-heterostructures. Int. J. Light Electron Optics: Optik **127**, 7274–7282 (2016)

16. Anjum, S.G., Sandhya, K., Khan, A.B., Khan, A.M., Siddiqui, M.J., Alvi, P.A.: Effects of variation of quantum well numbers on gain characteristics of type-I InGaAsP/InP nano-heterostructure. Bull. Electr. Eng. Inf. **6**(3), 301–310 (2017)

17. Vurgaftman, I., Meyer, J.R., Ram-Mohan, L.R.: J. Appl. Phys. **89**, 5815 (2001)

18. Kolokolov, K.I., Savin, A.M., Beneslavski, S.D., Ya, M.N., Hansen, O.P.: Phys. Rev. B **59**, 7537 (1999)

19. Ahmad, S., Raushan, M.A., Kumar, S., Dalela, S., Siddiqui, M.J., Alvi, P.A.: Modeling and simulation of GaN based QW LED for UV emission. Int. J. Light Electron Optics: Optik **158**, 1334–1341 (2018)

20. Muramoto, Y., Kimura, M., Nouda, S.: Development and future of ultravioletlight-emitting diodes: UV-LED will replace the UV lamp. Semicond. Sci. Technol. **29**, 084004 (2014)

Experimental Performance Evaluation of Cloud Servers in Ad Hoc Cloud Network

Vijaya Lakshmi Singh and Dinesh Rai

Abstract Wireless networks are gaining popularity in today's world and so are the technologies related to them such as cloud network and ad hoc network. However, not every user could adopt cloud for some reasons. In this paper, all these concepts are discussed, and an emerging concept derived from these two technologies, termed as ad hoc cloud, is implemented using multi-hop ad hoc network with cloud servers using services such as FTP, HTTP, E-mail, Database, and Print. In this simulation model, the two protocols of ad hoc network are used, namely, Ad hoc On-Demand Distance Vector (AODV) and Dynamic Source Routing (DSR) in two scenarios, respectively. The performance of the five cloud servers and the ad hoc cloud network using these two protocols is observed and compared. OPNET Modeler 14.5 is used for the simulation.

Keywords Ad hoc network · Ad hoc cloud · Cloud computing · AODV · DSR

1 Introduction

1.1 Ad Hoc Network

Wireless computer networks have induced much interest from the public. Nongovernmental and governmental organizations, agencies, armed forces, universities, and companies are now using this new technology. This wireless network can generally be classified into two categories: fixed infrastructure wireless network and ad hoc wireless network as shown in Fig. 1.

V. L. Singh (✉) · D. Rai
School of Engineering and Technology, Ansal University, Gurgaon, India
e-mail: vijayalakshmisingh@ansaluniversity.edu.in

D. Rai
e-mail: dineshrai@ansaluniversity.edu.in
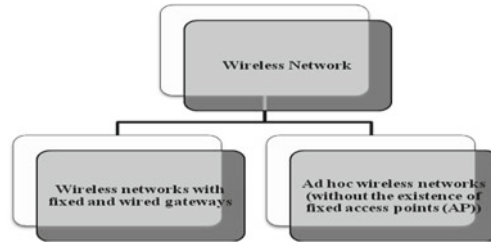
Fig. 1 Categories of wireless network
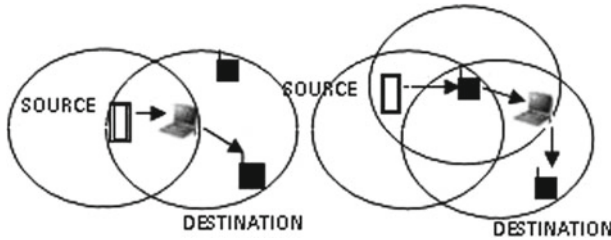
Fig. 2 Ad hoc network [2]

Ad hoc network [1] is self-configuring wireless network that consists of nodes which communicate with each other through wireless medium without any fixed infrastructure. In this network, each node acts as a router and forwards data for the other nodes. Hence, the network is ad hoc. Figure 2 shows the ad hoc network.

1.2 Cloud Computing

A new computing model that originated from distributed computing, grid computing, parallel computing, utility computing, virtualization technology, and other computer technologies. Cloud computing [2] is the only architecture that allows users to access data, application, or service without any add-on service infrastructure, hardware, or software without any location specification. It has various characters such as virtualization, large-scale computation and data storage, high reliability, high expansibility, and low price service that are also their key advantages. Figure 3 explains NSIT definition of cloud computing that represents Internet as cloud.
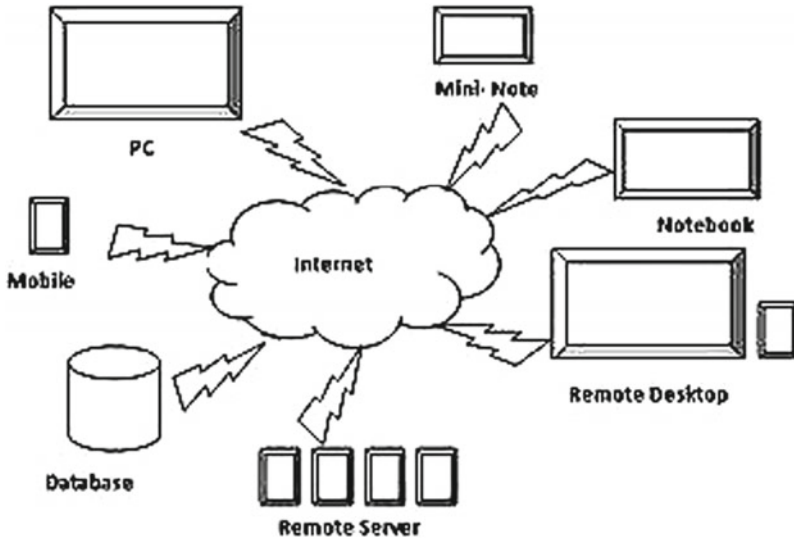
Fig. 3 The term cloud computing seems to originate from computer network diagram that represents the Internet as a cloud [2]

1.3 Ad Hoc Cloud

Personal infrastructure users with a number of underutilized computers, from startup companies to large-scale organizational infrastructures, can benefit by using the concept of ad hoc cloud. Ad hoc clouds produce resources from such existing infrequently available, inclusive (i.e., primarily used for some other purpose), and unreliable infrastructures.

In simple term, ad hoc cloud computing can be defined as running cloud services on ad hoc network. The ad hoc cloud architecture is shown in Fig. 4 [3]. This concept is useful:

- to improve user's infrastructure efficiency and utilization;
- to reduce costs by improving their return on IT investments;
- to those who are not able to use the commercial or private cloud.

There are various advantages and key difference of using ad hoc cloud over various other similar technologies as mentioned in Fig. 5. Ad hoc cloud is different with respect to grid and cloud in various ways, such as ad hoc cloud operates on infrequently available, inclusive (i.e., primarily used for some other purpose), and unreliable hosts that may not be predictable; not assume trust between the infrastructure provider and end user; and ensures job continuity over a number of unreliable host, low interference with executing host processes and targets more diverse workload or applications.
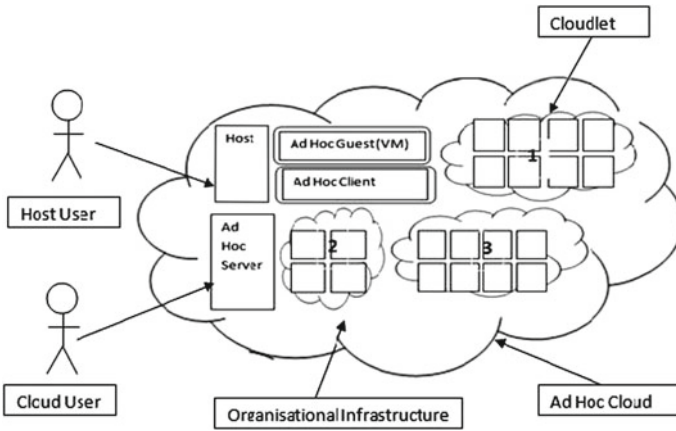
Fig. 4 The ad hoc cloud architecture [2]
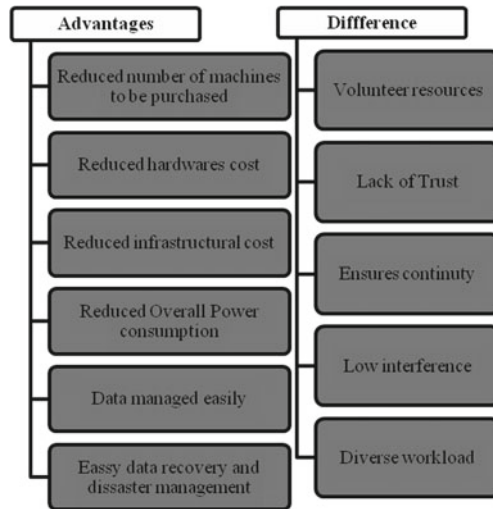
Fig. 5 Advantages and differences of ad hoc cloud over other similar technologies

In this paper, multi-hop ad hoc cloud network is implemented with five cloud servers and four subnets. The routing protocols used to implement this ad hoc network are AODV and DSR. Also, the performances of ad hoc servers in each scenario are observed and compared.

2 Foundation and Related Work

Kirby et al. [3] projected and outlined the ad hoc cloud model and its major challenges of implementation. They also outlined one approach to tackle them. The major claims made by this approach are as follows: to reduce IT costs for the organization; to achieve cloud computing advantages in new application areas; and to reduce IT activities net energy consumption.

Zia and Khan [4] conversed performance issues and parameters of cloud service models that are Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). In this work, they also critically analyzed various Quality of Service (QoS) issues and identified the key challenges in different areas for performance improvement.

Altamas and Niranjan [5] performed analysis in resource management of virtualized systems, examined various issues in performance modeling, and also described CloudSim as a new simulation framework built for virtual cloud computing infrastructures.

Khanghahi and Ravanmehr [6] analyzed and evaluated cloud performance in different scenarios based on the different performance factors of cloud computing are highlighted in this work. They used evaluation based on criteria and characteristics and evaluation based on simulation method in this work.

O'Loughlin and Gillam [7] discussed bad and good metrics for cloud performance. They presented transparent price per unit performance comparisons based on the kinds of performance measures that Cloud service Brokers (CBSs) should use for performance discovery.

Yildirim and Girici [8] implemented cloud and LAN model and compared their performance in FTP, HTTP, and E-mail traffic. Based on their experiment, they concluded that the cloud model shows better performance in all three traffic types. Also, cloud model performance improvement is gathered in all three traffic conditions by assigning Quality of Service (QoS) profile of the cloud model.

Lacuesta et al. [9] planned and developed a trusted algorithm to create a spontaneous ad hoc mobile cloud computing network. They used simulator Castalia to simulate and create such a network. Based on their simulation result, they concluded that their proposal presents good efficiency and network performance even by using a high number of nodes.

McGilvary et al. [10] deliberated the problems for ad hoc cloud computing solutions. They also sketch the architecture of ad hoc cloud based on BOINC.

In this paper, a simulation model is proposed and implemented for multi-hop network that has four subnets and five ad hoc servers using AODV and DSR protocol of ad hoc network. The performance of these servers and wireless LAN is analyzed and comparison of both the protocol is done based on this analysis.

3 Proposed Simulation Model

In the proposed simulation model multi-hop Ad hoc cloud network is implemented that have four subnets (India_Head_office, subnet_Bangkok, subnet_Sydney, and subnet_Tokyo) at different locations with five servers and 25 wireless nodes in India_Head_office and 25 nodes in each of the other three subnets as shown in Figs. 6 and 7. OPNET Modeler 14.5 is used for simulation in this paper. The five ad hoc servers are FTP server, HTTP server, E-mail server, Database server, and Print server, and all these servers run on cloud. There are two scenarios; each consists of these four severs however different protocol at a time. The first scenario implements AODV and second implements DSR. In this architecture, server and client, both work in ad hoc mode with different ad hoc routing protocols. Every second, large amount of traffic is sent by clients to access all services that run on cloud servers. In both the scenarios, simulation runs for 120 s. The performance metric and parameter considered in this work are listed below in Table 1.
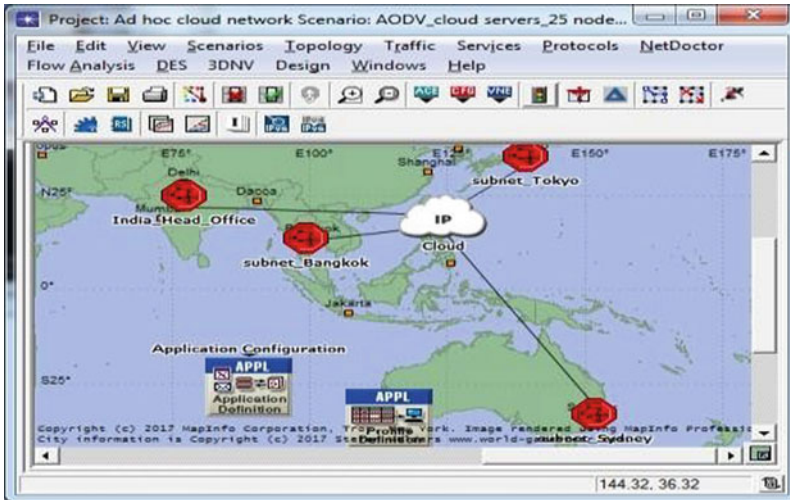
Fig. 6 Multi-hop ad hoc cloud network with four subnets
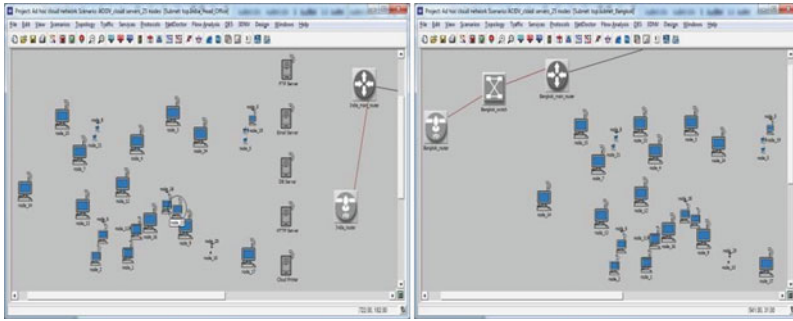
Fig. 7 India_Head_office subnet with five ad hoc servers and 25 nodes (left) and other subnets with 25 nodes (right)

Table 1 Performance metrics and parameters of the simulation model

Global statics	Metric and parameter
DB Query	Response time, traffic received, and traffic sent (bytes or packets per second)
E-mail	Download response time, upload response time, traffic received, and traffic sent (bytes or packets per second)
HTTP	Object response time, page response time, traffic received, and traffic sent (bytes or packets per second)
FTTP	Download response time, upload response time, traffic received, and traffic sent (bytes or packets per second)
Print	Traffic received and traffic sent (bytes or packets per second)
Wireless LAN	Data dropped, delay, load, network load, retransmission, and throughput

4 Results and Discussions

Graphs are generated on the basis of simulation performed. These graphs are used to analyze the experiment and deduce the result of the experiment. The graphs displayed and discussed in the following section based on the global statics, performance metric, and parameter listed in Table 1.

4.1 *Database Query (DB Query)*

This static has metrics such as response time and traffic sent and received (both in terms of bytes/s and packets/s). According to Fig. 8, the average response time of AODV keeps on increasing as compared to DSR that has stable response time. As it is seen from Figs. 9, 10, 11 and 12, the average traffic sent and received (both in terms of bytes/s and packets/s) of AODV is more than DSR.
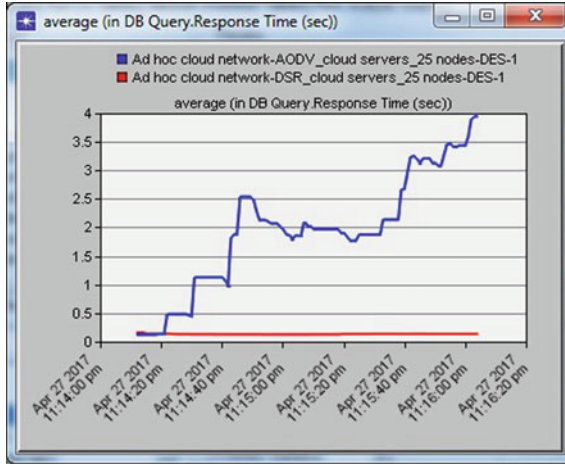
Fig. 8 DB query average response time of AODV (in blue) and DSR (in red)
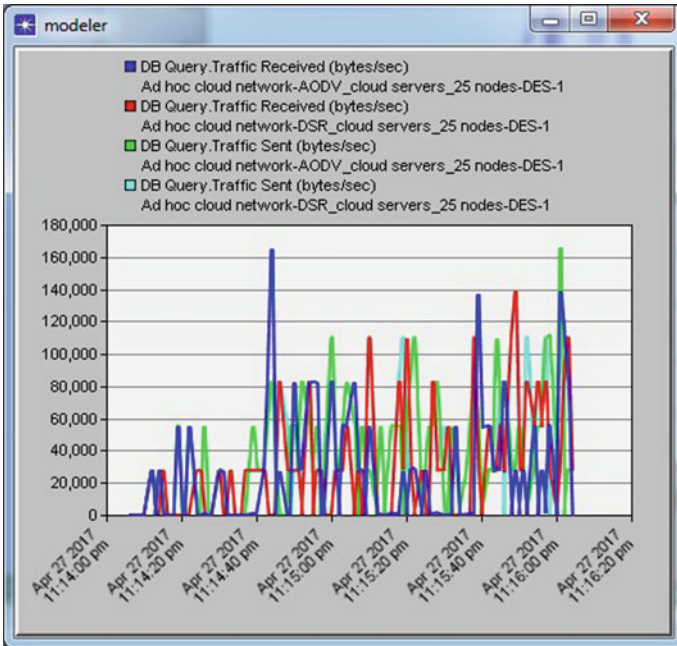
Fig. 9 DB query traffic received (bytes/s) of AODV (in blue) and DSR (in red) and traffic sent (bytes/s) of AODV (in green) and DSR (in sky blue)
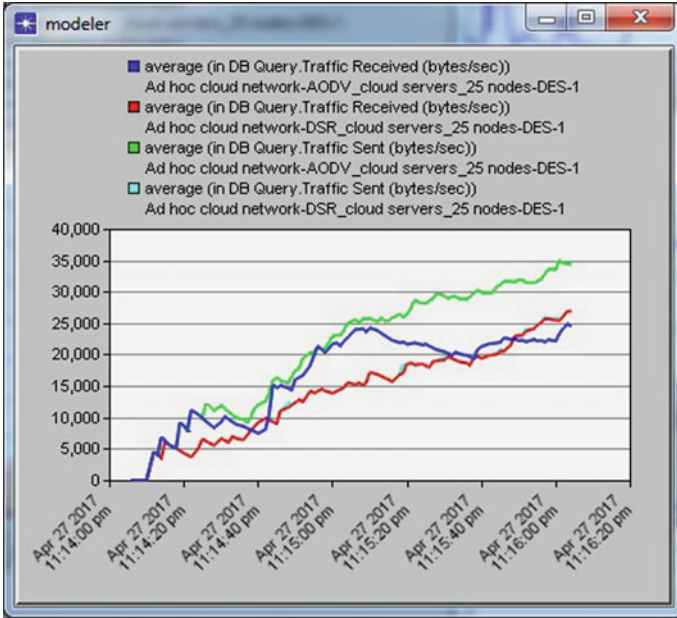
Fig. 10 DB query average traffic received (bytes/s) of AODV (in blue) and DSR (in red) and average traffic sent (bytes/s) of AODV (in green) and DSR (in sky blue)
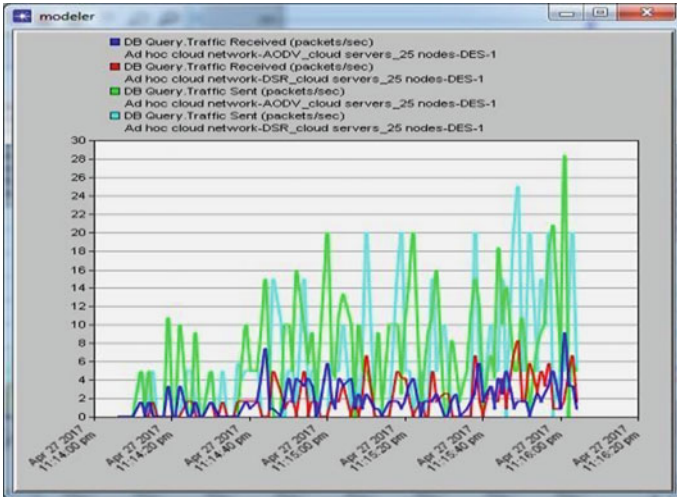
Fig. 11 DB query traffic received (packets/s) of AODV (in blue) and DSR (in red) and traffic sent (packets/s) of AODV (in green) and DSR (in sky blue)
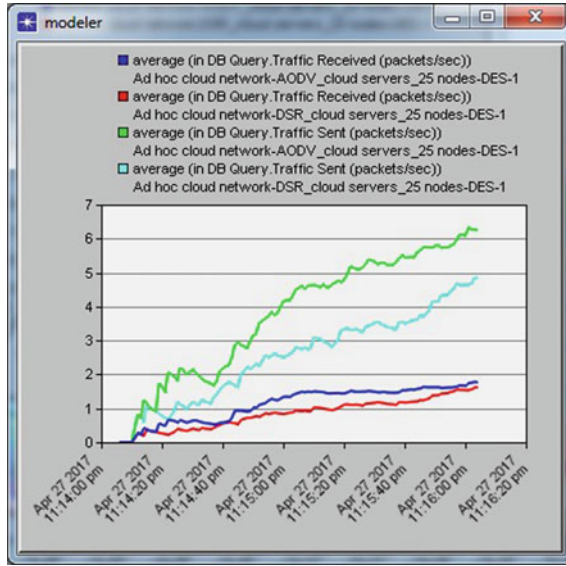
Fig. 12 DB query average traffic received (packets/s) of AODV (in blue) and DSR (in red) and average traffic sent (packets/s) of AODV (in green) and DSR (in sky blue)

4.2   E-mail

Upload and download response time and traffic sent and received (both in terms of bytes/s and packets/s) are the metrics that are considered in this static. In Fig. 13, it is observed that the average download response time of AODV range is 0–10 s while that of DSR is 0–0.1 s. Same is the case with average upload response time. According to Figs. 14, 15, 16 and 17, the average traffic sent and received (both in terms of bytes/s and packets/s) of DSR is more than AODV.

4.3   File Transfer Protocol (FTP)

In this statics, upload and download response time and traffic sent and received (both in terms of bytes/s and packets/s) are the metrics that are considered. In Fig. 18, it is observed that the average download response time of AODV range is 0–8 s while that of DSR is 0–0.3 s. Also, the graph increases in a stepwise fashion in case of AODV while it is constantly the same in case of DSR. Same is the case with average upload response time. According to Figs. 19, 20, 21 and 22, the average traffic sent and received (both in terms of bytes/s and packets/s) of DSR is more than AODV. The average traffic sent (packets/s) of AODV is gradually increasing and then at a point, it started decreasing while it is increasing in case of DSR.
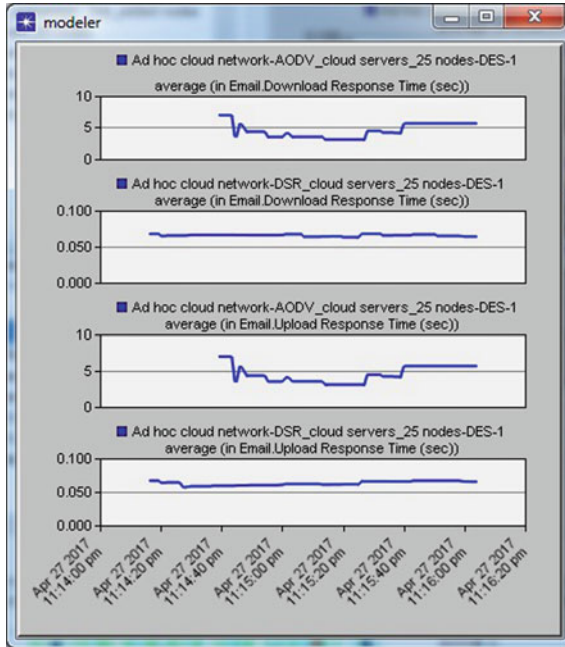
Fig. 13 E-mail average download response time and average upload response time (s) of AODV and DSR
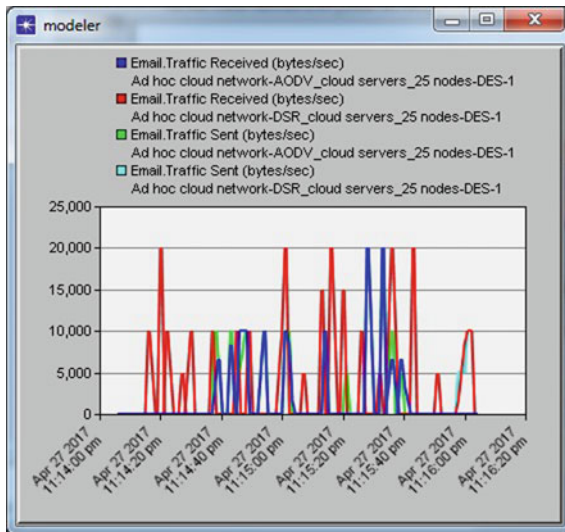
Fig. 14 E-mail traffic received (bytes/s) of AODV (in blue) and DSR (in red) and traffic sent (bytes/s) of AODV (in green) and DSR (in sky blue)
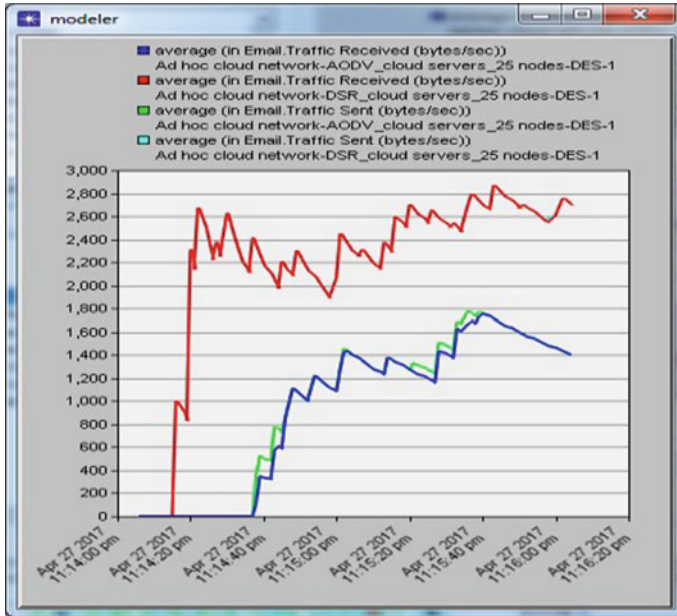
Fig. 15 E-mail average traffic received (bytes/s) of AODV (in blue) and DSR (in red) and average traffic sent (bytes/s) of AODV (in green) and DSR (in sky blue)

4.4   Hypertext Transfer Protocol (HTTP)

Metrics that are considered in this static are object response time, page response time, and traffic sent and received (both in terms of bytes/s and packets/s). According to Fig. 23, the average object response time of AODV range is 0–0.8 s while that of DSR is 0–0.1 s. Also, the average page response time of AODV range is 0–4.5 s while that of DSR is 0–0.5 s. As it is seen from Figs. 24, 25, 26 and 27, the average traffic sent and received (both in terms of bytes/s and packets/s) of DSR is more than AODV.

Fig. 16 E-mail traffic received (packet/s) of AODV (in blue) and DSR (in red) and traffic sent (packet/s) of AODV (in green) and DSR (in sky blue)
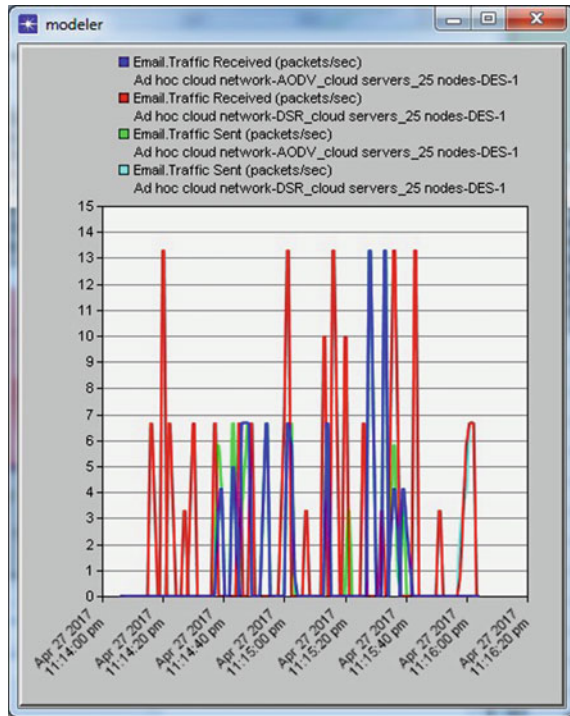
Fig. 17 E-mail average traffic received (packets/s) of AODV (in blue) and DSR (in red) and average traffic sent (packets/s) of AODV (in green) and DSR (in sky blue)
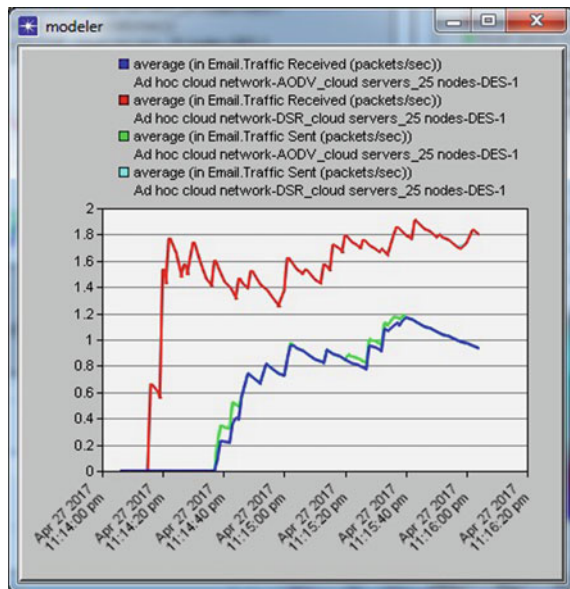
Fig. 18 FTP average
download response time and
average upload response
time (s) of AODV and DSR
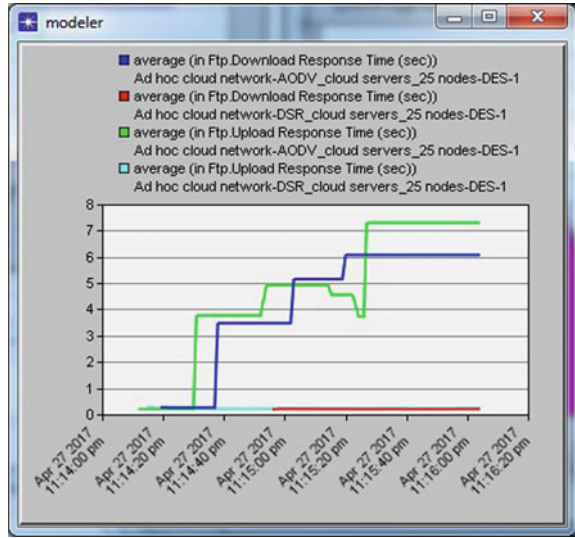
Fig. 19 FTP traffic received
(bytes/s) of AODV (in blue)
and DSR (in red) and Traffic
sent (bytes/s) of AODV (in
green) and DSR (in sky blue)
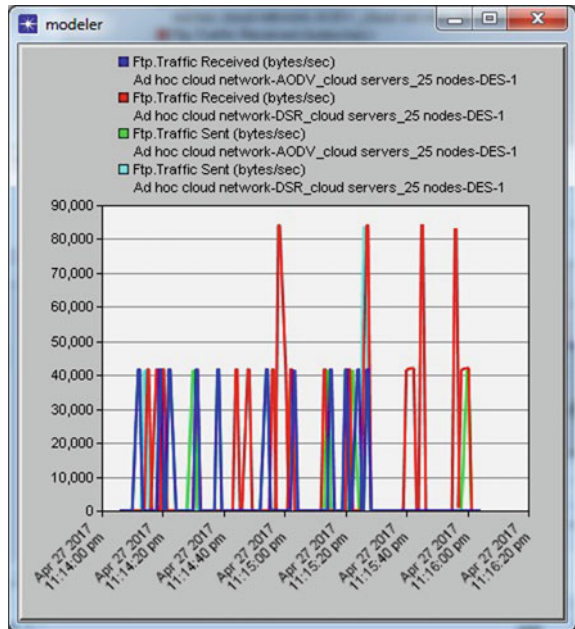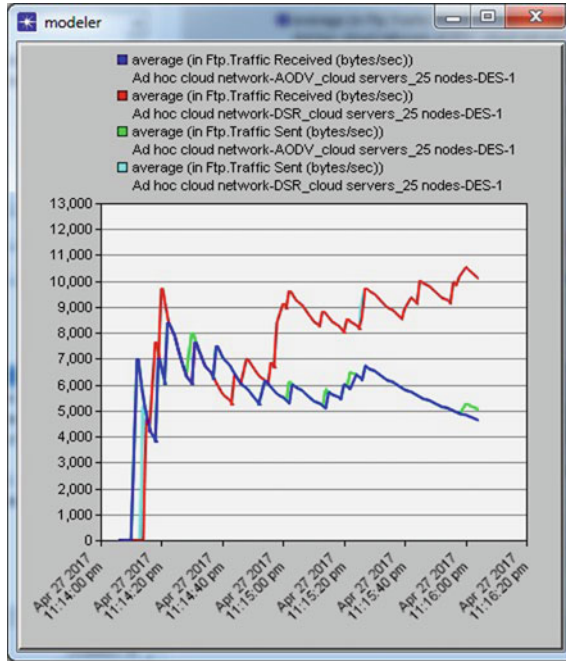
Fig. 20 FTP average traffic received (bytes/s) of AODV (in blue) and DSR (in red) and average traffic sent (bytes/s) of AODV (in green) and DSR (in sky blue)
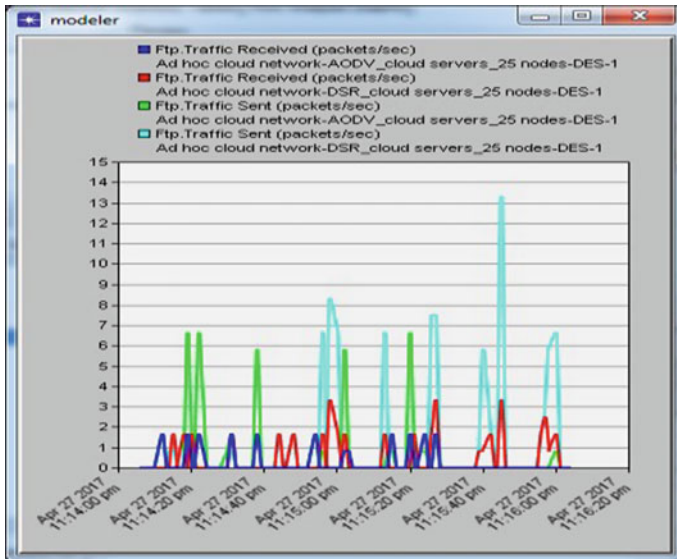
Fig. 21 FTP traffic received (packet/s) of AODV (in blue) and DSR (in red) and traffic sent (packet/s) of AODV (in green) and DSR (in sky blue)

Fig. 22 FTP average traffic
received (packets/s) of
AODV (in blue) and DSR (in
red) and average traffic sent
(packets/s) of AODV (in
green) and DSR (in sky blue)
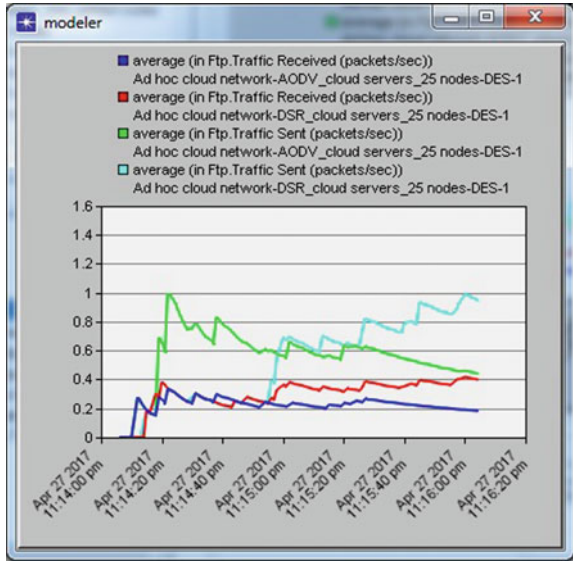
Fig. 23 HTTP average
object response time and
average page response time
(s) of AODV and DSR
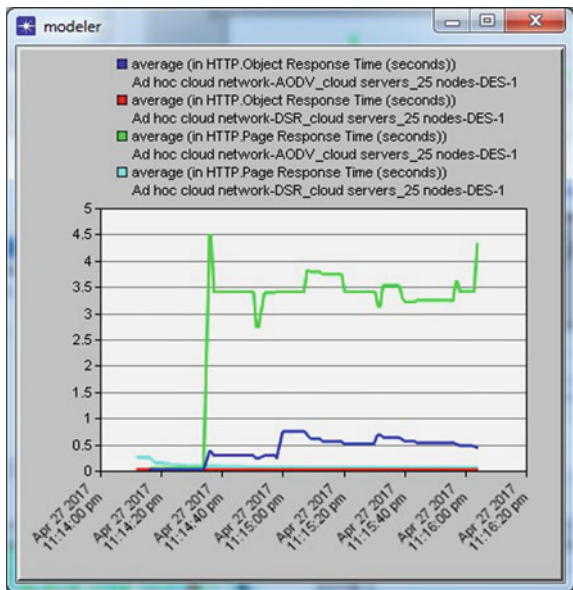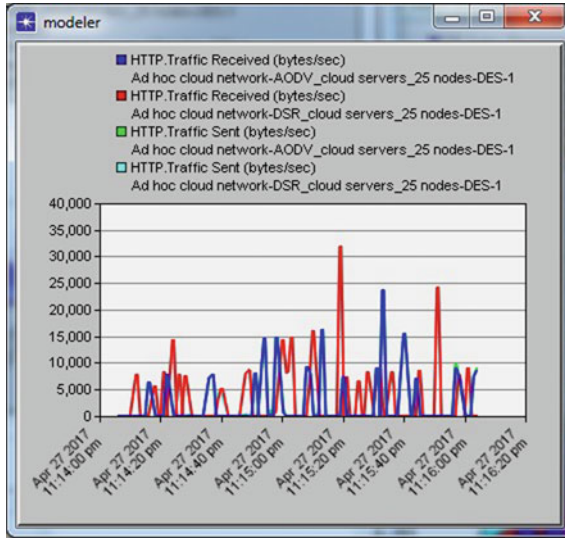
Fig. 24 HTTP traffic received (bytes/s) of AODV (in blue) and DSR (in red) and traffic sent (bytes/s) of AODV (in green) and DSR (in sky blue)
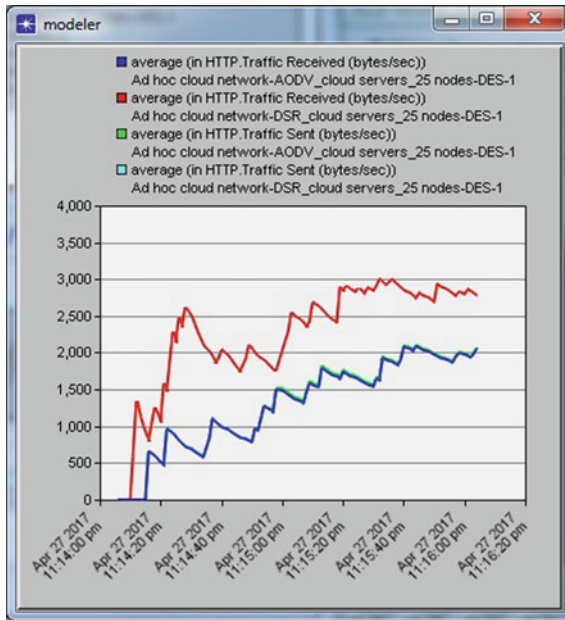
Fig. 25 HTTP average traffic received (bytes/s) of AODV (in blue) and DSR (in red) and average traffic sent (bytes/s) of AODV (in green) and DSR (in sky blue)

Fig. 26 HTTP traffic
received (packet/s) of AODV
(in blue) and DSR (in red)
and traffic sent (packet/s) of
AODV (in green) and DSR
(in sky blue)
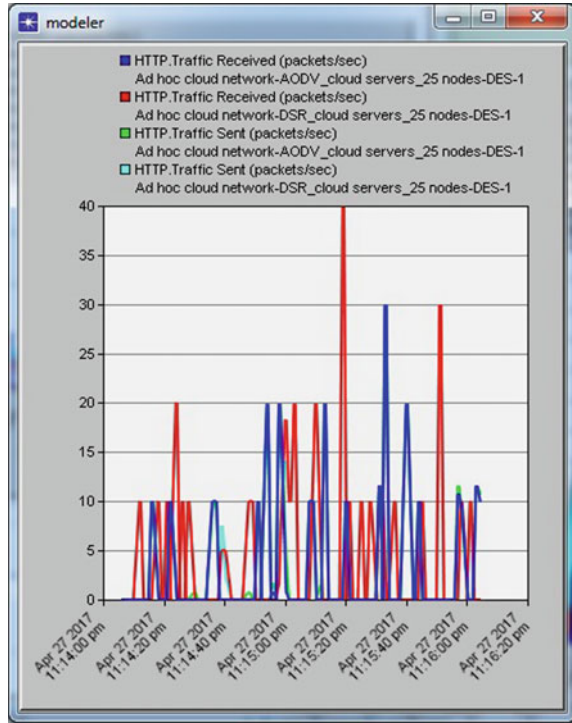
Fig. 27 HTTP average
traffic received (packets/s) of
AODV (in blue) and DSR (in
red) and average traffic sent
(packets/s) of AODV (in
green) and DSR (in sky blue)
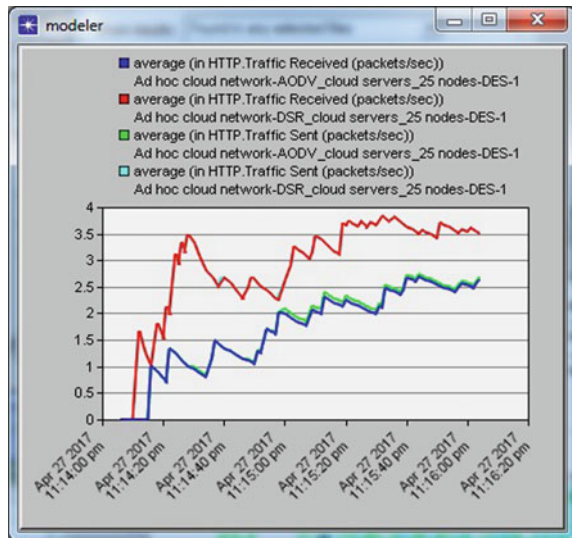
4.5 Print

Traffic sent and received (bytes/s) and traffic sent and received (packets/s) are the metrics that are considered in this static. According to Figs. 28, 29, 30 and 31, the average traffic sent and received (both in terms of bytes/s and packets/s) of AODV is more than DSR.
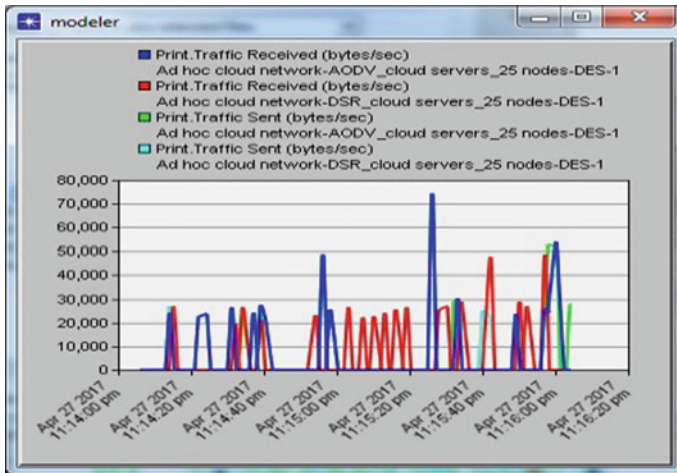
Fig. 28 Print traffic received (bytes/s) of AODV (in blue) and DSR (in red) and traffic sent (bytes/s) of AODV (in green) and DSR (in sky blue)
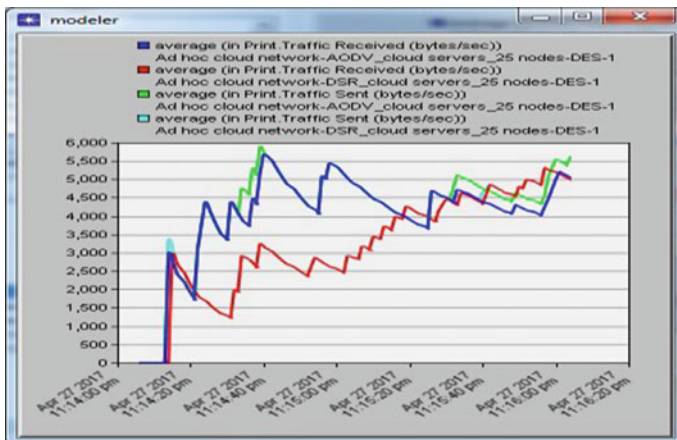
Fig. 29 Print average traffic received (bytes/s) of AODV (in blue) and DSR (in red) and average traffic sent (bytes/s) of AODV (in green) and DSR (in sky blue)
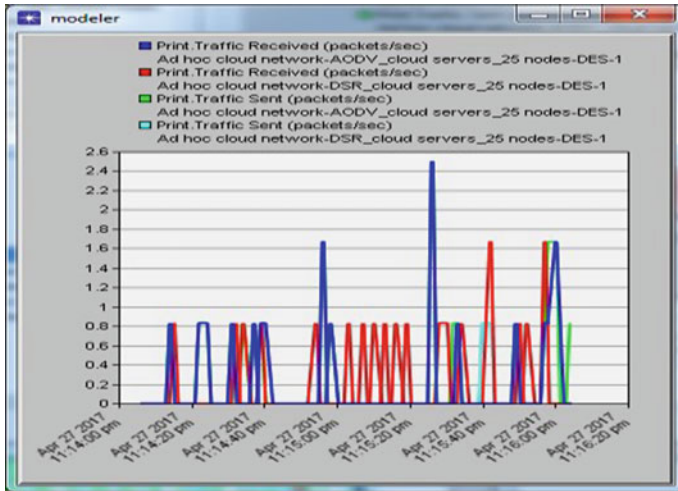
Fig. 30 Print traffic received (packet/s) of AODV (in blue) and DSR (in red) and traffic sent (packet/s) of AODV (in green) and DSR (in sky blue)
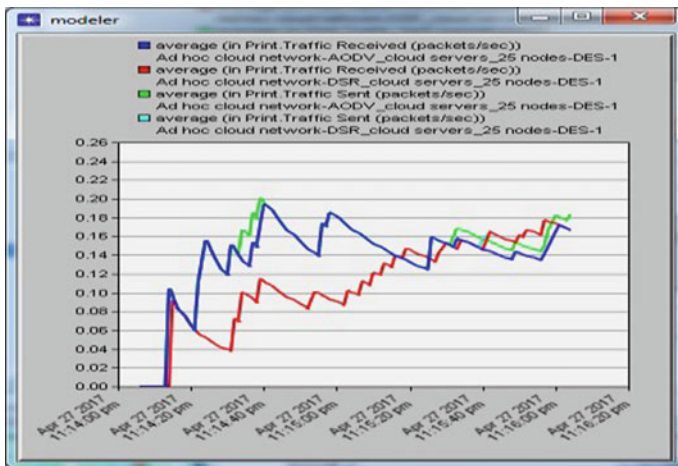
Fig. 31 Print average traffic received (packets/s) of AODV (in blue) and DSR (in red) and average traffic sent (packets/s) of AODV (in green) and DSR (in sky blue)

4.6 Wireless LAN

Wireless LAN has many metrics; however, in this paper, delay, media access delay, load, network load, retransmission attempts, and throughput are considered.
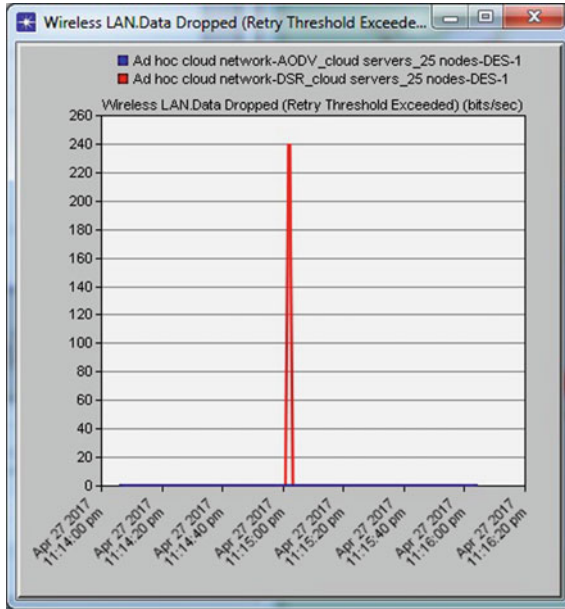
Fig. 32 Wireless LAN data dropped (retry threshold exceeded) of AODV (in blue) and DSR (in red)
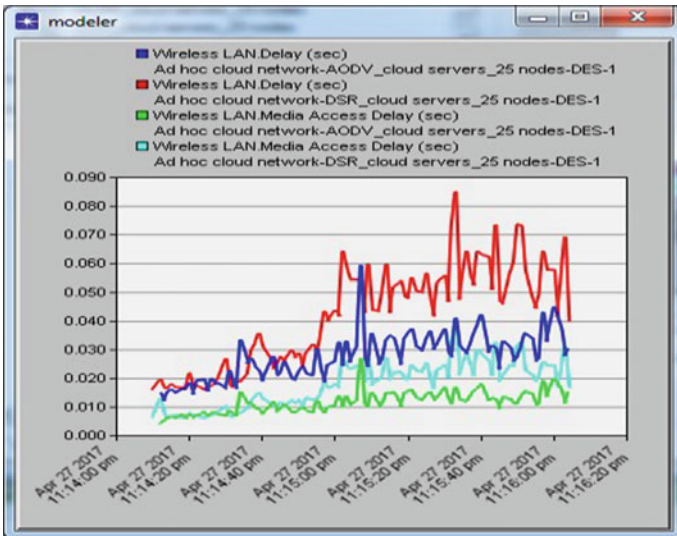
Fig. 33 Wireless LAN delay of AODV (in blue) and DSR (in red) and wireless LAN media access delay of AODV (in green) and DSR (in sky blue)
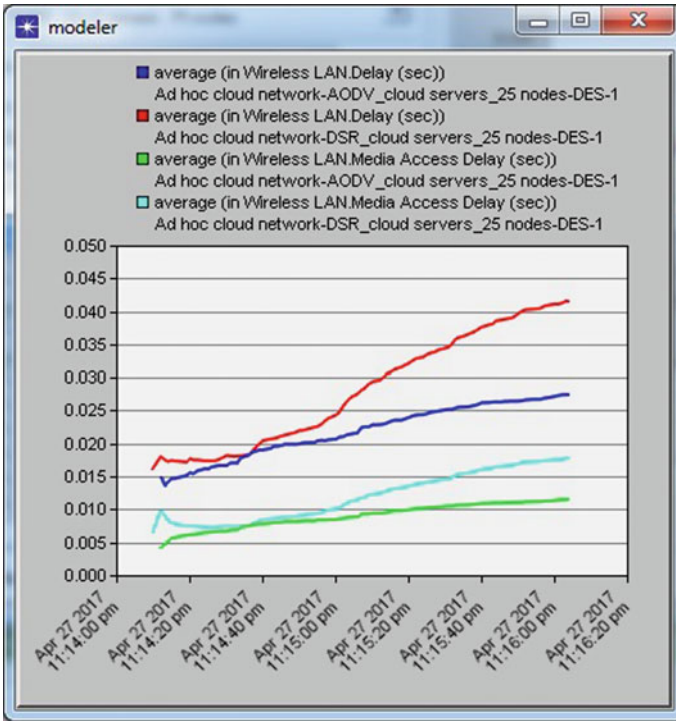
Fig. 34 Wireless LAN average delay of AODV (in blue) and DSR (in red) and wireless LAN average media access delay of AODV (in green) and DSR (in sky blue)
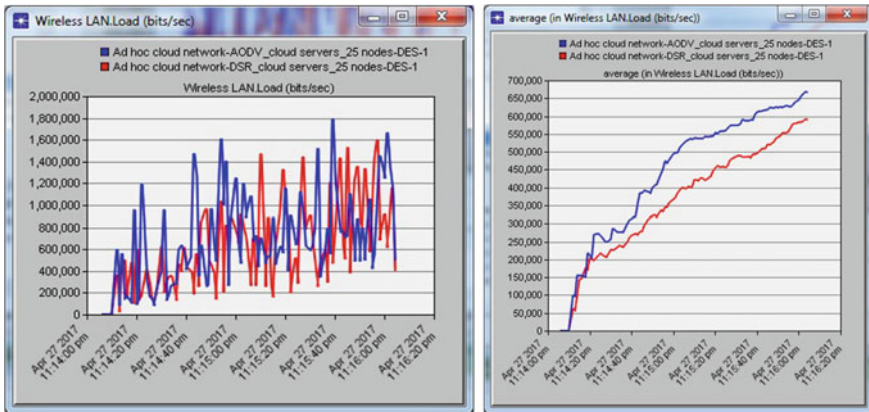
Fig. 35 Wireless LAN load of AODV (in blue) and DSR (in red) (left) and wireless LAN average load of AODV (in blue) and DSR (in red) (right)

Based on the simulation and the graphs from Figs. 32, 33, 34, 35, 36, 37, 38 and 39, the summary of the simulation result for wireless LAN static is listed in Table 2.

Fig. 36 Wireless LAN network load of India_Head_office of AODV (in blue) and DSR (in red); wireless LAN Network load of subnet_Bangkok of AODV (in green) and DSR (in sky blue); wireless LAN network load of subnet_Sydney of AODV (in yellow) and DSR (in dark pink); wireless LAN network load of subnet_Tokyo of AODV (in gray) and DSR (in light pink)
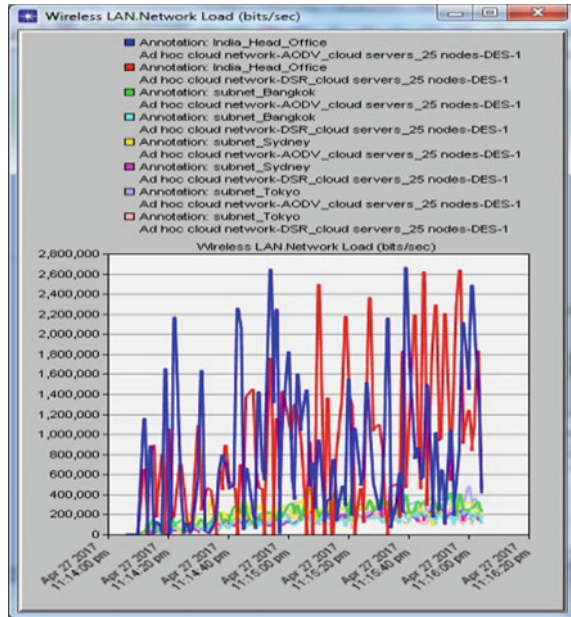
Fig. 37 Wireless LAN average network load of India_Head_office of AODV (in blue) and DSR (in red); wireless LAN average network load of subnet_Bangkok of AODV (in green) and DSR (in sky blue); wireless LAN average network load of subnet_Sydney of AODV (in yellow) and DSR (in dark pink); wireless LAN average network load of subnet_Tokyo of AODV (in gray) and DSR (in light pink)
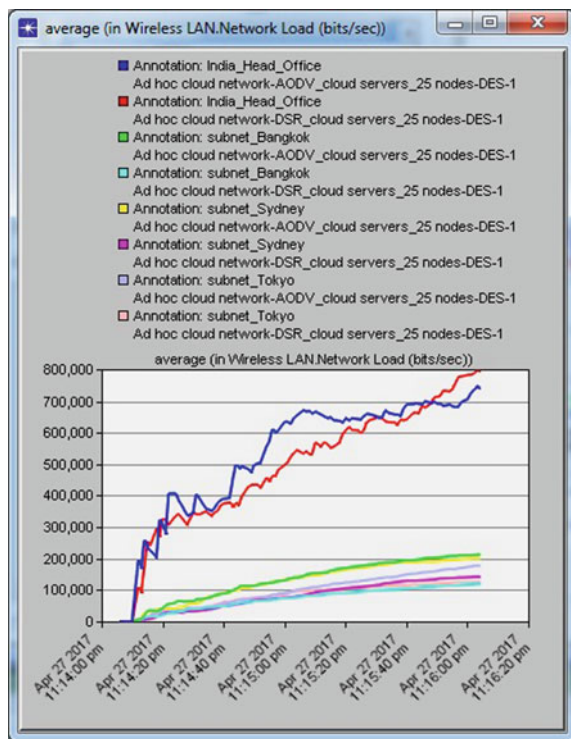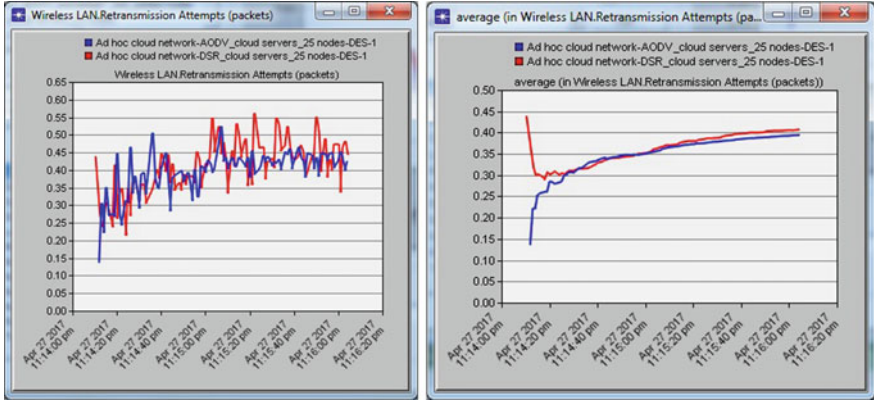
Fig. 38 Wireless LAN retransmission attempt of AODV (in blue) and DSR (in red) (left) and wireless LAN average retransmission attempt of AODV (in blue) and DSR (in red) (right)
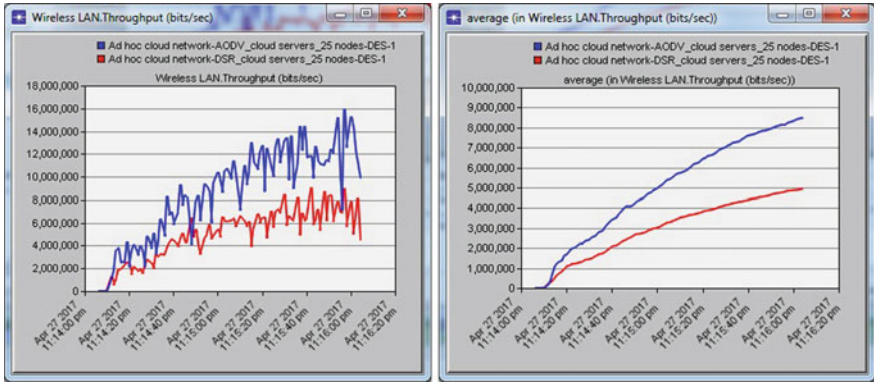
Fig. 39 Wireless LAN throughput of AODV (in blue) and DSR (in red) (left) and wireless LAN average throughput of AODV (in blue) and DSR (in red) (right)

Table 2 Summary of simulation graph and result of wireless LAN statics

Protocol	Average	Maximum	Minimum
Wireless LAN delay (s)			
AODV	0.027447	0.059314	0.012364
DSR	0.041521	0.084931	0.016041
Wireless LAN network load India_Head_office (bits/s)			
AODV	741,907	2,665,293	0
DSR	793,368	2,640,173	0
Wireless LAN network load subnet_Bangkok (bits/s)			
AODV	213,246	470,773	0
DSR	119,223	229,120	0
Wireless LAN network load subnet_Sydney (bits/s)			
AODV	200,546	395,893	0
DSR	141,850	359,680	0
Wireless LAN network load subnet_Tokyo (bits/s)			
AODV	177,363	488,107	0
DSR	125,458	324,027	0
Wireless LAN data dropped (retry threshold exceeded) (bits/s)			
AODV	0	0	0
DSR	2.40	240.00	0.00
Wireless LAN retransmission attempt (packets)			
AODV	0.39407	0.52498	0.13793
DSR	0.40708	0.56155	0.21503
Wireless LAN throughput (bits/s)			
AODV	8,476,900	16,022,560	0
DSR	4,917,678	9,048,920	0
Wireless LAN load (bits/s)			
AODV	666,583	1,798,220	0
DSR	590,000	1,606,167	0
Wireless LAN media access delay (s)			
AODV	0.011533	0.027045	0.004124
DSR	0.017731	0.039090	0.006089

5 Conclusion

In this paper, multi-hop ad hoc cloud network is proposed and implemented for those users who cannot choose cloud due to some reason but however want to take benefit of such technologies. So multi-hop ad hoc cloud network with cloud server using two protocols of the ad hoc network, AODV and DSR, is implemented, and their performance is analyzed. According to the analysis, Database server and Print server perform better using AODV as compared to DSR while for the HTTP server, FTP server, and E-mail server, the performance is enhanced using DSR as compared to AODV. As it is seen in Table 2, AODV performs better than DSR as overall except that in case of wireless LAN network load India_Head_office. So AODV is good choice to implement ad hoc cloud network if the application service is Database or Print; however, DSR is a better choice to implement ad hoc cloud network if the application service is HTTP, FTP, or E-mail. The future scope of this work is to implement other protocols of the ad hoc network in such network with an increased number of nodes and application services and then compare them to identify which protocol is best to implement ad hoc cloud.

References

1. Subbarao, M.W.: Ad Hoc Networking Critical Features and Performance Metrics. Wireless Communications Technology Group, NIST (1999)
2. Singh, V.L.: Performance metrics, parameters and factors of ad hoc, cloud and ad hoc cloud network. Eng. Sci. Technol. Int. J. (ESTIJ) **6**(4) (2016)
3. Kirby, G., Dearle, A., Macdonald, A., Fernandes, A.: An approach to ad hoc cloud computing (2010). arXiv:1002.4738
4. Zia, A., Khan, M.N.A.: Identifying key challenges in performance issues in cloud computing. Int. J. Mod. Edu. Comput. Sci. **4**(10), 59–68 (2012)
5. Altamas, Md.S., Niranjan, P.Y.: A survey of identifying key challenges of performance modeling in cloud computing. Int. J. Comput. Sci. Inf. Technol. Res. **1**(1), 33–41 (2013)
6. Khanghahi, N., Ravanmehr, R.: Cloud computing performance evaluation: issues and challenges. Computer **5**(1), 29–41 (2013)
7. O'Loughlin, J., Gillam, L.: Good performance metrics for cloud service brokers. In: Proceedings of CLOUD COMPUTING 2014, The Fifth International Conference on Cloud Computing, GRIDs, and Virtualization, pp. 64–69 (2014)
8. Yildirim, A.S., Girici, T.: Cloud technology and performance improvement with intserv over diffserv for cloud computing. In: 2014 International Conference on Future Internet of Things and Cloud (FiCloud), pp. 222–229. IEEE (2014)
9. Lacuesta, R., Lloret, J., Sendra, S., Peñalver, L.: Spontaneous ad hoc mobile cloud computing network. Sci. World J. (2014)
10. McGilvary, G.A., Barker, A., Atkinson, M.: Ad hoc cloud computing. In: 2015 IEEE 8th International Conference on Cloud Computing (CLOUD), pp. 1063–1068. IEEE (2015)

Analyzing Performance of Apache Pig and Apache Hive with Hadoop

Krati Bansal, Priyanka Chawla and Pratik Kurle

Abstract Big Data is the term used for huge datasets which are very complex in nature and difficult to be processed using traditional devices. The current requirement is for a new technology for analyzing these huge datasets. One of the best options is Apache Hadoop as it consists of various components which work simultaneously to provide an efficient and robust Hadoop ecosystem. Apache Pig and Apache Hive are core components of Hadoop ecosystem that facilitate specification and search of processing tasks. Apache Hive facilitates to run queries and manage huge datasets using simple commands similar to SQL. Apache Pig is a scripting platform which creates MapReduce programs utilized with Hadoop. In our previous work, we had analyzed and compared both these components to identify benefits and drawbacks on the basis of some parameters. We have showcased analysis of previously conducted research by various researchers. In this paper, we have carried out the analysis by utilizing both these components installed on Hadoop with large dataset as an input.

Keywords Apache Pig · Pig Latin · Big Data · Apache Hive · MapReduce Query

1 Introduction

Currently, one of the big issues in front of companies and individuals is very rapidly increasing data and traditional technologies unable to analyze existing data to pro-

K. Bansal (✉)
Indicosmic capital Pvt. Ltd., Mumbai, India
e-mail: kratibansl1992@gmail.com

P. Chawla
School of Computer Science and Engineering, Lovely Professional University, Phagwara, India
e-mail: priyankachawla.cse@gmail.com

P. Kurle
Mumbai University, Mumbai, Maharashtra, India
e-mail: Pratik.j.kurle@gmail.com

vide the desired information. This age of rapid growth in data can be termed as age of Big Data. Analysis of complex and unstructured data requires new and innovative techniques that can retrieve relevant information in lesser processing time. IT companies are investing a large amount of money in the development of tools that can help in overcoming challenges while analyzing Big Data. Hadoop is an open source framework that is among the top technologies recommended to handle Big Data [1]. Hadoop is a framework utilized to build applications which can run on huge datasets distributed across cluster of computers or servers. Apache Hive and Apache Pig are among the core components of Hadoop ecosystem. Apache Hive is data warehouse infrastructure installed on the topmost layer of Hadoop to provide data analyses, summarization, and query. Apache Pig is a high-level platform which uses Pig Latin as scripting language for creation and execution of MapReduce jobs on Big Data. In the research aspect, both these components can be utilized to analyze Big Data. In our research paper, we have performed an analysis on dataset installed on Hadoop. In this research work, we have focused to analyze and portray benefits and shortcomings of Apache Hive and Apache Pig on basis of processing time and coding lines. In this research study, we have installed huge dataset on Hadoop to run various queries to extract relevant data using Apache Pig and Apache Hive. In last section, we conclude our analysis and discussed about some more techniques that can be worked in future.

2 Theoretical Analysis

Hadoop ecosystem has multiple components having different functions such as data management, data security, data access, and data integration. All the components of Hadoop components have their own importance and have a dedicated role in solving various types of problems [2].

2.1 Apache Pig

Apache Pig is a scripting platform used for transforming and analyzing datasets [1, 3]. YAHOO was creator of Pig and later acquired by Apache foundation in 2007. Apache Pig utilizes procedural and Pig Latin as scripting languages to create job.
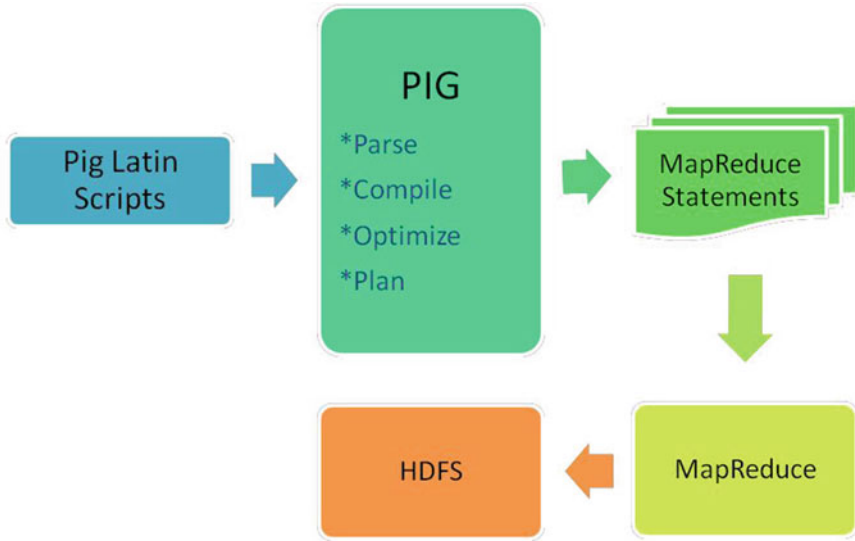
Fig. 1 Pig architecture

A Pig job is a chain of operations processed in pipelines which later automatically get converted into MapReduce job [1]. Pig Latin conceptualizes the programming from MapReduce (Java) expression into a script that transforms MapReduce into an high-level program. Extract Transform Model (ETL) is used by Pig for data extraction from different sources. Pig transforms and stores the data in HDFS. Refer Fig. 1 for Pig Architecture [4].

2.2 Apache Hive

Hive was initially developed by Facebook for providing developers with data ware-house interface for MapReduce programing similar to the traditional data warehouse techniques. Hive Query Language (HiveQL or HQL) statements are similar to stan-dard SQL statements [2].

Architecture: In the Hive architecture, it has three parts in the architecture. In the metadata, it stores information about the stored data. HiveQL is similar to other SQL it uses familiar relational database HiveQL is similar to SQL for querying on schema info on the metastore [5]. It is one of the replacements of traditional approach for MapReduce program. Instead of writing MapReduce program in Java, we can write a query for MapReduce job and process it. HDFS is the Hadoop distributed file system which is stored the data into the CSV Format (Fig. 2).

Hive works as structured query and runs on the batch processing system with high latency. Hive works as summation query analysis and is not able to work on the real
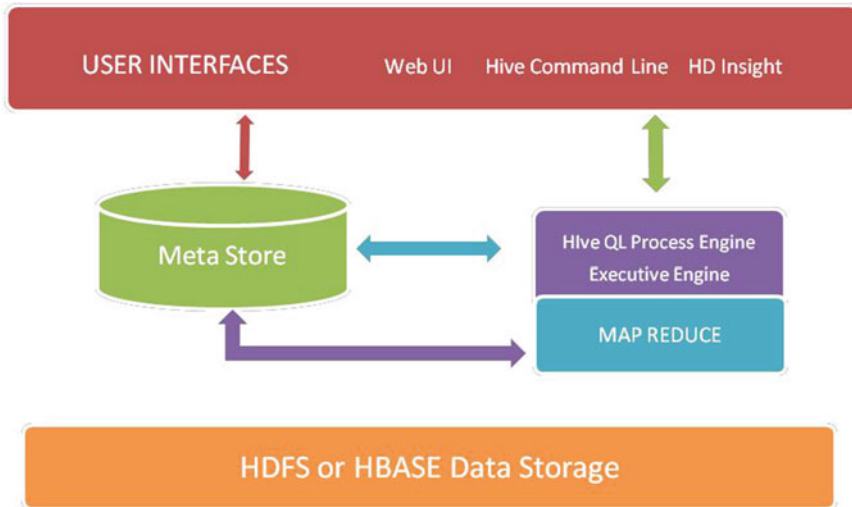
Fig. 2 Hive architecture

time data base query [5]. Hive is not designed for OLTP. Hive gives the best result in batch processing. It also works on key/pair value. Hive QL is similar to SQL but is not SQL. Hive does not support all queries of SQL; it includes multi-table inserts and creates table on the index level. It does not support real-time data like transaction. It supports full ACID properties of the database. The compiler translates HIVE QL statements into DAG graph of MapReduce.

3 Case Study

For analysis, we have used the 125,087 records medical misleading doctor of different states of United State of America Comprising in the below given table, which shows various medical boards in the United States of America publishing disciplinary actions against medical practitioners. These actions are taken by the government in the different years and different medical boards. The main reason is to collect this dataset; it unstructured and different in format, and these datasets have eight tuples and these tuples having different data types. We made datasets and execute the load, arithmetic, group, join, and filter operation on the datasets. It consists of the information of data extraction from a dummy database and virtual graph of data extracted. We extracted the data from various State Medical Professional Boards in United States which gives information regarding the misconduct. The source of information is represented in Table 1.
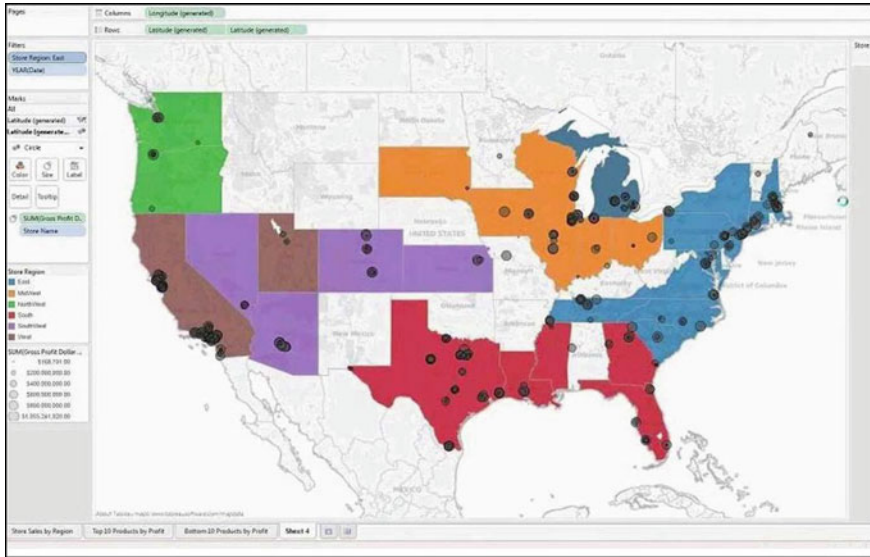
Fig. 3 Analysis on data

We collected 125,873 tuples data which is having the different attributes corresponding to entity name (Fig. 3).

Table 1 Medical Professional Boards in United States [6]

1. California Medical Board	9. Massachusetts Board of Registrant of Medicine	17. Florida Board of Nursing	25. Maryland Board of Physician
2. Ohio Medical Board	10. Ohio Board of Nursing	18. Florida Board of Chiropractic Medical Board	26. Minnesota Board of Medical practice
3. Federation of State Medical Boards	11. South Carolina Board of Medical Education	19. Florida Board of Dentistry	27. New Jersey State Board of Medicine
4. Florida Board of Nursing	12. South Carolina Board of Nursing	20. Florida Board of Medicine	28. North Carolina Medical Board
5. New Jersey Board of Nursing	13. Virginia Board of Nurse Aide	21. Florida Board of Osteopathic Medicine	29. Virginia Board of Optometry
6. New Jersey State Board of Dentists	14. Virginia Board of Dentistry	22. Florida Board of Respiratory Care	30. Virginia Board of Pharmacy
7. North Carolina Board of Nursing	15. Virginia Board of Medicine	23. Alabama Board of Medical Examine	31. Virginia Board of Physical
8. Kentucky Board of Nursing	16. Virginia Board of Nursing	24. Arkansas State of Board	32. Virginia Board of Therapy

In This Phase We Collect the Data Different Source and Performing the Operation

The main processing tasks are as follows:

1. Extraction of data as per the name.
2. Extraction of data as per License name.

Formatting and Cleaning Up the Input Data

Some operations are as follows:
Loading and storing the cleaning (filtering) the data.
Grouping the data.
Performing the operation.
Joining and distinct.
Grouping the data.

Datasets are considered very simple to maintain; however, in the real world, handling so large amount of data is very complex [7]. This section provides a comparative analysis between Pig and Hive after utilizing these components for Big Data analysis. In our analysis, we performed some operations to know about the fact of Hive and Pig.

Hive Query

```
See current tables we used the
Hive> Show Tables;
Crete the database
Hive> Create database;
Check the creation of database
hive> Create database name;
Create table into the database
CREATE TABLE medical_fraudex_sample (
name            string
city_state      struct<city:string,state:string>
license_number     string
entity_type        string
type_of_action     string
medical_board      string
date_of_actions    string
) row format delimited fields terminated by ',' collection items terminated by ':'
Stored as textfile; Load the data into the database
Load data inpath "/kriti/originaldata.csv" into table
medical_fraudex_sample;
```

1. Select distinct(name) from medical_fraudex_sample; Store output to hdfs in a directory.
   Use INSERT statement to populate data into a table from another HIVE table. Since query results are usually large it is best to use an INSERT clause to tell HIVE where to store your query.

2. INSERT OVERWRITE DIRECTORY "/kriti_project" select distinct (name) from medical_fraudex_sample;

3. Select name from medical_fraudex_sample where name LIKE "%Graham Beaumont"; Or

4. Select name from medical_fraudex_sample where license_number = 66828;

Hive>
See current tables we use the
hive>Show Tables;
Crete the database
hive> Create database;
Check the creation of databse
hive> Create database **name**;
Create table into the database
CREATE TABLE medical_fraudex_sample(
name   string
city_state   struct<city:string,state:string>
license_number   string
entity_type   string
type_of_action   string
medical_board   string
date_of_actions   string
) row format delimited fields terminated by ',' collection items terminated by ':' Stored as
textfile; Show table
Hive > show table;
Ok
Med_fraud
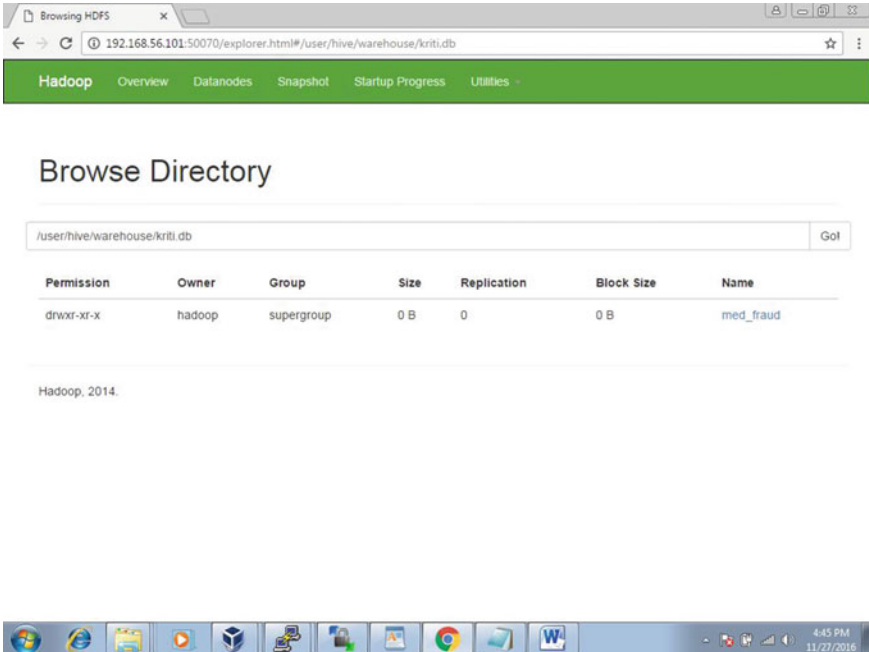
Load data into table

```
Hive> Load data local inpath "/opt/project/data/originaldata.csv" into table
med_fraud; Loading data to table kriti.med_fraud
Table kriti.med_fraud stats: {numFiles=1, totalSize=14406391}
Ok
Hive> select * from med_fraud limit10;
OK
HIVE> describe med_fraud;
name             string
city_state            struct<city:string,state:string>
license_number    string
entity_type       string
type_of_action    string
medical_board     string
date_of_actions   string
) row format delimited fields terminated by ',' collection items terminated by ':'
Stored as textfile; Hive> select distinct (name) from med_fraud;
```

Filter the data as per requirement

PIG Query

Grunt>

Load the database

```
Grunt > A = LOAD '/opt/project/data/original_data.csv' USING PigStorage (',')
AS (name:,city_state, license_number, entity_type, type_of_action, medical_board, date_of_actions);
```

Describe the A aliases

```
Grunt describe A;
A:(name:bytearray,city_state:bytearray,license_number:bytearray,entity_type:bytearray,
type_of_action:bytearray, medical_board:
bytearray,date_of_actions:bytearray)
```

Define the data type corresponded to the tuples

```
Grunt Describe C;
Grunt {C: group: bytearray}
```

Put Screen Shot

```
Grunt describe A;
A:(name:bytearray,city_state:bytearray,license_number:bytearray,entity_type:bytearray,
type_of_action:bytearray, medical_board:
bytearray,date_of_actions:bytearray)
Grunt > L=Limit A 10;
Grunt> dump L;
```

Use the filter to fetch the details as per requirement

```
Grunt describe A;
A:(name:bytearray,city_state:bytearray,license_number:bytearray,entity_type:bytearray,
type_of_action:bytearray, medical_board:
bytearray,date_of_actions:bytearray)
Grunt X = filter A name == „A Armeshi";
Grunt DumpX;
```

It provides all details in the database which pertained to A Armeshi.

3.1 Similarity in the Operation of PIG and Hive

To reduce the task and obtained the result Hive and Pig both are having the JOIN operation [3]. This function reduces the Average time of the CPU.

When the same operations are performed on both PIG and HIVE, the following outcomes are found [8]. PIG gives high speed to join the replicated and datasets.

In the runtime analysis in the group-based query, PIG has more time as compared to the HIVE. Hive in cluster jobs outperformed those produced by the PIG Compiler.

When we saw in the equal terms Hive is remain slower than PIG.

The performance difference between the HIVE and PIG does not scale linearly; when we increase the size of the datasets, Hive consistently slows. Hadoop was designed to run on clusters contains thousand nodes; therefore, running on the small scale data analysis may not give really do it any justice, which is true or gives the true analysis about the HIVE and PIG. Hive and Pig uses abstraction as it hides the complexity of the generated code.

3.2 Compatibility Pig and Hive in the Following Terms

Storage: In terms of storage, Hive is better as compare to Pig. It does not require the structure of table before importing the data; it allows project in tabular format. It consists of two component schemas on read and meta store. The biggest difference is Pig and Hive; Pig does store data in metadata [9].

Data Extraction: Pig is better than Hive for ETL tasks facts; in Hive, we need the tool which performs the ETL tools built their Hadoop integration on the top of Hive [10]. When and who use PIG: A technical sounded person who is familiar with java use Pig Latin, while Hive is used by SQL familiar person [11].

Cost Effective: Both tools are cost-effective when we define the SLA on the basis of cluster size, which is cost-effective and provides the flexibility to use the tools. Hive provides the statistical function. Hive can support the tester who queries and calculates the statistical function. Pig is having the standard function like average, sum, etc.

Time Analysis

See Graphs 1 and 2.
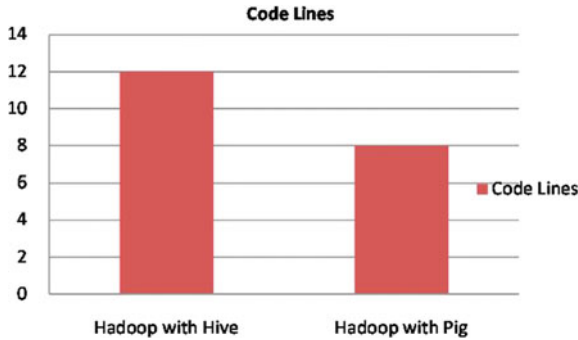
Code Lines

Graph 1 Comparison on code lines between Hadoop with Hive and Hadoop with Pig

Graph 2 Comparison on result time between Hadoop with Hive and Hadoop with Pig

Result Time
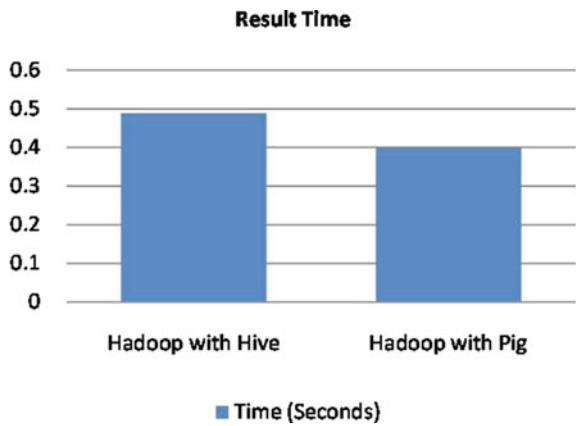
4 Research Findings

Big Data analytics is a big challenge to various organizations. Pig is one of the components of Hadoop which can be used to analyze Big Data. The data processing and analysis of big data are done through Pig and Hive. We have studied and explained principles of programming in the Pig Latin query language and its relation to the Map Reduce Programming model. Hive also gives the same analysis. The current trend in analysis of the real-time data is very difficult, and Hadoop is one of the frameworks which resolve this problem by using Pig, Hive, and another framework. Hadoop consists of multiple frameworks. PIG is the one which is fastest and easy to understand and to analyze the large data. Hive also has its benefits in various results of analysis of Big Data. Hive works on metadata and has the data warehouse. For finding the performance of Pig and Hive, we collected data from the various medical boards of the United States. This data we analysis the different queries, data analysis data the application was designed, implemented and tested, implementation and deployment are presented and result is presented by using graph. Pig is one

the best which provides the abstraction of the MapReduce. It allows writing data manipulation statements and querying in the HLL. It automatically distributes the data and works on the cluster. Hive is also one of the frameworks which is able to work on Hadoop platform, but it has some limitation; it is not able to work on real-time data analysis, and Pig is able to analyse the real-time data. Pig shows the high performance for the query processing; queries are run independently. In the worst case, it shows the out of memory error.

5 Conclusions and Future Work

In this work, we compared Hive and Pig w.r.t performance and complexity when large data is given as an input. On that basis, we tried to find out under what circumstances PIG and HIVE should be used. Experimental results show that Hive performs better and is less complicated than Pig and more convenient to use. In future, we would explore and apply other tools that are used on the top layer of Hadoop architecture. In future, we intend to carry out analysis on the varied sizes of data (ranged from MB to TB).

References

1. Pol, U.R.: Big data analysis: comparison of hadoop mapreduce, pig and hive. Int. J. Innov. Res. Sci. Eng. Technol. **5**(6) (2016)
2. Dave, K., Vania, J.: Survey on big data processing using hadoop component. IJSRD **3**(01) (2015)
3. Nawsher, I., Abaker, I., Hashem, T., Inayat, Z.: Big Data: Survey, Technologies, Opportunities, and Challenges, C Volume (2014)
4. Kumar, S., Goel, E.: Comparative analysis of mapreduce, hive and pig. Int. J. Eng. Sci. **17** (2016)
5. Laxmi Lydia, E., BenSwarup, M.: Analysis of big data through hadoop ecosysetm component like flume, hive, pig and mapreduce. Int. J. Comput. Sci. Eng. **5** (2016)
6. Hansen, M.M., Miron-Shatz, T.: Big Data in Science and Health Care. IMIA and Schattauer Gmbh, IMIA Year Book of Medical Informatics (2014)
7. Stella, C.: Apache pig for data science. In: Proceeding at Linuz Foundation, 9 April 2014
8. Ouaknine, K., Carey, M., Kirkpatrick, S.: The pig mix benchmark on pig, map reduce, and HPCC system. In: IEEE International Congress on Big Data (2015); Ramsingh, J., Bhuvaneswari, V.: An insight on big data analytics using pig script. Int. J. Emer. Trends Technol. Comput. Sci. (IJETTCS) **4**(6), 84–90 (2015)
9. Dhawan, S., Rathee, S.: Big data analytics using hadoop component like hive and pig. Am. Int. J. Res. Sci. Technol. Eng. Math. 88–93 (2013)
10. Mechine, J., Sriama, S.: Large Scale Data Analysis Using Apache Pig, Master Thesis, Tartu (2011)
11. Jakobus, B., McBrien, P.: Pig vs Hive: Benchmarking High Level Query Languages, IBM
12. Jalali, V., Leake, D.: Manual for bear big data ensemble of adaptations for regression version 1.0. General Public License Version 3, 5 Oct 2015
13. EMC2 "Data Lake For Data Science" EMC White Paper, May 2015
14. Kaisher, S., Frank Armour, J., Espinosa, A., Money, W.: Big data: issue and challenges moving forward. In: 46th Hawaii International Conference on System Science (2013)

Characterization of Path Loss for VHF Terrestrial Band in Aizawl, Mizoram (India)

Thaisa Jawhly and **Ramesh Chandra Tiwari**

Abstract This paper reports the field strength measurement and analysis using various propagation models such as the free space propagation model, Egli model, Hata model, Walfisch and Bertoni model, and Perez-Vega and Zamanillo model. The field strength of a 10 KW VHF transmitter operating in the frequency band III (203.25 MHz) has been analyzed using Anritsu SiteMaster with the standard dipole antenna. This study will help identify the propagation mechanism and proper method for finding the attenuation. Variation of signal transmission across the capital will be better understood which will aid in the location planning for future transmitting station or the receiving end.

Keywords Field strength measurement · Propagation losses · Path loss models
Radio propagation

1 Introduction

Maxwell's equations tell us that electromagnetic (EM) waves can self-propagate based on the principle that time-varying magnetic field produces an electric field and time-varying electric field produces a magnetic field. These equations shape our understanding of electromagnetic wave propagation. This is a well-developed theory where the prediction of EM wave transmission was first observed by Heinrich Rudolph Hertz in 1886 [1]. Wireless communication plays an important role in reshaping the way we live. From its preliminary implementation, in the early 1980s [2], mobile communication is one of the fastest growing areas of communication [3]. With this viable accomplishment of cellular communication, there has been a substantial increase in the requirements for planning in radio frequency (RF) trans-

T. Jawhly · R. C. Tiwari (✉)
Department of Physics, Mizoram University, Tanhril, Aizawl 796004, Mizoram, India
e-mail: ramesh_mzu@rediffmail.com

T. Jawhly
e-mail: thaisjly@gmail.com

mission and reception. Since understanding the spatial variability of field strength and optimizing it are a key concept in any wireless system, it is crucial to understand the different system propagation mechanisms through a medium by analyzing its various propagation parameters.

2 Methodology

Variation in the transmitted and received signal, expressed in decibel is termed as path loss. In this study, the field strength or the received power data from the VHF tower will be taken at different points along the north and south radial location from the transmitting tower. The field strength for the different points will be plotted against distance, and this result will be compared with the preexisting path loss models.

2.1 Experimental Details

The transmitter is a 10 KW (Doordarshan) VHF transmitter located at (23°45′58.6″N 92°44′17.5″E) Durtlang, Aizawl, Mizoram. SiteMaster S332E (Anritsu, Japan) and the standard dipole antenna MP534B (Anritsu, Japan) is used to measure the field strength from the transmitting tower which stands 42 m above ground. Data are collected in accessible areas with the dipole antenna at a fixed height of 1.8 m.

2.2 Area Under Study

Study Area: India—Mizoram.
Mizoram, one of the northeastern states of India, is a hilly terrain region with huge forest cover area. Measurements are taken along the road connected areas in the northern and southern region of the transmitting station (Doordarshan, Durtlang).

2.3 Data Collection

Field strength, at an approximate interval of 2 km aerial distance from the transmitting tower (shown by star mark in Fig. 1), is recorded using SS32E (Fig. 2) spectrum analyzer along with the coordinates for each of these locations. In each of these locations, the receiving antenna's orientation is set to receive peak signal available, and the subsequent peak variations of the signal in each spot are recorded a number of times.

Fig. 1 Map showing the study area, the transmitter (VHF antenna) with star points where measurement of field strength is taken
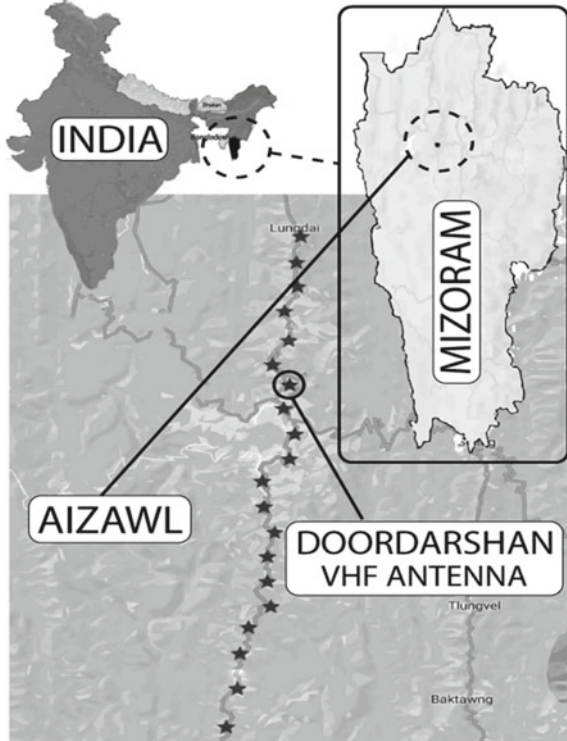
3 Path Loss Models

Path loss model is one of the few estimation choices, which is the step that ensures a more reliable and power efficient transmission providing alongside the predictability of interference and the information needed for the proper assignment of frequencies. Due to the high cost of site measurements, path loss models have been developed as an alternative cost-effective method. Prominent ones showing signs of good predictability within the specified frequency and the area considered will be highlighted.

3.1 Measured Propagation Loss

The power received by the mobile antenna is given by

$$P_r(dBm) = P_u\left(\frac{dBm}{m^2}\right) + 10\log_{10}\left(\frac{\lambda^2}{4\pi}\right) \tag{1}$$
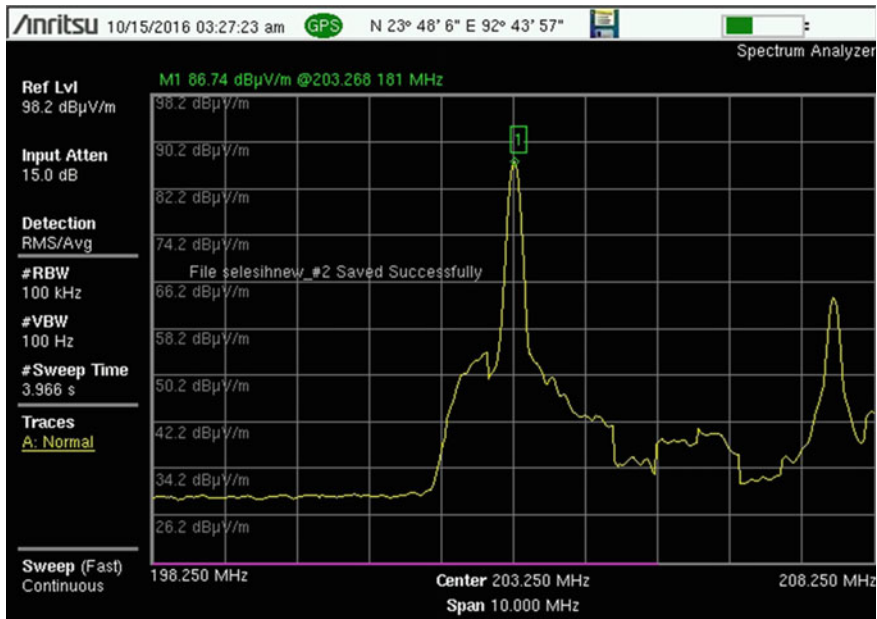
Fig. 2 Picture of the SiteMaster interface showing the field strength and location (GPS) data (courtesy of Doordarshan, Aizawl)

where $\left(\frac{\lambda^2}{4\pi}\right)$ is the absorption cross section of an isotropic antenna, and P_u is the received power density

$$P_u\left(dBm/m^2\right) = E(dB\mu V/m) - 10\log_{10}(120\pi) - 90 \qquad (2)$$

$$P_u\left(dBw/m^2\right) = E(dB\mu V/m) - 10\log_{10}(120\pi) - 120 \qquad (3)$$

Here, the field strength received is measured in $(dB\mu V/m)$ (Fig. 2). If the effective radiated power of an isotropic antenna is P_t, then the propagation path loss can be expressed as the difference in the radiated power and the received power [4]

$$L_p(dB) = P_t - P_r \qquad (4)$$

$$L_p(dB) = P_t(dBw) - E(dB\mu V/m) - 10\log_{10}\left(\frac{\lambda^2}{4\pi}\right) + 145.8 \qquad (5)$$

3.2 Free Space Propagation Model

The free space propagation model [5] is obtained using Friis free space equation. In this equation, the power received by the mobile antenna which is separated at a distance d from the transmitter is given by the relation

$$P_r(d) = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2 L} \tag{6}$$

P_t is the transmitted power, P_r is the received power, G_t is the transmitter antenna gain, G_r is the receiver antenna gain, d is the (transmitter–receiver) separation distance (meter), λ is the wavelength (meter), and L is the system loss factor.

The effective area of the antenna or the aperture of the antenna, A_e, is related to the gain of the receiver antenna by the relation

$$G = \frac{4A_e}{\lambda^2} \tag{7}$$

The path loss for the free space model with antenna gains included is given by

$$PL(dB) = 10 \log \frac{P_t}{P_r} = -10 \log \left[\frac{G_t G_r \lambda^2}{(4\pi)^2 d^2} \right] \tag{8}$$

And when the gains are not included, antennas are assumed to have unity gain and therefore

$$PL(dB) = 10 \log \frac{P_t}{P_r} = -10 \log \left[\frac{\lambda^2}{(4\pi)^2 d^2} \right] \tag{9}$$

3.3 Egli Model

Egli model is a reputed statistical model for predicting the loss in the urban or rural environment [6]. This model arises from wide-ranging measurements made over irregular terrain above 40 MHz till 1000 MHz with a transmitter and receiver separation up to 40 miles (64.37 km) [7]. This model includes the terrain factor which is derived empirically; however, it does not explicitly take into account the diffraction losses [8]. There are different interpretations of Egli's monograph expressed in path loss [6, 8, 9]. One of the such modified models is expressed as

$$
\begin{aligned}
L_E(dB) = & 20 \log_{10}(f_c) + 40 \log_{10}(R) - 20 \log_{10}(h_b) \\
& + \begin{cases} 76.3 - 10 \log_{10}(h_m), & h_m < 10 \\ 85.9 - 20 \log_{10}(h_m), & h_m \geq 10 \end{cases}
\end{aligned} \tag{10}
$$

where the frequency f_c is in megahertz (MHz).

h_b and h_m are the respective base antenna and mobile antenna height in meter (m).

R is the distance between the base station and the receiver in kilometers (km).

3.4 Okumura–Hata Model

Hata developed an empirical mathematical relationship that incorporates the graphical data given in Okumura's method [5] which is applicable in the frequency range of 150–1500 MHz. In this computational model, there are three different terrain-based formulas. Hata model is the commonly used prediction model for TV signal propagation and has shown to be fair in performance [10, 11].

1. For urban areas

$$L_{50}(dB) = 69.55 + 26.16 \log f_c - 13.82 \log h_t - ah_r + (44.9 - 6.55 \log h_t) \log d \tag{11}$$

In small and medium-sized city,

$$ah_r = (1.1 \log f_c - 0.7)h_r - (1.56 \log f_c - 0.8)\, \text{dB} \tag{12}$$

In large city

$$ah_r = 8.29(\log 1.54h_r)^2 - 1.10(dB); \ f_c \leq 200\, \text{MHz} \tag{13}$$

$$ah_r = 3.2(\log 11.75h_r)^2 - 4.97(dB); \ f_c \geq 200\, \text{MHz} \tag{14}$$

2. For suburban area

$$L_{50}(dB) = L_{50}(urban) - 2\{\log_{10}(f_c/28)\}^2 - 5.4\, (\text{dB}) \tag{15}$$

3. For open area

$$L_{50}(dB) = L_{50}(urban) - 4.78(\log_{10} f_c)^2 + 18.33 \log_{10} f_c - 40.94\, (\text{dB}) \tag{16}$$

where h_t is the base station antenna height in meters (ranging from 30 to 200 m), h_r is the mobile antenna height in meters (ranging from 1 to 10 m), d is the T-R separation distance in km and ahr is the correction factor the effective mobile antenna height [4].

3.5 *Walfisch and Bertoni Model*

This model is a diffraction-based theoretical model for predicting the average path loss in an urban environment in the frequency range of 300 MHz–3 GHz (UHF band). This is the first theoretical model that predicts the effect of buildings on the median propagation path loss [12], where the buildings considered are of nearly uniform height [13]. The expression for the excess path loss is given by

$$L_{ex} = 57.1 + \log f_c + A - 18 \log H + 18 \log R - 18 \log \left[1 - \frac{R^2}{17H} \right] \quad (17)$$

The last term accounts for the curvature of the earth. This is neglected (here) in the actual calculation to make the calculation simpler

$$A = 5 \log \left[\left(\frac{d}{2} \right)^2 + (h_b - h_m)^2 \right] - 9 \log d + 20 \log \left\{ \frac{\tan^{-1}[2(h_b - h_m)]}{d} \right\} \quad (18)$$

H is the transmitter height above the height of the building in meter [Walfisch],
f_c is the frequency in MHz,
R is the distance between the transmitting and mobile antenna in km [Walfisch],
h_m is the height of the mobile antenna in feet,
h_b is the height of the building in feet, and
d is the center to center spacing between buildings in meters (30–60 m).

If the antenna gains are neglected and considered unity, then the free space path loss (8) in terms of dB can be written as

$$L_0 = 32.4 + 20 \log f_c + 20 \log R \quad (19)$$

Now, the path loss of Walfisch and Bertoni could be expressed as

$$L_p = L_{ex} + L_0 \quad (20)$$

3.6 *Perez-Vega and Zamanillo Model*

Based on the data from FCC F(50, 50) curve, Perez-Vega and Zamanillo [14] developed a model which allows for the estimation of median path loss, received power or the electrical field strength. This model has shown fair prediction in some studies [15, 16]. In this model, the individual effect of the propagation mechanism like attenuation, reflection, diffraction, etc., is not quantified, instead, this empirical model takes in the effect of all the propagation loss mechanism and embeds them into one single term—the exponent of distance (n).

The isotropic power from the transmitter at a distance d is given by

$$P_{iso} = \frac{P_{rad}}{d^n} \left(\frac{\lambda}{4\pi} \right)^2 \text{W} \qquad (21)$$

P_{iso} is the effective isotropic radiated power, d is the distance between the transmitting and the receiving antenna in meters, λ is the wavelength of the transmitted signal in meters. Path loss in decibel (dB) is expressed as

$$L = 10 \log \left(\frac{P_{rad}}{P_{iso}} \right) \qquad (22)$$

Therefore, from Eq. (21), it can be calculated as

$$L = 10n \log_{10}(d) + L_0(dB) \qquad (23)$$

$$L_0 = 20 \log_{10} \left(\frac{4\pi}{\lambda} \right) \qquad (24)$$

where L_0 is the attenuation at 1 m in free space.

The exponent of distance n for the receiving antenna height of 9 m is given in [14].

4 Analysis of the Observed and Predicted Path Loss

The field strength measurement and its subsequent propagation loss of the transmitted signal from Doordarshan, Durtlang Aizawl is compared with various path loss models like the free space, Okumura–Hata, Egli, Walfisch Bertoni, and Perez-Vega and Zamanillo model and plotted in a Matlab.

From Fig. 3, it is seen that the measured path loss is in close agreement with the path loss provided by Perez-Vega and Zamanillo model and that of the Free space propagation model. While in Fig. 4, it is seen that the measured path loss varies between that of the Perez-Vega and Zamanillo model, Egli model and Okumura–Hata model. From the two plots, it is seen that there is an irregular variation of the measured path loss. This could be attributed to the shadowing effect of the terrain in the location under observation as there is roughly no line-of-sight (NLOS) from the cluster of points where the field strength measurements are made.

4.1 Comparison of the Prediction Model

Root mean square error (RMSE) [16] between the measured and the other prediction models is calculated, where
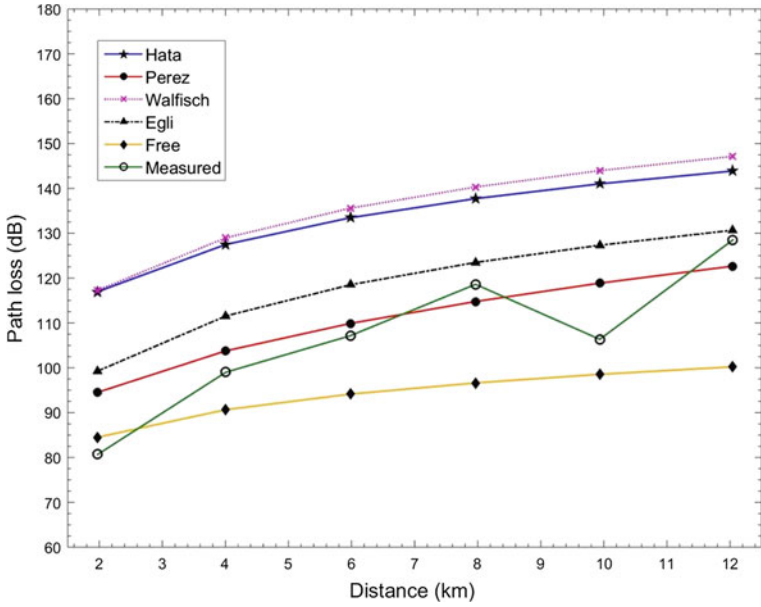
Fig. 3 Plot showing path loss versus distance for the northern part of the transmitter
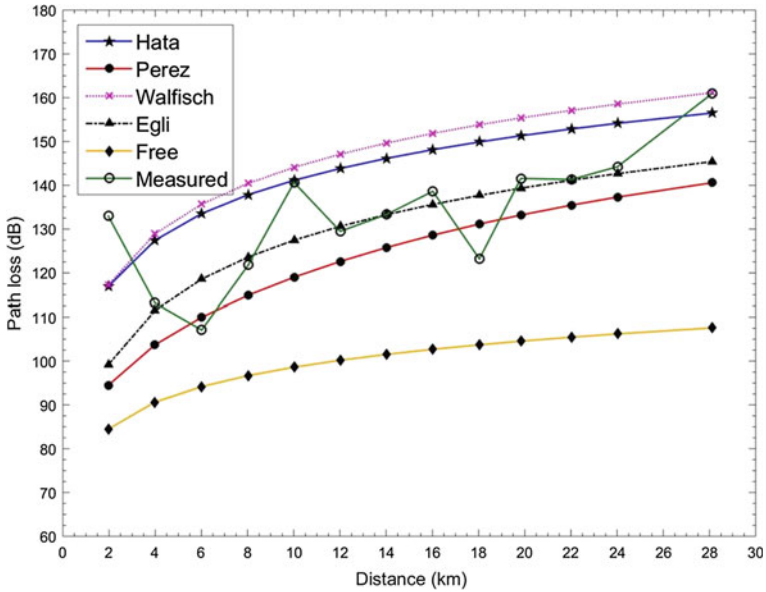
Fig. 4 Plot showing path loss versus distance for the southern part of the transmitter

Table 1 RMSE comparison of path loss models

Location	Free	Hata	Egli	Walfisch-B	Perez-Vega
North	16.325	27.762	13.544	29.699	8.4462
South	34.989	15.034	12.164	17.552	15.011

$$RMSE = \sqrt{\frac{\sum \left(L_m - L_p\right)^2}{N}} \tag{25}$$

L_m is the measured path loss (dB),
L_p is the predicted path loss (dB), and
N is the number of measurements taken.

RMSE comparison of path loss models with the measured data for the northern part shows that Perez-Vega and Zamanillo model gives a good agreement with the measured path loss while for the southern part Egli model gives the most appreciable prediction (Table 1). However, further measurements are required to establish the practicality of these models for the area selected.

5   Conclusion

In this paper, the characterization of path loss model based on measurement of field strength is presented. Path loss models applicable within the VHF frequency band are selected and evaluated and the RMSE evaluation of these models have shown that Perez-Vega and Zamanillo model gives good agreement with the measured path loss for the Northern part while Egli model for the Southern side. As indicated in [10], no single model could be the best fit model in each and every location, and even some reputed model could be inconsistent depending on the location settings.

References

1. Iskander, M.F., Yun, Z.: Propagation prediction models for wireless communication systems. IEEE Trans. Microw. Theory Technol. **50**, 662–673 (2002)
2. Sarkar, T.K., Ji, Z., Kim, K., Medouri, A., Salazar-Palma, M.: A survey of various propagation models for mobile communication. IEEE Antennas Propag. Mag. **45**, 51–82 (2003)
3. Rao, T.R., Rao, S.V.B., Prasad, M.V.S.N., Sain, M., Iqbal, A., Lakshmi, D.R.: Mobile radio propagation path loss studies at VHF/UHF bands in Southern India. IEEE Trans. Broadcast. **46**, 158–164 (2000)

4. Hata, M.: Empirical formula for propagation loss in land mobile radio services. IEEE Trans. Veh. Tech. **29**, 317–325 (1980)
5. Rappaport, Theodore S.: Wireless Communications: Principles and Practice, vol. 2. Prentice Hall, New Jersey (1996)
6. Delisle, G.Y., Lefevre, J.P., Lecours, M., Chouinard, J.Y.: Propagation loss prediction: a comparative study with application to the mobile radio channel. IEEE Trans. Veh. Tech. **34**, 86–96 (1985)
7. Egli, J.J.: Radio propagation above 40 MC over irregular terrain. Proc. IRE. **45**, 1383–1391 (1957)
8. Parsons, J.D.: The mobile radio propagation channel. Wiley (2000)
9. Phillips, C., Sicker, D., Grunwald, D.: A survey of wireless path loss prediction and coverage mapping methods. IEEE Commun. Surv. Tut. **15**, 255–270 (2013)
10. Faruk, N., Ayeni, A., Adediran, Y.A.: On the study of empirical path loss models for accurate prediction of TV signal for secondary users. Prog. Electromag. Res. B **49**, 155–176 (2013)
11. Maurya, G.R., Kokate, P.A., Lokhande, S.K., Shrawankar, J.A.: A review on investigation and assessment of path loss models in urban and rural environment. In: IOP Conference Series: Materials Science and Engineering, vol. 225, no. 1, p. 012219. IOP Publishing, Aug (2017)
12. Walfisch, J., Bertoni, H.L.: A theoretical model of UHF propagation in urban environments. IEEE Trans. Antennas Propag. **36**, 1788–1796 (1988)
13. Adawi, N.S., Bertoni, H.L., Child, J.R., Daniel, W.A., Dettra, J.E., Eckert, R.P., Forrest, R.T.: Coverage prediction for mobile radio systems operating in the 800/900 MHz frequency-range-IEEE-vehicular-technology-society committee on radio propagation. IEEE Trans. Veh. Tech. **37**, 3 (1988)
14. Perez-Vega, C., Zamanillo, J.M.: Path-loss model for broadcasting applications and outdoor communication systems in the VHF and UHF bands. IEEE Trans. Broadcast. **48**, 91–96 (2002)
15. Sridhar, B., Khan, M.Z.A.: RMSE comparison of path loss models for UHF/VHF bands in India. In: Reg. 10 Symps. IEEE, pp. 330–335. IEEE April (2014)
16. Bolli, S., Khan, M.Z.A.: A novel LMMSE based optimized Perez-Vega Zamanillo propagation path loss model in UHF/VHF bands for India. Prog. Electromagn. Res. B **63**, 17–33 (2015)
17. Faruk, N., Ayeni, A.A., Adediran, Y.A., Surajudeen-Bakinde, N.T.: Improved path-loss model for predicting TV coverage for secondary access. Int. J. Wirel. Mob. Comput. **7**(6), 565–576 (2014)

A Comparative Performance Evaluation of Beamforming Techniques for a 2 × 6 Coaxial Cavity Horn Antenna Array for MELISSA

Shweta Vincent ⓘ**, Sharmila Anand John Francis** ⓘ**, Om Prakash Kumar** ⓘ **and Kumudha Raimond** ⓘ

Abstract MELISSA is a GB-SAR system used for monitoring of the Tessina land-slide in Italy. It uses 12 horn antennas for transmission and 12 vivaldi antennas for reception of the transmitted signal. The 12 horn antennas which have been placed in a linear geometry in MELISSA, when placed in a 2 × 6 planar geometry, yield better performance results in terms of directivity and peak side lobe ratio. However, the side lobes in the radiation pattern of the transmitted beam need to be reduced in order to improve the half power beam width. This article describes the simulation results obtained by comparing different windowing techniques used for beamforming of the transmitted waveform and proposes the use of the Dolph-Chebyshev beamforming technique, based on the simulation results. The usage of this technique results in an increase in the half power beam width from 26.98° to 41°, which consecutively increases the area covered by the GB-SAR system for monitoring.

Keywords GB-SAR · MELISSA · Dolph-Chebyshev beamforming · Windowing

1 Introduction

The MIMO Enhanced Linear Short Synthetic Aperture Radar (MELISSA) system is a successfully deployed Ground-Based Synthetic Aperture Radar (GB-SAR) system for monitoring of terrestrial and structural deformation. MELISSA was first used for

S. Vincent (✉) · O. P. Kumar
Manipal Institute of Technology, MAHE, Manipal, India
e-mail: shweta.vincent@manipal.edu

O. P. Kumar
e-mail: omprakash.kumar@manipal.edu

S. A. J. Francis
King Khalid University, Abha, Saudi Arabia
e-mail: sharmilaanand2003@yahoo.co.in

S. Vincent · K. Raimond
Karunya Institute of Technology and Sciences, Coimbatore, India
e-mail: k.raimond@karunya.edu

Fig. 1 MELISSA. (Adapted from [1])
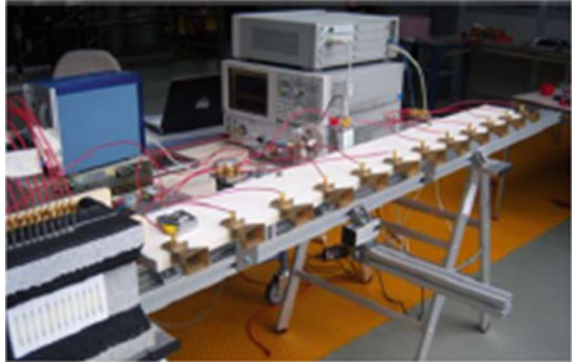
the monitoring of the Tessina landslide of Italy [1] and to monitor the sinking of the luxury cruise ship, the Costa Concordia [2]. The transmitting and receiving antenna subsystems of MELISSA present a huge scope for research and improvement in terms of the type of antenna used and the geometry of the antenna arrangement. Figure 1 illustrates the existing MELISSA system. MELISSA's transmitter and receiver antenna subsystems utilize the MIMO technique to improve the performance of the system, in terms of area coverage and data acquisition time, in comparison to its predecessor, the Linear Synthetic Aperture Radar (LISA [3]). Research has been carried out to improve the characteristics of the beam pattern radiating out of the MELISSA transmitter by using an alternate geometry of antennas placed in a 2×6 planar fashion [4]. The results of the simulation studies show an increase in the directivity and better Peak Side Lobe Ratio (PSLR) of the proposed 2×6 architecture of antenna arrays.

A comparative analysis of different shapes of horn antennas has also been carried out to arrive at the most appropriate shape of horn antenna to replace the pyramidal horn of the MELISSA transmitter sub-system [5]. The result of this simulation study shows that two types of horn antennas, namely, the cantenna and coaxial cavity horn antenna, have a potential to replace the pyramidal horn antenna of the MELISSA transmitter module due to their higher gain and Half Power Beam Width (HPBW) values, when placed in the 2×6 planar configuration [5]. Further, detailed simulations have been carried out to enhance the characteristics of the radiation pattern obtained using the 2×6 cantenna array and 2×6 coaxial cavity horn antenna array [6]. Extensive simulation studies carried out using the Antenna Magus tool suggest that the 2×6 coaxial cavity horn antenna array results in a gain of 10.07 dBi and the highest HPBW value of approximately 140° [6]. By tweaking the design of a single coaxial cavity horn antenna, the overall gain of the 2×6 array has been increased to 10.65 dBi using the Antenna Magus tool [7]. With this high gain transmitter sub-system, a Simulink model of a radar is generated, and the value of the Return Loss (RL) incurred using the 2×6 coaxial cavity horn antenna array is computed [7]. From this study, it is concluded that the 2×6 coaxial cavity horn antenna array results in a high value of return loss which is desirable by any radar system for image acquisition [7].

The radiation pattern obtained using the 2×6 coaxial cavity horn antenna array has side lobes which are undesirable when used in the application of image acquisition. The presence of side lobes results in a decrease in the power transmitted in the main lobe. This article presents a MatLab simulation study of the most appropriate beamforming technique which can be used to suppress the undesired side lobes in the radiation pattern of the 2×6 coaxial cavity horn antenna array and hence increase the HPBW. An increase in the HPBW would result in greater coverage of area [8].

Section 2 presents a comparative simulation study of different types of windowing techniques and the selection of the most suitable technique for our application. Section 3 of this article explains the theory behind the calculation of the Dolph-Chebyshev polynomial weights used for implementing the Dolph-Chebyshev beamforming in the 2×6 coaxial cavity horn antenna array. Section 4 concludes the article.

2 Comparative Analysis of Windowing Used in Beamforming

Beamforming is a principle used for reducing side lobe levels and focusing the transmitted power in the desired direction of the main lobe. This can be achieved by combining the individual elements of an antenna array in a pattern such that they interfere constructively to add up their radiative powers. There are several techniques used for beamforming out of which the most popular is the technique of Windowing. A window function (also known as a tapering function) is a mathematical model of a signal which has a value within an interval and zero value outside the interval. Due to this property of a window function, it can be used to suppress side lobes in undesired directions of a radiation pattern.

This section describes the different types of windowing techniques which can be used for beamforming. The Bartlett Window, also known as the Triangular window is represented in Eq. (1).

$$Bartlett(n) = 1 - \left| \frac{2n - N}{N} \right| \tag{1}$$

where $n = 1, 2, 3, \ldots N$.

Here N is the length of the window, and *Bartlett(n)* is the window function.

The Bartlett-Hanning window is formed by the integration of the Bartlett and Hanning windows. It is represented by Eq. (2)

$$Bartlett\_Hanning(n) = 0.62 - 0.48 \left| \frac{n}{N} - 0.5 \right| + 0.38 \cos\left(2\pi\left(\frac{n}{N} - 0.5\right)\right) \tag{2}$$

The Blackman window is an advanced version the Exact Blackman window and is represented in Eq. (3).

$$Blackman(n) = 0.42 - 0.50 \cos\left(\frac{2\pi n}{N}\right) + 0.08 \cos\left[2\left(\frac{2\pi n}{N}\right)\right] \tag{3}$$

where $n = 0, 1, 2, \ldots N - 1$ and N is the length of the window.

The Blackman-Harris window is yet another variant of the Exact Blackman window and is represented in Eq. (4).

$$Blackman\_Harris(n) = 0.42 - 0.5\cos\left(\frac{2\pi n}{N}\right) + 0.08\cos\left[2\left(\frac{2\pi n}{N}\right)\right] \quad (4)$$

where $n = 0, 1, 2, \ldots N - 1$ and N is the length of the window.

The Bohman window can be obtained by convolving two side lobes of half-length. This window is represented in Eq. (5).

$$Bohman(n) = \left[1 - \frac{|n - N/2|}{N/2}\right]\cos\left[\pi\frac{|n - N/2|}{N/2}\right] + \frac{1}{\pi}\sin\left[\pi\frac{|n - N/2|}{N/2}\right] \quad (5)$$

Here $n = 0, 1, 2, \ldots N - 1$ where N is the length of the window.

The Dolph-Chebyshev window is a special type of window function in which the side lobe level can be specified. Based on this value specified by the user, the side lobes get suppressed. The Dolph-Chebyshev window is shown in Eq. (6).

$$Dolph\_Chebyshev(n) = \frac{1}{N}\left[s + 2\sum_{k=1}^{(N-1/2)} C_{N-1}\left(t_0\cos\left(\frac{k\pi}{N}\right)\right)\cos\left(\frac{(2k\pi(n - (N-1))/2)}{N}\right)\right] \quad (6)$$

where $n = 0, 1, 2, \ldots N - 1$ and N is the length of the window.

The Gaussian window is used for time–frequency analysis as the Fourier transform of the Gaussian window and its derivative are both Gaussian functions. The Gaussian window is represented in Eq. (7).

$$Gaussian(n) = e^{-(n-m)^2/2(\sigma N)^2} \quad (7)$$

for $n = 0, 1, 2, \ldots N - 1$ where N is the length of the window and σ is the standard deviation. The variable m is defined as $m = (N - 1)/2$.

The Hamming window is a modification of the Hanning window and has the shape of a cosine function. Equation (8) shows the Hamming window.

$$Ham(n) = 0.54 - 0.46\cos\frac{2\pi n}{N} \quad (8)$$

where $n = 0, 1, 2, \ldots N - 1$ and N is the length of the window.

The Hanning window has the shape of the half cycle of a cosine wave. Equation (9) describes the Hanning window.

$$Hann(n) = 0.5 - 0.5\cos\frac{2\pi n}{N} \quad (9)$$

where $n = 0, 1, 2, \ldots N - 1$ and N is the length of the window.

The Kaiser window is also known as the Kaiser–Bessel window. The Kaiser window is used in applications where a feedback of the output of a system is provided

to the input in order to fine-tune the performance of the system to the desired value. For larger values of β, the Kaiser window has lower side lobe levels. Equation (10) describes the Kaiser window.

$$Kaiser(n) = \frac{I_0\left(\pi\beta\sqrt{1 - \left(\frac{2n}{N-1} - 1\right)^2}\right)}{I_0(\pi\beta)} \tag{10}$$

where $n = 0, 1, 2, \ldots N - 1$ and N is the length of the window.

The Nuttall is a highly directional window and has a very small HPBW. Equation (11) describes the Nuttall window.

$$Nuttall(n) = 0.355 - 0.487\cos\left(\frac{2\pi n}{N-1}\right) + 0.144\cos\left(\frac{4\pi n}{N-1}\right) - 0.012\cos\left(\frac{6\pi n}{N-1}\right) \tag{11}$$

Here $n = 0, 1, 2, \ldots N - 1$ where N is the length of the window.

The Parzen window is formed by the convolution of four rectangles each of one-fourth length or two triangles each of half a length. Equation (12) describes the Parzen window.

$$Parzen(n) = \begin{cases} 1 - 6\left(\frac{n-N/2}{N/2}\right)^2 + 6\left(\frac{|n-N/2|}{N/2}\right)^3 & \text{for } 0 \leq \left|n - \frac{N}{2}\right| \leq \frac{N}{4} \\ 2\left(1 - \frac{|n-N/2|}{N/2}\right)^3 & \text{for } \frac{N}{4} \leq \left|n - \frac{N}{2}\right| \leq \frac{N}{2} \end{cases} \tag{12}$$

where $n = 0, 1, 2, \ldots N - 1$ and N is the length of the window.

A Rectangular window has a value of 1 over the entire length of the window and applying it is equivalent to not applying any window at all. Equation (13) specifies the Rectangular window.

$$R(n) = 1 \tag{13}$$

The difference between a Taylor and Dolph-Chebyshev window is that unlike the Dolph-Chebyshev window, the Taylor window has monotonically decreasing side lobe levels, whereas in the former, the side lobe levels have the same amplitude.

The Tukey window is a tapered cosine window with a variable α which decides when the window becomes a Rectangular window ($\alpha = 0$) and when it becomes a Hanning window ($\alpha = 1$). Equation (14) describes the Tukey window.

$$Tukey(n) = \begin{cases} \frac{1}{2}\left[1 + \cos\left(\pi\left(\frac{2n}{\alpha(N-1)} - 1\right)\right)\right] & \text{for } 0 \leq n \leq \frac{\alpha(N-1)}{2} \\ \frac{1}{2}\left[1 + \cos\left(\pi\left(\frac{2n}{\alpha(N-1)} - \frac{2}{\alpha} + 1\right)\right)\right] & \text{for } (n-1)\left(1 - \frac{\alpha}{2}\right) \leq n \leq (N-1) \end{cases} \tag{14}$$

A comparative analysis of the various windowing techniques, used in beamforming, has been performed to reduce the side lobes of the radiation pattern obtained, for the proposed 2×6 planar coaxial cavity horn antenna array, to replace the
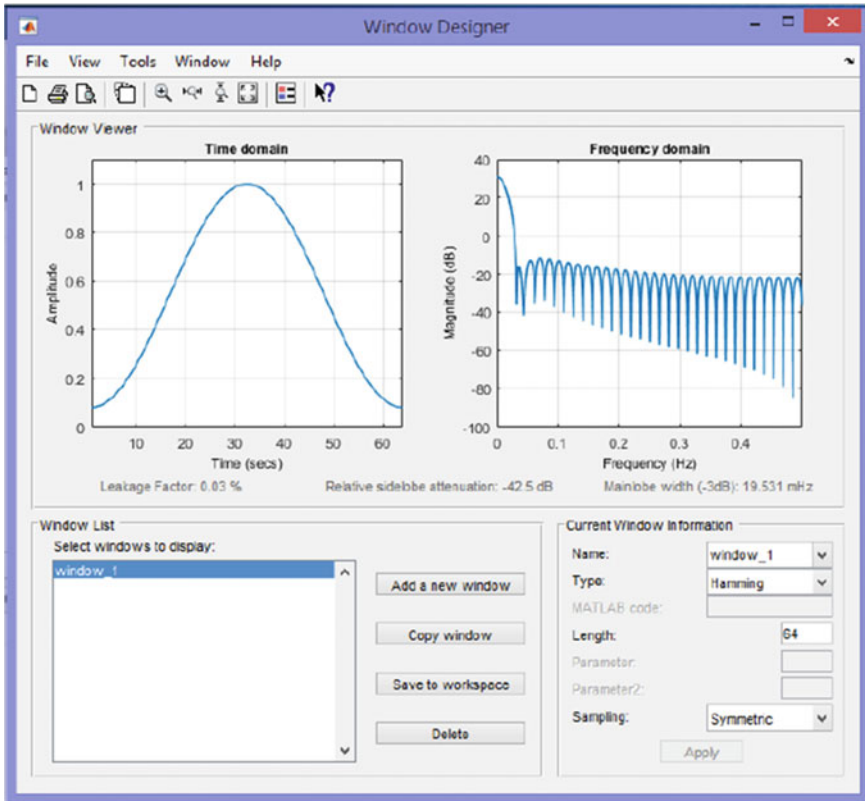
Fig. 2 Window designer tool of MatLab

transmitter module of the MELISSA GB-SAR system. This comparative analysis has been carried out using the Windows Designer tool in MatLAB 2016b version as shown in Fig. 2. The different parameters used for this comparative analysis are described below.

Relative side lobe attenuation: This is defined as the difference in the power level of the main lobe peak to the greatest side lobe peak power level.

Leakage factor: The leakage factor is defined as the ratio of the power in the side lobes to the total window power.

HPBW: HPBW corresponds to the half power beam width of the main lobe. It is computed by computing the width of the main lobe at the 3 dB or half power points. This is done, as the width of the main lobe is largest at the 3 dB or half power points.

Table 1 shows the results of the comparative analysis of the various windowing techniques used for beamforming for a 2×6 planar antenna array configuration.

It has been observed from Table 1, that the Dolph-Chebyshev beamforming technique yields the lowest side lobe level value. This is ideal for the case of a GB-SAR

Table 1 Comparative analysis of windowing techniques used in beamforming

Type of window	Relative side lobe attenuation (dB)	HPBW (dB)	Leakage factor
Bartlett	−26	0.22656	0.29%
Bartlett-Hanning	−35.1	0.25	0.03%
Blackman	−60.3	0.2968	0%
Blackman-Harris	−86.5	0.343	0%
Bohman	−46.1	0.304	0%
Dolph-Chebyshev	−50	0.141	0%
	−100	0.281	0%
Gaussian	−47.8	0.242	0% for $\alpha = 2.5$
Hamming	−37.4	0.2265	0.04%
Hanning	−31.6	0.257	0.05%
Kaiser	−13.5	0.148	8.36% for $\beta = 0.5$
Nuttall	−80.7	0.335	0%
Parzen	−59.4	0.296	0%
Rectangular	−13.1	0.140	9.27%
Taylor	−29.7	0.179	0.37%
Tukey	−15.2	0.203	3.6% for $\alpha = 0.5$

system such as MELISSA, where, a low side lobe level would result in a greater HPBW of the main lobe and in turn yield a higher coverage area.

Another advantage of the Dolph-Chebyshev technique is that the side lobe level can be specified according to the requirement of the system.

3 Application of Dolph-Chebyshev Beamforming to 2 × 6 Coaxial Cavity Horn Antenna Array

The Dolph-Chebyshev beamforming technique was proposed by C. L. Dolph in 1946. He derived a single family of equations for current feeds to the antenna elements of a Broadside array such that, all the side lobes have an equal magnitude and the HPBW to the first null has a minimum value for all the patterns. This set of equations was termed as the Chebyshev polynomials and is shown in Eq. (6),

A weighting method is used to allot weights that cause the signals to get multiplied from the individual antenna elements in an antenna array. The total output radiation field of an antenna array is the field produced by a single antenna element of the array multiplied by a constant termed as the Array Factor (AF) of the array. The AF varies with the change in the geometry of the array and change in the excitation phase of each element of the array.

The array factor for a planar $m \times n$ antenna array is mathematically defined as

$$AF = \sum_{b=0}^{m-1} \sum_{a=0}^{n-1} e^{-j\pi \sin\theta (a\cos\phi + b\sin\phi)} \tag{15}$$

Here, the angles θ and ϕ correspond to the vertical and horizontal scan angles used to describe a point in three-dimensional space in a radiation pattern. The Dolph-Chebyshev beamforming technique has been used extensively for antenna array designs where the side lobe level can be specified. It results in an increased HPBW value and in turn a decrease in the directivity. This trade-off between the HPBW and directivity is termed as a compensation loss or distribution loss (Recommendation ITU-R BT. 1195-1). The amount of compensation loss which is permissible varies from one system to another and should be carefully planned while using the Dolph-Chebyshev beamforming technique.

The Dolph-Chebyshev beamforming technique has been applied to 2×6 coaxial cavity horn antenna array using the Antenna Magus tool. The Array Factor (AF) shown in Eq. (15) is used to customize it to the 2×6 antenna array geometry. The variable a ranges from 0 to 5 corresponding to the 6 elements in the column of the 2×6 planar array and b ranges from 0 to 1 corresponding to the 2 elements in the row of the 2×6 planar array. The resulting array factor for the 2×6 planar antenna array is shown in Eq. (16).

$$AF = \sum_{b=0}^{1} \sum_{a=0}^{5} e^{-j\pi \sin\theta (a\cos\phi + b\sin\phi)} \tag{16}$$

The Dolph-Chebyshev polynomials, which define the weights to be allotted to each of the elements in the 2×6 planar array, are defined as expressed in Eq. (17)

$$T_{m(z)} = \begin{cases} 1 \ for \ m = 0 \\ 1 \ for \ m = 1 \\ 2zT_{m-1}(z) - T_{m-2}(z) \ for \ m = 2, \ 3, \ \ldots \end{cases} \tag{17}$$

These polynomials with known coefficients are used to match them to the array factor equations with unknown coefficients, i.e. the weights of the individual antenna array elements. Therefore, the array factor for the 2×6 antenna array can be expressed as shown in Eq. (18), where x is the number of elements of the antenna array, w_m is the corresponding weight of the mth element, and s is the inter-element spacing of the antenna array.

$$AF = \sum_{m=1}^{x} w_m e^{-jk(2m-1)\frac{s}{2}\cos\theta} + \sum_{m=-x}^{-1} w_m e^{-jk(2m-1)\frac{s}{2}\cos\theta} \tag{18}$$

Using the complex-exponential form of the cosine function, Eq. (18) can be rewritten as Eq. (19).

Table 2 Dolph-Chebyshev weights of 2×6 planar coaxial cavity horn array

Element position	Weight number	Weight value
(0, 0) and (0, 1)	w_1	5679.58
(1, 0) and (1, 1)	w_2	3293.83
(2, 0) and (2, 1)	w_3	1414.82
(3, 0) and (3, 1)	w_4	356.508
(4, 0) and (4, 1)	w_5	41.84
(5, 0) and (5, 1)	w_6	2.024
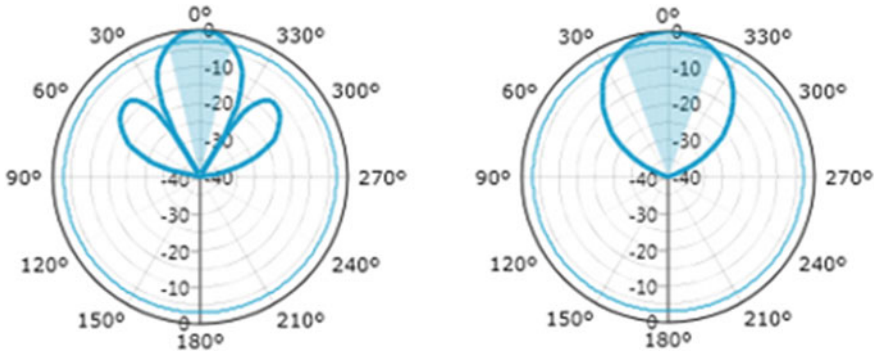
Fig. 3 2×6 coaxial cavity horn planar array, before and after Dolph-Chebyshev side lobe suppression

$$AF = \sum_{m=1}^{x} \cos(2m-1)g \qquad (19)$$

where $g = kscos\frac{\theta}{2}$. Using trigonometric identities for the cosine function, Eq. (19) for the proposed 2×6 planar antenna array can be expanded as shown in Eq. (20).

$$
\begin{aligned}
AF = {} & w_1 \cos g + w_2(4\cos^3 g - 3\cos g) + w_3(16\cos^5 g - 20\cos^3 g + 5\cos g) + w_4(64\cos^7 g \\
& - 112\cos^5 g + 56\cos^3 g - 7\cos g) + w_5(256\cos^9 g - 576\cos^7 g + 432\cos^5 g - 120\cos^3 g + 9\cos g) \\
& + w_6(1024\cos^{11} g - 2816\cos^9 g + 2816\cos^7 g - 1232\cos^5 g + 220\cos^3 g - 11\cos g)
\end{aligned} \qquad (20)
$$

The polynomial coefficients for Eq. (20) were computed for the 2×6 coaxial cavity horn antenna array and the values obtained are shown in Table 2.

The corresponding radiation pattern for the 2×6 coaxial cavity horn planar array after application of the Dolph-Chebyshev beamforming is plotted and illustrated in Fig. 3. The plot shows a rise in the HPBW of radiation pattern from an earlier value of 26.98° to 41°, with a complete suppression of the side lobe levels.

4   Conclusion

Suppression of side lobes in an antenna array's radiation pattern is a requirement in applications which need a high coverage area. Landslide monitoring using GB-SAR systems is one such application. This leads to the need for a suitable beamforming technique which could reduce unwanted side lobes in an antenna's radiation pattern.

This article presented a comparative analysis of various techniques used for beamforming of antenna array radiation patterns. Simulations have been performed in the Antenna Magus tool to carry out Dolph-Chebyshev beamforming for the proposed 2×6 coaxial cavity horn array. From the simulation results, it was observed that the HPBW of the 2×6 coaxial cavity horn antenna array increased from $26.98°$ to $41°$.

References

1. Tarchi, D., Oliveri, F., Sammartino, P.F.: MIMO radar and ground-based SAR imaging systems: equivalent approaches for remote sensing. IEEE Trans. Geosci. Remote Sens. **51**(1) (2013)
2. Broussolle, J., Kyovtorov, V., Basso, M., Guido Ferraro Di Silvi, E.C., Figueiredo Morgado, J., Giuliani, R., Oliveri, F., Francesco Sammartino, P., Tarchi, D.: MELISSA, a new class of ground based InSAR system. An example of application in support to the Costa Concordia emergency. ISPRS J. Photogram. Remote Sens. (2014)
3. Tarchi, D., Leva, D., Casagli, N., Fanti, R., Luzi, G., Pieraccini, M., Pasuto, A., Silvano, S.: Landslide monitoring by using ground-based SAR interferometry. An example of application to the Tessina landslide in Italy. J. Eng. Geol. **68**(1/2), 15–30 (2003)
4. Vincent, S., Francis, S.A.J., Rajsingh, E.B.: An alternate antenna array geometry for a GB-SAR system used in landslide monitoring. J. Ind. Soc. Remote Sens. **43**(4), 761–768. Springer (2015). ISSN: 0255-660X
5. Vincent, S., Francis, S.A.J., Kumar, O.P., Rajsingh, E.B.: A comparative study of horn antennas suitable for the transmitting antenna array module of MELISSA architecture. In: Proceedings of the International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER-2016), 13–14 Aug 2016, pp. 28–34. Surathkal, India, IEEE-Xplore Digital Library (2016)
6. Vincent, S., Francis, S.A.J., Kumar, O.P., Rajsingh, E.B.: Design of a planar antenna array for the transmitting module of MELISSA. Int. J. Appl. Eng. Res. (ISSN: 0973-4562), through the Joint International Conference on Artificial Intelligence and Evolutionary Computing in Intelligent Systems (ICAIECES – 2017) and Power Circuit and Information Technologies (ICPCIT-2017), 27–29 April 2017 **12**(1). Madanapalle, India (2017)
7. Vincent, S., Francis, S.A.J., Kumar, O.P., Raimond, K.: Optimization of gain and return loss of a 2×6 planar coaxial cavity horn antenna array for MELISSA. In: 2nd International Conference on Antenna Innovations & Modern Technologies for Ground, Aircraft and Satellite Applications, (IEEE iAIM 2017) conference, to be held at Bangalore from 24th–26th of Nov (2017)
8. Rodriguez, V.: Half power beamwidth and high power measurements: the dangers of using far field approximations in the near field. In: Proceedings of IEEE International Symposium on Electromagnetic Compatibility, EMC (2008)

The Impacts of Exposure to Low Frequencies in the Human Auditory System—A Methodological Proposal

Juliana Araújo Alves, Lígia Torres Silva and Paula Remoaldo

Abstract The aim of this paper is to evaluate the impact of exposure to low-frequency noise in residential areas, where there are power poles and power lines, in the human auditory system. A methodology to assess discomfort due to the low-frequency noise as well as audiometric tests exclusively for the low frequencies is proposed. Two predominantly urban areas were defined in Northwest Portugal to test the methodology. An "exposed" and "unexposed" study was used; the first group was highly exposed to the source under study and the second had no record of exposure to high voltage lines. To develop the research, a methodology was used to assess the discomfort due to low-frequency noise using audiometric tests (based on ISO-8253-1/2010) to determine the hearing threshold for pure sounds and recorded sound, as well as cognitive tests (Mini-Mental State Examination—MMSE). The average hearing threshold for recorded sound of the eight individuals tested in the "exposed" group was 51.3 dB, ranging from 40 to 65 dB, while the mean of the six subjects in the "unexposed" group was 24.1 dB, ranging from 20 to 30 dB. Based on the results obtained, the "exposed" group seems to be less sensitive to the low frequencies when compared to the "unexposed" group. The methodology used is adequate for a subjective assessment of the discomfort due to low-frequency noise.

Keywords Low-frequency noise · Human auditory system · Hearing threshold
Audiometric test

J. A. Alves · P. Remoaldo
Lab2PT – Landscape, Heritage and Territory Laboratory, University of Minho,
Campus Azurém, 4800-058 Guimarães, Portugal
e-mail: jalves.geografia@gmail.com

P. Remoaldo
e-mail: premoaldo@geografia.uminho.pt

J. A. Alves · L. T. Silva (✉)
Department of Civil Engineering, CTAC – Centre of Territory, Environment and Construction,
University of Minho, Campus Gualtar, 4710-057 Braga, Portugal
e-mail: lsilva@civil.uminho.pt

1   Introduction

Discomfort has been reported as the most frequent effect due to exposure to low-frequency noise in humans [1–4]. Different to "auditory effects", which directly affect the human auditory system, "nonauditory effects" are the most difficult to prove as they result exclusively from exposure to noise. Assessing noise discomfort is generally centered on medium and high frequencies. In addition, the evaluation methodologies adopted in many European countries focus only on objective parameters, while discomfort is essentially a subjective evaluation parameter, which varies from individual to individual. By convention, a frequency A-weighting filter is used in these evaluations because the human auditory system is not very sensitive to low frequencies. However, an A-weighting filter is not suitable for assessing the discomfort of low-frequency noise [1, 5–7]. The application of this weighting filter considerably reduces the noise levels emitted, often contributing to the resulting values remaining below the levels considered as annoying or harmful, when compared to the reference curves used in different countries. Furthermore, the various existing methods consider a limited frequency range, which generally covers the frequency bands between 50 and 200 Hz. Thus, the aim of this article is to propose a methodology to evaluate discomfort due to low-frequency noise, from subjective parameters, using audiometric tests adapted to determine the hearing threshold of the "exposed" and "unexposed" volunteers, as well as the assessment of noise perception by these individuals. To test the methodology, two predominantly urban areas were selected in Northwest Portugal. The "exposed" area was the village of Serzedelo (southwest of the municipality of Guimarães). It was classified as "exposed" because it is in a place with the highest concentration of power poles and power lines, and there is a substation nearby with the highest installed power at the national level (Substation of Riba d'Ave). The "unexposed" area was the village of Abação (São Tomé—located to the southeast of the same municipality) that has no record of high voltage lines. Thus, eight audiometric tests were carried out on the "exposed" population and six on the "unexposed" population so as to determine the hearing threshold for pure sounds and recorded sound and evaluate the perception of discomfort due to low-frequency noise in an environment that was not the natural one.

2   The Impacts of Exposure to Low-Frequency Noise

A large number of studies on health impacts due to occupational and environmental noise exposure can be observed at an international level. The most cited effects on human health refer to emotional changes ranging from agitation and distraction [8–11], as well as associating low-frequency noise with cognitive alterations to developing cardiovascular diseases [12], sleep disorders [13] and arterial hypertension.

In the field of occupational medicine, there is a large body of research that affirms that low-frequency noise is an agent that interferes with the performance of labor

tasks [14, 15]. In addition to these changes, noise can be an agent that affects mental and physical health.

However, there are still few studies focusing exclusively on health impacts and discomfort due to low-frequency noise. One of the main reasons is the low sensitivity of the human auditory system to the low frequencies. On the other hand, this type of noise presents very particular characteristics and causes acute discomfort and long-term "nonauditory effects" [2].

Low-frequency noise sources are of particular concern to the population living close to them, such as wind turbines and areas with power lines. In this case, when the surrounding noise is reduced, such as during nighttime, low frequencies dominate the perceived noise spectrum. Thus, the impact of exposure to this type of noise is particularly worrying in these cases both for the quality of life of the population and the sustainability of the places [2, 4, 16–19].

Exposure to low-frequency noise has harmful effects and is a risk factor for human health. Some authors have treated these effects under the name of vibroacoustic disease [20], vibronoise pathology, i.e., systemic pathology encompassing the entire organism characterized by abnormal proliferation of extracellular matrices and caused by excessive and prolonged exposure to low-frequency noise [21].

In Portugal, the first reference to this type of study refers to the health of workers from the Portuguese Air Force at an aircraft maintenance, repair, and manufacturing plant (*Oficinas Gerais de Manutenção Aeronáutica*—OGMA in Portuguese) [22], which led to defining three clinical stages of vibroacoustic disease: Stage I—Slight (1–4 years)—mild mood swings, indigestion and heartburn, oropharyngeal infections, and bronchitis; Stage II—Moderate (4–10 years)—chest pain, mood changes, low back pain, fatigue, skin infections, inflammation of the gastric surface, pain and blood in the urine, conjunctivitis, and allergic processes; and Stage III—Severe (>10 years)—psychiatric disorders, conjunctiva and nasal and digestive epithelial hemorrhages, varicose veins and hemorrhoids, duodenal ulcers, spastic colon, decreased visual acuity, headaches, intense joint and muscle pain, and neurological changes [21, 22].

Concerning "nonauditory effects", discomfort has been reported as the most frequent effect due to exposure to low-frequency noise in humans [1–3]. The discomfort may vary from individual to individual and depends not only on the recorded sound pressure levels but also on the exposure time and low-frequency components found in the measured sound levels, i.e., noise levels containing low frequencies tend to be more troublesome than noise without such components [1, 6].

Thus, noise pollution cannot be treated merely as a problem of acoustic discomfort. On the contrary, it is currently one of the main environmental pollutants. The World Health Organisation (WHO) considers environmental noise as the second largest form of environmental pollution preceded only by air pollution. From 2000 onward, the WHO started to recognize low-frequency noise as an environmental problem and warned that a large proportion of low-frequency components in noise can greatly increase adverse health effects. In addition, the health impacts generated by the low-frequency components on noise are estimated as being more severe [2].

3 Methodologies for Assessing Discomfort Due to Low-Frequency Noise

Some guidelines for low-frequency noise control in residential areas adopted in some European countries such as Germany and the United Kingdom [22] are based on 1/3 octave analysis of sound pressure levels measured and compared with reference curves. However, in the case of some methods, corrections are applied, such as A-weighting filter. Using this filter leads to a significant reduction in the sound levels emitted, often contributing to resulting values remaining below the levels considered as annoying or harmful, when compared with criterion curves used in various countries (Fig. 1).

It can be observed that the existing methods are based on quantitative parameters to analyze discomfort due to low-frequency noise, while discomfort is essentially a subjective parameter, which can vary from individual to individual. These methods consider a limited frequency range and generally span the frequency range between 25 and 200 Hz.

The current legislation for ambient noise adopted in Portugal, NP ISO 1996 standard, establishes that noise assessment should use frequency A-weighting, with fast integration time and frequency bands of 1/3 octave to vary between 50 and 10,000 Hz [18, 19].

In the study of discomfort due to low-frequency noise, there is a need to lower the minimum threshold of evaluation, i.e., below 50 Hz.

Reference Curve, dB	10	12,5	16	20	25	31,5	40	50	63	80	100	125	160
dB Leq (Polish)	80.4	73.4	66.7	60.5	54.7	49.3	44.6	40.2	36.2	32.5	29.1	26.1	23.4
dB Leq (Swedish)						56	49	43	41.5	40	38	36	34
dB Leq (Dutch)				74	64	55	46	39	33	27	22		
dB Leq (British)	92	87	83	74	64	56	49	43	41.5	40	38	36	34
dB Leq (HTL)				78.5	68.7	59.5	51.1	44	37.5	31.5	26.5	22.1	17.9
dB Leq (S40)					70	61	54	48	46.5	45	43	41	39
dB Leq (S60)					80	71	64	58	56.5	55	53	51	49
dB Leq (S80)					90	81	74	68	66.5	65	63	61	59

Fig. 1 Reference curves used to assess discomfort due to exposure to low-frequency environmental noise, according to the various methods available

4 Audiometric Tests as an Instrument to Analyze the Discomfort Due to Low-Frequency Noise—A Methodological Proposal

Currently, investigating hearing thresholds for high frequencies is significantly important for audiological practices. However, research on hearing thresholds for low frequencies is not often addressed in the scientific literature as it is restricted to investigations of hearing loss for high frequencies (with hearing thresholds affecting one or more frequencies). Nevertheless, the effects of exposure to noise are not only limited to hearing, although it is within this organ that the effects may be more remarkable. Its effects may also interfere with the quality of life of the population, such as sleep quality, concentration, and work performance [4, 7, 23, 24].

The Threshold of hearing or 0 dB SPL (Sound Pressure Level) corresponds to the threshold of each individual for each sound frequency, corresponding to the intensity of the lowest sound oscillation that can be perceived by the human auditory system [4, 7, 23, 24]. Determination of the hearing threshold generally comprises the range of 250 and 8000 Hz [23, 25]. It should be noted that, in the international literature on the subject, there is no audiometric evaluation methodology oriented to the low frequencies and, due to this aspect, the protocols developed were adapted based on ISO 8253-1/2010.

The protocols developed for audiometric tests adjusted to the low frequencies comprised of three stages: determining the hearing threshold, evaluating the discomfort, and carrying out cognitive tests.

To determine the hearing threshold of pure sound, it was reproduced by frequency bands from 10 to 90 Hz with a variation in the signal intensity. The hearing threshold was also determined for the recorded sound from the real sound, with varying signal intensity. After being exposed to the reproduction of pure sounds and recorded sounds, the participants took part in a small survey consisting of three questions.

The subjective and complementary dimension to determine the hearing threshold consisted of conducting a small survey addressing discomfort (using a Likert scale), as well as doing a cognitive test (Mini-Mental State Examination).

The tests were carried out in an audiometric booth Optac—Aumec Horprufkabine brand, and two acoustic stimuli were presented: pure sounds (tracks from a high-fidelity CD of the Nordost brand—System Set-Up and Tuning Disc) and the sounds previously recorded at the place of exposure (in the "exposed" area 5 m from a 400 kV pole). The sound was reproduced inside the audiometric booth through headphones (HE4OOS—Hifiman) with a good response at low frequencies. Connections were made to allow the passage and control of the signal by the AD28—Interacoustics audiometer.

The pure sounds were played by Windows Media Player, and the recorded sound was played by *software dBTrait*.

5   Results and Discussion

In this section, the results and discussion of the audiometric tests will be presented, as well as the evaluation of the perception, for "exposed" and "unexposed" volunteers, of the discomfort due to the noise.

5.1   Determining the Hearing Threshold for Pure Sounds and Recorded Sound

According to the mentioned methodology, audiometric tests were run with eight individuals from the "exposed" group aged 24–68, who had a distinct professional profile (students, various occupations in the textile industry and an administrative assistant). In the "unexposed" group, six individuals aged between 44 and 60, who had a differentiated professional profile (unemployed, drivers, and construction workers), were interviewed.

Concerning pure sounds, both for the "exposed" group (Fig. 2) and for the "unexposed" group (Fig. 3), in all frequency bands, the hearing threshold was evaluated and different intensities were presented, varying from individual to individual (Fig. 4).

The mean hearing threshold for recorded sound of the eight individuals tested in the "exposed" group was 51.3 dB, ranging from 40 to 65 dB, while the mean auditory threshold of the six subjects tested in the "unexposed" group was 24.1 dB, ranging from 20 to 30 dB. The differences recorded for the hearing threshold for pure sounds and recorded sound were significant, showing differences between the hearing threshold means of the "exposed" and "unexposed" groups which was 33.9 dB for the 18 Hz frequency; 41.2 dB for the 21 Hz frequency; 38.9 dB for the 24 Hz frequency; 40.6 dB for the 30 Hz frequency; 16.4 dB for the 39 Hz frequency; 5 dB for the 51 Hz frequency; 2.5 dB for the 60 Hz frequency; -1.3 dB for the 80 Hz frequency; and 1.4 dB for the 90 Hz frequency (Fig. 5). For recorded sound, the difference recorded between the averages corresponds to 27.2 dB.

Fig. 2 Hearing threshold for pure sounds of the "exposed" group

Frequency Band, Hz	18	21	24	30	39	51	60	80	90
Max	100	100	90	90	70	50	40	30	35
Min	75	75	75	70	25	25	20	15	15
Average	85.6	88.8	85.6	80.6	50.6	35.0	30.0	23.8	23.1
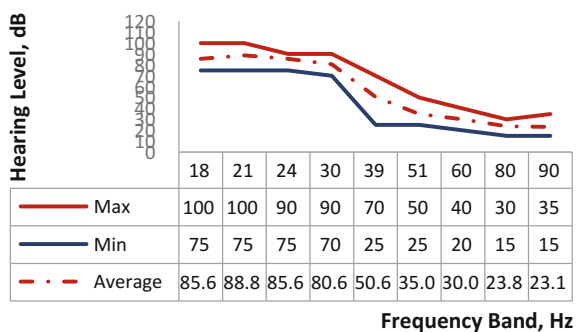
Fig. 3 Hearing threshold for pure sounds of the "unexposed" group

	18	21	24	30	39	51	60	80	90
Max	65	60	60	55	50	45	40	35	30
Min	45	40	40	35	20	25	20	20	15
Average	51.7	47.5	46.7	40.0	34.2	30.0	27.5	25.0	21.7

Frequency Band, Hz
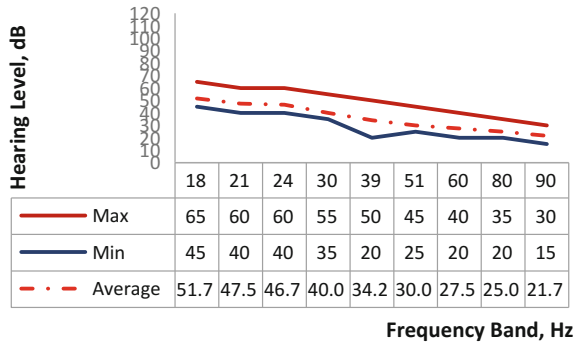
Fig. 4 Average hearing threshold for pure sounds of the "exposed" and "unexposed" groups

	18	21	24	30	39	51	60	80	90
"Exposed"	85.6	88.8	85.6	80.6	50.6	35.0	30.0	23.8	23.1
"Unexposed"	51.7	47.5	46.7	40	34.2	30	27.5	25	21.7

Frequency Band, Hz
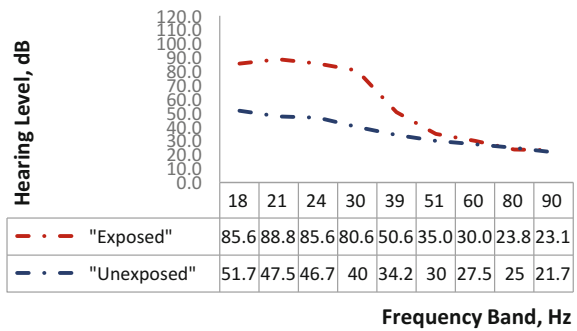
Fig. 5 Hearing threshold for recorded sound of the "exposed" and "unexposed" the groups

	T1	T2	T3	T4	T5	T6	T7	T8
"Exposed"	40	45	50	50	60	50	65	50
"Unexposed"	25	30	25	25	20	20		

Total Frequency (Real Sound)
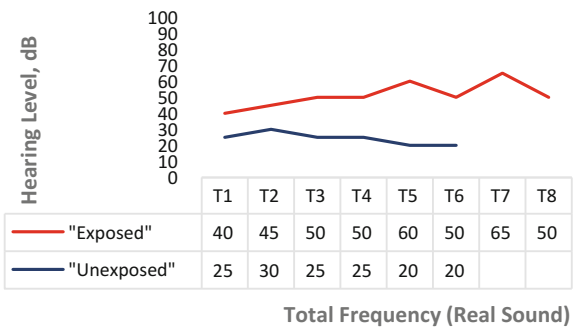
These data reveal that the "exposed" group is not very sensitive to the low frequencies, and corroborates with the results shown in Figs. 2 and 3, which clearly show the low sensitivity of the "exposed" group below 39 Hz, when compared to the "unexposed" group (Fig. 4).

Finally, the tests carried out revealed that the noise collected *in loco* (recorded sound) was audible at sound levels above 40 dB for the "exposed" group and 20 dB for the "unexposed" group.

Table 1 Minimum hearing threshold recorded for pure sounds, by frequency range, for the "exposed" and "unexposed" groups

Frequency (Hz)	18	21	24	30	39	51	60	80	90
"Exposed"	75	75	75	75	25	25	20	15	15
"Unexposed"	45	40	40	35	20	25	20	20	15

The performance of the audiometric tests helped assess the discomfort due to the noise collected and also to demonstrate that pure low-frequency sounds were audible to sound levels above those presented in Table 1 by frequency range.

Finally, the tests carried out revealed that the recorded sound was audible at sound levels above 40 dB for the "exposed" group and 20 dB for the "unexposed" group.

5.2 Evaluating the Perception of Discomfort Due to Noise

The perception of discomfort due to noise was evaluated based on 10 questions using a five-level Likert scale, ranging from "strongly disagree" to "strongly agree".

Some observations are valid to evaluate the perception of discomfort due to noise in the two villages studied:

1. The perception of hearing tended to be higher for the "unexposed" group when compared to the "exposed" group.
2. The perception of day-to-day noise discomfort was similar for the two study groups, i.e., most volunteers responded not feeling discomfort.
3. Only volunteers in the "exposed" group responded that noise has affected their lives. Although four of the eight volunteers from the same group "disagreed" with this statement;
4. Regarding the statement "I awake easily with minimum noise", the group's perception of the "exposed" responses were not unanimous (three "disagreed", one "undecided", and four "agreed" or "strongly agreed"). In the "unexposed" group, most volunteers (n = 4) "agreed" or "strongly agreed".
5. In the statement "I can easily get used to most of the noise", most volunteers from the "unexposed" group (n = 4) "agreed" or "strongly agreed". In the case of the "exposed" group, only three out of the eight volunteers "agreed" with this statement.
6. About the statement "I feel annoyed with the noise", in both groups, most volunteers "agreed" or "strongly agreed" with this statement.
7. Regarding the statement, "I feel discomfort with the noise coming from outside my home", for the "exposed" group, four volunteers "agreed" or "strongly agreed", while in the case of the "unexposed" most volunteers (n = 5) "disagreed" or "strongly disagreed".
8. For the statement, "I have difficulty concentrating in noisy environments", as far as the "unexposed" group is concerned, most volunteers (n = 4) "disagreed"

or "strongly disagreed". In the "unexposed" group, most volunteers (n = 6) "agreed" or "strongly agreed".

9. When presented with the statement "I have difficulty in relaxing in noisy environments", most volunteers (n = 6) from the "exposed" group "agreed" or "strongly agreed". In the case of the "unexposed" group, the perception was divided into three who "agreed" and three who "disagreed" or "strongly disagreed".

10. Finally, regarding the statement "I am aware of the impact of noise on my health", in both groups, most "exposed" (n = 6) and "unexposed" (n = 4) volunteers "agreed" or "strongly agreed" with this statement.

For the "exposed" group, for most participants, the noise reproduced inside the audiometric booth corresponds to a "humming" noise.

Participants described this type of noise as "continuous noise", "humming", "sea waves" and "annoying and irritating". Regarding "discomfort", four participants responded, "not bothered", one respondent felt "sleepiness", and another volunteer felt "annoyed".

After being exposed inside the audiometric booth, the Mini-Mental State Examination (MMSE) was carried out to evaluate the cognitive status of the volunteers.

Only one participant in the "exposed" group scored 22 points, which may reveal a cognitive deficit. One of the participants obtained 25 points, which represents a non-suggestive change in deficit. For the remaining participants, three scored 26 points and the other three participants 27 points, which corresponds to a preserved cognitive function.

Participant 5 (male, 68 years old and with 4 years of schooling), who had a cognitive deficit, had been a textile carder for 44 years and retired 14 years ago. Although this participant revealed that he "strongly agreed" with the statement "I usually hear well" and that the noise had little impact on his quality of life, the participant's hearing threshold for pure sounds was above average for low frequencies. For the real sound recorded, the hearing threshold of this participant was 60 dB.

For the "unexposed" group, participant number 2 (male, 44 years old and 2 years of schooling) scored 17 points, which reveals a cognitive deficit. This participant had worked in civil construction for 5 years and unemployed for 10 years. The hearing threshold was 30 dB.

6 Conclusion

The developed methodology proved to be appropriate for the intended purpose as it evaluated the perception to low-frequency noise and showed the need to extend the minimum interval of frequency to evaluate the impact of this type of noise on the population.

The distance between the sound level meter and the source may interfere with the degree of discomfort. For the presented results, for the real sound, 5 m of projected

distance was measured horizontally between the receiver and the source. The mean hearing threshold for recorded sound was 51.3 dB for the "exposed" group and 24.1 dB for the "unexposed" group. For both groups, the exposure to the sound recorded inside the audiometric booth generated discomfort (in the case of the "exposed" group, one participant reported having a "headache", three were "irritated", one felt "scared", and three felt " sleepiness", while in the "unexposed" group, four participants were "not bothered", one was "annoyed", and one participant felt "sleepiness".

The values recorded were perceived as being uncomfortable by volunteers. This shows the fragility of the existing methods to assess noise discomfort, which are generally based on objective parameters. In addition, the results obtained from the audiometric tests show that the population of Serzedelo ("exposed") may have acquired a "psychological defense". When living with this type of noise on a daily basis (despite classifying it as annoying) it prevented them from perceiving/hearing with the same intensity as the population of Abação (São Tomé, "unexposed"). The results obtained in audibility tests for the pure sounds carried out with the population of the "exposed" and the "nonexposed" corroborate this statement.

References

1. Waye, K.P.: On the effects of environmental low frequency noise. Dissertation Thesis. Gothenburg University, Gothenburg, Sweden (1995)
2. Berglund, B., Hassmén, P., Job, R.F.: Sources and effects of low-frequency noise. J. Acoust. Soc. Am. **99**(5), 2985–3002 (1996)
3. Pawlaczyk-Luszczyńska, M., Szymczak, W., Dudarewicz, A., Sliwiñska-Kowalska, M.: Proposed criteria for assessing low frequency noise annoyance in occupational settings. Int. J. Occup. Med. Environ. Health (2006). https://doi.org/10.2478/v10001-006-0022-9
4. Alves, J.A., Silva, L.T., Remoaldo, P.C.: The influence of Low-frequency noise pollution on the quality of life and place in sustainable cities: a case study from Northern Portugal. Sustainability (2015). https://doi.org/10.3390/su71013920
5. Kjellberg, A., Goldstein, M.: Loudness assessment of band noise of varying bandwidth and spectral shape. An evaluation of various frequency weighting networks. J. Low Freq. Noise Vib. **4**, 12–26 (1985)
6. Leventhall, G.: Low frequency noise and annoyance. Noise Health **6**, 59–72 (2004)
7. Alves, J.A., Silva, L.T., Remoaldo, P.C., Arezes, P., Neto Paiva, F.M.: Proposta Metodológica para Avaliação Audiométrica e da Incomodidade ao Ruído de Baixa Frequência. Acta do 9° Iberian Congress and the 47th Spanish Congress on Acoustics TECNIACUSTICA®, pp. 175–176. Porto, Portugal (2016)

8. Karpova, N.I., Alekseev, S.V., Erokhin, V.N., Kadyskina, E.N., Reutov, O.V.: Early response of the organism to low-frequency acoustical oscillations. Noise Vib. Bull. **11**(65), 100–103 (1970)
9. Brown, J.E., Thompson, R.N., Folk, E.D.: Certain non-auditory physiological responses to noise. Am. Ind. Hyg. Assoc. J. **36**, 285–291 (1975)
10. Job, R.: The role of psychological factors in community reaction to noise. In: Vallet, M. (ed.) Noise as a Public Health Problem, vol. 3, pp. 47–79. INRETS, Arcueil Cedex, France (1993)
11. Pawlaczyk-Luszczyńska, M., Dudarewicz, A., Waszkowska, M., Sliwiñska-Kowalska, M.: Assessment of annoyance from low frequency and broadband noises. Int. J. Occup. Med. Environ. Health **16**(4), 337–343 (2003)
12. Passchier-Vermeer, W., Passchier, W.F.: Noise exposure and public health. Environ. Health Perspect. **108**(1), 123–131 (2000)
13. Ising, H., Kruppa, B.: Health effects caused by noise: evidence in the literature from the past 25 years. Noise Health **6**(22), 5–13 (2004)
14. Waye, K.P., Rylander, R., Benton, S., Leventhall, H.: Effects on performance and work quality due to low frequency ventilation noise. J. Sound Vib. **205**(4), 467–474 (1997)
15. Waye, K.P., Bengtsson, J., Kjellberg, A., Benton, S.: Low frequency noise "pollution" interferes with performance. Noise Health **4**(3), 33–49 (2001)
16. Berglund, B., Berglund, U., Lindvall, T. (eds.): Adverse Effects of Community Noise—Research Needs. Nordic Council of Ministers, Oslo, Norway (1984)
17. Waye, K.P., Björkman, M.: Annoyance due to low frequency noise and the use of the dB(A) scale. J. Sound Vib. **127**, 491–497 (1988)
18. Silva, L.T.: Environmental quality health index for cities. Habit. Int. (2015). https://doi.org/10.1016/j.habitatint.2014.06.020
19. Silva, L.T., Oliveira, I.S., Silva, J.F.: The impact of urban noise on primary schools. Perceptive evaluation and objective assessment. Appl. Acoust. (2016). https://doi.org/10.1016/j.apacoust.2015.12.013
20. Castelo Branco, N., Lopez, E.: The vibroacoustic disease—an emerging pathology. Aviat. Space Environ. Med. **70**(3), A1-6 (1999)
21. Castelo Branco, N., Alves-Pereira, M.: Vibroacoustic disease. Noise Health **6**, 3–20 (2004)
22. Moorhouse, A., Waddington, D., Adams, M.: Procedure for the assessment of low-frequency noise complaints. University of Salford, DEFRA (2011)
23. Arezes, P.M.: Percepção do risco de exposição ocupacional ao ruído. Escola de Engenharia da Universidade do Minho (2002)
24. Arezes, P.M., Barbosa, S., Miguel, A.S.: Noise as a cognitive impairment factor: a case study amongst teachers. In: 20th International Congress on Acoustics, pp. 1–5. ICA (2010)
25. Canalis, R.F., Lambert, P.R.: The Ear—Comprehensive Otology. Lippincott Williams & Wilkins, Philadelphia (2000)

Behavior of Single Pylon of Air Cooled Condenser Support Structure Under Seismic and Wind Forces

Sanish Panchal, Kushang Prajapati and Suhasini M. Kulkarni

Abstract Air cooled condenser (ACC) is the setup of machinery and motors, which is used in thermal power plant to condensate hot steam coming out from turbine. It is an approach to reduce the cooling water demand using direct dry cooling phenomena, which requires no consumptive water and also reduces a power plant water demand by up to 90%. The ACC support structure consists of wind wall at top, truss frames, columns, and bottom raft, respectively. A typical ACC support structure can be of any height depending upon wind and topographic conditions, which is subjected to concentrated load so generally it acts like an inverse pendulum. Due to high sensitivity of this type of structures toward lateral and dynamic loads, it requires meticulous engineering analysis and deliberate sectional design which is optimized too. This paper presents the detailed optimized design criteria for a 6×8 fan air cooled condensers using design moments and axial force from seismic analysis (response spectrum method) and wind load analysis (gust effectiveness factor method) using IS 875:2002 (Part-3). Both analyses are done on a single pylon to find out the governing load using STAADPro.

Keywords Air cooled condenser (ACC) support structure · Seismic analysis
Optimized design · Pylon design · Wind analysis

S. Panchal · K. Prajapati · S. M. Kulkarni (✉)
Department of Civil Engineering, Parul University, Vadodara, Gujarat, India
e-mail: suhasini.kulkarni@paruluniversity.ac.in

S. Panchal
e-mail: sanishraj1994@gmail.com

K. Prajapati
e-mail: kushangprajapati18@gmail.com

1 Introduction

Air cooled condensers (ACC) are the machinery setup placed at a deck type support structure, which are built to greater heights provided with long columns. The primary function of ACC is to condensate the steam (low pressure) coming out from turbine and recycled back to where it was heated at subatmospheric pressure [1]. As the ACC does not require cooling water as in case of water cooled condenser, so it reduces the burden of having power plant [2] near cooling water source and also helps in easy installation at any place. From water scarcity point of view in the power industry, use of Air cooled condenser plays a major prime role to get better efficiency from any individual plant. Air cooled condenser support structure is one of the important components which is to be designed considering different parameters, as loads of deck are large, which are transferred to support structure [3]. So to withstand these loads, support structure is to be designed.

The air cooled condenser support structure in this paper is analyzed [4] and designed for following recommendations:

Height of structure $= 57 + 3$ m (lump mass as cap),
External diameter $= 5$ m,
Diameter of single fan $= 9.75$ m,
Dimension of single fan deck $= (10$ m $\times 10$ m$)$,
Space left for pathway $= 1$ m,
Single opening at base of the column $= (2$ m $\times 1.2$ m$)$,
No of pylons $= 16$, and
Assumed height of truss frames $= 7$ m. Also assumed that "space for duct pipes and accessories" is also assigned with same loading (Fig. 1).

Fig. 1 3D rendering view of model
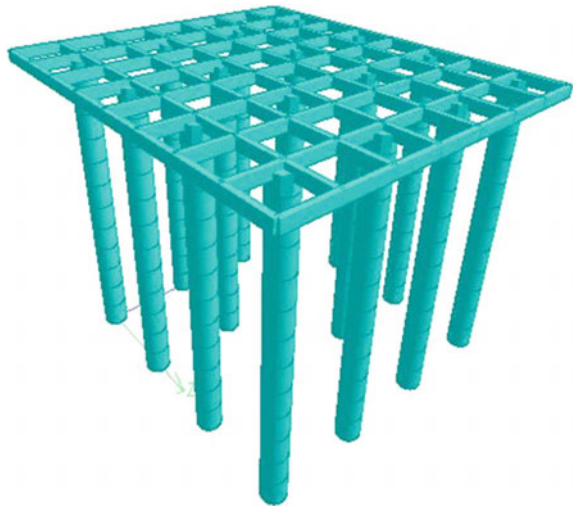
2 Manual Analysis

2.1 Earthquake Load Estimation

Response spectrum method is used for earthquake analysis in which calculation of earthquake loads was carried out by using IS 1893 (Part-4): 2005 [5].

Moment and Shear force Calculation
For the evaluation of moment and shear force, the following are required:

1. The fundamental time period of the free vibration is

$$T = \frac{Ct\sqrt{W \cdot h}}{3.1\sqrt{Ec \cdot Ac}}$$

where

Ct   Coefficient depending on slenderness ratio of the structure.
W   Total weight of structure.
Ac   Area of cross section at base of the structural shell.
Ec   Modulus of elasticity of concrete.
H   Total height of the structure.

2. By using the fundamental time period T, the horizontal seismic coefficient is obtained by

$$\alpha h = \frac{\beta \times I \times \overset{\circ}{F} \times Sa}{g}$$

where

ß   Damping factor.
I   Importance factor.
$\overset{\circ}{F}$ Zone factor.
Sa   Average acceleration (Figs. 2 and 3).

2.2 Wind Load Estimation

For calculation of wind load, according to IS 875 (Part-3):1987 [6], gust effectiveness factor method is used for single self-standing pylon structure. The following values are obtained and calculated using given formulation in the IS code:

Location: Bhuj, Gujarat (Zone-3),
Terrain category: II,
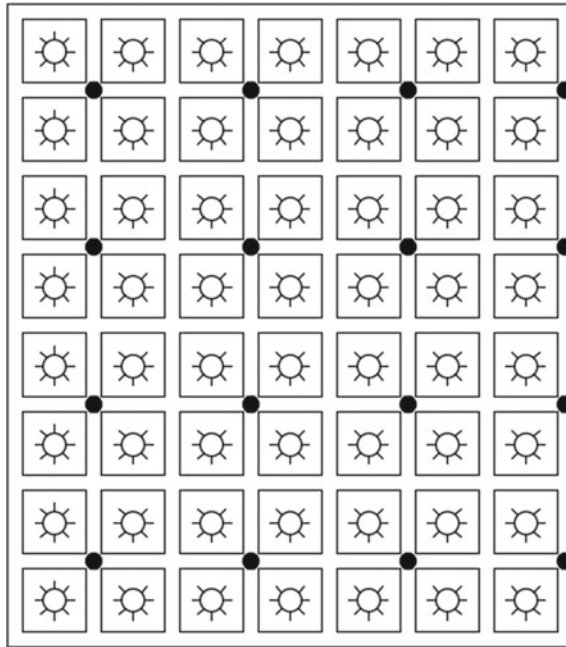
Fig. 2 Top view of support structure
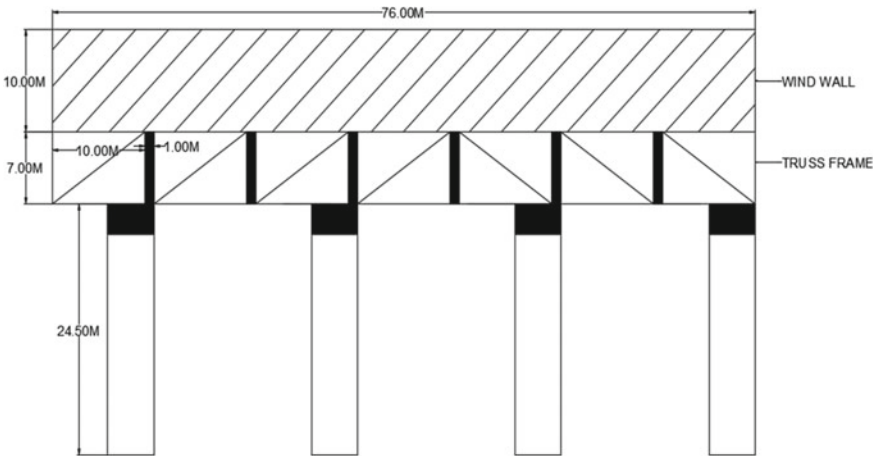
Fig. 3 Side view of support structure

Height of structure: 60 m,
Basic wind speed (Vb): 50 m/s,
Probability factor (K1): 1.08,
Terrain category factor (K2): 0.86, and

Topography factor (K3): 1.0.

Hourly mean wind speed $(\overline{Vz}) = $ Vb × K1 × K2 × K3
$(\overline{Vz}) = 46.65$ m/s.
Along wind load at any height is given by

$$(\overline{Fz}) = Cf \times Ae \times (\overline{Pz}) \times G$$

where

Cf   Force coefficient for building $= 0.92$.
Ae   Effective frontal area $= (5 \times 60) = 300$ m^2.
$(\overline{Pz})$ $0.6 \times (\overline{Vz})^2 = 1300.7$ N/m^2.
G   Gust factor $= 2.072$.

which is obtained by $G = 1 + gf \cdot r\sqrt{[B(1 + \varnothing)^2 + \frac{SE}{\beta}]}$.
where

gf   Peak factor.
r   Roughness factor.
B   Background factor.

    Hence, along wind load at 60 m height is obtained as

$$(\overline{Fz}) = 743.83 \text{ kN}.$$

2.3   Overall Wind Load

As from earthquake analysis and wind load analysis on a single pylon, we found that wind load is governing load for the design criteria. Hence, overall wind load on complete support structure is obtained by IS 875 (Part-3):1987 without considering gust factor (due to whole structure consideration).

Wind load on wind wall

Height of A-frame $= 8.66$ m,
Diameter of steam duct pipe $= 1$ m,
Total height $= 9.66$ m (Assuming it to be 10 m), and
Drag coefficient for hoarding/walls $= 1.24$.

    Wind load on wind wall is given by

$$\begin{aligned}
(Fz) &= Cf \times Ae \times Pz \\
&= 1.24 \times (76 \times 10) \times [0.6 \times (50 \times 1.08 \times 1.19 \times 1.0)] \\
(Fz) &= 2334.89 \text{ kN}
\end{aligned} \tag{1}$$

Wind load on truss frames

Assumed height of truss frames = 7 m,
Assumed solidity ratio = 0.3, and
Drag coefficient = 1.7.

Wind load on first frame is given by Eq. (1), i.e.,

$$(Fz) \; = \; 649.81 \text{ kN}$$

According to IS code cl.6.3.3.4, structure having two or more parallel frames may have a shielding effect on the next frames.

We have shielding factor $(\eta) = 0.9$.
Loading on second frame is given by
= (Wind load on first frame × shielding factor)
= 0.9 × 649.81 kN
= 584.83 kN

Total Wind load on wind wall
As we have considered 16 pylons, so loading on single pylon will be equal to total wind load divided by no. of pylons.

Hence, total wind load = 2334.89 + 649.81 + (584.83 × 8)
= 7668.84 kN.
Load on single pylon = 478.958 kN.
Assumed wind load on single pylon = 480 kN.

3   Software Analysis

In STAAD.Pro, both earthquake analysis and wind load analysis are done using the same dimensions and specifications. The results obtained by software analysis are used in the design procedure (Fig. 4).

4   Design Methodology

For design procedure, working stress methodology is considered. As we require variation of parameters to obtain the optimized design of pylons, doing calculation manually on paper would be deliberate and delaying process. Hence, the formation of spreadsheet from Microsoft Excel is an alternative method to obtain design.

Stepwise Procedure
Grade of concrete = M30,
Grade of steel = Fe500,

Thickness (t) = 500 mm or 0.5 m,
Diameter = 5 m, Radius (r) = 2.5 m,
Axial load (W) = 14835.74 kN (as calculated in STAAD.Pro),
Moment load (M) = 44244.675 kN m,
Modular Ratio (m) = 9.33 (IS 456 recommendation),
Reinforcement % (p) = 1.5% {Assumed},
As our single opening makes an angle of 22.91°,
Which represents 2ß = 22.91°,
So, ß = 11.45°,
Assume ß = 15° (due to minimum ß value chart is for 15°),
And μ = 0° (no another opening is there except one)
Eccentricity (e) = M/W = 2.98.
{e/r} ratio = 1.193.
Steel ratio parameter (ηp) is given by

$$(\eta p) = \frac{[\text{Modular ratio} \times (\% \text{ of reinforcement})]}{[1 - (\% \text{ of reinforcement})]}$$

$$(\eta p) = 0.142$$

Values of C and Ø are obtained from charts given in reference book [4].

$$C = 0.82 \text{ and } Ø = 88°$$

Concrete stress at center of the shell is given by

$$Fc = \frac{C \times W}{r \times t \times (1 - p)}$$

	Beam	L/C	Node	Fx kN	Fy kN	Fz kN	Mx kN-m	My kN-m	Mz kN-m
Max Fx	40	3 DL+WL	43	14835.740	988.994	-3.006	-0.002	59.627	43987.773
Min Fx	1	2 WL	1	-175.205	991.228	4.586	0.000	-90.923	44031.974
Max Fy	131	2 WL	141	-56.244	1002.221	-0.162	0.000	3.348	44244.675
Min Fy	131	5 DL+RSX	142	13940.595	-695.137	-4.060	-0.000	-41.299	-32237.993
Max Fz	1	4 RSX	1	149.984	687.047	7.431	0.000	126.742	35543.654
Min Fz	1	5 DL+RSX	2	13840.212	-687.047	-7.431	-0.000	-89.968	-32108.608
Max Mx	315	2 WL	263	-2.947	16.855	0.124	1284.408	-1.235	168.553
Min Mx	270	2 WL	41	-2.947	-16.855	-0.124	-1284.408	0.000	0.000
Max My	13	2 WL	13	-175.205	474.977	4.586	0.000	184.264	-2116.869
Min My	169	2 WL	181	-175.205	474.977	-4.586	-0.000	-184.261	-2116.869
Max Mz	131	2 WL	141	-56.244	1002.221	-0.162	0.000	3.348	44244.675
Min Mz	131	5 DL+RSX	142	13940.595	-695.137	-4.060	-0.000	-41.299	-32237.993

Fig. 4 Results obtained by software analysis

We have Fc = 9880 kN/m^2 or Fc = 9.88 N/mm^2 (SAFE condition is satisfied)
Maximum steel stress is given by

$$Fs = \frac{S \times m \times W}{r \times t \times (1 - p)}$$

We have Fs = 102469 kN/m^2 or Fs = 102.47 N/mm^2 (SAFE condition is satisfied)

Building the Spreadsheet
A large set of parameters, some of which are interrelated, is involved in the design procedure. In order to make easy the interaction between the designer and the spreadsheet, the cell values which requires an input by the user are highlighted in orange color while the ones which will be automatically generated are highlighted in red color. Calculation is highlighted using blue color. Using Excel spreadsheet, we optimized this structure by thickness variation and % of reinforcement variation using height ranges.

Optimization of support structure
For the optimization procedure, we created three different cases by creating height ranges from

1. 0–20 m,
2. 20–40 m, and
3. 40–60 m.

The axial force and moment at this stage are calculated for three different positioned columns.

1. Outside columns,
2. Inside columns, and
3. Extreme side columns.

Case: (1) Outside Columns {M30, Fe500}

[Dead Load and Wind Load Governing]
If Fc < 10, then design method shows "PASS"
If Fs < 230, then design method shows "PASS" (Figs. 5 and 6a).

Case: (2) Inside Columns {M30, Fe500}

[Dead Load and Wind Load Governing]
If Fc < 10, then design method shows "PASS"
If Fs < 230, then design method shows "PASS" (Table 1, Fig. 6b).

Case: (3) Extreme side Columns {M30, Fe500}

[Dead Load and Wind Load Governing]
If Fc < 10, then design method shows "PASS"
If Fs < 230, then design method shows "PASS" (Tables 2 and 3, Fig. 6c).

Fig. 5 Different height ranges
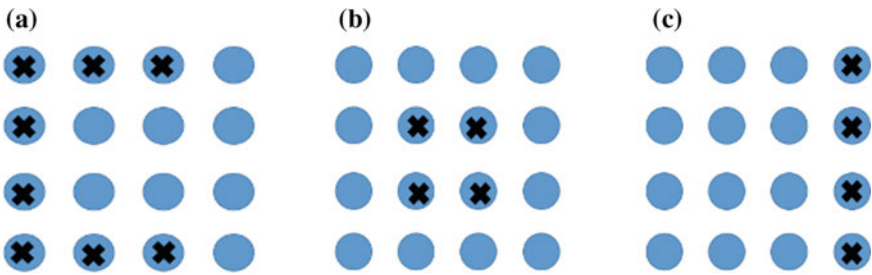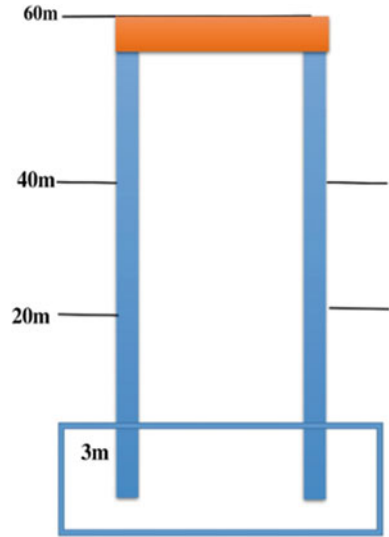
Fig. 6 **a** Outside columns. **b** Inside columns. **c** Extreme side columns

Table 1 Results for case: 3

Height range (m)	Moment (kN m)	Axial load (kN)	Concrete stress (Fc) (N/mm^2)	Steel stress (Fc) (N/mm^2)	Thickness (mm)	% of rein-forcement (%)
0–20	41020.78	14427.14	8.91	103.63	500	1.5
20–40	22507.96	11703.18	5.88	31.16	450	1
40–60	7533.94	8979.21	4.13	15.17	350	0.6

5 Results

By creating these three cases, we found that optimized design for 500 mm thickness and 5 m diameter can also be obtained by varying thickness and percentage of reinforcement by different stage heights.

Table 2 Results for case: 2

Height range (m)	Moment (kN m)	Axial load (kN)	Concrete stress (Fc) (N/mm^2)	Steel stress (Fc) (N/mm^2)	Thickness (mm)	% of reinforcement (%)
0–20	41237.98	14312.91	9.30	90.27	500	1.5
20–40	22460.65	11588.95	5.41	28.65	450	1
40–60	7221.82	8864.98	4.08	14.97	350	0.6

Table 3 Results for case: 3

Height range (m)	Moment (kN m)	Axial load (kN)	Concrete stress (Fc) (N/mm^2)	Steel stress (Fc) (N/mm^2)	Thickness (mm)	% of reinforcement (%)
0–20	41020.78	14427.14	8.91	103.63	500	1.5
20–40	22507.96	11703.18	5.88	31.16	450	1
40–60	7533.94	8979.21	4.13	15.17	350	0.6

Table 4 Optimized design results

Height range (m)	Thickness (mm)	% of reinforcement
0–20	500	1.5
20–40	450	1.0
40–60	350	0.6

Hence, from safety and reliability point of view, the optimized design for given axial load and moment is tabled (Table 4).

6 Conclusions

An optimized design with satisfactory project requirements is always advantageous from the economical point of view. With respect to the above optimized design, here are some more conclusions which are also beneficial.

1. The height of the fan deck is one of the most important factors that need to be selected with respect to wind conditions in order to have a maximum air cooled condenser efficiency.
2. The increment in fan deck height is not always favorable.
3. The earthquake analysis shows effective results which prove its importance to design the structure.

References

1. Cao, S., Wang, J., Cao, J., Zhao, L., Chen, X.: Experimental study of wind pressure acting on a cooling tower exposed to stationary tornado-like vortices. J. Wind Eng. Ind. Aerodyn. (2015)
2. Butler, J.C., Grimes, R.: The effect of wind on the optimal design and performance of a modular air-cooled condenser for a concentrated solar power plant. J. Energy (Elsevier) (2014)
3. Baiju, A., Geethu, S.: Analysis of Tall RC Chimney as per Indian Standard Code. Int. J. Sci. Res. **5**(9) (2016)
4. Manohar, S.N.: Design of Tall Chimneys
5. IS 1893 (Part 4): 2005 Criteria for earthquake resistant design
6. IS 875 (Part 3): 1987, Code of Practice for Design Loads (Other than Earthquake) for Buildings and Structures

Rice Moisture Detection Based on Oven Drying Technique Using Microstrip Ring Sensor

Sweety Jain, Pankaj Kumar Mishra and Vandana Vikas Thakare

Abstract A new microstrip ring moisture sensor is proposed for determining the moisture of rice grains using the oven drying method at multiple frequencies, i.e., 3.7, 5.8, 6.9, 8.4 GHz, with low insertion loss as well as good return loss -18, -17, -22, -27 dB. The proposed sensors parameters of ground length and width are 30 mm and 25 mm, as well as the outer and inner radii are 8 mm and 4 mm. The design is simulated by the CST software, fabricated on the FR4 substrate, which is cheaper and measured by the vector network analyzer (Model No. Field fox N9925A).

Keywords Microstrip ring sensor · Vector network analyzer · Oven drying technique · Rice moisture detection · Reflection coefficient

1 Introduction

Rice is the main source of nutrition for human beings. Moisture content inside rice grain is a crucial parameter for grain processing, i.e., harvesting storage, quality control, and transportation [1–4]. Mostly, rice grain is usually harvested between 19 and 25% MC for maximum grain yields and needs to be dried to 14% or less, depending on the season and the weather, for safe storage [5–7]. Also, the ideal moisture content for milling is 14% in order to maximize the head rice. So, moisture

S. Jain (✉)
Department of Electronics, ASET, Amity University, Gwalior, Madhya Pradesh, India
e-mail: 1502sweety@gmail.com

P. K. Mishra
Department of Applied Physics, Amity School of Pure and Applied Science, Amity University, Gwalior, Madhya Pradesh, India
e-mail: pmishra@gwa.amity.edu

V. V. Thakare
Department of Electronics, Madhav Institute of Technology and Science, Gwalior, Madhya Pradesh, India
e-mail: vandana@mitsgwalior.in

99

Fig. 1 Proposed microstrip
ring sensor configuration
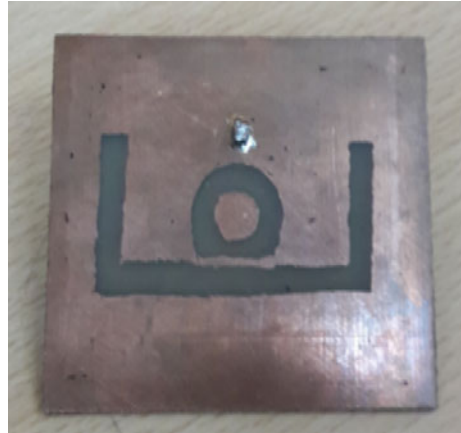
determination of rice is very important [8–11]. Water contained in rice is called rice moisture. The water is held within the rice pores. Rice water is the major component of the rice in relation to human beings growth [12, 13]. Hence, rice moisture content has a quite significant influence on engineering, industrial quality monitoring, and hydrological behavior [14–19] of the rice mass. Furthermore, it has a major role to play as far as the human being growth, the determination of rice water content is of the vital importance efforts to improve growth and water efficiency in agriculture.

In the agriculture sector, the application of adequate and timely moisture for irrigation, depending upon the rice moisture, storage place, and essential percentage of moisture in rice grains [20–22].

Earlier researchers have developed several techniques for measuring the rice moisture such as time-domain reflectometry, frequency-domain reflectometry, tensiometers, oven drying method, etc. Oven drying method is easy and cheaper method which is very helpful to measure the moisture of rice grains. Oven drying technique is cheap, less time-consuming, as well as accuracy [23, 24–26]. The proposed microstrip ring sensor configuration is shown in Fig. 1.

2   Sensor Configuration and Description

The proposed microstrip ring moisture sensor is designed at different frequencies simulating by CST Software and measuring by the vector network analyzer. The parameters of length and width of sensors are 30 and 25 mm as well as the outer radius and inner radius are 8 and 4 mm, and other parameters are shown in Table 1.

Table 1 Parameters and dimensions of microstrip ring sensor

Parameters	Dimensions
ε_r	4.4
h	0.038
$\tan\delta$	0.001

2.1 Simulation Results

The proposed microstrip ring moisture sensor is designed with the help of CST software and get the good return loss as at different frequencies as well as analyzed all the parameters which is useful for the microstrip ring sensor such as return loss, magnitude, phase, gain surface current, smith chart, axial ratio, etc., as shown in Figs. 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12.
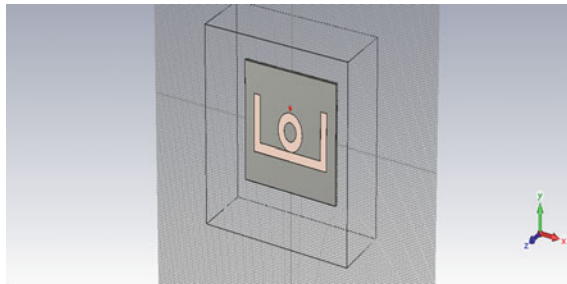
Fig. 2 Structure of microstrip ring sensor
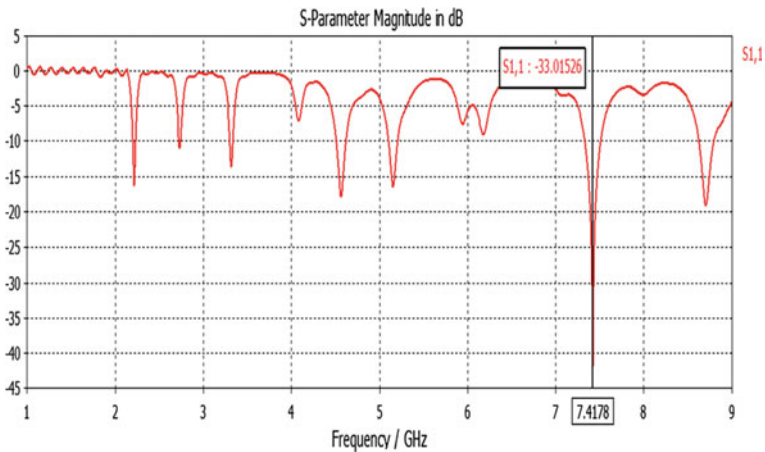
Fig. 3 Return loss of microstrip ring sensor
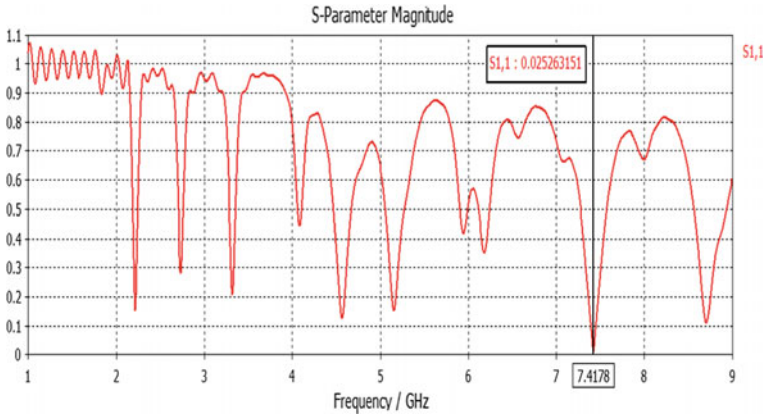
Fig. 4 Magnitude of microstrip ring sensor
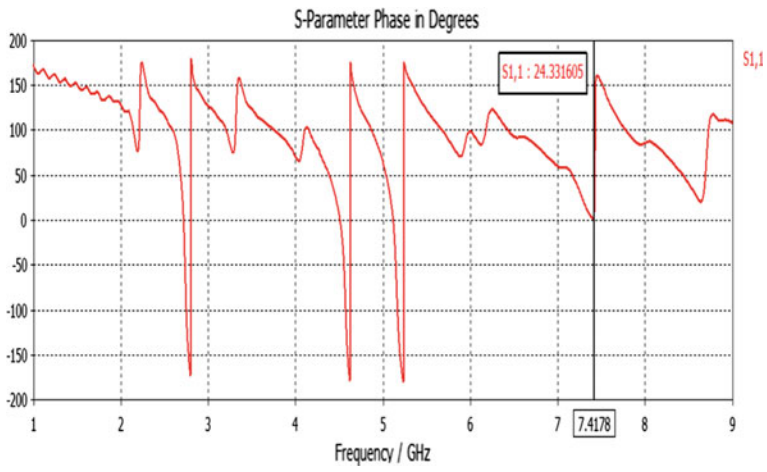
Fig. 5 Phase of microstrip ring sensor
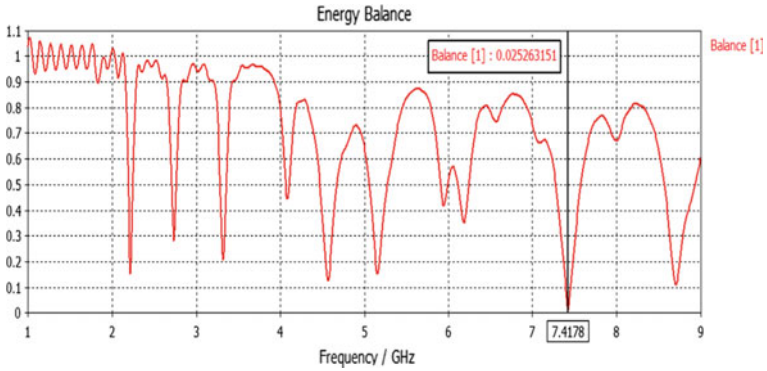
Fig. 6 Energy balance of microstrip ring sensor
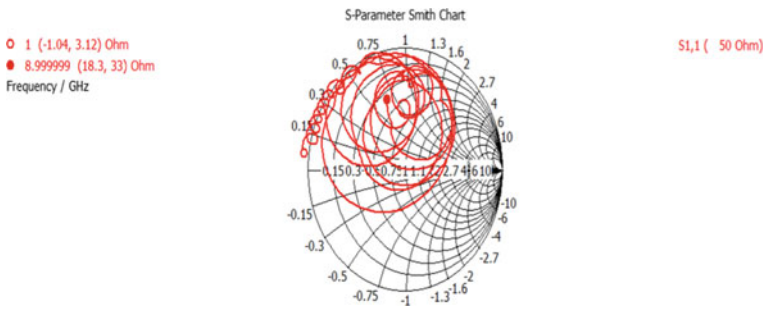
Fig. 7 Smith chart of microstrip ring sensor
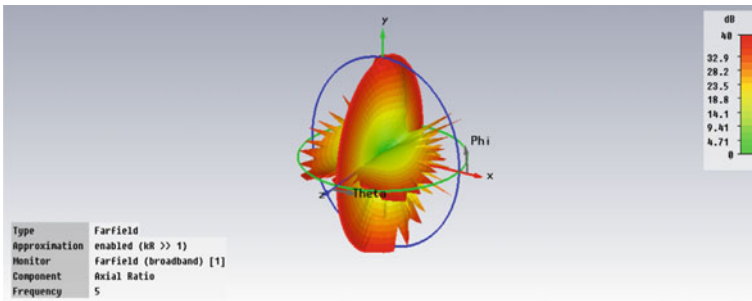
Fig. 8 Axial ratio of microstrip ring sensor
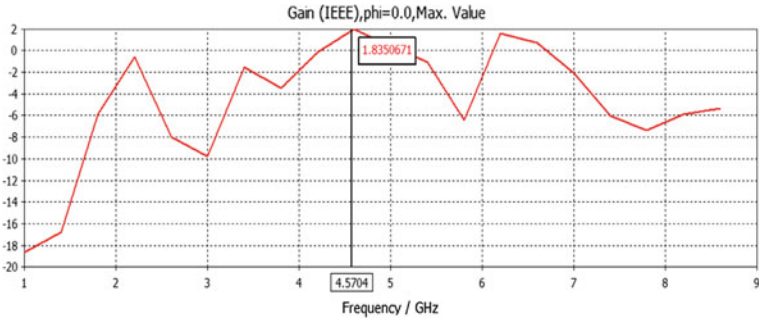
Fig. 9 Gain of microstrip ring sensor
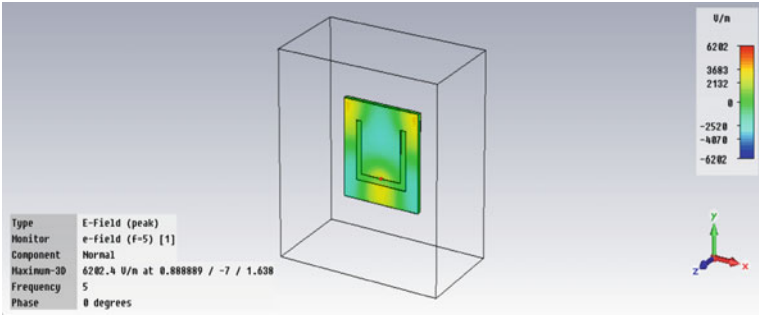
Fig. 10 E-field of microstrip ring sensor
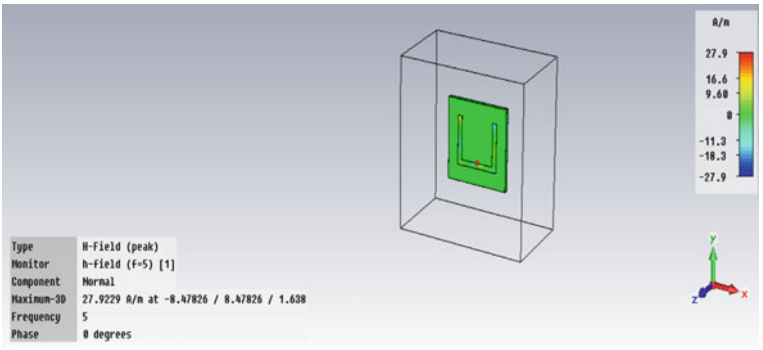
Fig. 11 H-field of microstrip ring sensor
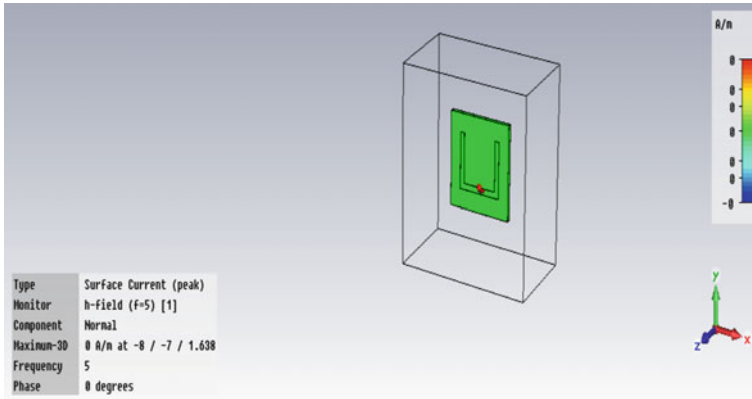
Fig. 12 Surface current of microstrip ring sensor

2.2 Experimental Results

The proposed microstrip ring moisture sensor is measured with the help of vector network analyzer at different frequencies. The dry rice measured the moisture at 3.7, 5.8, 6.9, and 8.4 GHz with return loss −18, −17, −22, −27 dB, after measure the moisture of wet rice at same frequencies 3.7, 5.8, 6.9, and 8.4 GHz with return loss −17, −16, −16, −24 dB as shown in Table 2. The comparison of the measured moisture content at different frequencies, the dry rice increased the return loss but when measured the wet rice of moisture then return loss id decreased after calibration of dry and wet rice then the percentage of moisture can be determined by this formula

Table 2 Summary of dry and wet rice measured the moisture content by vector network analyzer

Frequency	Dry rice	Wet rice
1	−0.66	−0.77
2	−1.65	−3.17
3	−11	−4.6
3.7	−18	−17
4	−3.2	−3.5
5	−2.7	−3.2
5.8	−17	−16
6	−5.5	−6.03
6.9	−22	−16
7	−10	−8
8.4	−27	−24
9	−7	−9.8
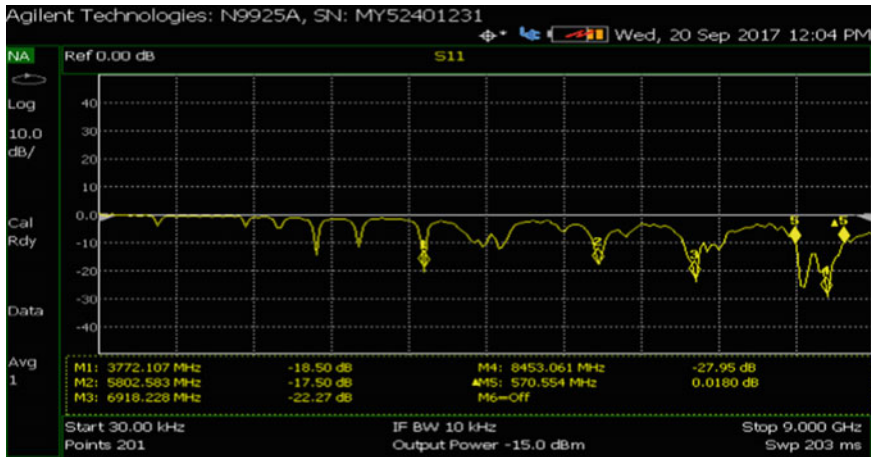
Fig. 13 Measured moisture of dry rice with microstrip ring sensor
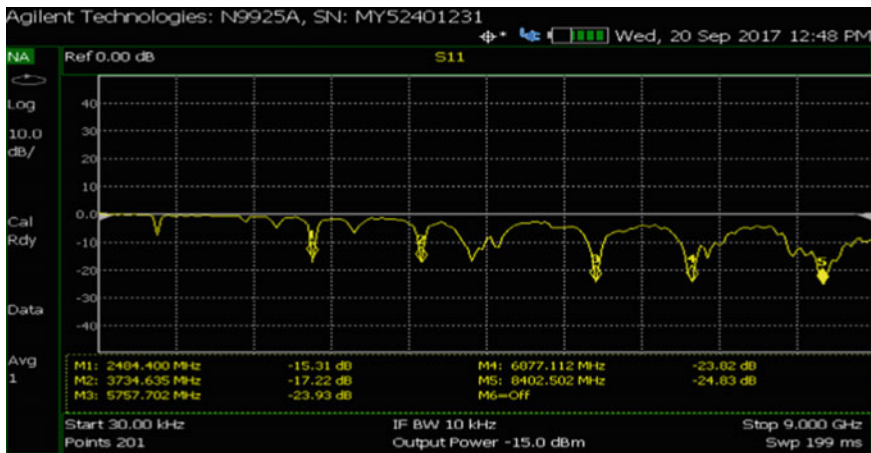
Fig. 14 Measured moisture of wet rice with microstrip ring sensor

as given below. The experimental results of dry and wet rice are shown in Figs. 10 and 11.

The actual moisture content is determined using the standard oven drying technique.

Moisture content (%) = (mass of water/dry mass of sample) * 100
Mass of water = wet of mass – dry of mass

where

m_{wet} is the initial mass before and after drying, and
m_{dry} is the final mass before and after drying.

Fig. 15 a Measured dry rice with microstrip ring sensor. **b** Measured wet rice with microstrip ring sensor. **c** Comparison of dry rice and wet rice measured moisture by microstrip ring sensor
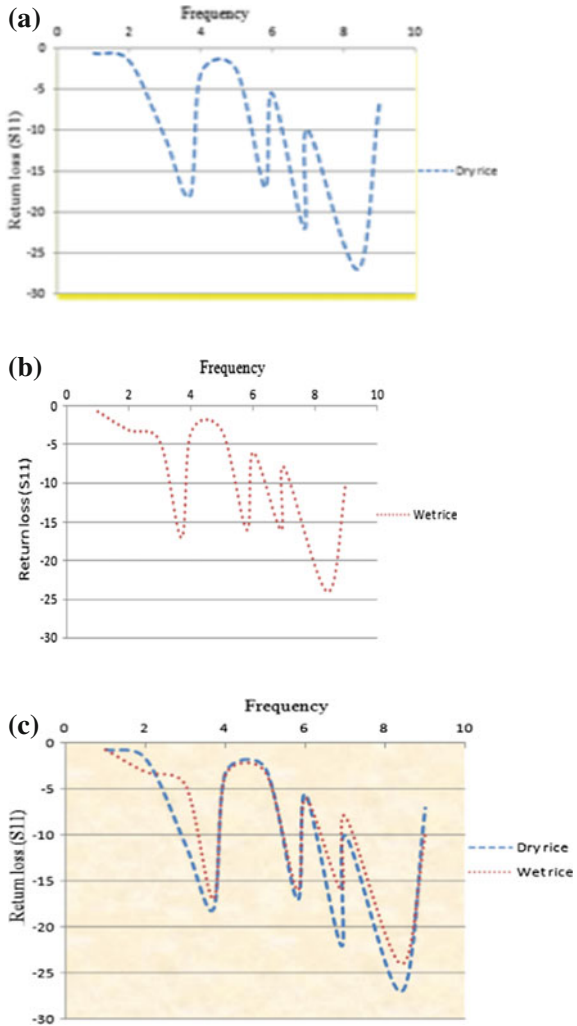
The experimental results of dry and wet rice measured the moisture content are shown in Figs. 13 and 14.

The comparison of moisture content of dry and wet rice with different frequencies is shown in Fig. 12.

3 Conclusion

A fast and accurate method of measurement of moisture content of rice grains using a microstrip ring sensor and a new microstrip ring moisture sensor is proposed using the oven drying method at multiple frequencies, i.e., 3.7, 5.8, 6.9, 8.4 GHz, with low

insertion loss as well as good return loss -18, -17, -22, -27 dB simulated using the CST software and measured by the vector network analyzer (Model No. Field fox N9925A). The proposed design is very cheap, versatile, small size, accuracy, reliability, sensitivity, time nonconsuming, as well as determination of moisture at dual and triple frequencies (Figs. 13, 14 and 15).

References

1. Mun, H.K., You, K.Y., Dimon, M.N.: Broken rice detection based on microwave measurement technique using microstrip wide sensor and microstrip coupled line sensor. AJCS, 2079–2090 (2013)
2. Nelson, S.O., Kraszewski, A.W., Samir, T., Kurt, C.L.: Using cereal grain permittivity for sensing moisture. IEEE Trans. Meas. **49**, 222–230 (2000)
3. Khamis, N.H., Ramli, H.S., Ean, T.C., Bakar, W.Z.: Application of microwave technology for home industry. In: Asia-Pacific Conference on Applied Electromagnetic, pp. 20–21 (2005)
4. Khalid, K., Hassan, J., Abbas, Z., Hamami, M.: Microwave dielectric properties of hevea rubber latex, oil palm fruit and timber and their application for quality assessment. In: Electromagnetic Aquametry. Springer (2005)
5. Jain, S., Mishra, P.K., Thakare, V.V.: Analysis of moisture sensor based on microstrip patch antenna: a review. In: International Conference on Futuristic Trends in Engineering Science, Humanities and Technology (FTESHT-16), vol. 3, no. 1 (2016)
6. Singh, D.K., Kumar, P., Rizvi, N.Z.: Microstrip transmission line sensor for rice quality detection. IJERGS **2**, 589–598 (2014)
7. Jain, S., Mishra, P.K., Thakare, V.V.: Design and analysis of microstrip moisture sensor for the determination of moisture content in soil. Ind. J. Sci. Technol. **9**(43) (2016). https://doi.org/10.17485/ijst/2016/v9i43/101065
8. You, K.Y., Salleh, J., Abbas, Z., You, L.L.: A rectangular patch antenna technique for the determination of moisture content soil. In: PIERS Proceedings, pp. 850–854. Cambridge, USA (2015)
9. Jain, S., Mishra, P.K., Thakare, V.V.: A Rectangular microstrip moisture sensor based on microwave measurement technique. SKIT Res. J. **7**(1) (2017)
10. Jain, S., Mishra, P.K., Thakare, V.V.: Bandwidth and gain enhancement in microstrip moisture sensor for agricultural applications. Ind. J. Sci. Eng. Res. **1**(1) (2017)
11. Jyoti, M., Kumar, M.P., Khare, P.K.: Study of characterization of pure and malachite green samples using spectroscopic studies. AIP Conference Proceedings, vol. 1349, pp. 561–562 (2010)
12. Rachana, K., Kumar, M.P., Jyoti, M.: Preparation, characterization and cyclic voltammetric studies on PVDF-PMMA polymer nanocomposite blends. Adv. Sci. Lett. **21**(9), 2943–2946 (2015)
13. Jain, S., Mishra, P.K., Thakare, V.V.: Design and development of microstrip sensor with triple frequency for determination of rice grains moisture content. Int. J. Mech. Prod. Eng. Res. Develop. (IJMPERD) **7**(5), 375–380 (2017)
14. Jain, S., Mishra, P.K., Thakare, V.V.: Design and analysis of dual-frequency microwave moisture sensor based on rectangular microstrip antenna. In: 4th International Workshop/Conference on Computational Condensed Matter Physics and Materials Science (IWCCMP-2016) (2016)
15. Kumar, M.P., Jyoti, M.: Investigation of the molecular mechanisms, trapping and charge storage behavior of polymer. Quant. Matter **5**(3), 335–338 (2016)
16. Abegaonkar, M.P., Karekar, R.N., Aiyer, R.C.: A microwave microstrip ring resonator as a moisture sensor for biomaterials: application to wheat grains. Meas. Sci. Technol. **10**, 195–200 (1999). https://doi.org/10.1088/0957-0233/10/3/014

17. Ansarudin, F., Abbas, Z., Hassan, J., Yahaya, N.Z., Ismail, M.A.: A simple insulated monopole sensor technique for determination of moisturecontent in hevea rubber latex. Meas. Sci. Rev. **12**, 249–254 (2012). https://doi.org/10.2478/v10048-012-0034-5

18. Zainuddin, M.F., Abbas, Z., Hafizi, M.H.M., Jusoh, M.A., Razali, M.H.H.: Monopole antenna technique for determining moisture content in the Dioscorea Hispida tubers. Aust. J. Crop Sci. **7**, 1–6 (2013)

19. ASABE, ASAE S352.2: Moisture measurement—unground grain and seeds. In: American Society of Agricultural and Biological Engineers, Michigan (2006)

20. Chang, K., Hsieh, L.-H.: Microwave Ring Circuits and Related Structures, 2nd edn. Wiley, Hoboken, New Jersey (2004)

21. You, K.Y., Salleh, J., Abbas, Z., You, L.L.: Cylindrical slot antennas for monitoring the quality of milled rice. In: Progress in Electromagnetics Research Symposium, pp. 1370–1373. Suzhou, China (2011)

22. Kim, K.-B., Kim, J.-H., Lee, S.S., Noh, S.H.: Measurement of grain moisture content using microwave attenuation at 10.5 GHz and moisture density. IEEE Trans. Instrum. Meas. **51**(1), 1196–1200 (2002)

23. Bhattacharya, K.R.: Rice Quality: A Guide to Rice Properties and Analysis. Woodhead Publishing, New Delhi, India (2011)

24. Trabelsi, S., Nelson, S.O., Lewis, M.A.: Microwave nondestructive sensing of moisture content in shelled peanuts independent of bulk density and with temperature compensation. Sens. Instrum. Food Qual. Safety **3**, 114–121 (2009). https://doi.org/10.1007/s11694-009-9071-7

25. Sumesh Sofin, R.G., Aiyer, R.C.: Measurement of dielectric constant using a Microwave Microstrip Ring Resonator (MMRR) at 10 GHz irrespective of the type of overlay. Microw. Opt. Technol. Lett. **47**(2), 11–14 (2005). https://doi.org/10.1002/mop.21066

26. Jain, S., Mishra, P.K., Thakare, V.V.: The analysis and design of circular microstrip moisture sensor for rice grain. In: 4th International Workshop/Conference on Computational Condensed Matter Physics and Materials Science (IWCCMP-2016) (2016)

A Novel Fabric Adhesive UWB Magnetoelectric Dipole Antenna

Neetu and Vivekanand Tiwari

Abstract Planting antennas over fabrics is a known secure technique to help human beings in various applications like health monitoring, firefighting, rescue work, and space and military personal communications. This paper provides a novel opportunity to design a lightweight and robust fabric adhesive magnetoelectric dipole antenna covering 5.0–10.0 GHz. The antenna possesses ultrawide impedance bandwidth, dual circular polarization, and high gain in the entire band of frequency operation. This antenna has been pasted over the clothes in the region of minimal bending, and good electrical and radiation characteristics have been exhibited in the frequency region meant for satellite communication.

Keywords Adhesive antenna · ME dipole · Circular polarization · 3-dB axial ratio · Satellite communication

1 Introduction

Rooting wireless structures over clothes under the name on-body antenna is a possible and qualified technique that promotes user safety, awareness, convenience, and operability in various applications like health monitoring [1], firefighting [2], rescue operation [3], and space and military personal communications [4]. The literature provides detailed information for both on-body [5–7] as well as off-body [8, 9] embedded wireless techniques. The on-body antennas are also known as Body-Centric (BC) antennas.

The purpose of placing the antenna on user's clothes is to make an antenna conformable and least obstructive as possible. With the movements of user, the antenna shall not lose its connectivity as well as its functionality. However, literature confirms

Neetu (✉) · V. Tiwari

Department of Electronics and Communication, Manipal University Jaipur, Jaipur, India

e-mail: neetu03@rediffmail.com

111

that the shape and placement of antenna are two major constraints for the underrated performance of the antenna [10]. Hence, choosing and designing the appropriate shape of the antenna and its placement on body parts with minimum bending possibilities are two ways to improve the efficiency of the BC antenna. In 2006, an improvised version of the complementary antenna was presented and designated as magnetoelectric (ME) dipole antenna. Most of the ME dipole antennas have been designed on copper sheet persisting a nonplanar geometry. Here, the planar electric dipole is mounted on quarter wave shorted magnetic dipole. Despite having a bigger structure, the antenna has the capability to provide wide impedance bandwidth, identical E-plane and H-plane radiation patterns, stable gain, and low back radiations [11, 12]. A planar circularly polarized ME dipole antenna has been proposed for X-band with 21.1% impedance bandwidth and 3-dB axial ratio bandwidth of 9.52% [13].

This paper provides the key strategy to construct a fixed and lightweight BC ME dipole antenna with high gain and left-handed circular polarization to be used for satellite communication applications. The antenna is made operational in an outdoor environment to allow normal activities like walking, running, and sitting, while the antenna would perform its transmission and reception process.

2   Antenna Design

In order to provide an adequate solution for integrating the antenna onto clothing, a planar ME dipole antenna has been selected as the baseline for this research because of its inherent low profile. Moreover, keeping the purpose of BC antenna in mind where the size of the ground plane has to be small, and this is meticulously done by maintaining the ground plane as 48 mm × 2 mm (GL × GW). Figure 1 shows the geometry of the proposed antenna. This antenna has been printed on RT Duroid 5880 substrate, having permittivity of 2.2 and thickness 0.78 mm. In its basic geometry, the antenna consists of horizontal T-shaped electric dipole, two directors, an inverted U-shaped feed line, a pair of shorted stubs, and a truncated rectangular-shaped ground plane having dimension 48 mm × 2 mm (GL × GW). The truncated ground plane is combined with shorted stubs to form a magnetic dipole. The electric dipole, feeding structure, directors, and shorted stubs are printed on the front side of the substrate whereas the truncated ground plane is printed on the backside of the substrate. Here, the concept of directors in Yagi–Uda antenna for gain enhancement has been utilized and ME directors are printed on the front side of the substrate. Figure 2 shows the final prototype of the proposed antenna according to the dimensions indicated in Table 1.
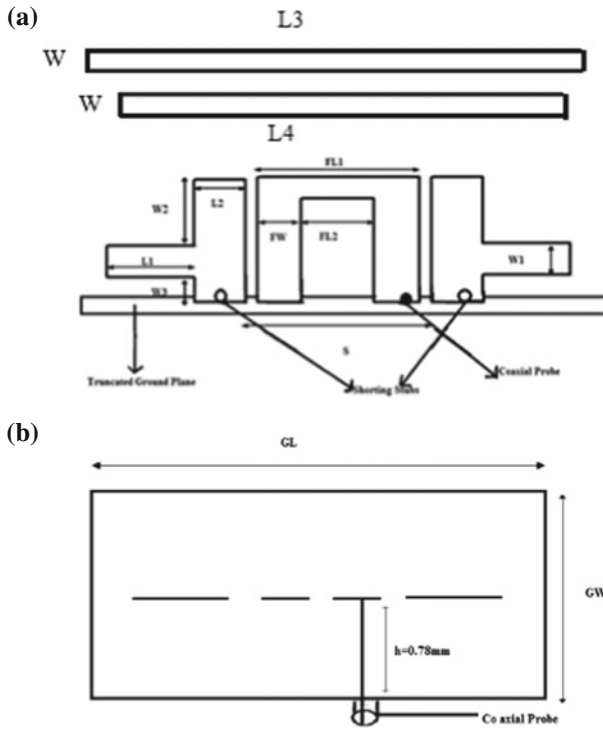
Fig. 1 **a** Top view of the proposed antenna, and **b** side view of proposed antenna
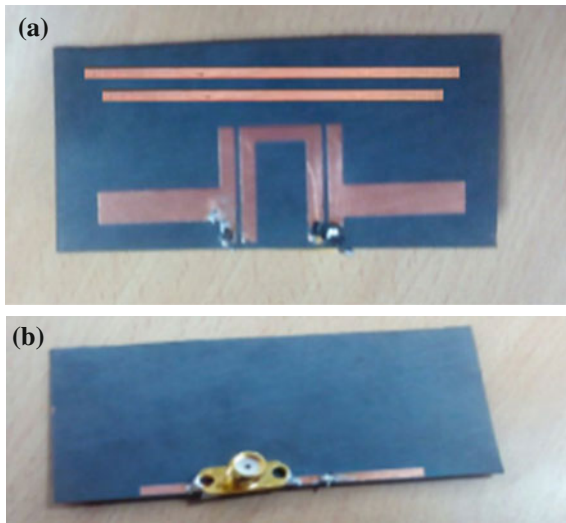
Fig. 2 **a** Front side of the proposed antenna, and **b** backside of proposed antenna

Table 1 Optimized dimensions of the proposed antenna

Parameter	L1	L2	W1	W2
Value (mm)	20	2	5.5	10.5
Parameter	CL	CW	CH	GL
Value (mm)	80	40	10	48
Parameter	W3	FL1	FL2	L3, W
Value (mm)	4	12	7	60, 1.5
Parameter	GW	FW	S	L4
Value (mm)	2	2.5	14	50

3 Results and Discussion

The proposed circularly polarized ME dipole antenna is designed and analyzed using MOM-based full-wave electromagnetic simulator IE3D. The variation in simulated and measured reflection coefficient with frequency is presented in Fig. 3. The proposed antenna delivers an ultrawide impedance bandwidth of 66.6% in the frequency range 5.0–10.0 GHz. Figure 4 represents the variation in axial ratio in broadside direction with frequency. It has been clearly showcased from the figure that the antenna has achieved circular polarization in dual bands. The axial ratio taken at (45,0) degree remains below 3-dB from 5.2 to 6.3 GHz hence giving 3-dB axial ratio bandwidth of 19.1%. When the axial ratio is measured at (60,0), the antenna is still below 3-dB in the frequency range 9.0–10.0 GHz, thereby giving 3-dB axial ratio impedance bandwidth of 10.52%. The peak gain of 7.5 dBi has been confirmed from simulated results, and a stable gain of almost 6.5 dBi in the entire operating range of frequency operation has been indicated in Fig. 5. To prove that the proposed ME dipole antenna is having circular polarization, current distribution pattern at 5.5 GHz has been highlighted in Fig. 6. The figure gives insight about the antenna at quarter time period interval. The overall current in the structure at $t = 0$ is moving toward right-hand side while it is in downward direction at $t = T/4$. Similarly, the vector current rotates in anticlockwise direction, suggesting left-handed circular polarization. One of the important characteristics of ME dipole antenna is to show identical E-plane and H-plane radiation pattern. It is observed that almost identical E-plane and H-plane radiation patterns are shown by the antenna at 5.5, 6.5, 7.5, 8.5, and 9.5 GHz, which is shown in Fig. 7. The prototype of the proposed antenna has been made and embedded over clothes to make a BC antenna, which has been shown in Fig. 8.

Fig. 3 Measured and simulated reflection coefficient of the proposed antenna in off-body state
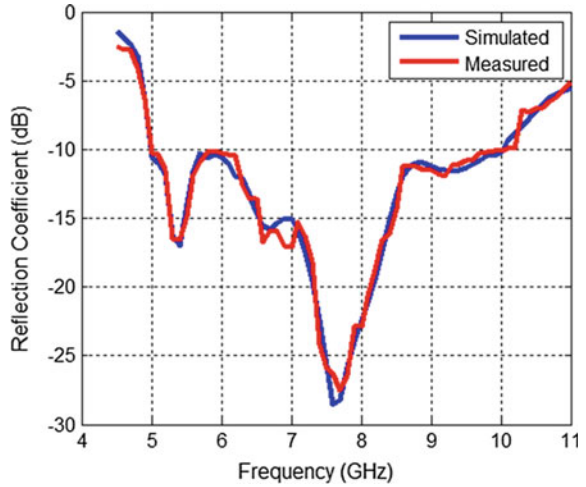
Fig. 4 Measured and simulated 3-dB axial ratio of proposed antenna in off-body state
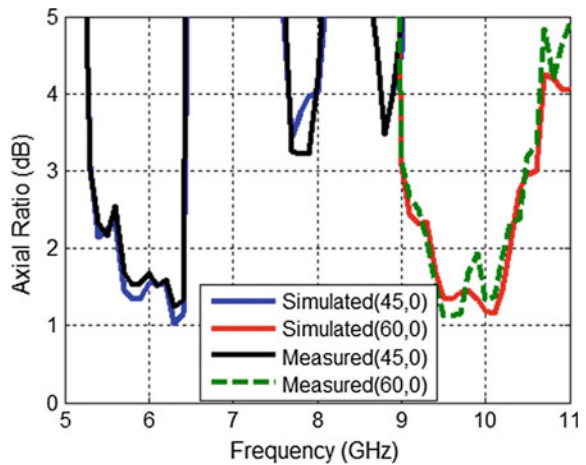
4 Antenna Measurements On-body State

Placement of the antenna over the human body will effectively reduce its gain and radiation efficiency due to the change in impedance matching caused by the lossy human tissues. Keeping this in mind, the designed antenna has been kept over the human body with cotton padding having a thickness of 6 mm between them. Measurement of reflection coefficient when antenna has been placed over arm (P1) and back (P2) is shown in Fig. 9. Measured results indicate good agreement between results of reflection coefficient for antenna in free space and when placed at P2. At position P1, the change in impedance bandwidth is significant. For the position P1, the antenna is resonating in dual bands in the frequency range 5.3–5.8 and 6.3–8.9 GHz, which will
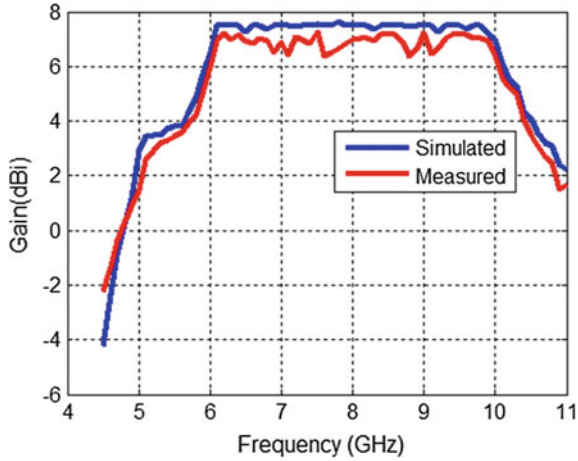
Fig. 5 Measured and simulated gain of proposed antenna in off-body state

(a) t=0

(b) t=T/4

(c) t=T/2

(d) t=3T/4
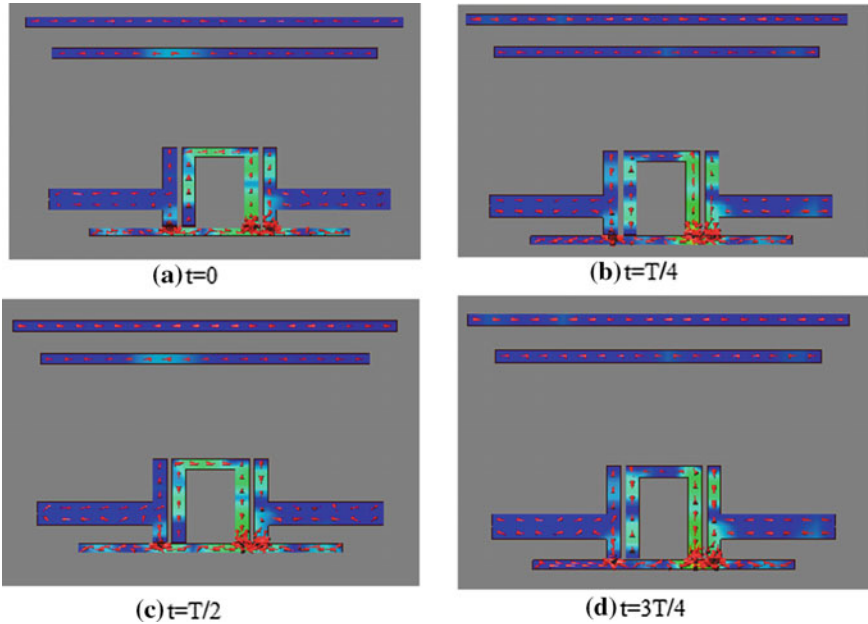
Fig. 6 Current distribution indicating left-handed circular polarization at 5.5 GHz in off-body state

lead to overall reduction in impedance bandwidth. However, at position P2, there is a drift of 300 MHz and the antenna is resonating at frequency 5.3–10.0 GHz. Due to the wave impedance mismatching between low (fat) and high (skin, muscle) water content level in various tissues, significant reflections may occur for far-field exposure. This effect can lead to the increased specific absorption rate (SAR), and hence
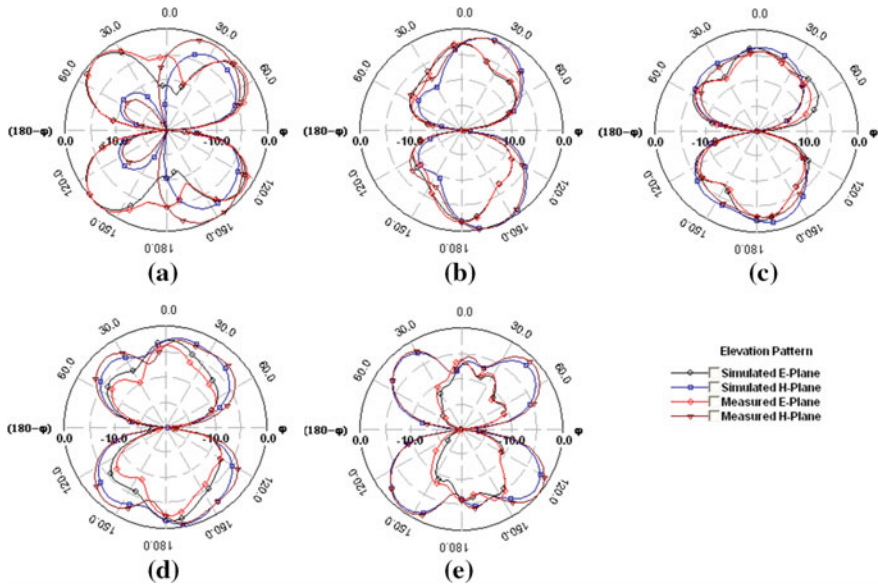
Fig. 7 Measured and simulated E-plane and H-plane radiation pattern in off-body state at **a** 5.5 GHz, **b** 6.5 GHz, **c** 7.5 GHz, **d** 8.5 GHz, and **e** 9.5 GHz

Table 2 Gain performance of the proposed antenna at different body positions at different frequency regions

Frequency band (GHz)	Total gain (dBi)		
	Free space	P1	P2
5.5	3.8	5.2	6.12
6.5	7.25	9.43	10.25
7.5	7.55	9.64	10.59
8.5	7.5	9.52	10.45
9.5	7.4	9.47	10.32

interfering the results obtained for reflection coefficient. Figure 10 gives the information about E-plane and H-plane radiation pattern of the antenna kept in position 1 and position 2. The antenna has shown nearly omnidirectional radiation pattern at 7.5 and 8.5 GHz and bidirectional (quasi-end-fire) radiation pattern at 5.5, 6.5 and 9.5 GHz. Omnidirectional and quasi-end-fire radiation patterns help the human by enforcing less penetration of the radiations into the body as compared to broadside radiation pattern where maximum radiations are absorbed by the body. As the antenna is circularly polarized at 5.5 and 9.5 GHz, which will further lead to fewer losses to the human body. The gain of antenna is also changed when the antenna has been kept at different positions on the body which has been summarized in Table 2.

Fig. 8 **a** Proposed body-centric antenna at position 1 (P1), and **b** proposed body-centric antenna at position 2 (P2)

(a)

(b)
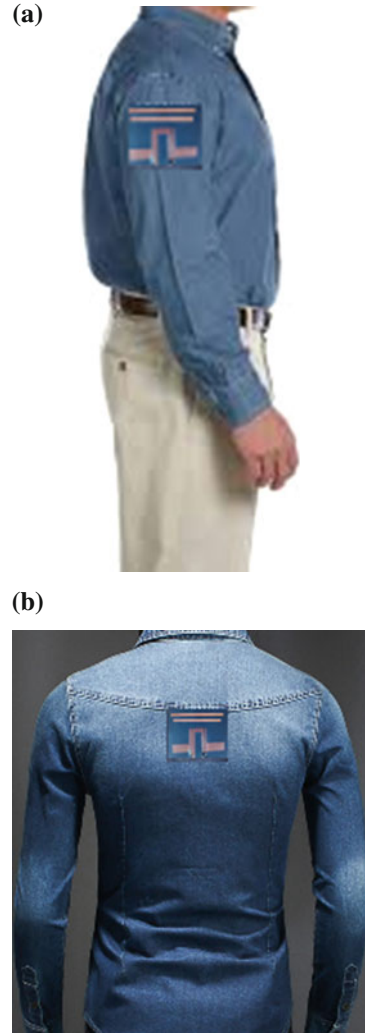
5 Conclusion

A novel and simple circularly polarized body-centric magnetoelectric dipole antenna has been proposed, designed, and fabricated. The measured and simulation results confirm that the antenna exhibits ultrawide impedance bandwidth of 66.6% in the frequency region 5.0–10.0 GHz. This antenna is also capable to provide 3-dB axial ratio bandwidth of 19.1% and 10.52% covering 5.2 GHz–6.3 GHz and 9.0 GHz–10.0 GHz, respectively. The antenna has shown radiation pattern with almost symmetrical E-plane and H-plane radiation patterns and provides a peak gain of 7.5 dBi. The pro-
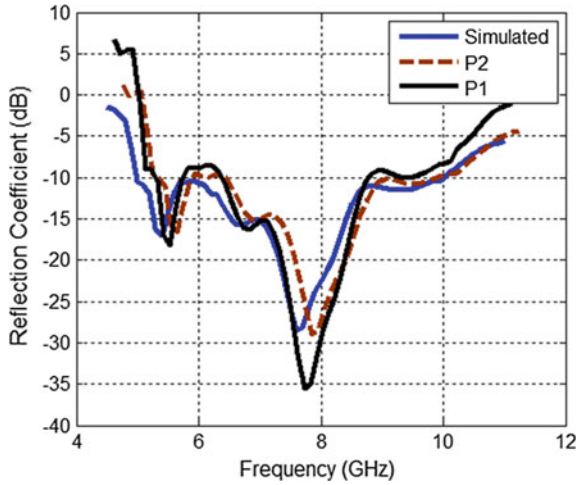
Fig. 9 Measured reflection coefficient of the proposed antenna in on-body state
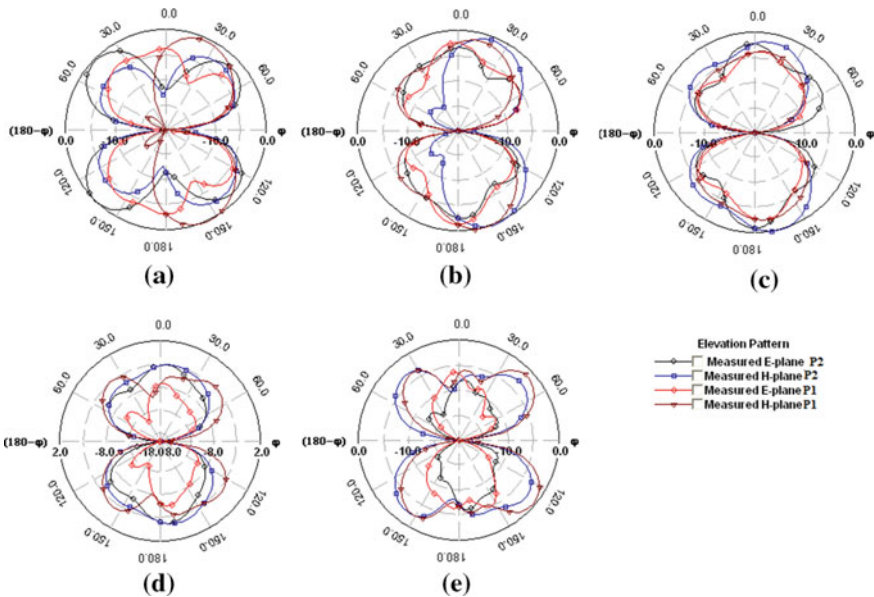
Fig. 10 Measured E-plane and H-plane radiation patterns of the proposed antenna in on-body state at **a** 5.5 GHz, **b** 6.5 GHz, **c** 7.5 GHz, **d** 8.5 GHz, and **e** 9.5 GHz

posed antenna has been kept at two different minimal displacement positions, and the antenna has shown good radiation and electrical characteristics, while maintaining circular polarization in the bands meant for satellite communication.

References

1. Alomainy, A., Hao, Y., Pasveer, F.: Numerical and experimental evaluation of a compact sensor antenna for healthcare devices. IEEE Trans. Biomed. Circuits Syst. **1**(4), 242–249 (2007)
2. Vallozzi, L., Van Torre, P., Hertleer, C., Rogier, H., Moeneclaey, M., Verhaevert, J.: Wireless communication for firefighters using dual-polarized textile antennas integrated in their garment. IEEE Trans. Antennas Propag. **58**(4), 1357–1368 (2010)
3. Curone, D., Secco, E.L., Tognetti, A., Loriga, G., Dudnik, G., Risatti, M., Whyte, R., Bonfiglio, A., Magenes, G.: Smart garments for emergency operators: the ProeTEX project. IEEE Trans. Inf. Technol. Biomed. **14**(3), 694–701 (2010)
4. Salonen, P., Rahmat Samii, Y.: Textile antennas: effects of antenna bending on input matching and impedance bandwidth. IEEE Aerosp. Electron. Syst. Magazine **22**(12), 18–22 (2007)
5. Hall, P.S., Hao, Y., Nechayev, Y.I., Alomalny, A., Constantinou, C.C., Parini, C., Kamarudin, M.R., Salim, T.Z., Hee, D.T.M., Dubrovka, R., Owadally, A.S., Song, W., Serra, A., Nepa, P., Gallo, M., Bozzetti, M.: Antennas and propagation for on-body communication systems. IEEE Antennas Propag. Magazine **49**(3), 41–58 (2007)
6. Khan, I., Hall, P.S., Serra, A.A., Guraliuc, A.R., Nepa, P.: Diversity performance analysis for on-body communication channels at 2.45 GHz. IEEE Trans. Antennas Propag. **57**(4), 956–963 (2009)
7. Psychoudakis, D., Volakis, J.L.: Conformal asymmetric meandered flare (AMF) antenna for body-worn applications. IEEE Antennas Wirel. Propag. Lett. **8**, 931–934 (2009)
8. Hertleer, C., Rogier, H., Vallozzi, L., Van Langenhove, L.: A textile antenna for off-body communication integrated into protective clothing for fire-fighters. IEEE Trans. Antennas Propag. **57**(4), 919–925 (2009)
9. Cotton, S.L., Scanlon, W.G.: Measurements, modeling and simulation of the off-body radio channel for the implementation of bodyworn antenna diversity at 868 MHz. IEEE Trans. Antennas Propag. **57**(12), 3951–3961 (2009)
10. Kellomaki, T., Heikkinen, J., Kivikoski, M.: Effects of bending GPS antennas. In: Proceedings of Microwave Conference, pp. 1597–1600 (2006)
11. Luk, K.M., Wong, H.: Complementary wideband antennas, US No. 7843389, 30 Nov (2010)
12. Luk, K.M., Wong, H.: A new wideband unidirectional antenna element. Int. J. Microw. Opt. Technol. **1**(1), 34–44 (2006)
13. Neetu, G.P.P., Tiwari, V.N.: Design of circularly polarized planar magneto-electric dipole antenna. Prog. Electromag. Res. C **67**, 21–29 (2016)

Experimental Investigation Using Laser Vibrometer and Finite Element Modeling for Modal Analysis of Camshaft

Jai Kumar Sharma and Sandeep Kumar Parashar

Abstract Vibration analysis of the internal combustion (IC) engine components is extremely helpful in optimizing the design of these components and has direct bearing on the engine life. This paper presents an experimental and numerical modal analysis of IC engine camshaft. The experimental modal analysis of camshaft excited using impact hammer is performed with the help of a laser vibrometer. The signals from the experiments are obtained in the time domain and are later converted into the frequency domain using fast Fourier transformation (FFT) with the help of MATLAB program. The numerical modal analysis is performed using finite element modeling (FEM) software ANSYS Workbench 14.5. The purpose of this experiment is to validate the finite element model of camshaft using the experimental data, so that the same FEM model of camshaft can be used with confidence for more complex boundary conditions or real operating conditions. To achieve this aim, the natural frequencies and mode shapes of camshaft obtained through experiment and FEM are compared. The methodology presented and the results are useful in optimizing the design of camshaft to be used in IC engine.

Keywords Experimental modal analysis · Camshaft · Finite element modeling
Fast Fourier transformation · Laser vibrometer

1 Introduction

The experimental modal analysis of structures and components has received a lot of attention nowadays. The experimental modal analysis technique can be used to determine the dynamic behavior of structure in terms of modal parameters such as natural frequency, mode shapes, and damping. In recent days, the experimental modal

J. K. Sharma (✉) · S. K. Parashar
Department of Mechanical Engineering, Rajasthan Technical University, Kota, India
e-mail: jaikumarsharma94@gmail.com

S. K. Parashar
e-mail: skparashar@rtu.ac.in

analysis can be performed using laser vibrometer. It is a noncontact device which eliminates the impact of the sensor on the behavior of test structure and components. Currently, finite element modeling plays a very vital role for modal analysis of complex engineering structure and components.

This paper presents a modal analysis of camshaft through experiment and numerical method. The literature gives a great number of works on the modal analysis of engineering structure and components. Nagalo et al. [1] investigated vibration response of freight wagon. The equation of the modal analysis was obtained using Lagrange's equation. In this paper, complex-exponential method was used for determining the modal parameter. Nikhil et al. [2] developed and designed a test rig for estimating the vibration of the beam. They investigated the transverse vibration of the flexible rectangular beam (aluminum alloy and mild steel) by applying the centrifugal force. The dynamic behavior of the tire in rotating condition was investigated through experiments by Diaz et al. [3]. Giagopouls and Natsiavas [4] investigated dynamic behavior of a complex vehicle, creating a superstructure through a combination of experimental and numerical modal analysis. Bertini et al. [5] performed an experimental modal analysis of bladed wheel using laser Doppler vibrometer. Teter and Gawryluk [6] determined natural frequencies and mode shapes of rotor blade using laser vibrometer. In this paper, the blades were made of glass–epoxy unidirectional laminate. The experimental results were compared using finite element modeling software package ABAQUS. Chandravanshi and Mukhopadhyay [7] determined the modal parameters of the vertically tapered frame using experimental modal analysis, and the experimental results were validated through finite element modeling software ANSYS Workbench.

In the present paper, experimental and finite element modal analysis of camshaft is carried out. The experimental modal analysis is performed using the laser vibrometer and for finite element modeling, ANSYS Workbench is used. The purpose of this analysis is to validate the finite element modal by experiment. Once the modal has been validated, one can apply more complex boundary conditions in finite element modeling.

2 Experimental Modal Analysis

The experimental study of engineering structure and components always provide a better understanding and control of the vibration characteristics encountered in practice [8]. In the present work, the laser vibrometer (Polytech NLV 2500-05) is used for the measurement of vibrations. This laser vibrometer has two units, one is scanning head and the other one is data acquisition unit [9]. During the measurement of the natural frequency of camshaft, the reference light is reflected from the camshaft. The frequency of reflected light which is changed by the Doppler's effect is compared with reference signal frequency. The schematic diagram of experimental setup is shown in Fig. 1. First, the camshaft is excited at the selected point by the impact hammer. Then, the vibration signals are captured by the laser vibrometer (NLV-2500-
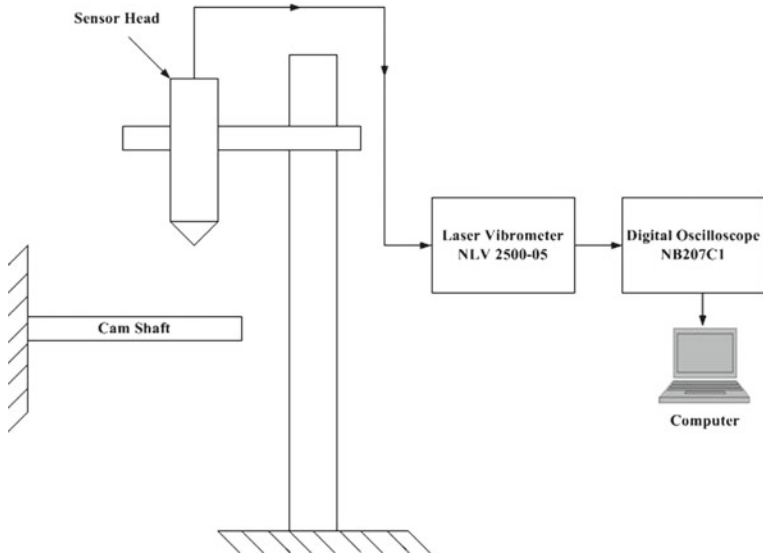
Fig. 1 Diagram of measuring free vibration

Fig. 2 Actual experiment setup

05) and fed into digital oscilloscope (NB207C1). From the digital oscilloscope, the time-domain signal is saved in USB drive. These signals obtained are of continuous type. These continuous signals are discretized by the process called sampling. At the sampling rate (8×10^{-5}), 20,482 samples are taken in the present study. The time-domain signals are converted into frequency-domain signal using FFT algorithm with the help of MATLAB program. The actual photograph of experiment is shown in Fig. 2.

3   Results and Discussion

We can determine the natural frequency of the camshaft up to first mode using the impact hammer. We cannot excite higher mode of the camshaft using the impact hammer in the fixed-free boundary condition because the camshaft is made of chilled cast iron [10]. The chilled cast iron is a brittle material and so we cannot capture the vibrating signal for higher mode because the signal dies out very early due to high damping nature of the material. The cantilever boundary conditions are also specifically chosen as it was easier to excite the first bending mode by impact hammer in comparison to the other boundary conditions such as free–free.

Table 1 displays the experimentally determined natural frequency of the camshaft. The experimentally obtained time-domain signal (from digital scope) and frequency-domain signal (from MATLAB) are shown in Fig. 3 and Fig. 4, respectively. We can clearly see in the frequency-domain graph the first peak that occurs at the first natural frequency of 63.48 Hz.

Table 1   Experimental results of camshaft

Mode	Experimental natural frequency (Hz)
Mode 1	63.48
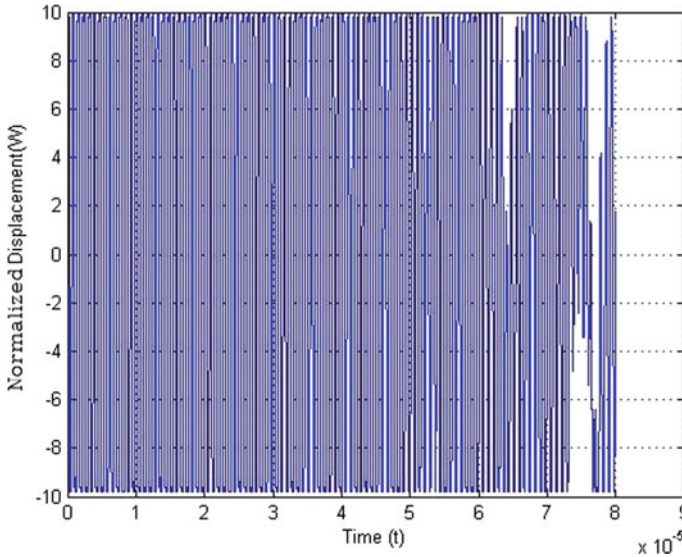
Fig. 3   Time-domain signal for camshaft
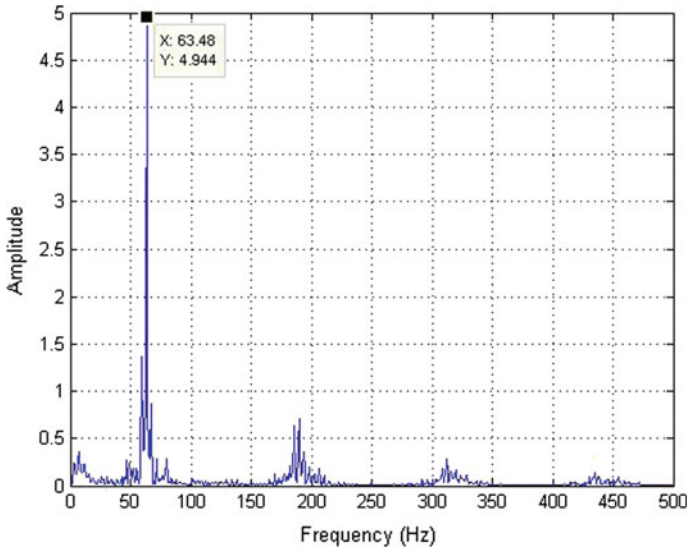
Fig. 4 Frequency-domain signal for camshaft

4 Finite Element Modal Analysis of Camshaft

The three-dimensional model of the camshaft is constructed in ANSYS Workbench 14.5, and then computational modal analysis is performed to generate natural frequencies and mode shapes for fixed-free boundary conditions. The camshaft is made of chilled cast iron, and the mechanical properties of cast iron are given in Table 2.

The geometric parameters of the camshaft are displayed in Fig. 5, while the actual camshaft used in the experiments is shown in Fig. 6. The default meshing and elements for analysis are taken in the ANSYS workbench. For the fixed-free boundary conditions, one end of the camshaft is fixed while the other end is free as displayed in Fig. 7.

The cam angles are carefully measured and are taken for the analysis from right to left as 0, 111, 276, 20, 186, 66, 290, and 172 in degrees, respectively.

The numerically obtained natural frequencies of camshaft are shown in Table 3. The numerical mode shapes and natural frequencies up to two modes are shown in Fig. 8.

Table 2 Mechanical properties of chilled cast iron

Material parameter	Values
E	1.1×10^{11} N/m^2
ρ	7200 kg/m^3
υ	0.28

A = 34, B = 34, C = 18, D = 26, E = 18, F=26, G=18, H=26, I= 18, J=32,
K=18, L = 26, M=18, N=26, O =18, P= 26, Q = 18, R = 26, S = 28, T =
26, U = 21.68 All dimensions in mm
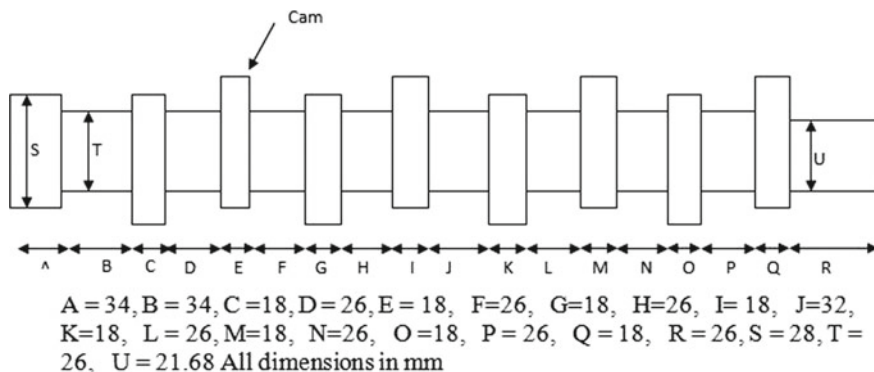
Fig. 5 Schematic diagram of camshaft

Fig. 6 Actual camshaft

Table 3 Numerical natural frequency of camshaft

Mode	Numerical natural frequency (Hz)
Mode 1	69.495
Mode 2	441.46

5 Comparisons Between Numerical and Experimental Results for Camshaft

The results of the numerical natural frequencies and experimental natural frequencies of chilled cast iron camshaft in fixed-free boundary conditions are calculated using the material properties (Table 2) and dimensions (Fig. 5) of the camshaft. The finite element natural frequencies are determined using the finite element modeling soft-
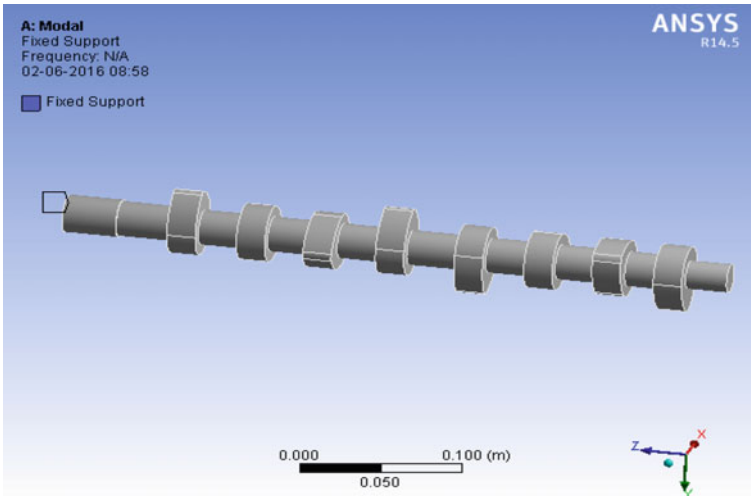
Fig. 7 Camshaft in fixed-free boundary conditions
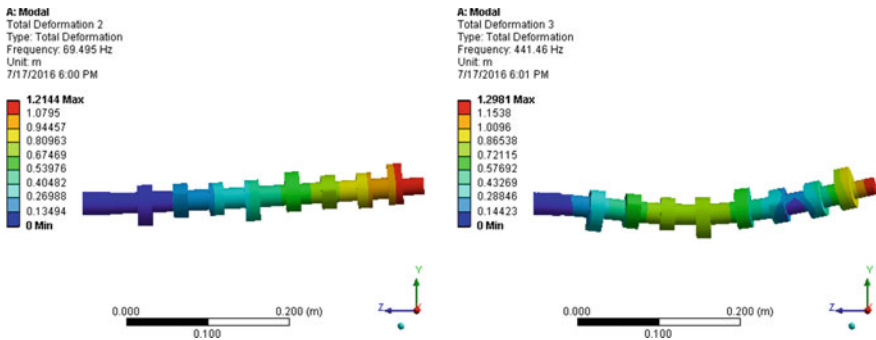
Fig. 8 Mode shapes and corresponding natural frequencies of camshaft

ware ANSYS workbench 14.5 and experimental natural frequencies are determined using impact hammer excitation method.

It was observed that in the experiment it was not possible to excite the second and higher modes of the camshaft using the impact hammer because camshaft is made of cast iron so the vibration signal dies out very quickly. For determination of higher modes, the forced excitation method such as the sine sweep test can be more effective. It can be observed from Table 4 that the experimental and numerical results for the camshaft are having good correlation. Some error between the results is introduced due to the fact that the geometry of camshaft is very complex and while obtaining physical dimensions minor error has occurred. However, the results obtained from the two methods are well within the acceptable range of error.

Table 4 Comparison between the numerical and experimental natural frequency

Mode	Numerical natural frequency (Hz)	Experimental natural frequency (Hz)
Mode 1	69.445	63.48
Mode 2	441.46	–

6 Conclusion

In this paper, the experimental and finite element modal analysis of camshaft was performed. The experimental modal analysis of camshaft was carried out using laser vibrometer and impact hammer is used for excitation. It was not possible to excite the second mode of camshaft in the experiment in fixed-free boundary condition using impact hammer excitation method because the camshaft is made of chilled cast iron. The cast iron is a brittle material, and it has high damping characteristics so excited vibration dies out very early. The finite element modeling was carried out using ANSYS Workbench 14.5. The finite element modeling and experimental results are found to have a good correlation. The cam has a complex geometry and it is difficult to measure physical dimension. Furthermore, it is also difficult to develop a three-dimensional model in CAD software. Due to these facts, some error has cropped in between the experimental and the numerical results. The overall analysis of camshaft shows that the results are all good within a reasonable error. The vibration analysis technique (experimental and numerical) presented in this paper may be helpful in developing and optimizing the design of automobile components.

References

1. Nangolo, N.F., Soukup, J., Rychlíková, L., Skočilas, J.: A combined numerical and modal analysis on vertical vibration response of railway vehicle. Proc. Eng. **96**, 310–319 (2014)
2. Nikhil, T., Chandrahas, T., Chaitanya, C., Sagar, I., Sabareesh, G.R.: Design and development of a test-rig for determining vibration characteristics of a beam. Proc. Eng. **144**, 312–320 (2016)
3. Diaz, C.G., Kindt, P., Middelberg, J., Vercammen, S., Thiry, C., Close, R., Leyssens, J.: Dynamic behaviour of a rolling tyre: experimental and numerical analyses. J. Sound Vib. **364**, 147–164 (2015)
4. Giagopoulos, D., Natsiavas, S.: Dynamic response and identification of critical points in the superstructure of a vehicle using a combination of numerical and experimental methods. Exp. Mech. **55**, 3 (2015)
5. Bertini, L., Neri, P., Santus, C., Guglielmo, A.: Automated experimental modal analysis of bladed wheels with an anthropomorphic robotic station. Exp. Mech. **57**, 273–285 (2017)
6. Teter, A., Gawryluk, J.: Experimental modal analysis of a rotor with active composite blades. Comp. Struct. **153**, 451–467 (2016)

7. Chandravanshi, M.L., Mukhopadhyay, A.K.: Modal analysis of a vertically tapered frame. Int. J. Struct. Stab. Dy. **16**, 1–16 (2016)
8. Ewins, D.J.: Modal Testing: Theory, Practice and Application, 2nd edn. Wiley, ISBN: 978-0-86380-218-8
9. Parashar, S.K., von Wagner, U., Hagedorn, P.: Finite element modeling of nonlinear vibration behavior of piezo-integrated structures. Comput. Struct. **119**, 37–47 (2013)
10. Kumruog̃lu, L.C.: Mechanical and microstructure properties of chilled cast iron camshaft: Experimental and computer aided evaluation. Mat. Design **30**, 927–938 (2009)

Performance Evaluation of Cognitive Internet of Things in Asynchronous Distributed Space-Time Block Codes over Two-Wave Diffuse Power Fading Channel

Varsha Vimal Sood, Surbhi Sharma and Rajesh Khanna

Abstract The internet of things (IoT) is the strongest potential candidate for future internet services. It bridges the real and the virtual worlds. Cooperative communication is an enabling technology for the interconnection of the distributed heterogeneous entities. Cognitive IoT empowers the system to behave intelligently as the human brain does. Application of the human cognition model onto the IoT reduces the degree of human intervention. In order to make a decision by the system, reliability is a key factor which can be obtained from cooperative diversity technique such as distributed space-time block codes (DSTBC). In this work, we analyze the average bit error rate (ABER) of asynchronous distributed space-time block codes (ADSTBC) cooperative scheme in two-wave diffuse power (TWDP) fading channel. Closed-form expression of the performance metric is mathematically derived, analyzed, and verified by Monte Carlo simulation.

1 Introduction

Internet of things (IoT) heralds a new era in the Internet service and has become a highly researched topic [1]. It is considered the network of the future where the heterogeneous entities like people, things, and services get interconnected [2]. IoT

V. V. Sood (✉) · S. Sharma · R. Khanna
Electronics and Communication Engineering Department,
Thapar Institute of Engineering and Technology (TIET), Patiala, India
e-mail: varsha.vimal@thapar.edu

S. Sharma
e-mail: surbhi.sharma@thapar.edu

R. Khanna
e-mail: rkhanna2001@gmail.com

finds applications in medical sciences, robotics, traffic management, cognitive radios, smart homes and cities, and many more which the future shall unravel [3–6]. Along with interconnections, researches are going on to add human cognitive like ability to the IoT which calls for reliable information to be acted upon [7, 8]. Cognitive IoT (CIoT) equipped with sensors, processors, and actuators is a distributive system which can take intelligent decisions and/or perform suitable action. Cooperative communication systems play a very significant role in providing a reliable and untethered wireless communication [9]. It can create a virtual or a distributed space-time block codes (DSTBC) by the mutual participation of two or more single (or multiple) antenna equipment [10]. A cooperative CIoT is capable of offering diversity gains for reliable communication [11–13]. The wirelessly connected CIoT nodes are geographically dispersed, therefore susceptible to fading and timing delays. Due to the distributed locations of the relaying nodes and different times of transmission from them, the cooperative CIoT system is basically an asynchronous system [14]. In the existing literature, several delay-tolerant codes were proposed for different fading scenarios such as Rayleigh, Rician, Nakagami, etc. [15–18]. Future developments in IoT draw inspiration from the working of the human brain and its support system [19, 20]. The brain acts as the central data system. The spinal cord serves as the distributed network of data processing nodes and smart gateways, whereas the nerves can be viewed as the networking components and sensors. Such a mapping of the human brain model to the IoT holds immense potential in the healthcare system [21]. In this work, we investigate the performance of asynchronous distributed space-time block codes (ADSTBC)-based CIoT system in the two-wave diffuse power (TWDP) fading channel [22]. TWDP fading is used to model severe fading environment. The present work analyzes the performance of a selective decode and forward (DF)-based ADSTBC for IoT in TWDP fading channel.

2   Cognitive Internet of Things (CIoT) System Model

Cognitive Internet of Things (CIoT) here refers to the knowledge acquired by the IoT system through efficient sensors and processors and is able to take decisions and action thereafter through suitable actuators. For this type of decision-making by an entity, such as device or human, it is imperative for the information being communicated be reliable. IoT can be implemented with the help of cooperative communication, where the IoT nodes relay each others' information to the destination. DSTBC is a highly reliable system for the cooperative communication to obtain diversity. The distributed topology of the IoT nodes and their different times of transmission, the cooperative CIoT is asynchronous in nature. In order to make the system delay tolerant, we apply an asynchronous distributed space-time block codes (ADSTBC) system which provides diversity gain in spatially located IoT nodes prone to delay.
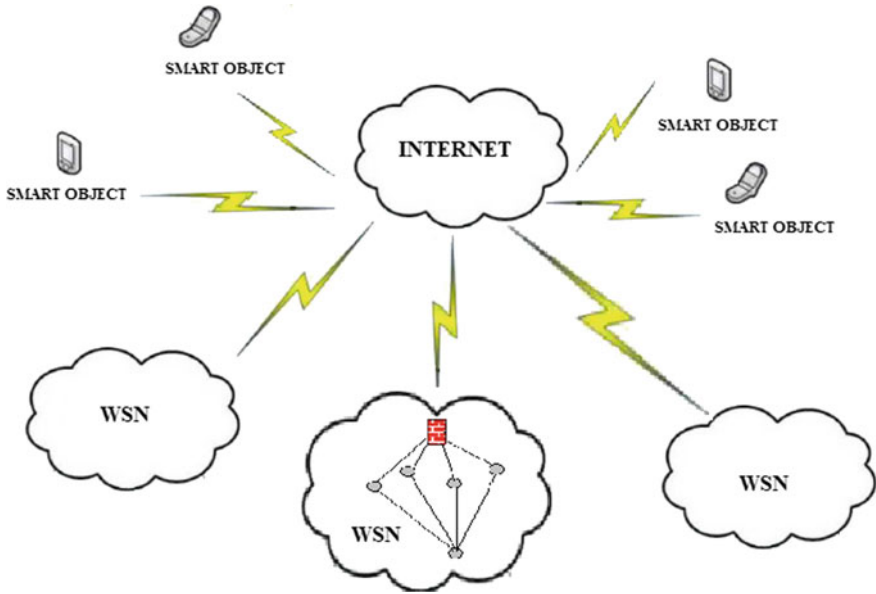
Fig. 1 Illustration of an IoT system

In this work, we study an ADSTBC system which can serve as a model for a wireless sensor network (WSN) where high reliability is necessary. In other cases, plain relaying-based cooperative system can suffice. CIoT basically refers to the interconnection of various systems and subsystems which are themselves supporting CIoT in a localized area such as a WSN. We in this work refer to CIoT with respect to its constricted scope pertaining to a single WSN consisting of several IoT nodes. Let an ADSTBC-based CIoT system constitutes of $R + 2$ nodes, where the transmitter, receiver, and each of the R relay nodes are designed for single antenna system as shown in Fig. 1. It is assumed that the destination is a cognitive element (CE), which not only perceives the information coming from different CIoT source nodes one at a time, duly relayed by intermediate CIoT relay nodes in the form of ADSTBC but also takes a decision accordingly. The information bits from a particular CIoT source node are binary phase shift keying (BPSK) modulated as $\mathbf{s} = [s_1, ..., s_N]'$ in the codebook $\{\mathbf{s}_1, ..., \mathbf{s}_L\}$ where $s_N = 2b_N - 1$ broadcasted to the relays with average power $P_1 []'$. At the CIoT relay nodes, the received signal is decoded. In selective DF, only those CIoT relay nodes participate in the relaying phase, which have successfully decoded the information and have received signal to noise ratio (SNR) greater than a threshold. The decoded and re-encoded symbol vector $\mathbf{s}_k$ is transmitted by the kth CIoT relay node to the destination (CE).

The transmitted signal $\mathbf{t}_k$ from kth CIoT relay node is a combination of a certain linear dispersion (LD) matrix $\mathbf{A}_k$ and the re-encoded signal $\mathbf{r}$.

$$\mathbf{t}_k = \sqrt{P_2}\mathbf{A}_k\mathbf{r} \tag{1}$$

The dispersion matrix $\mathbf{A}_k$ distributes the N symbols among T channel uses. The LD matrices used in this paper are represented using the theory of unitary matrices and frame theory [23].

2.1 Received Signal at the Destination

In the CIoT systems, there may be occasions when the distributed CIoT relay nodes incur timing delays and thus lead to an asynchronous scenario. These delays cause degradation of diversity the DSTBC system at the destination. The received signal $\mathbf{y}(t)$ at time t, at the destination with delayed symbols, may be given by

$$\mathbf{y} = \sum_{k=1}^{R} \sqrt{P_2}\mathbf{A}_k\mathbf{r}(t - \delta_k)g_k + \mathbf{w} \tag{2}$$

$\mathbf{w}$ is the additive white Gaussian noise with variance N_0. δ_k is the delay observed for the kth relay at the receiver relative to the first relay. g_k is the channel fading coefficient for the kth CIoT to CE link. The delay profile can be estimated from the arrival of the pilot signals. The CE is synchronized to the first relay, i.e., $\delta_1 = 0$ and $\delta_{max} \triangleq \max\{\delta_1, \ldots, \delta_R\}$. Making the codes delay tolerant, we introduce the following structure of the received signal at the destination as

$$\mathbf{y} = \sum_{k=1}^{R} \sqrt{P_2}\Big(\widehat{\mathbf{H}}\mathbf{D}(\delta_k)\mathbf{A}_k\mathbf{s}\Big) + \mathbf{w} \tag{3}$$

where $\mathbf{H} = \big[g_1, \ldots, g_R\big]$, $\widehat{\mathbf{H}} = \mathbf{H} \otimes \mathbf{I}$, $\otimes$ denotes the kronecker product and $\mathbf{I}$ is an identity matrix of size $T \times T$, $\mathbf{D}(\delta_k) = \begin{bmatrix} \mathbf{O}_{\delta_k \times T} \\ \mathbf{I}_T \\ \mathbf{O}_{(\delta_{max} - \delta_k) \times T} \end{bmatrix}$, $\mathbf{O}$ denote a zero matrix.

3 TWDP Channel

TWDP channels are perfect small-scale fading models to depict the channel behavior in small local area as is the case for the IoT. Owing to the non availability of the exact closed expression, an approximate probability density function (PDF) for the instantaneous SNR is given as [24]

$$f_\gamma(x) = \frac{\kappa}{2\bar{\gamma}} \sum_{i=1}^{L} \sum_{j=0}^{1} a_i e^{-P_{2i-j} - \frac{\kappa x}{\bar{\gamma}}} I_0\left(2\sqrt{X_{i,j}}\right) \tag{4}$$

where $\kappa = K + 1$, where K denotes the total specular power to the diffuse power, and $\bar{\gamma}$ is the average SNR of all relay to destination links. $P_{2j} = (\kappa - 1)(1 + \alpha_i)$, $P_{2i-1} = (\kappa - 1)(1 - \alpha_i)$, where $\alpha_i = \Delta \cos\left(\frac{\pi(i-1)}{2L-1}\right)$ and Δ is the relative strength of the specular components. L defines the order of the TWDP fading. a_i are constant coefficients related with the fading and first five values are tabulated in [22]. I_0 is the modified zeroth-order Bessel function of the first kind and $X_{i,j} = \frac{P_{2i-j}\gamma\kappa}{2\bar{\gamma}}$.

Joint PDF of the instantaneous SNR from the two relays through the TWDP channel can be expressed as the product of the marginal PDFs, owing to statistical independence between them

$$f_{x,y}(x, y) = f_x(x)f_y(y)$$
$$= \left(\frac{\kappa}{2\bar{\gamma}} \sum_{i=1}^{L} \sum_{j=0}^{1} a_i e^{-P_{2i-j} - \frac{\kappa x}{\bar{\gamma}}} I_0\left(2\sqrt{X_{i,j}}\right)\right)\left(\frac{\kappa}{2\bar{\gamma}} \sum_{k=1}^{L} \sum_{l=0}^{1} a_i e^{-P_{2k-1} - \frac{\kappa y}{\bar{\gamma}}} I_0\left(2\sqrt{Y_{k,l}}\right)\right) \tag{5}$$

The joint PDF in Eq (5) can be approximated using [Eq. (9.6.47)] of [25] as

$$f_{x,y}(x, y) = \frac{1}{4\bar{\gamma}^2}\kappa^2 \left(\sum_{i=1}^{L} \sum_{j=0}^{1} a_i e^{-P_{2i-j} - \frac{\kappa x}{\bar{\gamma}}} F_1\left(; 1; 0.25\left(2\sqrt{X_{i,j}}\right)^2\right)\right)$$
$$\left(\sum_{m=1}^{L} \sum_{l=0}^{1} a_m e^{-P_{2m-1} - \frac{\kappa y}{\bar{\gamma}}} F_1\left(; 1; 0.25\left(2\sqrt{Y_{m,l}}\right)^2\right)\right) \tag{6}$$

An approximate result for the above PDF may be obtained by expanding the confluent hypergeometric function to its first two values.

3.1 Upper Bound on CDF

The upper bound on the CDF may be obtained as follows [26]:

$$
\begin{aligned}
F_{up}(z) &= \int_0^{\frac{z}{2}} \int_y^{z-y} f_{x,y}(x,y)\,dxdy \\
&= \int_0^{\frac{z}{2}} \int_y^{z-y} \left(\frac{1}{\check{N}} \sum_{i=1}^{L} \sum_{j=0}^{1} a_i e^{-P_{2i-j}-\frac{\check{x}}{N}} I_0\left(2\sqrt{X_{i,j}}\right) \right) \\
&\quad \times \left(\frac{1}{\check{N}} \sum_{k=1}^{L} \sum_{l=0}^{1} a_i e^{-P_{2k-1}-\frac{\check{y}}{N}} I_0\left(2\sqrt{Y_{k,1}}\right) \right) dxdy \\
&= \frac{1}{4\check{N}} \sum_{i=1}^{L} \sum_{m=1}^{L} \sum_{j=0}^{1} \sum_{l=0}^{1} \frac{1}{24} e^{-\frac{z-NP_{2i-j}+NP_{2m-1}}{\check{\gamma}}} N_i a_m (N N ((-1 + e^{\frac{\check{z}}{N}}) N - z^\vee) + (6(-1 + e^{\frac{\check{z}}{N}}) N^2 \\
&\quad - 6 N^\vee - 3z^{2\vee}) P_{2m-1}) + P_{2i-j}(N(2(-1+e^{\frac{\check{z}}{N}}) N^2 - 2N^\vee - z^{2\cdot 2}) \\
&\quad + 2(6(-1 + e^{\frac{\check{z}}{N}}) N^3 - 6N^2 z^\vee \\
&\quad - 3N^{2\cdot 2} - z^{3\cdot 3}) P_{2m-1}))
\end{aligned}
\tag{7}
$$

We obtain the upper bound on the joint PDF by differentiating Eq. (7) w.r.t z as

$$
\begin{aligned}
f_{ub}(z) &= \sum_{i=1}^{L} \sum_{m=1}^{L} \sum_{j=0}^{1} \sum_{l=0}^{1} \frac{1}{96\bar{Y}^4} (e^{-\frac{z\kappa + \bar{\gamma}P_{2i-j} + \bar{\gamma}P_{2m-1}}{\bar{\gamma}}} \bar{\gamma} a_i a_m (3\bar{\gamma}(4\bar{\gamma}(-\kappa + e^{\frac{z\kappa}{\bar{Y}}}\kappa) + (-6\bar{\gamma}\kappa + 6e^{\frac{z\kappa}{\bar{Y}}}\bar{\gamma}\kappa \\
&\quad - 6z\kappa^2)P_{2m-1}) + P_{2i-j}(3\bar{\gamma}(-2\bar{\gamma}\kappa + 2e^{\frac{z\kappa}{\bar{Y}}}\bar{\gamma}\kappa - 2z\kappa^2) + 2(-6\bar{\gamma}^2\kappa + 6e^{\frac{z\kappa}{\bar{Y}}}\bar{\gamma}^2\kappa \\
&\quad - 6\bar{\gamma}z\kappa^2 - 3z^2\kappa^3)P_{2m-1})) - e^{-\frac{z\kappa + \bar{\gamma}P_{2i-j} + \bar{\gamma}P_{2m-1}}{\bar{\gamma}}} \bar{\gamma} a_i a_m (3\bar{\gamma}(4\bar{\gamma}((-1 + e^{\frac{z\kappa}{\bar{Y}}})\bar{\gamma} - z\kappa) \\
&\quad + (6(-1 + e^{\frac{z\kappa}{\bar{Y}}})\bar{\gamma}^2 - 6\bar{\gamma}z\kappa - 3z^2\kappa^2)P_{2m-1}) + P_{2i-j}(3\bar{\gamma}(2(-1 + e^{\frac{z\kappa}{\bar{Y}}})\bar{\gamma}^2 - 2\bar{\gamma}z\kappa \\
&\quad - z^2\kappa^2) + 2(6(-1 + e^{\frac{z\kappa}{\bar{Y}}})\bar{\gamma}^3 - 6\bar{\gamma}^2 z\kappa - 3\bar{\gamma}z^2\kappa^2 - z^3\kappa^3)P_{2m-1})))
\end{aligned}
\tag{8}
$$

Assuming the signals from the two relays arrive at the destination via two resolvable paths, the PDF of the delay profile is given by $\dfrac{1}{\left(\delta_{max} - \delta_{min}\right)^2}$ [27]. The overall PDF of the ADSTBC system is given as

$$
\begin{aligned}
f_{sys}(z) &= \frac{1}{(\delta_{max} - \delta_{min})^2} \sum_{i=1}^{L} \sum_{m=1}^{L} \sum_{j=0}^{1} \sum_{l=0}^{1} \frac{1}{96\acute{Y}^4} (e^{-\frac{z\kappa + \bar{\gamma}P_{2i-j} + \bar{\gamma}P_{2m-1}}{\bar{\gamma}}} \bar{\gamma} a_i a_m (3\bar{\gamma}(4\bar{\gamma}(-\kappa + e^{\frac{z\kappa}{\bar{Y}}}\kappa) \\
&\quad + (-6\bar{\gamma}\kappa + 6e^{\frac{z\kappa}{\bar{Y}}}\bar{\gamma}\kappa - 6z\kappa^2)P_{2m-1}) + P_{2i-j}(3\bar{\gamma}(-2\bar{\gamma}\kappa + 2e^{\frac{z\kappa}{\bar{Y}}}\bar{\gamma}\kappa - 2z\kappa^2) + 2(-6\bar{\gamma}^2\kappa \\
&\quad + 6e^{\frac{z\kappa}{\bar{Y}}}\bar{\gamma}^2\kappa - 6\bar{\gamma}z\kappa^2 - 3z^2\kappa^3)P_{2m-1})) - e^{-\frac{z\kappa + \bar{\gamma}P_{2i-j} + \bar{\gamma}P_{2m-1}}{\bar{\gamma}}} \bar{\gamma} a_i a_m (3\bar{\gamma}(4\bar{\gamma}((-1 \\
&\quad + e^{\frac{z\kappa}{\bar{Y}}})\bar{\gamma} - z\kappa) + (6(-1 + e^{\frac{z\kappa}{\bar{Y}}})\bar{\gamma}^2 - 6\bar{\gamma}z\kappa - 3z^2\kappa^2)P_{2m-1}) + P_{2i-j}(3\bar{\gamma}(2(-1 \\
&\quad + e^{\frac{z\kappa}{\bar{Y}}})\bar{\gamma}^2 - 2\bar{\gamma}z\kappa - z^2\kappa^2) + 2(6(-1 + e^{\frac{z\kappa}{\bar{Y}}})\bar{\gamma}^3 - 6\bar{\gamma}^2 z\kappa - 3\bar{\gamma}z^2\kappa^2 - z^3\kappa^3)P_{2m-1}))
\end{aligned}
\tag{9}
$$

4 Performance Analysis

4.1 Average Bit Error Rate

The average bit error rate (ABER) can be obtained by averaging the received error conditioned of the system w.r.t. the PDF of the system.

$$P_e = \int_0^\infty P_e(z) f_{sys}(z) dz \tag{10}$$

$P_e(z)$ is conditioned on the fading channel. An approximate expression for the conditional error is given by [28] as

$P_e(z) = \frac{1}{6} e^{-z} + \frac{1}{2} e^{\frac{-4z}{3}}$ for BPSK modulation.

$$
\begin{aligned}
P_e = \int_0^\infty &\left(\frac{1}{6} e^{-z} + \frac{1}{2} e^{\frac{-4z}{3}} \right) \frac{1}{(\delta_{max} - \delta_{min})^2} \sum_{i=1}^L \sum_{m=1}^L \sum_{j=0}^1 \sum_{l=0}^1 \frac{1}{96\bar{\gamma}^4} (e^{-\frac{z\kappa + \bar{\gamma}P_{2i-j} + \bar{\gamma}P_{2m-1}}{\bar{\gamma}}} \bar{\gamma} a_i a_m (3\bar{\gamma}(4\bar{\gamma}(-\kappa \\
&+ e^{\frac{z\kappa}{\bar{\gamma}}} \kappa) + \left(-6\bar{\gamma}\kappa + 6e^{\frac{z\kappa}{\bar{\gamma}}} \bar{\gamma}\kappa - 6z\kappa^2 \right) P_{2m-1}) + P_{2i-j} (3\bar{\gamma}(-2\bar{\gamma}\kappa + 2e^{\frac{z\kappa}{\bar{\gamma}}} \bar{\gamma}\kappa - 2z\kappa^2) \\
&+ 2 \left(-6\bar{\gamma}^2\kappa + 6e^{\frac{z\kappa}{\bar{\gamma}}} \bar{\gamma}^2\kappa - 6\bar{\gamma}z\kappa^2 - 3z^2\kappa^3 \right) P_{2m-1})) \\
&- e^{-\frac{z\kappa + \bar{\gamma}P_{2i-j} + \bar{\gamma}P_{2m-1}}{\bar{\gamma}}} \bar{\gamma} a_i a_m (3\bar{\gamma}(4\bar{\gamma}(\left(-1 + e^{\frac{z\kappa}{\bar{\gamma}}} \right) \bar{\gamma} - z\kappa) + (6 \left(-1 + e^{\frac{z\kappa}{\bar{\gamma}}} \right) \bar{\gamma}^2 - 6\bar{\gamma}z\kappa \\
&- 3z^2\kappa^2) P_{2m-1}) + P_{2i-j} (3\bar{\gamma}(2 \left(-1 + e^{\frac{z\kappa}{\bar{\gamma}}} \right) \bar{\gamma}^2 - 2\bar{\gamma}z\kappa - z^2\kappa^2) + 2(6 \left(-1 + e^{\frac{z\kappa}{\bar{\gamma}}} \right) \bar{\gamma}^3 \\
&- 6\bar{\gamma}^2 z\kappa - 3\bar{\gamma}z^2\kappa^2 - z^3\kappa^3) P_{2m-1})) dz \\
= &\frac{1}{96\bar{\gamma}^4 (\delta_{max} - \delta_{min})^2} \sum_{i=1}^L \sum_{m=1}^L \sum_{j=0}^1 \sum_{l=0}^1 e^{-P_{zi-j} - P_{2m-1}} \kappa^2 a_i a_m (2\bar{\gamma}^4 (\frac{1}{(\bar{\gamma} + \kappa)^2} + \frac{27}{4\bar{\gamma} + 3\kappa}) \\
&+ (\frac{3\bar{\gamma}^4\kappa}{(\bar{\gamma} + \kappa)^3} + \frac{9\bar{\gamma}\kappa}{(\frac{4}{3} + \frac{\kappa}{\bar{\gamma}})^3}) P_{2m-1} + \kappa P_{2i-j} (\bar{\gamma}^4 (\frac{1}{(\bar{\gamma} + \kappa)^3} + \frac{81}{(4\bar{\gamma} + 3\kappa)^3}) + (\frac{2\bar{\gamma}^4\kappa}{(\bar{\gamma} + \kappa)^4} + \frac{6\kappa}{(\frac{4}{3} + \frac{\kappa}{\bar{\gamma}})^4}) P_{2m-1})) \tag{11}
\end{aligned}
$$

This is the closed-form equation for the ABER of the system.

5 Results

In this section, we discuss the error performance of the ADSTBC-based CIoT system in the TWDP fading channel. Results pertaining to the ABER are plotted. It is assumed that destination has channel and delay information, where the latter is fed back to the CIoT relay nodes. Selective decode and forward (DF) protocol is used to transmit signals from the relays to the destination where only those nodes participate in the relaying, which have successfully decoded the signal received from the source and have instantaneous SNR greater than a threshold. It is assumed that the maximum delay incurred by the CIoT nodes to destination links is two symbol duration. TWDP channel is characterized by two factors; K and Δ. The former
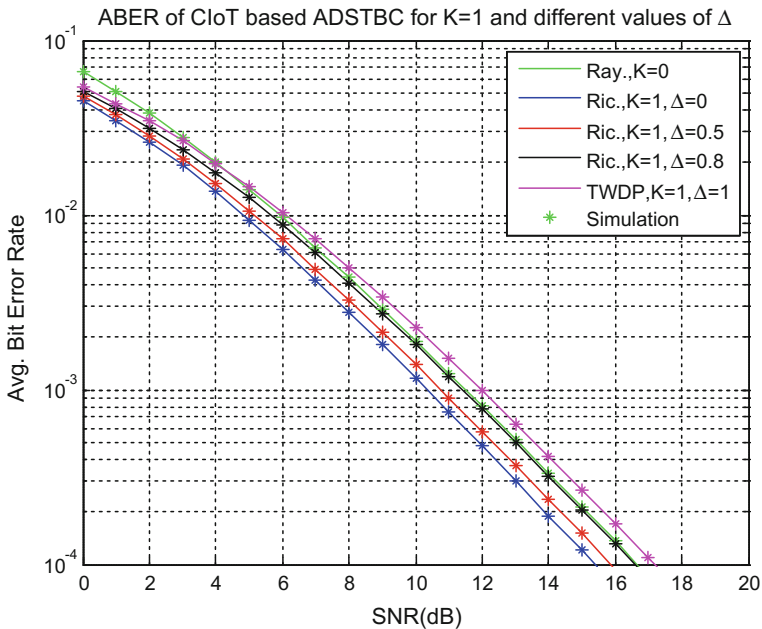
Fig. 2 ABER for ADSTBC-based CIoT system for K = 1

denotes the total specular power to the diffuse power, whereas the latter refers to the relative strength of the specular components. Monte Carlo simulation is done to verify the analytical results. Figure 2 depicts the ABER for different values of K. The TWDP fading encompasses the Rayleigh and Rician fading as special cases. By decreasing the value of K to zero, one approaches scenario closer to Rayleigh, whereas Rician fading can be modeled by reducing the value of Δ to zero. ABER curves for the above mentioned fading channels have been plotted along with other cases of TWDP fading. In Fig. 2, we plot the ABER curves for K = 1 and differing values of Δ w.r.t. the Rayleigh case obtained by putting K = 0. It is observed that as we decrease the value of Δ, i.e., move closer to the Rician type, the ABER reduces. By decreasing Δ from 0.8 to 0.5, we observe a gain of 0.5 dB for 10 dB average SNR for ABER of 10^{-4}. We observe that for $\Delta = 1$, i.e., the TWDP case is depicting the worst-case scenario.

In Fig. 3, ABER results for K = 3 and different values of Δ are plotted. We observe from the figure that as we decrease the values of Δ from 1 to 0, the performance improves. Comparing Figs. 2 and 3, we observe that this improvement becomes more pronounced when K increases from 1 in Figs. 2 and 3.

Figure 4 illustrates the ABER results for K = 5 and various values of Δ. We see that the performance improves while decreasing Δ. Also, comparative analyses of Figs. 2, 3, and 4 make clear that as K increases the improvement in the performance
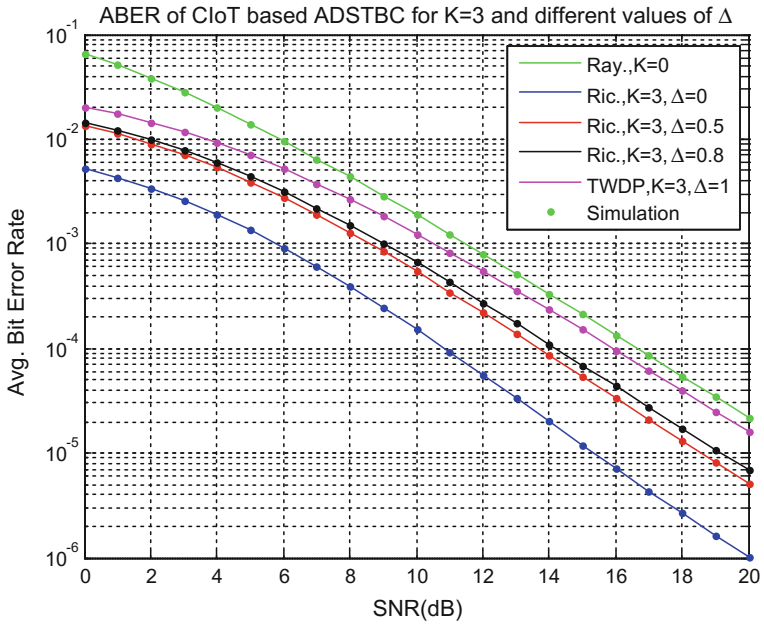
Fig. 3 ABER for ADSTBC-based CIoT system for $K = 3$
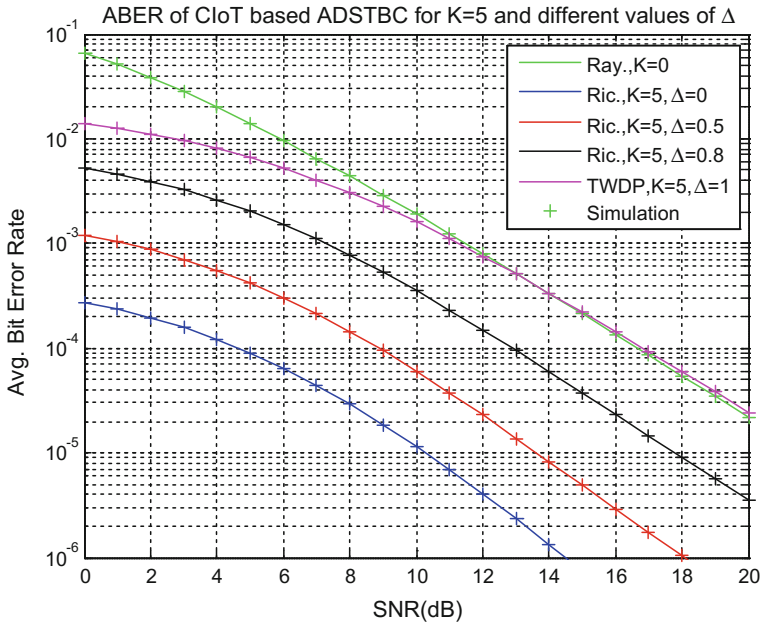
Fig. 4 ABER for ADSTBC-based CIoT system for $K = 1$

is registered in the form of coding gain. We also observe that for K = 5, for SNR below 13 dB, TWDP fading shows better results as compared to the Rayleigh fading.

6 Conclusions

In this work, we study an cooperation-based CIoT system consisting of entities which relay information for each other to an entity known as cognitive element (CE) which perceives the received information and takes suitable decision and action. We consider such cases of WSN where high degree of reliability is required in cooperation-based CIoT. Asynchronous distributed space-time block codes using selective DF protocol for cooperative communication are suitable candidate for providing diversity benefits in asynchronous environments. To cater high fading scenario, TWDP fading channels are assumed for communication which have Rayleigh and Rician fading as special cases. The error performance of proposed system is analyzed using ABER. From the results, we observe that the performance of the system in TWDP channel improves by increasing the values of K and/or by decreasing the value of Δ.

References

1. Ashton, K.: That 'internet of things' thing in the real world, things matter more than ideas. RFID J. **1** (2009)
2. Gao, Y., Qin, Z., Feng, Z., Zhang, Q., Holland, O., Dohler, M.: Scalable and reliable IoT enabled by dynamic spectrum management for M2M in LTE-A. IEEE Internet Things J. **3**, 1135–1145 (2016)
3. Islam, S.M.R., Kwak, D., Kabir, M.H., Hossain, M., Kwak, K.S.: The internet of things for health care: a comprehensive survey. IEEE Access. **3**, 678–708 (2015)
4. Razafimandimby, C., Loscri, V., Vegni, A.M.: A neural network and IoT based scheme for performance assessment in internet of robotic things. In: 2016 IEEE First International Conference on Internet-of-Things Design and Implementation (IoTDI), Berlin, pp. 241–246 (2016)
5. Paul, A., Daniel, A., Ahmad, A., Rho, S.: Cooperative cognitive intelligence for internet of vehicles. IEEE Syst. J. **11**, 1249–1258 (2017)
6. Wu, Q., Ding, G., Xu, Y., Feng, S., Wang, J., Long, K.: Cognititive internet of things: a new paradigm beyond connection. IEEE Int. Things **1**, 129–143 (2014)
7. Zhang, M., Zhao, H., Wu, Q., Wei, W.: Cognititive internet of things: concepts and application example. Int. J. Comput. Sci. **9**, 151–158 (2012)
8. Macciuliene, M.: Power through things: following traces of collective intelligence in internet of things. Soc. Technol. **4**, 168–178 (2014)
9. Laneman, J.N., Wornell, G.W.: Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks. IEEE Trans. Inf. Theory **49**, 2415–2425 (2003)
10. Jing, Y., Hassibi, B.: Distributed Space-Time Coding in Wireless Relay Networks. IEEE Trans. Wirel. Commun. **5**, 3524–3536 (2006)
11. Bai, J., Sun, Y., Phillips, C.: CRRP: a cooperative relay routing protocol for IoT networks. In: 2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Valencia, pp. 1–6 (2016)
12. Nessa, A., Kadoch, M.: Joint network channel fountain schemes for machine-type communications over LTE-advanced. IEEE Internet Things J. **3**, 418–427 (2016)

13. Nair, A.K., Asmi, S., Gopakumar, A.: Analysis of physical layer security via co-operative communication in internet of things. Procedia Tech. **24**, 896–903 (2016)
14. Wei, S., Goeckel, D.L., Valenti, M.C.: Asynchronous cooperative diversity. IEEE Trans. Wirel. Commun. **5**, 1547–1557 (2006)
15. Wang, H.M.: Full-diversity uncoordinated cooperative transmission for asynchronous relay networks. IEEE Trans. Veh. Tech. **66**, 468–480 (2017)
16. Damen, M.O., Hammons, A.R.: Delay-tolerant distributed TAST codes for cooperative diversity. IEEE Trans. Inf. Theory **53**, 3755–3773 (2007)
17. Li, Y., Xia, X.G.: Full diversity distributed space-time trellis codes for asynchronous cooperative communications. In: International Symposium on Information Theory. pp. 911–915 (2005)
18. Wang, W., Zheng, F.C., Fitch, M.: Design of delay-tolerant space-time codes with limited feedback. IEEE Trans. Veh. Tech. **64**, 839–845 (2015)
19. Ning, H., Wang, Z.: Future internet of things architecture: like mankind neural system or social organization framework? IEEE Comm. Lett. **15**, 461–463 (2011)
20. Toyonaga, S., Kominami, D., Murata, M.: Virtual wireless sensor networks: adaptive brain-inspired configuration for internet of things applications. Sensors **16**, 1–27 (2016)
21. Rahmani, A.M., Thanigaivelan, N.K., Gia, T.N., Granados, J. Negash, B., Liljeberg, P., Tenhunen, H.: Smart e-health gateway: bringing intelligence to internet of things based ubiquitous healthcare systems. In: 2015 12th Annual IEEE Consumer Communications and Networking Conference (CCNC). pp. 826–834 (2015)
22. Durgin, G.D., Rappaport, T.S., de Wolf, D.A.: New analytical models and probability density functions for fading in wireless communications. IEEE Trans. Commun. **50**, 1005–1015 (2002)
23. Heath Jr., R.W., Paulraj, A.J.: Linear dispersion codes for MIMO systems based on frame theory. IEEE Trans. Signal Process. **50**, 2429–2441 (2002)
24. Singh, D., Joshi, H.D.: BER performance of SFBC OFDM system over TWDP fading channel. IEEE Commun. Lett. **20**, 2426–2429 (2016)
25. Gradshteyn, I.S., Ryzhik, I.M.: Tables of Integrals, Series and Products, 6th edn. Academic Press, San Diego, CA, USA (2000)
26. Eddaghel, M.M., Mannai, U.N., Chen, G.J., Chambers, J.A.: Outage probability analysis of an amplify-and-forward cooperative communication system with multi-path channels and max-min relay selection. IET Commun. **7**, 408–416 (2013)
27. Charash, U.: Reception through Nakagami fading multipath channels with random delays. IEEE Trans. Commun. **27**, 657–670 (1979)
28. Chiani, M., Dardari, D., Simon, M.K.: New exponential bounds and approximations for the computation of error probability in fading channels. IEEE Trans. Wirel. Commun. **2**, 840–845 (2003)

Wavefunctions and Optical Gain in In$_{0.3}$Ga$_{0.7}$As/GaAs$_{0.4}$Sb$_{0.6}$ Type-II Double Quantum Well Nanoheterostructure Under External Uniaxial Strain

Amit Kumar Singh, Amit Rathi, Md. Riyaj and P. A. Alvi

Abstract Variations in wavefunction confinement under external uniaxial strain are observed to affect the optical gain obtained in type-II quantum well nanodimension heterostructures. This paper reports the wavefunctions and optical gain realized in In$_{0.3}$Ga$_{0.7}$As/GaAs$_{0.4}$Sb$_{0.6}$ type-II double QW heterostructure under uniaxial strain along [001]. Energy bands, wavefunctions of confinement states in the structure and optical gain of the heterostructure under electromagnetic field perturbation are presented. The 6×6 k·p Hamiltonian matrix is considered, and Luttinger–Kohn model has been applied for the electronic band structure calculations. Optical gain spectra of the double QW nanoheterostructure under external uniaxial strain of 1, 2 and 5 GPa, respectively, is calculated. The optical gain curve shows a significant improvement in gain under external uniaxial strain along [001] at 300 K. For a charge carrier injection of 8×10^{12}/cm^2, the optical gain is 9170 in x polarization. The heterostructure is seen to be operating in the energy range of 0.65–0.8 eV (1549–1907 nm). Thus, a wide range wavelength tuning can be realized.

Keywords Heterostructure · Type-II · Optical gain · InGaAs · GaAsSb

A. K. Singh (✉) · A. Rathi · Md. Riyaj
Department of Electronics & Communication Engineering,
Manipal University Jaipur, Jaipur 303007, Rajasthan, India
e-mail: amitkumarsingh89@gmail.com

A. Rathi
e-mail: amitrathi1978@gmail.com

Md. Riyaj
e-mail: roziriyaj@gmail.com

P. A. Alvi
Department of Physics, Banasthali University, Vanasthali 304022, Rajasthan, India
e-mail: drpaalvi@gmail.com

1 Introduction

Advancements in fabrication techniques, research in novel compounds and alloys and improved designs have led to the progress in lasers that can find application in diverse areas. Lasers based on III–V semiconductors operating in SWIR range (0.9–1.7 μm) are widely utilized in areas such as advanced imaging, surveillance systems, range finding systems, spectroscopy, etc. Several type-II InGaAs/GaAsSb nanodimension heterostructures have recently been modeled and studied extensively for application in IR region. InAs/AlSb/GaSb type-II double quantum wells under electric and magnetic fields have been presented [1]. Tobias Gruendl et al. have investigated InGaAs/GaAsSb type-II superlattice structures [2]. Stephan Sprengel et al. have demonstrated InP-based InGaAs/GaAsSb type-II QW nanoheterostructure lasers for wavelength range up to 2.7 μm [3]. Electrically and optically pumped InP-based type-II mid-infrared lasers have been demonstrated [4]. Chia-Hao Chang et al. have investigated InP-based InGaAs/GaAsSb W-type quantum wells [5]. Mid-infrared InP-based "M"-type InGaAs/GaAsSb nanoscale lasers have been demonstrated [6]. In [7], optical gain tuning by external pressure in type-II InGaAs/GaAsSb nanoheterostructure has been studied. Optically pumped type-II "W" InGaAs/GaAsSb QW nanoheterostructure on InP substrate has been investigated by Pan et al. [8]. k·p theory has been used to compute the wavefunction overlap and transition wavelength in InGaAs/GaAsSb quantum well heterostructures [9]. Optical gain in InGaAs/GaAsSb type-II nanoheterostructures has been studied using k·p theory under uniaxial strain [10]. Optical properties of InGaAs/GaAsSb heterostructures grown by MOVPE with varying InGaAs layer thickness has been investigated [11]. InP-based type-II "W"-type InGaAs/GaAsSb quantum well nanoheterostructures have been theoretically studied using k·p theory [12]. InGaAs/GaAsSb type-II QW heterostructures have been modeled and studied using k.p model by Baile Chen et al. [13]. Optical properties of InGaAs/GaAsSb superlattice structures have been studied [14]. E. V. Bogdanov et al. have experimentally reported polarization tuning in AlGaAs/GaAsP heterostructures [15]. Electroluminescence in AlGaAs/GaAsP heterostructures has been studied under uniaxial compression [16]. Computational and experimental studies on tensile-strained AlGaAs/GaAsP heterostructures under uniaxial stress have been presented [17]. E.V. Andreev et al. have reported electroluminescence in AlGaAs/GaAsP heterostructures under uniaxial compression [18]. Optical properties of GaAsP/AlGaAs heterostructure subjected to external uniaxial strain have been studied by utilizing k·p theory [19].

The objective of this paper is to report band structure, wavefunctions, and optical gain in $In_{0.3}Ga_{0.7}As/GaAs_{0.4}Sb_{0.6}$ type-II double quantum well nanoheterostructure under external uniaxial strain along [001]. The optical gain in the QW nanoheterostructure subjected to external uniaxial strains of 1, 2 and 5 GPa along [001] is computed. In the following sections, design and theory of the $In_{0.3}Ga_{0.7}As/GaAs_{0.4}Sb_{0.6}$ heterostructure are presented following which the simulation results are discussed. Finally, the inferences from the work are presented.

2 Structure and Theory

The designed $In_{0.3}Ga_{0.7}As/GaAs_{0.4}Sb_{0.6}$ heterostructure under study is composed of two of $GaAs_{0.4}Sb_{0.6}$ hole wells sandwiched between three $In_{0.3}Ga_{0.7}As$ electron wells. $In_{0.3}Ga_{0.7}As$ and $GaAs_{0.4}Sb_{0.6}$ layers are each 2 nm with the entire structure being of length 12 nm along the growth direction. Band structure with valence subbands, heavy hole (HHB) and light hole (LHB) is shown in Fig. 1. The type-II heterostructure is based on InP substrate.

Effective mass equation for electrons and holes in perturbed periodic structure is discussed in [20, 21]. Valence band structure can be represented by a block diagonalized system Hamiltonian that includes strain effects as given in (1) [22].

$$H_{6\times6}(k) = \begin{pmatrix} H_{3\times3}^{+} & 0 \\ 0 & H_{3\times3}^{-} \end{pmatrix} \tag{1}$$

where $H_{3\times3}^{+}$ and $H_{3\times3}^{-}$ can be further expanded as (2) where $\mathbb{Z} = -$ or $+$ correspond to lower and upper matrix blocks [23, 24].

$$H_{3\times3}^{\mathbb{Z}} = - \begin{pmatrix} P+Q-V_h(Z) & R(k) \pm iS(k) & \sqrt{2}R(k) \pm \dfrac{i}{\sqrt{2}}S(k) \\[2mm] R(k) \pm iS(k) & P-Q-V_h(Z) & \sqrt{2}Q \pm i\sqrt{\dfrac{3}{2}}S(k) \\[2mm] \sqrt{2}R(k) \pm \dfrac{i}{\sqrt{2}}S(k) & \sqrt{2}Q \mp i\sqrt{\dfrac{3}{2}}S(k) & P+\delta_{so}-V_h(Z) \end{pmatrix} \tag{2}$$

Fig. 1 Band diagram of $In_{0.3}Ga_{0.7}As/GaAs_{0.4}Sb_{0.6}$ double QW heterostructure
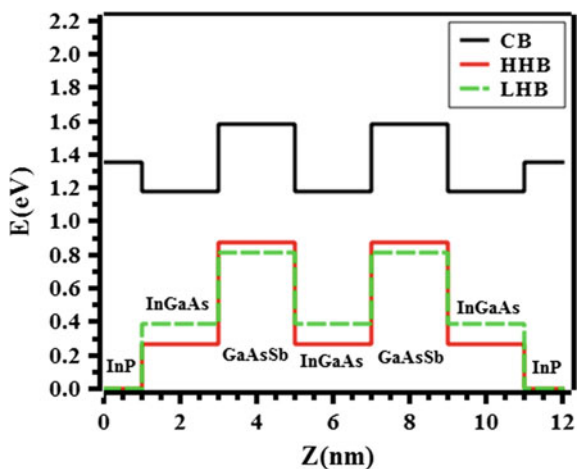
In (2), $V_h(Z)$ denotes the unstrained valence band edge, δ_{so} is the spin–orbit split-off energy, $P = P(k) + P(\varepsilon)$, $Q = Q(k) + Q(\varepsilon)$, $R(k)$ and $S(k)$ given as in Eq. 3a, 3b, and 3c [23, 25].

$$P(k) = \left(\frac{\hbar^2}{2m}\right)\gamma_1\left(k_t^2 + k_z^2\right), \qquad P(\varepsilon) = -a_v\left(\varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz}\right) \tag{3a}$$

$$Q(k) = \left(\frac{\hbar^2}{2m}\right)\gamma_2\left(k_t^2 - 2k_z^2\right), \quad Q(\varepsilon) = -\frac{b}{2}\left(\varepsilon_{xx} + \varepsilon_{yy} - 2\varepsilon_{zz}\right) \tag{3b}$$

$$S(k) = \left(\frac{\hbar^2}{2m}\right)\sqrt{3}\left(\frac{\gamma_2 + \gamma_3}{2}\right)k_t^2, \quad R(k) = \left(\frac{\hbar^2}{2m}\right)2\sqrt{3}\gamma_3 k_t k_z \tag{3c}$$

Fermi's golden rule as given in [26, 27] is used for the numerical computation of optical gain coefficient of the nanoscale heterostructure.

$$G(\hbar\omega) = \mathbb{C}_O \sum_{\eta,\mathbb{Z}} \sum_{n,m} \int \left| \left(\xi \cdot M_{nm}^{\eta\mathbb{Z}}(k_t)\right) \right|^2 \times (\aleph_\omega)\frac{k_t dk_t}{2\pi}$$

$$\text{where } \aleph_\omega = \frac{\left(f_n^c(k_t) - f_{\mathbb{Z}m}^v(k_t)\right)\left(\frac{\gamma}{\pi}\right)}{\left(E_{\mathbb{Z},nm}^{cv}(k_t) - \omega\hbar\right)^2 + \gamma^2},$$

$$\mathbb{C}_O = \frac{\pi e^2}{nc\varepsilon\omega Lm_e^2}, \quad f_n^c(k_t) = \left(e^{\left(\frac{E_n^c(k_t)-F_c}{k_B T}\right)} + 1\right)^{-1}$$

$$\text{and } f_{\mathbb{Z}m}^v(k_t) = \left(e^{\left(\frac{E_{\mathbb{Z}m}^v(k_t)-F_v}{k_B T}\right)} + 1\right)^{-1} \tag{4}$$

In (4), ε is free space permittivity, ξ is optical electric field polarization vector, and $M_{nm}^{\eta\mathbb{Z}}(k_t)$ elements of momentum matrix. $f_{\mathbb{Z}m}^v(k_t)$ and $f_n^c(k_t)$ represent the Fermi–Dirac distribution functions. Various material parameters for simulation are obtained from [28]. Based on the simulations, Figs. 2 and 3 show the wavefunctions in $In_{0.3}Ga_{0.7}As/GaAs_{0.4}Sb_{0.6}$ type-II heterostructure. In the following section, simulation results are presented and discussed.

3   Simulation Results and Analysis

Optical gain calculation requires the Fermi–Dirac distribution functions, momentum matrix elements and other parameters. For the computation of wavefunctions in the QW heterostructure, 6×6 Luttinger–Kohn model and self-consistent computation are used. Interband transitions e1-h1 along with e1-h2 are chosen for the simulations.

From Fig. 2, it can be seen that the wavefunctions of the first energy level electron (e1 with spin up) are confined to the InGaAs region. Also, the wavefunctions of first energy level holes (h1-spin up and h2-spin up) are confined in the GaAsSb layer.

Fig. 2 Wavefunctions in
$In_{0.3}Ga_{0.7}As/GaAs_{0.4}Sb_{0.6}$
heterostructure
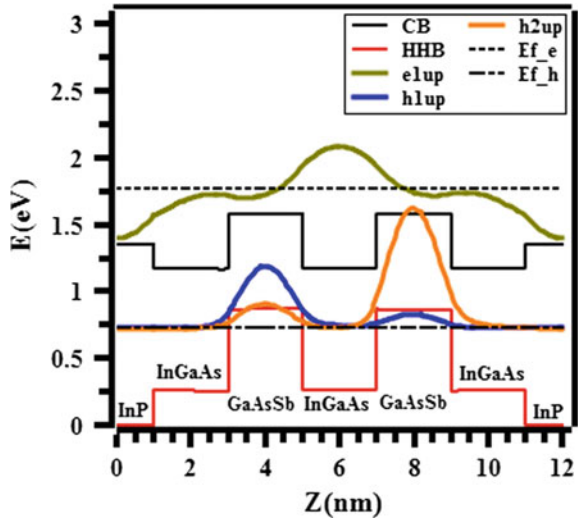
Fig. 3 Net wavefunctions in
$In_{0.3}Ga_{0.7}As/GaAs_{0.4}Sb_{0.6}$
heterostructure
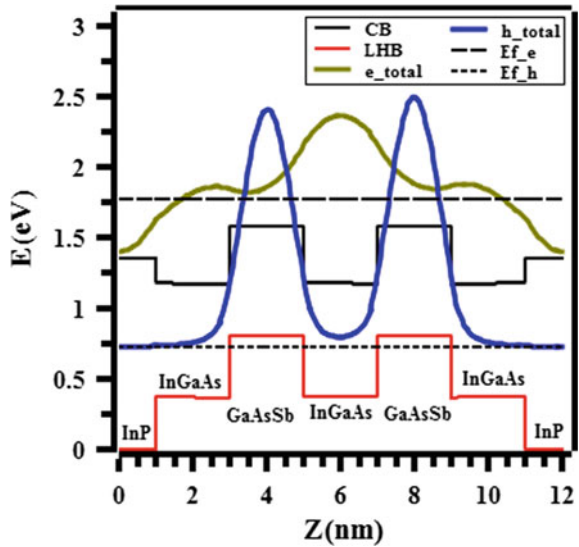
Figure 3 shows the net hole and electron wavefunctions along with the Fermi levels, conduction band and light hole bands in the $In_{0.3}Ga_{0.7}As/GaAs_{0.4}Sb_{0.6}$ heterostructure. From Fig. 3, it is established that the overall hole density maximizes only in the $GaAs_{0.4}Sb_{0.6}$ layer and the electron density maximizes only in the $In_{0.3}Ga_{0.7}As$ region. As a result, the optical transitions occur between electrons in the InGaAs layers and the holes present in the GaAsSb layers leading to a high optical gain.

In Fig. 4, the gain spectrum of $In_{0.3}Ga_{0.7}As/GaAs_{0.4}Sb_{0.6}$ nanoheterostructure in x polarization with respect to wavelength (λ) is shown. The optical gain is found

Fig. 4 Gain spectrum of $In_{0.3}Ga_{0.7}As/GaAs_{0.4}Sb_{0.6}$ nanoheterostructure in x polarization
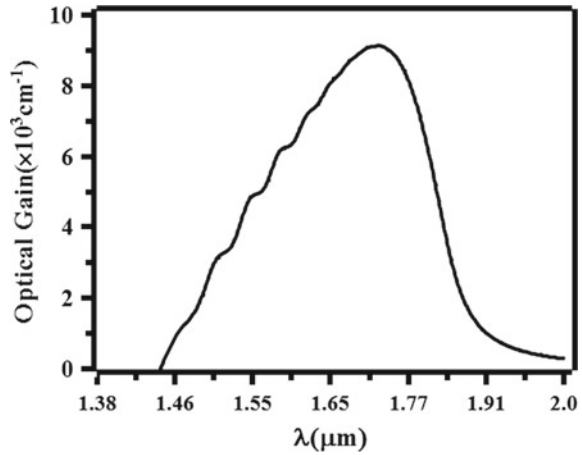
Fig. 5 Gain spectra of $In_{0.3}Ga_{0.7}As/GaAs_{0.4}Sb_{0.6}$ nanoheterostructure in x polarization versus energy
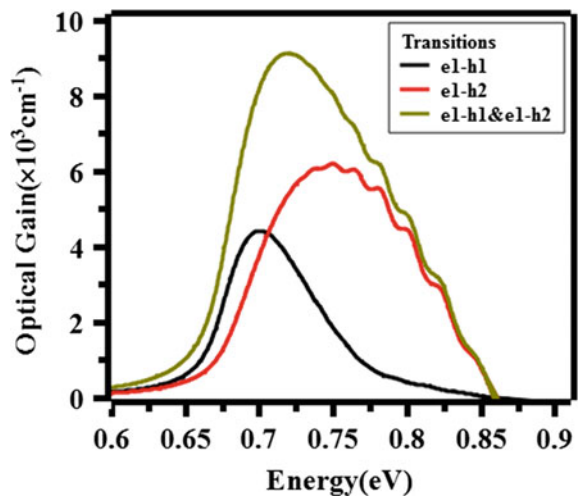
in the energy range of about 0.65–0.8 eV (1549–1907 nm). Figure 5 shows the optical gain spectrum versus emission energy curve of $In_{0.3}Ga_{0.7}As/GaAs_{0.4}Sb_{0.6}$ QW heterostructure. e1-h2 transition has a higher contribution as compared to the e1-h1 transition.

Figure 6 shows the optical gain in x polarization under varying external strain conditions of 1, 2 and 5 GPa at 300 K. The optical gain of the heterostructure under no external strain is 9170 in x polarization. A left shift along with a steady rise is observed in the optical gain spectra for increasing external strain conditions at a 2D carrier injection of $8 \times 10^{12}/cm^2$.

The optical gain shows a very significant improvement in gain under very high external uniaxial strain along [001]. Such a characteristic can be utilized to tune the response of the structure for application in the SWIR region.

Fig. 6 Gain spectra of $In_{0.3}Ga_{0.7}As/GaAs_{0.4}Sb_{0.6}$ nanoheterostructure as external field variations
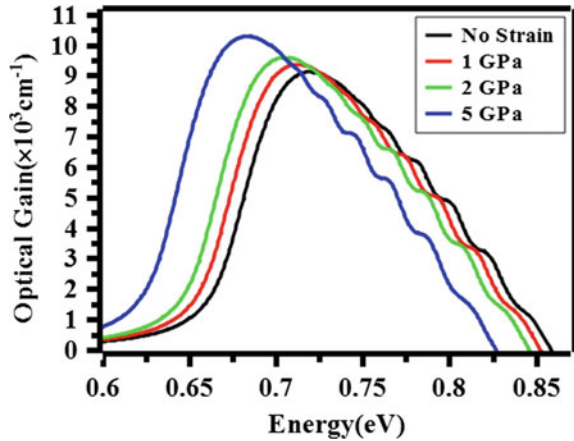
4 Conclusion

Optical gain in $In_{0.3}Ga_{0.7}As/GaAs_{0.4}Sb_{0.6}$ type-II heterostructure consisting of two hole and three electron wells under external electric field has been computationally examined. k·p Hamiltonian matrix is used for modeling the heterostructure, and Luttinger–Kohn model has been employed for wavefunction computations. Optical gain of the QW nanoheterostructure is investigated under external uniaxial strain. A left shift along with a steady rise is observed in the optical gain spectra when strain is increased. Based on the simulation results, the response of the InGaAs/GaAsSb heterostructure can be appropriately tweaked for applications in the SWIR region.

References

1. Lyo, S.K., Pan, W.: Excitons in coupled type-II double quantum wells under electric and magnetic fields: InAs/AlSb/GaSb. J. Appl. Phys. **118**(19), 195705 (2015)
2. Gruendl, T., Grasse, C., Sprengel, S., Vizbaras, K., Boehm, G., Meyer, R., Amann, M.C.: Type-II quantum wells for InP based mid-IR devices. In: Conference Proceedings of 'Mid Infrared Optoelectronics: Materials and Devices', pp. 115–116 (2012)
3. Sprengel, S., Veerabathran, G.K., Alexander, A., Köninger, A., Boehm, G., Grasse, C., Amann, M.C.: InP-based type-II heterostructure lasers for wavelengths up to 2.7 μm. Proc. SPIE **9382**, 93820U (2015)

4. Sprengel, S., Grasse, C., Wiecha, P., Andrejew, A., Gruendl, T., Boehm, G., Meyer, R., Amann, M.C.: InP-based type-II quantum-well lasers and LEDs. IEEE J. Sel. Top. Quantum Electron. **19**(4), 1900909 (2013)

5. Chang, C.H., Li, Z.L., Hong-Ting, L., Pan, C.H., Lee, C.P., Lin, G., Lin, S.D.: Low-threshold short-wavelength infrared InGaAs/GaAsSb 'W'-type QW laser on InP substrate. IEEE Photon. Technol. Lett. **27**(3), 225–228 (2015)

6. Chang, C.H., Li, Z.L., Pan, C.H., Hong-Ting, L., Lee, C.P., Lin, S.D.: Room-temperature mid-infrared "M"-type GaAsSb/InGaAs quantum well lasers on InP substrate. J. Appl. Phys. **115**(6), 063104 (2014)

7. Nirmal, H.K., Nisha, Y., Dalela, S., Rathi, A., Siddiqui, M.J., Alvi, P.A.: Tunability of optical gain (SWIR region) in type-II In0.70Ga0.30As/GaAs0.40 Sb0.60 nano-heterostructure under high pressure. Phys. E Low Dimens. Syst. Nanostruct. **80**, 36–42 (2016)

8. Pan, C.H., Chang, C.H., Lee, C.P.: Room temperature optically pumped 2.56-lasers with "W" type InGaAs/GaAsSb quantum wells on InP substrates. IEEE Photon. Technol. Lett. **24**(13), 1145–1147 (2012)

9. Chen, B., Jiang, W.Y., Holmes Jr, A.L.: Design of strain compensated InGaAs/GaAsSb type-II quantum well structures for mid-infrared photodiodes. Opt. Quant Electron **44**(3–5), 103–109 (2012)

10. Singh, A.K., Riyaj, M., Anjum, S.G., Nisha, Y., Rathi, A., Siddiqui, M.J., Alvi, P.A.: Anisotropy and optical gain improvement in type-II $In_{0.3}Ga_{0.7}As/GaAs_{0.4}Sb_{0.6}$ nano-scale heterostructure under external uniaxial strain. Superlattices Microstruct. **98**, 406–415 (2016)

11. Hu, J., Xu, X.G., Stotz, J.A.H., Watkins, S.P., Curzon, A.E., Thewalt, M.L.W., Matine, N., Bolognesi, C.R.: Type II photoluminescence and conduction band offsets of GaAsSb/InGaAs and GaAsSb/InP heterostructures grown by metal organic vapor phase epitaxy. Appl. Phys. Lett. **73**(19), 2799–2801 (1998)

12. Pan, C.H., Lee, C.P.: Design and modeling of InP-based InGaAs/GaAsSb type-II "W" type quantum wells for mid-Infrared laser applications. J. Appl. Phys. **113**(4), 043112 (2013)

13. Chen, B., Holmes, A.L., Khalfin, V., Kudryashov, I., Onat, B.M.: Modeling of the type-II InGaAs/GaAsSb quantum well designs for mid-infrared laser diodes by k·p method. Laser Technol. Def. Secur. VIII **8381**, 83810F. International Society for Optics and Photonics (2012)

14. Jin, C., QingQing, X., Chen, J.X.: Growth mechanism and optical properties of InGaAs/GaAsSb Su-perlattice structures. Sci. China Phys. Mech. Astron **58**(4), 1–5 (2015)

15. Bogdanov, E.V., Kissel, H., Kolokolov, K.I., Ya, N.: Minina: TM/TE polarization tuning and switching in tensile strained p-AlGaAs/GaAsP/n-AlGaAs heterostructures by uniaxial compression. Semicond. Sci. Technol. **31**(3) 035008 (2016)

16. Minina, N.Y., Bogdanov, E.V., Shirokov, S.S.: Uniaxial compression influence on valence sub-bands energy spectrum and electroluminescence in n-AlGaAs/GaAsP/p-AlGaAs diode structures. J. Phys. Conf. Ser. **377**(1), 012096. IOP Publishing (2012)

17. Bogdanov, E.V., Minina, N.Y., Tomm, J.W., Kissel, H.: Effect of uniaxial stress on electroluminescence, valence band modification, optical gain, and polarization modes in tensile strained p-AlGaAs/GaAsP/n-AlGaAs laser diode structures: numerical calculations and experimental results. J. Appl. Phys. **112**(9), 093113 (2012)

18. Andreev, E.V., Bogdanov, E.V., Kissel, H., Kolokolov, K.I., Minina, N.Y., Shirokov, S.S., Yunovich, A.E.: Electroluminescence and band structure in p-AlxGa1-xAs/GaAs1-yPy/n-AlxGa1-xAs under uniaxial compression. High Press. Res. **29**(4), 495–499 (2009)

19. Riyaj, M., Singh, A.K., Sandhya, K., Rathi, A., Alvi, P.A.: Optical properties of type-I GaAsP/AlGaAs nano-heterostructure under external uniaxial strain. In: AIP Conference Proceedings, vol. 1832, no. 1, p. 120022. AIP Publishing (2017)

20. Luttinger, J.M., Kohn, W.: Motion of electrons and holes in perturbed periodic fields. Phys. Rev. **97**(4) 869 (1955)

21. Luttinger, J.M.: Quantum theory of cyclotron resonance in semiconductors: General theory. Phys. Rev. **102**(4), 1030 (1956)

22. Chang, C.S., Chuang, S.L.: Modeling of strained quantum-well lasers with spin-orbit coupling. IEEE J. Sel. Topics Quantum Electron. **1**(2), 218–229 (1995)

23. Chuang, S.L.: Physics of Optoelectronic Devices. Wiley (1995)
24. Zhao, H., Arif, R.A., Ee, Y.K., Tansu, N.: Self-consistent analysis of strain-compensated InGaN–AlGaN quantum wells for lasers and light-emitting diodes. IEEE J. Quantum Electron. **45**(1), 66–78 (2009)
25. Sun, Y., Thompson, S.E., Nishida, T.: Strain effect in semiconductors: theory and device applications. Springer Science & Business Media (2009)
26. Harrison, P.: Quantum Wells, Wires and Dots: Theoretical and Computational Physics of Semiconductor Nanostructures. Wiley (2005)
27. Zory, P.S.: Quantum Well Lasers. Academic Press (1993)
28. Vurgaftman, I., Meyer, J.R., Ram-Mohan, L.R.: Band parameters for III–V compound semiconductors and their alloys. J. Appl. Phys. **89**(11), 5815–5875 (2001)

Design of Rectangular MIMO Antenna for Bluetooth and WLAN Applications to Reduce the Mutual Coupling

K. Vasu Babu and B. Anuradha

Abstract This article demonstrates a compact design of rectangular MIMO antenna that resonates at a frequency of 2.48 GHz for Bluetooth and WLAN applications. The antenna has a compact size of 50×40 mm^2 and spacing between the two rectangular patches is maintained as $0.05 \lambda_0$ which minimizes the effect of isolation (mutual coupling) between the antennas greatly 40 dB and improvement in the return loss of 43 dB at the resonant frequency. To improve the overall performance of diversity, MIMO antenna systems including the antenna parameters like channel capacity, capacity loss, total active reflection coefficient, envelope correlation coefficient, and total radiation efficiency are calculated. Due to the arrangement of minimal spacing between the antennas, resultant improvement in the directivity, impedance, and antenna radiation pattern of rectangular MIMO antenna is obtained. The rectangular MIMO antenna also exhibits an acceptable average peak gain of 6.044 dBi at that particular resonant frequency used in the Bluetooth and WLAN applications.

Keywords Rectangular MIMO antenna · Channel capacity · TARC · Capacity loss · Bluetooth

1   Introduction

MIMO technology plays an important role in the present wireless communication systems in the application areas of Wireless LAN, WiMAX, LTE, and satellite communication systems. Improvement of isolation is one of the methods used in hybrid coupling of E and H fields with a spacing between the antennas of $0.04 \lambda_0$ which produces an isolation of 50 dB [1]. The antenna is operated in frequencies between 6 and 16 GHz when edge-to-edge distance is maintained as $0.11 \lambda_0$ at resonant frequency

K. Vasu Babu (✉) · B. Anuradha
Department of ECE, Sri Venkateswara University, Tirupati, Andhra Pradesh, India
e-mail: vasubabuece@gmail.com

B. Anuradha
e-mail: anubhuma@yahoo.com

of 7.5 GHz [2]. Several methods have been implemented to reduce the isolation in the designs [3–5] at different operating frequencies. In [6], an edge-to-edge separating dimension of 0.097 λ_0 with an antenna size of 125 mm × 85 mm with higher than 15 dB isolation is maintained. The different types of decoupling techniques providing the reduction of mutual coupling between the design antennas [7, 8] have a very small size with S_{12} values of 32.5 and 29.1 dB which maintained a ρ value less than 0.1. In [9], sizes of antenna 22 mm × 36 mm with very low in size operated in UWB applications resonate at a frequency of 5.3 GHz with S_{11} and S_{12} produced values of 35 and 17 dB. Finally, in [10], wideband diversity MIMO which improves the impedance bandwidth 54.5% resonates at 3.5 GHz. In the proposed study, rectangular MIMO antenna is maintained with a size of 50 mm × 40 mm and consists of an overall area of 2500 mm^2, and edge-to-edge spacing maintained for low mutual coupling of 0.05 λ_0 produces a resonant frequency of 2.48 GHz in wireless LAN and Bluetooth.

2   Antenna Geometry and Analysis

The rectangular-shaped MIMO antenna is with a parameter ε_r (relative permittivity) = 4.4. The material FR-4 lossy is used for manufacturing of substrate. It consists of a 0.65 mm thickness and loss tangent of 0.02. The design dimensions are shown in Table 1 and comparison with other existing methods are shown in Table 2 (Fig. 1).

Table 1 Rectangular MIMO design parameters

Design parameters	L	W	a	b	c	d	e	f	g	h
Value (mm)	50	40	22	20	11	18	8	4	24	3

Table 2 Comparison with other existing methods

Published literature	Total occupied size (mm^2)	Operating frequency (GHz)	Edge-to-edge distance (λ_0)	Mutual coupling	Efficiency (%)
[1]	3600	2.5	0.04	50	85
[5]	5600	2.45	0.02	39	73
[6]	10625	0.9	0.09	35	82
[8]	4200	2.45	0.03	9.2	70.3
[10]	4200	2.45	0.04	29.1	74
Rectangular MIMO antenna	2000	2.48	0.05	40	87
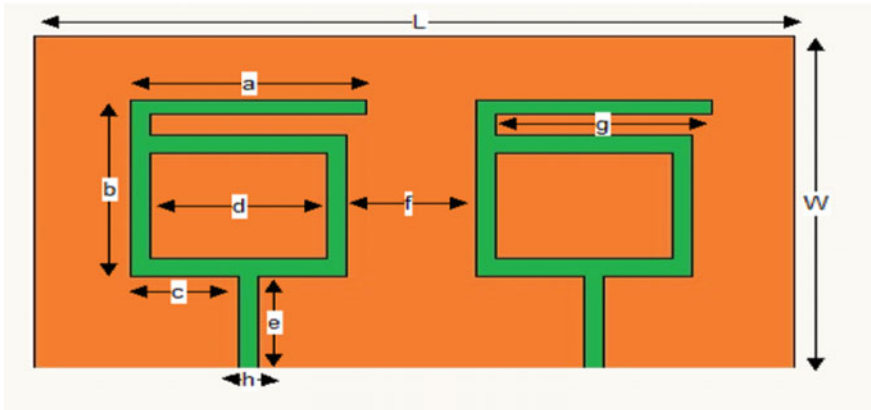
Fig. 1 Rectangular MIMO antenna

3 Multiple-Input Multiple-Output

3.1 The Total Active Reflective Coefficient (TARC)

TARC is defined as the combination of ratios of total reflected power square root divided by total incident power square root. Mathematically, it is represented by

$$TARC = \sqrt{\left|S_{11} + S_{12}e^{j\theta}\right|^2 + \left|S_{21} + S_{22}e^{j\theta}\right|^2} \Big/ \sqrt{2} \tag{1}$$

where θ value in the range of 0–2π.

3.2 The Envelope Correlation Coefficient (ECC)

ECC is defined as the analysis of scattering parameters (S-parameters) of S_{11} and S_{12} for a 2×2 MIMO system which is obtained using Eq. (2):

$$\rho = \frac{S_{11}^* S_{12} + S_{21}^* S_{22}^2}{\left(1 - S_{11}^2 - S_{21}^2\right)\left(1 - S_{22}^2 - S_{12}^2\right)} \tag{2}$$

3.3 The Channel Capacity

The parameters like signal-to-noise ratio (SNR), logarithmic function, number of antenna receivers (M), the transfer matrix of a given channel (H), Shannon channel capacity (C), and transpose of the channel matrix (H') are represented by Eq. (3):

$$C = \log_2\left(\det\left(1 + \frac{SNR}{M}HH'\right)\right) \qquad (3)$$

3.4 Capacity Loss

To obtain the capacity loss of a channel, in the case of large, signal-to-noise ratio is evaluated using Eq. (4):

where Ψ^R represents the correlation matrix of the receiving antenna.

$$C_{loss} = -\log_2 \det\left(\Psi^R\right) \qquad (4)$$

4 Results and Discussion

The parameter return loss (S_{11}) of a rectangular MIMO design is shown in Fig. 2 with a value of 43 dB. The mutual coupling or isolation of the designed antenna which greatly reduces the value of 40 dB at the resonant frequency 2.48 GHz used in the application areas like Bluetooth and WLAN is shown in Fig. 3. The comparison graph of S-parameters is depicted in Fig. 4. The parameter VSWR is maintained ≤2 at a particular resonant frequency of a rectangular MIMO antenna depicted in Fig. 5. Directivity of the rectangular design is improved to 6.044 dBi at the resonant frequency consisting of 2.48 GHz as shown in Fig. 6. For MIMO design, envelope correlation coefficient is shown in Fig. 7. Finally, co- and cross-polarization of the proposed design is shown in Fig. 8.

In the design of rectangular MIMO antenna, the correlation coefficient value is 0.0219, and the total reflection coefficient at this frequency is −28.56 dB. Similarly, capacity of a channel and channel capacity is obtained using the relations of Eqs. (3) and (4) with a value of 4.045 and 0.26 bits/s/Hz. To find the radiation pattern of rectangular MIMO antenna, one port of the antenna (radiator) is excited and the other port of the antenna (radiator) is terminated by a characteristic impedance of 50 ohms. The antenna radiation patterns in xz plane are omnidirectional, and in the planes of xy and xz, they are nearly self-complementarity structures. It provides a good diversity in the approach of scattering environment, due to the arrangement of symmetry of both the antenna elements. The parameters of return loss and mutual
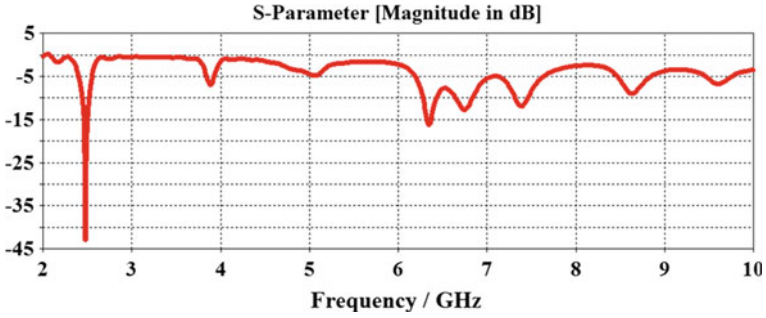
Fig. 2 Return loss (S_{11})
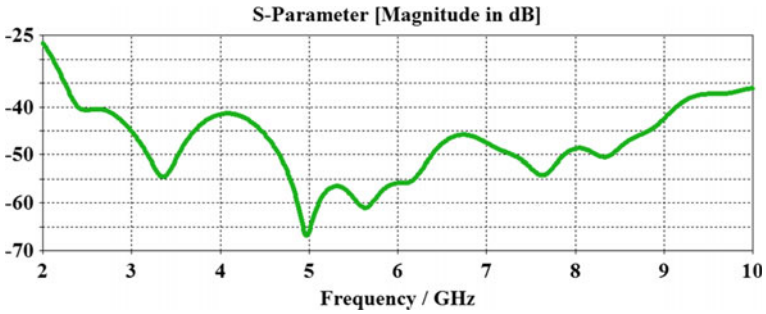
Fig. 3 Mutual coupling (S_{12})
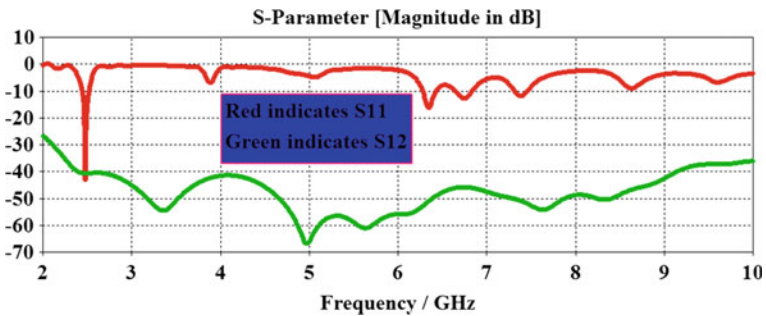
Fig. 4 S-parameters comparison

coupling are used to estimate the efficiency of an antenna using relation in Eq. (5) (Table 3).

$$\eta_{total} = \eta_{radiation}(1 - |S_{11}|^2 - |S_{21}|^2)$$

$$\eta_{total} = \text{Total efficiency and } \eta_{radiation} = \text{ Radiation efficiency} \tag{5}$$
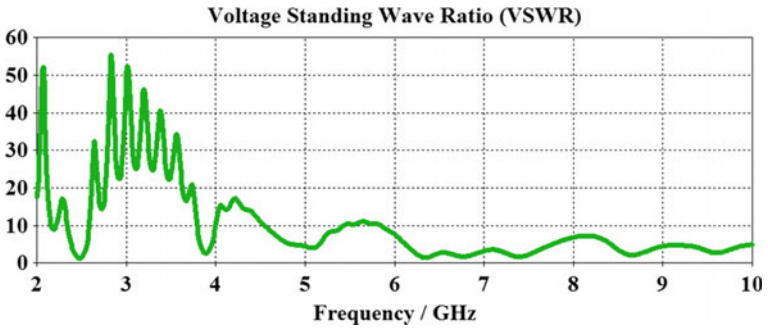
Fig. 5 VSWR graph
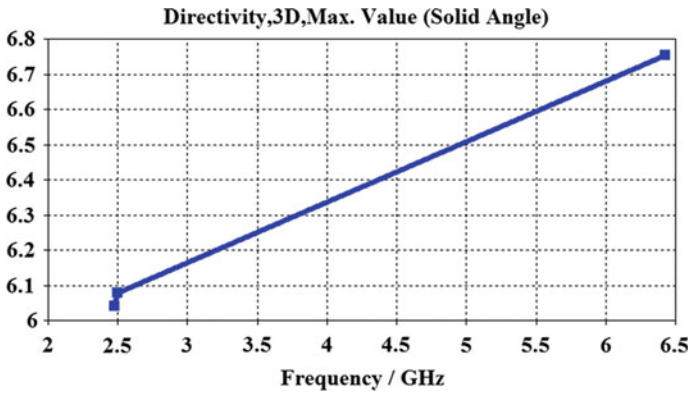
Fig. 6 Directivity of rectangular antenna
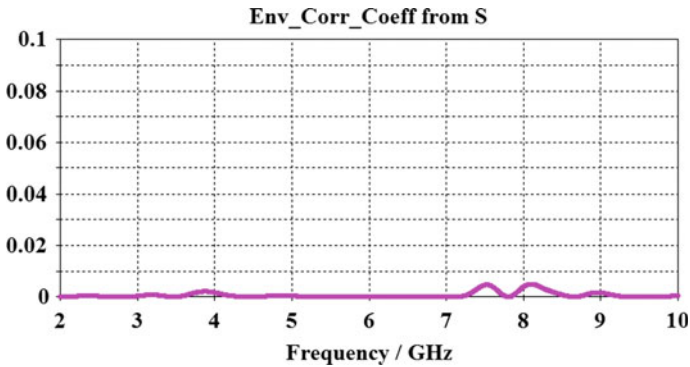
Fig. 7 Envelope correlation coefficient

Fig. 8 Co- and cross-polarization at 2.48 GHz
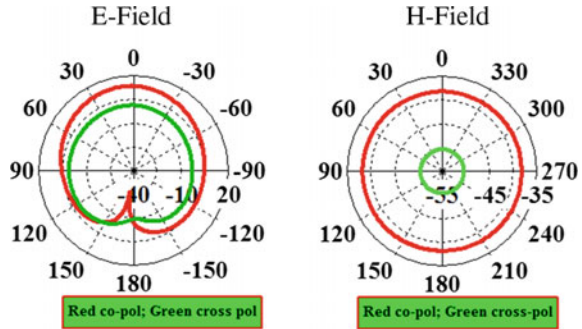
Table 3 Rectangular MIMO antenna parameters

S. no.	Frequency (GHz)	Method	Correlation coefficient (ρ)	TARC in dB	Channel capacity (bits/s/Hz)	Capacity loss in (bits/s/Hz)
1	2.5	MIMO antenna	0.0219	−28.56	4.045	0.26

We observed that within the 20 dB impedance, bandwidth of mutual coupling in the frequency 2.3–2.5 GHz is increased with 71–83% of antenna efficiency.

5 Conclusions

The rectangular MIMO antenna is designed for Bluetooth and WLAN applications. The rectangular MIMO antenna resonates at a frequency of 2.48 GHz that has been simulated. To choose a separation, distance of the two patches is minimum in order to obtain the low mutual coupling between the antennas presented here. Due to arrangement, spacing between the elements is 0.06 λ_0 with improvement in the MIMO parameters like TARC, channel capacity, correlation coefficient, and capacity loss presented here.

References

1. Xue, C.D., Zhang, X.Y.: MIMO antenna using hybrid electric and magnetic coupling for isolation enhancement. IEEE Trans. Antennas Propag. **65**, 5162–5169 (2017)
2. Alsath, M.G., Kanagasabai, M., Balasubramanian, B.: Implementation of slotted meander line resonators for isolation enhancement in microstrip patch antenna arrays. IEEE Antennas Wirel. Propag. Lett. **12**, 15–18 (2013)
3. Shafique, M.F., Qamar, Z., Riaz, L., Saleem, R., Khan, S.A.: Coupling suppression in densely packed microstrip arrays using metamaterial structure. Microwav. Opt. Technol. Lett. **57**, 759–763 (2015)

4. Khandelwal, M.K., Kanaujia, B.K., Dwari, S., Kumar, S., Gautam, A.K.: Analysis and design of wide band microstripline fed antenna with defected ground structure for Ku band applications. AEU—Int. J. Electron. Commun. **68**, 951–957 (2014)
5. Ketzaki, D.A., Yioultsis, T.V.: Metamaterial-based design of planar compact MIMO monopoles. IEEE Trans. Antennas Propag. **61**, 452–455 (2013)
6. Shoaib, S., Shoaib, I., Shoaib, N., Chen, X., Parini, C.G.: Design and performance study of a dual-element multiband printed monopole antenna array for MIMO terminals. IEEE Antennas Wirel. Propag. Lett. **13**, 329–332 (2014)
7. Toktas, A., Akdagli, A.: Wideband MIMO antenna with enhanced isolation for LTE, WiMAX and WLAN mobile handsets. Electron. Lett. **50**, 723–724 (2014)
8. Wu, C.-H., Chiu, C.-L., Ma, T.-G.: Very compact fully lumped decoupling network for a coupled two-element array. IEEE Antennas Wirel. Propag. Lett. **15**, 158–161 (2016)
9. Liu, L., Cheung, S.W., et al.: Compact MIMO antenna for portable UWB applications with band-notched characteristics. IEEE Trans. Antennas Propag. **63**, 1917–1924 (2015)
10. See, C.H., Abd-Alhameed, R.A., Abidin, Z.Z., McEwan, N.J., Excell, P.S.: Wideband printed MIMO/diversity monopole antenna for WiFi/WiMAX applications. IEEE Trans. Antennas Propag. **60**, 2028–2035 (2012)

Diabetes Data Analysis Using MapReduce with Hadoop

Sunil Kumar and Maninder Singh

Abstract Big data is the collection of complex and huge amount of data that comes from different sources such as social media, online transaction details, sensor data, etc. Such collection of voluminous data becomes hard to analyze using traditional processing applications. In healthcare system, doctors prescribed the insulin to the diabetic patients and the decision is based on the patient's previous record and measure the sugar level at the regular intervals. The aim of this paper is to analyze the medical database of diabetes patients using data mining algorithms such as decision tree and naïve Bayes. This analysis is done using UCI machine learning datasets of diabetes having four features for the training phase. The results have shown that the decision tree algorithm has the more accuracy, precision, recall, and F-measure than naïve Bayes.

Keywords Big data · Hadoop · Data mining · MapReduce

1 Introduction

In recent years, a healthcare industry has been generating a lot of data and this data is still growing at a very fast pace [1]. Huge amount of this data is hard to process and needs some new technology to process. Big data analytics is a growing area with the potential to provide useful insight into healthcare system [2]. The healthcare sector has produced a huge amount of data, and majority of this data is saved in the hard copies, but these days the trend is to digitize this large quantity of data [3]. In health care, Big data processing is similar to the processing of traditional health analytics project, having the difference in working of processing in which project is executed. In the traditional health analysis project, the analysis of data can be performed on a single standalone desktop or laptop [4]. In modern healthcare system, massive processing of unstructured health records, to access these records

S. Kumar (✉) · M. Singh
Department of Computer Science, Punjabi University, Patiala, India
e-mail: sunilkapoorldh@gmail.com

for analysis and make it active after analysis, will create more complexities. Due to this, health industry faces many difficulties to analysis of this kind of data and there is a need to develop the data analytics. The health industry has ongoing movement from investigating facts to newly discovered understanding, becoming a large data-driven healthcare organization. Big Data has the capacity to transform the full healthcare system chain from drug analysis to patients' caring quality.

Furthermore, open-source platforms such as Hadoop and MapReduce have encouraged the application of Big data in health care [5]. In Hadoop, MapReduce is a productive parallel programming model to process large volumes of data. This model was launched in the market by Google. Hadoop open-source implementation having a framework with HDFS is developed by Apache with the powerful features scalability, automatic parallelization, and fault tolerance. Big data is large by concept; data is processed in the form of clustering and run across the multiple nodes cluster in the network [6]. Basically, this is the concept of distributed processing and is used to handle large medical data sets. In MapReduce, there are chunks of each file having size 64 MB each. Role of chunks is to handle the node failure and rack failure. MapReduce is a parallel processing model and is implemented in the form of computer cluster. In the large-scale processing of Big data, many complex calculations are processed on computer cluster with an efficient manner.

The aim of extracting information is to take out data from the dataset and make a clear picture for its further use. Diabetic is a common problem which is on high rise and most of the people suffering from this disease in the whole world. The patient having the diabetic problem has number of side effects such as stroke, heart disease, nerve damage, kidney damage, etc. There is a need to detect this problem early which can secure the patient from other health problems. Doctors prescribed the insulin to the diabetic patients before taking the meals. They decide the insulin based on the patients previous record and measure the sugar level at the regular intervals. Many of the researchers carried out in this work on diabetic datasets.

Classification methods worked well in grouping dataset of blood donors, and the accuracy rate of these methods is around 89.9% by Arvind and Gupta [7]. WEKA tool and J48 method are utilized for the entire research work. Data is gathered from an EDP department of a blood bank, and J48 algorithm is used for the grouping of donors. Therefore, it helps the owner of a blood bank to make accurate and faster decisions. Pandeeswari and Rajeswari [8] proposed the hybrid approach which is a combination of naïve Bayes and K-means for the classification of diabetes patients. There are two stages of the model: K-means clustering is used in the first stage to locate and remove negative instances, and correctly clustered instances go as input to the naïve Bayes. The final results have shown that this hybrid model has improved the classification accuracy. Koklu and Unal [9] analyzed 768 diabetic patients using multilayer perceptron (MLP), naïve Bayes, and J48 classifiers. The analysis was carried out on Pima Indian Diabetes dataset using a Weka tool. Each layer of MLP contains multiple neurons which are interlinked with each other by weights. The previous layer's activation function receives the input and produces the output for the next layer. The results have shown that the naïve Bayes performed better than the other two classifiers with the accuracy of 76.30%. Ianchao et al. [10] used type 2

Pima Indians Diabetes dataset for analysis using Rapid Mine. The aim of the study was to gather data of patients with and without beginning diabetes, and the main focus was on construction of diabetes prediction model. E. G. Yildirim worked on type 2 diabetes dataset and used it for predictive data mining and changes of diabetic patients getting heart-related problem. The proposed system used diabetic diagnosis to predict diabetic patient getting a heart disease, and naive Bayes classifier is used in this system which provides the best possible prediction model.

1.1 Decision Tree Algorithm

Decision tree algorithm (or decision tree learning) is a supervised learning algorithm suitable for classifying categorical data based on their attributes. It uses a decision tree approach to formulate a learning model which can forecast the value of a target attribute (or variable) by studying the outcomes of the decision rules inferred from the training data. A tree can be "learned" by following the procedure called recursive partitioning. In recursive partitioning, the training data is split recursively into multiple subsets based on the outcomes of the attribute value test (or decision rules). This process is repeated for each of the newly derived subsets using different sets of attributes. Now there are two types of tree models, namely, classification trees and regression trees, based on the nature of values represented by the target attribute (or variable). In classification trees, the target attribute (or variable) can take a discrete set of values, whereas, in regression trees, the target attribute (or variable) can take continuous values (typically real numbers). Decision tree learning is a predictive modeling approach widely used in statistics, data mining, and machine learning.

A decision tree has a flowchart-like structure in which every internal node means a "test" on an attribute (e.g., if the salary is greater than 50 K or not), every branch shows the result of the test, and leaf nodes represent a class label. The classification rules are represented by paths from root to leaves. To predict the class label of an unknown observation (or record), we start by comparing the attributes' value of the queried (unknown) record with the root (node) followed with further internal nodes until a leaf node with the forecasted class label is reached. The selection of the next internal node for comparison is made by following the branch corresponding to the outcome of the comparison made at the previous node. Figure 1 is an example of a decision tree.

In data mining, data is represented as

$$(\mathbf{x}, \; Y) = (x_1, \; x_2, \; x_3, \; \ldots, \; x_n, \; Y)$$

Y is the target attribute that we want to classify (or in simpler words, understand) by using the values of input variables (or attributes) x_1, x_2, x_3, etc., represented by vector $\mathbf{x}$.

The biggest advantage offered by the decision trees is their simplicity and extremely simple to understand in comparison to other classification algorithms as
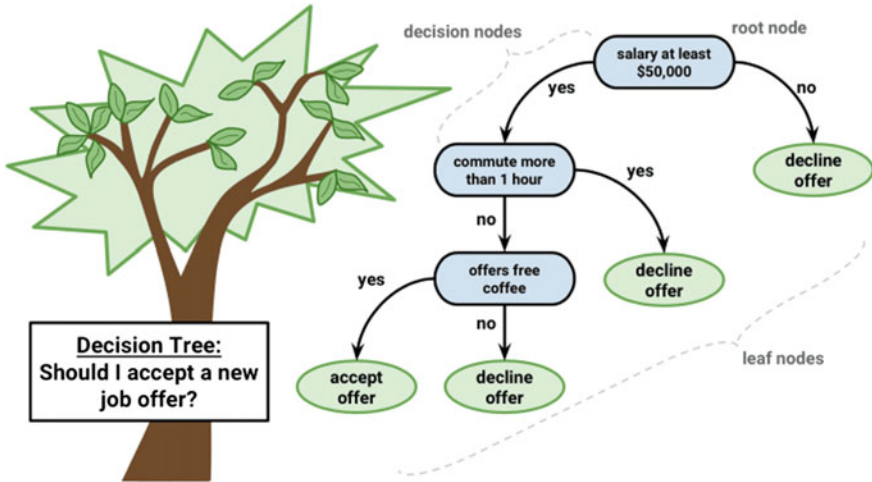
Fig. 1 Decision tree classifier

they use the similar method as human beings normally pursue for making decisions. The explanations of a complex decision tree model can be simplified, so that even an inexperienced person can easily understand the logic. However, the computation complexity of decision trees increases with the increase in the number of class labels.

1.2 Naive Bayes Algorithm

Naive Bayes classifier uses the Bayes theorem to determine the belongingness of the given record for each class, i.e., the probability that the queried (given) record belongs to a particular class. The class with the highest probability is considered the most likely class. Naive Bayes classifier works on the assumption that all the attributes (or features) are unrelated to each other, i.e., the effect of an attribute value on given class is independent of the values of the other attribute. The presence or absence of an attribute (or feature) does not influence the presence or absence of any other attribute. Thus, the naïve Bayes algorithm uses the attribute independence approach to uncouple the effect of multiple attributes on the prediction of target attribute's value. Naive Bayes classifiers are highly scalable and can be quickly modeled to learn from high-dimensional features with limited training data [11]. They are a well-performing model in the field of document classification and disease prediction.

- **Steps for Naive Bayes Algorithm**

1. Given a class variable y which we wish to estimate (or predict) using the set of (dependent) feature values x_1, x_2, x_3, ..., x_n, then the Bayes theorem states the following relationship:

$$P(y|x_1, \ldots, x_n) = \frac{P(y)P(x_1, \ldots, x_n|y)}{P(x_1, \ldots, x_n)}$$

2. The above relation can be simplified using the naive independence assumption to

$$P(y|x_1, \ldots, x_n) = \frac{P(y) \times P(x_1|y) \times P(x_2|y) \times \cdots \times P(x_n|y)}{P(x_1, \ldots, x_n)}$$

$$P(y|x_1, \ldots, x_n) = \frac{P(y) \prod_{i=1}^{n} P(x_i|y)}{P(x_1, \ldots, x_n)}$$

3. For the given input, $P(x_1, \ldots, x_n)$ is constant, therefore,

$$P(y|x_1, \ldots, x_n) \propto P(y) \prod_{i=1}^{n} P(x_i|y)$$

4. Based on the above relationship, the new classification rule can be written as

$$\hat{y} = \arg \max_{y} P(y) \prod_{i=1}^{n} P(x_i|y)$$

where $P(y)$ and $P(x_i|y)$ can be estimated using maximum a posteriori (MAP) estimation, $P(y)$ refers to the prior probability of the class, i.e., the relative frequency of occurrence of class y in the training set and $P(x_i|y)$ refers to the likelihood, i.e., the probability of occurrence of the feature values given the class. Different versions of naive Bayes classifiers differ in their assumptions regarding the distribution of $P(x_i|y)$.

2 Hadoop Framework Components

- **Apache Hadoop**: In 2005, Highly archived distributed object-oriented programming (Hadoop) was the development of Doug Cutting and Mike Cafarella. Hadoop is an open-source framework which supports distributed search engine project [12]. Hadoop supports Java open-source framework technology used for storing,

Fig. 2 Basic structure of
Hadoop
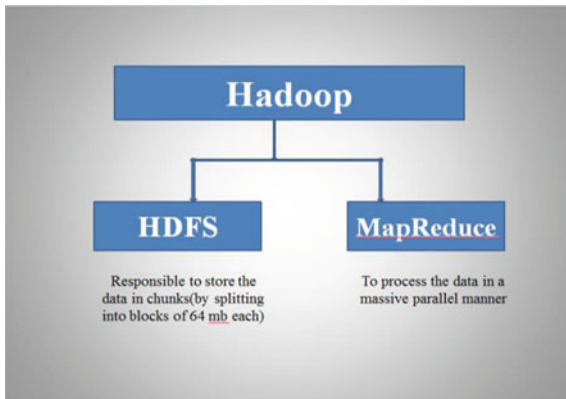
accessing, and analyzing large sources from the Big data having less cost and high
fault tolerance [13]. Hadoop is such software framework which is very trending
these days. As traditional data dealing software like DBMS, RDMS, etc. fails in
case of Big data, Hadoop uses MapReduce framework which deals with parallel
processing of data which requires a sequential read stream of large part of the
stored data [14]. Hadoop can be single node or multi-node in nature. It enhances
the concept of parallel processing on thousands of servers through MapReduce.
MapReduce is Google's programming model concept. Mapper and reducer are
two stages to be followed in Hadoop [15]. There are two main layers of Hadoop
as shown in Fig. 2.

- **HDFS**: Hadoop distributed file system is highly reliable to store file having high
volume in size and providing fault-tolerant file system [16]. Hadoop stores files on
HDFS which is distributed in nature and collects files of different nodes. Hadoop
system has the combined form of single node and multiple of data nodes [16]. Files
are stored as blocks in proper sequence, and these blocks are equal in size, i.e.,
64 MB. Functioning of name node stores metadata, i.e., location of each block,
name, and file attributes. Storage of metadata and file data is separated. Metadata
is stored in name node and application data is stored in data node [17].

- **MapReduce Stage Process**

Both data mining algorithms follow two major stages in MapReduce parallel
processing. The output of map phase is fed as input to reduce. It defines in two stages
as follows:

Mapper Stage: MapReduce is parallel processing programming which is used for
parallel processing over the group of nodes called cluster [18]. As the MapReduce
paradigm is based on the key value scenario so here also mapper takes the input
and divides the extracted data into key–value pairs as output [17]. The input to the
mapper phase is diabetes file of a particular year. Here, we extract two fields from
every record: observation date and the value of sugar of the patients. Quality code is
also tested to identify the missing value in the dataset if any. In mapper stage, every
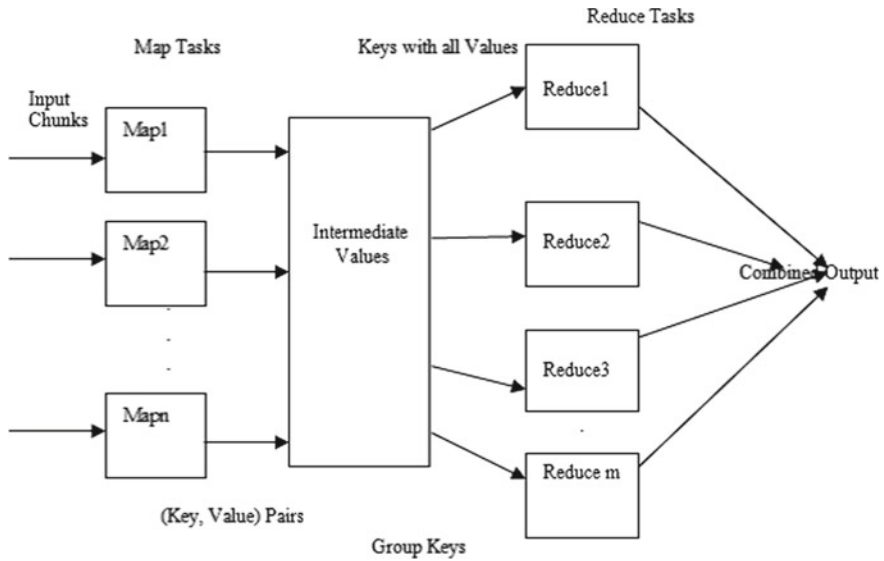
Fig. 3 MapReduce workflow

record processed sequentially and independently in parallel manner and originates pairs of key–value (Fig. 3).

$$\textbf{Map(k1, v1)} \rightarrow \textbf{list(k2, v2)}$$

Reduce Stage: Reduce phase accepts the output of mapper phase as its input. It receives the key–value pairs in text and IntWritable form, respectively, that is, the value of diabetes to the number of patients belongs to a particular observation date at the specified time and specific conditions [17]. Every key value is specific. Final output of the reduce stage is the combined form of all intermediate values which are collected after processing and repeatedly in the combination of key–value pairs [19]. After each phase of processing, the output gets sorted in orderly manner (Fig. 4).

$$\textbf{Reduce(k2, list(v2))} \rightarrow \textbf{list(k3, v3)}$$

3 Experiment Findings

Database Sources: The dataset was taken from https://archive.ics.uci.edu/ml/datasets/diabetes UCI machine learning. Diabetes patient records were obtained from two sources: an automatic electronic recording device and paper records. The automatic device had an internal clock to timestamp events, whereas the paper records only provided "logical time" slots.

- Attribute 1: Date in MM-DD-YYYY format
- Attribute 2: Time in XX.YY format
- Attribute 3: Code
- Attribute 4: Value

The Code field is deciphered as follows:

33 = Regular insulin dose
34 = NPH insulin dose
35 = UltraLente insulin dose
48 = Unspecified blood glucose measurement
57 = Unspecified blood glucose measurement
58 = Pre-breakfast blood glucose measurement
59 = Post-breakfast blood glucose measurement
60 = Pre-lunch blood glucose measurement
61 = Post-lunch blood glucose measurement
62 = Pre-supper blood glucose measurement
63 = Post-supper blood glucose measurement
64 = Pre-snack blood glucose measurement
65 = Hypoglycemic symptoms
66 = Typical meal ingestion
67 = More-than-usual meal ingestion
68 = Less-than-usual meal ingestion
69 = Typical exercise activity
70 = More-than-usual exercise activity
71 = Less-than-usual exercise activity
72 = Unspecified special event

Dataset Description: Diabetes dataset file has four fields in each record. Each field is separated by a tab and each record is separated by a newline.
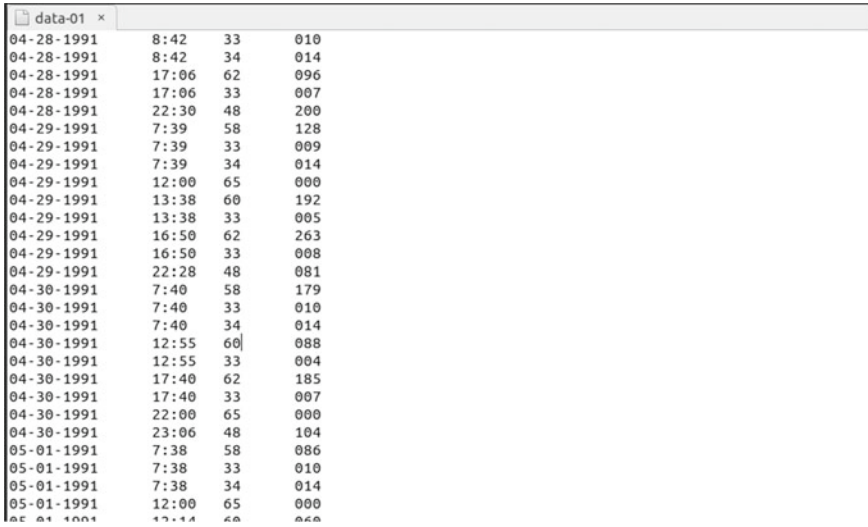
The dataset consists of 70 dataset files and a data code and a domain description file. Each file out of the 70 different files represents the data for diabetic patient. Each file contains four attributes: Date, Time, Code, and Value.

Date: The dataset covers several weeks' to months' worth of outpatient care on patients. The date is in MM-DD-YYYY format.

Time: These are the logical time slots which were fixed on paper records and assigned to breakfast (08:00), lunch (12:00), dinner (18:00), and bedtime (22:00). The time is in XX:YY format.

Code: The code field is deciphered by the amount of insulin given to the patient to control the diabetic value.

Value: This attribute represents a blood glucose (BG) concentration at a particular time after a particular activity (see Code). BG concentration varies even in individuals with normal pancreatic hormonal function.

```
 data-01  ×
04-28-1991      8:42    33      010
04-28-1991      8:42    34      014
04-28-1991      17:06   62      096
04-28-1991      17:06   33      007
04-28-1991      22:30   48      200
04-29-1991      7:39    58      128
04-29-1991      7:39    33      009
04-29-1991      7:39    34      014
04-29-1991      12:00   65      000
04-29-1991      13:38   60      192
04-29-1991      13:38   33      005
04-29-1991      16:50   62      263
04-29-1991      16:50   33      008
04-29-1991      22:28   48      081
04-30-1991      7:40    58      179
04-30-1991      7:40    33      010
04-30-1991      7:40    34      014
04-30-1991      12:55   60      088
04-30-1991      12:55   33      004
04-30-1991      17:40   62      185
04-30-1991      17:40   33      007
04-30-1991      22:00   65      000
04-30-1991      23:06   48      104
05-01-1991      7:38    58      086
05-01-1991      7:38    33      010
05-01-1991      7:38    34      014
05-01-1991      12:00   65      000
05-01-1991      12:14   60      060
```

Fig. 5 Diabetes dataset layout

4 Experimental Result

The experimental research is carried out on the basis of the execution of performed algorithm using MapReduce technique and results are measured based on confusion metrics.

4.1 Count the Number of Diabetic Patient

After the processing of the diabetic datasets, MapReduce generated the output file having the total count of diabetic patients with insulin dose. The output file has the count of diabetic patient's records with their insulin dose/blood glucose measurement. Figure 5 shows the working process of decision tree using Hadoop (MapReduce architecture). All the internal segments of MapReduce-like file system counters are shown in the figure, i.e., number of read operations, number of write operations, and bytes read for processed specific datasets.

Figure 6 shows the output of the algorithm having the number of patients of same quantity of diabetic. The outputs are diabetic code value 33 having 390 patients, 34 code value having 21 patients, 48 code value having 105 patients, and so on in the given dataset of diabetes (Figs. 9, 10, 11, 12, and 13).

Figure 7 shows the working process of naive Bayes algorithm using Hadoop (MapReduce architecture). All the internal segments of MapReduce-like file system
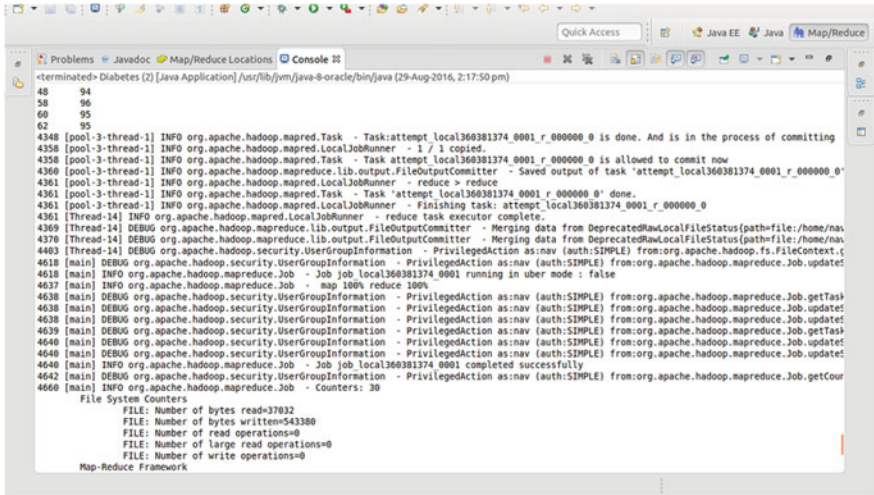
Fig. 6 Working process of decision tree in Hadoop

Fig. 7 Diabetic patient count in decision tree

counters are shown in the figure, i.e., map output bytes, map output records, reduce input records, and reduce output records for processed dataset.

Figure 8 shows the output of the algorithm having the number of patients of same quantity of diabetic. The outputs are diabetes code value 33 having 390 patients, 34 code value having 15 patients, 48 code value having 97 patients, 58 code value having 95 patients, and so on in the given dataset of diabetes.
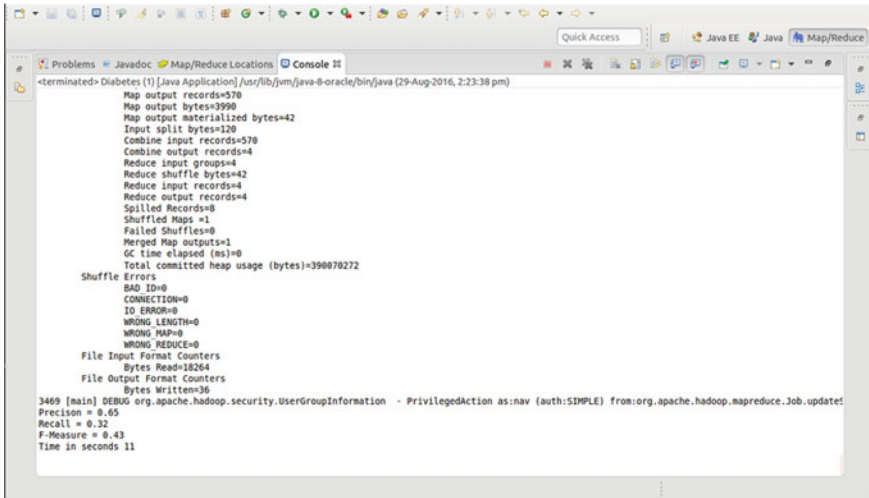
Fig. 8 Working process of naive Bayes in Hadoop

Fig. 9 Diabetic patient count in naive Bayes algorithm

4.2 Precision

Precision is used to represent the fraction of accessed data from connected datasets, which are relevant to search. Precision is the description of random errors or it is a measure of statistical variability. Precision is used to identify how many instances identified correctly from the dataset.

$$\text{Precision } P = \frac{\text{TP}}{TP + FP}$$

Where TP $=$ True Positive Rate and FP $=$ False Positive Rate.

Fig. 10 Graph of precision
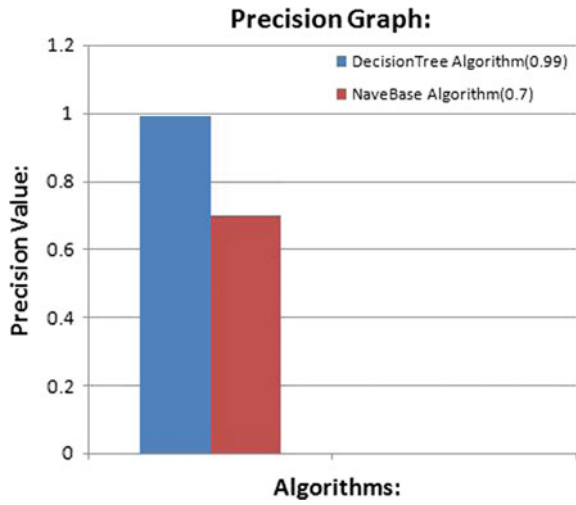
Fig. 11 Graph of recall
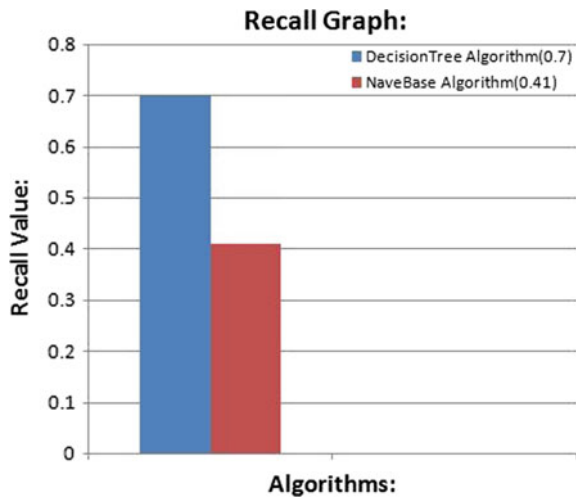
In the given graph of the dataset of diabetes, it shows that precision value of the decision tree is more than as compared to the naive Bayes algorithm, i.e., 1.

4.3 Recall

Recall (or called Sensitivity) is the proportion of the real positive situations that are correctly predicted positive. Recall is used to represent the fraction of accessed data from connected datasets, which are relevant to the query pass that is successfully

Fig. 12 Graph of F-measure
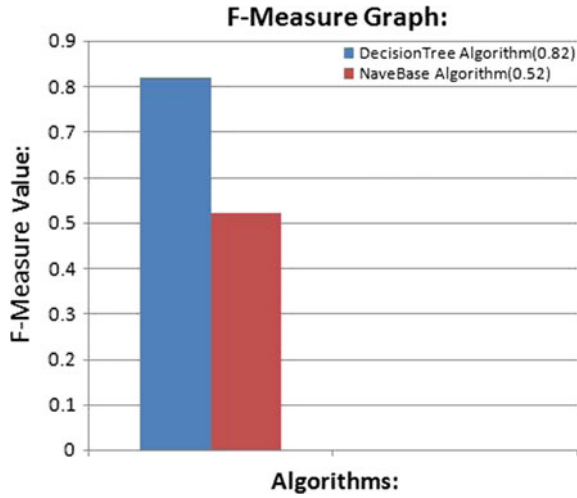
Fig. 13 Graph of execution time
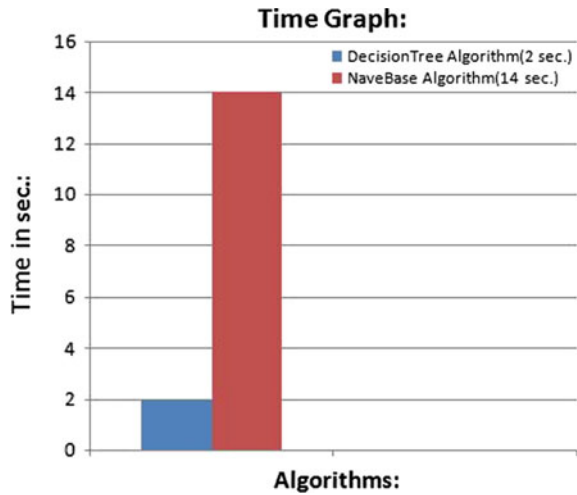
executed. It is a statistical measure or the fraction of material or value which is restored by research. Recall is computed as

$$\text{Recall} \quad R \ = \ \frac{TP}{TP + FN} \text{Where FN} \ = \ \text{False Negative Rate}$$

The given graph of the datasets of diabetes shows that the recall value of the decision tree algorithm is more than the recall value of the naive Bayes algorithm.

Table 1 Final result analysis

Algorithm	TP rate	FP rate	Precision	Recall	F-measure	Execution time (s)
Decision tree	1.00	0.037	1.0	0.79	0.82	2
Naïve Bayes	0.901	0.072	0.7	0.41	0.52	14

4.4 F-Measure

For efficient analyses, precision and recall are used together in the form of F-measure to provide the single measurement for a system. This result of F-measure is the weighted harmonic mean of precision and recall. F-Measure is computed as

$$F - Measure = 2 * \frac{Recall * Precision}{Precision + Recall}$$

Decision tree has the more F-measure value as compared to the naive Bayes in the given graph of the datasets of diabetes.

4.5 Execution Time

Execution time is a series of data points which show specific processing time of an algorithm. In Hadoop, algorithm needs less time for execution of the algorithm as compared to other technologies.

In the time graph of the dataset of diabetes, the time taken by the naive Bayes algorithm is more, i.e., 14 s as compared to that of time taken by the decision tree algorithm, i.e., 2 s (Table 1).

5 Conclusion

The overall objective of this analysis is to predict more accurately the presence of diabetes datasets to find the optimal solution for the patients based on the results. To perform the analysis, the platforms such as Hadoop and MapReduce are used in addition to the data mining algorithms, i.e., decision tree and naïve Bayes. The performance indicators of confusion matrices such as precision, recall, F-measure, and execution time are performed on the dataset. Accusation besides a proper data preprocessing technique can get better the accuracy of the classifier. Based on the results, decision tree is significantly superior to naïve Bayes algorithm in the classification of diabetic dataset. For improving the overall accuracy of the algorithm, there is a need for more datasets with the number of attributes with the best feature

selection method. In future, a new novel hybrid classification approach of decision tree and naïve Bayes with more number of attributes may produce the best results for the dataset.

References

1. Philip, C., Zhang, C.: Data-intensive applications, challenges, techniques and technologies: a survey on big data. Elsevier J. Inf. Sci. **275**, 314–347 (2014)
2. Labrinidis, A., Jagadish, H.V.: Challenges and opportunities with Big Data. ACM-Proc. VLDB Endow. **5**(12) (2012)
3. Raghupathi, W., Raghupathi, V.: Big Data analytics in healthcare: promise and potential. Health Inf. Sci. Syst. **2**, 1–10 (2014)
4. Herland, M., et al.: A review of data mining using big data in health informatics. J. Big Data, 1–35 (2014)
5. Larose, D.: Discovering Knowledge in Data: An Introduction to Data Mining. Wiley, New Jersey (2005)
6. Patel, J., Sharma, P.: Big data for better health planning. In: IEEE International Conference on Advances in Engineering & Technology Research, 1–2 Aug 2014
7. Arvind, S., Gupta, P.: Predicting the number of blood donors through their age and blood group by using data mining tool. Int. J. Commun. Comput. Technol. **1**(2), 6–10 (2012)
8. Pandeeswari, L., Rajeswari, K.: K-means clustering and Naïve Bayes classifier for categorization of diabetes patients. Int. J. Innov. Sci. Eng. Technol. (IJISET) **2**(1) (2015)
9. Koklu, M., Unal, Y.: Analysis of a population of diabetic patients databases with classifiers. World Acad. Sci. Eng. Technol. **7**(8) (2013)
10. Ianchao, J., Rodriguze, C., Beheshti, M.: Diabetes data analysis and prediction model discovery. In: International Conference on Future Generation Communication and Networking, pp 96–99 (2008)
11. Rokach, L.: Data Mining with Decision Trees: Theory and Applications, vol. 69. World Scientific (2007)
12. Sun, W., Chen, J., Li, J.: Decision tree and PCA-based fault diagnosis of rotating machinery. Mech. Syst. Signal. Process. **21**, 1300–1317 (2007)
13. Rahman, R.M., Afroz, F.: Comparison of various classification techniques. J. Softw. Eng. Appl. **6**, 85–97 (2013)
14. Huang, J., Lu, J., Ling, C.X.: Comparing naive bayes, decision trees and SVM with AUC and accuracy. In: Third IEEE International Conference on Data Mining, USA, 22 Nov 2003
15. Chen, M., Mao, S., Liu, Y.: Big data: a survey. Springer-Mob. Netw. Appl. **19**(2), 171–209 (2014)
16. Sugumaran, V., Muralidharan, V., Ramachandran, K.I.: Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing. Mech. Syst. Signal Process. **21**, 930–942 (2007)
17. Addin, O., Sapuan, S.M., et al.: A Naïve-Bayes classifier for damage detection in engineering materials. Mater. Des. **28**, 2379–2386 (2008)
18. Xhemali, D., Hinde, C.J., Stone, R.G.: Naïve bayes vs. decision trees vs. neural networks in the classification of training web pages. Int. J. Comput. Sci. **4**(1) (2009)
19. Xindong, W., Xingquan, Z., Gong-Qing, W., Wei, D.: Data mining with big data. IEEE Trans. Knowl. Data Eng. **26**, 97–107 (2014)
20. Sharma, S., Mangat, V.: Technology and trends to handle big data: survey. In: IEEE 5th International Conference on Advanced Computing & Communication Technologies, 21–22 Feb 2015
21. Shin, D., Choi, M., Kim, W.: Ecological views of big data: perspective and issues. Elsevier Telemat. Inf. **32**, 311–320 (2014)

22. Patnaik, D., Marwah, M., Sharma, R.K., Ramakrishnan, N.: Data mining for modeling chiller systems in data centers. Adv. Intell. Data Anal. **IX**, 125–136 (2010)
23. Augustine, D.P.: Leveraging big data analytics and hadoop in developing india's healthcare services. Int. J. Comput. Appl. **89**, 44–50 (2014)
24. Baldominos, E., Albacete, Y., Saez, P.: A scalable machine learning online service for big data real-time analysis. In: IEEE Symposium on Computational Intelligence in Big Data (CIBD), pp. 1–8, Dec 2014
25. Parthiban, G., Rajesh, A., Srivatsa, S.: Diagnosis of heart disease for diabetic patients using Naive Bayes method. Int. J. Comput. Appl. **24**, 7–11 (2011)
26. Chen, J., Huang, H., et al.: Feature selection for text classification with Naïve Bayes. Expert Syst. Appl. **36**, 5432–5435 (2009)
27. Ibrahim, S., Jin, H., Qi, L., Wu, S., Shi, X.: Evaluating MapReduce on virtual machines: the hadoop case. Springer: Cloud Computing Lecture Notes in Computer Science, Vol. 5931, pp. 519–528 (2009)
28. O'Driscoll, A., Daugelaite, J.: Big data, hadoop and cloud computing in genomics. Elsevier J. Biomed. Inf. **46**, 774–781 (2013)
29. Dede, E., Sendir, B., Kuzlu, P., Ramakrishnan, L.: Processing cassandra datasets with hadoop-streaming based approaches. IEEE Trans. Serv. Comput. **9** (2016)
30. Apache Hadoop. http://hadoop.apache.org/. Sept 2014
31. Borthakur, D.: The Hadoop Distributed File System: Architecture and Design (2007)
32. Dean, J., Ghemawat, S.: MapReduce: simplified data processing on large clusters. In: 6th Conference on Symposium on Operating Systems Design & Implementation. USENIX Association, Berkley, USA, pp. 137–150 (2004)

Theoretical and Experimental Modal Analysis of Beam

Jai Kumar Sharma

Abstract Vibration analysis of the beam components is extremely helpful in engineering analysis and design. Experimental modal analysis is the process to determine the modal parameters in the form of natural frequency, mode shape, and damping. This paper presents a theoretical, experimental, and numerical modal analysis of beam in free-free and simply supported boundary condition. The experimental modal analysis (EMA) of the beam excited using impact hammer is performed with the help of a Laser Vibrometer. With the help of MATLAB program, the signal from time domain to frequency domain is converted into the frequency domain using Fast Fourier Transform (FFT). The finite element modeling (FEM) software, ANSYS workbench 14.5, performed the numerical modal analysis of beam. Thus, the obtained natural frequency and mode shapes of the free-free and simply supported beam are compared with results obtained through theoretical. The comparison of all three results, i.e., theoretical results, numerical results, and experimental results shows that the results are all well within the reasonable error margin. It will help the researchers and engineers for the better design and development of engineering components.

Keywords Experimental · Impact hammer · Laser vibrometer

1 Introduction

The experimental and numerical modal analysis of engineering component always provides an extreme contribution to our effort for better understanding and to control many vibration problems encountered in practice [1]. The accurate boundary conditions, damping, and actual rigidity cannot forecast for complex engineering components using the theoretical modal analysis [2]. So, the determined results have certain error with real results. The EMA was first used in the 1940s, for analysis

J. K. Sharma (✉)
Department of Mechanical Engineering, ITM University, Gwalior, India
e-mail: jaikumarsharma94@gmail.com

of dynamic behavior of structures with help of sine dwell method. Using EMA, we can find the mode shapes at natural frequencies, the damping at natural frequency, and natural frequency of engineering components [3]. In the past two decades, this methodology based on FFT approach has received a lot of attention. This method also enables engineers and researchers to get a well understanding of dynamic properties of structures.

Experimental modal analysis using laser vibrometer is noncontact testing, based on vibration response of the structure. The impact hammer is widely used in EMA for excitation of engineering structure. Miguel et al. [4] determined the modal parameters, i.e., natural frequency, mode shapes, and damping of cantilever steel beam through experiment and compared the experimental results with theoretical and numerical results. Prashant et al. [5] investigated the modal parameter of the rectangular beam through experiment. Thus, the obtained results from the experiment are compared with theoretical and numerical results. The experimental modal analysis of bladed wheel using laser Doppler vibrometer was performed by Bertini et al. [6]. There are two units in laser vibrometer; one is laser sensor head and another one is an electronic unit. The first unit has CCD camera for better visualization and another unit which process the signal [7]. Prasad et al. [8] determined the frequency, mode shape, and damping of different materials beams. In this paper, the material used for beams are Steel, Copper, Brass, and Aluminum and the beam is excited using impact hammer. In the present paper, experimental, theoretical, and finite element modal analysis of the beam is carried out. The laser vibrometer is used for performing the experimental modal analysis and ANSYS Workbench software is used for finite element modeling.

2 Theoretical Modal Analysis

A beam is an inclined or horizontal structural member that is subjected to loads exerting transversely to the longitudinal axis. There are two types of beam theories which are commonly used namely Euler–Bernoulli's beam theory and Timoshenko beam theory. In the Euler–Bernoulli's beam theory, the rotation effects and the shear deformation are neglected, plane sections remain plane and is normal to longitudinal axis but in the Timoshenko beam theory, the plane section still remaining plane are no longer perpendicular to the longitudinal axis. Consider an Euler–Bernoulli's uniform beam undergoing transverse vibration conditions. For free vibration, the equation of motion of beam can be present as [9].

$$EI\frac{\partial^4 w}{\partial x^4} + \rho A \frac{\partial^2 w}{\partial t^2} = 0 \tag{1}$$

$$c^2 \frac{\partial^4 w}{\partial x^4} + \frac{\partial^2 w}{\partial t^2} = 0 \tag{2}$$

$$c = \sqrt{\frac{EI}{\rho A}} \tag{3}$$

In Eq. (1), w is the displacement t is the time x is the position in the beam 0 to L, E is the Young's modulus of the beam, ρ is the mass density of beam and A is the cross-section area of the beam. The solution of differential equation can be found using the method of variable separation [9].

$$w(x,\ t) = w(x)T(t) \tag{4}$$

Using Eqs. (1) and (4)

$$\frac{c^2}{w}\frac{d^4 w}{dx^4}T = -\frac{1}{T}\frac{d^2 T}{dT^2} = a \tag{5}$$

where $a = \omega^2$ is a constant. The Eq. (5) can be written in two forms, so the first form equation will be

$$\frac{d^4 w(x)}{dx^4} - \beta^4 w(x) = 0 \tag{6}$$

where

$$\beta^4 = \frac{\rho A \omega^2}{EI} \tag{7}$$

And, other form equation obtained from (5) will be

$$\frac{d^2 T(t)}{dt^2} + \omega^2 T(t) = 0 \tag{8}$$

The solution of Eq. (8) can be written as

$$T(t) = A\,cos\omega t + B\,sin\omega t \tag{9}$$

Using the initial boundary conditions can be determined the value of constant A and B in Eq. (9).

The solution of Eq. (6) is

$$W(x) = Ce^{sx} \tag{10}$$

Using Eqs. (6) and (11) obtained the general solution.

$$\begin{aligned} w(x) = {} & c_1\,(cos\beta x + cosh\beta x) + c_2\,(cos\beta x - cosh\beta x) \\ & + c_3\,(sin\beta x + sinh\beta x) + c_4\,(sin\beta x - sinh\beta x) \end{aligned} \tag{11}$$

The value of c_1, c_2, c_3 *and* c_4 can be determined from the boundary conditions of the beam. At ends, the displacement and the bending moment is zero in simply supported boundary condition. Thus, Eq. (11) becomes

$$sin\beta l = 0 \tag{12}$$

If the beam is free at both the ends then the shear force and bending moment must be zero at the free end. Thus, Eq. (11) becomes

$$cos\beta l cosh\beta l = 0 \tag{13}$$

From the Eq. (7), the natural frequency of beam ω_n can be written as

$$\omega_n = (\beta l)^2 \sqrt{\frac{EI}{\rho Al^4}} \tag{14}$$

The value of $(\beta l)^2$ is determined using Eqs. (12) and (13) for simply supported and free-free boundary conditions. The material and geometric parameters used for theoretical and finite element modal analysis of beam are shown in Table 1. Table 2 displays the first three natural frequencies of beams using Eq. (14).

Table 1 The material and geometric parameters of beam

Material parameter	Geometric parameter (m)
$E = 2.05 \times 10^{11}$ N/m^2	L = 1.044
$\rho = 7830$ Kg/m^3	B = 0.023
$\vartheta = 0.3$	t = 0.005

Table 2 The theoretically obtained natural frequencies of beam

End conditions	Mode	Theoretical frequency (Hz)
Simply supported	1	10.61
	2	42.47
	3	95.56
Free-free	1	24.07
	2	66.353
	3	130.07

3 Experimental Modal Analysis

The objective of experimental modal analysis of beam is to deduce natural frequencies up to the third mode and observe the system response subjected to small initial disturbance. Due to initial deflection in the system itself, the free vibration takes place in the absence of externally applied forces. The experimental setup was prepared to get free-free, and simply supported boundary conditions of the beam. For the free-free boundary condition, the beam is hanged on thin wires to carry out the experiment (Fig. 1). The care has been taken that wire stiffness is extremely small in comparison to the beam and the support does not offer any resistance to the motion providing virtually free boundary conditions. For simply supported boundary conditions, the beam is supported on the roller at one end and another end is hinged to carry out the experiment as shown in Fig. 1b. The test structure is mounted on the heavy frame to provide proper isolation from the other ground disturbances.

The experimental procedure for both boundary conditions is the same for determination of the natural frequency of the beam. First, the beam is excited (at different points) by using impact hammer. It sets up free vibrations in the beam. The ensuing vibrations of the beam are measured with the help of a laser vibrometer (Polytec laser vibrometer NLV-2500). The ray of laser light impacts on the beam and the displacement is measured with respect to time by selecting the displacement module on the controller. To provide proper reflection, a small miniscule reflecting sticker is attached to the beam at the point of measurement. The laser vibrometer has two units—one is an electronic unit which processes the signal and the other is laser sensor head having CCD camera for better visualization. Finally, the excited signal from the laser vibrometer is fed to a digital oscilloscope (NB207C1) in which the data is processed. The signals in the time domain are saved in the USB drive from the digital oscilloscope. The time domain signal is converted into the frequency domain
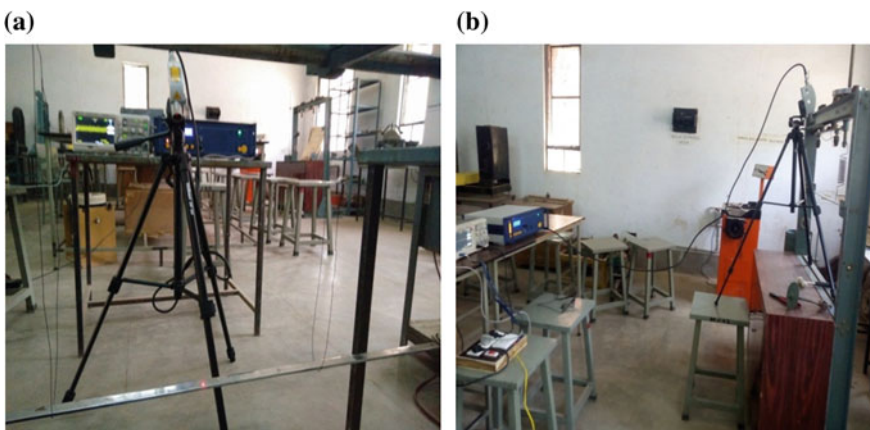
Fig. 1 Experimental setup of **a** free-free beam and **b** simply supported beam

(a)Time domain response   (b) Frequency domain response
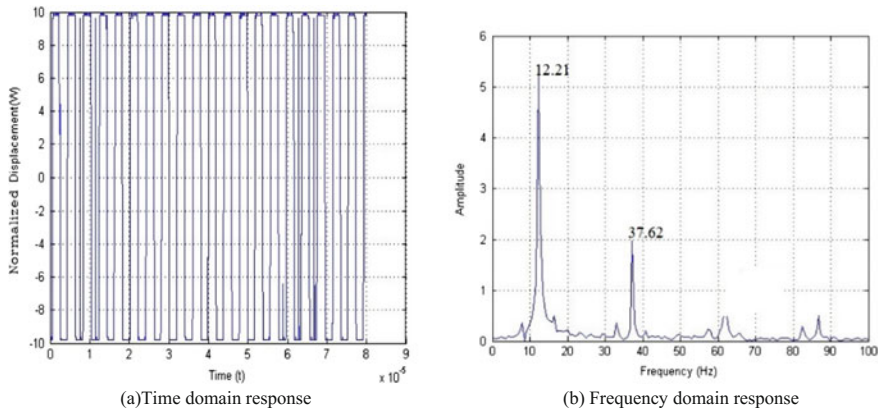
Fig. 2 Response of beam in **a** time domain and **b** frequency domain for an S-S beam

Table 3 Experimentally obtained natural frequencies of beam

End conditions	Mode	Experimental frequency (Hz)
Simply supported	1	12.21
	2	37.62
	3	–
Free-free	1	25.63
	2	77.51
	3	128.8

using Fast Fourier Transformation with the help of MATLAB program. The experimental setup is designed and fabricated to determine the natural frequency of beam in different boundary conditions. The boundary conditions considered for the present experimental work are free, free-free and simply supported as described above. The beam was excited with the help of impact hammer and response was captured using vibrometer. The experimentally obtained natural frequencies of the beam for various boundary conditions are tabulated in Table 3.

The time domain signal is converted into the frequency domain signal using the FFT with the help of MATLAB program. The time domain and frequency domain response for a simply supported beam are shown in Fig. 2.

We can see in the frequency domain graph of simply supported beam, the first peak occurs at the first natural frequency (12.21 Hz) and the second peak is at a second natural frequency (37.62 Hz). The time domain and frequency domain response for the free-free beam is shown in Fig. 3. From these figures, it can be seen that the first peak occurs at the first natural frequency (25.63 Hz), the second peak is at a second natural frequency (77.51 Hz), and the third peak is at a third natural frequency (128.8 Hz) for the free-free beam. In the case of simply supported beam, the higher modes of the beam, cannot be excited using impact hammer.

(a)Time domain response

(b) Frequency domain response
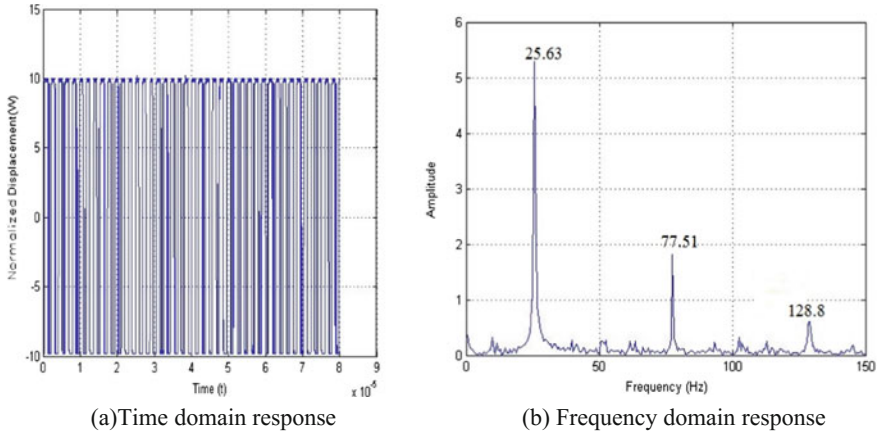
Fig. 3 Response of beam in **a** time domain and **b** frequency domain for free-free beam

4 Finite Element Modeling

The finite element modal analysis of beam using the ANSYS workbench 14.5 is carried out in three boundary conditions. The boundary conditions used for the analysis are free-free and simply supported. The material and geometric parameters is taken from Table 1. The three-dimensional finite element model of the beam is constructed in ANSYS workbench and then computational modal analysis is carrying out to generate natural frequencies and mode shapes. Thus, natural frequencies in bending mode obtained for beam using ANSYS workbench is shown in Table 4. The corresponding numerically obtained mode shapes for different boundary conditions of the beam are shown in Figs. 4 and 5.

Table 4 Numerical natural frequency (Hz)

End conditions	Mode	Numerical natural frequency (Hz)
Simply supported	1	10.64
	2	42.57
	3	95.79
Free-free	1	24.12
	2	66.505
	3	130.37

A: Modal
Total Deformation 3
Type: Total Deformation
Frequency: 10.644 Hz
Unit: m
7/17/2016 6:03 PM

1.4588 Max
1.2967
1.1346
0.97251
0.81042
0.64834
0.48625
0.32417
0.16208
2.481e-7 Min

A: Modal
Total Deformation 4
Type: Total Deformation
Frequency: 42.574 Hz
Unit: m
7/17/2016 6:03 PM

1.4593 Max
1.2971
1.135
0.97284
0.8107
0.64856
0.48642
0.32428
0.16214
1.8615e-8 Min

0.000   0.600 (m)
       0.300

0.000   0.600 (m)
       0.300

(a) First Mode

(b) Second Mode

A: Modal
Total Deformation 6
Type: Total Deformation
Frequency: 95.788 Hz
Unit: m
7/17/2016 6:04 PM

1.4601 Max
1.2978
1.1356
0.97338
0.81115
0.64892
0.48669
0.32446
0.16223
8.1303e-9 Min

0.000   0.600 (m)
       0.300

(c) Third Mode

Fig. 4 Mode shapes and corresponding natural frequency of simply supported beam

5   Comparison of Results

The theoretical natural frequencies are calculated using the Eq. (14) and finite element natural frequencies are determined using the FEM software ANSYS workbench 14.5. The experimental natural frequencies are determined using laser vibrometer with impact hammer excitation method. Table 5 displays the numerical and experimental natural frequencies of the beam in free-free and simply supported boundary conditions.

A: Modal
first mode
Type: Total Deformation
Frequency: 24.127 Hz
Unit: m
7/17/2016 5:53 PM

2.0628 Max
1.834
1.6051
1.3763
1.1474
0.91861
0.68977
0.46093
0.23208
0.0032439 Min

0.000   0.700 (m)
        0.350

(a) First mode

A: Modal
second mode
Type: Total Deformation
Frequency: 66.505 Hz
Unit: m
7/17/2016 5:54 PM

2.0629 Max
1.8337
1.6045
1.3753
1.1461
0.91686
0.68764
0.45843
0.22922
1.9741e-6 Min

0.000   0.700 (m)
        0.350

(b) Second Mode

A: Modal
third mode
Type: Total Deformation
Frequency: 130.37 Hz
Unit: m
7/17/2016 5:54 PM

2.0631 Max
1.8355
1.6078
1.3801
1.1524
0.9247
0.69701
0.46932
0.24163
0.013937 Min

0.000   0.700 (m)
        0.350

(c) Third Mode

Fig. 5 Mode shapes and corresponding natural frequency of free-free beam

6 Conclusions

In present work theoretical, experimental, and numerical modal analysis of beam performed in simply supported and free-free boundary conditions. It can be observed that in the experiment it was not possible to excite the third mode in simply supported boundary condition. Similarly, the fourth and higher modes cannot be excited in free-free boundary conditions of the beam using the impact hammer. For determination of higher modes, the sine sweep test can be more effective. The experimental and numerical results for the beam are found to have an extremely good correlation. Some error between the results is also introduced by physical dimensions of beam

Table 5 Comparison between theoretical, numerical, and experimental results for beam

End conditions	Mode	Numerical frequency (Hz)	Theoretical frequency (Hz)	Experimental frequency (Hz)
Simply supported	1	10.64	10.61	12.21
	2	42.57	42.47	37.62
	3	95.79	95.56	–
Free-free	1	24.12	24.07	25.63
	2	66.505	66.353	77.51
	3	130.37	130.07	128.8

and variation of Young's modulus. The overall analysis shows that the results are all well within the reasonable error margin. The modal analysis techniques presented in this paper may be helpful in dynamic analysis, developing and optimizing the design of complex engineering structure and component.

References

1. Ewins, D.J.: Modal Testing: Theory, Practice and Application, 2nd edn. Wiley. ISBN 978-0-86380-218-8
2. He, J.: Zhi-Fang Fu, Modal Analysis, Butterworth-Heinemann. Replika Press Pvt Ltd, Dehli, ISBN 0 7506 5079 6 (2001)
3. Lengvarský, P., Bocko, J.: Theoretical basis of modal analysis. Am. J. Mech. Eng. **1**, 173–179 (2013)
4. Miguel, F.L.F., Miguel, F.L.F., Thomas, K.C.: Theoretical and experimental modal analysis of a cantilever steel beam with a tip mass. Proc. IMechE Part C J. Mech. Eng. Sci. **223**, 1535–1541 (2009). https://doi.org/10.1243/09544062jmes1390
5. Prashant, S.W., Chougule, V.N., Mitra, A.C.: Investigation on modal parameters of rectangular cantilever beam using experimental modal analysis. Mater. Today Proc. **2**, 2121–2130 (2015). https://doi.org/10.1016/j.matpr.2015.07.214
6. Bertini, L., Neri, P., Santus, C., Guglielmo, A.: Automated experimental modal analysis of bladed wheels with an anthropomorphic robotic station. Exp. Mech. **57**(2), 273–285 (2016). https://doi.org/10.1007/s11340-016-0223-5
7. Parashar, S.K., von Wagner, U., Hagedorn, P.: Finite element modeling of nonlinear vibration behaviour of piezo-integrated structures. Comput. Struct. **119**(3), 37–47 (2013). https://doi.org/10.1016/j.compstruc.2012.12.026
8. Prasad, D.R., Seshu, D.R.: A study on dynamic characteristics of structural materials using modal anaylsis. Asian J. Civ. Eng. (Build. Hous.) **9**(2), 141–152 (2008)
9. Rao, S.S.: Mechanical Vibration. Pearson Education, Inc., publishing as Prentice Hall (2011)

Design and Analysis of Low Profile, Enhanced Bandwidth UWB Microstrip Patch Antenna for Body Area Network

Raghvendra Singh, Abhishek Singh Rathour, Vivek Kumar, Dambarudhar Seth, Sanyog Rawat and Kanad Ray

Abstract In this paper, a low profile and enhanced bandwidth, ultra-wideband microstrip patch antenna for wireless body area network (WBAN) has been proposed. Application of various bandwidth enhancement techniques result in the ultra-wideband ranging from 4.11 to 13.83 GHz in free space and 5.65–13.83 GHz in proximity of human tissue. The antenna consists of cascaded T-slots and rectangular notches in patch to enhance the operating bandwidth in ultra-wideband frequency range. In addition, fabrication and several analyses have been conducted in free space and close proximity to the canonical phantom for finding the effects of human body on the performance of WBAN antenna. The favorable results of return loss, bandwidth, radiation pattern, and directivity are revealing the proposed antenna as a promising candidate for wireless body area network.

R. Singh · D. Seth
Department of Electronics and Communication Engineering,
JK Lakshmipat University, Jaipur, Rajasthan, India
e-mail: planetraghvendra@gmail.com

D. Seth
e-mail: dseth@jklu.edu.in

A. S. Rathour
Department of Electronics and Communication Engineering,
Sobhasaria Engineering College, Sikar, Rajasthan, India
e-mail: abhisheksrathour@gmail.com

V. Kumar
Department of Electronics and Communication Engineering, Pranveer Singh
Institute of Technology, Kanpur, U.P., India
e-mail: rastogi0807@gmail.com

S. Rawat
Department of Electronics and Communication Engineering, Manipal University, Jaipur, India
e-mail: sanyograwat@gmail.com

K. Ray (✉)
Amity School of Applied Sciences, Amity University, Jaipur,
Kant Kalwar, NH-11C, Jaipur 303007, Rajasthan, India
e-mail: kanadray00@gmail.com

1 Introduction

The FCC (Federal Communications Commission) furnished guidelines for UWB emissions in the frequency range between 3.1 and 10.6 GHz [1]. Ultra-wideband technology has various unique advantages over the conventional wireless communications technology as low power consumption, high data rate, and less multipath propagation. Due to the aforesaid advantages, UWB technology always attracts researchers of various applications including wireless body area network (WBAN).

Various types of monopole antennas have been developed for ultra-wideband application, such as rectangular patch, circular disc, and many techniques of bandwidth enhancement have been introduced such as adding slot to the patch [2], using partial ground plane [3], etc. Using a key technology for designing WBAN antenna is always a challenge because of the fact that the antenna characteristics are affected significantly by the human body. In the vicinity of human tissue, the impedance bandwidth, radiation pattern, and gain of the antenna are affected.

In this paper, a small-scale cascaded T-slot UWB antenna for wireless body area network (WBAN) is presented. To design this UWB antenna, three techniques have been applied to the proposed antenna: (i) two steps of notches are situated at the two lower corners of the patch, (ii) a partial ground plane having triangular cut at edges and rectangular cut, (iii) T-slot, which can guide to a good impedance matching. By choosing these parameters, the proposed antenna can be tuned to operate in the (3.1–10.6) GHz frequency range. The simulation results in this paper are obtained using CST microwave suite. Approach of triangular cut at edges has been used to enhance the impedance bandwidth for microstrip patch antenna.

2 Wireless Body Area Network

Due to fast amendments in wearable technology as fast computing, microelectronics, and their miniaturization have made viable the development of wireless body area network (WBAN) in the past few years [4, 5]. The main concern of these devices is to upgrade the quality of life or the comfortness of human being. These types of the systems are deployable in real-time health monitoring system, entertainment, sports, and defense [6, 7]. Aforesaid intelligent devices can be integrated together to form WBAN [7, 8]. Nowadays, many researchers are doing work in this field to integrate the devices in proximity of body to enhance the working capability of sports persons, patients, and soldiers in the battle field [9, 10].

Wireless body area network is depicted in Fig. 1, which is showing many wearers, wearing many sensor nodes on their body and connected to a gateway node to transmit
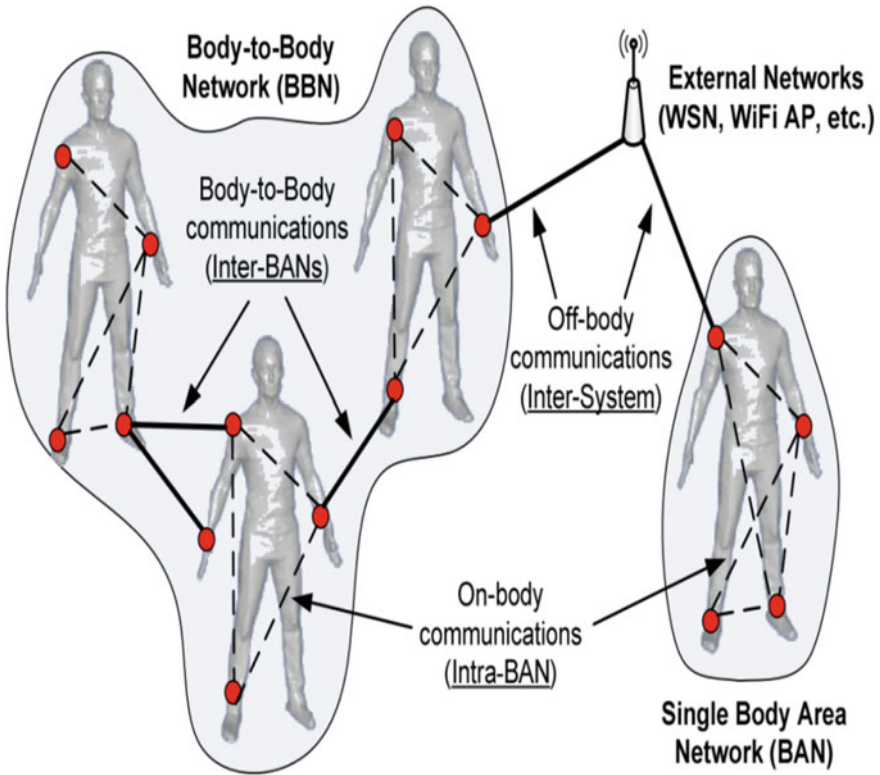
Fig. 1 Wireless body area network with on-body sensors [13]

or receive the data with the nearby external network. All sensor nodes on the body are wirelessly connected to the gateway node for retrieving data on the external network. On the other side of the external network, any controller or the physician can monitor the data in real time. The wireless body area network contains On-body and Off-body communication as depicted in Fig. 1. All sensor nodes are connected using on-body communication link and off-body communication link is connecting the gateway node and external network.

3 Antenna Design Strategy

A substrate of high permittivity, FR4 is selected and the rectangular patch is designed for operating frequency in UWB range. This microstrip patch antenna is designed and simulated in computer simulation software, i.e., CST Microwave Suite. Various researchers used one-layer skin model as in [11, 12], but in this work four-layered phantom model (skin, fat, muscle, bone) is used for making the study more viable.
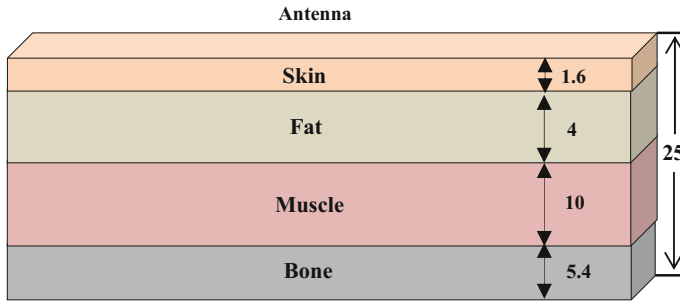
Fig. 2 Four-layered canonical model of human tissue (Phantom)

Table 1 Electromagnetic properties of different human tissues [14]

Tissue	Relative permittivity	Loss tangent	Conductivity (S/m)
Skin	35.3632	0.3200	3.4632
Fat	9.9382	0.2546	0.7742
Muscle	48.8831	0.3081	4.6091
Bone	15.6351	0.4225	2.0212

This electrical equivalent four-layered canonical model (Phantom) of the human tissue is more realistic than one-layer model for simulations as shown in Fig. 2.

Four-layered canonical model is designed for making an electrical equivalent of human body and observe the effect of the model on the characteristics of WBAN antenna. WBAN antennas are used in proximity of human body and the human body also affects the characteristics of antenna, hence the properties of different parts of human tissue (skin, fat, muscle and bone) are a very important concern for observation. Table 1 is depicting the various electrical properties of different layers of human tissue at frequency 5.5 GHz.

3.1 Design Configuration of Antenna

The geometrical parameters of the proposed antenna are given in Table 2. The material used in the patch region is lossy copper and substrate used in the proposed antenna is FR-4 with relative dielectric constant $\varepsilon_r = 4.7$, conductivity $\sigma = 1$ and tangent delta 0.00022. The dimensions are so chosen to be reasonable for body worn devices, surface area of antenna is 900 mm^2, which is miniaturized enough to be used for On-body applications in UWB band for WBAN.

The geometries of the antenna are shown in Fig. 3, (a) Simulated Front view with ground plane, (b) Simulated Side view, (c) Fabricated Back view and (d) Fabricated Front view. This proposed antenna is containing cascaded T-slots, staircase notches in lower part of the patch, two notches and triangular cuts in ground. The rectangular

Table 2 Parameter value of the proposed antenna

Parameter	Value (mm)
Length of substrate (L)	30
Width of substrate (W)	30
Thickness of substrate	1.6
Width of microstrip line	1.8
Length of rectangular patch (L_p)	15
Width of rectangular patch (w_p)	12
Length of ground (Lg)	30
Width of ground (Wg)	05
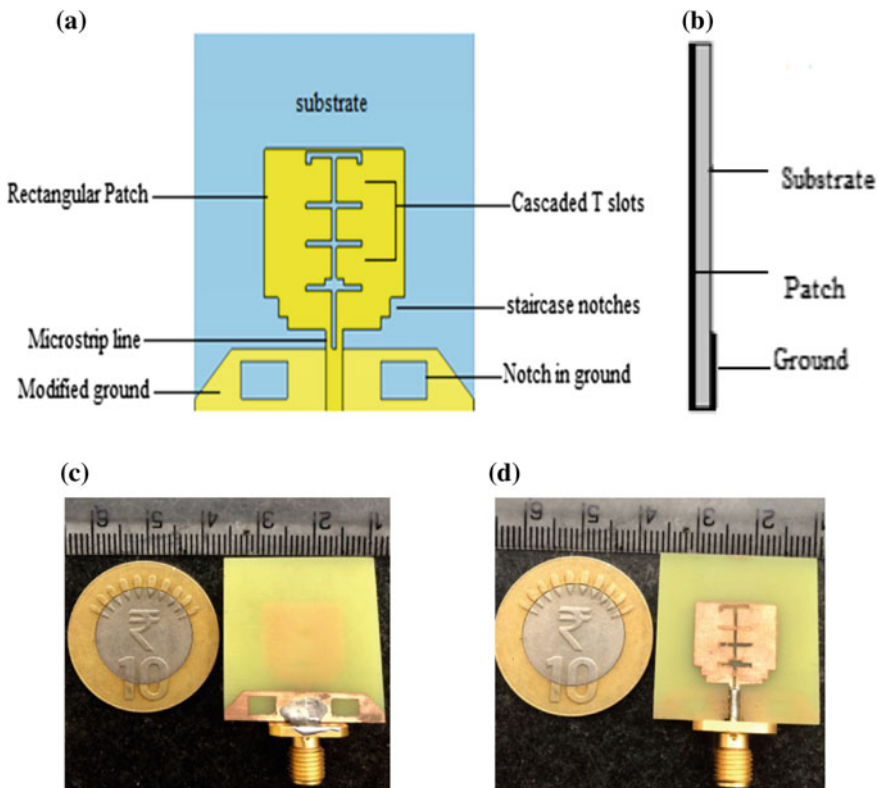
Fig. 3 Proposed design of WBAN antenna **a** simulated front view, **b** simulated back view, **c** fabricated back view, **d** fabricated front view

microstrip patch antenna is chosen here because it creates many advantages, such as being compact, economical, and lightweight and operating in ultra-wideband (UWB). Moreover, a lower bandwidth is a drawback of this shape. This study targeted to
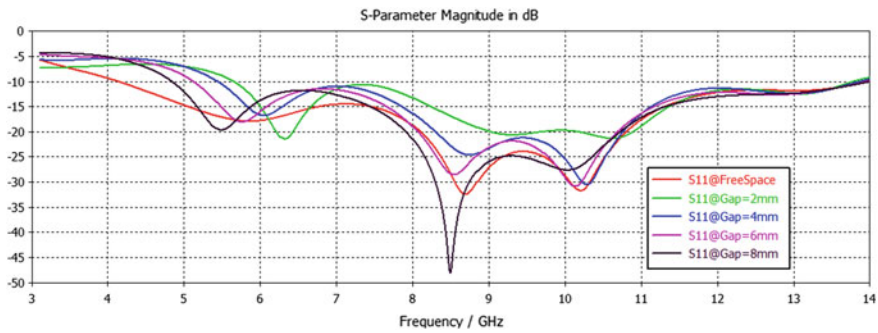
Fig. 4 Comparison of reflection coefficient versus frequency plots

Table 3 Comparison of reflection coefficient in proximity of phantom

Proximity cases	Lower freqency (GHz)	Upper frequency (GHz)	Bandwidth (%)
Free space	4.11	13.83	108.36
Air gap 2 mm	5.65	13.83	83.15
Air gap 4 mm	5.44	13.83	87.12
Air gap 6 mm	5.12	13.83	91.97
Air gap 8 mm	4.84	13.83	96.35

modify the shape and including the approach to improve the bandwidth for WBAN operating in ultra wide band. As follows, concise analysis of the parametric studies to attain the best values of return loss and bandwidth is discussed.

4   Results and Discussion

4.1   *Reflection Coefficient Versus Frequency*

Reflection Coefficient is obtained in various cases at the same port, the cases are; free space, air gap 2 mm, 4 mm, 6 mm, and 8 mm depicted by S_{11} @ free space, S_{11} @ 2 mm, S_{11} @ 4 mm, S_{11} @ 6 mm, S_{11} @ 8 mm, respectively, in Fig. 4. The purpose of inserting the air gap is to analyze more realistic results in proximity of human body.

The WBAN antenna is not in direct contact with the human body hence different cases are studied. Table 3 is showing the comparison of various cases in proximity of phantom. Lower and upper frequencies are mentioned in Table 3 for calculation of −10 db bandwidth for each case.
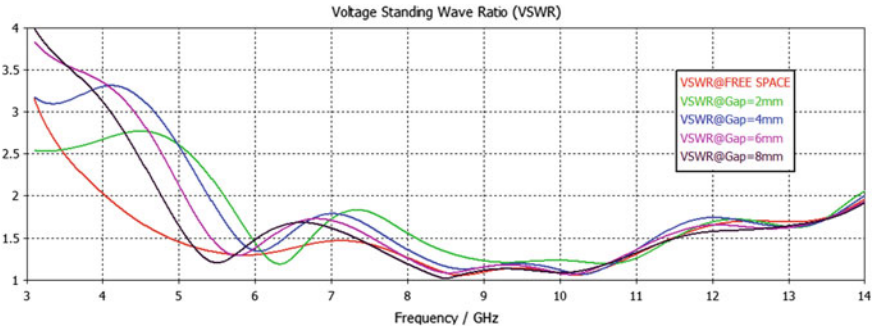
Fig. 5 VSWR versus frequency plots

4.2   VSWR Versus Frequency Comparison

In all cases the voltage standing wave ratio is below 2, it means that the antenna is suited for transmission in UWB and in proximity of phantom also. The VSWR plot of the proposed antenna is assuring the perfect operation in ultra-wideband in proximity of phantom. VSWR is obtained in various cases at the same port, the cases are; free space, air gap 2 mm, 4 mm, 6 mm, and 8 mm depicted by VSWR @ free space, VSWR @ 2 mm, VSWR @ 4 mm, VSWR @ 6 mm, VSWR @ 8 mm, respectively, in Fig. 5.

4.3   Radiation Pattern

Radiation pattern of the WBAN antennas is of primary importance because the radiation in body side is less preferred to decrease the penetration of radiation in human body. In the proposed study, the antenna is producing only one sided radiation pattern at frequency 5.5 and 10.5 GHz (directivity 9.131 dBi) depicted in Fig. 6a, c. Body side radiation is increased at frequency 8.5 GHz as in Fig. 6b.

4.4   Surface Current

The surface current at frequencies 5.5 GHz, 8.5 GHz, and 10.5 GHz of the proposed antenna is depicted in Fig. 7a, b, and c respectively.

Arrows are indicating the intensity and direction of current on the surface of patch, cascaded T-slots are drawing current significantly at thee frequency between 5.5 and 8.5 GHz, and the lower part of antenna is drawing the maximum current at 10.5 GHz with density 37.76 A/m. Hence, the lower and around part of antenna is drawing
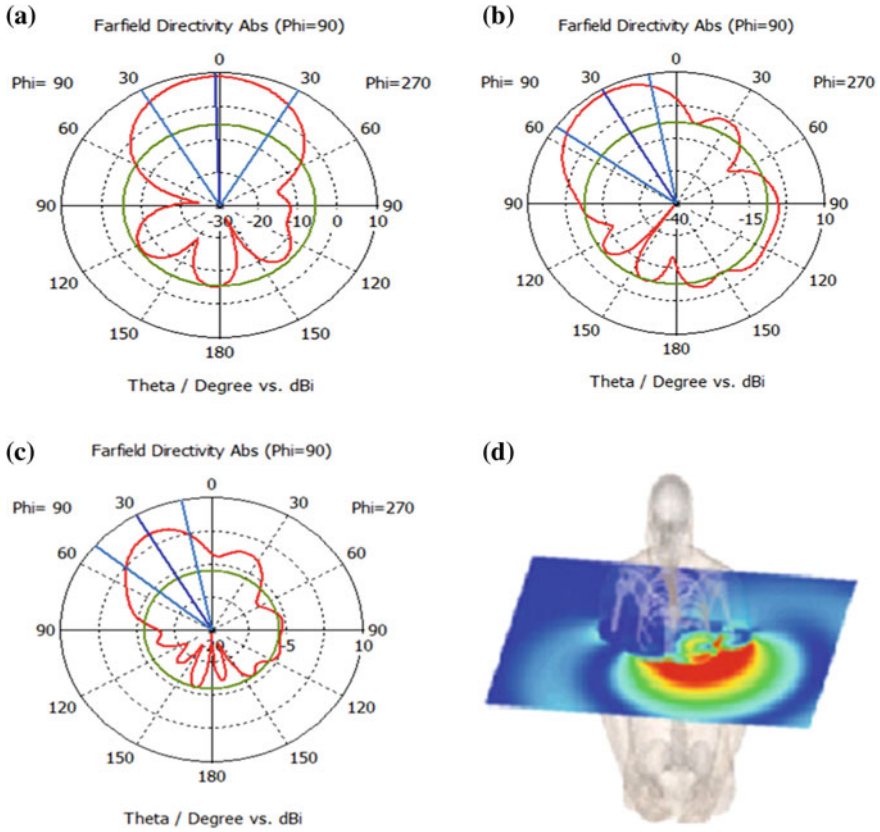
Fig. 6 Radiation pattern at **a** 5.5 GHz, **b** 8.5 GHz, **c** 10.5 GHz, **d** diagrammatic visualization of field on the human body
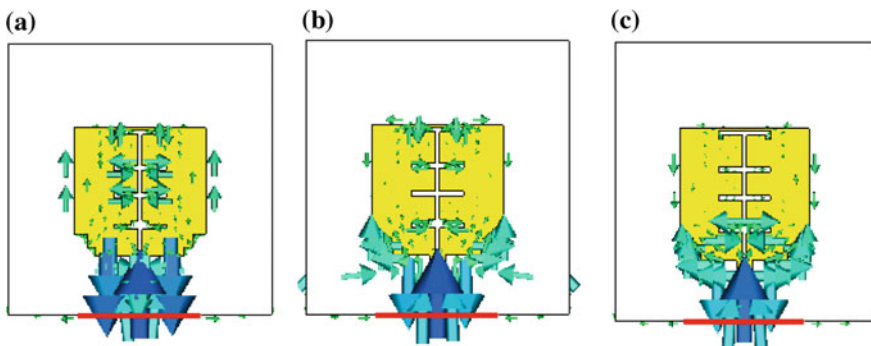
Fig. 7 Surface current at frequency **a** 5.5 GHz, **b** 8.5 GHz, **c** 10.5 GHz

maximum current at higher range of frequencies and upper part of T-slots at lower range of frequencies.

5 Conclusion

A low profile, robust microstrip patch antenna that is suitable for wireless body area network (WBAN) has been presented. Presented antenna occupies only 30 mm × 30 mm × 1.6 mm volume, accepted for On-body and Off-body communication, provides very wide band width with one-sided radiation pattern, which is away from the human body and favorable for integration with WBAN devices. The numerical studies of proposed antenna shows that the bandwidth more than 95% is achieved, directivity 9.131 dBi at 10.5 GHz, and surface current density 37.76 A/m in the close proximity ($d = 8$ mm) of the canonical human phantom. Based on the above study and facts, the proposed antenna is a promising candidate for WBAN.

References

1. Molisch, A.F.: Ultra Wideband Antennas and Propagation for Communications, Radar and Imaging. Wiley (2007)
2. Sze, J.Y., Wong, K.L.: Slotted rectangular microstrip antenna for bandwidth enhancement. IEEE Trans. Antenna wave Propag. **48**(8) (2000)
3. Alam, A.H.M.Z., Islam, M.R., Khan, S.: Design and analysis of UWB rectangular patch antenna. In: APACE 2007 on Applied Electromagnetic, 4–6 Dec 2007
4. Baber, C., et al.: Ergonomics of wearable computers. Mob. Netw. Appl. **4**, 15–21 (1999)
5. Guha, D., Antar, Y.M.M.: Microstrip and Printed Antennas: New Trends, Techniques and Applications. Wiley, London. ISBN-10 0470681926. Design Aspects of Body-Worn UWB Antenna for Body-Centric … 123 Author's personal copy (2011)
6. Hall, P.S., Hao, Y.: Antennas and Propagation for Body Centric Wireless Communications Systems. Norwood, Artech House (2006). ISBN-10 1-58053-493-7
7. Latre, B., Bream, B., Moerman, I., Blondia, C., Demeester, P.: A survey on wireless body area networks. Wirel. Netw. **17**(1), 1–18 (2011)
8. Patel, M., Wang, J.: Applications, challenges, and prospective in emerging body area networking technologies. IEEE Wirel. Commun. **17**(1), 80–88 (2010)
9. Lymperis, A., Dittmar, A.: Advanced wearable health systems and applications, research and development efforts in the European Union. IEEE Eng. Med. Biol. Mag. **26**(3), 29–33 (2007)
10. Hao, Y., Foster, R.: Wireless body sensor networks for health monitoring applications. Phys. Meas. **29**, R27–R56 (2008)
11. Guo, Y.X., Xiao, S., Changrong, L.: A review of implantable antennas for wireless biomedical devices. Forum Electromagn. Res. Methods Appl. Technol.
12. Hall, P.S., Bai, Q., Nechayev, Y., Constantinou, C.C., Bozzettil, M., Gallo, M.: Simulation and measurement of dynamic on-body communication channels. IEEE Trans. Antennas Propag. **59**, 623–630 (2011)
13. http://www.mdpi.com/2227-7080/5/3/43/htm
14. Gabriel, S., Lau, R.W., Gabriel, C.: The dielectric properties of biological tissues: II measurements in the frequency range 10 Hz to 20 GHz. Phys. Med. Biol. **41**, 2251–2269 (1996)

A Jaya Algorithm for Discrete Optimization Problems

Prem Singh and Himanshu Chaudhary

Abstract This study describes the discrete optimization algorithm based on Jaya algorithm. Originally, the Jaya algorithm is developed for continuous optimization problems. In the proposed algorithm, continuous domain of variables is converted into discrete domain applying bound constraint of middle point of corresponding two consecutive discrete values. The effectiveness of proposed algorithm is demonstrated through standard truss examples taken from the literature and optimum values are compared with discrete optimization algorithms. The proposed algorithm is computationally more efficient for the discrete optimization problems as tested by numerical example.

Keywords Jaya algorithm · Discrete variables · Truss · Weight

1 Introduction

Many practical engineering problems, optimum design variables are not considered as continuous due to the availability of standard values. For example, structural design, the number of bolts for a connection, balancing of rotor using set of the balance masses on each plane, etc. [1]. The existing optimization algorithms have been considered as continuous design variables. But, these algorithms are not efficient for the practical design problems. Therefore, in recent years, the discrete optimization techniques have been developed for practical engineering problems. However, classical and evolutionary have been applied for the discrete optimization problems. Sequential linear programming, branch and bound methods (BBM), penalty function approach, rounding-off techniques based on continuous variables, cutting plane

P. Singh (✉) · H. Chaudhary
Mechanical Engineering Department, Malaviya National Institute of Technology, Jaipur,
Rajasthan, India
e-mail: premsingh001@gmail.com

H. Chaudhary
e-mail: hchaudhary.mech@mnit.ac.in

techniques, zero-one variable techniques (integer programming), and Lagrangian relaxation are classical optimization technique [2]. These methods include more computational, low efficiency, and complexity in calculation of derivatives and Hessians of the objective function [3]. Further, these give local optimal solution due to convergence on optimal solution near to start point [4]. The evolutionary discrete optimization algorithms do not require the calculation of derivatives and Hessians as in case of classical discrete optimization techniques.

The hybrid particle swarm optimization (HPSO), genetic algorithm (GA), artificial bee colony (ABC), and simulated annealing (SA) are the evolutionary discrete optimization algorithms. Furthermore, these algorithms require algorithm-specific parameters for its convergence that affect their performance. The selection of these optimization parameters increases the complexity of algorithm [5–10]. However, Jaya algorithm is an evolutionary algorithm. This algorithm has been developed for continuous design variables by Rao [11, 12]. Furthermore, discrete Jaya algorithm has been proposed for discrete optimization design variables by transforming continuous variables into discrete variables. It uses initial population to find optimum global solution. Moreover, this algorithm finds the optimal solution rapidly and removes the worse solution in every iteration. However, the TLBO algorithm is also not required algorithmic parameter for its convergence, although teacher phase and learner phase (two phases) are required to solve the optimization problems. But Jaya algorithm has only one phase. The implementation of Jaya is simpler than other evolutionary discrete optimization techniques such as ABC, GA, SA, PSO, etc., and any algorithmic parameters are not required in this algorithm.

This paper is structured as follows, Sect. 2 presents discrete optimization problems formulation. A discrete Jaya optimization technique has been proposed in Sect. 3, while Sect. 4 demonstrates a numerical example based on truss design. Finally, Sect. 5 presents conclusions.

2 Formulation of Optimum Discrete Problem

This section describes the formulation of discrete optimization problems. Discrete variables are the set of predefined standard values. The discrete optimization formulation is given as

$$
\begin{aligned}
& Minimize \quad fn(x), \\
& subjected\,to: \; g_s(x) \leq 0, \quad s = 1, 2, \ldots, cn \\
& \qquad x_i \in D_i, D_i\big(d_{i1}, d_{i2}, d_{i3}, \ldots, d_{iA_i}\big), \quad i = 1, n_r \\
& \qquad d_{iA_i}^{L} \leq x_i \leq d_{iA_i}^{U}
\end{aligned}
\tag{1}
$$

where fn and g are objective and non-equality constraint functions, respectively. "n_r" and "cn" are discrete variables and numbers of total constraints, respectively, d_{ij} is the jth discrete value for the ith variable, A_i is the number of discrete values for the

ith variable, D_i is discrete values set for the ith variable. However, the number of discrete values may be different for each variable. $d_{iA_i}^U$ and $d_{iA_i}^L$ are upper and lower bounds of the discrete variable x_i.

The constraint optimization problem described in Eq. (1) is changed into an unconstrained problem using penalty formulation [13]. The objective function is penalized for infeasible solution for each constraint violation. Hence, the global optimum solutions are obtained while satisfying all the constraints. The original constrained problem is then posed as an unconstrained problem in which the first part represents the objective function and the second part is the penalty function. Finally, the discrete optimization problem is formulated as

$$\varphi(x) = fn(x) + \sum_{j=1}^{m} C_q * pn^s \tag{2}$$

$$d_{ip_i}^L \le x_i \le d_{ip_i}^U \tag{3}$$

where pn^s ($s = 1, 2, 3, \ldots, cn$) are the constants with the high value of order 10^5 to penalize the objective function if the constraints are violated and C_j is the Boolean Function [14] expressed as

$$C_q = \begin{cases} 0 \ if \ g_s(x) \le 0 \\ 1 \ \ otherwise \end{cases} \tag{4}$$

3   Discrete Optimization Based on Jaya Algorithm

Discrete algorithm based on Jaya is proposed for the discrete optimization problem formulated as in previous section. In this algorithm, the available discrete design variables are arranged in ascending order. Moreover, values of continuous design variables are converted into available values of discrete design variable using bound constraints of middle point of two consecutive discrete values in which the corresponding continues value lies as shown in Fig. 1. Further, best and worse solutions of objective function are compared with previous solutions at each iteration. The best solutions are stored and worse solutions are removed. The process of the algorithm continues until the stopping criteria are satisfied. The function evaluations and number of generations are considered as the stopping criterion for Jaya algorithm. Population size and the number of iterations are used to determine the number of function evaluations. The product of number of iterations and population size determine the number of function evaluations. Therefore, the number of design variables does not affect the function evaluations but algorithm computational time can be increased. The detailed procedure of this algorithm is explained by flowchart as shown in Fig. 1. Moreover, this algorithm reduces the computational effort then the
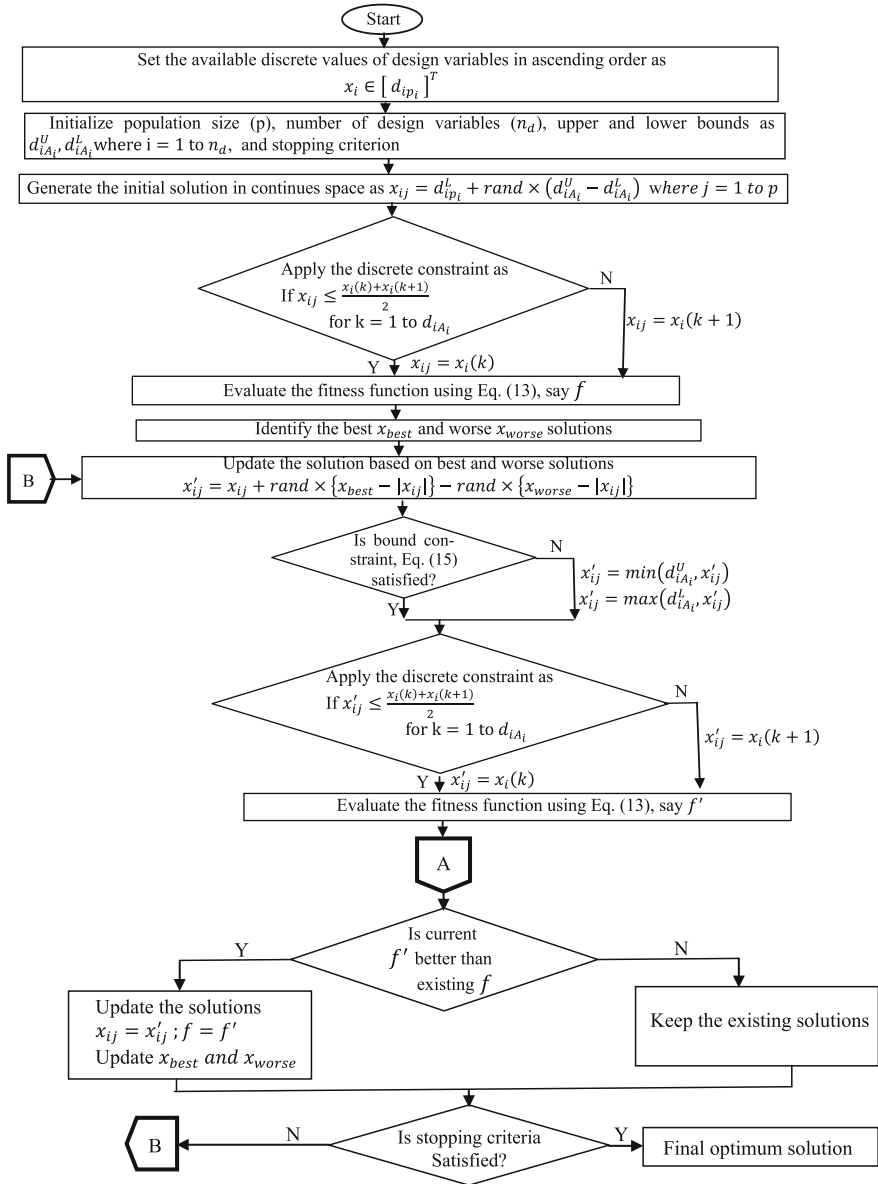
Fig. 1 Flowchart of discrete Jaya algorithm

other discrete optimization algorithms. However, for the first time, it is applied for discrete optimization problems in this study.

4 Numerical Example

In this section, discrete Jaya algorithm as described in previous section is tested by planar 10-bar truss design [10] as shown in Fig. 2. This truss design problem is solved using GA [5], SA [6], HPSO [7], ACO [8], ABC [9], and TLBO [10].

In this truss structure design, minimization of weight of truss is considered as objective function with constraints of displacements at each nodal point and the stress induced in each member. The cross-sectional areas of the each member are taken as a discrete design variable. The nodes 2 and 4 are subjected to a vertical nodal loads of 100 kips. Modulus of elasticity of material of each bar and density are taken as $E = 10,000$ ksi and $\rho = 0.1$ lb/in^3, respectively. The allowable displacements for the free nodes and the permissible stress for all members are taken as ± 2 in. in both directions and ± 25 ksi, respectively. Ten discrete design variables and their values are chosen from the standard set $D = \{1.62, 1.99, 1.80, 2.13, 2.38, 2.62, 2.63, 2.88, 3.09, 2.93, 3.13, 3.38, 3.47, 3.55, 3.63, 3.84, 3.87, 3.88, 4.18, 4.22, 4.49, 4.59, 4.80, 4.97, 5.12, 5.74, 7.22, 7.97, 11.5, 13,5, 14.2, 13.9, 15.5, 16.0, 16.9, 18.8, 1.99, 22.0, 22.9, 26.5, 30.0, 33.5\}$ (in^2) [10].

For the truss design, MATLAB code is developed for the proposed algorithm. The effectiveness of the algorithm is compared for same problem with discrete optimization algorithm as given in literature. The population size and the number of generations are taken as 10 and 95, respectively. The best objective function value corresponding to the design variables are found for 10 independent runs corresponding to each iteration. The optimal results comparison of planar 10 bar truss design under loading condition is given in Table 1. Convergence rates of best objective function are shown in Fig. 3. Table 1 shows that the Jaya algorithm takes less function evaluation for finding the best objective function compared to other discrete optimization techniques.

Fig. 2 A planar 10-bar truss structure [10]
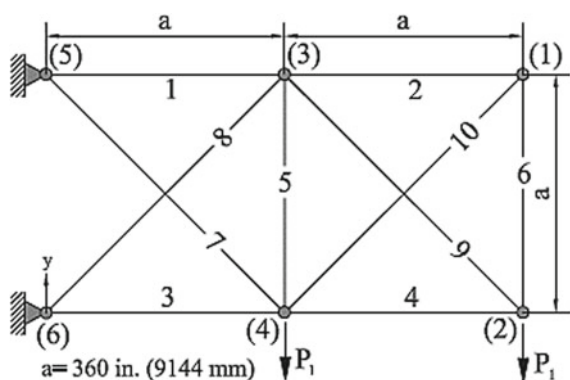
Table 1 Optimum design comparison for planar 10-bar truss structure

Design variables (in^2)	GA [5]	SA [6]	HPSO [7]	ACO [8]	ABC [9]	TLBO [10]	This study
A_1	33.5	33.5	30	33.5	33.5	33.5	**33.5**
A_2	1.62	1.62	1.62	1.62	1.62	1.62	**1.62**
A_3	22	22.9	22.9	22.9	22.9	22.9	**22.9**
A_4	15.5	14.2	13.5	14.2	14.2	14.2	**14.2**
A_5	1.62	1.62	1.62	1.62	1.62	1.62	**1.62**
A_6	1.62	1.62	1.62	1.62	1.62	1.62	**1.62**
A_7	14.2	7.97	7.97	7.97	7.97	7.97	**7.97**
A_8	19.9	22.9	26.5	22.9	22.9	22.9	**22.9**
A_9	19.9	22	22	22	22	22	**22**
A_{10}	2.62	1.62	1.8	1.62	1.62	1.62	**1.62**
W (lb)	5,613.84	5,490.74	5,531.9	5,490.74	5,490.74	5,490.74	**5,490.74**
Fun. Eval.	N/A	N/A	50,000	10,000	25,800	10,000	**9500**
Constraints violation	None	None	None	None	None	None	**None**
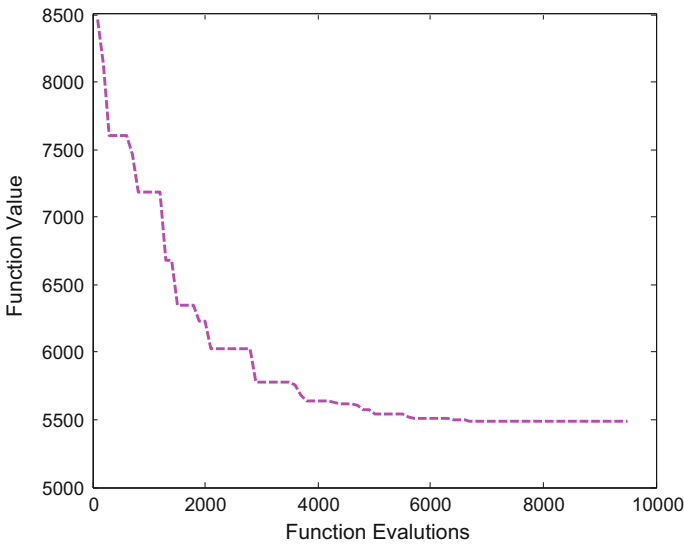
Fig. 3 Convergence of best objective function for planar 10-bar truss structure

5 Conclusion

In this paper, discrete Jaya algorithm is proposed for the discrete optimization problems. However, Jaya algorithm has been developed for continuous optimization problems. In this study, continuous variables are converted into discrete variable using discrete constraint. Furthermore, the efficiency of discrete Jaya algorithm is demonstrated by 10-bar planar truss structure problem taken from literature. Moreover, the optimum results obtained from proposed algorithm are compared with the results of well-known discrete optimization algorithms. The results shows that it minimizes the total weight of truss structure without violation of the design constraints and gives the better results compared to other discrete optimization algorithms. Other discrete optimization problems can be effectively solved using this algorithm.

References

1. Arora, J.S., Huang, M.W., Hsieh, C.C.: Methods for optimization of nonlinear problems with discrete variables: a review. Struct. Multidiscip. Optim. **8**(2), 69–85 (1994)
2. Arora, J.S.: Methods for discrete variable structural optimization. In: Burns, S.A. (ed.) Recent Advances in Optimal Structural Design. Technical Committee on Optimal Structural Design. ASCE, Reston, VA, pp. 1–35 (2002)
3. Arora, J.S.: Introduction to Optimum Design, 3rd edn. Reed Elsevier India Private Limited, India (2012)
4. Rao, R.V., Savsani, V.J.: Mechanical Design Optimization Using Advanced Optimization Techniques. Springer Science & Business Media (2012)
5. Rajeev, S., Krishnamoorthy, C.S.: Discrete optimization of structures using genetic algorithms. J. Struct. Eng. ASCE **118**(5), 1233–1251 (1992)
6. Kripka, M.: Discrete optimization of trusses by simulated annealing. J. Braz. Soc. Mech. Sci. Eng. **XXVI**, 170–173 (2004)
7. Li, L.J., Huang, Z.B., Liu, F.: A heuristic particle swarm optimization method for truss structures with discrete variables. Comput. Struct. **87**, 435–444 (2009)
8. Camp, C.V., Bichon, B.J.: Design of space trusses using ant colony optimization. J. Struct. Eng. ASCE **130**, 741–751 (2004)
9. Sonmez, M.: Discrete optimum design of truss structures using artificial bee colony algorithm. Struct. Multidiscip. Optim. **43**(1), 85–97 (2011)
10. Dede, T.: Application of teaching-learning-based-optimization algorithm for the discrete optimization of truss structures. ksce J. Civ. Eng. **18**(6), 1759–1767 (2014)
11. Rao, R.V.: Jaya: a simple and new optimization algorithm for solving constrained and unconstrained optimization problems **7**, 19–34, (2016). https://doi.org/10.5267/j.ijiec.2015.8.004
12. Rao, R.V., Waghmare, G.G.: A new optimization algorithm for solving complex constrained design optimization problems **273**, (2016). https://doi.org/10.1080/0305215x.2016.1164855
13. Singh, R., Chaudhary, H., Singh, A.K.: Defect-free optimal synthesis of crank-rocker linkage using nature-inspired optimization algorithms **116**, 105–122 (2017). https://doi.org/10.1016/j.mechmachtheory.2017.05.018
14. Singh, R., Chaudhary, H., Singh, A.K.: A new hybrid teaching–learning particle swarm optimization algorithm for synthesis of linkages to generate path. Sādhanā **42**(11), 1851–1870 (2017)

High-Gain L Probe-Fed Planar and Cylindrical Patch Antenna for X Band Applications

S. K. Kundu, Damanpreet Singh Walia, Shashank Jaiswal and P. K. Singhal

Abstract L probe provides a huge amount of impedance bandwidth by mounting the patch on a planar and cylindrical substrate. A parametric study and a comparative analysis between planar and cylindrical patch has been presented. Both planar and conformal patches have been designed in X band frequency. Around 2.65 GHz and 3 GHz impedance bandwidth and about 9.66 dBi and 8.37 dBi average gain has been achieved for both planar and conformal antenna, respectively. A stable radiation patterns across the pass band in both the antennas is observed. Design parameters and its effects have been thoroughly studied. Resonant input impedance is plotted. All the simulation work has been carried out in CST EM Simulation Software.

Keywords L probe · EPR · Co and cross-polarization · Input impedance Conformal antenna

1 Introduction

International Telecommunications Union (ITU) have assigned a part of X band frequency for space telecommunication, terrestrial communications and networking, terrestrial broadband, amateur radio, Electron Paramagnetic Resonance (EPR). X band frequency range from 8.0 to 12.0 GHz is a portion of the EM spectrum specified by IEEE. Not only the Telecommunication era, this band also supports the use of radar including continuous wave, pulsed, SAR, and other applications. Due to its short wavelength, it allows the higher resolution imagery for target identification and discrimination . Leading countries are using SATCOM in X band for difficult com-

S. K. Kundu (✉) · D. S. Walia · S. Jaiswal
Maharaja Agrasen Institute of Technology, New Delhi, India
e-mail: sunny_19662002@yahoo.co.in

P. K. Singhal
Madhav Institute of Technology and Science, Gwalior, Gwalior, India

munication systems. For the strategic requirement to dimension, physical size, the X band SATCOM is structured for naval and maritime platforms. This paper presents a theoretical as well as simulated result of L probe antenna in X band with wide bandwidth of conformal patch antenna in rectangular shape, embedded on a planar and cylindrical ground substrate. Cylindrical microstrip patch antennas have many applications in marine (naval) engineering, fighter aircrafts, and spacecrafts, where planar patches have constraint to be placed due to its polarization impurity [1]. One of the major weakness of the microstrip antenna is narrow bandwidth. Small relative bandwidth (BW), 2–10% has normally been observed in microstrip antenna [2]. There are a number of methods which have been studied to increase the bandwidth of microstrip antenna such as (i) using parasitic patch [3] (ii) use of thick substrate [4]. In [4], thicker substrate with co-axial- fed microstrip antenna will cause a huge cross-polarization in H plane and reduce the bandwidth. The probe inductance can be canceled out by inserting the capacitance produced by etching a circular slot in the patch, by etching U slot and capacitive feeding [5–7]. A novel approach was used to achieve a huge bandwidth using L probe (which itself acting as antenna) technique [8, 9]. L-shaped feed is the best option for thick substrate [8, 9]. With the use of L probe, the bandwidth can be enhanced up to 35% and average gain up to 7.5 dBi [10]. Patch is proximity fed, where L-shaped feed can be thought of a simple monopole antenna, whose resonance is considered to be nearly the same (exciting different modes) as that of patch for getting broadband operation [11]. It must be kept in mind that radiation patterns of both antennas (resonant dipole and patch) and the field which is coming out from the resonant L probe must be properly adjusted and merged to get high gain. The input impedance has been studied from [12–15]. The maturity of planar antenna compared with conformal antenna has been reached to its edge. Since the past decade, conformal antennas are getting its attention in terms of its theoretical works and patches on nonplanar surfaces like cylindrical, spherical, and conical bodies have been reported [16].

To enhance the bandwidth and radiation pattern on conformal surfaces, many techniques have been studied [17]. Patch antennas when used in surfaces of airplane, can radiate from the top and bottom surfaces of airframes, respectively. Patch antennas are placed on printed circuit boards that are connected to GPS and Radar Altimeter and are easier to scale down [18, 19]. Drag against wind is a serious issue for designing an antenna in the airplane, and fortunately, it reduces the extra drag as it is totally invisible because of its conformability with the surface. Lesser weight, ease in fabrication, and integration with MIC make it attractive for the application in airplanes [1, 19]. The rectangular planar patch as well as conformal cylindrical patch is fed electromagnetically by L probe, where the concept of cavity model with modal expansion techniques have been used to analyze both.

2   Theoretical Analysis

In Fig. 1, the geometry of the L probe fed with electromagnetically coupled planar and cylindrical patch and its equivalent circuit [12–15] is shown, where C_1 is the capacitance between horizontal part of the L probe and the ground plane, C_2 is the capacitance between horizontal portion of the strip and ground plane due to fringing effect at the end of strip. C_3 is capacitance of the horizontal portion of the L probe and the patch separated by a finite distance and C_4 is the capacitance between horizontal portion of the strip and patch due to fringing effect at the end of strip, and also R_s and L_s are the are resistance and inductance of the vertical part of the L probe. Low and unwanted radiation and the easiest modeling techniques with largest bandwidth is one of the advantages of proximity coupling. Between the parallel portions of L probe with that of the patch, a large amount of electromagnetic coupling is produced. The patch antenna is modeled as series R-L circuit with the combination of capacitor as parallel [20] and the L probe as a two-wire transmission lines resonant antenna, one is vertical with the patch and another one is horizontal. But these two ideal lines is no more planar but cylindrical in shape. Electromagnetic induction between the patch and the horizontal L probe can be thought of coupled lines (MACLIN component) [11]. As it is known from the basic analysis [21], wide bandwidth can be achieved using thick substrate at the cost of surface wave loss and increase of spurious feed radiation [14] Substrate of air ($\varepsilon_r = 1.00059$) with 3.2 mm has been used to achieve the desired results. Unlike the vertical part, the horizontal part will also offer some capacitance as shown.

Radial electric field along the feed line and magnetic field around feed line are responsible for the voltage and current developed in the probe [22]. In this case, the L probe is taken as cylindrical shape whose equivalent width can be thought of $2\pi a$, where a is the radius of the cylinder of L probe. The proposed L probe is nothing but as two-wire resonant dipole antenna with image ground plane, whose equivalent circuit of open-ended horizontal feed line is given below in Fig. 2 with one side opened.

Capacitances are responsible to nullify the stored energy due to the inductance effect of vertical probe and horizontal probe. The electric field radiated from the dipole (L Probe) will act as a source for patch, which is placed above horizontal portion of the L Probe. To obtain optimum wide bandwidth, the dimensions of the patch are adjusted with that of L probe horizontal feed line.

In case of cylindrical L probe feed, the metallic patch of curvature shape will have the different value of capacitance per unit length unlike planar patch and Laplace's equation can be solved by conformal mapping [16] to find the capacitance per unit length of the condenser parallel between patch and portion of L probe and between ground plane and portion of L probe. The equivalent capacitance of the patch C_{total} is because of capacitances and inductances given in Fig. 1c [12–15] as per transmission line model and using the relation, input impedance can be obtained from [13–15, 23]. The total impedance is given by [12–15]
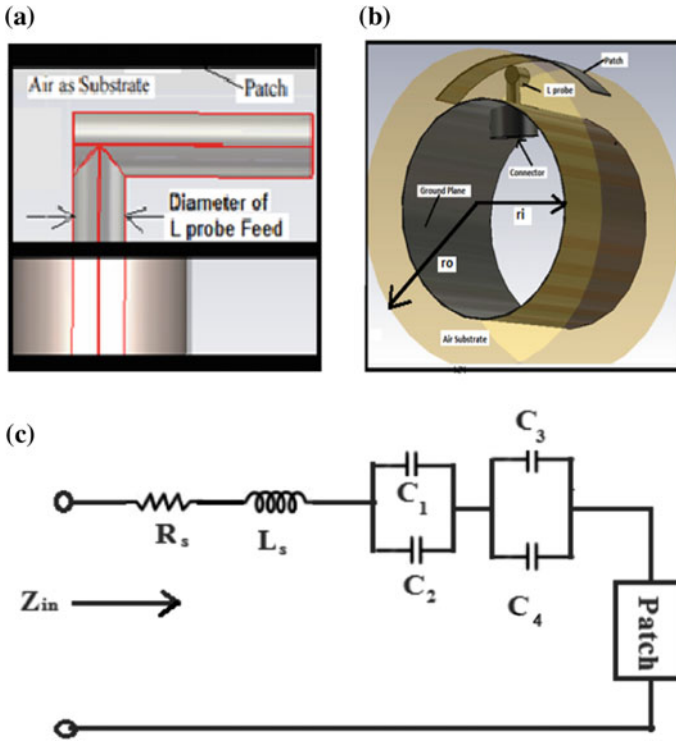
Fig. 1 Geometry of design proximity-fed **a** planar microstrip patch, **b** L-probe proximity-fed cylindrical antenna, **c** its equivalent circuit (of transmission line model) [12–15]
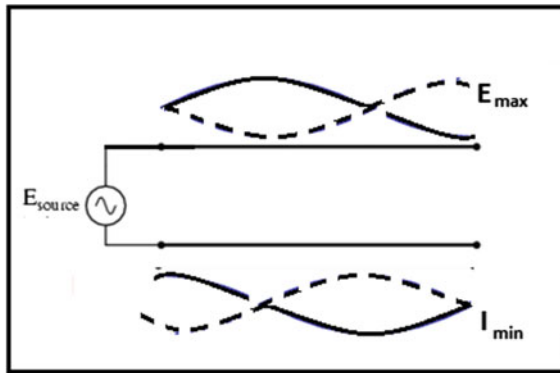
Fig. 2 Equivalent circuit of horizontal line with one-side opened acting as monopole

$$Z_{in} = R_s + j\omega L_s + \frac{1}{j\omega C_{total}} + Z_{patch} \qquad (1)$$

where C_{total} is the series–parallel combination of C_1, C_2, C_3, and C_4 [23] as per the equivalent circuit in Fig. 1c and fourth term is the impedance of the patch.

3 Antenna Configuration

Planar patch has the following design parameters as shown in Table 1.

Design parameters of the proposed cylindrical patch antenna are as shown in Table 2.

Table 1 Parameters for planar patch

Design parameters	Value
Height of the substrate	3.2 mm
Patch length	10.8391 mm
Patch width	14.9977 mm
Substrate dimension	44.99 mm × 44.99 mm
Permittivity of substrate; ε_r	1.00059
Vertical L probe length	1.964 mm
L probe radius	0.6 mm

Table 2 Design parameters for cylindrical conformal Patch

Design parameters	Value
Height of the substrate	3.2 mm
Patch length	10.8391 mm
Patch width	14.9977 mm
Axial length of copper cylinder	44.99 mm
Radius of the copper cylinder	7.16 mm
Permittivity of substrate; ε_r	1.00059
Vertical L probe length	1.964 mm
L probe radius	0.6 mm

4   Results and Discussion

The incident and reflected wave amplitudes at the port is shown in Fig. 3. The red curve indicates incident wave amplitude and the green curve indicates reflected wave amplitude. As compared, the planar patch has a less resonance effect than that of conformal cylindrical patch.

The plot of Fig. 4 also shows some important properties of feeding field. The upper side of the above figure shows the pattern of dominant mode TE_{11} modes with linear polarization aligned to the field lines. In an empty waveguide, the electric field components are orientated and are directed to ground, as can be seen in both figure on the left.
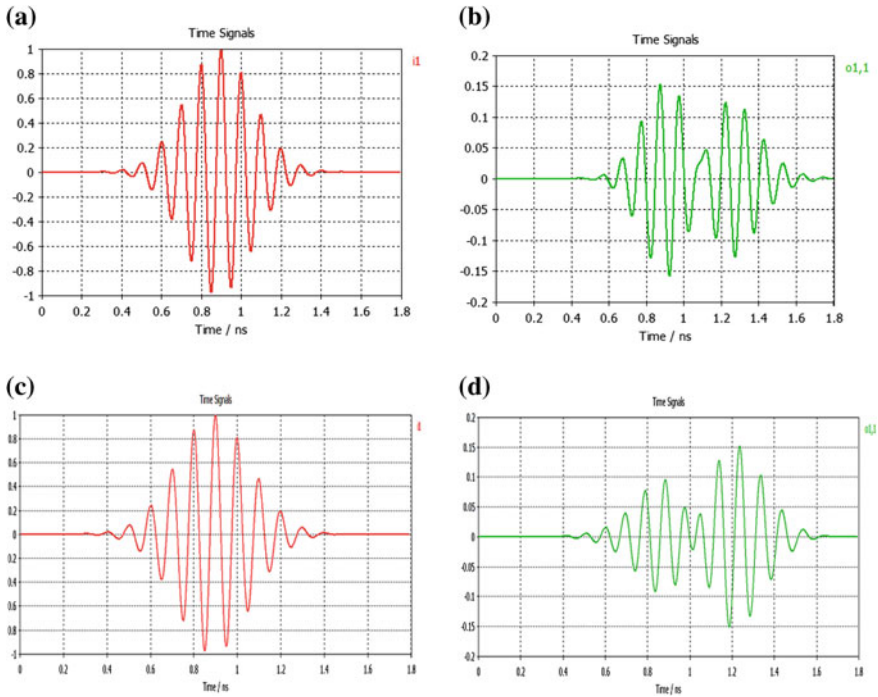
Fig. 3 Incident (red) and reflected (green) field amplitude distribution at the port only **a** planar patch antenna, **b** cylindrical patch antenna with base radius 7.16 mm
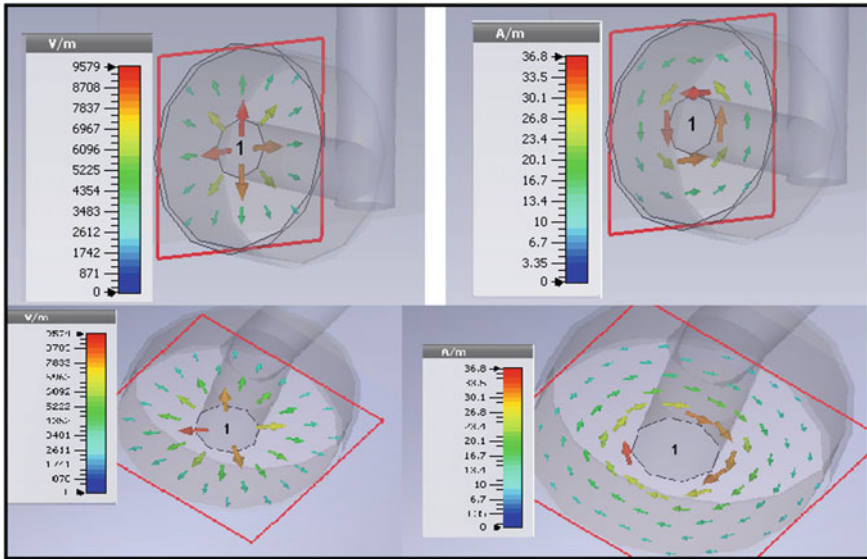
Fig. 4 E field and H field distribution at the port only of the planar and cylindrical patch antenna

Any other degenerate modes other than dominant mode are generated corresponding to the input port's coordinate system which is best suited for the empty waveguide port. Hence, E field and H Field will have the same input distribution.

Figures 5 and 6 indicate the real and imaginary part of input impedances and impedance bandwidth related to return loss in planar and cylindrical antenna. It is observed that with same dimension, the proximity-fed L probe of the planar patch has lower impedance bandwidth compared to that of cylindrical patch; this is because the inaccuracy in input impedance in planar patch. Four separate single-patch antennas with identical dimension are simulated at 10 GHz, having radius of $= \infty$ (planar), 20 mm, 10 mm and 7.16 mm, respectively, as shown in Fig. 6. Each patch has dimensions of 10.8391 × 14.9977 mm and is fed by constant-length section of L probe of cylindrical shape, whose open end is assumed to be of the shape of resonant-type dipole antenna. The axial length of the cylinders is 44.99 mm, which is equal to the dimension of ground plane of planar patch.
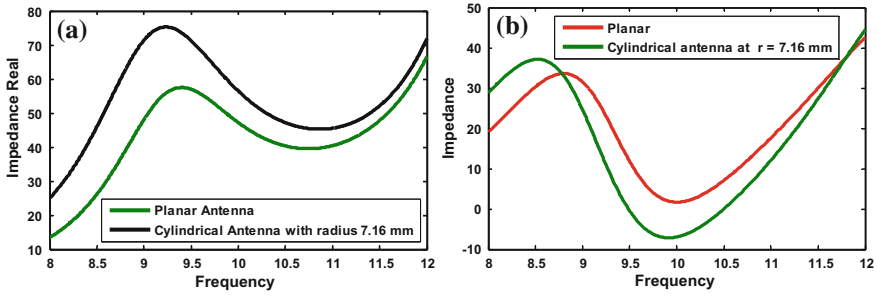
Fig. 5 Input impedance **a** real and, **b** imaginary part of L probe patch for planar Patch and conformal cylindrical Patch with radius of cylinder = 7.16 mm

Fig. 6 Simulated return loss for planar and cylindrical L probe patch antenna with the same dimension but different radius of cylinder
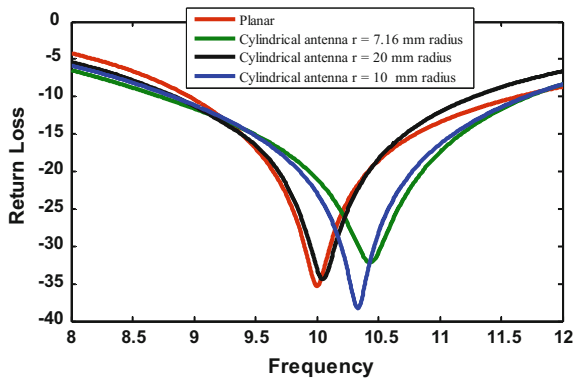
Figures 7 and 8 indicates much better co polarization effects on the radiation pattern at $\Phi = 0°$ and $\Phi = 90°$, respectively, in planar and cylindrical patch. In Figs. 9 and 10, the antenna has very high cross-polarization attenuation. In Figs. 9 and 10, the H plane shows more cross-polarization. Figure 5 shows that the input impedance of a planar patch and cylindrical patch antenna and indicates that more is the curvature of cylinder; the less is the input impedance of a polarized patch which is acceptable in engineering designs. The input impedance of the probe and patch is the combination
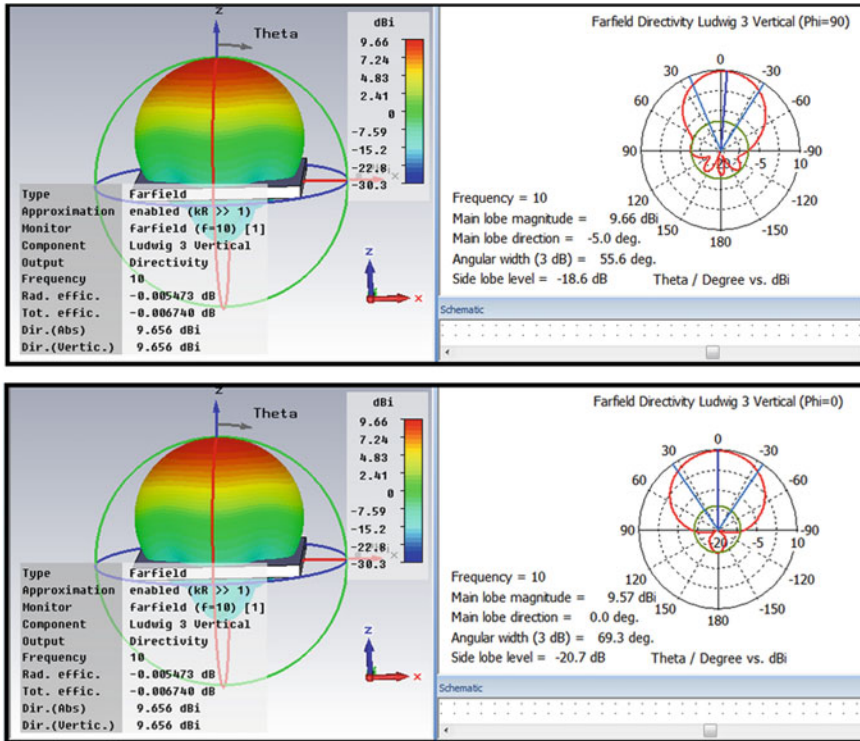
Fig. 7 Co-polar component: three-dimensional radiation patterns at 10 GHz of the proposed planar patch at $\Phi = 0^\circ$ and $\Phi = 90^\circ$ (using Ludwig-3 scheme)

of self-impedance of the probe due to current density within itself with absence of the patch and with presence of the patch [24]. The impedance due to curvature is approximated with that of planar patch when radius of curvature $R \gg$ the substrate thickness and wavelength and it varies with propagation constant along ρ and z direction and zero-order Bessel's function [25–27]. The impact due to curvature effect of the patch disrupts the match hence radiation and gain as well frequency with increase of curvature also decreases. The frequency is decreased as a result that the cylindrical antenna with exact dimension as planar may be applied where compactness is desirable.
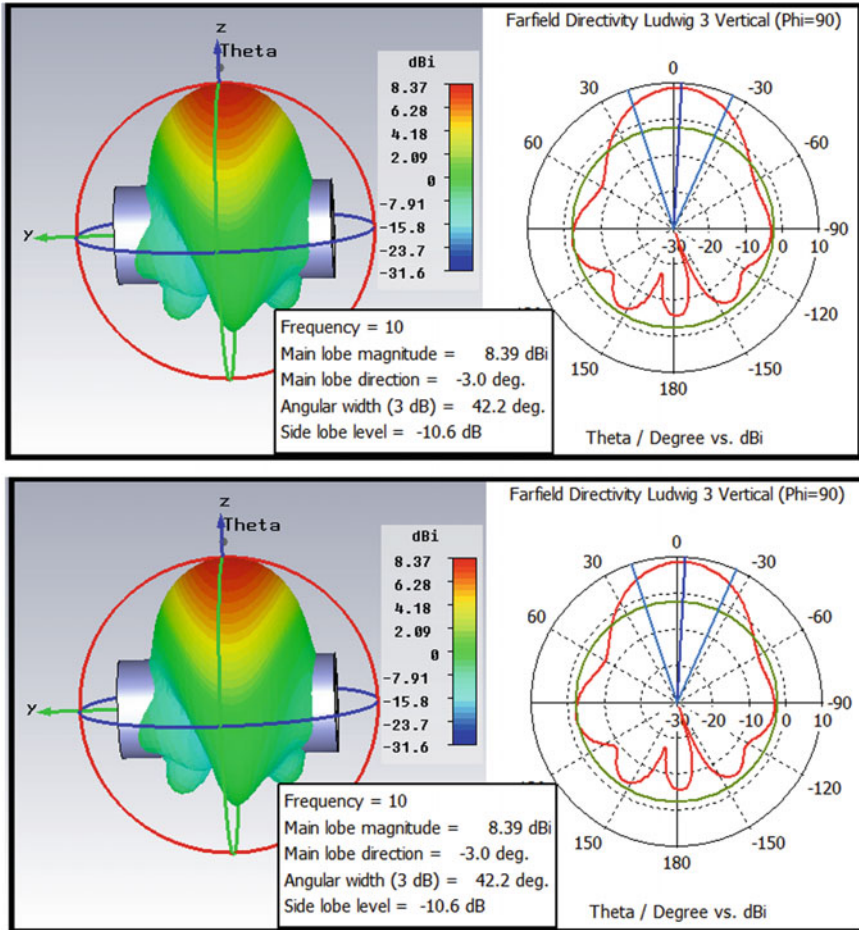
Fig. 8 Co-polar component: Three-dimensional radiation patterns at 10 GHz of the proposed cylindrical patch at $\Phi = 0°$ and $\Phi = 90°$ with radius of cylinder $= 7.16$ mm (using Ludwig-3 scheme)
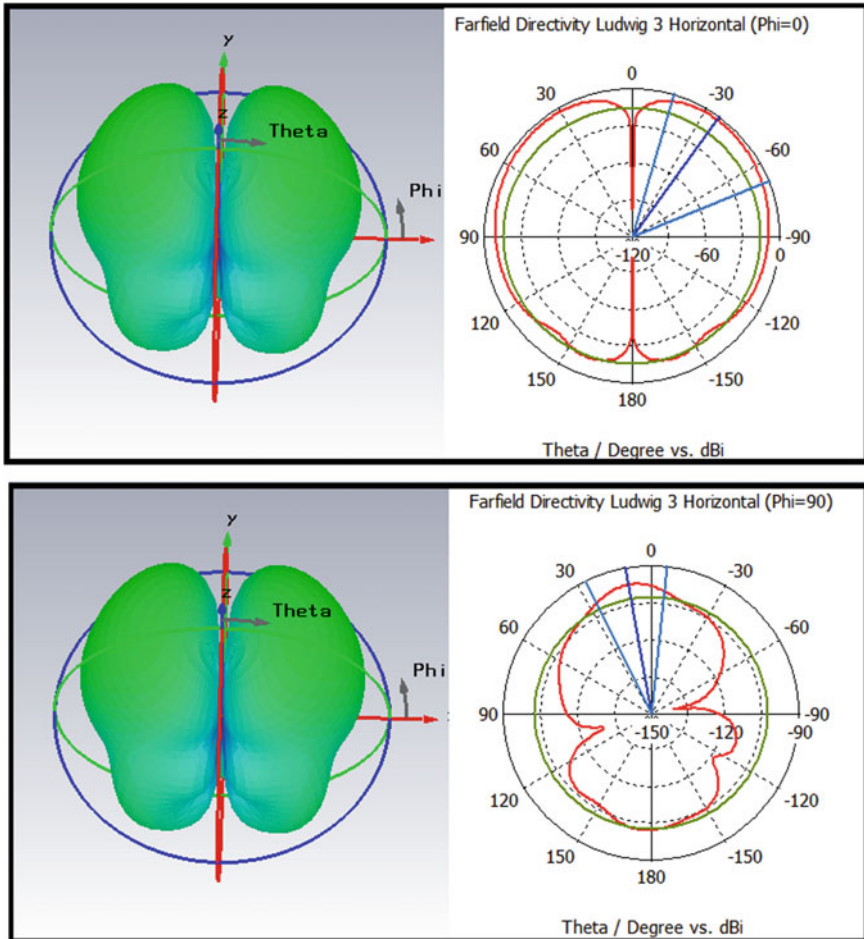
Fig. 9 Cross-polar component: three-dimensional radiation patterns at 10 GHz of the proposed planar patch at $\Phi = 0°$ and $\Phi = 90°$ (using Ludwig-3 scheme)
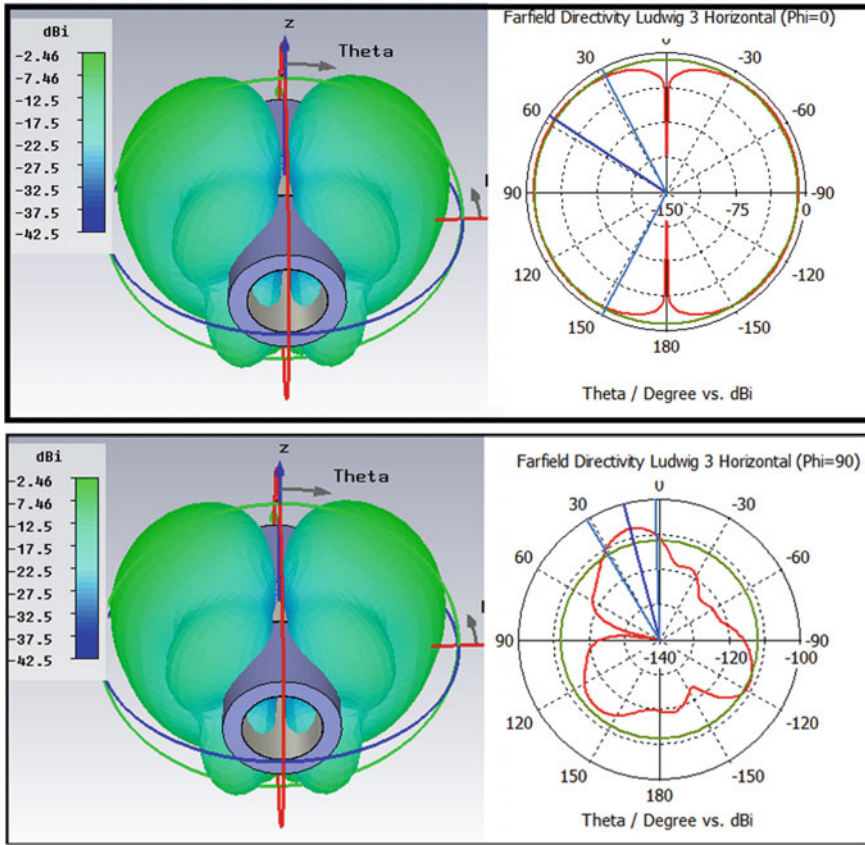
Fig. 10 Cross-polar component: three-dimensional radiation patterns at 10 GHz of the proposed cylindrical patch at $\Phi = 0°$ and $\Phi = 90°$ with radius of cylinder $= 7.16$ mm (using Ludwig-3 scheme)

5 Conclusion

Proposed L probe broadband single-element planar patch and conformal patch achieves a larger bandwidth. An impedance bandwidth of about 2.65 GHz and 3 GHz [S_{11} <10 dB] have been obtained for planar and conformal antennas respectively. The planar antenna and conformal antennas resonates at 10 GHz and 10.432 GHz with radius 7.16 mm, respectively, which will be best suited for space telecommunication, terrestrial communications, and radar. Both the antennas are compared with each other on the basis of each and every parameter and are also verified.

References

1. James, J.R., Hall, P.S.: Handbook for Microstrip Antenna. Peter Peregrinus, London
2. Balanis, C.A.: Antenna Theory, Analysis and Design. Wiley, New York (1997)
3. Lee, R.Q., Lee, K.F., Bobinchak, J.: Characteristics of a two-layer electromagnetically coupled rectangular patch antenna. Electron. Lett. **23**, 1070–1072 (1987)
4. Chang, E., Long, S.A., Richards, W.F.: Experimental investigation of electrically thick rectangular microstrip antennas. IEEE Trans. Antennas Propag. **AP-43**, 767–772 (1986)
5. Hall, P.S.: Probe compensation in thick microstrip patches. Electron. Lett. **23**, 606–607 (1987)
6. Lee, K.F., Luk, K.M., Tong, K.F., Shum, S.M., Huynh, T., Lee, R.Q.: Experimental and simulation studies of coaxially fed U-slot rectangular patch antenna. Inst. Elect. Eng. Proc. Microw. Antennas Propagat. **144**, 354–358 (1997)
7. Vandenbosch, G.A.E., Van De Capelle, A.R.: Study of the capacitively fed microstrip antenna element. IEEE Trans. Antennas Propagat. **42**, 1648–1652 (1994)
8. Luk, K.M., Mak, C.L., Chow, Y.L., Lee, K.F.: Broadband microstrip patch antenna. Electron. Lett. **34**, 1442–1443 (1998)
9. Mak, C.L., Luk, K.M., Lee, K.F.: Wideband L-strip fed microstrip antenna. In: IEEE AP-S International Symposium USNC/URSI National Radio Science Meeting, Orlando, FL, July 1999, pp. 1216–1219
10. Mak, C.L., Luk, K.M., Lee, K.F., Chow, Y.L.: Experimental study of a microstrip patch antenna with an L shaped Probe. IEEE Trans. Propag. **48**(5), May 2000
11. Hazdra, P., Mazánek, M., Čermák, J.: Wideband rectangular microstrip patch antenna using L probe feeding system. Radio Eng. **16**(3) (2007)
12. Shum, S.M., Luk, K.M.: FDTD analysis of probe-fed cylindrical dielectric resonator antennas. IEEE Trans. Antennas Propagat. **46**, 325–333 (1998)
13. Singh, A.K., Gangwar, R.K., Kanaujia, B.K.: Bandwidth enhancement of L probe proximity-fed annular ring microstrip slot antenna. In: 6th International Conference on Advanced Infocomm Technology (ICAIT) (2013)
14. Hoffman, R.K.: Handbook of Microwave Integrated Circuits. Artech House, Norwood, MA (1987)
15. Kala, P., Reena, P.: L probe fed planar rectangular microstrip patch antenna mounted on cylindrical ground surface. Int. J. Microw. Opt. Technol. **5**(2) (2010)
16. Wong, K.L.: Design of Nonplanar Microstrip Antennas and Transmission lines. Wiley, Inc. (1999)
17. Wong, K.L.: Compact Broadband Antenna. Wiley, Inc.
18. Macnamara, T.M.: Introduction to Antenna Placement and Installation, 1st edn. Wiley, Ltd. (2010)
19. Hamid, A.-K.: Multi-dielectric loaded axially slotted antenna on circular or elliptic cylinder. J. Electromagn. Waves Appl. **20**(9), 1259–1271 (2006)
20. Bahl, I., Bhartia, P., Garg, R.: Microstrip Antenna Design Handbook. Artech house (2001)
21. Kumar, G., Ray, K.P.: Broadband Microstrip Antennas. Artech House, Inc. (2003)
22. Guo, Y.X., Mak, C.L., Luk, K.M.: Analysis and design of L probe proximity fed-patch antennas. IEEE Trans. Antennas Propagat. **49**(2) (2001)
23. Edwards, T.C.: Foundations for Microstrip Circuit Design. Wiley (1983)
24. Habashy, T.M., Ali, S.M., Kong, J.A.: Input impedance and radiation pattern of cylindrical rectangular and wrapround microstrip antennas. IEEE Trans. Antennas Propagat. **38**(2), 722–731 (1990)
25. Chew, W.C., Kong, J.A., Shen, L.C.: Radiation characteristics of a circular microstrip disk antenna. J. Appl. Phys. **57**
26. Chew, W.C., Kong, J.A.: Analysis of circular microstrip disk antenna with thick dielectric substrate. IEEE Trans. Antennas Propagat. **AP-29**, 68–76 (1981)
27. Habashy, T.M., Ali, S.M., Kong, J.A.: Impedance parameters and radiation pattern of two coupled microstrip disk antennas. J. Appl. Phys. **54**, 493–509 (1983)

Mitigating Primary User Emulation Attacks Using Analytical Model

Ishu Gupta and O. P. Sahu

Abstract Cognitive radio has been a widely studied area to increase the efficiency of the spectrum. Most of the present work focuses on sensing techniques but very less work has been done for security in Cognitive Radio Networks (CRNs). Primary User Emulation Attack (PUEA) is a very prominent Denial of Service (DoS) attack that degrades the performance of the network to a large extent. This paper deals with an analytical model which depends on Neyman–Pearson Composite Hypothesis Test (NPCHT) to determine whether a PUEA is present in CRN. Log-normal shadowing and Rayleigh fading have been taken for the signals received from the primary transmitters as well as the attackers. The movement of good secondary user has been assumed in vertical direction ranging from an angle of $(-\pi/4)$ to $(\pi/4)$ with respect to primary transmitter. The performance of the system model has been evaluated by plotting the Receiver Operating Characteristic (ROC) curve. Variation in miss detection and false alarm probabilities has been studied for different angles. Results demonstrate that the likelihood of successful PUEA increments with the increasing distance between the good secondary user and primary transmitter.

Keywords Dynamic spectrum access · Cognitive radio · Primary user emulation attacks · Probability of false alarm · Probability of miss detection

1 Introduction

With increasing use of the multimedia applications, the demand for spectrum is continuously increasing. However, the spectrum available is very limited as well as costly. Hence, there is a need to utilize the spectrum very judiciously. The presently adopted

I. Gupta (✉) · O. P. Sahu
Department of Electronics and Communication Engineering, National Institute
of Technology, Kurukshetra, Kurukshetra 136119, Haryana, India
e-mail: ishugupta2203@gmail.com

O. P. Sahu
e-mail: ops_nitk@yahoo.co.in

Static Spectrum Allocation (SSA) policy leads to underutilization of spectrum [1]. J. Mitola suggested the Dynamic Spectrum Allocation (DSA) method to solve this problem [2]. Later, quantification of Mitola's work was done by Haykin [3]. Cognitive radio is a technology that is widely studied nowadays and can possibly eliminate the SSA problem in future. Spectrum sensing is a very crucial task to practically implement this technology. Various spectrum sensing methods have been discussed in [4]. Among all, energy detection has been widely adopted and accepted method for sensing of spectrum, because it does not require any prior knowledge regarding the features of the primary signal and is computationally simple to implement [5]. Most of the present work in Cognitive Radio Technology focuses on spectrum sensing and its optimization. However, for successful implementation for any technology, its security aspects need to be taken into account. A comprehensive survey on different security attacks and their detection has been presented in [6]. This paper focuses on one such Denial of Service (DoS) attack named Primary User Emulation Attack (PUEA) [7]. In such an attack, an adversary tries to imitate the characteristics of primary user so as to fool the good secondary user regarding the usage of channel. The good secondary user believing that channel is being used by the Primary User leaves the spectrum and hence making the use of DSA method trivial. The malicious users can use the channel for their own purpose. Various studies have been conducted on detection and elimination of this attack. The impact of PUEA on Cognitive Radio Networks has been studied in [8]. PUEA can be classified into two types—always on and smart [9]. Always on PUE attacks emulate the primary signal anytime without taking into account whether the primary signal is transmitting or not whereas smart PUE attackers take into account the transmission characteristics of primary signal. The impact of smart PUE attackers is more severe than always on PUE attackers. This paper focuses on always on PUEA.

This paper presents an analytical model for PUEA in which secondary users are assumed to move in a vertical direction. Effect on the probability of miss detection and false alarm has been considered for varying distances between real secondary user and primary user. The first analytical model was presented in [10], where the use of Markov inequality and Fenton's approximation was done so as to get a lower limit for the likelihood of successful PUEA. In [11], WSPRT is used for detection of PUEA by first formulating the probability density function (pdf) for the received power at the good secondary user from the malicious users. Jin et al. [12] compares the Wald's Sequential Probability Ratio test (WSPRT) and NPCHT for identification of PUEA using the same analytical model. It was found that WSPRT offers less likelihood of successful PUEA as compared to NPCHT because it allows two thresholds for both probability of false alarm and miss detection. This paper presents a similar model using NPCHT with the difference being that the good secondary user is allowed to move in a vertical direction that allows the study of probability of false alarm as well as miss detection with changing distance between the real secondary user and primary user. The simulated results match with the theoretical results that probability of successful PUEA increases with increasing distance between primary transmitter and the good secondary user.

The rest of the paper is structured as follows. Section 2 discusses the system model. Section 3 presents the analytical model. Section 4 describes the NPCHT approach for detecting PUEA. Results are analyzed in Sect. 5. Conclusion is provided in Sect. 6.

2 System Model

The system model in this paper has one secondary user surrounded by different randomly located malicious users in circular grid around it as given in Fig. 1 [11]. Let the radius of the circular grid be R. One primary transmitter is fixed at a distance greater than or equal to D_p from all the other users. The distance $D_p \gg R$. Energy-based detection technique has been used for detection of PUEA. If the power received at the secondary user exceeds a certain threshold say Λ, then the primary transmission is assumed else malicious transmission is assumed.

Two hypotheses have been taken as null and alternative hypothesis given as follows:

H0: Primary User Transmission. This includes the case when only primary user is transmitting the signal.

H1: Malicious User Transmission. This hypothesis considers the case when malicious and good secondary users are simultaneously transmitting the signal.

Following assumptions have been made for doing analytical study of the system model:

- Total M malicious users along with one good secondary user are present in the CRN.
- Minimum distance of primary transmitter from all other users should be D_p.
- The primary transmitter power is taken to be P_t.
- Power of each malicious user is taken to be P_m where $P_m \ll P_t$.
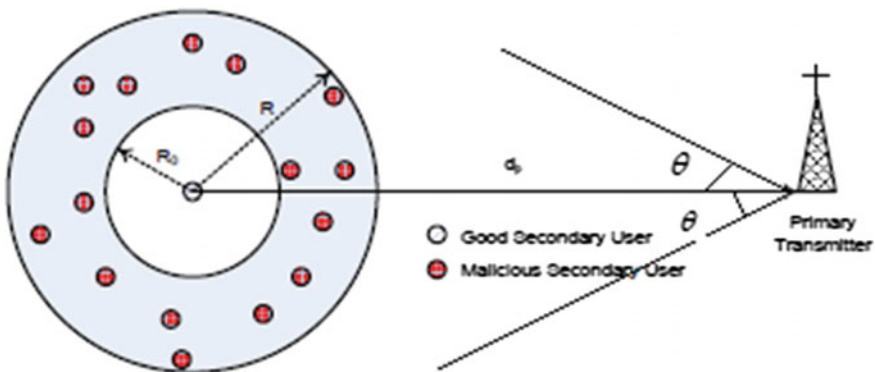
Fig. 1 Typical cognitive radio scenario

- The adversaries are uniformly distributed in a circular region around the good secondary user and their positions are independent to each other.
- The position of primary transmitter is fixed whereas the movement of secondary user is restricted in a vertical direction.
- The position of primary transmitter is known to the good secondary user and the adversaries.
- The incoming signals received from both the primary transmitter as well as malicious users undergo log-normal shadowing, path loss and Rayleigh fading.
- The mean, Δ of the Rayleigh random variable is assumed to be unity.
- The exponent for path loss is assumed to be 2 for propagation from primary transmitter and 4 from malicious users.
- There is an excluded distance R_0 from the good secondary user where there is no malicious user. Such condition is required since if there would be any malicious user present in this region, then the power received from the malicious users will be always higher than the power received from primary transmitter leading to successful PUEA every time.

3   Analytical Model

First of all, the probability density function (pdf) of the signal is analyzed. It is assumed that the good secondary user is fixed at origin (0, 0). It is also assumed that the motion of SU is restricted in a vertical direction ranging from a minimum angle of $-\pi/4$ to maximum angle of $\pi/4$ from primary transmitter. Since $Dp \gg R$, coordinates of primary transmitter can be assumed to be same for all the malicious SUs.

Let the coordinates of the M malicious users be (r_j, θ_j), where $1 \leq j \leq M$. The pdf of r_j is taken as

$$p(r_j) = \begin{cases} \frac{2r_j}{R^2 - R_0^2}, & R_0 \leq r_j \leq R \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

θ_j is taken as uniformly distributed in the range $(-\pi, \pi)$.

The power transmitted from primary transmitter to the SU can be taken as

$$P_r^{(p)} = P_t G_p^2 \left(\frac{d_p}{\cos(\theta)} \right)^{-2} \tag{2}$$

where $G_p^2 = 10^{\frac{\xi_p}{10}}$ and $\xi_p \sim N(0, \sigma_p^2)$. The corresponding pdf $p^{(P_r)}(\gamma)$ can be defined as log normal distribution which can be written as follows:

$$p^{(P_r)}(\gamma) = \frac{1}{A\sqrt{2\pi}\sigma_p\gamma} \exp\left\{\frac{-\left(10\log_{10}\gamma - \mu_p\right)^2}{2\sigma_p^2}\right\} \tag{3}$$

where $A = \frac{\ln 10}{10}$ and

$$\mu_p = 10\log_{10} P_t - 20\log_{10}\left(\frac{d_p}{\cos(\theta)}\right) \tag{4}$$

Since different malicious users are not dependent on each other, their power can be added which can be taken as

$$P_r^{(m)} = \sum_{j=1}^{M} P_m(d_j)^{-4} G_j^2 \tag{5}$$

where G_j^2 is the shadowing between the SUs and jth MU. $G_j^2 = 10^{\frac{\xi_j}{10}}$ and $\xi_j \sim N(0, \sigma_m^2)$. All the variables in above equation are log-normally distributed of the form $10^{\frac{\omega_j}{10}}$ and $\omega_j \sim N(\mu_j, \sigma_m^2)$. Here

$$\mu_j = 10\log_{10} P_m - 40\log_{10}(d_j) \tag{6}$$

By approximating the pdf from different malicious users using Fenton's approximation, pdf $p^{(m)}(\chi)$ can be given as

$$p^{(m)}(\chi) = \frac{1}{A\sqrt{2\pi}\sigma_\chi\chi} \exp\left\{\frac{-\left(10\log_{10}\chi - \mu_\chi\right)^2}{2\sigma_\chi^2}\right\} \tag{7}$$

The received power $P_r^{(m)}$ is assumed to be log-normally distributed random given by above equation. Here, μ_χ and σ_χ^2 can be taken as

$$\mu_\chi = \frac{1}{A} \ln\left[\frac{E^2\left[P_r^{(m)}\right]}{\left(\text{var}\left(P_r^{(m)}\right) + E^2\left[P_r^{(m)}\right]\right)^2}\right] \tag{8}$$

$$\sigma_\chi^2 = \frac{1}{A^2} \ln\left[\frac{\text{var}\left(P_r^{(m)}\right) + E^2\left[P_r^{(m)}\right]}{E^2\left[P_r^{(m)}\right]}\right] \tag{9}$$

The values of μ_χ and σ_χ^2 can be substituted in Eq. (7) to get the pdf for the malicious users. Further, the error probabilities, i.e., the probability of miss detection and false alarm can be analyzed by applying Neyman–Pearson criterion as defined in the following section.

4   Neyman–Pearson Criterion for Detection of Puea

In Neyman–Pearson criterion, to take any decision, two hypothesis are assumed which can be defined as follows:

H0: Primary User transmission (Null hypothesis).

H1: Malicious User transmission (Alternative hypothesis).

Any secondary user can undergo two types of errors during sensing of the channel, namely:

- Miss detection: This case corresponds to when the PU is actually transmitting the signal but the SU believes that it to be sent by the malicious user. The probability of detection, P_d is defined as $(1\text{-}P_m)$, where P_m is the likelihood of miss detection.
- False alarm: This error corresponds to case when the signal is transmitted by the MU but the good SU assumes the signal to be transmitted by the PU. This leads to interference to the PU and thus PU may impose a penalty to the good SU for further usage of the channel.

The Neyman–Pearson criterion aims at reducing the possibility of successful PUEA assuming a fixed false alarm probability say, α. The condition of fixed false alarm probability has to be imposed because both the error probabilities, i.e., miss detection and false alarm cannot be reduced simultaneously [13]. The ratio of two pdfs is taken and compared with a certain threshold, say Λ, which is taken as

$$\Lambda = \frac{p^{(m)}(x)}{p^{(P_r)}(x)} \tag{10}$$

where x can be taken as the measured power of the signal at SU. The final decision is based on criterion given as follows:

$$\Lambda \begin{cases} \leq \lambda & \text{D1: Primary Transmission} \\ \geq \lambda & \text{D2: Malicious User Transmission} \end{cases} \tag{11}$$

where λ can be obtained by a constraint on probability of miss detection $\Pr\left\{\frac{D2}{H0}\right\}$ which is fixed at a certain value α given by the following expression:

$$\Pr\left\{\frac{D2}{H0}\right\} = \int_{\Lambda \geq \lambda} p^{(P_r)}(x)dx = \alpha \tag{12}$$

Hence, Neyman–Pearson criterion aims at reducing the error probabilities while keeping a constraint as shown in the above equation.

5 Results and Discussion

The proposed work in this paper is simulated by assuming the following system parameters (Table 1).

The simulation is performed for 10,000 times. The performance of the model is illustrated by plotting probability of miss detection verses that of false alarm in Fig. 2. Plot of probability of detection versus that of false alarm, also called ROC is shown in Fig. 3. Finally, variation of probability of false alarm as well as miss detection verses ɵ can be analyzed in Fig. 4. The results in Fig. 2 are concordant with the theoretical results that both the errors probabilities cannot be reduced simultaneously and the mean probability of both false alarm as well as miss detection is found to be nearly varying from 0.01 to 0.05.

Similarly, the plot for Receiver Operating Characteristics (ROC) [13], which is generally used as another way for detection of a detector is shown in Fig. 3 and the mean P_d is found to be varying around 0.95–0.99.

Figures 4 and 5 demonstrate how the probability of false alarm and miss detection, respectively, varies with varying angles of the secondary user with the primary transmitter. Change in angles leads to change in distances between primary transmitter and real secondary user and thus, the error probabilities are minimum for ɵ = 0° and maximum for ɵ = −45° and +45°. Hence, as the distance between primary transmitter and secondary user increases, the likelihood of false alarm increases and vice versa.

Table 1 System parameters used for simulation

Parameter	σ_p (dB)	σ_m (dB)	D_p (km)	R (m)	R_0 (m)	P_t (kW)	P_m (W)	M	ɵ	Δ
Value	8	5.5	100	500	30	100	4	15	(−π/4): (π/2048): (π/4)	1

Fig. 2 Probability of miss detection versus false alarm
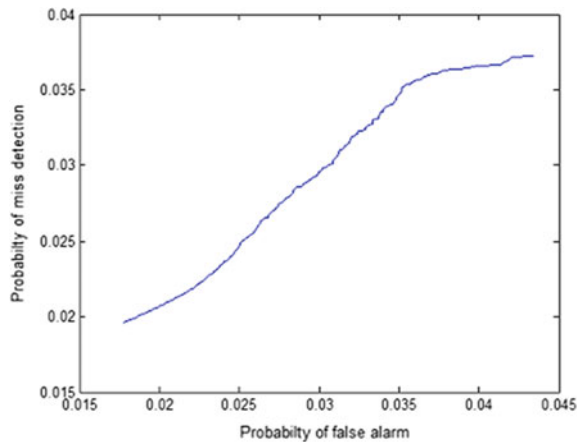
Fig. 3 Probability of
detection versus false alarm
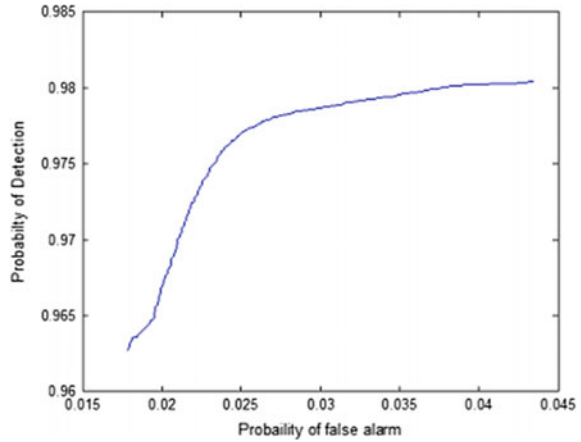
Fig. 4 Probability of false
alarm versus angle, ϴ
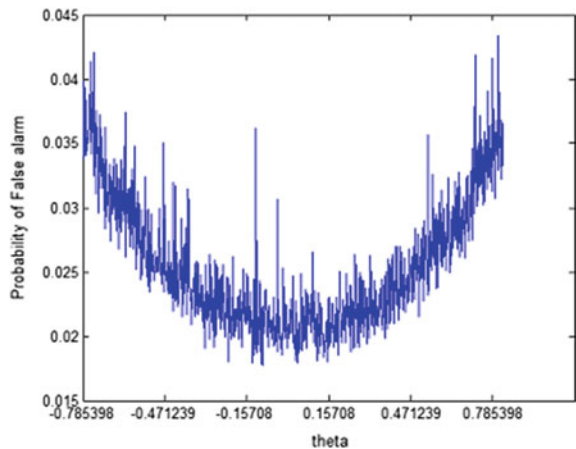
6 Conclusion

An analytical model for the practical Cognitive Radio Network has been proposed
based on Neyman–Pearson criterion and effects on probability of false alarm as
well as miss detection has been studied with varying distances between primary
transmitter and good secondary user. Movement of secondary user is assumed in a
vertical direction ranging from an angle of $-\pi/4$ to $\pi/4$ with the primary transmitter.
The simulated results show that error probabilities increments with the increasing
distances between primary transmitter and the secondary user. Future work can be
based on variable malicious users' power. Other analytical models can be developed
that deals with other distributions apart from uniform distributions. Also, different
techniques need to be developed that focuses on smart PUE attackers.

Fig. 5 Probability of miss detection versus angle, ⊖
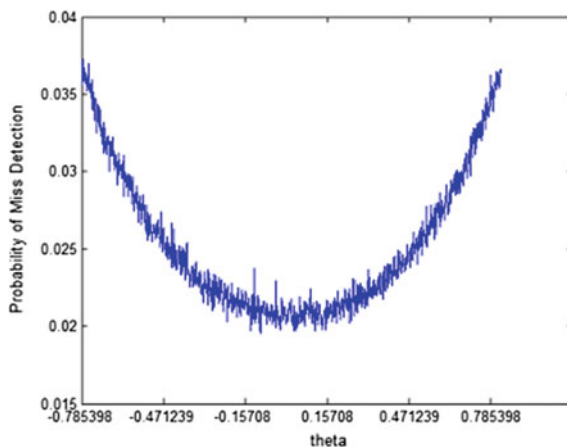
References

1. Force, S.: Spectrum policy task force report. Federal Commun. Comm. ET Docket 02, **135** (2002)
2. Mitola, J., Gerald, Q.: Maguire.: cognitive radio: making software radios more personal. IEEE Pers. Commun. **6**(4), 13–18 (1999)
3. Haykin, S.: Cognitive radio: brain-empowered wireless communications. IEEE J. Sel. Areas Commun. **23**(2), 201–220 (2005)
4. Yucek, T., Arslan, H.: A survey of spectrum sensing algorithms for cognitive radio applications. IEEE Commun. Surv. Tutor. **11**(1), 116–130 (2009)
5. Umar, R., Sheikh, A.U.H.: A comparative study of spectrum awareness techniques for cognitive radio oriented wireless networks. Phys. Commun. **9**, 148–170 (2013)
6. Fragkiadakis, A.G., Tragos, E.Z., Askoxylakis, I.G.: A survey on security threats and detection techniques in cognitive radio networks. IEEE Commun. Surv. Tutor. **15**(1), 428–445 (2013)
7. Chen, R., Park, J.M., Reed, J.H..: Defense against primary user emulation attacks in cognitive radio networks. IEEE J. Sel. Areas Commun. **26**(1) (2008)
8. Jin, Z., Anand, S., Subbalakshmi, K.P.: Impact of primary user emulation attacks on dynamic spectrum access networks. IEEE Trans. Commun. **60**(9), 2635–2643 (2012)
9. Shrivastava, S.: Security Issues in Cognitive Radios. Diss (2017)
10. Anand, S., Jin, Z., Subbalakshmi, K.P.: An analytical model for primary user emulation attacks in cognitive radio networks. In: 3rd IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks, DySPAN 2008. IEEE (2008)
11. Jin, Z., Anand, S., Subbalakshmi, K.P.: Detecting primary user emulation attacks in dynamic spectrum access networks. In: IEEE International Conference on Communications, ICC'09. IEEE (2009)
12. Jin, Z., Anand, S., Subbalakshmi, K.P.: Mitigating primary user emulation attacks in dynamic spectrum access networks using hypothesis testing. ACM SIGMOBILE Mob. Comput. Commun. Rev. **13**(2), 74–85 (2009)
13. Kay, S.M.: Fundamentals of Statistical Signal Processing, Volume 2: Detection Theory, Series. Prentice Hall Signal Processing Series, AV Oppenheim, Ed. Prentice Hall PTR (1998)

Wireless Technologies in IoT: Research Challenges

Sumit Singh Dhanda, Brahmjit Singh and Poonam Jindal

Abstract Internet of Things (IoT) is a technology to create smart world. To enable an IoT system, communication plays an important role. Everything in IoT depends on the flow of information. Many wireless technologies are available for the purpose but the choice for the right technology depends on the use case requirements. Reliability and availability of an IoT application is the outcome of reliable communication. Issues like scalability and heterogeneity present numerous challenges to the researchers. Mobility maintenance, packet delay, signal load, etc., are challenges that make reliable communication difficult and affect the quality of service. This paper presents the details of wireless technologies available for the IoT and discusses the open challenges and research issues in IoT.

Keywords Bluetooth · IoT · LTE-A · LoRaWAN · Research issues · Sigfox
Wi-Fi · WiMAX

1 Introduction

Way ahead of his time Nikola Tesla had suggested that one day wireless technology in its full bloom will convert the earth into one huge brain [1]. Today, world has started realizing this vision with the help of looking at a future nearby.

In 1999, Kevin Aston presented the idea for a smart world, where things can be connected to internet using Radio-Frequency Identification (RFID) and used the word Internet of Things (IoT) for the first time. Today, IoT has matured to complex systems that provide solutions to numerous problems on the go. The main aim of an

S. S. Dhanda (✉) · B. Singh · P. Jindal
National Institute of Technology, Kurukshetra, Kurukshetra, Haryana, India
e-mail: dhandasumit@gmail.com

B. Singh
e-mail: brahmjit.s@gmail.com

P. Jindal
e-mail: poonamjindal81@nitkkr.ac.in

IoT system is to provide intelligent decision-making for the optimum utilization of resources [2, 3]. In 2020, there will be 50 billion devices that will be connected to Internet. It was not possible to connect so many devices with IPv4 but with IPv6, it became possible to connect every object on the earth [4, 5]. Internet and Web are interconnected, where the former is responsible for the reliable, secure, and fast transmission of the information in network via switches and routers. Web operates on its top, providing an interface to convert the flowing information in usable form [6]. Web has undergone four stages of evolution until now. In the first phase, Web was primarily used by academia for research. In the second phase, focus was to provide the information of products and services to customers by companies. Third phase of evolution was marked by the sale and purchase of the products and services over the Web, the dot-com boom. The fourth stage is of "Social or experience web", the present time. While IoT marks the first real evolution of the Internet. But Internet is going through its first evolution via Internet of things (IoT).

Information is the key to efficient decision-making. It is collected with the help of sensors and communicated with the help of wireless technology to cloud. There are many technologies available to connect the entities of an IoT system with each other. Many reviews are available that provide a detailed analysis of the IoT systems but to the best of our knowledge, no review is available that discusses the wireless technologies and challenges associated with them in detail. It has been the motivation behind this work.

This paper has been divided into five sections. Elements of an IoT system are discussed in Sect. 2. In Sect. 3, various wireless technologies have been described in brief. In Sect. 4, challenges and research issues associated with them are discussed. Finally, conclusion is drawn in Sect. 5.

2 IoT Elements

To understand the functionality of IoT, one must understand its building blocks. IoT can be divided into six different elements: Identification, Sensing, Communication Computation, Services, and Semantics [7].

2.1 Identification

First and the most basic requirement for creating an IoT system is to provide the constituent objects/devices with a unique identification. Identification process can be divided into two steps namely, object ID and object address. Electronic Product Codes (EPC) and Ubiquitous Codes (u-Code) are used for the identification [8, 9]. IPv6 and IPv4 can be used for the addressing in IoT. IPv6 over Low-power Wireless Personal Area Networks (6LoWPAN) [4, 5] is a variant of IPv6 addressing, which is suitable for low-power wireless networks.

2.2 Sensing

Data is collected from the sensors/devices. It is passed on to the cloud to extract information for decision-making. Till date, wireless sensor networks, RFIDs were responsible for this part. But now, a new paradigm has come to the fore named as Crowd sourcing or Crowd sensing. In this, smartphone users willingly participate in sensing [10].

2.3 Communication

In an IoT system, various entities and domains are present. These entities and domains communicate with each other on a continuous basis. Communication links are needed between these entities and domains [11–13]. Various wired and wireless technologies will come into play for establishing these links. In wired technologies, communication is comparatively reliable and secure but they do not provide the advantage of the mobility. Wireless, on the other hand, is the natural and only choice for the mobile scenarios. Wireless technologies such as Wi-Fi, Bluetooth, IEEE 802.15.4, LTE-Advanced, and in the coming future 5G can be used for establishment of communication link. Based upon the use case scenario, one can divide the Wireless technologies into two parts long-range wireless connectivity and short-range wireless connectivity.

The long-range wireless connectivity is provided using LTE, LTE-Advanced, LTE-M, NB-IoT, WiMAX and in coming future 5G. The short-range wireless connectivity is provided by Bluetooth and its variants, Zigbee, UWB, RFID with EPC global, Wi-Fi. Selection of a technology depends on the matching of its characteristics with the use case requirements.

2.4 Computation

Processing of sensed data; extraction of knowledge from the received information; utilizing the knowledge for intelligent decision-making, all these require computational capabilities. IoT services are completely an outcome of computational abilities. At sensor level, the computation is required to convert the data into usable form. At cloud, to extract the knowledge from the received information, techniques of Big Data are required. Information extraction from the raw data is a challenging task. It will require a mechanism to find, extract, and characterized meaningful abstractions from the raw data. Data analytic techniques are required for intelligent decision-making [14].

2.5 Semantics

Data in raw form is not useful. For the right and efficient decision-making, it should be converted into knowledge. Recognition and analysis of data, modeling of information, discovery of resources, and their utilization comes under knowledge extraction. This ability to extract knowledge to provide the IoT services is referred to as Semantics [15]. Allocation of resource to the appropriate demand is done with the help of Semantic. It acts as brain of the IoT system. Extraction of knowledge and resource allocation are helped by Resource Description Framework (RDF), another option that can be utilized is Web Ontology Language (OWL). Efficient XML Interchange (EXI) format was adopted as a recommendation for web ontology by the World Wide Web Consortium (W3C) in 2011 [16]. Semantics are so important that their interoperability is required for the interoperability of the various IoT platforms. The key role in the interoperability would be played by W3C SSN ontology [17].

2.6 Services

Last and the most important element of an IoT system is services, that are provided to its users. Normally, these are also termed as use cases or applications. There are many use cases such as smart grids, smart buildings, smart cities, e-Health, smart vehicles, intelligent transportation systems, Industrie 4.0, smart agriculture, smart home, environmental monitoring, logistics and tracking, monitoring and security, e-Retail, etc. These are few applications of IoT systems. Other applications, in one way or another, are derivatives of these. Every application does a different type of work. Based on these, services can be categorized into four types [18, 19]:

(a)   Identity-related Services,
(b)   Information Aggregation Services,
(c)   Collaborative-Aware Services and
(d)   Ubiquitous Services.

3   Wireless Technologies for IoT

Communicating the data from sensors to the cloud or to any other computing facility is of utmost importance. Sensor data form the basis for the intelligent decision-making in IoT system. It puts immense responsibility on the protocols and technology to reliably transmit the sensed data to destination. IoT system is characterized by enormous uplink data and comparatively lesser data in downlink on sensor to cloud link. On the other hand, on user side in the IoT system, the requirement of downlink data is greater compared to uplink data. A developer must compare system requirements with the capabilities of the wireless technology. One shall be selected if it meets the

criteria. Technologies that can be used in IoT can be put into three categories first Cellular Technology (CT), second is Low-Power Wide Area Networks (LPWAN) (>200 m) and third and final are Short-Range Standards (SRS) for networking (up to 200 m).

In this section, different technologies that can be used for the creating IoT will be discussed.

3.1 Ultra-Wideband

UWB is a viable candidate for short-range communications for IoT. It can achieve high data rate in dense multipath environments, characterized by low signal-to-noise ratio (SNR). This becomes possible as UWB spreads its signal in very high bandwidth. UWB, due to simpler architecture and low transmission power, consumes very low power for its operations. It helps UWB communication devices to achieve longer battery life. Orthogonal codes used by various users helps in avoiding interference with the other surrounding communication and those operating in the same bandwidth with UWB communication. It has been used for designing RFIDs [20].

3.2 802.11—Wireless LAN

IEEE 802.11 has different standards like a, b, g, n, ac, and ad, operating from 2.4 to 60 GHz bands. Its communication range varies from 20 m (indoor) to 100 m (outdoor) while data rate ranges from 1 Mb/s to 6.75 Gb/s. A new standard of Wi-Fi for the IoT application was made available named as 802.11 ah. It works in sub-1 GHz license-exempt band. It provides 5 bandwidth ranges to support the IoT use cases from 1 to 16 MHz with maximum transmit power of 250 mW. The data rates vary from 150 kbps to 347 Mbps to support large-scale IoT [21–23].

3.3 WiMAX

IEEE 802.16 is a wireless broadband standard also known as Worldwide Interoperability for Microwave Access (WiMAX). It provides data rates from 8 to 80 Mb/s. Various standards of WiMAX provide for fixed access, nomadic access, and limited mobility support. It provisions high capacity for the stationary users as compared to mobile ones. In the updated version of WiMAX, i.e., (802.16 m), for the mobile and fixed stations the supported data rates are 100 Mbps and 1 Gbps, respectively. It can be verified from the working group's website. There is very limited literature that has utilized the WiMAX for IoT [24].

3.4 LR-WPAN

IEEE 802.15.4 defines standard for the Low-Rate Wireless Personal Area Networks (LR-WPAN). Higher level protocols such as Zigbee are based on its specifications. It is named low rate as its data rate varies from 40 to 250 kbps. It provides a low-cost communication solution for small sized devices. While operating at 868/915 MHz, it supports low data rates, on the other hand high data rate operations are carried out at 2.4 GHz frequencies [25, 26].

3.5 LTE, LTE-A, and Latest Versions

LTE and LTE-A has been developed for the main purpose of providing broadband capabilities and high capacity to mobile users. IoT systems require very high capacity in the uplink(UL,) but LTE and LTE-A provide high capacity for downlink(DL). 3GPP in its release 12 has come up with the 2 new versions of the LTE, namely, LTE-M and Narrow-Band IoT. Both these technologies are designed for specific need of IoT [26, 27].

3.6 Bluetooth

Bluetooth is based on the IEEE 802.15.1 standard. It is a short-range solution (8–10 m) for data exchange, with data rates of 1–24 Mbps (for various versions). One of its variants is named Bluetooth Low Energy (BLE/Bluetooth Smart), supports ultra-low-power operations. It was merged with its version v4.0 in 2010. Bluetooth is an important technology that is being utilized to create the IoT. It can help create many systems that can utilize the smartphone as sensor. Apple's iBeacon and Google's Eddystone are the examples [28–30].

3.7 LoRaWAN/LoRa

This is a Low-Power Wide Area Networks (LPWAN) standard to enable IoT. LoRaWAN (Long-Range Wide Area Network) is an open communication standard that also defines system architecture for the Network [31]. It works with the LoRa technology, which stands for Long Range, that defines the physical layer. It was acquired by SemTech in 2012. It utilizes ISM Bands of 433/867/915 MHz for operations, which are capable of penetrating deeply into structures. LoRa works on chirp spread spectrum modulation, which helps in increasing the range of communication while it maintains low-power characteristics. Gateways chipsets are capable of

demodulating on multiple channels and multiple data rates at once. LoRaWAN works on Star Topology, where the end device is connected to the gateways. LoRaWAN data rates range from 0.3 to 50 kb/s. LoRa operates in 868 and 900 MHz ISM bands. Battery life for the attached node is normally very long, up to 10 years. This is an effective technology but there are also some issues related with it that must be kept in mind while developing products for IoT scenarios [31–33].

3.8   Sigfox

Sigfox is a proprietary technology by Ingenu, which was earlier known as on Ramp Wireless. It transmits in the ISM band of 868 MHz. It utilizes the frequency as well as time diversity as it broadcasts a message thrice in three different frequency channels. It has ability to be demodulated at very small values such as –142 dBm. It has the functionality that its base station scans every channel for message and retrieves it. The modulation scheme used is BPSK. It suits a network where long range is required along with small amount of data with large number of devices. It provides the option of transmitting 12 bytes' payload in uplink and 8 bytes' payload in downlink with a protocol overhead of 26 bytes. Downlink mode works only as an acknowledgement-receiving window. Sigfox has four transmission configurations where transmission of per day messages is fixed from 1 to 140 messages per day. Benson et al. [34] have developed SCALE a low cost, smart, and safe home solution to all residents of a locality with the help of novel networking techniques, they have used Sigfox as one of the technologies.

4   Challenges and Research Issues for IoT

There is a plethora of open research challenges that are still required to be addressed.

Scalability is the main requirement of any IoT system. It poses an important challenge, as number of connected nodes in the system keeps on changing, the amount of data generated varies. It causes data variability issues.

Heterogeneity of the devices and technologies is the cause of many challenges. It creates the issue of veracity of the data. Every technology has its own data rates. This variation in data rates causes the variability in data speed. It also creates the issue of variability in the type of data, as different technologies have different data representation formats. Another challenge that arises from heterogeneity is reconfigurability of the hardware or protocols. Researchers need to design such sensors, protocols, and platforms that must be reconfigurable, as per application and scenario. Reconfigurability is the solution to compatibility issues as well. One such solution is presented in [35], in the form of a reconfigurable RF-ID tag and that is utilized as a generic sensing gateway.

Interoperability is another issue that is the byproduct of heterogeneity. Semantic interoperability is one option to provide new services from the knowledge of an existing IoT system. Another option can be the design of protocols that should make the interoperability of the technology possible [36].

Reliability and availability of an IoT service is an important challenge. It depends upon the reliability and availability of the communication. Wireless channel is very dynamic and unpredictable. Any loss of communication may result in the form of huge financial losses or it may even result in the loss of life, in case of Industrie 4.0 and intelligent transportation system, respectively. This problem may occur due to two reasons, either device failure or link failure. For link failure, there may be different solutions but provision of buffering or caching, and use of dynamic spectrum access are the prominent one that can be pursued. In case of device failure, redundancy in devices can be introduced. Its calculation and placement must be done during the planning phase [37]. The same solution can also help in the mobility maintenance of the services.

QoS is another very important challenge for IoT application. Radio resource control, discontinuous reception/packet loss, end-to-end packet delay, availability of bandwidth are important factors in determining the QoS. It is one of the biggest challenges that must be overcome by the researchers. However, it should be done in a way that suits the requirement of enabler. In [38], an analytic result has been derived for the expectation of the end-to-end packet delay. Another effort is done in cellular network, in LTE radio resource control and discontinuous reception are the two factors that affect power consumption, signal load, and delay. These factors can affect the QoS of the network heavily. Moreover, numerous interactions and the session setups in IoT will create the problem of Signal load in the network. In [39], a session management methodology for power saving and signal saving has been presented. In [40], probabilistic analysis on QoS provisioning has been provided.

Improvement in *link capacity* is an issue that will have a huge impact on the service mobility as well as QoS. Important thing to note about all existing communication technologies, wireless networking standards, or cellular standards like 2G/3G or 4G, etc., is that these standards have high capacity for downlink in comparison to uplink. Using these in machine-type communication for Industrie 4.0 can cause capacity and interference issues for the human type communication and degrade the Quality of Service for the human-type communication [41].

Energy Efficiency of a system is a very important topic for the researchers; any method that helps in reduction of power consumption must be given a thought if it can be incorporated in creation of an efficient system for the future. In [42], an efficient certificate-less access control scheme has been presented, which reduces the computation cost by 62% and energy consumption by 64%. In [43], a Zigbee-based smart home control system that provides smart interference and energy control has been implemented successfully.

5 Conclusion

An IoT system is comprised of different elements. However, communication is the most important. Without reliable communication, realization of an IoT system is not possible. Many technologies are available for connectivity in an IoT system. Every IoT application has its own set of requirements. Selection of the technology is made on the characteristics like range, capacity, number of devices that can be connected, etc. Selection of the wireless technology is made, if its characteristics match the requirements of the application. Heterogeneity is an inherent characteristic of IoT. It poses the challenges like reconfigurability and interoperability to researchers apart from creating issues of variability and veracity of data. Capacity, coverage, reliability, and availability of communication, QoS provided, energy efficiency of the system, mobility maintenance are the key issues which must be addressed. IoT is about efficient decision-making for the optimum utilization of resources and for that these challenges must be addressed.

References

1. Atzori, L., Iera, A., Morabito, G.: Understanding the internet of things: definition, potentials, and societal role of a fast-evolving paradigm. AdHoc Netw. **56**, 122–140 (2017)
2. Lopez Research: An Introduction to Internet of Things: Part 1 of IoT series. Lopez Research. https://www.cisco.com/c/dam/en_us/solutions/trends/iot/introduction_to_IoT_november.pdf (2013)
3. Internet-of-Things Architecture (IoT-A), Project Deliverable D1.2—Initial Architectural Reference Model for IoT. http://www.IoT-a.eu/public/public-documents/d1.2
4. Kushalnagar, N., Montenegro, G., Schumacher, C.: IPv6 Over Low-Power Wirel. Pers. Area Netw. (6LoWPANs): Overv. Assumpt. Probl. Statement, Goals RFC 4919. https://doi.org/10.17487/rfc4919. (2007)
5. Montenegro, G., Kushalnagar, N., Hui, J., Culler, D.: Transm. IPv6 Pack. IEEE 802.15. 4 Netw. Internet Eng. Task Force (IETF), Fremont, CA, USA, Internet Propos. Std. RFC 4944. https://doi.org/10.17487/rfc4944. (2007)
6. Evans, D.: The Internet of Things How the next Evolution of Internet is changing Everything. CISCO, San Jose, CA, USA, White Paper. https://www.cisco.com/c/dam/en_us/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf (2011)
7. Al-Fuqaha, A., Guizani, M., Mohammadi, M., Aledhari, M., Ayyash, M.: Internet of things: survey on enabling technologies, protocols, and applications. IEEE Commun. Surv. Tutor. **17**(4), 2347–2376 (2015). https://doi.org/10.1109/COMST.2015.2444095
8. The EPC global Architecture Framework, EPC global Final Version 1.3, 2009. www.epcglobalinc.org
9. Welbourne, E., Battle, L., Cole, G., Gould, K., Rector, K., Raymer, S., Balazinska, M., Borriello, G.: Building the internet of things using RFID: the RFID ecosystem experience. IEEE Internet Comput. **13**(3), 48–55 (2009). https://doi.org/10.1109/MIC.2009.52
10. Jaimes, L.G., Vergara-Laurens, I.J., Raij, A.: A survey of incentive techniques for mobile crowd sensing. IEEE Internet Things J. **2**(5), 370–380 (2015). https://doi.org/10.1109/JIOT.2015.2409151
11. Introduction to Architectural Reference Model for Internet of Things. http://www.IoT-a.eu/arm

12. The industrial Internet reference architecture, version 1.7, Ind. Internet Consortium, Needham, MA, USA, Tech. Rep. IIC: PUB: G1: V1.07: PB:20150601. http://www.iiconsortium.org/IIRA.html (2015)

13. Information Technology—Internet of Things Reference Architecture (IoT RA), International Organization for Standardization, ISO Central Secretariat, Geneva, Switzerland (2015)

14. Ganz, F., Puschmann, D., Barnaghi, P., Carrez, F.: A practical evaluation of information processing and abstraction techniques for the internet of things. IEEE Internet Things J. **2**(4), 340–354 (2015)

15. Barnaghi, P., Wang, W., Henson, C., Taylor, K.: Semantics for the internet of things: early progress and back to the future. Proc. IJSWIS **8**(1), 1–21 (2012)

16. Kamiya, T., Schneider, J.: Efficient XML Interchange (EXI) Format 1.0. World Wide Web Consortium, Cambridge, MA, USA, Recommend. REC-Exi-20110310 (2011). http://www.w3.org/TR/exi/

17. Ganzha, M., Paprzyckia, M., Pawowski, W., Szmeja, P., Wasielewska, K.: Semantic interoperability in the internet of things; an overview from the INTER-IoT perspective. J. Netw. Comput. Appl. **81**, 111–124 (2017)

18. Xiaojiang, X., Jianli, W., Mingdong, L.: Services and key technologies of the internet of things. ZTE Commun. Shenzhen, China **2**, 011 (2010)

19. Gigli, M., Koo, S.: Internet of things: services and applications categorization. Adv. Internet Things **1**(2), 27–31 (2011)

20. Kshetrimayum, R.S.: An introduction to UWB communication systems. IEEE Potentials **28**(2), 9–13 (2009). https://doi.org/10.1109/MPOT.2009.931847

21. Rajandekar, A., Sikdar, B.: On the feasibility of using WiFi white spaces for opportunistic M2M communications. IEEE Wirel. Commun. Lett. **4**(6), 681–684 (2015)

22. Park, M.: IEEE 802.11 ah: Sub-1-GHz license-exempt operation for the internet of things. IEEE Commun. Mag. **53**, 145–151 (2015). https://doi.org/10.1109/MCOM.2015.7263359

23. Qutab-ud-din, M., Hazmi, A., Badihi, B., Larmo, A., Torsner, J., Valkama, M.: Performance analysis of IoT-enabling IEEE 802.11 ah technology and its RAW mechanism with non-cross slot boundary holding schemes. In: 16th International Symposium on World of Wireless Mobile and Multimedia Networks. IEEE (2015). https://doi.org/10.1109/WoWMoM.2015.7158204

24. Park, J., Jeong, S., Cho, J., Choi, H., Roh, W., Pi, Z.: Interference level control in mobile WiMAX uplink system. IEEE Mob. WiMAX Symp. IEEE, 52–56 (2009). https://doi.org/10.1109/MWS.2009.32

25. Ndih, E.D.N., Cherkaoui, S.: On enhancing technology coexistence in the IoT era: Zig-Bee and 802.11 Case. IEEE Access **4**, 1835–1844 (2016)

26. Yang, J., Song, L., Koeppe, A.: LTE field performance for IoT applications. In: 84th Vehicular Technology Conference. IEEE (2016). https://doi.org/10.1109/VTCFall.2016.7881080

27. Lauridsen, M., Kovacs, I.Z., Mogensen, P., Sorensen, M., Holst, S.: Coverage and capacity analysis of LTE-M and NB-IoT in a rural area. In: 84th Vehicular Technology Conference. IEEE (2016). https://doi.org/10.1109/VTCFall.2016.7880946

28. Akinsiku, A., Jadav, D.: BeaSmart: a beacon enabled smarter work place. In: Network Operations and Management Symposium. IEEE, pp. 1269–1272 (2016). https://doi.org/10.1109/NOMS.2016.7503001

29. Alletto, S., Cucchiara, R., Del Fiores, G., Mainetti, L., Mighalli, V., Patrono, L., Serra, G.: An indoor location—aware system for an IoT based smart Museum. IEEE Internet Things J. **3**(2), 244–253 (2016)

30. Chiang, T.H., Chuang, Y.T., Ke, C.L., Chen, L.J., Tseng, Y.C.: Calorie map: an activity intensity monitoring system based on wireless signals. In: Wireless Communications and Networking Conference. IEEE (2017). https://doi.org/10.1109/WCNC.2017.7925708

31. LoRA Alliance: LoRAWAN^TM what is it? White Paper LoRA Alliance. https://docs.wixstatic.com/ugd/eccc1a_ed71ea1cd969417493c74e4a13c55685.pdf. (2015)

32. Dongare, A., Hasling, C., Bhatia, K., Balanutta, A.: OpenChirp: a low-power wide area networking architecture. In: First International Conference on Pervasive Computing and Communications Workshop IEEE (2017). https://doi.org/10.1109/percomw.2017.7917625

33. Neumann, P., Montavont, J., Noel, T.: Indoor deployment of LPWAN: a LoRaWAN case study. In: 12th International Conference on Wireless and Mobile Computing, Networking and Communications. IEEE (2016) https://doi.org/10.1109/wimob.2016.7763213

34. Benson, K., Fracchia, C., Wang, G., Zhu, Q., Almomen, S., Cohn, J., D'Arcy, L., Hoffman, D., Makai, M., Stamatakis, J., Venkatasubramanian, N.: SCALE: safe community awareness and alerting leveraging the internet of things. IEEE Commun. Mag. **53**(12), 27–34 (2015)

35. Khan, M.S., Islam, M.S., Deng, H.: Design of a reconfigurable RFID sensing tag as a generic sensing platform towards the future Internet of Things. IEEE Internet Things J. **1**(4), 300–310 (2014)

36. Yacchirema, D.C., Palau, C.: Smart IoT gateway for heterogeneous devices interoperability. IEEE Latin Am. Trans. **14**(8), 3900–3906 (2016)

37. Macedo, D., Guedes, L.A., Silva, I.: A dependability evaluation for the internet of things incorporating redundancy aspects. In: International Conference on Networking, Sensing and Control (ICNSC). IEEE, pp. 417–422 (2014). https://doi.org/10.1109/icnsc.2014.6819662

38. Li, J., Zhao, Y.Q., Yu, F.R., Huang, X.: Queuing analysis of two-hop relay technology in LTE/LTE-A networks with unsaturated and asymmetric traffic. IEEE Internet Things J. **3**(3), 378–385 (2016)

39. Wang, X., Chang, M.J., Shih, Y.Y., Chiang, M.: Internet of things session management over LTE-balancing signal load, power and delay. IEEE Internet Things J. **3**(3), 339–353 (2016)

40. Zhang, R., Wang, M., Shen, X., Xie, L.L.: Probabilistic analysis on QoS provisioning for IoT in LTE-A heterogeneous networks with partial spectrum usage. IEEE Internet Things J. **3**(3), 354–365 (2016)

41. de Andrade, T.P., Astudillo, C.A., da Fonseca, N.L.: Allocation of control resources for machine-to machine and human-to-human communications over LTE/LTE-A networks. IEEE Internet Things J. **3**(3), 366–377 (2016)

42. Li, F., Hong, J., Omabi, A.A.: Efficient certificateless access control for industrial IoT. Future Gener. Comput. Syst. **76**, 285–292 (2017)

43. Khan, M., Silva, B.N., Han, K.: Internet of things based energy aware smart home control system. IEEE Access (2016). https://doi.org/10.1109/acess.2016.2621752

Fog Computing for Ubiquitous Transportation Applications—A Smart Parking Case Study

Md. Muzakkir Hussain, Faraz Khan, Mohammad Saad Alam and M. M. Sufyan Beg

Abstract The current transportation architectures are heavily populated with smart devices and entities due to unfolded technological evolutions in Intelligent Transportation Systems (ITS). The ITS ecosystem, when introduced to Internet of Things (IoT) makes every object active and brings them online. Such devices generate data deluge that demand scalable storage and computational resources. Though centralized cloud-based solutions significantly circumvent those demands, but the current deployments still have silos and cease to meet the analytics and computational exigencies for such dynamic ITS subsystems. In this work, we investigate the current state of cloud-based solutions for fulfilling the mission-critical store and compute requirements of IoT-aided ITS architectures and revisit the motivations for adopting edge-centered fog computing paradigms. We also proposed a fog computing topology customized to ITS architectures. Further, the viability of proposed fog framework is demonstrated through a smart parking case study. Results show a significant improvement performance in terms of probabilistic QoS guarantees for private parking land owners, at the expense of a relatively small number of reserve premium spaces.

Keywords Internet of things (IoT) · Cloud computing · Fog computing
Smart transportation applications · Smart parking

Md. Muzakkir Hussain (✉) · F. Khan · M. M. Sufyan Beg
Department of Computer Engineering, ZHCET, AMU, Aligarh, Uttar Pradesh, India
e-mail: md.muzakkir@zhcet.ac.in

M. M. Sufyan Beg
e-mail: mmsbeg@cs.berkeley.edu

M. S. Alam
Department of Electrical Engineering, ZHCET, AMU, Aligarh, Uttar Pradesh, India
e-mail: saad.alam@zhcet.ac.in

1   Introduction

Since the past decade, the traditional transportation system is under progressive overhaul by data-driven Intelligent Transportation (ITS) infrastructures [1]. However, the computation and processing tasks have been exponentially raised due to the data-driven transportation application, thanks to advanced protocols for integrating Internet of things (IoT) into contemporary Information and Communication Technologies (ICT) [1–3]. Due to development in cellular and internet technologies, the whole transportation entities are being meshed up forming Intelligent Transportation Web (ITW). The Vehicular Ad hoc Networks (VANETs), platoons, connected vehicles; Vehicular Sensor Networks (VSN), etc., are notable among them [4]. The data-driven ITS services emerged to be efficient way of improving the performance of transportation systems, enhancing commuter's safety and security, improving both driver's and user's experiences by improving QoS and QoE [1]. Consequently, there is a significant change in the data generation and consumption landscapes, much more data are collected from multitudes of ITS sources and can be processed into various forms for multiple stakeholders. Thus, there is urgent need for architectures equipped with store, process and compute resources, that can act as supporting analytics framework for ITS applications [5].

Cloud computing due to its elastic virtualized storage capabilities, proves to be promising technology for fulfilling the storage and processing demands [6]. Also, with the advent of Internet of Things (IoT), the amount of data that must be collected through sensors and transportation telematics, sent through the communication networks, and processed in the cloud servers have grown tremendously at a very rapid rate. The Cloud-centric IoT are facing the problem of not only communication latencies, but also inefficient data processing and applications that were supposed to be smart are not meeting the expectations [5, 7]. For example, smart parking and smart traffic have yet to overcome not only infrastructure issues, but also security and performance issues related to data collection and processing. To address such range issues, fog/edge computing comes into play. Fog computing focuses on the collaboration of end-user devices for data storage, communication, control, configuration, and management such that computations that were carried out formerly in the clouds are now partially offloaded to the network edges (local nodes) such that the latencies between the end users and the cloud can be reduced significantly and the results of data processing obtained more efficiently [8]. Through such a computing service at the edge of networks, better quality-of-service (QoS) and data analytics can be achieved. In the age of wireless communications and mobile computing, fog computing will play a major role in terms of lower power usage (such as the latest Narrow-Band IoT), embedded intelligence (such as in cyberphysical systems), and software-defined functions (such as in Software-Defined Networking).

We propose a fog computing model customized to smart transportation applications, where the vehicles in the parked state, forms ad hoc fogs. The notion is to employ the vehicles as part of computational infrastructures by harnessing the store and compute resources latent as underutilized vehicular resources. Further, to

demonstrate the viability of the proposed framework, a smart parking management case study is performed. Before going to the proposed fog architecture, we in the next section analyze the computing needs of mission-critical smart transportation applications and assess the states of generic cloud models. Correspondingly, we also highlight how a paradigm shift from generic cloud-based centralized computation into geo-distributed fog computing model can turn to be a near ideal solution to carry up the mission-critical smart transportation applications.

2 Mission-Critical Computing Requirements of Smart Transportation Applications

Consider a typical traffic lighting use case where smart traffic lights will be able to adapt themselves to the real-time traffic circumstances within a particular region. In this case, the reaction time for one or several smart traffic lights is too short that it is virtually impossible to traffic all the application executions to a distant Cloud. Therefore, such traffic lights should be programmed in a way that they autonomously cooperate with each other and with all the locally available computing resources such as Roadside Units (RSU) to coordinate their operations. Other such examples may be vehicular search applications [9], vehicular crowd sourcing [10], smart parking, etc. [11, 12]. From such examples, it is perceived that there is need of computing frameworks that will provide ubiquitous and real-time analytics services for varying transportation domains. Here, we highlight some key data collection, processing, and disseminating requirements of smart transportation infrastructures.

A. **Modularity**: An intelligent transportation network is usually large and complex as it involves heterogeneous IoT and non-IoT devices with numerous data types demanding a wide set of processing algorithms. Thus, the software platform supporting the ITS applications should have characteristic modularity and flexibility support. The applications must be incrementally deployed in a way that the system should be self-evolving and fault tolerant, i.e., partial failures do not affect the whole system dynamics. Modularity also ensures different data processing algorithms to be designed and plugged into the system with minimal effort. This is important due to the diverse range of data streams generated in Smart Transportation infrastructures. Thus the application development process can be done in two independent stages, developing individual modules and developing module interconnection logics. The earlier can be done by component or module providers while the later can be done by Smart Transportation developers. The cloud platforms provides enough modularity and flexibility support for deploying ITS applications but the centralized execution strategies often lead to poor Quality of Experience (QoE) for the stakeholders.

B. **Scalability**: An ideal ITS architecture should be distributed and scalable enough to efficiently serve a large vehicle population. Though cloud provides scalable resource pools, due to the huge volume of real-time data generated by the ITS

environment, it might not able to sustain with the Smart Transportation applications' requirement with respect to the low latency requirement. Current cloud-based ITS applications often "embrace inconsistency", thus implementing consistency preserving computational structures constitute a promising investment domain for the research and development sector. The trend envisions a more flexible infrastructure, as in fog computing models where computation resources in dynamic objects such as moving vehicles can also participate in the application.

C. ***Context-Awareness and Abstraction Support***: As the ITS components such as vehicles and other infrastructures are mobile and sparsely distributed over large geography, fog computing will provide context-aware computing platforms for reliable transportation services. Further, the geo-distributed context information should be exposed to developers so that they can build context-aware applications. Because of the high level of heterogeneity and the large number of IoT devices in a typical ITS application, viz., smart parking requires high degree of abstraction of how the heterogeneous computations and processing are described, coordinated or interact with one another. The centralized cloud-based ITS solutions needs to be upgraded to dedicated fog solutions such that the model allows it to work with a pool of vehicles at once. For instance, such a programming abstraction should be able to describe the command like: "get the State of Charge (SoC) of these groups of cars in this location."

D. ***Decentralization***: Since the ITS applications usually operate over a large number of heterogeneous and dynamic transportation telematics such as mobile/autonomous vehicles or Roadside Units (RSU), decentralized execution or programming model is necessary. The centralized cloud-based application has to implement all sorts of conditions and exception handling to deal with such heterogeneity and dynamic nature. The fog platform will ensure scalable execution if the application can be developed in a modular way with components being distributed to the edge devices. Instead of relying on a remote cloud data centers, fog computing provides robust decentralization support to leverage the computing resources of the ITS components such vehicles, sensor, etc., to execute the application, in order to fulfill the latency requirement of the ITS applications.

E. ***Energy Consumption of Cloud Data Centers***: The energy consumption in mega data centers is likely to get tripled in the coming decade [13], thus adopting energy-aware strategies becomes an earnest need for computational folks. Offloading the whole universe of transportation applications into the cloud data centers causes untenable energy demands, a challenge that can only be alleviated by adopting sensible energy management strategies. Also, there are a plenty of ITS applications without significant energy implications and instead of overloading data centers with such trivial tasks, the analytics can be made ready at and within ITS fog nodes such as vehicular platoons, parked vehicular networks, RTUs, SCADA systems, Roadside Units (RSU), base stations, and network gateways.

Motivated by the above-mentioned mission-critical computing requirements of IoT-aware smart transportation applications, the downsides of current cloud computing infrastructures to meet those needs, and having the assumption that the transportation design community is not in a position to reinvent a dedicated Internet infrastructure or to develop computing platforms and elements from scratch that fulfill all those requirements, we in this work present a fog computing framework whose principle underlie on offloading the time and resource critical operations **F**rom c**O**re to ed**G**e. The argument here is not to cannibalize the existing centralized cloud support for SG, but to comprehend the applicability of fog computing algorithms to interplay with the core centered cloud computing support leveraged with a new breed of real-time and latency-free utilities. The objective is also to develop a viable computational prototype for an ITS architecture in the realm of IoT space, through proper orchestration and assignment of compute and storage resources to the endpoints and where the cloud and fog technologies tuned to interplay and assist one other in a synergistic way.

3 Fog Computing Model for Smart Transportation Applications

Figure 1 depicts a typical fog-assisted cloud architecture customized for smart transportation applications. It is a consensus that fog paradigm is not envisioned to cannibalize or replace the cloud computing platforms rather the notion is to realize fog platforms as perfect ally [11], or an extension or cooperative modules having an interplay with the cloud infrastructure [12]. In fact, as according to [4], properties like elasticity, distributed computation, etc., are defined commonly for both cloud as well as fog. However, since the computation-intensive tasks from resource-constrained entities such as sensor nodes are mapped to Computational Resource Blocks (CRBs) of dedicated fog nodes, the response time is appreciably reduced. The distinguishing geo-distributed intelligence provided by fog deployments makes it more viable for security constrained services as the critical and sensitive is selectively processed on local fog nodes and is kept within the user control instead of offloading to the vendor regulated mega data centers. The fog service models also improve the energy efficacy by offloading the power intensive computations to battery-saving modes [12]. Additional fog nodes can be dynamically plugged-in when and wherever necessary thereby removing the scalability issues that hinders the success of cloud computing models. The bandwidth issues are dramatically fixed as raw application requests are filtered, processed, analyzed, and cached in local computing nodes, thus reducing the data traffic across the cloud gateways. If robust and predictive caching algorithm is employed, the fog nodes would serve a significant portion of consumer requests from the local nodes only, thus liberating the reliance on data center connectivity [14]. The fog nodes can be efficiently programmed to incorporate context and situational awareness about the data thereby improving the dependability of the system.
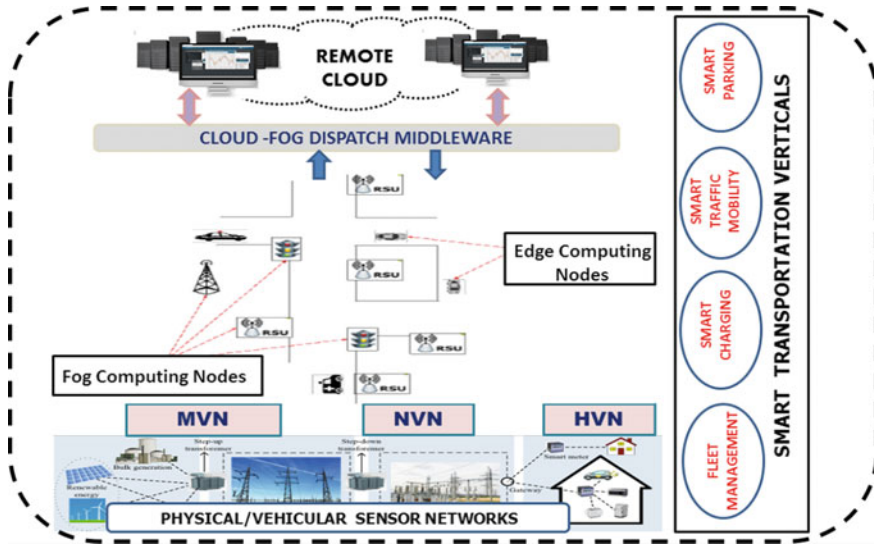
Fig. 1 Topology of FOG computing paradigm for smart transportation architectures
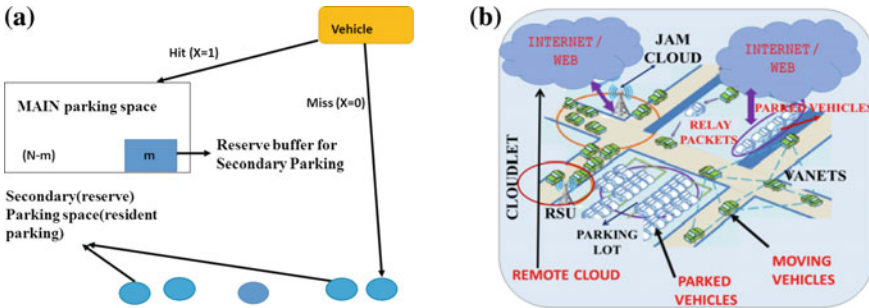
Fig. 2 **a** Resource pooling at the proposed parking scenario. **b** Fog computing at parked vehicular network

4 Fog-Based Smart Parking—A Case Study

In order to substantiate the claim regarding superiority of fog paradigms over generic cloud-based transportation applications, in this section we explore a smart parking model, where the vehicle fleet is managed through real-time fog-based distributed analytics. The vehicles leveraged with sense and actuate IoT devices are tracked in real time by the centralized cloud while the computations at deep in the network are performed through fogs. Figure 2b shows an instance of fog deployment at parked vehicular infrastructures.

The proposed smart parking prototype can be used in the existing suburb cities for better management of parking resources. We consider two parking spaces—MAIN

and Secondary, on the basis of daytime and night time users. During daytime there is overcrowding at one parking space while it may be surplus at other. The vehicular sensors and IoT devices deployed at vehicular transportation hubs, such as academic campuses, malls, hospital, etc., captures the real-time dynamics of fleet characteristics. Consider such a college building as a MAIN space and the nearby residential area as a secondary space. During day (business) hours, there is congestion at MAIN space and surplus at the secondary space (residential area). Most of the people from secondary space leave there houses for work during business hour, leaving spaces for academic stakeholder to utilize that space. Since those secondary spaces are not owned by the MAIN space authority but rather rented by the secondary space owner or landlords. To achieve this model, we should guarantee the QoS and QoE to the landlord of secondary space by reserving parking space in MAIN space for them. The sensing and analytics at the dedicated fog nodes must ensure that users of MAIN space will get equal opportunity to park their vehicle in MAIN space with complete utilization of the parking lots.

Consider a MAIN space with N number of parking spaces surrounded by Q secondary spaces (example residential buildings) as shown in Fig. 2a. On a normal working days when MAIN space capacity is unable to meet the demand of its users, the vehicles should be offloaded somewhere else for parking. At the same time, many nearby residents leave their home for work vacating their Q parking spaces. Our prime motive is how to club these independent parking entities together in a way that guarantees the QoS and QoE to the landlords of secondary space and ad hoc users of MAIN space. We assume a contract between landlord of the Q secondary space and the MAIN space authority such that within a fixed interval of daytime [0, W] the MAIN space users use the secondary space, outside this interval landlord take the possession of the secondary space. To guarantee the QoS to the landlords of secondary space, we reserve *m* spaces in the MAIN space for the event when landlord return home within the contracted daytime interval [0, W] or MAIN space user does not vacate the secondary space within the contracted daytime interval [0, W]. To guarantee the QoS to the MAIN space users, we have to optimally allocate the MAIN space and secondary space to each user. To ensure a guaranteed QoS for the landlords of secondary space, we consider two scenarios where:

A. The MAIN's user does not vacate the secondary space before the contracted daytime interval [0, W], call as *Overstay event*.
B. Accommodate the needs of landlords in a situation when a landlord returns home during/before the contracted time, call it *Home Early event*.

A. ***OVERSTAY EVENT*** (***Oi***)

Overstay event represents the situation when a MAIN space user using the secondary space for parking his/her vehicle, does not vacant the space within the contracted daytime interval [0, W]. To confront such circumstances, we select the rental window of optimal size. The overstay event will be defined as

$$O_i = \{W < T_i\} \cap \{T_i < A_i\}$$

where

T_i = random variable (i.i.d) which donates the time when ith landlord return home.

A_i = random variable (i.i.d) which donates the departure time of ith daytime user.

W = rental window duration.

Probability of overstay event will be

$$P\,(O_i)\;=\;P\,(A_i\,>\,W)$$
$$P\,(O_i)\;=\;P\,(W\,<\,T_i\,<\,A_i) \qquad (1)$$

Let the probability distribution of A_1 denote by K_A and let the probability distribution of T_1 will be denote by K_T. From Eq. 1, we have

$$P\,(O_i)\;=\;\int_{z=W}^{\infty} P(W\,<\,T_1\,<\,A_1|A_1=z)dK(z)$$
$$P\,(O_i)\;=\;\int_{z=W}^{\infty} P(W\,<\,T_1\,<\,z)\,dK(z)$$

Thus

$$P\,(O_i)=\int_{z=W}^{\infty} (K_T(z)-K_A(W))\,dK(z) \qquad (2)$$

Equation (2) shows that the overstay event probability is dependent on rental window. If the value of W is greater than or equal to A_i than the probability of the overstay event should be zero. To achieve the optimum rental window size, we have to exploit the previous parking data of the vehicle. The fog analytics will use the dynamic parking data to decide the rental window size such that probability of overstay event is minimum.

B. *HOME EARLY EVENT* (*HE_i*)

Home early event is the situation when the landlord wants his/her secondary space during the contracted daytime interval [0, W], the reasons being returning home early or does not leave his/her house during contracted daytime interval [0, W].

$$HE_i\;=\;\{T_i\in[0,\,W]\}\cap\{T_i\,<\,A_i\}$$

T_i = Random variable which donates the time when landlord returns home.

A_i = Random variable which donates the departure time of daytime user.

W = Rental window duration. For calculating the probability of home early event HE_1, we need a probability distribution of A_1 and probability distribution of T_1. Let the probability distribution of A_1 is denoted by H_A and let the probability distribution of T_1 will be denoted by H_T.

$$
\begin{aligned}
P\,(HE_1) &= P\,(\{T_1 \in [0,\,W]\} \cap \{T_1 < A_1\}) \\
&= P\,(\{T_1 < W\} \cap \{T_1 < A_1\})\,(by\ T_1 > 0) \\
&= P\,(\{T_1 < W\} \cap \{T_1 < A_1\}|\,A_1 \leq W)\,P\,(A_1 \leq W) \\
&\quad + P\,(\{T_1 < W\} \cap \{T_1 < A_1\}|\,A_1 > W)\,P\,(A_1 > W) \\
&= P\,(T_1 < A_1|\,A_1 \leq W)\,P\,(A_1 \leq W) + P\,(T_1 < W|\,A_1 > W)\,P\,(A_1 > W) \\
&= P\,(T_1 < A_1 \leq W) + P\,(T_1 < W)\,P\,(A_1 > W) \\
&= \int_{h=0}^{W} (P\,(h < A_1 \leq W))\,dH_T(h) + P\,(T_1 < W)\,P\,(A_1 > W)
\end{aligned}
$$

$$
P\,(HE_1) = \int_{h=0}^{W} (H_A(W) - H_A(h))\,dH_T(h) + H_T(W)\,(1 - H_A(h)) \tag{3}
$$

If data is not available to estimate the distribution of H_T and H_A, it is considered that the assumption that H_T and H_A are normally distributed and with the help of normal distribution we can calculate the probability of home early event from Eq. 3. There may be several reasons for early return of the landlord or not leave home during contracted daytime interval [0, W] such as sickness of landlord. To guarantee the QoS for the landlord in such case of emergency requirement of their secondary space, m number of spaces has been reserved in the MAIN space for landlords known as reserve buffer. The objective is to optimize **m**, i.e., how much space is sufficient to fulfill the need of landlord during Home Early event.

Let us derive a formula to calculate the optimum number of reserve parking space **m** from the available N MAIN spaces. Referring to Fig. 2a. (system architecture), we reserve m spaces from N MAIN spaces as a reserve buffer to meet the need of landlord during day hours. Consider the probability P (Q, m) of the episode that designate that more than m spaces are required to fullfill the secondary space owner's need for the reserve spaces in MAIN space during contracted daytime interval [0, W].

$$
P\,(Q,\,m) = P\sum_{i=1}^{Q} 1_{HEi} > m \tag{4}
$$

where 1_{HEi} is Bernoulli random variable taking value 1 when home early event occur and value 0 if not. The value of Bernoulli random variable 1_{HEi} will be 1 only when home early event occurs. That is mean probability P(Q, m) is entirely dependent on home early events probability. P(Q,m) can be characterized by P(HE_i).

Let $x = P\,(HE_i)$

$$
P\,(Q,\,m) = \sum_{k=m}^{Q} \binom{Q}{k} x^k (1-x)^{Q-k} \tag{5}
$$

Our aim is to find the value of **m** for which at given probability of home early event the value of P(Q, m) is zero. Now, we estimate the frequency of landlords to return home early or not leaving home at all. There are many reasons for landlords to return home early or not leave home. We estimate this frequency (P(HE_i)) of such
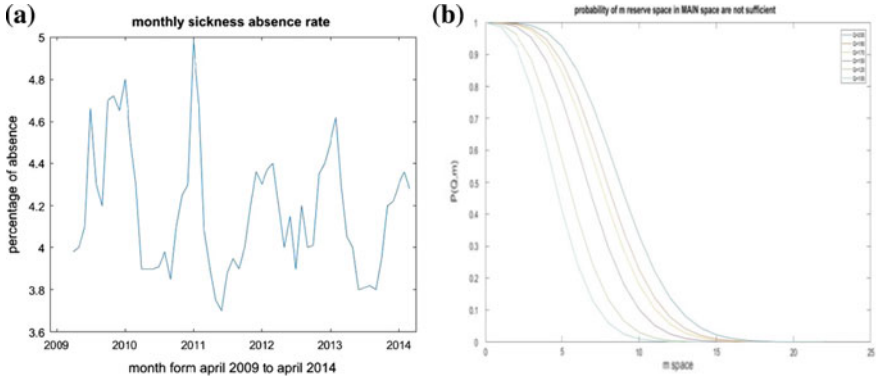
Fig. 3 **a** Monthly sickness absence rate [12]. **b** Probability that m reserve spaces are not sufficient for Q contracted spaces

days with the number equals to sick days from the available data of monthly sickness absence rate by [12] as in Fig. 3a. Figure 3b shows the percentage of the absence of staffs due to sickness that is mean value that they stayed at their home or return home early due to sickness on that particular day. Considering $T_i = 0$ for landlord remaining at home due to sickness and $T_i = W$ for rest, and for the fixed value of P (HE_i), we can find the value of m for which the probability $P(Q, m)$ will be zero. That value of m (reserve buffer in MAIN space) will be sufficient for the landlords in case home early events. In Fig. 3b, for example, all curves are generated for fixed value of P (HE_i) = 0.041828 and different values of Q (contracted secondary space).

Let assume a college building as a MAIN space and nearby residential apartment as secondary space. The objective of this algorithm is to provide fair chances to the users of MAIN space to park their vehicles in the MAIN space and make sure all the space is utilized that is not a single slot is left unallocated. Let us assume a population of 500 competing for 300 MAIN space slots. Each day user users are assigned the parking space in either MAIN space or Secondary space according to their past utilization. Figure 4 shows the convergence of the different user's average utilization of the MAIN space to the optimum point for the iteration of 700 days. As we see in Fig. 4, when the number of days increases all the user's average utilizations of MAIN space tends to converge at the optimum value which is the maximum utilization of the MAIN space by any user or we can say every member of MAIN space gets nearly equal opportunity to park their vehicle in MAIN space. That is, in the long haul, average access to the MAIN space of all users is approximately equal.
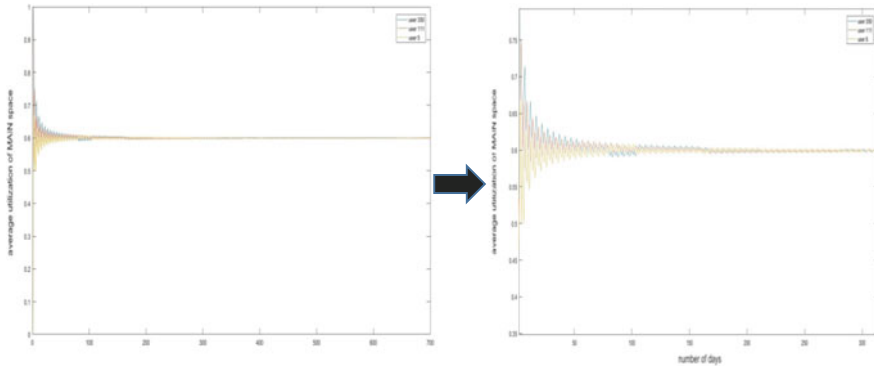
Fig. 4 Average allocation of the MAIN space to the users

5   Conclusion

In this paper, an edge-centered FOG computing model is proposed for ITS infrastructures. The work first investigates the inadequacies of pure cloud-based ITS solutions to meet the processing and computational requirements and proposes a fog computing framework that ensures proper distribution and offloading of computations across the cores (cloud) as well as edges. For assessing the applicability of the proposed framework, a smart parking use case is explored. An optimization framework is presented that improves the Quality of Service (QoS) and quality of Experience (QoE) of associated stakeholders. A significant performance improvement in terms of probabilistic QoS guarantees for private parking land owners has been achieved, incurring a relatively lesser and optimal number of reserve premium spaces.

References

1. Zhang, J., Wang, F., Wang, K., Lin, W., Xu, X., Chen, C.: Data-driven intelligent transportation systems: a survey. IEEE Trans. Intell. Transp. Syst. **12**(4), 1624–1639 (2011)
2. Perera, C., Qin, Y., Estrella, J.C., Reiff-marganiec, S., Vasilakos, A.V.: Fog Computing for Sustainable Smart Cities: A Survey. ACM Computing Surveys (2017) (Yet to be Published)
3. Pham, T.N.A.M., Tsai, M., Nguyen, D.U.C.B.: A cloud-based smart-parking system based on internet-of-things technologies. IEEE Access, 1581–1591 (2015)
4. Tubaishat, M., Zhuang, P., Qi, Q., Shang, Y.: Wireless sensor networks in intelligent transportation systems. Wirel. Commun. Mob. Comput. 287–302 (2009)
5. Hou, X., et al.: Vehicular fog computing: a viewpoint of vehicles as the infrastructures. IEEE Trans. Vehi. Technol. **65**(6), 3860–3873 (2016)
6. Yigit, M., Gungor, V.C., Baktir, S.: Cloud computing for smart grid applications. Comput. Netw. (2014)
7. Tang, B., et al.: Incorporating intelligence in fog computing for big data analysis in smart cities. IEEE Trans. Ind. Inf. **3203**(c), 1–11 (2017)
8. Paper, I.: Fog and IoT : an overview of. IEEE Internet Things J. **3**(6), 854–864 (2016)

9. Giang, N.K., Leung, V.C.M., Lea, R.: On developing smart transportation applications in fog computing paradigm. In: ACM, DIVANet'16, Malta, Malta, 13–17 Nov 2016, pp. 91–98
10. Wu, D., Zhang, Y., Bao, L., Regan, A.C.: Location-based crowd sourcing for vehicular communication in hybrid networks. IEEE Trans. Intell. Transp. Syst. **14**(2), 837–846 (2013)
11. Pham, T.N., Tsai, M.-F., Nguyen, D.B., Dow, C.-R., Deng, D.-J.: A cloud-based smart-parking system based on internet-of-things technologies. IEEE Access **3**, 1581–1591 (2015)
12. Kotb, A.O., Shen, Y.C., Zhu, X., Huang, Y.: IParker-a new smart car-parking system based on dynamic resource allocation and pricing. IEEE Trans. Intell. Transp. Syst. **17**(9), 2637–2647 (2016)
13. Yu, X., Xue, Y.: Smart grids : a cyber—physical systems perspective, 1–13 (2016)
14. He, W., Yan, G., Da Xu, L., Member, S.: Developing vehicular data cloud services in the iot environment. IEEE Trans. Ind. Inform. **10**(2), 1587–1595 (2014)

Vibration of Nonhomogeneous Square Plate with Circular Variation in Density

Amit Sharma

Abstract Here, an analysis of vibrational frequency of nonuniform and nonhomogeneous square plate is presented. In view of nonuniformity and nonhomogeneity in plates material, the author considered linear variation in thickness and circular variation in density. The variation of temperature is assumed bilinear (linear along both axes). To analyze the behavior of frequency modes under different variations in parameters, Rayleigh–Ritz technique is used.

1 Introduction

The failure of structures and noise are caused because of excessive dynamic behavior. To increase the safety and reliability of structures, it is required to have a better understanding of dynamic properties of structures. To do modal analysis of structures, it is essential to calculate natural frequencies and mode shapes. These parameters play an essential role in the designing of structures and determine the dynamic characteristics of existing structures.

Plates are very common components (structural), which have been widely applied in various applications of engineering and sciences, viz., civil engineering, aerospace engineering, ocean engineering, optical instruments, etc. Numerous studies have been carried out to analyze the characteristic of plates.

Chopra and Durvasula [1] provided frequency (natural) of vibration with mode shape of symmetric trapezoidal plate on simply supported condition. Free vibrations of rectangular (viscoelastic) tapered (linear variation in thickness) plate have been discussed by Gupta and Khanna [2]. Hutchinson and Zillmer [3] applied series solution and determined accurate natural frequency of vibration of rectangular parallelepiped having transaction-free surface. Rayleigh–Ritz method has been applied by Itakura [4] to analyze natural vibration of skewed plate with arbitrary boundary conditions. Khanna et al. [5] studied the behavior of temperature and exponential

A. Sharma (✉)
Amity University Haryana, Gurugram 122413, Haryana, India
e-mail: dba.amitsharma@gmail.com

tapering on vibration of square plate. Liu et al. [6] presented the vibration characteristic of rectangle thick plate (on arbitrary boundary) on Pasternak foundation. Sharma and Sharma [7] discussed the frequency of parallelogram plate (nonhomogeneous) with the help of mathematical model. An analysis of vibration of rectangular plate (tapered and nonhomogeneous) with thermal effect has been given by Sharma et al. [8]. Sharma et al. [9] presented the vibration (natural) of trapezoidal plate (nonhomogeneous and tapered). Sharma et al. [10] discussed natural vibration of nonuniform and nonhomogeneous square plate. Timoshenko and Woinowsky–Krieger [11] provided free vibration of plates and shells based on classical plate theory in their book. Wang et al. [12] investigated a unified solution procedure to analyze free vibration of rectangular plate (moderately thick orthotropic) with general edge condition. Zhang et al. [13] analyzed free vibration of rectangle plate (moderately thick laminated composite) with nonuniform boundary condition using improved Fourier series method. Natural vibration characteristic of rectangular thick plate resting on elastic foundation has been studied by Zhou et al. [14] using Ritz method.

In the present paper, the author calculates the first two modes of vibration for nonhomogeneous square plate with linear variation in thickness under bilinear thermal effect and presented in the form of tables. To support the evaluated results, a graphical comparison is also given with the existing results.

2 Analysis

The transverse equation of motion for plate is

$$\frac{\partial^2 M_x}{\partial x^2} + 2\frac{\partial^2 M_{xy}}{\partial x \partial y} + \frac{\partial^2 M_y}{\partial y^2} = \rho l \frac{\partial^2 \phi}{\partial t^2} \tag{1}$$

where

$$\left.\begin{array}{l} M_x = -D_1\left[\dfrac{\partial^2 \phi}{\partial x^2} + v\dfrac{\partial^2 \phi}{\partial y^2}\right] \\[3mm] M_y = -D_1\left[\dfrac{\partial^2 \phi}{\partial y^2} + v\dfrac{\partial^2 \phi}{\partial x^2}\right] \\[3mm] M_{xy} = -D_1\left(1 - v\right)\dfrac{\partial^2 \phi}{\partial x \partial y} \end{array}\right\} \tag{2}$$

Using Eqs. (2), (1) becomes

$$
\begin{aligned}
&\left[D_1 \left(\frac{\partial^4 \phi}{\partial x^4} + 2\frac{\partial^4 \phi}{\partial x^2 \partial y^2} + \frac{\partial^4 \phi}{\partial y^4} \right) + 2\frac{\partial D_1}{\partial x} \left(\frac{\partial^3 \phi}{\partial x^3} + \frac{\partial^3 \phi}{\partial x \partial y^2} \right) + 2\frac{\partial D_1}{\partial y} \right. \\
&\left(\frac{\partial^3 \phi}{\partial y^3} + \frac{\partial^3 \phi}{\partial y \partial x^2} \right) + \frac{\partial^2 D_1}{\partial x^2} \left(\frac{\partial^2 \phi}{\partial x^2} + v\frac{\partial^2 \phi}{\partial y^2} \right) + \frac{\partial^2 D_1}{\partial y^2} \left(\frac{\partial^2 \phi}{\partial y^2} + v\frac{\partial^2 \phi}{\partial x^2} \right) \\
&\left. + 2\left(1 - v\right) \frac{\partial^2 D_1}{\partial x \partial y} \frac{\partial^2 \phi}{\partial x \partial y} \right] + \rho l \frac{\partial^2 \phi}{\partial t^2} = 0
\end{aligned}
\tag{3}
$$

For solution of Eq. (3), deflection $\phi(x, y)$ can be taken as

$$
\phi(x, y, t) = \Phi(x, y) * T(t)
\tag{4}
$$

Using Eq. (4), (3) becomes

$$
\begin{aligned}
&T\left[D_1 \left(\frac{\partial^4 \Phi}{\partial x^4} + 2\frac{\partial^4 \Phi}{\partial x^2 \partial y^2} + \frac{\partial^4 \Phi}{\partial y^4} \right) + 2\frac{\partial D_1}{\partial x} \left(\frac{\partial^3 \Phi}{\partial x^3} + \frac{\partial^3 \Phi}{\partial x \partial y^2} \right) + 2\frac{\partial D_1}{\partial y} \right. \\
&\left(\frac{\partial^3 \Phi}{\partial y^3} + \frac{\partial^3 \Phi}{\partial y \partial x^2} \right) + \frac{\partial^2 D_1}{\partial x^2} \left(\frac{\partial^2 \Phi}{\partial x^2} + v\frac{\partial^2 \Phi}{\partial y^2} \right) + \frac{\partial^2 D_1}{\partial y^2} \left(\frac{\partial^2 \Phi}{\partial y^2} + v\frac{\partial^2 \Phi}{\partial x^2} \right) \\
&\left. + 2\left(1 - v\right) \frac{\partial^2 D_1}{\partial x \partial y} \frac{\partial^2 \Phi}{\partial x \partial y} \right] + \rho l \Phi \frac{\partial^2 T}{\partial t^2} = 0
\end{aligned}
\tag{5}
$$

Now, separate the variables using variable separable technique

$$
\begin{aligned}
&\left[D_1 \left(\frac{\partial^4 \Phi}{\partial x^4} + 2\frac{\partial^4 \Phi}{\partial x^2 \partial y^2} + \frac{\partial^4 \Phi}{\partial y^4} \right) + 2\frac{\partial D_1}{\partial x} \left(\frac{\partial^3 \Phi}{\partial x^3} + \frac{\partial^3 \Phi}{\partial x \partial y^2} \right) + 2\frac{\partial D_1}{\partial y} \right. \\
&\left(\frac{\partial^3 \Phi}{\partial y^3} + \frac{\partial^3 \Phi}{\partial y \partial x^2} \right) + \frac{\partial^2 D_1}{\partial x^2} \left(\frac{\partial^2 \Phi}{\partial x^2} + v\frac{\partial^2 \Phi}{\partial y^2} \right) + \frac{\partial^2 D_1}{\partial y^2} \left(\frac{\partial^2 \Phi}{\partial y^2} + v\frac{\partial^2 \Phi}{\partial x^2} \right) \\
&\left. + 2\left(1 - v\right) \frac{\partial^2 D_1}{\partial x \partial y} \frac{\partial^2 \Phi}{\partial x \partial y} \right] \Big/ \rho l \Phi = -\frac{1}{T}\frac{\partial^2 T}{\partial t^2} = \omega^2
\end{aligned}
\tag{6}
$$

where two expressions are set to constant ω^2 because both the expressions are independent to each other.

Now taking first and last expression of Eq. (6), therefore,

$$
\begin{aligned}
&\left[D_1 \left(\frac{\partial^4 \Phi}{\partial x^4} + 2\frac{\partial^4 \Phi}{\partial x^2 \partial y^2} + \frac{\partial^4 \Phi}{\partial y^4} \right) + 2\frac{\partial D_1}{\partial x} \left(\frac{\partial^3 \Phi}{\partial x^3} + \frac{\partial^3 \Phi}{\partial x \partial y^2} \right) + 2\frac{\partial D_1}{\partial y} \right. \\
&\left(\frac{\partial^3 \Phi}{\partial y^3} + \frac{\partial^3 \Phi}{\partial y \partial x^2} \right) + \frac{\partial^2 D_1}{\partial x^2} \left(\frac{\partial^2 \Phi}{\partial x^2} + v\frac{\partial^2 \Phi}{\partial y^2} \right) + \frac{\partial^2 D_1}{\partial y^2} \left(\frac{\partial^2 \Phi}{\partial y^2} + v\frac{\partial^2 \Phi}{\partial x^2} \right) \\
&\left. + 2\left(1 - v\right) \frac{\partial^2 D_1}{\partial x \partial y} \frac{\partial^2 \Phi}{\partial x \partial y} \right] - \rho l \Phi \omega^2 = 0
\end{aligned}
\tag{7}
$$

Eq. (7) represents differential equation of motion of plate with variable D_1 (rigidity of flux). The expression for flexural rigidity of the plate is $D_1 = \frac{Yl^3}{12(1-v^2)}$.

3 Frequency Equation and Assumptions

In order to obtain frequency equation and frequency modes, Rayleigh–Ritz technique is used, therefore,

$$\delta(V_s - T_s) = 0 \tag{8}$$

where

$$V_s = \frac{1}{2} \int_0^a \int_0^a D_1 \times \left[\left(\frac{\partial^2 \Phi}{\partial x^2} \right)^2 + \left(\frac{\partial^2 \Phi}{\partial y^2} \right)^2 + 2v \frac{\partial^2 \Phi}{\partial x^2} \frac{\partial^2 \Phi}{\partial y^2} \right.$$
$$\left. + 2(1-v) \left(\frac{\partial^2 \Phi}{\partial x \partial y} \right)^2 \right] dy dx \tag{9}$$

$$T_s = \frac{1}{2} \omega^2 \int_0^a \int_0^a \rho l \Phi^2 dy dx \tag{10}$$

Vibration of plate is a very vast area. It is not possible to study vibration at once; therefore, the present study requires some limitations in the form of assumptions.

1. For nonuniformity, linear thickness variation in plate is considered in one dimension as shown in Fig. 1

Fig. 1 Square plate with one dimensional linear thickness variation
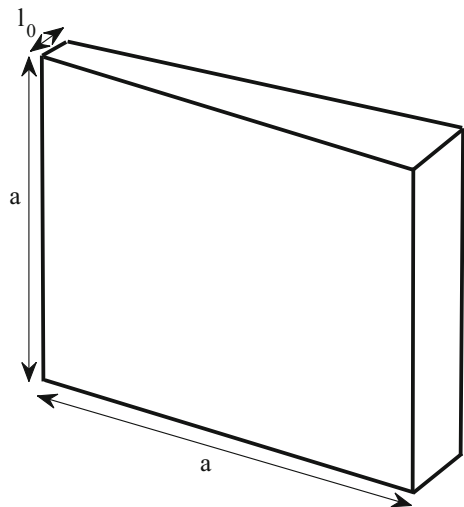
$$l = l_0 \left[1 + \beta \frac{x}{a} \right] \tag{11}$$

where β, $(0 \leq \beta \leq 1)$ is known as tapered parameter of the plate and $l = l_0$ at $x = 0$.

2. For consideration of nonhomogeneous material, author assumed circular variation in density as

$$\rho = \rho_0 \left[1 + m \left(1 - \sqrt{1 - \frac{x^2}{a^2}} \right) \right] \tag{12}$$

where m, $(0 \leq m \leq 1)$ is known as nonhomogeneity constant.

3. Here, variation in temperature on the plate is linear along both the axes, therefore,

$$\tau = \tau_0 \left(1 - \frac{x}{a} \right) \left(1 - \frac{y}{a} \right) \tag{13}$$

where τ and τ_0 represent temperatures above the mention temperature at any point on the plate, and at origin, i.e., $x = y = 0$. The modulus of elasticity for engineering material is

$$Y = Y_0 \left(1 - \gamma \tau \right) \tag{14}$$

where Y_0 is the Young's modulus at mentioned temperature, i.e., $\tau = 0$ and γ is slope of variation.

Using Eq. (13), (14) becomes

$$Y = Y_0 \left[1 - \alpha \left(1 - \frac{x}{a} \right) \left(1 - \frac{y}{a} \right) \right] \tag{15}$$

where α, $(0 \leq \alpha < 1)$ is known as temperature gradient, which is the product of temperature at origin and slope of variation, i.e., $\alpha = \tau_0 \gamma$. Using Eqs. (11) and (15), the flexural rigidity becomes

$$D_1 = \frac{Y_0 l_0^3 \left[1 - \alpha \left(1 - \frac{x}{a} \right) \left(1 - \frac{y}{a} \right) \right] \left[1 + \beta \frac{x}{a} \right]^3}{12 \left(1 - v^2 \right)} \tag{16}$$

4. Here frequency modes are computing on $C - C - C - C$, (i.e., all edges are clamped). Therefore, boundary conditions for clamped edges are

$$\Phi = \frac{\partial \Phi}{\partial x} = 0 \quad at \quad x = 0, a \tag{17}$$

$$\Phi = \frac{\partial \Phi}{\partial y} = 0 \quad at \quad y = 0, a$$

The deflection function, which satisfies Eq. (17) could be

$$\Phi\,(x,\,y) = \left[\left(\frac{x}{a}\right)^2 \left(\frac{y}{a}\right)^2 \left(1 - \frac{x}{a}\right)^2 \left(1 - \frac{y}{a}\right)^2\right]$$
$$\times \left[F_1 + F_2 \left(\frac{x}{a}\right)\left(\frac{y}{a}\right)\left(1 - \frac{x}{a}\right)\left(1 - \frac{y}{a}\right)\right] \tag{18}$$

where F_1 and F_2 represents arbitrary constants.

4 Solution for Frequency Equation

Now, introduce nondimensional variable X and Y as

$$X = \frac{x}{a}, \quad Y = \frac{y}{a} \tag{19}$$

Using Eqs. (11), (12), (16), (19), (9) and (10) becomes

$$V_s^* = \frac{Y_0 l_0^3}{24a^2(1 - \nu^2)} \int_0^1 \int_0^1 \left[[1 - \alpha\,(1 - X)\,(1 - Y)]\,[1 + \beta X]^3 \right.$$
$$\left. \times \left(\frac{\partial^2 \Phi}{\partial X^2}\right)^2 + \left(\frac{\partial^2 \Phi}{\partial Y^2}\right)^2 + 2\nu \frac{\partial^2 \Phi}{\partial X^2} \frac{\partial^2 \Phi}{\partial Y^2} + 2\,(1 - \nu) \left(\frac{\partial^2 \Phi}{\partial X \partial Y}\right)^2 \right] dY dX \tag{20}$$

$$T_s^* = \frac{1}{2}\omega^2 \rho_0 l_0 a^2 \int_0^1 \int_0^1 \left[1 + m\left(1 - \sqrt{1 - X^2}\right)\right] (1 + \beta X)\, \Phi^2 dY dX \tag{21}$$

Using Eq. (20), (21) and (8) become

$$\delta\left(V_s^* - \lambda^2 T_s^*\right) = 0 \tag{22}$$

Here, $\lambda^2 = \frac{12(1-\nu^2)\rho_0\omega^2 a^4}{Y_0 l_0^2}$ is known as frequency parameter.

Equation (22) consists of two unknowns constants, i.e., F_1, F_2 because of substitution of deflection function Φ. These constants can be determined by

$$\frac{\partial\left(V_s^* - \lambda^2 T_s^*\right)}{\partial F_i} = 0, \quad i = 1, 2 \tag{23}$$

After simplifying Eq. (23), homogeneous system of equation is obtained.

$$\begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{24}$$

where d_{11}, $d_{12} = d_{21}$ and d_{22} contain parametric constants and frequency parameter. To obtained solution, the determinant of coefficient matrix obtained from Eq. (24) is zero, i.e.,

$$\begin{vmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{vmatrix} = 0 \tag{25}$$

Equation (25) is quadratic equation from which two values (modes) λ_1, λ_2 are obtained.

5 Results Discussion

For numerical calculation, the following parameters are considered.

$$\rho_0 = 2.80 \times 10^3 \, \text{Kg/m}^3, \, \nu = 0.345, \, l_0 = 0.01 \, \text{m}$$

Table 1 shows the vibrational frequency of first two modes corresponding to nonhomogeneity (m) in the material for the following three cases.

$$Case1 \quad \alpha = \beta = 0, \quad Case2 \quad \alpha = \beta = 0.4, \quad Case3 \quad \alpha = \beta = 0.8$$

A decrement is observed in frequency modes for all the three cases when nonhomogeneity (m) in plate material increases. The rate of decrement is much lesser because of circular variation in density. When temperature (α) and tapering (β) increases, the frequency modes also increases.

Table 2 provides the frequency modes of vibration corresponding to temperature (α) variation on the plate for the following three cases.

$$Case4 \quad m = \beta = 0, \quad Case5 \quad m = \beta = 0.4, \quad Case6 \quad m = \beta = 0.8$$

Table 1 Nonhomogeneity (m) variation in plate material versus vibrational frequency (λ)

m	$\alpha = \beta = 0.0$		$\alpha = \beta = 0.4$		$\alpha = \beta = 0.8$	
	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2
0.0	35.99	140.88	41.79	163.56	47.91	187.82
0.2	35.46	138.48	41.14	160.63	47.17	184.31
0.4	34.94	136.21	40.53	157.84	46.45	180.99
0.6	34.45	134.04	39.95	155.20	45.77	177.85
0.8	33.98	131.98	39.38	152.69	45.11	174.87
1.0	33.53	130.01	38.85	150.30	44.49	172.05

Table 2 Temperature (α) variation on the plate versus vibrational frequency (λ)

α	$m = \beta = 0.0$		$m = \beta = 0.4$		$m = \beta = 0.8$	
	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2
0.0	35.99	140.88	42.40	165.09	49.07	190.02
0.2	35.08	137.31	41.47	161.50	48.11	186.35
0.4	34.15	133.65	40.53	157.84	47.14	182.60
0.6	33.19	129.28	39.57	154.09	46.14	178.77
0.8	32.19	126.01	38.57	150.24	45.11	174.87

Table 3 Thickness (taper constant β) variation in the plate versus vibrational frequency (λ) for fixed $m = 0.4$

β	$\alpha = 0.0$		$\alpha = 0.4$		$\alpha = 0.8$	
	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2
0.0	34.94	136.21	33.15	129.22	31.25	121.83
0.2	38.56	150.21	36.74	143.12	34.82	135.66
0.4	42.40	165.09	40.53	157.84	38.57	150.24
0.6	46.40	180.62	44.48	173.18	42.46	165.40
0.8	50.52	196.66	48.53	188.99	46.45	180.99
1.0	54.74	213.09	52.67	205.16	50.51	196.91

For the all the three cases, the frequency modes of vibration decreases with the increasing value of temperature (α). On the other hand, when the combined increasing value of nonhomogeneity (m) and tapering (β) varies from 0 to 0.8, frequency modes increases.

Frequency modes corresponding to thickness variation (β) in the plate with fixed value of nonhomogeneity ($m = 0.4$) in the plate material is presented in Table 3 for the following cases:

$$Case7 \quad \alpha = 0, \quad Case8 \quad \alpha = 0.4, \quad Case9 \quad \alpha = 0.8$$

A continuous increment is noted in the frequency modes for the all the cases when the thickness parameter (β) increases. When the temperature (α) increases on the plate, the frequency mode decreases.

6 Comparison of Results

A comparison for frequency modes of the present paper with [5] is given in the graphical form. Figure 2 provides the comparison of frequency modes of present paper and [5] corresponding to tapering parameter (β) for fixed values of temperature
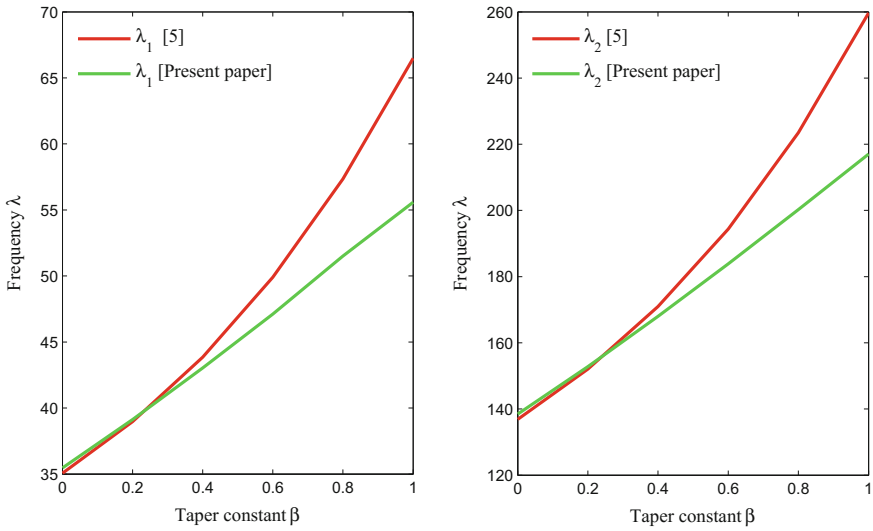
Fig. 2 Comparison of frequency modes of present paper with [5] corresponding to taper constant
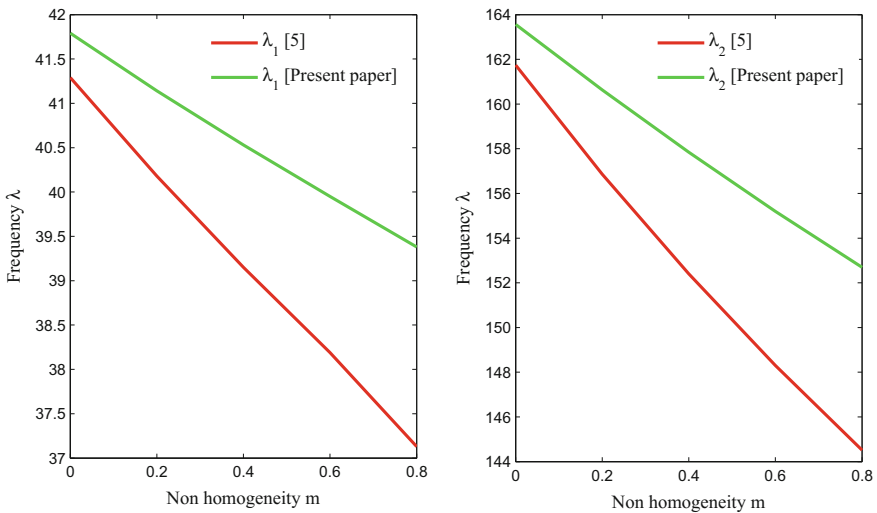
Fig. 3 Comparison of frequency modes of present paper with [5] corresponding to nonhomogeneity constant

($\alpha = 0$) and nonhomogeneity ($m = 0.2$) in the plate material. From Fig. 2, one can easily get that frequency mode for the present paper is less when compared to [5].

Figure 3 displays the comparison of frequency modes of present paper and [5] corresponding to nonhomogeneity (m) in the plate material for fixed temperature ($\alpha = 0.4$) and tapering ($\beta = 0.4$). Although the frequency of present paper is higher

when compared to [5] as shown in Fig. 3 but the frequency variation (rate of decrement) is less in present paper in comparison with [5].

7 Conclusion

Based on the graphical comparison and results discussion, the author concludes the following points.

1. Frequency modes are less in case of linear variation (present paper) in thickness when compared to exponential variation in thickness [5].
2. Nonhomogeneity of the material affects the vibrational frequency modes. The frequency modes are higher in case of circular variation (present paper) when compared to parabolic variation in density [5].
3. Frequency variation (rate of decrement) is less in case of circular variation in density (present paper) when compared to parabolic variation in density [5].
4. When temperature on the plate increases, vibrational frequency modes decreases.
5. By taking appropriate variation in the parameters, frequencies can be controlled actively.

References

1. Chopra, I., Durvasula, S.: Vibration of simply supported trapezoidal plates I: Symmetric trapezoids. J. Sound Vib. **19**, 379–392 (1971)
2. Gupta, A.K., Khanna, A.: Vibration of visco-elastic rectangular plate with linear thickness variation in both directions. J. Sound Vib. **301**(3–4), 450–457 (2007)
3. Hutchinson, J.R., Zillmer, S.D.: Vibration of a free rectangular parallelepiped. J. Appl. Mech. Trans. ASME **50**, 123–130 (1983)
4. Itakura, K.: Free vibration analysis of thick skewed plates having arbitrary boundary conditions. J. Struct. Constr. Eng. (Transaction of AIJ) **62**, 37–45 (1997)
5. Khanna, A., Deep, R., Kumar, D.: Effect of thermal gradient on vibration of non-homogeneous square plate with exponentially varying thickness. J. Solid Mech. **7**, 477–484 (2015)
6. Liu, H., Liu, F., Jing, X., Wang, Z., Xia, L.: Three dimensional vibration analysis of rectangular thick plates on pasternak foundation with arbitrary boundary conditions. Shock Vib. **2017**, 10 (2017)
7. Sharma, A., Sharma, A.K.: Mathematical modeling of vibration on parallelogram plate with non homogeneity effect. Roman. J. Acoust. Vib. **13**, 53–57 (2016)
8. Sharma, A., Sharma, A.K., Raghav, A.K., Kumar, V.: A modeling on frequency of rectangular plate. Int. J. Control Theory Appl. **9**, 277–282 (2016)
9. Sharma, A., Sharma, A.K., Kumar V., Raghav, A.K.: Computation of vibrational frequency on trapezoidal plate with non homogeneity effect. In: International Conference on Computing, Communication and Automation, pp. 1470–1473. IEEE (2017)
10. Sharma, A., Sharma, A.K., Rahgav, A.K., Kumar, V.: Vibrational study of square plate with thermal effect and circular variation in density. Roman. J. Acoust. Vib. **13**, 146–152 (2016)
11. Timoshenko, S., Woinowsky-Krieger, S.: Theory of plates and shells (1959)

12. Wang, Q., Shi, D., Shi, X.: A modified solution for the free vibration analysis of moderately thick orthotropic rectangular plates with general boundary conditions, internal line support on elastic foundation. Meccanica **51**, 1985–2017 (2016)
13. Zhang, H., Shi, D., Wang, Q.: An improved Fourier series solution for free vibration analysis of the moderately thick laminated composite rectangular plate with non-uniform boundary conditions. Int. J. Mech. Sci. **121**, 1–20 (2017)
14. Zhou, D., Cheung, Y.K., Lo, S.H., Au, F.T.K.: Three dimensional vibration analysis of rectangular thick plates on Pasternak foundation. Int. J. Numer. Methods Eng. **59**, 1313–1334 (2004)

To Calculate the Usability of Healthcare Mobile Applications Using Cognitive Walkthrough

Charu Bisht, Deepti Mehrotra and Parul Kalra

Abstract Usability is a measure of determining the ease of use of an application. It is for sure an exceptionally crucial part from a client's and business point of view. This paper means to discover strategies to improve the ease of use of an application. It utilizes the CW strategies to delineate user's viewpoint to the applications usability criteria. It familiarizes with how an application is seen by individuals of fluctuating age. The primary thought is to check the ease of use of an application and upgrade it, if required, such that it serves useful for all scope of clients.

Keywords Cognitive walkthrough (CW) · Learnability · Usability of mobile applications

1 Introduction

Websites and mobile applications have advanced ever since they were created. Necessity of customers have expanded as far back as at that point, thus has the scope of these applications. For instance, we have numerous medicinal mobile applications which serve a similar target. As a result, in order to showcase an application at its best, various factors are taken into consideration. Features and usability are the most recognized factors. Usability of a website/mobile application refers to how effortlessly it can be used without any guidance or a manual. Overtime some standard usability criteria have been developed.

C. Bisht · D. Mehrotra (✉) · P. Kalra
Amity University, Noida, Uttar Pradesh, India
e-mail: dmehrotra@amity.edu

C. Bisht
e-mail: chia.bisht@gmail.com

P. Kalra
e-mail: parulkalra18@gmail.com

Usability of an application can be addressed by the following criteria:

1. Does the website follow a simple pattern?
2. Are the tasks painless and palpable to finish?
3. Can information on the webpage be easily navigated?
4. Does the website grow with the user?
5. Does user has control and freedom to explore?
6. How strong an application is to the blunders?
7. Does the application have decent face esteem?

Usability of an application is closely related to the learnability of that application. If the application is well perceived by the user, we say that the application is more usable.

Thus, user plays a crucial role in deciding the usefulness of an application, which subsequently can derive a range of attributes like learnability and informativeness. There are many ways in which usability of a website can be checked. Cognitive walkthrough among those is a standout method. The technique takes into account distinctive user perspectives and mixes it with itself to anticipate the usability and learnability of website/mobile application.

2 Background

Cognitive walkthrough is one of those methods which examine a websites/mobile application in a way a novice user would. It is completed by considering the accompanying inquiries, which are as follows:

1. Will the user work in the direction to reach the goal?
2. Will the user locate the correct activity?
3. Will the user have the capacity to recognize progress post an activity?
4. Will the user interface the correct activity with the goal?

The above condenses how CW method can be related to a website/mobile application. Over the course of time, several researches have been conducted on usability of websites/mobile applications. "*Comparative Analysis of Cognitive Walkthrough Approaches for Measuring Usability of Web*" [1] discusses about the broadened scope of CW models. Jennifer in "*Examining the Usability of Web Site Search*" [2, 3] pressed on the use of meta-data to navigate in-between search queries. In "*Evaluating the Usability of Educational Websites Based on Students Preferences of Design Characteristics*" [4], Layla Hasan investigated about usability criteria and found that content and navigation were most essential criteria. NielsEbbe and Bonnie E. in "*Two Case Studies in Using Cognitive Walkthrough for Interface Evaluation*" [5, 6] contributed a detailed case study on interface evaluation using cognitive walkthrough.

This case study converges usability criteria with the user's viewpoint and gives a turn in determining the comparative usability among two or more mobile applications. A detailed case study is talked about in later sections.

2.1 CW on Mobile Application

Numerous speculations have come up and related CW techniques with usability and learnability of applications. Because of its expanding interest and consideration, this study will perform CW on healthcare mobile applications [7].

3 Research Assumptions

Considering the prerequisites, a precognitive survey was conducted which yielded with the following assumptions:

1. Informativeness of health apps is positively associated with the age factor.
2. Cognitive skill of an individual will contribute more toward performing the task successfully over the knowledge of the same individual.
3. Tech-friendly partakers will navigate the application with ease when compared to the others.
4. Length of the text is directly proportional to the informativeness quotient of an application.
5. By integrating the cognitive principle with the usability criteria, improvement in the expediency quotient of these applications is anticipated.

4 Methodology

Three average-rated healthcare mobile applications were picked at random from the Google Play store for the case study. For convenience, they were labeled as A, B, and C respectively. They were precisely watched and some intermediate data was inferred. This data was blended with the precognitive survey done beforehand and a questionnaire was shaped. A survey of the above was conducted, the overview of which is discussed underneath.

4.1 Survey

A questionnaire was prepared using the information deduced in the precognitive survey. It comprised of 11 questions, which had five checkboxes each. Each question depicted a task and each checkbox related itself as a sub-task. The tasks ranged from the basics of the mobile application (like downloading and installing it) to the specifics like booking appointments with the doctors. Partakers were requested to fill the questionnaire, where each tick against the checkbox met all requirements for score "1" and the otherwise fit the bill for score "0". Add up to score of question was

Fig. 1 Survey scale

added up to number of tick detriments for the checkboxes. Least score per question was 1 and most extreme was 5.

The scoring of the questionnaire can be explained by the following illustration: MY ORDERS

- WAS IT EASY TO FIND THE MEDICINES ON THE MENU?
- WERE THERE ALTERNATIVES TO THOSE MEDICINES?
- WERE YOU ABLE TO ORDER THE MEDICINES?
- WERE YOU ABLE TO SELECT MODES OF TRANSACTION?
- WERE YOU ABLE TO FIND OTHER INFORMATION REGARDING THE MEDICINES?

The above is one of the few questions, which were enlisted in the survey. Assume the client could perform three out of five tasks; at that point the score ends up noticeably three for the above inquiry. Figure 1 shows the scale provided to participants.

4.2 Partakers

A sum of 180 people had contributed to this study. Their age ranged from 15–70 yrs and they were grouped as young, grown-up, and old. The mass was differing in nature as partakers had individual take on technology. Some were novice. Some had analytical skills yet not learning about the application. The diverse crowd was a requirement as each of them has different cognitive style and each would contribute distinctively toward the end result of the case study. Figure 2 depicts the graphs showing different number of participants per age group.

5 Case Study

5.1 Survey

The survey (discussed in Sect. 4.1), which was made utilizing the Google forms. It was converted into an Excel file and scores were ascertained for every question

Fig. 2 Participants

Fig. 3 Dataset

for each application. The dataset was further molded and was inputted for further calculation. Figure 3 shows a screenshot of sample from the data set.

5.2 Partakers

180 people became a part of this study. They ranged from 15–70 yrs and included both male and female gender. People were selected at random so that their individual cognitive styles could be assessed.

Fig. 4 Score for "young"
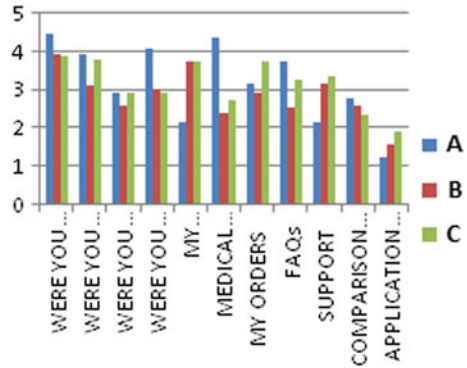
5.3 Technique

This case study uses an extended version of CW3. It is likewise alluded as "CW for Web" [1, 8]. Employing this technique, we evaluate the usability from the user's standpoint. The precognitive survey and the survey questionnaire were based on this technique and how it is further utilized is explained in the approaching segment.

5.4 Procedure

All partakers were approached to fill the survey for any of the mentioned healthcare applications, as per the tasks suggested by the form. Format for all the applicants was same. The data generated by this survey was compiled and sorted by application name. It was additionally grouped into age, namely young, grown-up, and old. The scores were aggregated for each of these divisions and the results were analyzed.

6 Result and Analysis

The accompanying outcomes were acquired after the aggregated categories (discussed in Sect. 5.4) were compared. The outcomes were shown in the form of charts in order to make the comparison simple.

Figures 4, 5, and 6 give a comparison analysis of each of these applications. After further streamlining, we obtain the following:

By interpreting Fig. 7, it was observed that the "young" category favored C the best followed by A followed by B.

It was additionally observed that for "grown-up" and "old" categories, A conveyed best outcomes took after by B and C respectively.

Fig. 5 Score for "middle aged"
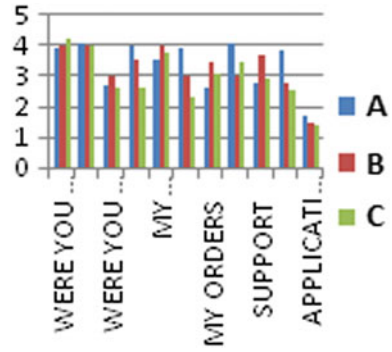
Fig. 6 Score for "old"
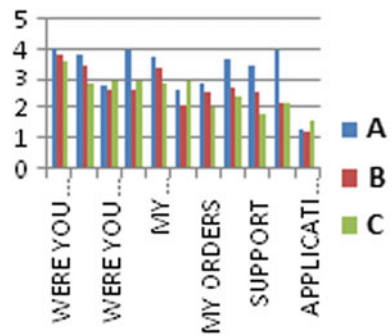
Fig. 7 Average score
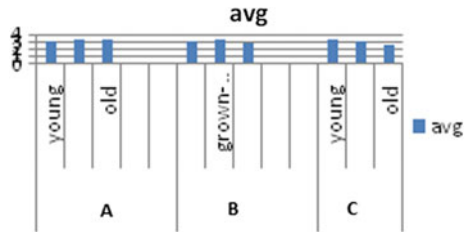
So, for a large portion of the experiments, A remained over the other applications. Hence, we can state A was most usable among the three followed by B and C.

7   Conclusion

The principle target of this contextual study was to use cognitive walkthrough to discover the ease of use of the application. In particular, this contextual analysis focused on three anonymous healthcare mobile applications, namely A, B, C. By methods for utilizing extension of CW3, user's viewpoint about these applications was recorded. This data was further extended and analyzed. The data was mainly contemplated on

age of the user. It was observed that the outcome contrasted reasonably with these components. "Old" was a venerable in correlation with the other two. The overall study also concluded that each age category viewed the applications differently. For "young" C was most liked application though the other two were much the same as A. Taken in general, it was tended to that extraordinary clients had one of kind psychological ability and for that measure, three of the applications were assessed distinctively.

Research Assumptions

Considering the prerequisites, a precognitive survey was conducted which yielded with the following assumptions:

1. Informativeness of health apps is positively associated with the age factor.
2. Cognitive skill of an individual will contribute more toward performing the task successfully over the knowledge of the same individual.
3. Tech-friendly partakers will navigate the application with ease when compared to the others.
4. Length of the text is directly proportional to the informativeness quotient of an application.

References

1. Rohatgi, P., Nagpal, R., Mehrotra, D.: Comparative Analysis of Cognitive Walkthrough Approaches for Measuring Usability of Web, Amity School of Engineering and Technology, Amity University Noida, India
2. Rong Tang Cognitive Styles: Demographic attributes, task performance, and affective experiences: an empirical investigation into astrophysics data system (ADS) core users. J. Lib. Inf. Stud. **14**(1) (2016)
3. Wharton, C., Rieman, J., Lewis, C., Polson, P.: Institute of Cognitive Science, University of Colorado Boulder, Colorado 80309-0344. The Cognitive Walkthrough Method: A Practitioner's Guide
4. Task-Centered User Interface Design: A Practical Introduction by Clayton Lewis and John Rieman. http://hcibib.org/tcuid/chap-4.html
5. Blackmon, M.H.: Cognitive walkthrough. In: Bainbridge, W.S. (ed.), Encyclopedia of Human-Computer Interaction, 2 volumes, vol. 1, pp. 104–107. Berkshire Publishing Group, Great Barrington, MA (2004)
6. Preben Hansen Swedish Institute of Computer Science: Evaluation of IR User interface-Implications for User Interface Design
7. Yuan, X. (Systems University at Albany), Liu, J. (State University of New York, School of Library and Information Science, University of South Carolina): Relationship Between Cognitive Styles and Users' Task Performance in Two Information
8. How to Conduct a Cognitive Walkthrough. https://www.interaction-design.org/literature/article/how-to-conduct-a-cognitive-walkthrough

WUCA: An Analysis of Web Usability and Content Accessibility of Webpages with Respect to Ailment People

Abid Ismail and K. S. Kuppusamy

Abstract The present age is the age of Information Technology (IT) and the Governments across the globe focuses on e-governance services. The websites serve as an excellent information dissemination platform, which is used by both disabled and non-disabled population. This study investigates the web usability and content accessibility called *WUCA* of webpages with a case study of affiliated college websites of Pondicherry University (PU), which is a leading central university in India. It also explores whether they are ranked high in terms of WUCA and whether these two measures are correlated. We have used automatic evaluation tools such as *TAW, Pingdom, and PageSpeed tools* to measure the accessibility and usability of eighty-three (83) college websites affiliated with Pondicherry University. Based on evaluation results, it has been observed that the overall performance grade (out of A, B, C, D) of these websites falls in category *B* and 12% websites fall in lower grade performance category *D*. It is found that the desktop device user experience is better than mobile device user experience. It is also found that out of 83 websites selected, 12 websites are highly optimized, 20 websites are less optimized, and 51 websites are not optimized. The correlation between accessibility and usability rank cases (SPSS Procedure) indicates that the accessibility is highly correlated with *load-time* and *page-size*. It was also inferred from the results that there is comparatively better compliance with accessibility and a relatively low web usability for most of the websites. Based on these results, some suggestions are framed to improve the WUCA of said websites for efficient access by persons with disabilities (PwDs).

Keywords Web accessibility · WCAG · Usability · Evaluation tools · PwDs
Rank cases · Bivariate correlation · Accessible guidelines

A. Ismail · K. S. Kuppusamy (✉)
Department of Computer Science, School of Engineering & Technology,
Pondicherry University, Pondicherry 605014, Puducherry, India
e-mail: kskuppu@gmail.com

A. Ismail
e-mail: abidpu2015@gmail.com

1   Introduction

Presently, we live in the era of information technology in which most of the day-to-day activities and information needs are served by the web platform. The web should be accessible for all types of people, which certainly includes *persons with disabilities* (PwDs). In 2016, the statistical profile of India about disabled persons [1] indicated that out of 121 crore population, about 2.68 crore people have some sort of disability which is 2.21% of the total population of the country as per census 2011. The break-up of this disability data is as follows: 20% are locomotor disabled, 19% hearing, 19% seeing, and 8% multiple disabilities as per census 2011 [1].

1.1   *Web Usability and Content Accessibility*

The central aim of Web Usability is to make the web interfaces more user friendly and easy to use for web users especially PwDs.

The cardinal aim of Web Accessibility, also called web content accessibility is to address the needs of PwDs such as *blind, partially blind, auditory (hearing), motor, seizures, cognitive or intellectual*, etc. It also refers to the *inclusive practice* means to access websites without any barrier for PwDs. All users have equal access to information and functionality when websites are correctly designed, developed, and maintained as per accessibility standards.

Following are the key points about website accessibility and usability [8, 16, 20]:

- A website is not usable unless the web is accessible [16].
- Accessibility is a subset of Usability [8].
- An accessible website benefits all types of users whether the user is abled or disabled one [20].
- Usability implies accessibility but the opposite is not necessarily true [8].

Generally, the accessibility refers to items that are designed to be usable by persons with disabilities. The products and environments which is designed to be usable by all types of users to the greatest extent possible, and the universal design expands accessibility to include the needs of a much broader audience.

The Universal Design (UD) addresses the scope of accessibility. It requires an understanding and consideration of the wide range of human abilities throughout the life period. It also suggests to make all components accessible and usable irrespective of the type of users to the greatest extent possible.

The fundamental key objectives of this paper are as under:

1. To find the web accessibility status of websites covered by the case study.
2. To find the website content accessibility covered by the case study.
3. To determine the correlations between accessibility and usability of websites.
4. To provide proper suggestions to increase the ratings of accessibility and usability of websites.

2   Literature Review

A study was carried out by Zaphiris and Ellis [24] on top USA universities in terms of usability and web content accessibility. They used two automatic assessment tools to find the accessibility and usability status of these top USA universities websites and then find their correlations also. It was found that the accessibility ratings are significantly correlated with overall usability ratings of websites. Ismail et al. [13, 14] have carried out the study on website content accessibility and readability of Indian universities websites and top government websites in India by using different online automatic evaluation tools. In addition to this, they provide some useful suggestions for the enhancement of accessibility and ease of read status of websites.

Ma et al. [18] investigate the United Kingdom's e-government websites in terms of their usability and content accessibility by using automatic evaluation tools. It was found that the UK's e-government websites are relatively in high compliance with web content accessibility guidelines (WCAG) and relatively low usability ratings. Another study was carried out [12] to find the accessibility status of Northeastern India region websites by using WAVE and EvalAccess automatic tool. It was found that some common list of errors should be minimized to enhance their accessibility status.

Content accessibility of websites has been conducted by Beaudin [6] to analyze the 66 government websites of Australia, UK, and the USA for disabled persons based on W3C accessibility guidelines by using different automatic evaluation tools. During analysis, they also considered the websites in terms of font size, navigation, and general usability perspectives. Similar work was also carried out by researchers such as Lazar et al. [17], Mankoff et al. [19], Huang et al. [11], Ismail et al. [15], etc., on accessibility and usability of web contents. They have also provided some useful suggestions for improvements.

Mustafa et al. [21] evaluated the Jordan Universities' websites in terms of web usability perspective. They used two automatic online evaluation tools to measure the internal attributes of websites. It was found that the usability of Jordan universities websites is acceptable and also, there is a need to further enhance the usability of these websites. Petrie et al. [22] carried out the study based on evaluation of usability, accessibility, and user experience. They outlined the various methods with respect to their appropriate use, strengths, and weaknesses for evaluating the accessibility, usability, and user experience. Another study was carried out by Aziz et al. [5] on usability and accessibility of higher education websites of Malaysia. They also used page size, speed, and broken link attributes for the usability evaluation of higher education websites of Malaysia.

A review work on usability evaluation was carried out by researchers [10] based on 31 research papers to find the user's design feedback. Another work carried out by Blanco-Gonzalo et al. [7] focused on analyzing the usability and accessibility of a face recognition system used by six people. In addition, a framework was proposed by researchers [23] to make the websites better accessible and usable to persons with disabilities particularly blind web users.

3   Motivational Questions and Methodology

Motivational Questions: For the purpose of performing correlation study between usability and accessibility, we considered the 83 affiliated college websites of Pondicherry University (PU), which is a leading central university of India. The scope of this paper is to find the solution of the following questions:

Q1: Are PU-affiliated college websites rated highly in terms of accessibility and usability?

Q2: Is the result of accessibility related to the result of usability evaluation?

Methodology: During February–June 2017, we have collected [4] a list of 83 affiliated colleges of Pondicherry University. These colleges are categorized into nine categories[1] based on arts, science, education, law, medical and research fields, etc. The two automatic evaluation tools namely *TAW* [9] *and Pingdom* [3] used to evaluate them in terms of Accessibility and Usability. In addition to this, we have also checked their performance for mobile and desktop devices in terms of three categories namely *Good, Needs Work, and Poor* by using *PageSpeed* tool [2]. The brief description of these tools are as follows.

TAW-Test Accessilibilidad Web Tool: It was developed by CTIC Foundation that allows the monitoring over time of different criteria like accessibility, quality of development, search engine positioning, standards, etc., on different regulations like WCAG 1.0, WCAG 2.0, etc. We have used WCAG 2.0 analyzer under Level AA of this tool to find out the accessibility evaluation status of these 83 college websites. TAW tool also categorizes the accessibility errors into three categories namely *Problems—corrections are needed, Warnings—a human review is necessary and Not reviewed—completely human review*.

Pingdom Tool: It is a website speed test tool used for monitoring the website's uptime, performance, and load time. Also, its reciprocal action for a better end-user experience and is easy tool to help web masters; and everywhere the web developers optimize the performance of websites. The attribute *Requests* in Pingdom is by content type or by domain, and the *Load time* attribute involves nonidentical stages of a requests. Based on requests, test from (almost all options) and size attributes of the tool, we have calculated the performance, size, and download time of the website.

PageSpeed Tool: It is also called *PageSpeed Insights*. It is used for mobile and desktop devices to measure the performance of a page. It fetches the URL of the site twice and provides a score of website from 0 to 100 points under three performance statuses with respect to particular categories namely Good category (0 to 65)—good user experience and applies most performance best practices, Needs Work category

---

[1] 1. Arts and Science Colleges, 2. Engineering Colleges, 3. B.Ed. Colleges, 4. Medical Colleges, 5. Nursing and Paramedical Colleges, 6. Dental Colleges, 7. Veterinary Colleges, 8. Research Institutions/Centres, 9. Law Colleges.

(66 to 85)—slow user experience and missing some common performance optimizations, and Poor category (86 to 100)—slow user experience and is not optimized. It also considers network-independent aspects of page performance.

4   Result Analysis and Discussion

For giving an answer to motivational question Q1, we have to calculate the mean and standard deviation of accessibility and usability findings of these 83 college websites affiliated with Pondicherry University.

The mean and standard deviation of accessibility and usability data of 83 college websites analyzed by using automatic evaluation tools namely *TAW and Pingdom* are listed in Tables 1, 2 and 3.

In Table 1, we present the mean and standard deviation of all websites individually with respect to TAW tool problems, TAW tool warnings and TAW tool not reviewed under the Perceivable (P), Operable (O), Understandable (U), Robust (R), and Success Criteria (SC) violations, respectively, obtained from the automatic evaluation TAW tool. In Table 2, we present the sum of results of all websites of P, O, U, R, and SC violations under TAW tool problems, TAW tool warnings, and TAW tool not reviewed attribute, respectively.

The overall total of all number of Perceivable, Operable, Understandable, and Robust violations of all websites under TAW tool problems, TAW tool warnings, and TAW tool not reviewed attribute, respectively, along with their means and standard deviations are also presented in Table 2. In addition, we present the mean and standard deviation of TAW tool evaluation per website involving the problems, warnings, and not reviewed attributes of TAW tool with respect to Perceivable, Operable, Understandable, Robust, and Success Criteria also.

In Table 3, we present the mean and standard deviation per website obtained from the Pingdom Tool with respect to performance of the webpage, page load time in seconds, speed of page in percentage, page size of the web page in MB (Mega Bytes), and requests of the page. In addition, the Pingdom Tool involves the performance insights based on different features, response codes, requests by content type, content size by domain and by content type, requests by domain, file requests, etc., features to calculate the result of the website. Thus, the Pingdom tool provides a report in the form of performance score, load time, speed, and requests. In addition to a general grading, Pingdom assigns a grading of *A (90–100), B (80–90), C (70–80), and D (less than 70)* also. The results of web page size, load time, and performance grade of websites obtained from automatic evaluation tool called *Pingdom* are presented in Table 4.

From Table 1, it is found that the *perceivable* problems and *robust warnings* are higher than others and the success criteria violations are also high in *not reviewed* category. The mean of *warnings* is also high as presented in Table 2. These are the major hurdles for the accessibility status of websites so need to be minimized them. The overall accessibility score of these websites are good but need to be

Table 1 TAW tool:type-wise mean and standard deviation accessibility status of websites (83) affiliated with Pondicherry University

Accessibility (N = 83)	Problems					Warnings					Not reviewed				
	P	O	U	R	SC	P	O	U	R	SC	P	O	U	R	SC
Mean	18.08	5.71	1.59	13.04	5.06	54.2	22.8	2.39	187.94	8.82	4.0	6.58	4.9	0.6	16.12
STDEV	20.16	7.08	3.25	24.42	1.66	59.3	31.3	4.86	320.64	3.57	0.0	1.16	0.2	0.5	1.51

Where P means Perceivable, O means Operable, U means Understandable, R means Robust, SC means Success Criteria, STDEV means Standard Deviation, N means Number of websites used by TAW tool

Table 2 Overall TAW tool accessibility score chart of eighty-three (83) college websites affiliated with Pondicherry University

Accessibility status-TAW tool (N = 83)	P	O	U	R	SC	Total (POUR)	Mean	St. deviation
Problems (T1)	1501	474	132	10822	420	**3189**	38.42	38.58
Warnings (T2)	4501	1896	198	15599	732	**22194**	267.40	323.91
Not reviewed (T3)	332	546	410	49	1337	**1337**	16.2	1.51
Total (T1, T2, T3)	6334	2916	740	16730	2489	**26720**		
Mean	76.31	35.13	8.92	201.57	29.99	**321.93**		
St. deviation	66.27	31.99	7.77	321.70	3.51	**327.48**		

Where P = Perceivable, O = Operable, U = Understandable, R = Robust, SC = Success Criteria

Table 3 Result status of 83 websites under Pingdom website evaluation tool

Usability (Pingdom) N = 83	Performance score (out of 100)	Load time (S)	Speed (% age)	Page size (MB)	Requests
Mean	81.47	8.81	49.69	4.02	51.02
Standard deviation	8.64	15.29	32.29	12.53	46.41

improved further because of these perceivable problems and robust warnings should be minimized further.

From Table 3, the overall usability score falls in *category—B*. The mean and standard deviation indicate that performance score is good but load time and speed need to be optimized further. Also, page size and request attribute deviation are too high and need to be minimized so that the usability score or ratings will be improved. Figure 1 represents the performance grade result of 83 websites by using Pingdom tool. The result indicates that 11% web sites fall in Grade A (90–100), 58% websites falls in Grade B (80–90), 19% websites falls in Grade C (70–80), and 12% websites falls in lower Grade D (less than 70) performance in terms of usability evaluation process. Furthermore, Table 4 represents the mean and standard deviations in terms of *web page size* and *load time*. It also presents the overall performance grade of these webpages obtained as Grade "B" in terms of said performance attributes.

From the individual college website evaluations by using *PageSpeed Tool based on user experience and page performance*, Table 5 represents the mean and standard deviations along with their overall qualifying percentage type of performance. The graphical representation of mobile- and desktop-based score of websites in terms of *poor, needs work, and good user experiences* are in Fig. 2.

For mobile device user experience, it is found that there are 69.88% websites are Poor, 22.89% websites are Needs Work means intermediate, and only 7.23% websites are Good in user experience and performance best practices. Similarly, for desktop device user experience, it is found that 60.24% websites are Poor, 22.89%
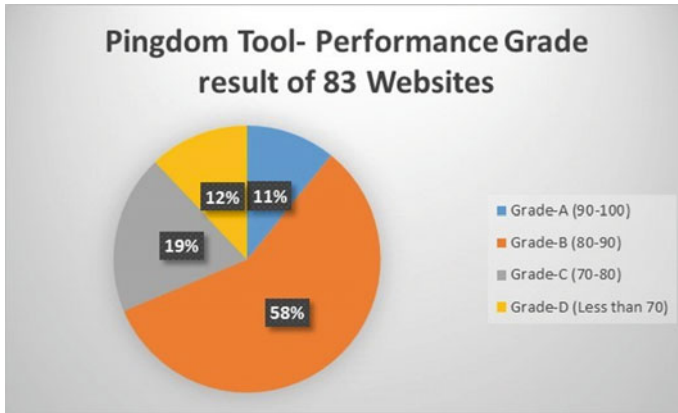
Fig. 1 Result of Pingdom tool in terms of performance grade along with their percentage score

Table 4 Pingdom tool—mean and standard deviation of webpage size and load time along with overall performance grade of 83 websites

Pingdom tool, N = 83	Mean	STDEV
Size (Mb)	4.02	12.53
LoadTime (s)	8.81	15.29
Performance grade (A, B, C, D)	B	

websites are Needs Work, and 16.87% websites are Good in terms of user experience and performance practices.

Hence, in general, the result indicates that desktop device user experience is better than mobile device user experience. In addition, these 83 websites are classified into three categories namely Good for highly optimized websites (12 in number), Needs work for less optimized websites (20 in number), and Poor for not optimized websites (51 in number) based on the result obtained by using PageSpeed evaluation tool. Hence, there are 61.45% websites Poor, 24.10% websites Needs work, and 14.46% websites Good in user experience and performance practices.

For motivational question Q2, we have to find the Bivariate Correlation between accessibility and usability involving TAW, Pingdom, and PageSpeed automatic tools. Therefore, to answer this question Q2, Table 6 represents the bivariate correlation between accessibility and usability rank cases generated by using SPSS procedure. The result of correlation in Table 6 indicates that the correlations inside the variables are strong but with outside variables are less. That is, when comparing the correlation between PageSpeed and Pingdom variables are strongly correlated and the correlation between PageSpeed or Pingdom with TAW variables are weakly positively correlated.

Table 5 PageSpeed tool: user experience based on mobile and desktop devices

User experience (PageSpeed, N = 83)	Mobile			Desktop		
	Poor	Needs work	Good	Poor	Needs work	Good
Total	58	19	6	50	19	14
Percentage	69.88	22.89	7.23	60.24	22.89	16.87
Mean	**52.19**			**57.74**		
STDEV	**21.94**			**23.29**		
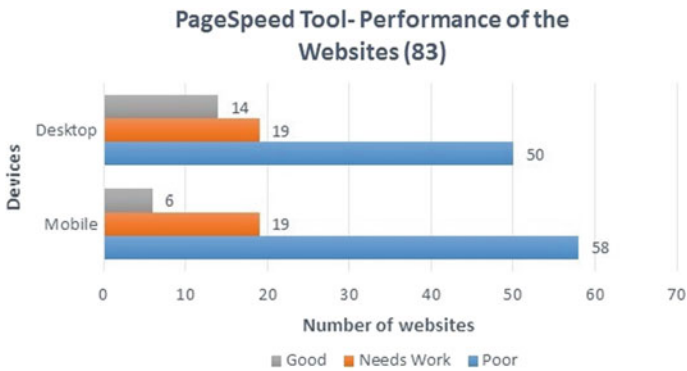
Fig. 2 Device-based result of PageSpeed insights tool with their category status

Table 6 Bivariate Pearson's correlation between accessibility and usability variables

Bivariate correlations

		PageSpeed	Pingdom	TAW
PageSpeed	Pearson correlation	1	0.129	1.000**
	Sig. (2-tailed)		0.246	0.000
	N	83	83	83
Pingdom	Pearson correlation	0.129	1	0.129
	Sig. (2-tailed)	0.246		0.246
	N	83	83	83
TAW	Pearson correlation	1.000**	0.129	1
	Sig. (2-tailed)	0.000	0.246	
	N	83	83	83

** Correlation is significant at the 0.01 level (2-tailed)

Also, the accessibility ranking correlates significantly with the usability as shown in Table 6. Furthermore, the size and load time means PageSpeed are highly correlated with accessibility but not with usability.

5 Suggestions

Based on our web content accessibility and usability results of websites, we provide the corresponding suggestions and measurements to improve the performance of websites in terms of accessibility along with their usability of websites.

Following are the possible optimization measures to enhance the performance of these sites as presented in Table 7.

In addition to this, how to make sites better in terms of website accessibility and website usability, the following measures presented in Table 8 shall be taken into consideration.

Table 7 Possible optimization measures to improve the performance of websites

S. No.	Optimization measures
1	Enable compression
2	Optimize images
3	Resistance browser caching
4	Elimination of render blocking CSS and Javascript in above the fold content
5	Minification of Javascript
6	Minification of CSS
7	Minification of HTML and
8	Reduce server response time

Table 8 Measures for better accessibility and usability

S. No.	Measures for accessibility and usability
1	To provide a search engine within the website
2	To make easy navigation within the website
3	To provide a contact email address
4	Use of thematic colors and design throughout the site
5	To provide a text-based option on the main page

6   Conclusion

Accessibility is a subset of Usability. Nowadays, Usability often assumes accessibility. The inclusion of more people in more situations is the main motive of accessibility. To summarize, it was found that the college websites have plenty of scope for improvement in both web accessibility and usability dimensions. The correlations between the web accessibility and usability is also found weak but internally they are strongly correlated. The size and load time are highly correlated with accessibility but not with usability.

Based on results, these college websites are classified into three categories namely 14.46% college websites falls in *Good* category means highly optimized in user experience and performance, 24.10% college websites falls in *Needs work* category means less optimized in terms of user experience and performance, 61.45% college websites falls in *Poor* category means not optimized in user experience and performance. The overall performance grade of websites falls in grade "B" category out of four grades (A, B, C, D) as per Pingdom tool evaluation report.

The mean and standard deviation indicate that load time, size, speed, requests, etc., of webpages need to be minimized so that usability will be enhanced. In addition to this, the perceivable problems and robust warnings as per TAW tool are high, so need to be minimized to enhance accessibility score too.

In addition to this, the perceivable problems and robust warnings as per TAW tool are high, so need to be minimized to enhance accessibility score too.

References

1. Mospi. http://mospi.nic.in/sites/default/files/publication_reports/Disabled_persons_in_India_2016.pdf (2017). Accessed 16 June 2017
2. Pagespeed insights tool. https://developers.google.com/speed/pagespeed/insights/ (2017). Accessed 05 June 2017
3. Pingdom tool. https://tools.pingdom.com/ (2017). Accessed 05 June 2017
4. Pondicherry university. http://www.pondiuni.edu.in/affiliate-colleges (2017). Accessed 13 Feb 2017
5. Aziz, M.A., Isa, W.A.R.W.M., Nordin, N.: Assessing the accessibility and usability of Malaysia higher education website. In: 2010 International Conference on User Science and Engineering (i-USEr), pp. 203–208. IEEE (2010)
6. Beaudin, D.: A content analysis of disability access on government websites in Australia, the United Kingdom, and the United States. A Master's Paper for the MS in LS degree. SILS, University of North Carolina, Chapel Hill. Nov (2001). http://ils.unc.edu/MSpapers/2722.pdf. Accessed 30 Dec 2009
7. Blanco-Gonzalo, R., Poh, N., Wong, R., Sanchez-Reillo, R.: Time evolution of face recognition in accessible scenarios. Hum. Cent. Comput. Inf. Sci. **5**(1), 24 (2015)
8. Brajnik, G.: Automatic web usability evaluation: what needs to be done. In: Proceedings of Human Factors and the Web, 6th Conference (2000)
9. CTIC, F.: Taw tool (2017). http://www.tawdis.net/ingles.html?lang=en. Accessed 14 Mar 2017
10. Følstad, A.: Users' design feedback in usability evaluation: a literature review. Hum. Cent. Comput. Inf. Sci. **7**(1), 19 (2017)

11. Huang, C.J.: Usability of e-government web-sites for people with disabilities. In: 2003 Proceedings of the 36th Annual Hawaii International Conference on System Sciences, p. 11. IEEE (2003)
12. Ismail, A., Kuppusamy, K.S.: Accessibility analysis of north eastern india region websites for persons with disabilities. In: 2016 International Conference on Accessibility to Digital World (ICADW), pp. 145–148 (2016). https://doi.org/10.1109/ICADW.2016.7942530
13. Ismail, A., Kuppusamy, K.S.: Accessibility of Indian universities' homepages: an exploratory study. J. King Saud Univ. Comput. Inf. Sci. (2016)
14. Ismail, A., Kuppusamy, K.S., Kumar, A., Ojha, P.K.: Connect the dots: Accessibility, readability and site ranking-an investigation with reference to top ranked websites of government of India. J. King Saud Univ. Comput. Inf. Sci. (2017)
15. Ismail, A., Kuppusamy, K.S., Nengroo, A.S.: Multi-tool accessibility assessment of government department websites: a case-study with JKGAD. Disabil. Rehabil. Assist. Technol. pp. 1–13 (2017)
16. Krug, S.: Don't Make Me Think!: a common sense approach to Web usability. Pearson Education India (2000)
17. Lazar, J., Dudley-Sponaugle, A., Greenidge, K.D.: Improving web accessibility: a study of webmaster perceptions. Comput. Hum. behav. **20**(2), 269–288 (2004)
18. Ma, H.Y.T., Zaphiris, P.: The usability and content accessibility of the e-government in the UK. In: Proceedings of Human Computer Interaction International Conference, Greece (2003)
19. Mankoff, J., Fait, H., Tran, T.: Is your web page accessible?: a comparative study of methods for assessing web page accessibility for the blind. In: Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 41–50. ACM (2005)
20. Matera, M., Rizzo, F., Carughi, G.: Web usability: principles and evaluation methods. In: Web Engineering, pp. 143–180 (2006)
21. Mustafa, S.H., Al-Zoua'bi, L.F.: Usability of the academic websites of jordan's universities an evaluation study. In: Proceedings of the 9th International Arab Conference for Information Technology, pp. 31–40 (2008)
22. Petrie, H., Bevan, N.: The Evaluation of Accessibility, Usability, and User Experience (2009)
23. Verma, P., Singh, R., Singh, A.K.: A framework to integrate speech based interface for blind web users on the websites of public interest. Hum. Cent. Comput. Inf. Sci. **3**(1), 21 (2013)
24. Zaphiris, P., Ellis, R.D.: Website usability and content accessibility of the top usa universities. In: WebNet, pp. 1380–1385 (2001)

CORDIF: A Machine Learning-Based Approach to Identify Complex Words Using Intra-word Feature Set

Muralidhar Pantula and K. S. Kuppusamy

Abstract Identification of complex words is an interesting research problem with various application scenarios such as text simplification. There are various approaches to identify complex words either by incorporating the complete sentence in which the word appears or by focusing only on the word. This paper falls under the later category, which employs intra-word features in classifying a word either as simple or complex. A model termed CORDIF (COmplex woRD identification with Intra-word Features). The proposed methodology incorporates *19* intra-word features. These features are harnessed to train a machine learning model. A dataset termed as *CWIdataset* is built with the proposed set of intra-word features. With the proposed feature-set, an accuracy level of 84.75% was achieved. Later using this model, we have tested for identifying the complex words for nonnative persons. As a result, we concluded that for identifying complex words, personalized systems are needed.

1 Introduction

In recent years, many research studies have focused on complex word identification problem for simplifying the text and make it understandable to the people with low-literacy skills [1]. Complex word identification was first introduced as a subtask of Lexical Simplification pipeline [2].

Last year International Workshop on Semantic Evaluation (SemEval 2016) has given complex word identification as one of the tasks [1]. They specified that they have designed the first dataset to train machine learning algorithms to identify the complex words. Their goal was to develop a system to find the individual word difficulty from the words classified by some set of individuals. A total of 42 systems

M. Pantula · K. S. Kuppusamy (✉)
Department of Computer Science, Pondicherry University,
Pondicherry, India
e-mail: kskuppu@gmail.com

M. Pantula
e-mail: vijayamuralisarma@gmail.com

were developed using machine learning algorithms from the annotated data to test the individual's vocabulary. They specified that word frequencies is the major factor in the complex word identification.

This paper focuses on predicting the word complexity based on intra-word features. We created a CWIdataset (Complex word Identification dataset) with the words gathered from the websites where many specified that those words are complex. Similarly, Google easy words are shown in Table 1. We have trained a system with the CWIdataset and intra-word features. The extracted words from *SemEval* testing dataset was used to annotate by different persons and compared the results. As a result, this paper concludes that the word complexity varies with person's individual vocabulary.

The main objectives of this study are the following:

- Building a model termed *CORDIF* to classify the complex word and easy word using intra-word features.
- Train the model with the CWIdataset and intra-word features to identify the complex words and easy words.
- Predicting one's vocabulary from the dataset developed.

2 Related Studies

Research on Identification of Complex words for making Lexical simplification is gaining increased significance in the recent years [3–6], even though lexical simplification research has started before two decades [7, 8].

At first, the experts developed readability ease formulas to calculate the readability of others' writings [9]. Various readability ease metrics were developed for identifying the level of persons to perceive the content. Various research studies were conducted on analyzing the readability measures for the given content like books [10, 11], websites [12], captions [13], medical records [14, 15], etc.

Next comes the task of simplifying the document for making it accessible to low-literacy persons. Lexical simplification pipeline is an alternative, in which the system identifies the complex words and will be replaced by simpler and easy words with equivalent meaning. The alternative word for complex words will be identified with the help of Wordnet database. Complex word identification has become a major task to simplify the text. The approaches for complex word identification were

- Threshold-based approach
- Train the system to classify
- Simplify Everything.

In threshold-based approach, a threshold is placed on the features to identify the complex words. In the second approach, a model will be trained with the features to optimize the results. The third method is simplifying the complete document.

Table 1 Systems for complex word identification

Team	System	Accuracy	Precision	Recall	F-score
SV000gg	Soft	0.779	0.147	0.769	0.246
TALN	RandomForest WEI	0.812	0.164	0.736	0.268
UWB	All	0.803	0.157	0.734	0.258
PLUJAGH	SEWDF	0.795	0.152	0.741	0.252
JUNLP	NaiveBayes	0.767	0.139	0.767	0.236
HMC	DecisionTree25	0.846	0.189	0.698	0.298
JUNLP	RandomForest	0.795	0.151	0.73	0.25
MACSAAR	RFC	0.825	0.168	0.694	0.27
TALN	RandomForest SIM	0.847	0.186	0.673	0.292
Pomona	NormalBag	0.604	0.095	0.872	0.171
Melbourne	runw15	0.586	0.091	0.87	0.165
UWB	Agg	0.569	0.089	0.885	0.161
Pomona	GoogleBag	0.568	0.088	0.881	0.16
IIIT	NCC	0.546	0.084	0.88	0.154
LTG	System2	0.889	0.22	0.541	0.312
MAZA	A	0.773	0.115	0.578	0.192
Melbourne	runw3	0.513	0.08	0.895	0.147
ClacEDLK	ClacEDLK-RF 0.6	0.688	0.081	0.548	0.141
PLUJAGH	SEWDFF	0.922	0.289	0.453	0.353
IIIT	NCC2	0.465	0.071	0.86	0.131
ClacEDLK	ClacEDLK-RF 0.5	0.751	0.09	0.475	0.152
MAZA	B	0.912	0.243	0.42	0.308
AmritaCEN	w2vecSim	0.627	0.061	0.486	0.109
GARUDA	SVMPP	0.796	0.099	0.415	0.16
AIKU	native1	0.583	0.057	0.512	0.103
AKTSKI	wsys	0.587	0.056	0.49	0.1
AKTSKI	svmbasic	0.512	0.053	0.558	0.097
USAAR	entropy	0.869	0.148	0.376	0.212
Sensible	Combined	0.737	0.072	0.39	0.122
BHASHA	SVM	0.844	0.119	0.363	0.179
CoastalCPH	NeuralNet	0.693	0.063	0.398	0.108
LTG	System1	0.933	0.3	0.321	0.31
USAAR	entroplexity	0.834	0.097	0.305	0.147
AmritaCEN	w2vecSimPos	0.743	0.06	0.306	0.1
GARUDA	HSVM&DT	0.88	0.112	0.226	0.149
CoastalCPH	Concatenation	0.869	0.08	0.171	0.109

As a goal to identify an individual nonnative English speaker word difficulty from a set of nonnative speaker vocabulary, SemEval builds a dataset from three corpuses namely CWcorpus, LexMTurk corpus, and Simple Wikipedia [1]. The set of systems developed on the dataset are placed in Table 1. Out of all these systems, the complex words are evaluated either by the probability of the word in the sentence, word embedding, word frequency, and the position of the words.

In this paper, CORDIF system is built to classify the targeted word as either complex or easy. The complexity of the word can be analyzed from the features of the sentence in which the word exist and from the internal features of the word.This paper focuses on finding the complexity based on the intra-word features explained in Sect. 4.

3 Dataset for Complex Word Identification

Paetzold and Specia created a dataset based on the words annotated by a set of nonnative English speakers [1]. The dataset having set of sentences, in each sentence they surveyed the complex words from nonnative users. The systems developed on the basis of this dataset is placed in Table 2.

We have designed a new dataset based on websites listed in Table 2. In the list of websites, anonymous persons specified their difficult words and the reason of their difficulty. Here, the users are anonymous we have got 2171 unique complex words and selected the same number of easy words out of 49,569 unique easy words extracted from Google easy words and other sources listed.

The variations in the datasets with respect to length, syllables, and synsets are mentioned in Table 3. It is clearly mentioned that there is no difference in any of the data.

4 Intra-word Features

Feature selection is very important in classifying any data. In this paper, we build a CORDIF system to classify the targeted word as complex or easy word. It seems like a simple binary classification. The feature extraction for this task can be done by either extracting the features from the sentence or from the features of the word itself. The features for classifying the data are:

4.1 Length

The number of characters in the word is taken as one feature $w = c_1 c_2 c_3, \ldots, c_n$, the word w is of length n. This paper assumes that the word with higher length is relatively complex.

Table 2 Dataset-extracted websites

S.No	Name of the Website
	Difficult words
1	https://www.vocabulary.com/lists/191545
2	http://blog.oxforddictionaries.com/2015/08/words-everyone-spells-wrong/
3	https://quizlet.com/6072175/75-more-difficult-english-words-flash-cards/
4	http://www.ef.com/blog/language/10-most-difficult-words-in-english/
5	http://mentalfloss.com/article/58543/10-words-difficult-remember-meanings
6	http://www.manikjoshi.com/z-manik-joshi-book-ewp-20-ddw.html#.WX7_oUfLc8o
7	https://www.memrise.com/course/383017/difficult-english-words-4/
8	http://www.yougowords.com/10-syllables
9	https://howtheyplay.com/team-sports/Top-10-Most-Difficult-Names-in-Football
10	http://www.listal.com/list/most-difficult-names-spell-andor
11	https://www.buzzfeed.com/robynwilder/ball\protect\kern+.1667em\relaxa-hoolish?utm_term.vyWg1N2V7#.wr5xQYP6E
12	https://www.prdaily.com/Main/Articles/24_complex_wordsand_their_simpler_alternatives_8750.aspx
13	https://www.scribd.com/doc/6879132/List-of-100-difficult-words-that-appear-most-frequently-in-GRE
14	http://destinationfemme.com/100-words-know-expand-vocabulary/
15	https://www.dailywritingtips.com/100-exquisite-adjectives/
16	http://www.hexco.com/what-are-the-top-100-hardest-words-to-spell/
17	http://www.word-buff.com/vocabulary-words.html
18	http://writingexplained.org/synonyms-for-difficult-list
19	http://wordwarriors.wayne.edu/list.php
20	https://www.learnthat.org/word_lists/view/37440
21	https://www.proprofs.com/flashcards/story.php?title=infrequently-used-words-vocabulary-expansion
22	https://justenglish.me/2014/03/10/important-infrequently-used-words-to-know/
	Easy words
1	https://www.vocabulary.com/lists/190098
2	http://www.esldesk.com/vocabulary/basic
3	https://github.com/first20hours/google-10000-english
4	https://www.easypacelearning.com/all-lessons/learning-english-level-1/1090-english-greetings-basic-lesson
5	https://www.easypacelearning.com/all-lessons/learning-english-level-1/691-fruit-vegetables-and-berries-with-pictures-english-lesson-pdf
6	https://www.easypacelearning.com/all-lessons/learning-english-level-1/871-asking-and-answering-english-basics-questions
7	http://www.esldesk.com/vocabulary/basic
8	http://teacherjoe.us/Vocab200.html
9	http://ogden.basic-english.org/word2000.html
10	TextStat python module easy words
11	http://www.home-speech-home.com/multiple-meaning-words.html

Table 3 Variation of three datasets

Parameter	CWIdataset						SemEvalDataset						Both					
	Easy		Complex				Easy		Complex				Easy		Complex			
	Mean	Variance	Mean	Variance			Mean	Variance	Mean	Variance			Mean	Variance	Mean	Variance		
Length	7.164	5.875	8.309	6.266			7.824	7.338	7.985	7.835			7.286	6.386	8.159	7.061		
Syllables	2.330	1.013	2.890	1.171			2.565	1.302	2.641	1.309			2.368	1.074	2.770	1.252		
Synset	2.373	14.00	1.853	2.324			4.902	34.63	2.086	8.342			2.303	13.86	1.957	5.262		

Table 4 Frequency of alphabets

Character	A	B	C	D	E	F	G	H	I	J	K	L	M
Frequency	8.2	1.5	2.8	4.3	12.7	2.2	2.0	6.1	7.0	0.2	0.8	4.0	2.4
Character	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Frequency	6.7	7.5	1.9	0.1	6.0	6.3	9.1	2.8	1.0	2.4	0.2	2.0	0.1

4.2 Syllables

A syllable is simply a stress needed in the word to pronounce clearly (or) a unit of pronunciation having one vowel sound, with or without surrounding consonants, forming the whole or a part of a word. The word with high syllables are treated as complex.

4.3 Rare Character

The percentages of the letters of the alphabet appear in English are listed in Table 4. Based on this, the letters with low frequency will be less used. This paper assumes the words with these characters are less known. From these, this paper assumes the words start with rare character, end with rare character, rare character by length, frequent characters by length, rare character by frequent character, number of rare characters in the word, and number of frequent characters in the words are some of the features.

4.4 Common Bigrams and Trigrams

In most English words, some combinations of character will most frequently appear and they are most frequent bigrams and trigrams.

Common bigrams: TH, EA, OF, TO, IN, IT, IS, BE, AS, AT, SO, WE, HE, BY, OR, ON, DO, IF, ME, MY, UP, SS, EE, TT, FF, LL, MM, and OO

Common Trigrams: THE, EST, FOR, AND, HIS, ENT, and THA

4.5 Parts of Speech

Every word belongs to the some category of parts of speech, which categorizes the words which have similar grammatical properties.

4.6   Continuous Vowels

The words which are having continuous vowels is taken as a feature.

4.7   Synset, Hypernym, and Hyponyms

Wordnet in NLTK(Natural language toolkit), is an English language lexical database. In which, a set of synonyms of the word is called synset, Hypernyms specifies the more general meaning of the word. Hyponyms have more specific meanings of the words.

4.8   Vowels and Consonants

Every word is a combination of vowels and consonants. The ratio of vowels and consonants is taken as one of the features in this paper.

In this paper, classification of complex word is performed using intra-word features. When the target word is given as an input based on the features, the system has to predict the classification.

5   CORDIF Model

This paper discusses the CORDIF model for classifying the target words into complex words and easy words. It is simply a binary classification system.

For this model, CWIdataset was designed which was explained in Sect. 3. The machine learning models are trained using intra-word features and CWIdataset. To optimize the classification, dimensionality reduction as done on the intra-word features is explained in Sect. 6.

Upon the testing set, where the words are annotated by 11 nonnative english speakers was tested with the classifier which produces high accuracy. The results concluded that the complexity varies with person's vocabulary. The model designed for classification is shown in Fig. 1.
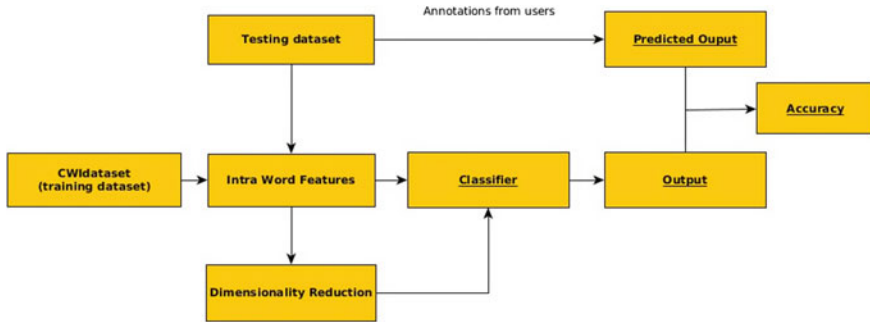
Fig. 1 CORDIF Model

6 Experimental Study

As a case study, to create a system for classification, we used the scikit-learn module [16] for machine learning algorithms and trained them using new dataset called CWIdataset. We have implemented it using Python and NLTK module [17] to extract the intra-word features from the CWIdataset.

We trained supervised learning algorithms with CWIdataset. In Fig. 2a, the *accuracy score, f1score, precision,* and *recall* of the various machine learning algorithms have shown. Among those, Linear Regression model has given high accuracy.

We have done tenfold cross-validation on training data using Random Forest algorithm, we got 82.57% accuracy.

(a) Scores of Machine learning algorithms

(b) Scores of Machine learning algorithms after dimensional reduction
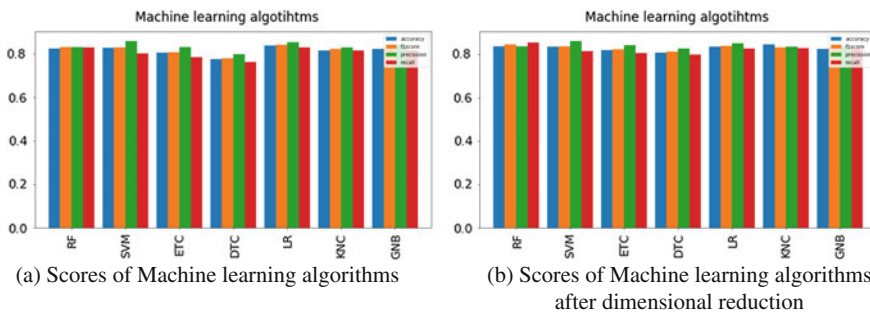
Fig. 2 Scores of Machine learning algorithms before and after dimensionality reduction

6.1 Dimensionality Reduction

Using intra-word features, we have classified the complex word identification by training the machine learning algorithms. we started analyzing to find the best features to improve accuracy. We have implemented Python program by removing each feature and analyzed the results and identified that length, rare character versus length, rare character end, and common bigram words are reducing the accuracy.

In Fig. 2b, after removing those four features the accuracy, F1 score, precision, and recall have increased. The accuracy of Linear Regression model increases to 84.75%, giving high accuracy when comparing with other machine learning models. When we try to do the same with other features the accuracy level is reducing.

Again we have done tenfold cross-validation with existing features and observed that the accuracy has increased to 83.26% accuracy.

We have taken 500 words from the testing set of SemEval testing dataset, which are not in our training data and asked the annotators for whom English is nonnative language, to annotate complex words for them, the words for which they cannot identify the meaning. The annotators were clearly informed about the process. They participated in the task with their consent and as a volunteering activity. We have identified that complex words for every person are not the same. The accuracy of training data is almost the same for every person but there is a high change in the precision, recall, and F1 score scores. In Fig. 3, the scores of individual persons have shown. This results that complex word identification for a person can be done with personalized systems than from combined complex words.
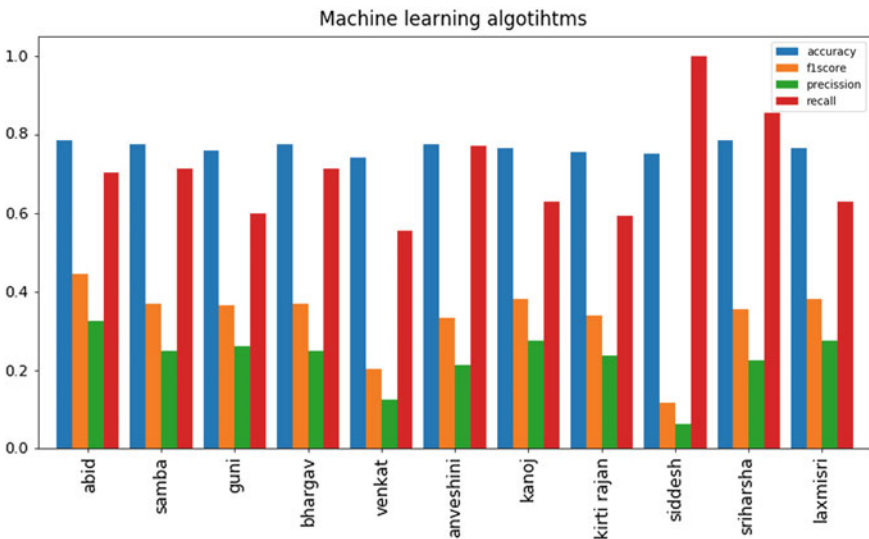
Fig. 3 Scores of LR model for different persons

7   Conclusion

Readability ease measure is nothing but finding a person's level of understandability with respect to a text. The persons with low-literacy level feel difficult to access the text. Many systems have been developed to identify the complex words from the sentences using the features like position of words in the sentence, probability of word within the corpus, frequency of the word in the corpus.

The first training dataset of semantic evaluation workshop is having only 1495 unique words, which is a combination of complex words and easy words. In this paper, we focus on intra-word features that make the target word to classify among the classes complex word and the easy word.

A new dataset, CWIdataset, was created from the websites listed in the paper. The supervised machine learning algorithm is trained with the dataset having 4342 unique words, having 50% each type of words.

The classification of testing data has been done using intra-word features and perform dimensionality reduction upon the features and best features are selected based on accuracy scores of machine learning algorithms. As a testing, we asked annotator to annotate the words by different nonnative persons and observed that the complex words are not unique for all. It is based on their personal vocabulary skills in the language.

Identifying the complex words in the caption files and simplifying the text makes the captions easy and designing a web browser extension to simply the web documents will be done as a future extension.

References

1. Paetzold, G., Specia, L.: SemEval 2016 Task 11: Complex Word Identification. In: SemEval@NAACL-HLT, pp. 560–569 (2016)
2. Shardlow, M.: Out in the Open: Finding and categorising errors in the lexical simplification pipeline. In: LREC, pp. 1583–1590 (2014)
3. Kauchak, D.: Improving text simplification language modeling using unsimplified text data. ACL **1**, 1537–1546 (2013)
4. Paetzold, G.H., Specia, L.: Text simplification as tree transduction. In: Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology, pp. 116–125 (2013)
5. Shardlow, M.: The CW corpus: a new resource for evaluating the identification of complex words. In: Proceedings of the 2nd Workshop on Pre-dicting and Improving Text Readability for Target Reader Populations (2013)
6. De Belder, J., Moens, M.-F.: A dataset for the evaluation of lexical simplification. In: Computational Linguistics and Intelligent Text Processing, pp. 426–437 (2012)
7. Carroll, J.A., Minnen, G., Pearce, D., Canning, Y., Devlin, S., Tait, J.: Simplifying text for language-impaired readers. In: EACL, pp. 269–270 (1999)
8. Devlin, S.: The use of a psycholinguistic database in the simplification of text for aphasic readers. Linguistic databases (1998)
9. Chall, J.S., Dale, E.: Readability Revisited: the new dale-chall readability formula (1995)
10. Ismail, A., Yusof, N.: Readability of ESL picture books in Malaysia. J. Nusant. Stud. (JONUS) **1**(1), 60–70 (2016)

11. Ismail, A., Yusof, N., Yunus, K.: The readability of malaysian english children books: a multilevel analysis. Int. J. Appl. Linguist. Engl. Lit. **5**(6), 214–220 (2016)
12. Karhu, M., Hilera, J., Fernández, L., Ríos, R.: Accessibility and readability of university websites in Finland. J. Access. Des. For All **2**(2), 178–189 (2012)
13. Pantula, M., Kuppusamy, K.S.: A model to measure readability of captions with temporal dimension. In: Proceedings of First International Conference on Smart System, Innovations and Computing
14. Friedman, D.B., Hoffman-Goetz, L.: A systematic review of readability and comprehension instruments used for print and web-based cancer information. Health Educ. Behav. **33**(3), 352–373 (2006)
15. Shepperd, S., Charnock, D., Gann, B.: Helping patients access high quality health information, BMJ **319**(7212), 764–766 (1999)
16. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V.: Scikit-learn: machine learning in Python. J. Mach. Learn. Res. **12**(Oct), 2825–2830 (2011)
17. Bird, S.: NLTK: the natural language toolkit. In: Proceedings of the COLING/ACL on Interactive presentation sessions, pp. 69–72 (2006)

Neural Network Approach for Classification of Human Emotions from EEG Signal

G. S. Shashi Kumar, Niranjana Sampathila and Harikishan Shetty

Abstract Emotions play an important role in human cognition, perception, decision-making, and interaction. In this paper, Neural Network (NN) based system for human emotions classification by extracting features from Electroencephalogram (EEG) signal is proposed. EEG data for the classification of emotions is obtained from the DEAP database. Extracted more than 30 features from EEG and they are used for the emotion classification. Totally, 33 varieties of features are extracted from EEG data. However, there are reports on voice-based, facial-image-based study of expressions to recognize their emotions. However, emotion identification using both methods can be biased as they can be faked. In order to overcome this difficulty, many researchers analyze brain physiological signals to represent the changing patterns during emotional fluctuations. Neural networks have widely been used in emotion classification. Reported here is the classification with the backpropagation artificial neural network. Experimental results have shown an average accuracy above 94.45% is achieved for all the subjects and regions combined.

Keywords Emotion · EEG · Neural network · Backpropagation · Classification

G. S. Shashi Kumar (✉)
Department of Electronics & Communication Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education (MAHE), Manipal 576104, Karnataka, India
e-mail: shashi.gs@manipal.edu

N. Sampathila
Department of Biomedical Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education (MAHE), Manipal 576104, Karnataka, India
e-mail: niranjana.s@manipal.edu

H. Shetty
Manipal Institute of Technology, Manipal Academy of Higher Education (MAHE), Manipal 576104, Karnataka, India
e-mail: harkshan69@gmail.com

1 Introduction

Generally, emotion assessment is done by analyzing the user's emotional expressions and/or physiological signals. The emotional expression of a person refers to the observable verbal and nonverbal behavior that are used to communicate emotion. Emotion assessment studies often focus on analyzing facial expressions, speech, and gestures to identify a person's emotional status. Humans interact specially with speech, but he can add other way like body gestures and looks, to better display its emotions. There has been a lot of research recently on the recognition of emotional information using physiological signals such as blood volume pulse, skin temperature, Electromyogram (EMG), Electrocardiogram (ECG), Galvanic Skin Resistance (GSR), and Electroencephalogram (EEG) [1]. So far limited works on human emotion detection have been reported in the current literature on brain physiological signal.

Normally, there are two means to classify emotional states; the *discrete* approach and the *dimensional* approach. In the discrete approach, emotions are classified by defining emotional states using discrete categories and identifying basic emotions. These basic emotions can be combined to form other emotions. Many different sets of basic emotions were proposed over the years. The very first such model was proposed by Ekman and Friesen [2] with six basic emotions namely *anger, disgust, fear, happiness, surprise, and sadness*.

Dimensional approach is an alternative way of representing different emotions. Russell proposed a two-dimensional bipolar model using Valence (V) and Arousal (A) parameters for the classification of emotional states [3]. 2D model is widespread and most popularly used for emotion classification. It can discover the basic emotions in its space along with many other emotions which could be without any discrete emotional labels. The quadrant-wise description of modified Valance-Arousal model using Russell's emotion model [4] is shown in Fig. 1.

Electroencephalography is a noninvasive signal acquisition method that captures the behavioral activities of the brain. Deducing emotional states from EEG has received considerable attention as EEG directly reflects the changes in the brain in response to a stimulus and thus, can detect the changes in emotional status [2].

2 Related Work

During the past couple of decades, facial expressions were widely used to recognize emotions. Silva and Hui [5] developed a real-time facial emotion recognition system to distinguish four emotions namely happy, sad, surprise, and anger. They used neural network classifier for the classification.

Punitha and Geetha [6] extracted textural features from the images taken from a facial expression database to classify happy, disgust, and surprise emotions. They used Support Vector Machine (SVM) classifier and obtained a classification accuracy
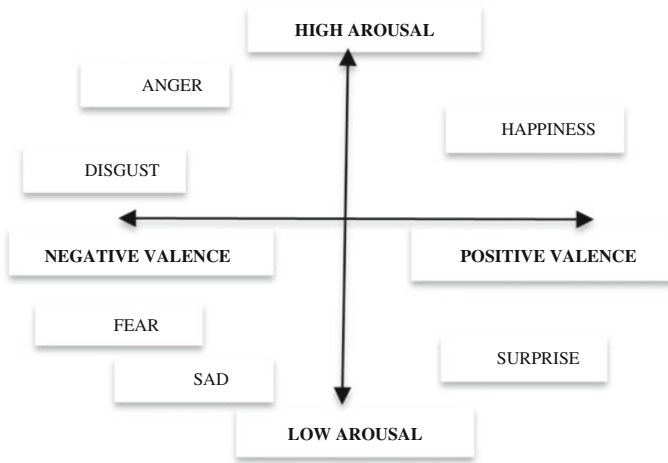
Fig. 1 Valance-arousal-dimensional model for basic emotions [3]

of 87%. However, emotion identification using facial expressions can be biased as they can be faked. In order to overcome this hurdle, many researchers chose to analyze physiological signals to represent the changing patterns during emotional fluctuations.

Katsis et al. [7] designed wearable system to identify the emotional condition of race car drivers. They attempted an automated approach for emotion recognition using several biosignals. The wearable system included sensors to measure facial EMG, respiration, ECG, and skin conductance. SVM classifier was used to differentiate emotional states such as "high stress", "low stress", "disappointment", and "euphoria" experienced by the drivers.

Soleymani et al. [8] attempted classification of emotion from EEG, where they took EEG recordings from 24 participants and by using linear discriminant criterion, they classified three emotions in valence and arousal plane individually with an accuracy of 68.5% and 76.4%, respectively. Chanel et al. [9] used fusion technique to combine the collected peripheral signals along with EEG and achieved an accuracy of 63% classification.

Frantzidis et al. [10] used Mahalanobis classifier to classify four emotions induced from pictures selected from International Affective Picture System (IAPS). Meza-Kubo et al. [11] used neural networks to classify "pleasant" and "unpleasant" emotions in older adults. EEG signals were recorded when the participants were exposed to visual stimuli using affective pictures from IAPS. Using Fast Fourier Transform, the signals were split into alpha, beta, and theta sub-bands from 14 electrodes. Features such as magnitude of the signal, maximum frequency, and spectral power were extracted from each sub-band and were fed to the designed neural network.

Work done by Oh et al. [12] showed that Hjorth parameters extracted from EEG signals can be highly useful features, as they have information in both time and frequency domain. Jenke et al. [13] in his review of effective features for the classification of emotions using EEG suggested that, features such as Hjorth parameters, fractal dimension, band power, and entropies of the different frequency bands gave good results. Petrantonakis and Hdjileneontiadis [14] used Higher Order Crossings (HOC) to classify Ekman's six basic emotions. They selected 3 channels: Frontopolar (Fp1), Mid-frontal (F3), Mid-frontal (F4), Frontopolar (Fp2). They also compared HOC-based features with statistical features and wavelet-based features. HOC showed promising results in comparison to other features extracted from EEG. They reported on Quadratic Discriminant Analysis (QDA) and SVM classifiers.

Frantzidis et al. [10] used C4.5 decision tree algorithm with Mahalanobis distance to classify four emotions based on 2D valence-arousal model. Images from IAPS were shown to 28 participants and their EEG responses were recorded. Event-related potentials and event-related synchronization/desynchronization computation was carried out to precisely track the changes in local neuronal activity. A success rate of 77.68% was achieved to classify the emotions residing in four different quadrants of the valence-arousal model.

Filho et al. [15] attempted to classify 2 emotional states namely "calm" and "stress" defined by some threshold values of Valence and Arousal. K-Nearest Neighbor (KNN) classifier was used to analyze and compare three feature extraction techniques.

Chung and Yoon [16] used Bayes' classifier and supervisory learning technique to classify emotions. Along with 32 EEG channels present in DEAP dataset, they added 61 additional virtual channels by bipolar montage and hence they derived 93 channels. They extracted power spectral density from different frequency bands for all the channels. Accuracies of classification of valence and arousal when two affective states were considered was 66.6% and 66.4%, respectively.

Cristian et al. [17] used Hidden Markov Model for emotion classification. They considered grouping of other physiological signals available in the database along with EEG. They divided arousal and valence into 2 groups named "LOW" (values from 1–5 in the scale) and "HIGH" (values from 5–9 in the scale). 75% accuracy was obtained for arousal classification when the combination of EEG along with physiological signals (heart rate, skin resistance, temperature) was used. A highest accuracy of 54% for the classification of valence was obtained when only EEG signals were considered. These results showed that emotion recognition can be achieved using multimodal approach as well.

3   Methodology

The methodology adopted for the classification of emotions is described by extracting features from the EEG signals of the subjects. The dataset is obtained from the online data source called Database for Emotion Analysis using Physiological (DEAP) sig-

nals. After acquiring permission from the authors, the dataset was downloaded from their website [18]. The aim of the authors in creating the dataset was to explore the possibility of analyzing and classifying human affective states, which are induced as a response to music video stimulus [19]. DEAP is a multimodal dataset containing peripheral physiological signals as well as EEG recordings, which were collected from 32 subjects (17 male and 15 female) while they were viewing 40 1-minute long excerpts of music videos. EEG was recorded at a sampling rate of 512 Hz using 32 active AgCl electrodes placed according to the international 10–20 system. The peripheral nervous system signals recorded were GSR, respiration amplitude, skin temperature, ECG, blood volume by plethysmograph, electromyograms of Zygomaticus and Trapezius muscles, and electrooculogram. For 22 of the 32 participants, facial expressions were also recorded when they were watching the videos. Participants later rated each video in terms of the level of Valence, Arousal, Dominance (all 3 in continuous scale of 1–9), Familiarity, and Like/Dislike (latter 2 in discrete scale of 1–5) with the help of Self-assessment Manikins (SAM) [20].

3.1 EEG-Based Emotion Classification

Happiness and Sad emotions were classified using EEG signal. The procedure for the recognition of the said emotions is shown in Fig. 2.
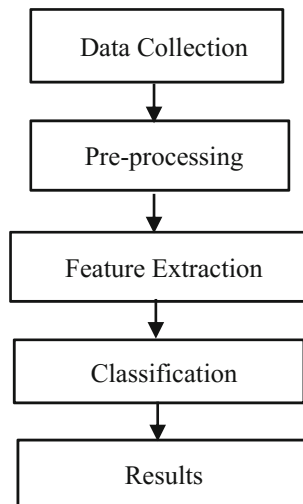
Fig. 2 Flow diagram for classification of emotion using EEG

3.2   Data Collection

Crucial task to elicit emotional reactions is selection of effective stimulus. The Valence-Arousal plane (V-A plane) was divided into 4 quadrants. 15 stimuli were selected for each quadrant from both manual and affective tags [10]. These selected 120 videos were then subjected to online volunteer rating, where an average of 15 people rated each of the 120 videos in terms of their level of valence and arousal. Normalized score for valence and arousal was calculated by taking the ratio of the mean and standard deviation of the public rating of every video. Then for each quadrant in the normalized V-A space, 10 videos that lie closest to the extreme corners of the quadrant were selected [11].

From quadrant-1 (high valance-high arousal) and quadrant-3 (low valence-low arousal), 1 video stimuli out of 40 stimuli were selected for its emotional content analysis. The video stimulus in the first quadrant is expected to elicit a highly positive emotion and the stimulus music video from third quadrant a very negative emotion. The assumption made here is that the positive emotion is "happy" and the negative emotion is "sad".

3.3   Preprocessing of EEG

Preprocessing of data is very much essential for improving signal-to-noise ratio of EEG data. Eye blinks, facial and neck muscle activities, and body movement are the major EEG artifacts. To address these artifacts, preprocessing of EEG signals was done by the authors of the database before making it publically accessible [21]. Following are the preprocessing steps taken:

- EEG was common average referenced and downsampled from 512 to 256 Hz and high-pass filtered with a 2 Hz cut-off frequency using EEGLAB software.
- Eye blink artifacts were removed from the signal using blind source separation technique.
- A baseline of 5 s before each video was recorded. The power of the baseline was calculated to nullify the stimulus unrelated variations in power over time. Frequency power of trials and baselines of 3–47 Hz frequency range was extracted with Welch's method with windows of 256 samples. The baseline power was then subtracted from the EEG signal power; yielding the change of power relative to pre-stimulus period.
- Muscle and movement artifacts are eliminated by further downsampling the EEG signal to 128 Hz and a passband filter of 4–45 Hz was applied [10].

EEG signal of a subject obtained from the database is shown in Fig. 3.
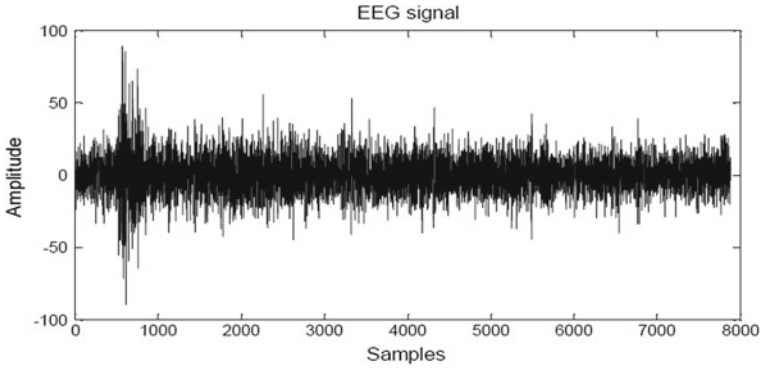
Fig. 3 EEG signal of a participant

3.4 Feature Extraction

Out of the 32 electrodes available, not all electrode channels were used for the classification of emotions in order to save computation time. The following electrodes were selected for emotion classification—Frontopolar (Fp1), Mid-frontal (F3), Mid-frontal (F4), Frontopolar (Fp2), Occipital (O1), Occipital (O2), Mid-temporal (T7), Mid-temporal (T8), Parietal (P3), and Parietal (P4) [4]. The selected electrodes represent all the lobes of the brain.

From the selected ten channels, features are extracted for both *happy* and *sad* emotions from all the 32 subjects. The feature vector designed is composed of both time and frequency domain features. Following are the important features extracted from EEG for emotion classification.

- Time domain features—Mean, Hjorth parameters, and higher order crossings.
- Frequency domain features—Bandpower and entropy.

Mean: It is the average value of the signal. It is calculated as shown in Eq. (1) for a signal $x(t)$.

$$Mean = \frac{1}{T} \sum_{t=1}^{T} x(t) \tag{1}$$

where T is the total length of the signal.

Hjorth parameters: Hjorth parameters describe the general characteristics of an EEG signal in few quantitative terms. The parameters are

Activity
The activity parameter represents the variance of a time signal. It is calculated as shown in Eq. (2) for a signal *x(t)*

$$Activity = \frac{1}{T} \sum_{t=1}^{T} (x(t) - \mu)^2 \tag{2}$$

where μ is the mean of the signal.

Mobility

The mobility parameter is an estimate of the mean frequency. It is determined as the square root of ration of variance of the first derivative of the signal to the signal. It is calculated as shown in Eq. (3).

$$Mobility = \sqrt{\frac{var(\dot{x}(t))}{var(x(t))}} \tag{3}$$

where $\dot{x}(t)$ is the first derivative of the signal $x(t)$ and var(x) is the variance of the signal.

Complexity

The complexity parameter represents the change in frequency. The parameter compares the similarity of the signal with a pure sine wave and if the signal is more similar, the value converges to 1 [22]. Complexity parameter is expressed in Eq. (4).

$$Complexity = \sqrt{\frac{Mobility(\dot{x}(t))}{Mobility(x(t))}} \tag{4}$$

These parameters together categorize EEG patterns in terms of amplitude, time scale and randomness.

Higher Order Crossings (HOC)

Zero crossings of a signal are the count of number of times the signal crosses the horizontal x-axis. Applying a specific filter to a time series, changes the oscillations of the signal and hence, its zero crossing counts as well. So, when a specific sequence of filters is applied to a time series, the corresponding zero crossing counts obtained is called HOC sequence [14, 23].

Algorithm to obtain HOC sequence is as follows:

1. The number of zero crossings for the original signal is counted.
2. The original signal is filtered and the number of zero crossings is counted.
3. Step-2 is repeated "z" times.

Here, sequential high-pass filters are applied to the signal by differentiating the signals sequentially and HOC is obtained. HOC reveals the oscillatory patterns of EEG signal and provides a feature set that conveys the emotional information to the classifier. The value of "z" which determines the length of the HOC sequence is determined by iterative process.
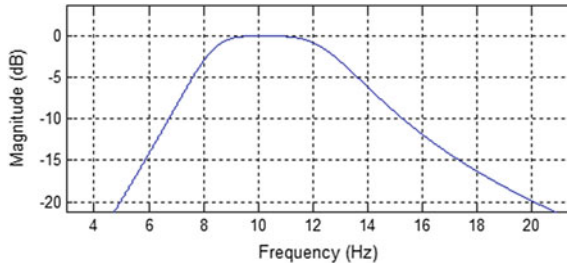
Fig. 4 Magnitude response of designed passband filter to obtain alpha frequency band

Splitting of EEG Signals into Rhythmic Bands
EEG signal was band-pass filtered to obtain different brain rhythms to extract features from them. EEG can be either filtered with Finite Impulse Response (FIR) or Infinite Impulse Response (IIR) digital filters. FIR filters are known to have linear phase response, while IIR filters are characterized by their nonlinear phase response. For a given set of filter specifications, IIR filters show small amount of sidebands than FIR filters, hence, where application demands a sharper cut-off, IIR filters are mainly used. For obtaining the different brain rhythms, IIR filters were used as they provide sharper cut-offs. IIR filters can be implemented as Butterworth, Chebyshev, and Elliptic filters. In electrophysiology, when phase disturbance is not an issue, commonly Butterworth and elliptic IIR filters are applied. IIR Butterwoth filters have no passband and stopband ripple. They have the shallowest roll-off near the cut-off frequency compared to other IIR filters like Chebyshev and Elliptic. Filter cut-offs can be made sharper by using higher order of the filter. To obtain different brain rhythms, IIR butterworth filter of order 4 was used. The magnitude plot of the butterworth filter was designed to get alpha band (8–13 Hz) and is shown in Fig. 4.

Similarly different passband filters were designed to split EEG into theta, beta, and gamma frequency bands to extract features from them. Different brain rhythms obtained from the signal in Fig. 3 after filtering are as shown in Fig. 5a–d.

Bandpower
From the band-filtered signals, power is calculated for each of the bands using Eq. 5.

$$Power = \frac{1}{N} \sum_{n=0}^{N} (x(n))^2 \tag{5}$$

where N is the length of the sequence and $x(n)$ is the time series data.

Entropy
Entropy of a signal can be defined as the amount of information content in the signal. Entropy can be used to analyze the complexity of the signal as it measures the randomness of the signal. Here, histogram method is used to compute the empirical
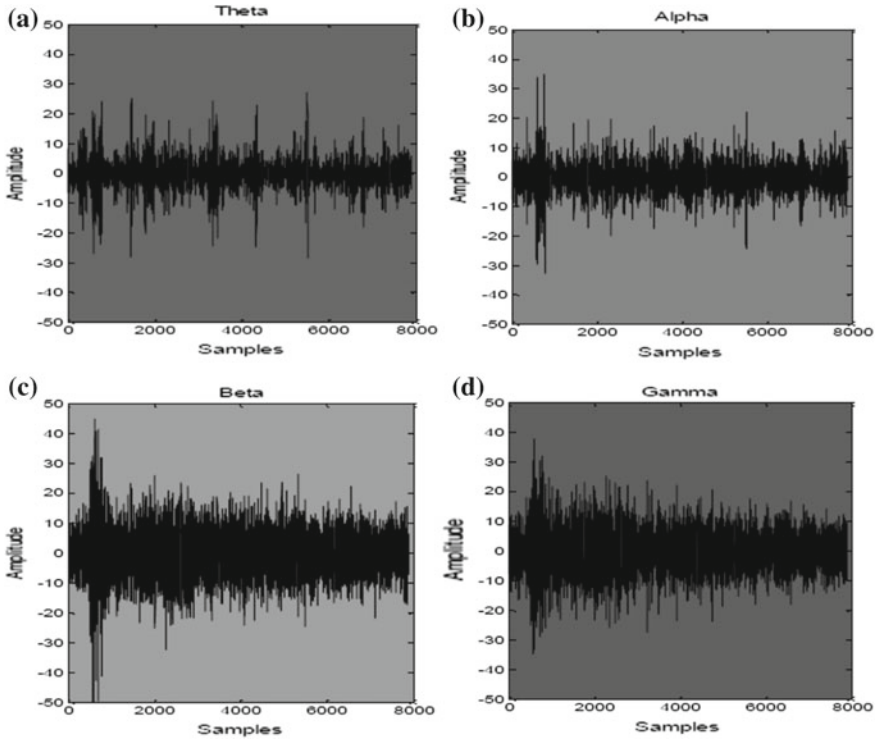
Fig. 5 EEG bands obtained after filtering: **a** theta, **b** alpha, **c** beta, **d** gamma

probability density function. If histogram has M bins, represents the probability of each bin [24]. Using Eq. 6 entropies of different bands were calculated.

$$Entropy = -\sum_{t=1}^{M} p_1 \log_2 p_t \tag{6}$$

3.5 Classification

Backpropagation Neural Network classifier used in this work to distinguish the first and third quadrant emotions using EEG signals.

Backpropagation Neural Network (BPNN)
Artificial Neural Networks (ANN) are statistical learning models inspired by biological neural networks and use a set of input neurons and hidden neurons which are repetitively weighted and tuned to activate output neurons for achieving an intended

task. BPNN is a supervised learning algorithm implemented in ANN where error propagates backwards from output neurons to input neurons.

The neural network toolbox in MATLAB software is used to create the artificial neural network [25]. The input dataset containing the extracted features are divided into three partitions namely training data, validation data, and testing data. The training data is used to train the neural network. During the training phase, the neural network is provided with a set of input feature vector along with the respective target vectors to distinguish the output classes. After the network is trained, the validation set is used to validate the trained network before testing it with the testing data. The network designed for 66 features is considered as input to the network, there are 66 input neurons. Total number of hidden neurons selected is 4 and the output of the network has 2 output neurons for a two class problem. The overall accuracy for training, testing, and validation is then calculated by averaging the individual accuracies. The classifier performance [26, 27] is analyzed by studying various error rates.

4 Results and Discussion

The results of classification of "happy" and "sad" emotion from EEG data and the performances of the designed classifier for the extracted feature vector is discussed as follows: To classify the first and third quadrant emotions (happy and sad respectively) features were extracted from the EEG signals. The feature vector used for the classification of "happy" and "sad" emotion include: Mean, Hojrath parameters, HOC, Bandpower, and Entropy.

A comparative study was done for the classification of the said emotions based on the electrode location and gender. Electrodes residing on different brain lobes were grouped to compare the performances of individual lobes in differentiating emotions. The lobes were frontal, parietal, temporal, and occipital. The electrodes representing each of the lobes are shown in Table 1. Features extracted from the electrodes representing a lobe were combined together before giving it as input to the classifier. Combination of all the electrodes was also considered to represent the entire brain region and the extracted features from each electrode were combined.

Table 1 Electrodes representing each lobe

Lobes	Electrodes
Frontal	Fp1, Fp2, F3, F4
Parietal	P3, P4
Temporal	T7, T8
Occipital	O1, O2
All	All 10 electrodes

Fig. 6 a Confusion matrix
for frontal region considering
all subjects. **b** Confusion
matrix for occipital region
considering all subjects
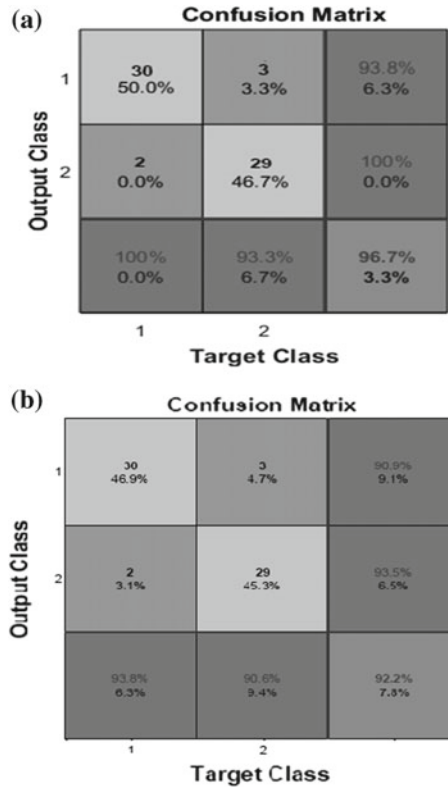
BPNN Classifier

The electrodes in a particular region were combined and features were extracted from them. The feature vectors representing each brain regions were then split into training, validation and testing sets before feeding them to the neural network classifier to identify *happy* and *sad* emotions. The partitioning of data was done four times with different ratios. The number of hidden layers is optimized. The overall average accuracy of the neural network classifier, when all the 32 subjects are considered and accuracy details are calculated. Among the other regions, Frontal and occipital region showed an accuracy of 96.7% and 92%, respectively. Figure 6a, b represents the confusion matrix for the classification of happy and sad emotion for frontal region and occipital region.

5 Conclusion

In this work to identify emotions, features extracted from EEG signals were used to represent the physiological changes in the subjects. Mean of the signal, HOC, Hjorth features were extracted along with entropy and bandpower from each bands to represent the characteristics of the EEG signals for differentiating two emotions namely *happy and sad*. Feature vectors of different brain regions were extracted to observe the significance of brain regions in emotion classification. The feature vectors were classified using Backpropagation Neural Network classifier. The developed neural network classifier has demonstrated an average accuracy above 94.45%.

References

1. Frijda, N.H.: Laws of emotion. J. Am. Psychol. Assoc. **43**(5), 349–358 (1998)
2. Ekman, P., Friesen, W.V.: Constants across cultures in the face and emotion. J. Personal. Soc. Psychol. **17**(2), 124–129 (1971)
3. Russell, J.A.: Affective space is bipolar. J. Pers. Soc. Psychol. **37**(3), 345–356 (1979)
4. Russell, J.A.: A circumplex model of affect. J. Pers. Soc. Psychol. **39**(6), 1161–1178 (1980)
5. Silva, L.D., Hui, S.C.: Real-time facial feature extraction and emotion recognition. In: 4th International Conference on Information, Communications and Signal Processing, pp. 1310–14. Massey University, New Zealand (2003)
6. Punitha, A., Geetha, M.K.: Texture based emotion recognition from facial expressions using support vector machines. Int. J. Comput. Appl. **80**(5), 1–5 (2013)
7. Katsis, C.D., Katertsidis, N., Ganaiatsas, G., Fotiadis, D.I.: Toward emotion recognition in car-racing drivers: a biosignal processing approach. IEEE Trans. Syst. Man Cybern. Part A Syst. Hum. **38**(3), 502–512 (2008)
8. Soleymani, M., Pantic, P., Pun, T.: Multimodal emotion recognition in response to videos. IEEE Trans. Affect. Comput. **3**(2), 211–223 (2012)
9. Chanel, G., Rebetez, C., Bétrancourt, M., Pun, T.: Emotion assessment from physiological signals for adaptation of game difficulty. IEEE Trans. Syst. Man Cybern. Part A Syst. Hum. **41**(6), 1052–1063 (2011)
10. Frantzidis, C.A., Bratsas, C., Klados, M.A., Konstantinidis, E., Lithari, C.D.: On the classification of emotional biosignals evoked while viewing affective pictures: an integrated data-mining-based approach for healthcare applications. IEEE Trans. Inf. Technol. Biomed. **14**(2), 309–318 (2010)
11. Meza-Kubo, V., Moran, A.L., Carrillo, I., Garcia-Canseco, E.: Assessing the user experience of older adults using a neural network trained to recognize emotions from brain signals. J. Biomed. Inf. **62**, 202–209 (2016)
12. Oh, S.H., Lee, Y.R., Kim, H.N.: A novel EEG feature extraction method using Hjorth parameter. Int. J. Electron. Electron. Eng. **2**(2), 106–110 (2014)
13. Jenke, R., Peer, A., Buss, M.: Feature extraction and selection for emotion recognition from EEG. IEEE Trans. Affect. Comput. **5**(3), 327–339 (2014)
14. Petrantonakis, P.C., Hadjileontiadis, L.J.: Emotion recognition from EEG using higher order crossings. IEEE Trans. Inf. Technol. Biomed. **14**(2), 186–197 (2010)

15. Filho, T.F.B., Ferreira, A., Atencio, A,C., Arjunan, A., Kumar, D.: Evaluation of feature extraction techniques in emotional state recognition. In: 4th International Conference on Intelligent Human Computer Interaction, pp. 1–6. Kharagpur, India (2012)

16. Chung, S.Y., Yoon, H.J.: Affective classification using Bayesian classifier and supervised learning. In: 12th International Conference on Control, Automation and Systems, JeJu Island, Korea, pp. 1768–1771 (2012)

17. Cristian, A., Valencia, T., Hernan, F., Arias, G., Mauricio, A.: Comparative analysis of physiological signals and EEG for multimodal emotion recognition using generative models. In: XIX Symposium on Image, Signal Processing and Artificial Vision, Armenia, pp. 1–5 (2014)

18. Koelstra, S., Muehl, C., Soleymani, M., Lee, J.S., Yazdani, A., et al.: Deap: a database for emotion analysis using physiological signals. IEEE Trans. Affect. Comput. **3**(1), 18–31 (2012)

19. DEAP dataset. http://www.eecs.qmul.ac.uk/mmv/datasets/deap/index.html (2015). Accessed 20 Nov 2015

20. Bradley, M.M., Lang, P.J.: Measuring emotion: the self-assessment manikin and the semantic differential. J. Behav. Ther. Exp. Psych. **25**(1), 49–59 (1994)

21. Alotaiby, T., Abd El-Samie, F.E., Alshebeili, S.A.: A review of channel selection algorithms for EEG signal processing. J. Adv. Signal Process. **66**, 1–21 (2015)

22. Hjorth, B.: EEG analysis based on time domain properties. Electroencephalogr. Clin. Neurophysiol. **29**, 306–310 (1970)

23. Petrantonakis, P.C., Hadjileontiadis, L.J.: Emotion recognition from brain signals using hybrid adaptive filtering and higher order crossings analysis. IEEE Trans. Inf. Affect. Comput. **1**(2), 81–97 (2010)

24. Li, P., Liu, C., Li, K., Zheng, D., Liu C: Assessing the complexity of short-term heartbeat interval series by distribution entropy. Med. Biol. Eng. Comput. **53**(1), 77–87 (2015)

25. Neural Network toolbox, www.in.mathworks.com/help/release/R2015a/pdf_doc/nnet_ug.pdf

26. Upadhyaya, P., Muralidhar Bairy, G., Vidya, S.: Electroencephalograms analysis and classification of neurological and psychological disorders. J. Med. Imag. Health Inf. **5**, 1–5. https://doi.org/10.1166/jmihi.2015.1545 (2015)

27. Mangalagowri, S.G., Cyril Prasanna Raj, P.: EEG feature extraction and classification using feed forward backpropagation algorithm for emotion detection. In: IEEE International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT), pp. 183–187 (2016)

Frequency Measurement of Resonator for Vibrating Gyroscope

M. Gopala Krishnamurthy, D. Dinakar, I. M. Chhabra, P. Kishore,
N. V. N. Rao Pasalapudi and K. C. Das

Abstract A novel and low-cost noncontact solution is realized for measuring the resonator frequency of vibration by using a sensitive microphone. This hemispherical resonator is operated at a specific mode (mode 2), which is useful for the Coriolis vibratory gyroscope applications. This mode of vibration is excited in the resonator by using electrostatic force. The experiments were conducted to measure vibrational frequency of the hemispherical metallic resonator by using the proposed method. It is observed that the experimental results are in good agreement with the simulated results.

Keywords Gyroscope · Noncontact · Mode-2 vibration · Hemispherical resonator · Accelerometer · LDV · Microphone

1 Introduction

Gyroscope is an instrument widely used in the navigation systems for sensing inertial angular motion about its input axis without any external reference. Over the past few decades, a host of physical laws were used to develop operational gyroscopes, which are currently in operation [1]. Development of Coriolis Vibratory Gyroscope (CVG) addresses all the aerospace applications needs [2]. In this type of gyroscopes, a vibrating hemisphere is excited to one of its resonant vibration mode called 'mode 2' at the fixed amplitude. When the resonator rotates about a particular fixed axis of a body, a Coriolis force acts on the vibrating resonator element and excites another resonant mode. The rate of exchange of energy can be measured in terms of change of amplitude and phase of the signals at different locations around the hemispherical resonator [3].

M. Gopala Krishnamurthy · I. M. Chhabra · N. V. N. Rao Pasalapudi · K. C. Das
Research Center Imarat, Hyderabad, TS, India

D. Dinakar (✉) · P. Kishore
National Institute of Technology, Warangal, Warangal, TS, India
e-mail: dinakar.anu@gmail.com

In the present work, an electrostatic forcing scheme is being used to establish wave pattern on to the resonator at its resonance frequency. For very precise and noncontact measurement of frequency, Laser Doppler Vibrometer (LDV) is generally used. The laser beam from LDV is directed toward the surface of resonator, by using Doppler shift, the amplitude and frequency are measured. The LDV setup requires careful alignment of laser beam and vibrating portion of the resonator. Although LDV gives a very good measurement accuracy, it is an expensive equipment. In this article, an alternate method is demonstrated to measure the frequency of a resonator by using sensitive microphone. The experimental results are in good agreement with the simulated results.

2 Working Principle

In a CVG, mode-2 vibration also called driving mode is forced electrostatic field in the hemispherical resonator at fixed amplitude [4]. When the platform on which CVG mounted rotates about a particular fixed axis, the Coriolis forces acting on the resonator excites a different resonant mode in another axis also called sensing mode [5]. The rate at which the energy is transferred between the driving mode and sensing mode is a measure of the rotation rate of the platform about that axis. The arctangent of the ratio of the both mode amplitudes change in response to the input rotation rate. In most cases, the driving mode frequency and sensing mode frequencies are identical. Examples are vibrating string, hemispherical shell, and in the Foucault pendulum [5].

In CVG, the driven and the readout mode amplitude satisfies coupled oscillator equations with Coriolis effect [6–8]. The equations are either same as two-dimensional oscillator or can be transformed into desired two-dimensional oscillator equations. When the thin-walled hemispherical resonator is forced to 'mode 2' vibration, the vibration pattern looks as shown in the Fig. 1 at the equatorial plane of the hemisphere [9]. In the first half cycle, the resonator deforms to its greatest ellipsoidal geometry and then returns to its spherical shape. In the next half cycle, a similar deformation takes place but spatially shifted by 90° in azimuth. This mode of vibration leads to four antinodes (A, B, C, D) with maximum displacement of resonator lip of the shell and four nodes (E, F, G, H) with zero displacement is shown in Fig. 1.

3 Resonator Design and Simulation

A thin wall of 1 mm aluminium hemispherical resonator with 12 g mass and 30 mm diameter is designed by using SOLIDWORKS®CAD Software with high geometrical tolerances of $\pm 5\,\mu$m sphero city and CNC machinability. The dimensions of the designed metallic hemispherical resonator are as shown in Fig. 2a. Nearly 6 mm-
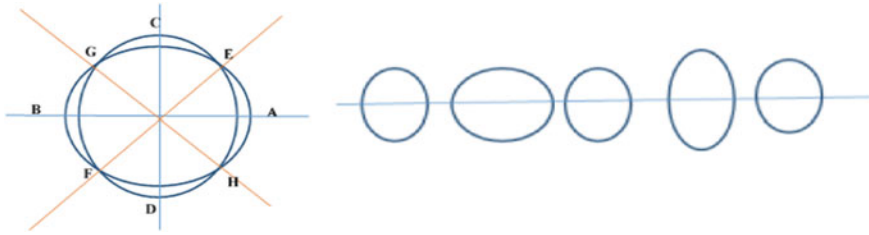
Fig. 1 The mode pattern of the resonator
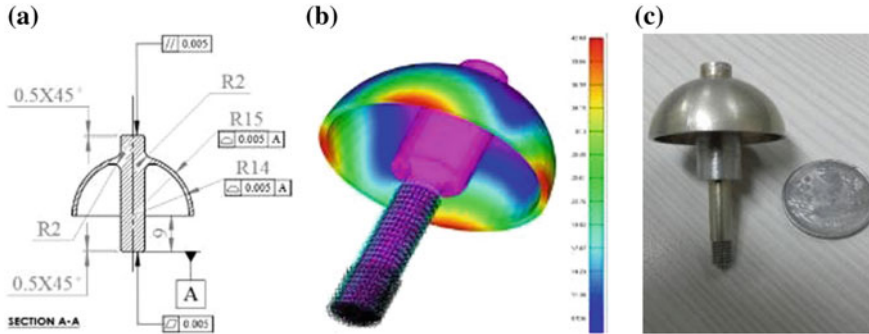
Fig. 2 Designed vibration sensor shell **a** dimensions, **b** simulated model and **c** photograph

thick cylindrical central stem holds the hemisphere firmly to eliminate other modes of vibration when it is forced. Finite Element map (FEMAP) with NASTRAN simulation software is used to analyze and understand the various modes of vibration, on the resonator and their frequencies. Simulation helped to fix the frequency of mode 2 vibration (as shown in Fig. 2b) and various possible cross coupling of other modes during the forcing. The selection of material, fixing method of resonator stem are some more critical parameters in tuning the frequency of resonator. A resonator with very high dimensional accuracy is realized as shown in Fig. 2c with the designed frequency of 6259 Hz.

4 Experimental Results and Discussion

During the mode-2 vibration, resonator's equatorial plane only participates in the displacement. The contribution of the resonator's mass in vibration is around 3 g. Initially, a sensitive accelerometer is used to measure the frequency of the resonator. During the experiment, it is observed that the mass of the accelerometer (appr. 1 g) is loading the resonator, leads to incorrect measurement of the frequency. Hence, a non-contact method of frequency measurement is required to characterize the resonator and frequency measurement.
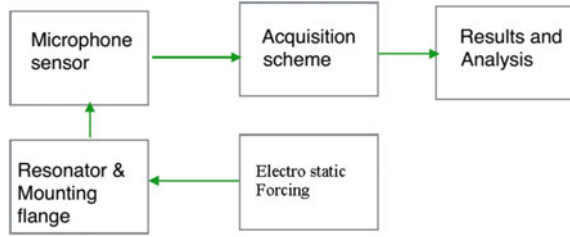
The experimental setup block diagram to measure the frequency is as shown in Fig. 3. The resonator is fixed on to the base firmly. Forcing electronics generates the electrostatic forcing pulse to excite the resonators to its mode 2 vibration. Mode 2 vibration is confirmed by the pick off electronics at the nodes and antinodes locations. When the resonator vibrates, an audible sound is heard from the resonator.

A calibrated GRAS makes 40 AQ random-incidence microphone with IEC 61,094 standard in the frequency range of 3–6 kHz with sensitivity of 50 mV/Pa being used to detect the frequency of vibrating resonator. Microphone is fixed very close to the resonator to pick the sound from it in anechoic chamber. When the forcing pulse is applied on the electrode, the hemispherical resonator generates audible frequency from it.

DEWE 4011 Data Acquisition System is used to acquire the signal from the microphone sensor at 50 k samples/s. The analog signal from the sensor with appropriate pre amplifier is fed to the inbuilt 16 bit ADC. The acquisition software have been used to display the real time data from the sensor for further analysis. The temporal response and frequency response of the sensor are shown in Figs. 4 and 5, respectively. The temporal response of the sensor clearly illustrates the damping vibration for applied force pulse at resonance frequency. The amplitude of the vibration suddenly reduced and gradually deceases to zero (low Q-factor of the material). Frequency response plot of the sensor shows that most of the vibration energy is at the frequency of 6279 Hz which is closely matched with the simulated frequency of 6259 Hz.

5 Conclusion

A novel and low-cost method is realized to measure vibrational frequency of hemispherical metallic resonator at specific mode with high accuracy by using high-sensitive microphone. The obtained experimental resonator frequency results are well matched with the simulated results. This sensor is less weight, miniature size and low cost. The operating frequency can be altered by changing the various parameter of the sensor, because the accuracy of the gyroscope is highly dependent on the frequency. This sensor is useful for gyroscopic applications to measure rotation rate of the satellites and aerospace vehicles.
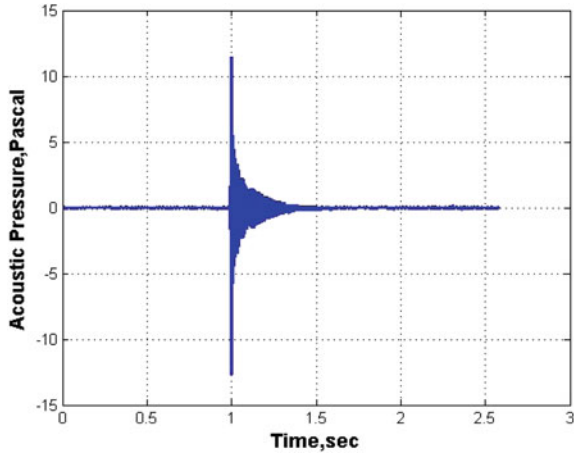
Fig. 4 Temporal response of the sensor using microphone
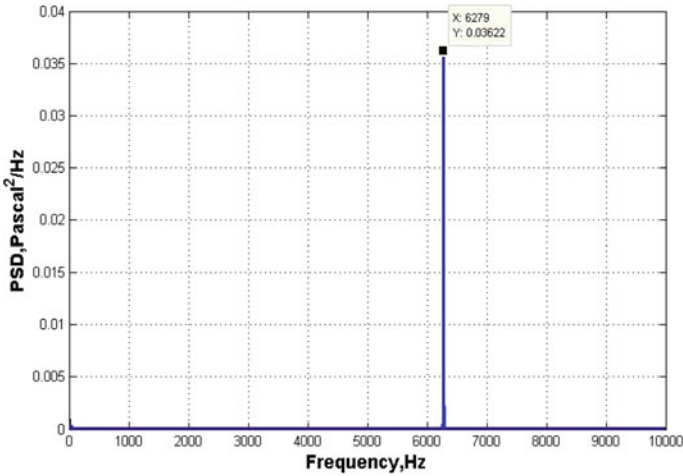
Fig. 5 Frequency response of the sensor shell

References

1. Dzhandzhgava, G.I., Bakhonin, K.A., Vinogradov, G.M., Trebukhov, A.V.: Strapdown inertial navigation system based on a hemispherical resonator gyro. Gyroscopy Navigat. **1**(2), 9197 (2010)
2. Matthews, A., Rybak, F.J.: Comparison of hemispherical resonator gyro and optical gyros. IEEE Aerosp. Electron. Syst. Mag. **l**7(5), 40–46 (1992)
3. Rozelle, D.M.: The hemispherical resonator gyro: from wineglass to the planets. In: Proceedings of 19th AAS/AIAA Space Flight Mech. Meeting, pp. 1157–1178 (2009)
4. Matthews, A., Varty, G.T., Li, C.M., Lynch, D.D.: Vibratory rotation sensor with whole angle tracking. U.S. Patenet 5,801,310 (1998)
5. Lynch, D.D., Savaya, R.R., Campanile, J.J.: Hemispherical resonator gyro control. U.S. Patent No. 2006/0248,953 (2006)
6. Lynch, D., Matthews, A., G.T.V arty, Innovative mechanization to optimize inertial sensors for high or low rate operations. In: Proceedings of Symposium Gyro Technology, vol. 6, pp. 16–17. Stuttgart, Germany (1997)
7. Roy Choudhury, D., Jain, S.B.: Linear Integrated Circuits. New Age Science Limited (2011)
8. Matthews, A., Varty, G.T., Nguyen, P.C., Lynch, D.D.: Extraction of double-oscillation-frequency sinusoids in vibratory rotation sensors. U.S. Patent 6,158,282 (2000)
9. Bose, Amitava, Puri, Somnath, Banerjee, Parithosh: Modern Inertial Sensors and Systems. PHI Learning Pvt. Ltd., First Edition, India (2008)

Known-Plaintext Attack on Cryptosystem Based on Fractional Hartley Transform Using Particle Swarm Optimization Algorithm

Phool Singh, A. K. Yadav and Kehar Singh

Abstract Known-plaintext attack based on particle swarm optimization algorithm is mounted successfully on an image encryption scheme that uses single-fractional Hartley transform. The results point out that the proposed algorithm successfully retrieved the fractional orders of the transform, which forms the secret keys. Robustness of the algorithm is tested on noisy and missing data encrypted images. The secret keys are successfully retrieved in the presence of Gaussian noise of considerable strength. This study brings out the success of particle swarm optimization algorithm in mounting known-plaintext attack on an optical image encryption scheme in fractional Hartley domain.

Keywords Known-plaintext attack · Particle swarm optimization · Fractional Hartley transform

1 Introduction

Optical image encryption approaches for information security are an emerging next-generation technology. High processing speed, large key space, parallel-processing capability, computer simulation of optical realization are a few advantages of optical image encryption over the digital image encryption. Double random phase encoding (DRPE) is a well-known image encryption algorithm developed by Refregier and

P. Singh (✉) · K. Singh
Department of Applied Sciences, The NorthCap University,
Sector 23A, Gurugram, Haryana, India
e-mail: phool.singh24@gmail.com

P. Singh
Department of Mathematics, Avvaiyar Government College
for Women, Karaikal, Puducherry, India

A. K. Yadav
Department of Applied Mathematics, Amity School of Applied Sciences, Amity University,
Gurgaon, Haryana, India

Javidi [1] in 1995. It uses random phase masks in the spatial as well as the Fourier domains, to encrypt an input image. Researchers have shown that the double random phase encoding is vulnerable to some of the basic attacks like known-plaintext attack (KPA) [2], chosen-plaintext attack (CPA) [3], ciphertext-only attack (COA) [4]. To enhance the security of the DRPE scheme by introducing additional encryption parameters, several other transforms have been attempted in the literature, namely fractional Hartley transform (FRHT) [5–7], gyrator transform [8], fractional Fourier transform [9], etc. Security of the schemes vis-à-vis these additional parameters have been investigated by researchers. Qin and Peng [10] discussed KPA on fractional Fourier transform and reported that fractional Fourier transform is also vulnerable. Fractional Hartley transform (FRHT) has been used in many encryption algorithms, comprising grayscale images [6, 11–13], double images [14], and color images [15]. No study on cryptanalysis of image encryption schemes based on FRHT is reported in the literature so far.

Motivated by the social behavior of swarm (fish or bird), Kennedy and Eberhart in 1995 [16] developed a stochastic optimization technique known as Particle Swarm Optimization (PSO) which is population based. The PSO algorithm is easy to implement as compared to other evolutionary algorithms such as evolutionary programming, genetic algorithm, and evolution strategies. Like genetic algorithm, PSO algorithm is also initialized with random solutions called population, and make iterative search for the optimal solution. In particle swarm optimization, solutions called particles follow the current iteration's best solution. It is successfully applied to artificial neural network training, fuzzy system control, function optimization, and many other areas, where genetic algorithm could be applied. PSO algorithm has the following advantages [17]: (1) It is easier to implement (2) Fewer parameters to regulate; (3) Effective memory capability as each particle recollects its previous best; (3) Efficient in maintaining diversity of the swarm as each particle uses information connected to the most successful particle. In the field of optical image processing, Sang et al. [18] successfully applied PSO algorithm in obtaining the parameter of gyrator transform. They showed that the recovery of the parameter of a single gyrator transform is much faster with improved particle swarm optimization algorithm as compared to exhaustive search method (brute force attack).

In this paper, we show how a single-fractional Hartley transform-based scheme, which provides a larger key space as against gyrator transform, is susceptible to known-plaintext attack. This paper is organized as follows: in Sect. 2, fractional Hartley transform and particle swarm optimization are briefly introduced, and the encryption scheme with its validation is presented. Section 3 presents PSO-based KPA algorithm to retrieve the secret keys, that is, the orders of the fractional Hartley transform. The final conclusions are presented in Sect. 4.

2 Related Background

2.1 Fractional Hartley Transform

Fractional Hartley transform (FRHT) with fractional order α and β, of the plaintext $I(x, y)$ is defined [6] as

$$
\begin{aligned}
H^{\alpha,\beta}(u, v) = {} & \frac{\sqrt{(1 - i \cot \theta_1)(1 - i \cot \theta_2)}}{2\pi} \exp\left[i\pi \left(\frac{u^2 \cot \theta_1}{\lambda g_{s1}} + \frac{v^2 \cot \theta_2}{\lambda g_{s2}} \right) \right] \\
& \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(\frac{i\pi x^2 \cot \theta_1}{\lambda g_{s1}} + \frac{i\pi y^2 \cot \theta_2}{\lambda g_{s2}} \right) \left\{ \frac{1 - i \exp[i(\theta_1 + \theta_2)/2]}{2} \right. \\
& \times cas\left(\frac{ux \csc \theta_1}{\lambda g_{s1}} + \frac{vy \csc \theta_2}{\lambda g_{s2}} \right) + \frac{1 + i \exp[i(\theta_1 + \theta_2)/2]}{2} \\
& \left. \times cas\left(-\frac{ux \csc \theta_1}{\lambda g_{s1}} - \frac{vy \csc \theta_2}{\lambda g_{s2}} \right) \right\} I(x, y) dx dy \quad (1)
\end{aligned}
$$

where g_{s1} and g_{s2} are focal lengths of lenses, respectively, in the x and y directions, λ is the wavelength of the input light when implemented optically, $cas = cos + sin$, $\theta_1 = \alpha \pi / 2$ and $\theta_2 = \beta \pi / 2$.

The two-dimensional FRHT can be written in terms of the fractional Fourier transform (FrFT) as follows [6]:

$$
H^{\alpha,\beta}(u, v) = \frac{1 + exp[i(\theta_1 + \theta_2)/2]}{2} F^{\alpha,\beta}(u, v) + \frac{1 - exp[i(\theta_1 + \theta_2)/2]}{2} F^{\alpha,\beta}(-u, -v). \quad (2)
$$

2.2 The Scheme

In the optical encryption process, a host image/plaintext $I(x, y)$, is bonded with a random phase mask (RPM), and placed on the input plane. It is then subjected to the fractional Fourier transform with orders (p, q), resulting in an encrypted image $En(x, y)$, as shown in Fig. 1a. The decryption process (Fig. 1b) consists of applying fractional Hartley transform of order $(-p, -q)$ on the encrypted image and taking the amplitude part of the resulting image to recover the host image. The scheme has been validated with the input grayscale images (Lena and Boy Fig. 2a, b) of size 256×256 pixels, by performing simulation on MATLAB 9.3 (2017b). Application of fractional Hartley transform of order $p = 1.5678$ and $q = 1.7342$ on the input images produces encrypted images, whose real and imaginary parts are shown in Figs. 2c–f. The corresponding decrypted images obtained from algorithm in Fig. 1b are shown in Fig. 2g, h.
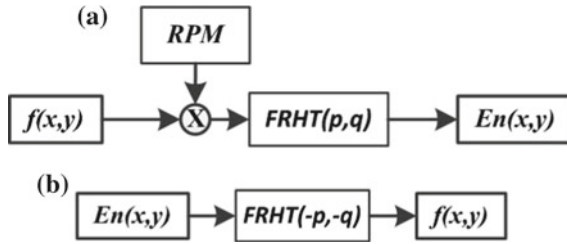
Fig. 1 **a** Encryption and **b** decryption scheme based on a single FRHT
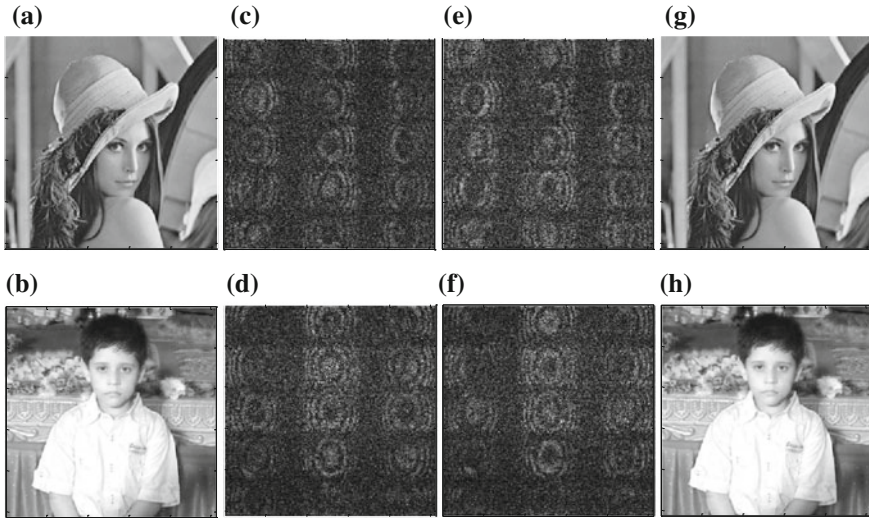
Fig. 2 **a**, **b** Input images; **c**, **d** and **e**, **f** are, respectively, the real and imaginary parts of the encrypted images; **g**, **h** are the recovered images

2.3 Particle Swarm Optimization Algorithm

Motivated by the schooling and flocking patterns of swarms (fish and birds), Particle Swarm Optimization (PSO) was developed by Kennedy and Eberhart in 1995 [16]. They simulated bird's swarm behavior by particles in a search space. Each particle has its peculiar velocity and position, and represents a potential solution. The PSO algorithm is initiated with random values of particles. On the basis of value of the fitness function, the position and the velocity of every particle are updated on every iteration of PSO. At each iteration, values of variables get adjusted according to the value that is nearest to the target. Consider a group of birds hovering over a region, where they can feel the odour of an out-of-sight source of foodstuff. The one which is nearest to the foodstuff chirps the loudest and the other birds start following in the direction of loudest sound. If any of the other circling birds approaches near

to the target as compared to the first, it chirps louder than first and the others go around over toward it. This converging pattern toward the target continues until one of the birds gets the foodstuff. This algorithm is simple to comprehend and easy in its implementation. After finding the two best positions, one local best (pbest) and another global best (gbest), the velocity v and the position p of each particle are updated by the formulae as in [19]:

$$v[\text{updated}] = w * v[\text{current}] + c_1 * \text{rand}() * (\text{pbest} - p[\text{current}])$$
$$+ c_2 * \text{rand}() * (\text{gbest} - p[\text{current}])$$
$$p[\text{updated}] = p[\text{current}] + v[\text{updated}] \tag{3}$$

where w is inertia weight, c_1 and c_2 are learning factors and rand() is a random number between (0, 1).

The following values of PSO parameters are considered in our simulations: maximum number of iterations is 100, fitness function is the mean-squared-error between the plaintext and the recovered image, range for FRHT orders is (0–2), inertia weight w is 0.8, and the learning factors c_1 and c_2 are taken as 1.4 and 2 respectively.

3 Known-Plaintext Attack Based on PSO

In this section, we look at how the FRHT-based image encryption scheme (Fig. 1a) can be susceptible to the known-plaintext attack (KPA). In KPA, it is anticipated that attacker knows the encryption scheme and tries to recover the encryption keys from the plaintext–ciphertext pair. The fractional orders of FRHT are additional secret keys for encryption which the attacker would attempt to recover. Herein, we present KPA on a scheme based on a single-fractional Hartley transform, using particle swarm optimization. A brute force attack would pose a serious challenge to the attacker in recovering the encryption keys provided by the orders (p, q) of the FRHT. However, due to linearity of FRHT, it is reasonable to believe that the fractional Hartley transform has a security flaw.

PSO-based KPA algorithm uses the plaintext–ciphertext pair to retrieve the secret keys corresponding to the encryption scheme Fig. 1a. Mean-squared-error (MSE) is considered as the fitness function for the PSO algorithm and is given by the formula as follows:

$$\text{MSE} = \frac{1}{N \times N} \sum_{x=1}^{N} \sum_{y=1}^{N} |I_o(x, y) - I_r(x, y)|^2 \tag{4}$$
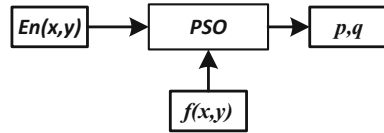
Fig. 3 Proposed PSO-based KPA

Table 1 Retrieved fractional Hartley transform orders (p, q) using PSO algorithm, and correlation coefficient between the plaintext and the decrypted image using retrieved keys

No. of iterations	Fractional Hartley transform orders (p, q)	Correlation coefficient
1	(1.5409, 2.2486)	0.163
10	(1.5649, 2.2452)	0.280
20	(1.5667, 1.7276)	0.512
30	(1.5669, 1.7344)	0.804
40	(1.5681, 1.7341)	0.958
50	(1.5678, 1.7341)	0.999
60	(1.5678, 1.7342)	1
70	(1.5678, 1.7342)	1
80	(1.5678, 1.7342)	1
90	(1.5678, 1.7342)	1
100	(1.5678, 1.7342)	1

Here, $I_r(x, y)$ are the pixels values of the recovered image obtained through the retrieved orders of FRHT from PSO algorithm and $I_o(x, y)$ are those of the original image. A flow diagram of the proposed KPA is shown in Fig. 3.

Table 1 shows the retrieved fractional Hartley transform orders (p, q) using particle swarm optimization after 1, 10, 20, ..., 100 iterations. Another statistical metric used in our analysis is correlation coefficient (CC), which is defined in terms of covariance (cov) and standard deviation (σ) as

$$CC = \frac{\text{cov}(I_o(x, y), I_r(x, y))}{\sigma(I_o(x, y))\sigma(I_r(x, y))}. \tag{5}$$

The correlation coefficient between the input and the recovered images, as a result PSO algorithm for 1, 10, 20, ..., 100 iterations is also displayed in Table 1. Figure 4 shows recovered images of Lena obtained by using the proposed KPA algorithm after 10, 30 and 50 iterations. Results indicate that the image is recognizable after 30 iterations of PSO algorithm with correlation coefficient more than 0.804. The drop in MSE for grayscale images as seen from the curves Fig. 5 plotted of CC (for Lena and Boy) against number of iterations of PSO algorithm. After 30 iterations, the PSO algorithm shows faster convergence as indicated by a rapid fall in the MSE (Fig. 5).

(a) **(b)** **(c)**
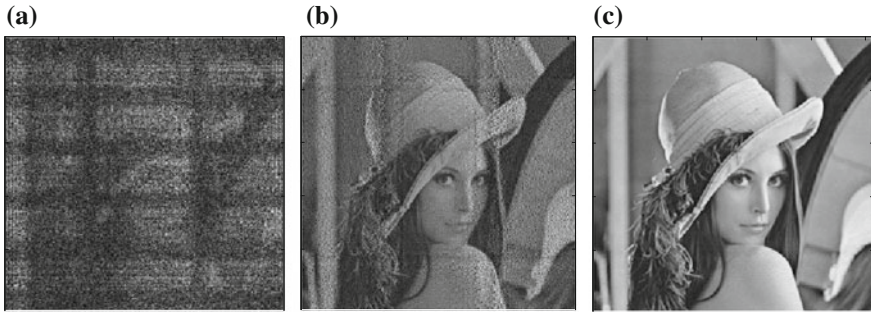
Fig. 4 **a–c** Recovered images by using retrieved FRHT orders from the proposed KPA algorithm after 10, 30, and 50 iterations, respectively
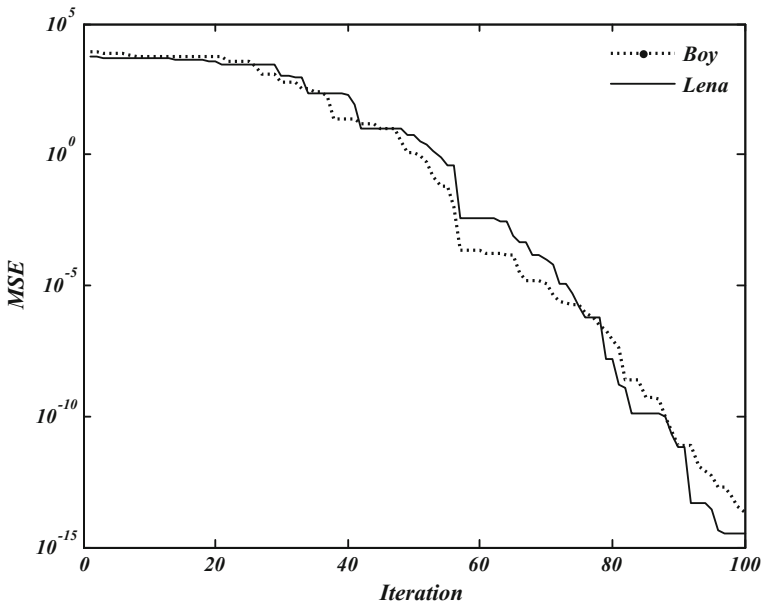
Fig. 5 Plot of MSE curves versus number of iterations of PSO algorithm

3.1 Robustness of the Algorithm in Presence of Noise in the Encrypted Image

We have tested our algorithm on noisy encrypted images E'. Gaussian noise (G) of strength (k) with mean zero and standard deviation unity is added to the image En according to the formula [20, 21]:

$$E' = En(1 + kG). \tag{6}$$
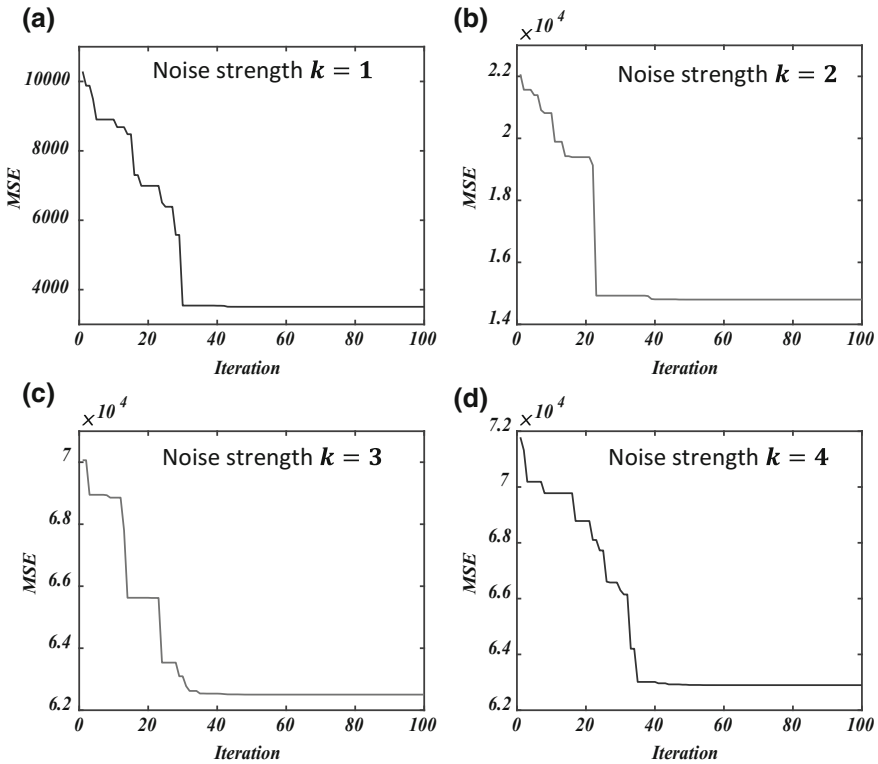
Fig. 6 Plots of MSE versus number of iterations of PSO algorithm for noise strength $k = 1, 2, 3, 4$, respectively, in (**a**–**d**)

Results indicate that the algorithm can successfully retrieve the encryption keys (FRHT orders), from a highly noise-affected encrypted image, with noise strength as high as $k = 4$. Plots of MSE versus number of iterations for different noise strength have been shown in Fig. 6. For noise strength $k = 1$ (Fig. 6a), the minimum value of MSE achieved by the algorithm is 3470. This high value of MSE is due to the noise present in the encrypted image. Similarly, high values of MSE achieved by the PSO-based KPA algorithm in plots Fig. 6b, c, d, is due to higher noise strength $k = 2, 3, 4$ respectively. Figure 7 shows convergence of our algorithm in presence of different noise strengths for retrieving FRHT orders. It is observed from the plots that irrespective of extent of noise present in the encrypted image, our algorithm successfully retrieves the secret keys within 40 iterations.
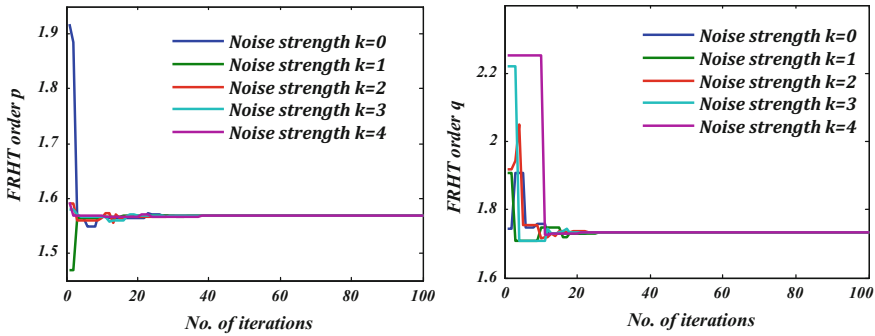
Fig. 7 Plots of convergence of our algorithm for FRHT orders (p, q). Curves correspond to different noise strength k = 0, 1, 2, 3 and 4
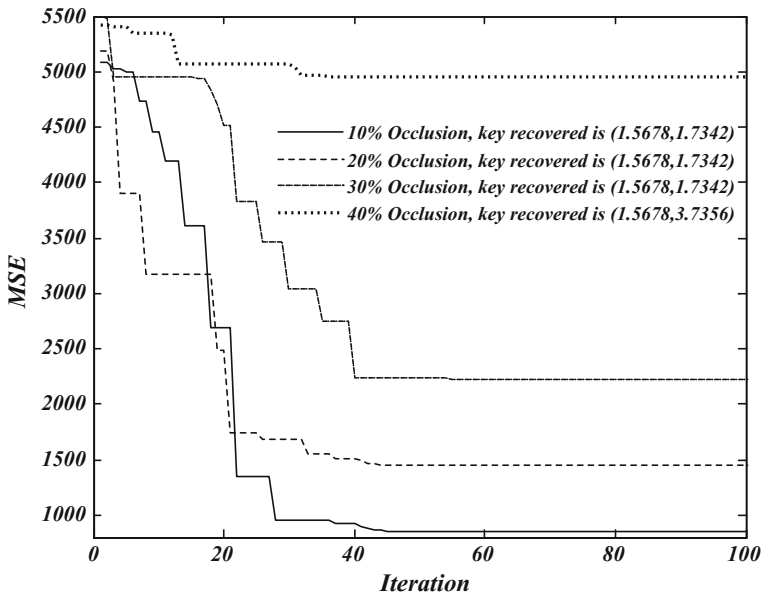
Fig. 8 Plot of MSE curves with no. of iterations for occlusion attack

3.2 Robustness Against Missing Data of the Encrypted Image

We have also performed an attack wherein we have shown that the algorithm can retrieve the FRHT orders with some missing encrypted data (occlusion attack). The encrypted image En has lost 10, 20, 30 and 40% data, and the corresponding MSE curves with retrieved secret keys (FRHT orders) are shown in Fig. 8 for Lena. Results show that up to 30% occlusion, the proposed algorithm successfully retrieved the secret keys, the orders of FRHT, i.e., $p = 1.5678$ and $q = 1.7342$. Although MSE

of the recovered images improves with the number of iterations of PSO, it does not change after 45 iterations.

4 Conclusions

Particle swarm optimization algorithm-based known-plaintext attack is successfully mounted on an image encryption scheme that uses single-fractional Hartley transform. The results reveal that the proposed algorithm successfully retrieved the secret keys that are the fractional orders of the transform for the grayscale images. The algorithm is tested for its robustness for noisy and occluded images. Simulations show that irrespective of the extent of noise present in the encrypted image, our algorithm successfully retrieved the secret keys. Our algorithm also shows successful retrieval of the secret keys, for occluded encrypted image up to 30% occlusion.

References

1. Refregier, P., Javidi, B.: Optical image encryption based on input plane and Fourier plane random encoding. Opt. Lett. **20**, 767–769 (1995)
2. Peng, X., Zhang, P., Wei, H., Yu, B.: Known-plaintext attack on optical encryption based on double random phase keys. Opt. Lett. **31**, 1044–1046 (2006)
3. Carnicer, A., Montes-Usategui, M., Arcos, S., Juvells, I.: Vulnerability to chosen-cyphertext attacks of optical encryption schemes based on double random phase keys. Opt. Lett. **30**, 1644–1646 (2005)
4. Jiao, S., Li, G., Zhou, C., Zou, W., Li, X.: Special ciphertext-only attack to double random phase encryption by plaintext shifting with speckle correlation. JOSA A. **35**, A1–A6 (2018)
5. Pei, S.-C., Tseng, C.-C., Yeh, M.-H., Shyu, J.-J.: Discrete fractional Hartley and Fourier transforms. IEEE Trans. Circuits Syst. II: Analog Digit. Signal Process. **45**, 665–675 (1998)
6. Zhao, D., Li, X., Chen, L.: Optical image encryption with redefined fractional Hartley transform. Opt. Commun. **281**, 5326–5329 (2008)
7. Singh, P., Yadav, A., Singh, K.: Color image encryption using affine transform in fractional Hartley domain. Opt. Appl. **47**, 421–433 (2017)
8. Yao, L., Yuan, C., Qiang, J., Feng, S., Nie, S.: An asymmetric color image encryption method by using deduced gyrator transform. Opt. Lasers Eng. **89**, 72–79 (2017)
9. Unnikrishnan, G., Joseph, J., Singh, K.: Optical encryption by double-random phase encoding in the fractional Fourier domain. Opt. Lett. **25**, 887–889 (2000)
10. Qin, W., Peng, X.: Vulnerability to known-plaintext attack of optical encryption schemes based on two fractional Fourier transform order keys and double random phase keys. J. Opt. Pure Appl. Opt. **11**, 075402 (2009)
11. Yadav, A.K., Singh, P., Singh, K.: Cryptosystem based on devil's vortex Fresnel lens in the fractional Hartley domain. J. Opt. (2017). https://doi.org/10.1007/s12596-017-0435-9

12. Singh, P., Yadav, A.K., Singh, K., Saini, I.: Optical image encryption in the fractional Hartley domain, using Arnold transform and singular value decomposition. AIP Conf. Proc. **1802**, 020017 (2017)
13. Singh, P., Yadav, A.K., Singh, K.: Phase image encryption in the fractional Hartley domain using Arnold transform and singular value decomposition. Opt. Lasers Eng. **91**, 187–195 (2017)
14. Vilardy, J.M., Torres, C.O., Jimenez, C.J.: Double image encryption method using the Arnold transform in the fractional Hartley domain. In: Proceedings of SPIE 8785, 8th Iberoamerican Optics Meeting and 11th Latin American Meeting on Optics, Lasers, and Applications, vol. 87851R, pp. 87851R-5 (2013)
15. Singh, P., Yadav, A.K., Singh, K.: Color image encryption using affine transform in fractional Hartley domain. Opt. Appl. **47**, 421–433 (2017)
16. Kennedy, J., Eberhart, R.: Particle swarm optimization. In: 1995 Proceedings of IEEE International Conference on Neural Networks (1995)
17. Li, J.: An optimized watermarking scheme using an encrypted gyrator transform computer generated hologram based on particle swarm optimization. Opt. Express **22**, 10002 (2014)
18. Sang, J., Zhao, J., Xiang, Z., Cai, B., Xiang, H.: Security analysis of image encryption based on gyrator transform by searching the rotation angle with improved pso algorithm. Sensors **15**, 19199–19211 (2015)
19. Poli, R., Kennedy, J., Blackwell, T.: Particle swarm optimization: An overview. Swarm Intell. **1**, 33–57 (2007)
20. Sharma, N., Saini, I., Yadav, A., Singh, P.: Phase-image encryption based on 3D-Lorenz chaotic system and double random phase encoding. 3D Res. 8, 1–17 (2017)
21. Anjana, S., Saini, I., Singh, P., Yadav, A.K.: Asymmetric Cryptosystem Using Affine Transform in Fourier Domain. In: Bhattacharyya, S., Chaki, N., Konar, D., Chakraborty, U.K., Singh, C.T., (eds.) Advanced Computational and Communication Paradigms. Springer, Singapore (2018)

DUCA: An Approach to Elongate the Lifetime of Wireless Sensor Nodes

S. B. Bore Gowda and G. Nayak Subramanya

Abstract Clustering the wireless sensor network (WSN) is a persuasive methodology for increasing the sensor nodes' lifetime. The clustering algorithms normally selects nodes as a cluster head only if the node has abundant remaining energy and quick energy depletion in the nodes is balanced through rotating the cluster heads at regular intervals. Mostly, the WSN is divided into small areas of identical size called clusters, which lead to non-uniform energy loads among the heads of the cluster and member nodes of the cluster. The head nodes of the cluster closer to the sink have a larger burden of the energy load. In order to prolong the premature node failure, it is very much essential to manage the energy load of the nodes. We propose a Distributed Unequal Clustering Algorithm (DUCA), which addresses hotspot issues and imbalanced energy consumption. In the proposed algorithm, the energy load balancing is dealt with by making the cluster size small which is near to sink against the nodes farthest from the sink. For inter-cluster communication and direct routing with the sink, the multi-hop routing technique is implemented. The behavior of the proposed protocol is simulated against the LEACH protocol. The simulation results prove that the proposed unequal clustering method balances uniform depletion of energy in the network and also effectively addresses the hot spot issues. The proposed DUCA enhances the greater improvement in the lifespan of the WSN.

Keywords Clustering · Energy · Routing · Lifetime · Energy efficiency
Wireless sensor network

1 Introduction

The technical advancements in the field of communication and sensing lead to the developments of devices with communication capability to monitor continuously

S. B. Bore Gowda (✉) · G. Nayak Subramanya
Department of E & C Engineering, Manipal Institute of Technology,
Manipal Academy of Higher Education, Manipal 576104, Karnataka, India
e-mail: bore.sb@manipal.edu

some physical phenomena results in the evolution of WSNs. WSN as a promising domain of networked systems contains a huge number of low-powered wireless nodes with certain constraints in CPU and an on-chip memory employed to gather the certain physical attributes from the environment. The wireless nodes are tiny, less expensive, multi-operational devices deployed in the sensor field. The environment being sensed by the sensor nodes shapes the data through the processor to convert it to a more readable form and also transmit the sensed attribute to the neighbouring nodes or to the head nodes. The sensor nodes communicate with the nodes situated at different distances using the wireless channel. The WSNs are primarily employed to take participation to perform assigned jobs, for example, administering the environs, safety monitoring, analysis of the battleground, security and health care.

The WSN is one of the promising technologies that can be used for gathering information. The WSN is capable of creating a universal environment suitable for distant sensing, observing and regulating a media or environmental state. The goodness of WSN is quite evident that it is capable of realizing adequate tracking of events occurring at a faraway locality that may or may not be inaccessible.

2   Literature Review

The data dissemination in WSN is primarily dependent of the kind of routing paths used to transmit the data to the intended destination from the source of information. The WSNs are highly constrained in energy, processing and communication. Many challenges raised during the routing process are addressed by plenty of the proposed algorithms in the literature to manage the many challenges raised during routing. Basically, the routing protocols for WSN are grouped depending on two factors: network structure and network operation. In a hierarchical network topology, certain factors and specific preconditions play important roles in the organization of sensor nodes in the field into clusters. Different tasks are being executed by different nodes in the WSNs [1].

Clustering is a process of organizing the sensing devices called nodes into a hierarchical structure, which exists in their relative vicinity. The lifespan of the nodes is boosted by the well-known scheme called clustering. Basically, the clustered WSNs consist of two kinds of heterogeneous or homogenous nodes in each individual cluster namely: cluster head and member nodes of the cluster. The member nodes of the cluster sense the data from the surrounding sensing area at a specified interval and finally, forward it to the cluster head (CH). Depending on the type of the application, the data aggregation of collected one from the members of the cluster is performed by the cluster heads. The aggregated data is forwarded to the sink either by single-hop or multi-hop mode of communication technique.

Hybrid Energy-Efficient Distributed Clustering (HEED) [2] is one of the most popular power conservative protocol. HEED is an example of a hierarchical clustering protocol in which member nodes and cluster heads communicate by single-hop mode, whereas between cluster heads and sink employs a multi-hop mode of com-

munication. The basic factors used for selection of cluster heads are each node's remaining battery power and the energy expense involved in intra-cluster communication.

One of the distributed hierarchical clustering algorithms, which employs k-hop communication for elongating the lifetime of the WSN is an energy-efficient hierarchical clustering (EEHC) [3]. The EEHC protocol operation begins as follows: all sensor nodes with probability "*p*" are chosen as cluster heads. These nodes broadcast their selection to the nearby nodes that fall into their communication range. The cluster heads selected this manner are termed as the volunteer cluster heads.

The PEGASIS [4] is a communication protocol for WSN. The nodes form the chain, which is initiated by the farthest node. In PEGASIS, nodes communicate only with adjacent neighbouring nodes. The node gets a chance to become a cluster head sequentially for data communications to the sink. Some of the features of PEGASIS are: random node placement, performing data fusion and the ability of the nodes to detect the vital data, supporting the wireless media of communication.

TEEN [5] is another protocol, which employs hierarchical-and cluster-based routing. TEEN forms small clusters by grouping the deployed sensor nodes. The clusters are managed by a particular cluster head. The member nodes send their sensed data to the respective cluster heads. The aggregated data is routed to a next level cluster head for the sink.

The most popular protocol in WSNs, Low-Energy Adaptive Clustering Hierarchy (LEACH) [6] is a hierarchical routing protocol. The succeeding cluster and routing protocols for WSN are inspired by LEACH. The primary goal of LEACH in the clustering process is that it rotates the selection of sensor nodes as cluster heads periodically. The criteria used for the election of cluster head depend on: set threshold and node should not be a cluster head in any of the previous rounds, i.e. the lifetime of the network. The threshold values set for the election of cluster head is based on the nodes probability. The randomly elected cluster heads form the clusters by broadcasting the information through advertisement message to the neighbouring nodes. In order to ensure the almost uniform distribution of energy load, the protocol selects the eligible node as new cluster heads at each and every round.

3 Proposed DUCA Protocol

3.1 Energy Model for Sensor Nodes

A simple radio energy hardware model is considered to calculate the energy expenditure to execute the receive/transmit electronics [6]. The radio energy model for the WSN is given in Fig. 1. The proposed protocol is simulated either by considering free space (fs) or multi-path (mp) fading model depending upon the remoteness of the receiver and transmitter. For the less threshold distances, the free space model is suitable and for higher threshold distances multi-path model is used [6].
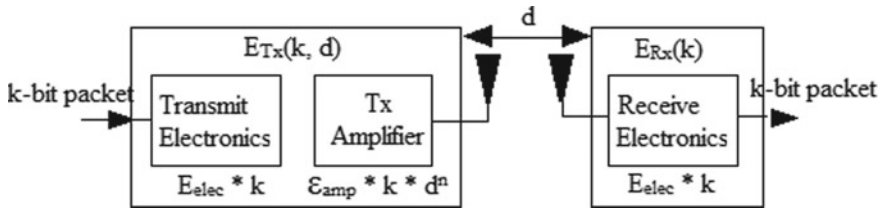
Fig. 1 Typical radio energy model of the sensor nodes [6]

If the distance from the transmitter to receiver node is d, the amount of power dissipated to send a message of length k-bit is computed as follows:

$$E_{Tx}(k; d) = E_{Tx\_elec}(k) + E_{Tx\_amp}(k; d) \tag{1}$$

where E_{Tx_amp} and E_{Tx_elec} are the energy dissipation incurred to run power amplifier and transmit electronics, respectively.

Based on the distance from the transmitter to the receiver, the transmitter energy dissipation can be controlled by setting the power amplifier. The radio energy model to compute the power dissipation of multi-path fading (d^4 power loss) and free space (d^2 power loss) channel are given as follows:

$$E_{Tx}(k, d) = \begin{cases} k E_{elec} + k \in_{fs} d^2, d < d_0 \\ k E_{elec} + k \in_{mp} d^4, d \geq d_0 \end{cases} \tag{2}$$

The energy expended by the receiving sensor node to receive a message of length k-bit is computed as follows:

$$E_{Rx}(k) = E_{Rx\_elec}(k) = k E_{elec} \tag{3}$$

$$d_0 = \sqrt{\frac{\in_{fs}}{\in_{mp}}} \tag{4}$$

where E_{elec} is the electronics energy dissipation and it depends on the following factors: filtering type, digital coding, spreading of the signal, modulation technique. The acceptable bit error rate (BER) and the separation between the receiver and the transmitter decide the amplifier energy dissipation, $\in_{fs} d^2$ or $\in_{mp} d^4$. The threshold transmit distance for the amplification circuit is represented by the symbol d_0.

3.2 Methodology of the Proposed Protocol

The proposed protocol employs competent distributed approach for maximizing the lifespan of the network. The clusters are created such that the energy load distribution is uniform across the deployed sensor nodes. To choose a node as a cluster head, the proposed protocol considers the residual energy of the node as the primary criteria. This helps in network scalability and data compression to decrease the quantity of sensed information to be transmitted to the main collecting centre, i.e. base station. Large amount of energy is conserved by employing data fusion locally and multi-hop transmission. The TDMA type of access technique minimizes the during inter and intra-cluster communication. These strategies employed in the proposed algorithm enhance the energy conservation by a greater extent, thereby maximizing the network life time.

The proposed DUCA protocol operation is classified into four different phases: Cluster setup phase, Cluster head advertisement phase, Scheduling phase and Data transmission phase.

Cluster setup phase: The sink collects the necessary information of every node through the dissemination of *REQ* message into WSN. The information includes: node ID and localization details such as (x, y) coordinates of the node's position in the WSN. Every node in the sensor field reacts to the sink by transmitting the requested information. The WSN is partitioned into the desired number of unequal clusters by sink depending on the localization information gathered from the nodes in the network. In unequal clustering, smaller sized clusters are formed which are near to the sink as compared to clusters created farthest from the sink. The sink computes the cluster radius RC_i to form clusters of unequal size using the maximum cluster radius R_i [7].

$$RC_i = \left[1 - C \cdot \frac{d_{\max} - d_i(c_i, S)}{d_{\max} - d_{\min}} \right] R_i \tag{3}$$

where d_{max} represents the maximum distance to cluster from the sink and d_{min} is the minimum distance, $d(C_i, S)$ is the separation from the cluster centre C_i to the sink, a constant coefficient C value range between 0 and 1, and the maximum cluster radius is R_i. By varying the constant C, the cluster radius can be varied. The cluster radius decreases, if the C value is increased from 0 to 1. The sink partition the complete sensor network into a predetermined number of clusters of unequal size.

After the completion of network portion, the sink broadcasts the cluster information *CLUST* which contains the cluster centre to every node in WSN and ID of the cluster. The decision of joining the cluster they belong to is made by obtaining the cluster ID information in the message *CLUST*. The cluster ID is very much useful to group the nodes into the cluster.

Cluster head advertisement phase: During this phase, selection of node as a cluster head for every cluster formed is processed. The critical parameter to be considered to

decide the cluster head for the subsequent round is the remaining energy of the node battery. In the beginning of every new round of protocol operation, node with more residual energy and very near to the cluster centre is elected as the head of cluster. Subsequently, all the member nodes of the cluster will receive an advertisement message *CH_ADV* from the elected cluster head. The *CH_ADV* message from the adjacent clusters is also being received by the nodes. Every node determines their probability to be the member of the cluster or not depending on the strength of the signal being received. If the nodes are inside the communication range of the cluster head, then the cluster heads receive *JOIN* message from them. Through this process, nodes intimate their willingness to be the member of that cluster to the closest cluster head. The cluster head gathers all the *JOIN* messages from the member nodes.

The collisions due to intra- and inter-cluster communication are avoided by employing TDMA scheduling technique. In this TDMA scheduling technique, each is allotted the specific time slot and is free to transmit only during their specific time slot and other nodes will be in the sleep or idle state.

TDMA Scheduling phase: Depending on the number of cluster members, the distance to the cluster head, the cluster head creates TDMA schedule. The cluster head creates TDMA time slot by taking the number of nodes and distances into account. These TDMA time slots convey the information when they can transmit the data to the cluster head. When the TDMA time slot of all the member nodes is decided by the cluster head, then it broadcasts the time schedule to the cluster member nodes.

Data Transmission phase: The member nodes of the cluster send their sensed information to the cluster heads in their allotted TDMA time slots. The member nodes can turn off their transceiver and can enter into sleep mode

The member nodes of the cluster can communicate only with the respective cluster heads and transmit their data to the immediate cluster head during the assigned TDMA time slots. After knowing their TDMA time slot assigned to them, the member nodes turn off their transceiver and goes to sleep mode. The energy of the nodes is greatly conserved by putting them in sleep mode untill their TDMA time slot. The node enters into wake up mode when their time slot arrives, and collects the data by sensing the media and transmits it to the cluster head. The data being sent by the member node during their time slot is collected by the cluster heads and data aggregation is performed for transmission. The redundant data being sent to the sink is reduced by the data aggregation techniques. This technique greatly contributes towards the conservation of energy in the sensor nodes.

Data from the network reaches the sink node by the multi-hop mode of transmission. In multi-hop mode, the data from the farthest cluster head reaches the sink through the nearest intermediate next level cluster head in the path of the sink. The data being received by all the intermediate cluster heads from higher cluster levels and as well as its own data undergoes data aggregation by every intermediate cluster heads and send the same to the next level cluster head. Finally, the data is forwarded to the sink by the mode of multi-hop transmission.

4 Simulation Results

The proposed protocol's performance is evaluated using MATLAB. The energy depletion in each node is calculated by considering the energy expended during transmission and aggregation. The performance of the DUCA is analyzed with the LEACH protocol. The simulation parameters [6] considered for the proposed protocol are listed in Table 1. The protocol was simulated by deploying 200 nodes in $100 \, \text{m} \times 100 \, \text{m}$ square region. The nodes in the LEACH deplete their energy quickly, since the CHs reach the base station in one hop, which is more energy-consuming process. Also, in LEACH, at the end of every round, the nodes make fresh negotiation to elect CH nodes. This increases the communication overhead and consumes more node energy. The nodes deplete their energy quickly and die faster in LEACH than UDCA. From Table 2, it is evident that the first node died during round 755 in LEACH, but in the proposed DUCA protocol it is reported in the round 891. The nodes death rate is prolonged in UDCA, since it uses energy-efficient techniques to extend the network lifetime. From Table 3, it is observed that last node died early in the LEACH. i.e. during the round 1358, but in proposed DUCA protocol it is in the round 1439.

In WSN, network longevity is one of the important measures for the performance evaluation of the whole network. If the node depletes the battery power very slowly without compromising the network operation, this increases the lifetime of the WSN. The network lifetime is measured by taking the number of nodes that are alive in each round into considerination and it is shown in Fig. 2. From Fig. 2, we observed that

Table 1 Simulation parameters

Parameter	Symbol	Values
Simulation area	M × M	100 m × 100 m
Total number of nodes	N	200
Base station position	(x, y)	(50, 175)
Packet size	P	500 bytes
Propagation delay	Td	50 μ s
Transmit/Receive electronic	Eelec	50 nJ/bit
Amplifier constant	$\in_{fs}$	10 pJ/bit/m2
	$\in_{fs}$	0.00013 pJ/bit/m2
Initial energy	E0	0.5 J
Energy for data aggregation	EDA	5 nJ

Table 2 First node death

Protocol	Round number
LEACH	755
DUCA	891

Table 3 Last Node Death

Protocol	Round number
LEACH	1358
DUCA	1439
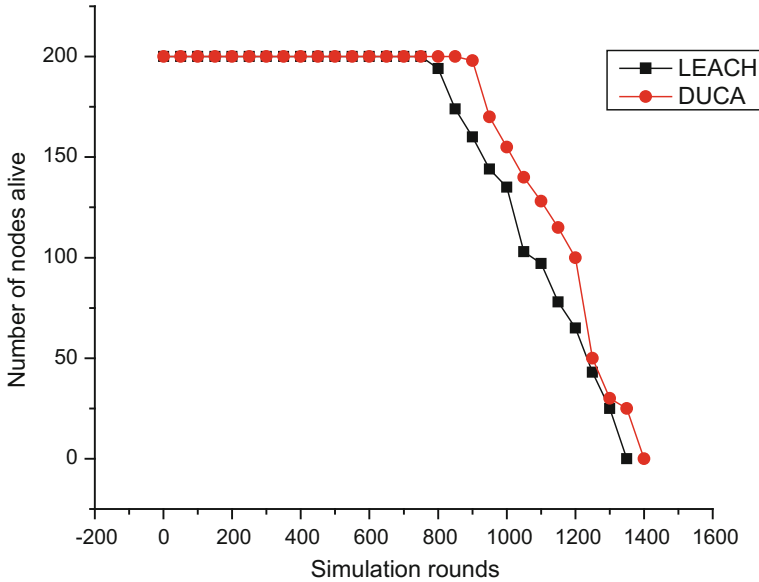
Fig. 2 Network lifetime

the proposed DUCA protocol maximizes the network life by controlling the death rate of the nodes. The algorithm employs distributed CH selection methodology as compared to the LEACH.

5   Conclusion

In this research work, a novel distributed unequal clustering protocol is presented. The distributed unequal cluster formation and multi-hop communication greatly contribute in extending the network lifetime. The hot spot problem is common in WSN with equal cluster size and also energy load among the nodes is also non-uniform. The hot spot problem is taken care by the method of unequal clustering, which also distributes the energy loads almost uniformly among the nodes.

In the presented DUCA algorithm, the cluster size is smaller near to the sink node compared to faraway clusters. This enables the CHs near to base station will have less energy loads from the member nodes of the cluster. In addition, these nodes

will have more loads from the faraway CHs. Since the energy load is spread evenly among the sensor nodes, so the nodes will have a lower death rate in the proposed algorithm compared to LEACH.

References

1. Liu, X.: A Survey on Clustering Routing Protocols in Wireless Sensor Networks. Sensors **12**, 11113–11153 (2012)
2. Younis, Ossama, Fahmy, Sonia: HEED: A Hybrid, Energy-Efficient, Distributed Clustering Approach for Ad Hoc Sensor Networks, pp. 366–379. IEEE Transactions, Mob. Comput. (2004)
3. Kumar, Dilip, Aseri, Trilok C., Patel, R.B.: EEHC: Energy efficient heterogeneous clustered scheme for wireless sensor networks. Comput. Commun. **32**, 662–667 (2009)
4. Lindsey, S., Raghavendra, C.: PEGASIS: power-efficient gathering in sensor information systems. In: Proceeding of IEEE Aerospace Conference, pp. 1125–1130 (2002)
5. Manjeswar, A., Agrawal, D.P.: TEEN: A protocol for enhanced efficiency in wireless sensor networks. In: Proceedings of 1st International Workshop on Parallel and Distributed Computing Issues in Wireless Networks and Mobile Computing, 189 p. San Francisco, CA, USA (2001)
6. Heinzelman, W., Chandrakasan, A., Balakrishnan, H.: Energy-efficient communication protocol for wireless microsensor networks. In: 2000 Proceedings of the 33rd Annual Hawaii International Conference on System science. IEEE (2000)
7. Guihai, C., Li, C., Ye, M., Wu, J.: An unequal cluster-based routing protocol in wireless sensor networks. Wirel. Netw. **15**(2) 193–207 (2009)

Video Image Retrieval Method Using Dither-Based Block Truncation Code with Hybrid Features of Color and Shape

Pradeep Yadav, Rishi Gupta and Sandeep Kumar

Abstract This paper presents the different approaches by which the video image retrieval systems can become more efficient. In today's world large database not only create the problem but also increases the complexity in terms of time as well as size. Traditional methods are now not so efficient to handle such problems, like computational time, response time, and complexity. In such a scenario dither-based block truncation method with hybrid features of color and shape together provides the better solution than BTC and all other methods. It not only limits the complexity but also provides the best compression and retrieval solutions.

Keyword Dither-based block truncation code · Hybrid · Color · Complexity

1 Introduction

People are producing millions of color images and videos daily, these may be of animals, plants, buildings, people, related to medical field, etc. From these images, we can collect huge amount of information and this information [1] provide us the different aspect and behavior of the data (images). Traditional image retrieval system was based on the features such as keywords, serial no., indexing, title, etc. This kind of system was useful for small databases, but as technology advances, database also expanded, so in this environment advanced system is needed for the retrieval that should be less time consuming and should be more efficient. Thus, many research scholars are working in this direction [2–10]. Now, the question comes how we can

P. Yadav (✉) · R. Gupta · S. Kumar
Department of CSE, Jagannath University, Jaipur, Rajasthan, India
e-mail: er.pradeepyadav0610@gmail.com

R. Gupta
e-mail: genieousrishi@gmail.com

S. Kumar
e-mail: sandpoonia@gmail.com

gather different information from the images and videos, as we have basic three features by which we can collect huge amount of information.

Color is a simple and basic prime feature to represent an image, different color extraction methods [11] have been developed and much research has been done in this field.

The second and most important feature of an image is shape, by using shape we can extract the features of an image, shape can vary differently, we use different edges to extract the shape features. Different algorithms have been developed such as Canny edge detector [12] used for shape feature detection, whereas edge histogram descriptor feature of MPEG-7 [13], in which different edges were used like horizontal, vertical, diagonal, etc.

The third important feature which plays an imperative role is texture, which shows the physical properties of the images. The physical properties of the image are calculated by such as roughness, contrast, directionality, degree of regularity, roughness. There are a number of texture features which have been used commonly like the Wavelet transform feature [14], Tamura Texture feature [15], Simultaneous Auto-Regressive (SAR) models [16], and Gabor texture features [17].

These are the different features by which we can find out the information about the images and videos. Nowadays many of the researches use hybrid methods for the information retrieval. Some of them use color and shape features other are using shape, textures, etc.

2   Literature Survey

Huang and Dai [18] asserted that accuracy and efficiency are two important factors that need to be focused while designing the content-based image retrieval system. To increase the efficiency and performance they use composite sub-band gradient vector and the energy distribution pattern string. So, these are the some of the basic parameters by which we can improve and make the most efficient system.

Lu and Chang [19] use the mean value, standard deviation, color distribution as parameters to represent the feature of the image. Hence, retrieval of related images from the digital database becomes more efficient. They also become successful in reducing the memory space to store the features of the image database compared to the other methods.

Lai and Chen [20] asserted that to precisely and effectively retrieve the similar image from the database, we need to follow some of the best mechanisms, and one of them is to use interactive genetic algorithm in which standard deviation, bitmap, and mean value possibly will be used as a color feature of the images or we may say that by using these attributes we can retrieve the features of the images. Similarly, edge histogram and the entropy are used as the texture features.

Due to the huge volume of the graphics information, there was a need of some efficient image retrieval algorithms, so that within a time frame efficient retrieval can

be achieved. Block truncation code is one of the best methods used for efficiently coding of images and also for indexing and retrieval purposes.

Guoping Qiu proposed block truncation code [21] for retrieval of the images from the large image database. Block co-occurrence matrix and block pattern histogram were the two features used to recover the images from the database. Here, block truncation method is not only used for compression but also used for indexing as well as retrieval. In this method, the mean value is calculated for each block.

3 Methods and Algorithms

3.1 Block Truncation Code Method

Block truncation code method [21] is one of the best compression methods. It is not only used for compression but also used for indexing and retrieval purposes. In this method, first of all, an image is to be divided into small blocks that may vary from size to size. Then, each block will be coded one at a time. Each pixel within the block is coded by a binary bitmap. The following steps will be followed:

Step 1 Split the image into different size blocks.
Step 2 Blocks may vary in size ($M \times N$), where $M = 1$ to n
And $N = 1$ to n.
Step 3 Let $X_1, X_2, X_3, \ldots, X_n$ are the values of each pixel in the block.
Step 4 Compute the mean pixel value of each block of frame
Mean = ratio of the sum of the values to the total number of elements in the set.

$$X_{mean} = (X_1 + X_2 + X_3 + X_n)/\ n \qquad (1)$$

Step 5 Compare each pixel value with the mean value, mean value will be treated as a threshold value(T).
Step 6 Encoding
If the value of a pixel is greater than or equal to the mean value or threshold value(T), then the bitmap value of that pixel will be 1 else the bitmap value of that pixel will be 0.
Step 7 Calculate the standard deviation value for each block.
Step 8 Decoding
For each block,replace the pixel bitmap value by the mean value.

As there are different color space models are available, RGB and CMY are one of them, to create bitmap in RGB color space we use three components in RGB; therefore, initially for each pixel position we have

$$P_{ij} = \frac{1}{3}\left[R_{ij} + G_{ij} + B_{ij}\right] \qquad (2)$$

where i $=$ 1 to m and j $=$ 1 to n.

After calculating the value P_{ij} (value of each pixel) then threshold is computed T. If the value of P_{ij} is greater than T then replace the value of each pixel by binary bitmap 1 on the other hand if the value is less than T then replace it by 0. Where M $\times$ N represents the block size.

After getting the binary bitmap value a binary matrix will be generated.

For example, if we have an image of 600×800 resolutions. Split the image into small blocks, let the block size be of M $\times$ N $= 4 \times 4$, as we have three color components for each pixel position, now the matrix of each color component is given as follows:

$$
\underbrace{\begin{bmatrix} 47\ 65\ 35\ 42 \\ 35\ 38\ 16\ 26 \\ 12\ 41\ 52\ 36 \\ 24\ 36\ 43\ 58 \end{bmatrix}}_{\dot{R}_{ij}}
\underbrace{\begin{bmatrix} 36\ 55\ 35\ 72 \\ 56\ 27\ 82\ 54 \\ 63\ 37\ 52\ 58 \\ 36\ 57\ 40\ 63 \end{bmatrix}}_{\dot{G}_{ij}}
\underbrace{\begin{bmatrix} 35\ 65\ 35\ 52 \\ 52\ 38\ 26\ 22 \\ 19\ 43\ 35\ 73 \\ 34\ 31\ 47\ 98 \end{bmatrix}}_{\dot{B}_{ij}}.
$$

After getting the matrix, we have P_{ij} $\begin{bmatrix} 39\ 61\ 35\ 53 \\ 47\ 34\ 41\ 34 \\ 31\ 40\ 46\ 55 \\ 31\ 41\ 43\ 73 \end{bmatrix}$.

Compute the threshold value (T) or mean value or $X_{mean} = 44$.

Now, compare the mean value with the threshold value (T).

If $(P_{ij}) \geq$ T replace it with 1.

Otherwise place 0.

By replacing, we will get following binary bitmap matrix

$$
\begin{bmatrix} 0\ 1\ 0\ 1 \\ 1\ 0\ 0\ 0 \\ 0\ 0\ 1\ 1 \\ 0\ 0\ 0\ 1 \end{bmatrix}.
$$

By this, we get the bitmap value of each block and after computing all the blocks we will get the bitmap representation of an image, this process is known as decoding.

Block truncation method is not only a compression method but also used for image indexing and image retrieval, several methods have been developed with the help of BTC to retrieve images on the basis of different features, like color, shape, and texture. Now, in this era technology is developing day by day, to improve the performance researchers developed some advanced method to overcome the limitation of this algorithm, one of the methods that Jing-Ming Guo and Ming-Feng Wu proposed in their research [22] that is known as ordered dither block truncation coding (Fig. 1).
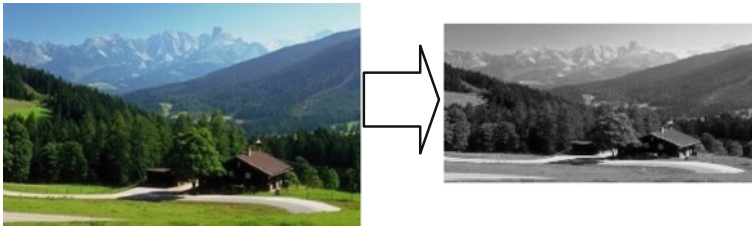
Fig. 1 Conversion from color to grayscale using BTC algorithm [23]

3.2 Dither-Based Block Truncation Code

To improve the performance of the block truncation code method dither-based BTC has been introduced. As we know that large database (specially images and video collection) become common and retrieval of the data from the large database becoming complex and time consuming day by day. In order to reduce the computational complexity and to reduce the processing time, dither-based block truncation code [24] plays an imperative role.

Dither-based block truncation code yields better image quality by adding dither value effect.

Let the dither array be

$M_{array} = [M_0, M_1, M_2, \ldots, M_n]$ where $0 < n < 255$.

As we have the minimum and maximum values of the block, say X_{min} and $X_{max.}$

By using the dither array table in which different scaled dither array stored ranging from 0 to 255. These different scaled dither arrays are obtained by the given formula [22, 24]

$$M_{P_{ij}} = K \times \frac{M(P_{ij}) - M_{min}}{M_{max} - M_{min}} \tag{3}$$

where P_{ij} represents the pixel position and M_{max} and M_{min} show the maximum and minimum values in the dither array [25], where k represents the scaling factor ranging from 0 to 255.

Therefore, by using dither-based block truncation code method a better quality image can be reconstructed.

If we want to retrieve the images, then we can use color as well as shape feature by using dither approach. This is the best method for retrieval proposed in [24].

3.3 Similarity Algorithms

There are different kinds of similarity measurement algorithms available, which are used to find the similarity between the elements, if the distance value between the two

elements is zero then the images appear similar and similarity factor becomes one. Several algorithms are available to measure the similarity, some of them are Euclidean distance measurement algorithm, min-max ratio algorithm, Hamming distance, etc.

The distance between the two vectors is calculated by the sum of squared differences

$$D(x, y) = \sum_{n=1}^{N} (x_n - y_n)^2 \tag{4}$$

$$S(x, y) = 1 - D(x, y) \tag{5}$$

x   the feature vector of query image.
y   the feature vector of one of the image present in the database.
n   the nth coefficient of the feature vector.
S   Similarity value of the image

The distance between the two vectors is calculated on the basis of min-max ratio. The similarity between the two images X and Y is given by

$$S(X, Y) = \frac{\sum_{n=1}^{N} \min(x_n, y_n)}{\sum_{n=1}^{N} \max(x_n, y_n)} \tag{6}$$

4 Hybrid Method

In this method, we can take the video and by using the color and shape together as a feature we can create the database. As we know that video is a collection of images, we get the sequence of images similar to each other. As an individual image define its property by color, shape, and texture. Hence, here color and shape provide the basic information about an image. By combining the color (C) and shape (H), we get CH feature by using it we create the database.

By using the sequence of images, we have database of several videos. Here, we have two videos one of the video is given in Fig. 2 [26] in which the video is converted into sequence of frames and in the other video we have a man playing golf.

Now, by using the color (C) and shape (H), i.e. CH we will create the database. Now, by using CH as a feature of query image or sequence of images by using similarity measurement algorithm we will compare the database. This method will be more efficient than other methods.

By the above results, we can see that the video retrieval is more accurate by using color and shape feature together. In Fig. 3 [27], a person with a red t-shirt is our query video, we are getting only four similar videos by color feature on the other hand by using bit-pattern feature or shape we will get only two similar video frames, and on the other hand if we combine both the feature we will get four similar video frames,

Fig. 2 Sequence of images (Video)
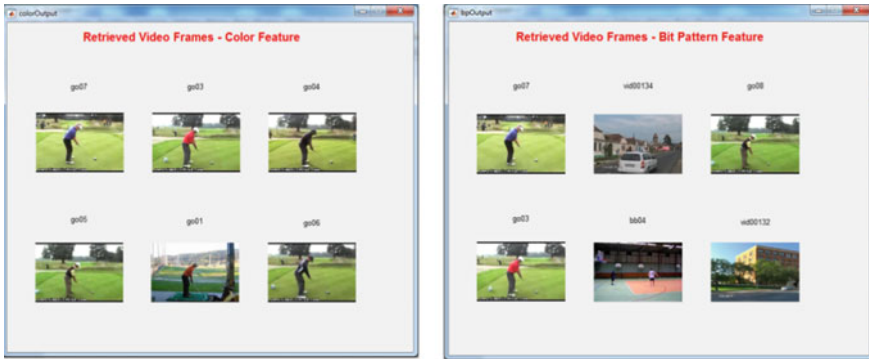
Fig. 3 Retrieved video frames on the basis of color and shape, respectively

this shows that the CH hybrid method is more efficient than other methods (Fig. 4) [27].

Fig. 4 Retrieved video frames on the basis of CH

5 Performance Evaluation

Different evaluation parameters are used to evaluate the performance as well as efficiency of the system. Some of the parameters are as follows:

(1) Precision: The numbers of similar images retrieved to the total number of images are retrieved.

$$P = Ps/Ts \tag{7}$$

(2) Recall: The number of similar images retrieved to the total number of similar images in the whole database.

$$R = Rs/Tr \tag{8}$$

By these two parameters, we evaluate the performance.

6 Conclusion

Hybrid features using the dither algorithm on videos is a more efficient and best retrieval method, it reduces the complexity and computational time, it provides the best solution for reconstructing the quality based images. By using hybrid feature selection, we can achieve better results. We can use color and shape together to create

database of the video images, we can vary the range of colors and different shape patterns so better results can be achieved.

References

1. Brunelli, R., Mich, O.: Image retrieval by examples. IEEE Trans. Multimed. **2**, 164–171 (2000)
2. Huang, P.W., Dai, S.K.: Image retrieval by texture similarity. Pattern Recogn. **36**(3), 665–679 (2003)
3. Wu, S., Rahman, M.K.M., Chow, T.W.S.: Content-based image retrieval using growing hierarchical self-organizing quadtree map. Pattern Recogn. **38**, 707–722 (2005)
4. Moghaddam, H.A., Khajoie, T.T., Rouhi, A.H., Tarzjan, M.S.: Wavelet correlogram: a new approach for image indexing and retrieval, Pattern Recogn. **38**, 2506–2518 (2005)
5. Brnuelli, R., Mich, O.: Histograms analysis for image retrieval. Pattern Recogn. **34**, 1625–1637 (2001)
6. Fuertes, J.M., Lucena, M., Peres de la Blanca, N., Chamorro-Martinez, J.: A Scheme of color image retrieval from databases. Pattern Recogn. **22**(3), 323–337 (2001)
7. Chun, Y.D., Seo, S.Y., Kim, N.C.: Image retrieval using BDIP and BVLC moments. IEEE Trans. Circuits Syst. Video Technol. **13**(9), 951–957 (2003)
8. Ko, B.C., Byun, H.: FRIP: a region-based image retrieval tool using automatic image segmentation and stepwise Boolean and matching. IEEE Trans. Multimed. **7**(1), 105–113 (2005)
9. Chan, Y.K., Chen, C.Y.: Image retrieval system based on color-complexity and color-spatial features. J. Syst. Softw. **71**(1–2), 65–70 (2004)
10. Chang, C.C., Chan, Y.K.: A fast filter for image retrieval based on color features, SEMS2000, pp. 47–51. German, Baden-Baden (2000)
11. Xue; B., Wanjun, L.: Research of image retrieval based on color. In: International Forum on Computer Science-Technology and Applications, IFCSTA'09., vol.1, pp. 283–286. 25–27 Dec (2009)
12. Ramamurthy, B., Chandran, K.R.: Content based image retrieval for medical images using canny edge detection algorithm. Int. J. Comput. Appl. (0975–8887), 17(6) 2012
13. Agarwal, S., Verma, A.K., Singh, P.: Content based image retrieval using discrete wavelet transform and edge histogram descriptor. In: International Conference on Information Systems and Computer Network (ISCON), pp. 19–23. 9–10 March 2013
14. Daubechies, Ingrid: The wavelet transform, time-frequency localization and signal analysis. IEEE Trans. Inf. Theory **36**(5), 961–1005 (1990)
15. Pentland, A., Picard, R.W., Sclaroff, S.: Photobook: tools for content-based manipulation of image databases. In: Storage and Retrieval for Image and Video Databases (SPIE) pp. 34–47 (1994)
16. Mao, Jianchang, Jain, Anil K.: Texture classification and segmentation using multiresolution simultaneous autoregressive models. Pattern Recogn. **25**(2), 173–188 (1992)
17. Manjunath, B.S., Ma, W.-Y.: Texture features for browsing and retrieval of image data. IEEE Trans. Pattern Anal. Mach. Intell. **18**(8), 837–842 (1996)
18. Huang, P.W., Dai, S.K.: Image retrieval by texture similarity. Pattern Recognit. **36**(3), 665–679 (2003)
19. Lu, T.-C., Chang, C.-C.: Color image retrieval technique based on color features and image bitmap. Inf. Process. Manage. **43**(2), 461–472 (2007)
20. Lai, C.-C., Chen, Y.-C.: A user-oriented image retrieval system based on interactive genetic algorithm. IEEE Trans. Instrum. Meas. **60**(10), 3318–3325 (2011)
21. Qiu, G.: Color image indexing using BTC. IEEE Trans. Image Process. **12**(1), 93–101 (2003)
22. Guo, J.-M., Wu, M.-F.: Improved block truncation coding based on the void-and-cluster dithering approach. IEEE Trans. Image Process. **18**(1), 211–213 (2009)

23. Corel photo collection color image database. [Online]. Available: http://wang.ist.psu.edu/docs/realted/. Accessed 2001
24. Guo, J.-M., Prasetyo H.: Content-based image retrieval using features extracted from halftoning-based block truncation coding. IEEE Trans. Image Process. **24**(3), 1010–1024 (2015)
25. Yadav, P., Gupta, R., Kumar, S.: Content based image retrieval using dither block truncation coding with similarity comparison algorithm. In: International Conference on Computer, Communication and Electronics (Comptelix), 1–2 July 2017
26. Trec video retrieval track video database, 2005. [Online]. Available: http://www-nlpir.nist.gov/projects/trecvid/
27. Index of data, Vision.eecs.ucf.edu. [Online]. Available: http://vision.eecs.ucf.edu/data/

Significance of Haralick Features in Bone Tumor Classification Using Support Vector Machine

M. V. Suhas and B. P. Swathi

Abstract Accurate classification of bone lesions into benign and malignant tumors plays a key role in determining the treatment course (surgical intervention or radiation), an essential part of radiologists work. In this work, we investigate the significance of Haralick textural feature components in an application to computer-aided classification of the bone tumor into benign and malignant. The Haralick features from Computed Tomography (CT) images are extracted to form the dataset. Support Vector Machine (SVM) with medium Gaussian kernel function is trained and tested with the dataset consisting of various malignant and benign tumors. The dataset prepared by extracting the Haralick features are subjected to Correlation-Based Feature Subset (CFS) Selection. The accuracy of the classifier is measured in each case. The study reveals an increased accuracy post-feature selection.

Keywords Haralick features · Support vector machine · Computed tomography
Bone cancer · Correlation based feature subset selection

1 Introduction

Cancer being a genetic disease is mainly caused due to several factors including environmental. Approximately, 60–70% of patients with cancer develop spinal metastases [1] that ultimately leads to neurological problems such as numbness of limbs, paralysis, and metastatic epidural spinal cord compression [2]. Bone lesions are the anomalies in the growth or structure of bones that have been categorized into three

M. V. Suhas
Department of ECE, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, Karnataka, India
e-mail: suhas.mv@manipal.edu
URL:https://manipal.edu/mit/department-faculty/faculty-list/suhas-mv.html

B. P. Swathi (✉)
Department of I & CT, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, Karnataka, India
e-mail: swathi.BP@manipal.edu

classes: benign (noncancerous), malignant (cancerous), and general bone cysts by the British Institute of Radiology. Imaging spine caters different needs such as identifying the early tumor involvement of bone, determining the extent of the disease in bone, assessment of accompanying complications such as cord compression and fractures and monitoring response to therapy [3–8]. Various anatomical and functional imaging modalities assist in the assessment of benign and malignant bone involvement. CT is one such anatomical modality which exhibits the distinctness between malignant and benign vertebral collapse. CT is more sensitive in detecting bone lesions prior to extensive destruction or involvement of marrow and can recognize a metastatic bone lesion up to 6 months in advance compared with X-ray [9]. Classification of benign and malignant bone tumors by radiologists is one of the important tasks which is tedious, time-consuming, and also leads to misclassification [10] when performed manually. The malignancy is often suggested through the detection of soft tissue accompaniment or posterior vertebral body elements involvement [11]. The ill-defined and the irregular sclerotic tumor boundary image features on CT defines the tumors are malignant. Meanwhile, benign tumors on CT image pose a well-defined profile with the edges being regularly sclerotic. These important and useful features which radiologists use in interpreting and classifying the tumors visually are embedded as textures on CT images. These textural features are extracted in a preparation to build a dataset for the work. The built dataset is utilized in training and testing a classification algorithm to classify benign and malignant bone tumors. The dataset is also subjected to a feature selection process to cross-validate the variation in classification accuracy. The usefulness of 36 textural features for bone tumor classification is investigated using Support Vector Machine (SVM) with medium Gaussian kernel function. The following section explains briefly the background theory of techniques used in this work.

2 Background Theory

This section deals with the necessary theory required for the implementation of the methodology. It explains, in brief, the techniques such as textural features, SVM, and feature selection methods.

2.1 Textural Features

Images can be interpreted using three features such as Spectral, Textural, and Contextual [12]. The consideration of textures and tones are important in the context of independent processing of small image areas in monochrome photographs. Textural features from CT images are extracted for the presented work because they possess spatial distribution information of variations in tone within a band. The variation in the shades of gray of resolution cells in an image defines the tone and is judged

as fine, coarse, smooth, lineated, irregular, and so forth. While texture symbolizes statistical or spatial distribution of gray tones in an image and describes the darkness or lightness of a particular area in an image. Texture and tone have a highly complex and intricate relationship among each other [12]. Gray-tone spatial-dependence matrix describes the frequency of appearance of gray tone to another gray tone in a specified spatial relationship on the image. Textural features can be discrete histograms, scalar numbers, or empirical distributions. The domination of texture increases with an increase in the number of discrete gray-tone distinguishable features of within the small area patch, which is true in the case of tumor appearance on the CT images.

2.2 Support Vector Machine and Gaussian Kernel Function

The linear classifier SVM show a good performance during classification or regression when compared to other machine learning techniques [13]. It can be extended to a nonlinear classier by mapping input data $S = \{X\}$ into a high-dimensional (possibly in nite-dimensional) feature space $F = f(X)$ by selecting an adequate mapping function f, a kernel. SVM aims at minimizing generalization error upper bound by maximizing the hyperplane and the data margin [14]. A radial basis function (RBF) is a real-valued function whose value depends only on the distance from the origin, so that $f(x) = f(||X||)$. The Gaussian RBF kernel makes a good default kernel in absence of expert knowledge about data and domain because it subsumes polynomial and linear kernel. Linear Kernels and Polynomial Kernels are a special case of Gaussian RBF kernel. The value of sigma in Gaussian kernels plays a key role by defining how the kernel will handle the neighbor and points that are far away. A high value of sigma takes into account all points but does not emphasize neighborhood [15]. Medium Gaussian kernel takes advantage of both low and high values of sigma

2.3 Feature Selection

Machine learning algorithms, inclusive of decision trees are known to deteriorate in performance when exposed to too many features in the training and test dataset that are not necessary for prediction. In the extracted features, some of the features are strongly correlated with each other. Hence, feature selection technique is applied on the prepared dataset to determine, if a subset of textural features can yield better accuracy in Support Vector Machine. An approach to feature subset selection called the filter is used proven for its speed compared to wrappers and hence can be applied to large datasets containing many features [16].

3 Methodology

The methodology of the work is briefly illustrated in the block diagram shown in the Fig. 1. The proposed work was conducted in a workstation with Intel® Core™ i5-6200U CPU 4 GB RAM and 2.30 GHz clock frequency with 64-bit operating system. The algorithm is implemented on MATLAB® version R2017a.

Fig. 1 Flowchart of Methodology
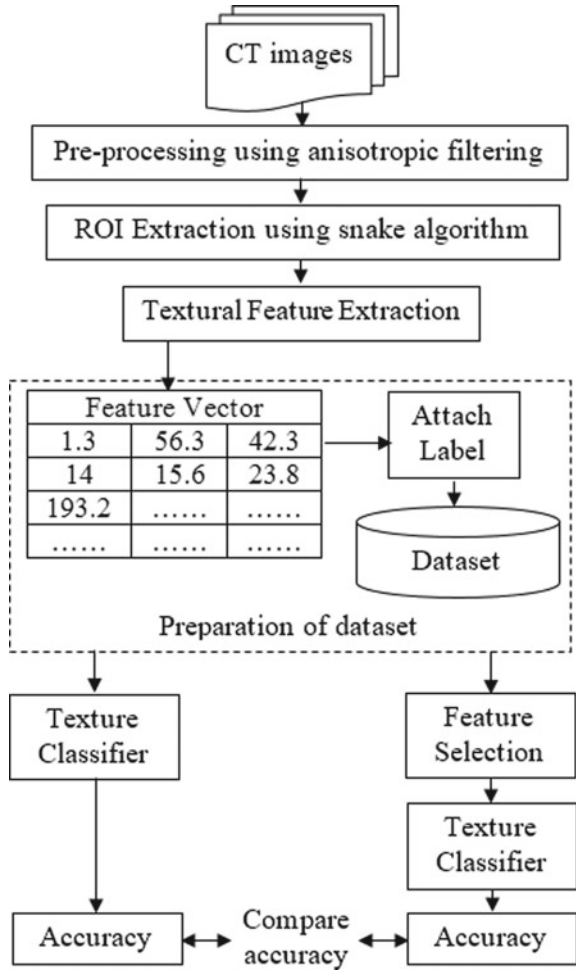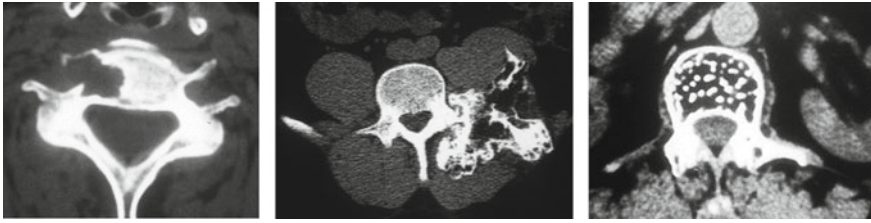
Fig. 2 CT images of osteoblastoma, osteochondroma, hemangioma from left to right
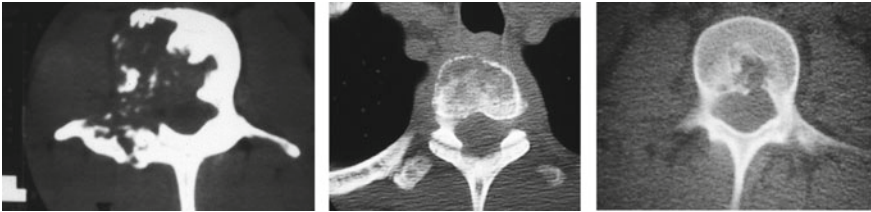
Fig. 3 CT images of Chondrosarcoma, Ewings Sarcoma, Chordoma from left to right

3.1 Dataset

The CT images were obtained from Manipal Hospital Bangalore with the consent of patients and ethical clearance. The existence of bone tumors was confirmed by the radiologists in the obtained CT scans by visual examination. In assistance with the radiologists, the CT images containing tumors were grouped into benign and malignant classes. The final image dataset consisted of 29 malignant bone tumors and 36 various benign tumors. The Figs. 2, and 3 shows the appearance of benign and malignant tumors on spine.

3.2 Preprocessing

The predominant presence of noise in biomedical images that deviate the result makes filtering a necessary and obvious process. In the method, the classified images are initially preprocessed to remove noise that might reduce the quality of features to be extracted. CT images are affected by noise due to thin sections and high-resolution image reconstruction algorithm. The preprocessing step comprises filtering CT images in order to remove the inherent noise and enhance the edge. While filtering the CT images consisting of tumor, the challenge is to retain the tumors of smaller dimension if any. Applying Gaussian filter dilates the smaller tumors, thereby leading to false negative results. Based on these challenges, the decision is made to apply nonlinear anisotropic filter. The applied anisotropic filter smoothens

the homogeneous or uniform regions thereby removing noise and preserving its appearance and enhances the discontinuous regions such as edges. Anisotropic filter [17] is defined mathematically as

$$I_t = div(c(x, y, t)\nabla I) = c(x, y, t)\Delta I + \nabla c \cdot \nabla I \tag{1}$$

where div is the divergence operator, ∇ is the gradient operator, and $\triangle$ is the Laplacian operator. The diffusion function of $c(x, y, t)$ depends on the image intensity gradient magnitude unlike linear filter where the diffusion coefficient is considered constant independent of space location. The diffusion function used in our method is defined as

$$c(\nabla I) = e^{(-(|\nabla I|/K)^2)} \tag{2}$$

The selection of the diffusion function is based on its ability to enhance edges of high contrast over low contrast ones. The control parameter K is set manually depending on the amount of noise to be filtered.

3.3 Extraction of Region of Interest (ROI)

The ROI is extracted from the preprocessed images with the help of an active contour model (or snake). Active contours are curves that minimize the energy functional by moving through the spatial domain of an image. This energy minimizing technique pulls the snake near the edges, lines, and subjective corner contours being affected by three types of forces: external constraint forces, internal or image forces, and external forces. The expression of energy functional is

$$E = \int_0^1 \frac{1}{2}(\alpha \mid x'(s) \mid^2 + \beta \mid x''(s) \mid^2) + E_{ext}(x(s))d \tag{3}$$

where the α controls the tension and β controls the rigidity of the contour $x'(s)$ denotes the first derivative and $x''(s)$ denotes the second derivative of $x'(s)$. The internal force controlled by α and β discourage the stretching and bending of the contour and external potential force pulls the contour towards the desired contour of tumors. Due to prominent appearance of bone tumor boundary on CT, a simple snake algorithm provides better segmentation result.

3.4 Feature Extraction and Selection

The segmented ROI is then passed on to feature extraction. A statistical technique called Gray-Level Co-occurrence Matrix (GLCM) is used for feature extraction that

tabulates the frequency of different combinations of pixel gray levels in an image. An unknown sample image is assigned to one of a set of known texture classes. A set of 36 textural features are extracted from the gray-tone spatial-dependence matrix [12] and are defined by the following equations.

Energy:

$$f_1 = \sum_i \sum_j p(i, j)^2 \quad (4)$$

Contrast:

$$f_2 = \sum_{n=0}^{N_g-1} n^2 \{ \sum_{i=0}^{N_g} \sum_{j=1}^{N_g} p(i, j) |i - j| = n \} \quad (5)$$

Correlation:

$$f_3 = \frac{\sum_i \sum_j (i, j) p(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (6)$$

Varience:

$$f_4 = \sum_i \sum_j (i - \mu)^2 p(i, j) \quad (7)$$

Homogeneity:

$$f_5 = \frac{1}{1 + (i - j)^2} p(i, j) \quad (8)$$

Sum Average:

$$f_6 = \sum_{i=2}^{2N_g} i p_{x+y}(i) \quad (9)$$

Sum Variance:

$$f_7 = \sum_{i=2}^{2N_g} (i - f_8)^2 p_{x+y}(i) \quad (10)$$

Sum Entropy:

$$f_8 = \sum_{i=2}^{2N_g} p_{x+y}(i) \{ p_{x+y}(i) \} \quad (11)$$

Entropy:

$$f_9 = \sum_i \sum_j p(i, j) log(p(i, j)) \quad (12)$$

Difference Variance:

$$f_{10} = variance of\ P_{x-y} \quad (13)$$

Sum Entropy:

$$f_{11} = \sum_{i=0}^{N_g-1} p_{x-y}(i) \{ p_{x+y}(i) \} \quad (14)$$

Autocorrelation:

$$f_{12} = \sum_i \sum_j (i, j) p(i, j) \quad (15)$$

Dissimilarity:

$$f_{13} = \sum_i \sum_j |i, j| p(i, j) \quad (16)$$

Cluster Shade:

$$f_{14} = \sum_i \sum_j (i + j - \mu_x - \mu_y)^3 p(i, j) \quad (17)$$

Cluster Prominence:

$$f_{15} = \sum_i \sum_j (i + j - \mu_x - \mu_y)^4 p(i, j) \quad (18)$$

Maximum Probability:

$$f_{16} = \max_{i,j} p(i, j) \quad (19)$$

Where HX and HY are entropies of P_x and P_y

Information Measure of Correlation:

$$f_{17} = \frac{HXY - HXY1}{\max\{HX, HY\}} \quad (20)$$

$$f_{18} = (1 - exp[-2(HXY2 - HXY)])^{\frac{1}{2}} \quad (21)$$

$$HXY1 = -\sum_i \sum_j p(i, j) \log\{p_x(i)p_y(j)\} \quad (22)$$

$$HXY2 = -\sum_i \sum_j p_x(i)p_y(j) \log\{p_x(i)p_y(j)\} \quad (23)$$

Four angular gray-tone spatial-dependency matrices are calculated for a given distance d which obtains a set of 4 values for every 18 measures. The classifier inputs consist of a set of 36 features that are obtained from the mean and range of 18 measures, averaged over the 4 values. The extracted 36 features were inspected to observe that there are no outliers, missing values, and noise, hence a clean dataset is obtained.

The applied feature selection technique is Correlation-Based Feature Subset Selection (CFS) which works in two stages are as follows and is shown in Fig. 4. Searching the Feature Subset Space: for every n initial possible features, there are 2n possible subsets. One of the heuristic search strategies called Best First [18] is often applied and is also faster than other heuristic strategies [16]. Correlation-Based Feature Selection: CFS is feature selection method which uses a search algorithm along with a function M_S that evaluates the merit of a feature subset S that contains k features as follows:

$$M_s = \frac{k\overline{r_{cf}}}{\sqrt{k + k(k-1)\overline{r_{ff}}}} \quad (24)$$

where $\overline{r_{cf}}$ is the mean correlation between feature and class $f \in S$ and $\overline{r_{cf}}$ is the average correlation between features. The numerator of equation provides an indication of a features class predictiveness; the denominator provides with the redundancy
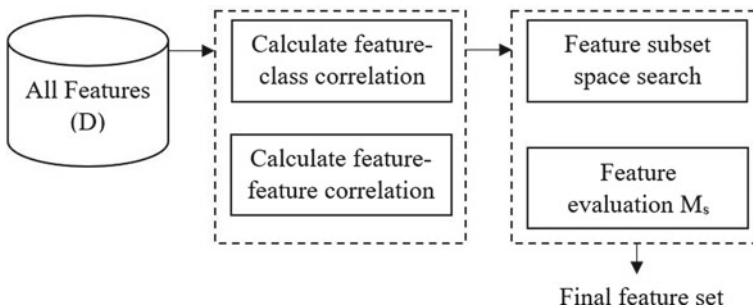
Fig. 4 Flowchart of feature selection

among features. To find the correlation between the two continuous valued attributes, Pearsons correlation is used [19]. The equation is stated as

$$r_{xy} = \frac{\sum xy}{n\sigma_x\sigma_y'} \tag{25}$$

where X and Y are expressed in terms of standard deviation and are continuous valued variables. When one of the attribute is continuous and the other is discrete a weighted Pearsons correlation is calculated, as represented in Eq. 26

$$r_{xy} = \sum_{i=1}^{k} p(X = x_i) r_{X_{bi}Y} \tag{26}$$

where X_{bi} is a binary attribute which assumes a value 1 with X taking a value x_i and 0 otherwise.

The features hence extracted and selected are separately used to train SVM with Gaussian the radial basis function which on two samples x and x is defined as

$$K(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right) \tag{27}$$

$\|x - x'\|^2$ is the squared Euclidean distance between the two feature vectors and is a free parameter. Medium Gaussian kernel is defined as

$$k_{Mixed}(x, y) = \lambda_H, k_G^{\sigma H}(x, y) + \lambda_L, k_G^{\sigma L}(x, y) \tag{28}$$

The following section analyzes the obtained result comparing the performance of SVM under the two different dataset.

4 Result Analysis

The resulting analysis is performed in two categories, a general analysis of ROC and confusion matrix and the variation in accuracy due to variation in kernel function.

4.1 ROC and Confusion Matrix

Since the available dataset is small, the accuracy of prediction is examined using a fivefold cross-validation method. The dataset is partitioned into fivefold disjoint sets. The model is trained using out-of-fold observations and testing is performed using in-fold dataset and average test error is calculated over all folds. Figure 5 shows the
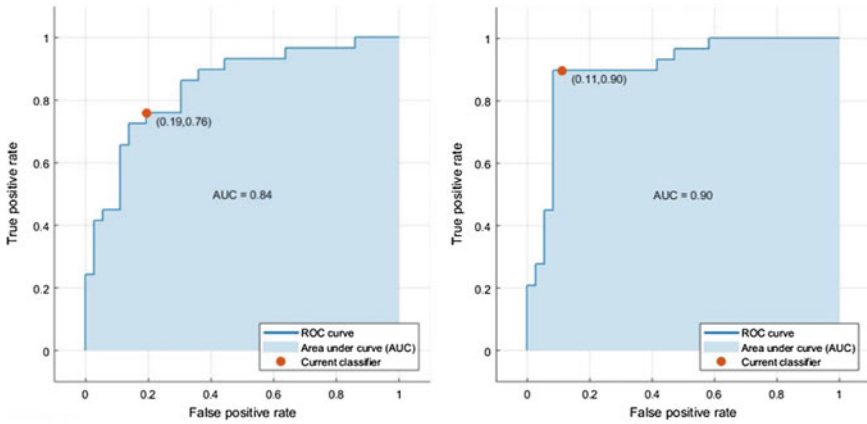
Fig. 5 ROC: left: Complete dataset, right: CFS dataset

Table 1 Confusion Matrix for complete dataset

		Predicted class		Total
		Benign	Malignant	
True class	Benign	29	7	36
	Malignant	7	22	29
	Total	36	29	65

Table 2 Confusion Matrix for CFS data

		Predicted class		Total
		Benign	Malignant	
True class	Benign	32	4	36
	Malignant	3	26	29
	Total	35	30	65

obtained ROC curve for complete and CFS dataset. We observed an increased area under the curve of 0.90 in case of CFS dataset classification compared to 0.84 of the complete dataset

The confusion matrices obtained for maximum accuracy in case of complete and CFS dataset are shown in Tables 1 and 2 denoting an increased detection rate for the latter case.
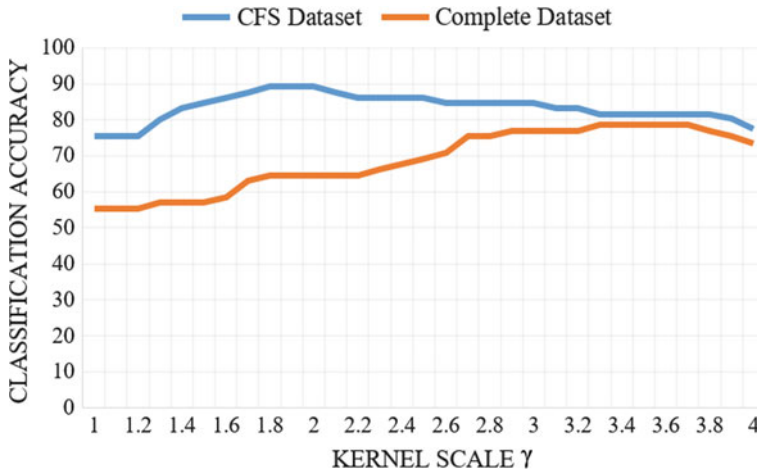
Fig. 6 Graph of Accuracy Variation with kernel scale

4.2 Accuracy Variation Due to Kernel Scale

The accuracy was also calculated at different values of kernel scale γ which is tabulated as shown in the graph. The graph compares the variation in accuracy between CFS dataset and complete dataset for a γ that varies from 1 to 4. These variations are graphically shown in Fig. 6 According to the observations, classification accuracy with CFS dataset was consistently higher than that with complete dataset. A maximum accuracy of 89.2% was observed in case of CFS dataset at a γ corresponding to 2 whereas, the complete dataset showed a maximum classification accuracy of 78.5% at a γ corresponding to 3.6. The observed prediction speed for complete dataset had 700 obs/s which increased to 2500 obs/s with CFS dataset showing an improvement in the prediction speed with reduced dataset

4.3 Precision and Recall

The precision of classification is measured and compared using the formula $Precision = \frac{TP}{TP+FP}$. We observed a precision of 75.86% in the case of complete dataset and an increased precision of 86.66% for CFS dataset. Similarly, the sensitivity or recall was also measured using the formula $sensitivity = \frac{TP}{TP+FN}$. A sensitivity of 75.86% was observed for complete dataset and observed sensitivity increased to of 89.65% for CFS dataset.

5 Conclusion

The work though mainly does not compare any existing works in the interest of SVM for classification of benign and malignant bone tumors, it tries to compare the efficiency of SVM with and without feature selection over various SVM kernel scale. The variations in the accuracy of SVMs is a clear indicator of variations in classifier accuracy due to feature selection. Also, the observations of prediction speed improvement suggest the essence of the feature selection process. The work through its results concludes the necessity of the feature selection process for the increased classification accuracy of SVM.

The future work mainly concentrates on increasing the dataset and thereby validate the results obtained in the current work and also improve the accuracy. We also intend to verify various other classification algorithms in bone tumor classifications and thereby chooses the best classification algorithm. The classification efficiency of SVM can also be tested for different preprocessing techniques.

References

1. Shah, L.M., Salzman, K.L.: Imaging of spinal metastatic disease. Int. J. Surg. Oncol. 12 (2011)
2. Lee, C.-S., Jung, C.-H.: Metastatic spinal tumor, Asian Spine J. **6**(1), 71–87 (2012)
3. Siegel, M.J.: MRI of bone marrow (2011). Disponible en www.arrs.org/shopARRS/products/pdf.cfm
4. Townsend, D.: Physical principles and technology of clinical pet imaging. Ann. Acad. Med. Singap. **33**(2), 133145 (2004)
5. Vogler 3rd, J., Murphy, W.: Bone marrow imaging. Radiology **168**(3), 679–693 (1988)
6. Salvo, N., Christakis, M., Rubenstein, J., de Sa, E., Napolskikh, J., Sinclair, E., Ford, M., Goh, P., Chow, E.: The role of plain radiographs in management of bone metastases. J. Palliat. Med. **12**(2), 195–198 (2009)
7. Daffner, R., Lupetin, A., Dash, N., Deeb, Z., Sefczek, R., Schapiro, R.: MRI in the detection of malignant infiltration of bone marrow. Am. J. Roentgenol. **146**(2), 353–358 (1986)
8. Uchida, N., Sugimura, K., Kajitani, A., Yoshizako, T., Ishida, T.: MR imaging of vertebral metastases: evaluation of fat saturation imaging. Eur. J. Radiol. **17**(2), 91–94 (1993)
9. Salvo, N., Christakis, M., Rubenstein, J., et al.: The role of plain radiographs in management of bone metastases. J. Palliat. Med. **12**(2), 195–198 (2009)
10. Mitra, S., Shankar, B.U.: Medical image analysis for cancer management in natural computing framework. Inform. Sci. **306** 111–131 (2015)
11. Even-Sapir, E.: Imaging of malignant bone involvement by morphologic, scintigraphic, and hybrid modalities*. J. Nucl. Med. **46**(8), 1356–1367 (2005)
12. Haralick, R.M., Shanmugan, K., Dinstein, I.: Texture features for image classification. IEEE Trans. Syst. Man Cybern. **3**(6) 610–621 (1973)
13. Meyer, D., Leisch, F., Hornik, K.: The support vector machine under test. Neurocomputing **55**(1), 169–186 (2003)
14. Amari, S., Si, W.: Improving support vector machine classifiers by modifying kernel functions. Neural Net. **12**(6), 783–789 (1999)

15. Blanc-Talon, J., Bone, D., Philips, W., Popescu, D., Scheunders, P.: Advanced Concepts for Intelligent Vision Systems. In: 2010, Proceedings of 12th International Conference, ACIVS 2010, vol. 6474. Springer, Sydney, Australia, Dec 13–16 (2010)
16. Hall, M.A., Lloyd A.S: Practical Feature Subset Selection for Machine Learning, pp. 181–191 (1998)
17. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. IEEE Trans. Pattern Anal. Mach. Intell. **12**(7), 629–639 (1990)
18. Rich, E., Knight, K.: Artificial Intelligence. McGraw-Hill (1991)
19. Scholkopf, B., Sung, K.K., Burges, C.J., Girosi, F., Niyogi, P., Poggio, T., Vapnik, V.: Comparing support vector machines with Gaussian kernels to radial basis function classifiers. IEEE Trans. Signal Process. **45**(11), 2758–2765 (1997)

Time-Series Outlier Detection Using Enhanced K-Means in Combination with PSO Algorithm

Neha Kant and Manish Mahajan

Abstract Outlier detection has been used in various fields due to multi-dimensional data sets, No. of fraud cases, etc. Outlier detection is used to detect and predict air pollution, fraud detection, anomaly detection in several areas of applications. Outliers in air quality time-series data is used in finding asthma exacerbation rate, pollution percentage, effects of smoking, etc. In this paper, we proposed a new method for detecting outliers using Enhanced K-means Clustering algorithm in combination with PSO (Particle Swarm Optimization) Plotbowl on the air quality time-series dataset. Data points were grouped into similar groups by using K-means, weight-based centre approach, and standard deviation approach was applied to improve the efficiency of K-means algorithm, so that outliers can be treated efficiently. To optimize this work, PSO plotbowl was used in combination with Enhanced K-means, so that we can get more refined set of data points.

Keywords PSO · K-means · WBC · Plotbowl · GRUBBS

1 Introduction

Outlier detection has been used in many areas such as Fraud detection, Intrusion detection, Healthcare, Fault detection, etc., where detection of outliers is based on the different characteristics of data or datasets. Detection of outliers is also called as detection of anomaly, which has been retained as a research problem in many areas like data mining, statistical fields [1]. Outlier detection finds abnormal patterns in any dataset that do not fit in with expected behaviour. These non-adjusting patterns are frequently referred to as anomalies, novelties, outliers or exceptions. Outlier detection is a major issue emerging in data mining since a couple of studies have been directed

N. Kant (✉) · M. Mahajan
Graphic Era University, 566/6 Bell Road, Clemet Town, Dehradun, Uttarakhand, India
e-mail: nhkant5@gmail.com

M. Mahajan
e-mail: manish.mhajn@gmail.com

on anomaly discovery for huge dataset in this field. Outlier detection can be applied in many practical applications such as transportation, ecology, public safety, public health, location-based services, etc., since outlier identification provides meaningful and useful information. Recently, researchers have developed an interest in detection of outliers in the time-series databases [2]. Time-series database is simply a database that shows the state of different data objects in terms of time [3]. In this paper, a time-series or temporal data was analysed to detect outliers present in the datasets [4]. Air quality data set was used as a time-series data and it tried to find the anomalies present in the dataset.

We have proposed a new method for detecting outliers using PSO in combination with Enhanced K-means on the air quality time-series datasets. Enhanced K-means was used to improve the efficiency and PSO, so that an optimal solution can be obtained.

Rest of the paper is organized as follows: Sect. 2 describes the methodology used, Sect. 3 describes the proposed algorithm, Sect. 4 explores implementation and result analysis and finally, the work is concluded in Sect. 5 and also Sect. 6 describes some future scopes.

2 Methodology

In the proposed method, outliers are detected using PSO-aided enhanced K-means algorithm. First, we applied K-means algorithm to get desired set of groups or clusters and then to improve the efficiency of K-means, we used WBC (Weight-based approach) and GRUBBS test which was used to calculate the critical value, so that outliers can be detected. Finally, we applied PSO plotbowl algorithm to get more refined set of data points.

2.1 System Architecture

See Fig. 1.

2.1.1 Input Dataset

Air quality data set was used in this paper and it is collected from UCI machine learning repository.
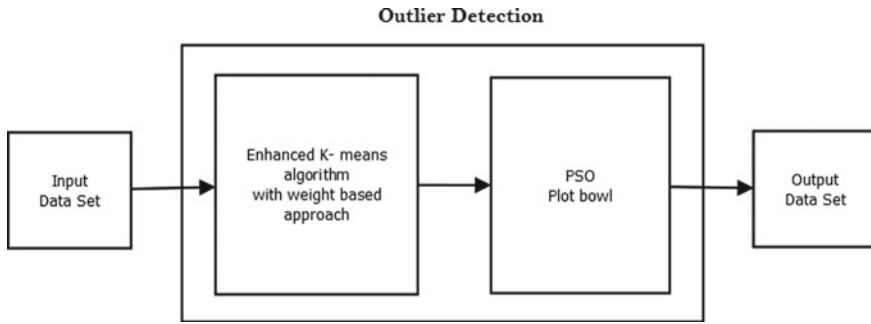
Fig. 1 System architecture

2.1.2 K-Means Algorithm

K-means clustering algorithm is the most widely used clustering algorithm. The k in the K-means refers to the fact that the algorithm is going to look for k different clusters, which means when applied on a data set, the algorithm is going to break the data set into k different clusters [5]. The value of k has to be specified to the algorithm before it starts, so that clusters are known before starting the clustering process.

Working of algorithm [6]:

1. To identify the seeds, i.e. value of centroids and can be chosen randomly.
2. To assign all the other observations to one of the chosen seeds based on their proximity to the seeds.
3. Repeat until convergence.

2.1.3 Enhanced K-Means Algorithm

Enhanced K-means algorithm works in three steps. In the first phase, traditional K-means algorithm was applied to find the number of clusters than Weight-based centre approach with standard deviation was used to find outliers. Enhanced K-means algorithm was used to eliminate some drawbacks of Traditional K-Means [7]. We applied WBC approach and calculated the critical value using GRUBBS test because traditional K-means does not work well with clusters with elongated shaped.

2.1.4 PSO Algorithm

PSO algorithm was used to optimize the result, so that we can get more refined set of data points. PSO is simple yet powerful optimization algorithm, and it is successfully applied to enormous applications in various fields of science and engineering like

machine learning, image processing, data mining, operational research and many other fields [8].

2.1.5   Outlier Detection Using PSO-Aided Enhanced K-Means Algorithm

The main objective of outlier detection is to detect a data object or data point in any data set, which is different from the other data objects in the dataset.

First, we apply K-means algorithm to get the desired number of clusters and outliers found in this step must be removed.

Then, calculate the weight of each cluster using the following equation:

$$Weight_k = \sum_{x=1}^{n} w_x \tag{1}$$

where k is the No. of clusters and w_x is No. of data points (d_1, d_2, d_3, …, d_n) in a particular cluster k.

Next, find the mean value and maximum value for each cluster using the following equations:

$$Mean_k = \sum_{x=1}^{n} w_x \big/ n \tag{2}$$

$$X_{max} = Max(d_1, d_2, d_3, \ldots, d_n) \tag{3}$$

where k is the No. of clusters, w_x is No. of data points in a particular cluster k, n is total No. of data points in any dataset.

Calculate the standard deviation using the following equation:

$$S = \sqrt{\frac{\sum (X_i - Mean_k)^2}{(n-1)}} \tag{4}$$

where X_i is a data item, M_k is a mean of particular cluster and n is total No. of data points in any dataset. Apply GRUBBS test on data points in each cluster [9].

We use the following two equations to find outliers:

$$G_{test} = \frac{(X_{max} - Mean_k)}{S} \tag{5}$$

$$G_{critical} > \frac{(n-1)}{\sqrt{n}} \sqrt{\frac{t_{\alpha/(2n),n-2}^2}{(n-2) + t_{\alpha/(2n),n-2}^2}} \tag{6}$$

where n is the total No. of data points, $t^2_{\alpha/(2n),n-2}$ is the upper critical vale of a t-distribution with n – 2 degrees of freedom.

Compare G_{static} with $G_{critical}$ for each data point.

If $G_{static} > G_{critical}$, then data point is treated as an outlier. If $G_{static} < G_{critical}$, then data point lies in the cluster. All the outliers were removed from the data set and new data set contains the normal points which may be called inliers. Finally, PSO algorithms were applied to optimize the result and to get more outliers.

3   Proposed Algorithm

See Fig. 2.

Algorithm: Time-series Outlier detection using Enhanced K-Means Algorithm in combination with PSO plotbowl Algorithm:

Step 1:   Select input data and No. of clusters.
Step 2:   Calculate distance with each data points and assign each data points to the nearest cluster.
Step 3:   Recalculate the cluster centres.
Step 4:   Repeat the process in Steps 2 and 3 until the convergence criteria is achieved.
Step 5:   Showing the clustering results.
Step 6:   Calculate the weight-based centre for each cluster.
Step 7:   Calculate the mean value, max value and standard deviation for each cluster.
Step 8:   Calculate the Critical value using GRUBBS test.
Step 9:   Compare each data item in a cluster with the critical value.
Step 10:   If the data value is found greater than critical value, then the given data item in a particular cluster is the outlier.
Step 11:   Remove the outliers from the cluster.
Step 12:   Repeat Steps 7–11 s for each resultant cluster.
Step 13:   Randomly initialize particle position $x^i_0 \in D$ in IR^n for $i = 1, …, p$ present in the resultant cluster.
Step 14:   Randomly initialize particle velocities $0 \leq v^i_0 \leq v^{max}_0$ for $i = 1, …, p$ present in the resultant cluster.
Step 15:   Set constants k_{max}, C_1, C_2 and $k = 1$.
Step 16:   Evaluation function value f^i_k using design space coordinates x^i_k.
Step 17:   If $f^i_k \leq f^i_{best}$ then $f^i_{best} = f^i_k$, $p^i_{k} = x^i_k$ and If $f^i_k \leq f^g_{best}$ then $f^g_{best} = f^i_k$, $p^g_{k} = x^i_k$.
Step 18:   If stopping condition is satisfied, then goto Step 22.
Step 19:   Update all particle velocities vik for $i = 1, …, p$.
Step 20:   Update all particle positions xik for $i = 1, …, p$.
Step 21:   Increment k and finally go to Step 16.
Step 22:   Terminate.
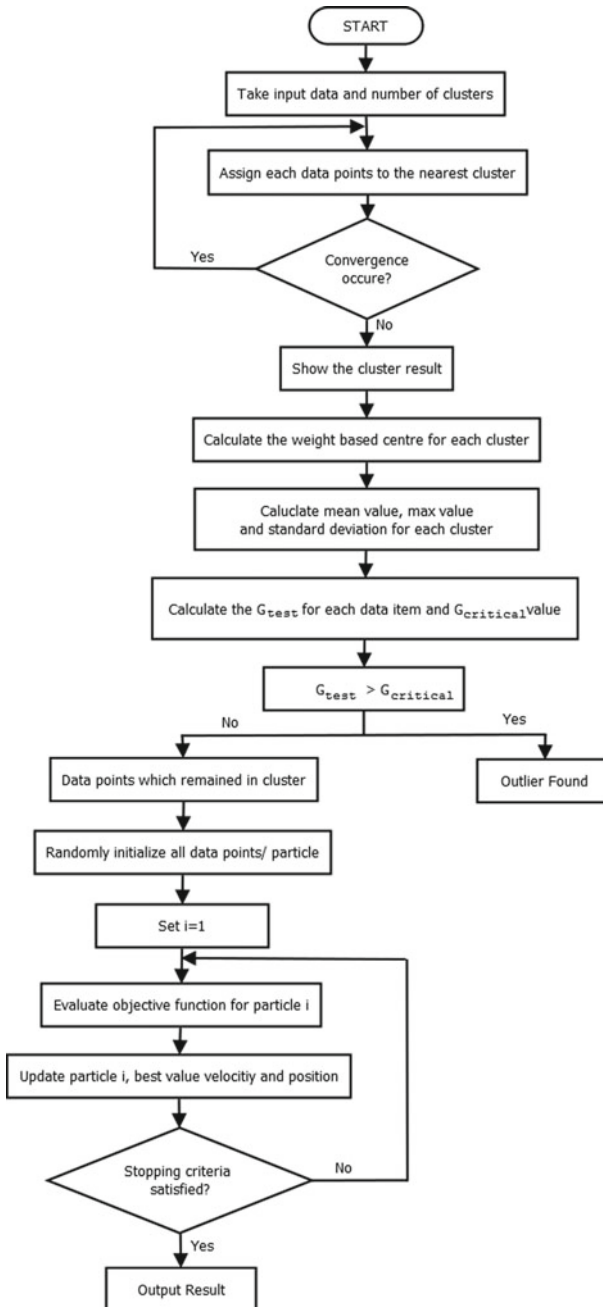
Fig. 2 Flowchart of the proposed algorithm

4 Experimental Result

In this section, air quality data is collected from UCI machine learning repository which has various parameters such as CO, SO_2, O_3, NO_2, etc., and instances.

Results show that the proposed method is better than traditional K-means algorithm for detecting outliers. We got more outliers using a new method than traditional algorithm.

We first applied K-means algorithm and we got eight No. of clusters which is shown in Fig. 3.

We also got 60 No. of outliers using traditional K-means algorithm which is shown in Fig. 4.

Fig. 3 Clusters found after applying K-means
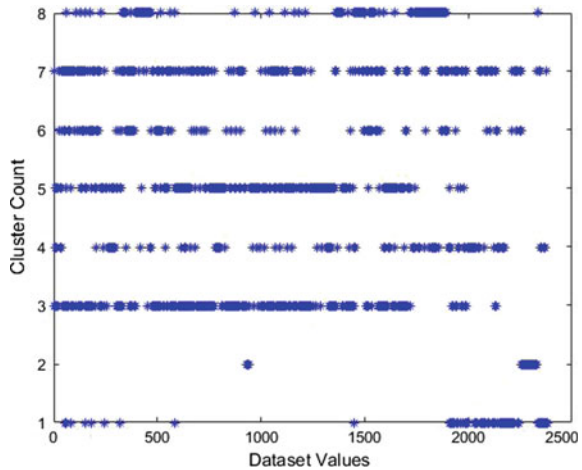
Fig. 4 Outliers found after applying K-means
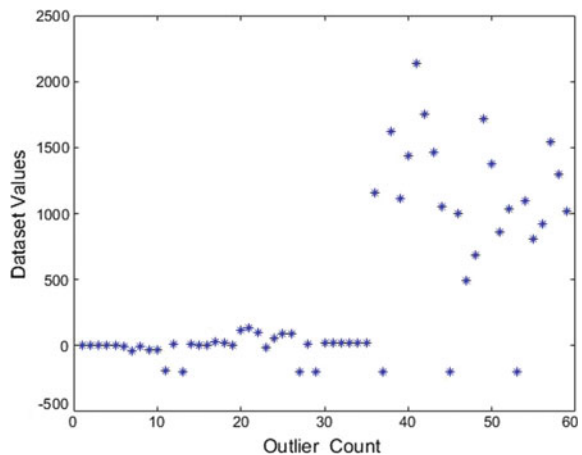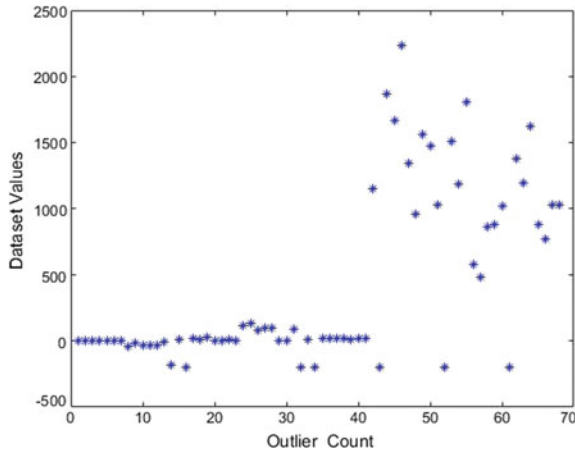
Fig. 5 Outliers found after applying enhanced K-means

Table 1 Comparative study of the proposed and existing algorithms

Algorithm	Data set used	No. of outliers
K-means algorithm	Air quality data set	60
Enhanced K-means algorithm	Air quality data set	68
PSO algorithm	Air quality data set	40
PSO-aided enhance K-means algorithm	Air quality data set	78

Number of clusters we found using K-Means with elongated shaped, and we already know K-means does not properly work with cluster with elongated shaped, so we applied Enhanced K-Means algorithm which properly works on elongated-shaped clusters and empty clusters [10]. Figure 5 shows 68 No. of outliers found using enhanced k-means algorithm which is greater than traditional k-means algorithm.

After finding sufficient no of outliers we applied PSO algorithm on those points which remain in the cluster after applying enhanced k-means. We used PSO to get more refined set of data and found 78 outliers which are more than what we got using enhanced K-Means alone.

Figure 6 shows the final result of PSO-aided enhanced K-means. It shows the global best for data points with state variance and objective variance. State variance and objective variance were used to obtain the critical value for data point.

We also implemented these three algorithms separately and found that PSO-aided enhanced K-means works better than traditional K-means and Enhanced K-means alone. Result in Table 1 shows the number of outliers obtained after applying traditional K-means, Enhanced K-means, PSO and PSO-aided enhanced K-means (Fig. 7).
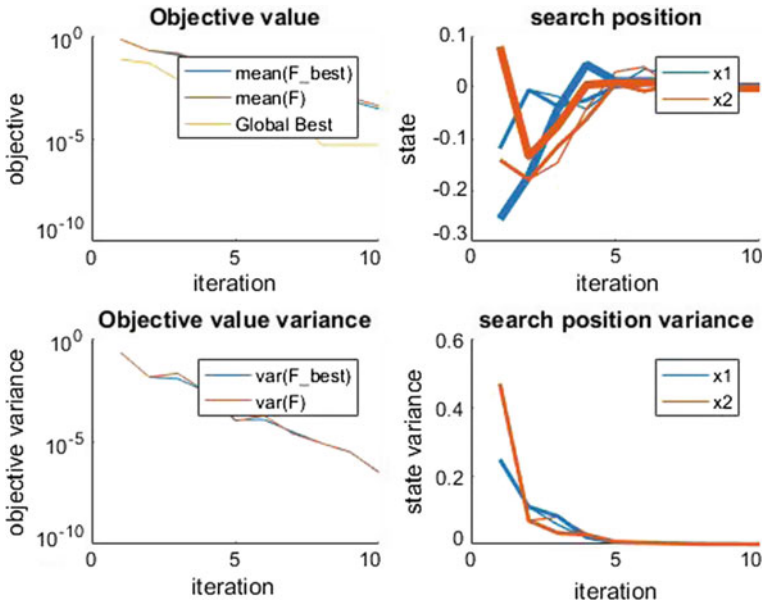
Fig. 6 Result of PSO-aided enhanced K-means
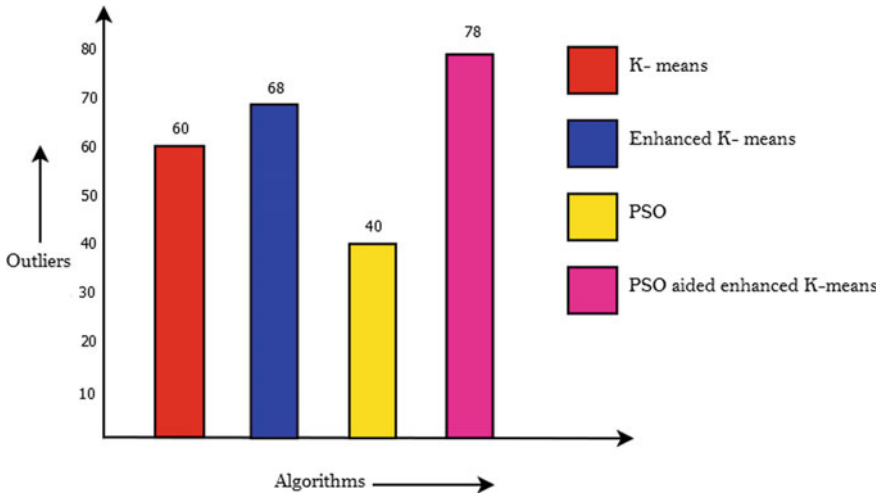
Fig. 7 Comparison between the proposed and existing algorithms

5   Conclusion

Outlier detection has been used in various fields due to new challenges like data set having multi dimensions, No. of fraud cases, etc. Outliers in air quality time-series data have many applications such as finding asthma exacerbation rate in air, pollution percentage in air, effects of smoking, etc. In this dissertation, a time-series or temporal data was analysed to detect outliers present in the datasets. Air quality data set was used as a time-series data and it tried to find the anomalies present in the dataset. We have proposed a new method for detecting outliers using PSO in combination with Enhanced K-means on the air quality time-series datasets. Enhanced K-means was used to improve the efficiency and PSO was used, so that an optimal solution can be obtained. This new technique was applied on air quality dataset and found that this algorithm is efficiently working on it. We proved that the method used in this paper is very useful in identifying outliers in the air quality dataset and also showed some comparative analysis of our work with Traditional K-means. The dissertation also proved that this method identifies a sufficient number of outliers presented in the dataset. This new method can be applied in different domains like healthcare, fraud detections, levels of organic compound found in the air, network-wide traffic. This paper also proved that this method identifies sufficient number of outliers present in the dataset.

6   Future Work

This new method applies only to numerical data. Further modifications in this method can be done by making it applicable to text mining as well. The future work has a scope for varying datasets.

References

1. Kamble, B., Doke, K.: Outlier detection approaches in data mining. Int. Res. J. Eng. Technol. (IRJET, March) **3**(3) 2017
2. Zhang, J.: Advancement of outlier detection: a survey. ICST Trans. Scalable Inf. Syst. (2013)
3. Ma, J., Perkins, S.: Online Novelty Detection on Temporal Sequences. Aureon Biosciences corporation Yonkers NY, USA (2003)
4. Gupta, M., Gao, J., Aggarwal, C.C., Han, J.: Outlier detection for temporal data: a survey. IEEE Trans. Knowl. Data Eng. (2014)
5. Yadav, J., Sharma, M.: A review of K-mean algorithm. Int. J. Eng. Trends Technol. (IJETT) **4**(7) 2013
6. Manoharan, J.J., Ganesh, D.S.H.: Initialization of optimized K-means centroids using divide-and-conquer method. ARPN J. Eng. Appl. Sci. **11**(2) (2016). ISSN 1819–6608

7. Kanungo, T., Mount, D.M., Netanyahu, N.S., Piatko, C.D., Silverman, R., Wu, A.Y.: An efficient k-Means clustering algorithm: analysis and implementation. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7) (2002)
8. Mohemmed, A.W., Zhang, M., Browne, W.: Particle swarm optimisation for outlier detection. School of Engineering and Computer Science, Victoria University of Wellington (2010)
9. Grubbs, F.E.: Procedures for detecting outlying observations in samples. Am. Stat. Assoc. Technometr, **11**(1) (1969)
10. Manoharan, J.J., Ganesh, S.H.: A framework for enhancing the efficiency of k-means clustering algorithm to avoid formation of empty clusters. Middle-East J. Sci. Res (MEJSR)

Asymmetric Image Encryption Using Gyrator Transform with Singular Value Decomposition

Jaideep Kumar, Phool Singh, A. K. Yadav and Anoop Kumar

Abstract   Motivated by the endurance of phase truncated asymmetric optical cryptosystems against basic attacks, single-channel encryption technique for a color image is proposed, that uses singular value decomposition, and affine transform in gyrator domain. Color 8 images have been used to validate the scheme. The input color image is split into its indexed components-RGB (red, green and blue). The red component of the input image is used as plaintext, and the other two components are used as phase in the encryption scheme. Experimental results presented in the paper are based on computation on MATLAB. The results show that the scheme endures occlusion and noise attacks. The simulation also confirmed the scheme's sensitivity to the encryption parameters of affine transform and gyrator transform. The overall results show that the proposed encryption scheme exhibits high levels of security.

Keywords   Affine transform · Gyrator domain · Singular value decomposition

1   Introduction

Over the past few years, there has been an exponential increase in the Internet usage. Although, with enhanced usage, better security methods have been developed and improvised, miscreants are always looking for ways to break these cryptographic methods or protocols. In order to minimize damages, the computer scientists and cryptographers are pushing their limits to develop highly secure solutions to over-

J. Kumar (✉) · A. Kumar
School of Engineering and Technology, K R Mangalam University, Gurugram, India
e-mail: jaideep.rohtak@gmail.com

P. Singh
Department of Mathematics, Avvaiyar Government College for Women,
Karaikal, Puducherry, India

A. K. Yadav
Department of Applied Mathematics, Amity School of Applied Sciences, Amity University
Haryana, Gurgaon, India

come these threats and problems. Although the digital security systems have their own limitations, which are constantly addressed by adding physical parameters to increase the security. Optical cryptosystems have multiple degrees of freedom and offer fast processing and reliable security. Double-random-phase-encoding (DRPE) is reportedly the first optical image encryption scheme proposed by Refregier and Javidi [1]. After that, several optical encryption systems have been proposed which are primarily based on DRPE and use different optical setups [2–4]. Ongoing efforts of improving security and examining the vulnerability of schemes for basic attacks such as known-plaintext attack (KPA) [5], chosen-plaintext attack (CPA) [6], ciphertext-only attack (COA) [7], etc., reveal that DRPE is not safe as it is a symmetric cryptosystem. In a symmetric cryptosystem, both encryption as well as decryption keys are the same whereas in asymmetric cryptosystem, encryption keys are different from the decryption keys. Qin and Peng [8] proposed a phase-truncated Fourier transform (PTFT)-based asymmetric cryptosystem. They generated real-valued ciphertext, which looks like a stationary white noise. Nonlinear operation in phase truncation makes a cryptosystem resistant against KPA and CPA. Aburtab [9] proposed an asymmetric cryptosystem for color images in gyrator transform (GT), which is based on phase-truncation and an optical-coherent superposition method. The proposed optical design is free from axial movement.

To increase security further, researchers used chaotic maps to permute the input images. Sui et al. [10] proposed multiple image asymmetric encryption in fractional Fourier transform domain that is based on coupled logistic maps. This paper proposes a secure color image asymmetric encryption algorithm that uses singular value decomposition and affine transform in gyrator domain. The affine transform is used in shuffling the positions of the image pixels in the spatial domain.

2 Related Background

2.1 Affine Transform

Affine transform (AT) is one of the pixel-permuting operations, which randomly shuffles the pixel position in any given image. Mathematically, it can be explained as [11]

$$\left(u', v'\right) = AT\{(u, v), N\} = (p + ru, q + sv) \, mod \, M \tag{1}$$

where (u, v) represents a pixel position of an image of size $M \times M$ pixels, $\left(u', v'\right)$ is the pixel position obtained after affine transform, mod is the modulo operation, p and q are two random numbers having values between 1 and M, and r and s are relative prime to M. After the affine transform, pixel value remains the same and the total energy of the plaintext remains unchanged.

(a) **(b)** **(c)**
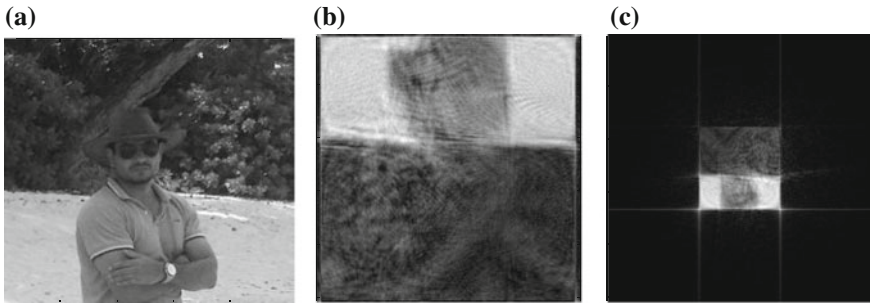
Fig. 1 Gyrator transform of grayscale image of **a** Man with angle, **b** 3.14 and **c** 1.25

2.2 Singular Value Decomposition

Singular value decomposition (SVD) [12] is one of the most robust and reliable methods for matrix decomposition. For a matrix A of size $r \times s$, there exist orthogonal matrices U and V, of sizes $r \times r$ and $s \times s$, respectively, such that $A = USV^T$, where V^T is the transpose of V and $S = diag(a_1, a_2, \ldots a_i \ldots, a_t)$, a_i being eigenvalues ($a_i \geq a_{i+1}$, where $i = 1, 2, \ldots, r - 1$, and $t = min(m, n)$) of A. The SVD method is successfully used in watermarking and digital image encryption.

2.3 Gyrator Transform

Gyrator transform, a linear canonical transform is used for image encryption [9, 13]. The gyrator transform for an image I(u, v) is defined as [14]

$$G(s, t) = g^\beta[I(u, v)](s, t)$$
$$= \frac{1}{|\sin \beta|} \iint I(x, y) \exp\left[i2\pi \frac{(st + uv)\cos \beta - (ut + vs)}{\sin \beta}\right] dudv \quad (2)$$

where β is a gyrator angle parameter. The function G(s, t) is the output of the transform. Figure 1b, c show the effect of gyrator transform of an input image of Man (Fig. 1a) of size 256×256 pixels.

3 Proposed Scheme

The proposed scheme is presented through a schematic diagram in Fig. 2. In the encryption process (Fig. 2a), an input color image is first decomposed into its indexed channels namely red, green, and blue, and then subjected to affine transform with w

(a)

(b)
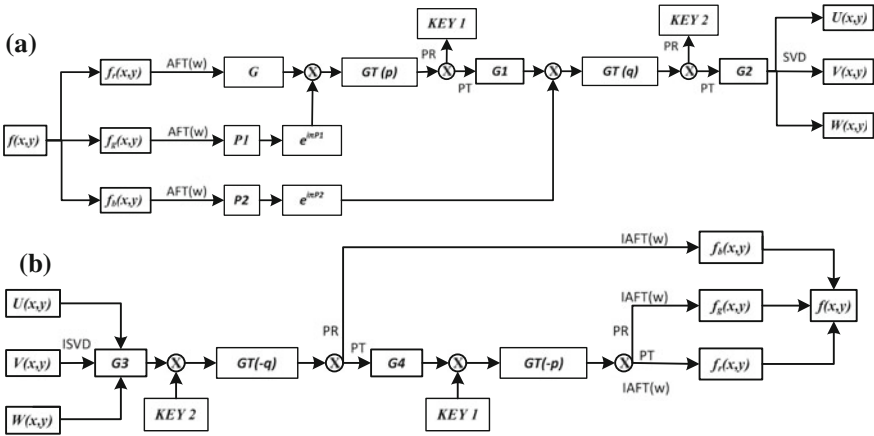
Fig. 2 Flowchart of the proposed scheme, **a** encryption process; **b** decryption process

iterations. The transformed red channel image is considered as an amplitude image whereas green and blue channels are regarded as phase masks. The red channel image is bonded with green channel phase mask and transformed using a p-order gyrator transform (GT). The phase-reserved part of the resulting image will serve as decryption key 1, whereas the phase-truncated part $G1$ is bonded with the blue channel phase mask and followed by another gyrator transform of order q. The new phase truncated part $G2$ is decomposed into three components (U, S, V) using singular value decomposition, whereas the new phase-reserved part will serve as decryption key 2. These three components of the encrypted image are then decrypted using the decryption process displayed in Fig. 2b.

4 Results and Discussion

The proposed scheme for color images has been validated through computer simulations. Results are obtained through the computations performed on MATLAB 7.14 (Intel-i7 processor, 16 GB RAM, 3.40 GHz CPU speed, Windows 8, 64-bit operating system), for an input color image of Man shown in Fig. 3a of size $256 \times 256 \times 3$ pixels. The original image is decomposed in three channels namely red, green, and blue (Fig. 3b shows red channel of the original image). Figure 3c shows the affine transformed image of red channel and following the scheme given in Fig. 2a, we get the encrypted image components (Fig. 3d–f). The decrypted image obtained using the proposed scheme is the same as the input image. We observe that the encrypted image components U and V are completely random and appear as stationary white noise, except for S component which is diagonal dominant. The values of parameters of the AT and GT taken in this study are $w = 30$, and $p = 0.6$, $q = 0.4$, respectively.
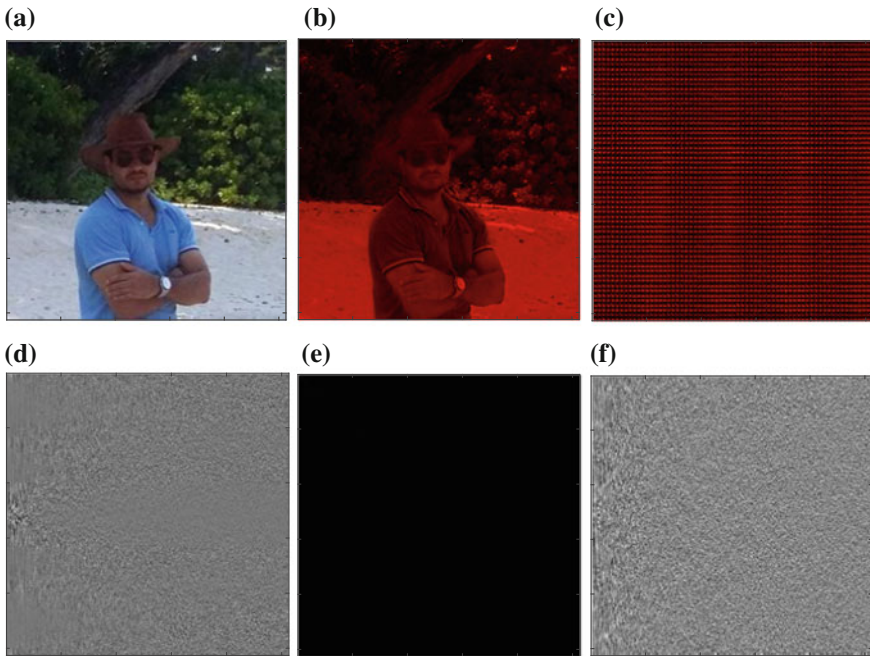
Fig. 3 Results of validation of the scheme for **a** input color image (Man) 256 × 256 × 3 pixels; **b** is red channel of (**a**); **c** is affine-transformed image of (**b**); **d**–**f** are encrypted images

4.1 Statistical Analysis

The effectiveness of an image encryption scheme can be analyzed from 3D plots of the encrypted images generated by the scheme. Our scheme provides encrypted images (ciphertext) in component form, whose 3D plots appear to be fairly random. Figure 4c gives 3D plot of the U component of encrypted image and when compared with 3D plot of the corresponding component of input image (Fig. 4a), it appears entirely different and random. Figure 4b shows the affined transformed 3D plot of the input image, which also establishes the efficacy of affine transform.

Another metric used for measurement of the encryption efficacy is the correlation distribution of adjacent pixels of the encrypted image in three directions, namely horizontal, vertical, and diagonal. For this, 5,000 pairs of adjacent pixels are arbitrarily selected from the red channel of the input color image of Man and U component of its encrypted image in the horizontal direction. Figure 5a displays correlation distribution for the plaintext in the horizontal direction whereas Fig. 5b displays the corresponding plot of U component of the encrypted image. It is evident from the figures that there is a total loss of correlation in the encrypted image, which resembles a random distribution. Hence, the present scheme provide effective encryption, as ascertained by the above statistical analysis.
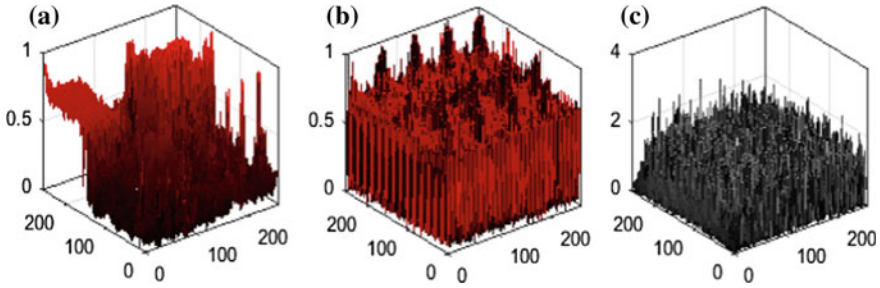
Fig. 4 3D plots of **a** red channel of the input image; **b** 3D plots after affine transform of (**a**); **c** encrypted image (*U* components of singular value decomposition)
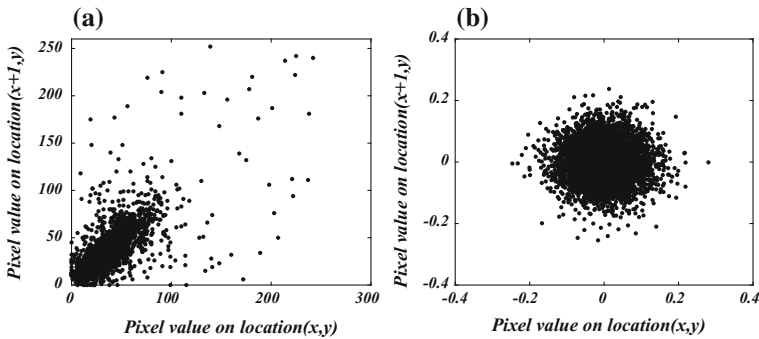
Fig. 5 Correlation distribution in horizontal direction of the plaintext (**a**) and its encrypted image's *U* component (**b**)

4.2 Key Sensitivity Analysis

If a scheme is highly sensitive to its encryption parameters then it is considered to be secure. In our scheme, the encryption parameters involve the gyrator angles *p* and *q* of the gyrator transform and parameter of the affine transform (w). The sensitivity analysis has been performed on these above mentioned parameters and the results are presented in plots of mean squared error (MSE), which is defined as follows:

$$\text{MSE} = \frac{1}{M \times M} \sum_{u=1}^{M} \sum_{v=1}^{M} |I_o(u, v) - I_r(u, v)|^2 \tag{3}$$

where $I_r(u, v)$ and $I_o(u, v)$ represent the pixel values of the recovered image and the original image, respectively. $M \times M$ is the size of the input image. Plots of MSE versus affine parameter, and orders of gyrator transform for the indexed channel red of Man reveal that scheme is sensitive to very minute changes in the values of these parameters (Fig. 6). These results establish the scheme's sensitivity to the encryption parameters, namely, the orders of *AT* and G*T*.
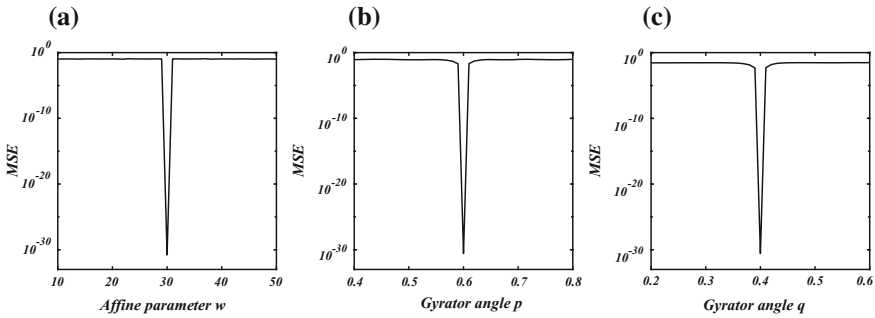
Fig. 6 Plots of sensitivity (MSE) relative to **a** parameter *w* of affine transform (AT); **b**, **c** parameters p and q of gyrator transform (GT) respectively
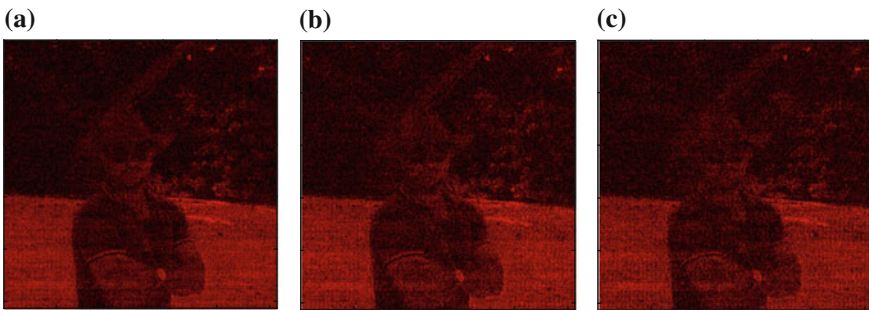
Fig. 7 Recovered images when the encrypted image has salt and pepper noise of strength (**a–c**) $k = 0.3$, 0.6, and 0.9, respectively

4.3 Robustness to the Added Noise Data

Robustness of an encryption scheme can be verified by testing the scheme against various attacks such as noise attack, occlusion attack, known-plaintext attack, chosen-plaintext attack, etc. We have tested the robustness of the proposed scheme against salt and pepper noise attack by varying degrees of noise strength [12, 15].

Figure 7a–c provide recovered images for the red channel when the encrypted image components subjected to inverse singular value decomposition, are affected by the noise present in the channel with growing strength respectively $k = 0.3$, 0.6, and 0.9. It is observed that the quality of the decrypted image is reasonably good, even in presence of substantial noise in the encrypted image. Similar results were also obtained for the other two channels (green and blue). Thus, we have demonstrated that the scheme is fairly resistant to by the noise attack.
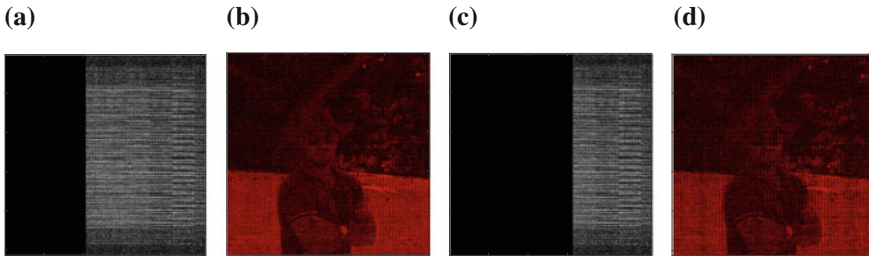
Fig. 8 Occlusion results for the encrypted images (after applying inverse singular value decomposition) for (**a**, **c**), respectively, with 40% and 60% blockage; figure (**b**, **d**) show the corresponding decrypted images of Man

4.4 Robustness Against Missing Data of the Encrypted Image

The present study is further investigated for its robustness to lost data of the encrypted image of Man when the encrypted image components are subjected to inverse singular value decomposition. Figure 8a, c show the encrypted images of Man with data loss of 40% and 60%, and the corresponding recovered images are, respectively, provided in Fig. 8b and 8d. When 40% and 60% of the data of encrypted images are blocked, the original images can be recovered with correlation coefficient 0.8695 and 0.7783, respectively. Although the quality of decrypted image drops with the increase in missing encrypted data, it is still identifiable even when missing encrypted data is as high as 60%. The proposed scheme spread the signature of the original image over the entire encrypted image, thereby providing robustness against the loss of data.

5 Conclusions

We have proposed a new asymmetric encryption technique for single-channel color images based on phase and amplitude truncation operations. It uses an affine transform for pixel scrambling of each channel of the original color image, and singular value decomposition after applying two gyrator transforms in spatial and frequency domain for encryption. The scheme has been validated for color images through simulations on MATLAB (2017b). Its effectiveness is investigated through statistical analysis based on 3D plots, and correlation distribution. The scheme also demonstrates its endurance to the widely reported attacks of occlusion and noise. Investigation of sensitivity analysis shows that the new scheme is highly sensitive to its encryption parameters such as the transform orders and the decryption keys obtained from amplitude truncation.

References

1. Refregier, P., Javidi, B.: Optical image encryption based on input plane and Fourier plane random encoding. Opt. Lett. **20**, 767–769 (1995)
2. Sharma, N., Saini, I., Yadav, A., Singh, P.: Phase-image encryption based on 3D-Lorenz chaotic system and double random phase encoding. 3D Res. **8**, 1–17 (2017)
3. Singh, P., Yadav, A.K., Singh, K., Saini, I.: Optical image encryption in the fractional Hartley domain, using Arnold transform and singular value decomposition. AIP Conf. Proc. **1802**, 020017 (2017)
4. Yadav, A.K., Singh, P., Singh, K.: Cryptosystem based on devil's vortex Fresnel lens in the fractional Hartley domain. J. Opt. (2017). https://doi.org/10.1007/s12596-017-0435-9
5. Peng, X., Zhang, P., Wei, H., Yu, B.: Known-plaintext attack on optical encryption based on double random phase keys. Opt. Lett. **31**, 1044–1046 (2006)
6. Carnicer, A., Montes-Usategui, M., Arcos, S., Juvells, I.: Vulnerability to chosen-cyphertext attacks of optical encryption schemes based on double random phase keys. Opt. Lett. **30**, 1644–1646 (2005)
7. Jiao, S., Li, G., Zhou, C., Zou, W., Li, X.: Special ciphertext-only attack to double random phase encryption by plaintext shifting with speckle correlation. JOSA A. **35**, A1–A6 (2018)
8. Qin, W., Peng, X.: Asymmetric cryptosystem based on phase-truncated Fourier transforms. Opt. Lett. **35**, 118–120 (2010)
9. Abuturab, M.R.: Color information cryptosystem based on optical superposition principle and phase-truncated gyrator transform. Appl. Opt. **51**, 7994–8002 (2012)
10. Sui, L., Duan, K., Liang, J., Zhang, Z., Meng, H.: Asymmetric multiple-image encryption based on coupled logistic maps in fractional Fourier transform domain. Opt. Lasers Eng. **62**, 139–152 (2014)
11. Singh, P., Yadav, A.K., Singh, K.: Color image encryption using affine transform in fractional Hartley domain. Opt. Appl. **47**, 421–433 (2017)
12. Singh, P., Yadav, A.K., Singh, K.: Phase image encryption in the fractional Hartley domain using Arnold transform and singular value decomposition. Opt. Lasers Eng. **91**, 187–195 (2017)
13. Mehra, I., Nishchal, N.K.: Optical asymmetric image encryption using gyrator wavelet transform. Opt. Commun. **354**, 344–352 (2015)
14. Liu, Z., Guo, Q., Xu, L., Ahmad, M.A., Liu, S.: Double image encryption by using iterative random binary encoding in gyrator domains. Opt. Express **18**, 12033–12043 (2010)
15. Anjana, S., Saini, I., Singh, P., Yadav, A.K.: Asymmetric cryptosystem using affine transform in Fourier domain. In: Bhattacharyya, S., Chaki, N., Konar, D., Chakraborty, U.K., Singh, C.T. (eds.) Advanced Computational and Communication Paradigms. Springer Singapore, Singapore (2018)

Identification of Empirical Model and Tuning of PID Controller for a Level Control System

Moby S. Philip, Bipin Krishna and S. Meenatchisundaram

Abstract Control of level in a process tank is most common in process and chemical industry. Development of model based control algorithms provides a better control under nonlinearities and disturbances. Identification is the process of estimating the model under given boundary conditions. Data driven empirical models are relatively easy to estimate and best suited within the process limitations. This paper deal with the design of an empirical model for a level control process and to tune the PID controller using different tuning method. The tuning method are compared between each other and the best control parameters are selected by analyzing the response.

Keywords Empirical model · Identification · Model estimation · PID tuning

1 Introduction

There are various parameters that an industry depends on like, flow, level, pressure and temperature. All the parameters provide a great impact on the productivity. For a better optimizing and better running, these parameters should be taken care with utmost precision and accuracy. In this paper, the model of a level process is obtained using experimental data (empirical model) and the model is used to tune the PID controller parameters.

PID controllers had a great application in the modern industries in the field of process control [1]. The execution of the feedback control system can be enhanced by legitimately tuning the controller parameters. If the tuning is uncalled for, at that point

M. S. Philip · B. Krishna (✉) · S. Meenatchisundaram
Department of Instrumentation and Control Engineering, Manipal Institute of Technology,
Manipal Academy of Higher Education, Manipal 576104, Karnataka, India
e-mail: bipin.monayi@gmail.com

M. S. Philip
e-mail: mobysphilip2016@gmail.com

S. Meenatchisundaram
e-mail: meenasundar@gmail.com

the closed-loop system deals with instability. So basically, tuning is fundamental for continuously monitoring the level. This paper dealt with the five different tuning methods that are Ziegler–Nichols for closed-loop method, Cohen–Coon method, Chein–Hrones–Reswick method, Wang–Juang–Chan method and AMIGO method [2]. The empirical identification strategy gives us the dynamic connection between the given input variables and obtained output variables.

The process of empirical identification is done by six-step procedures. This technique guarantees that legitimate information is produced through cautious exploratory outline and execution. This empirical technique includes planned experiments, amid which the procedure is given perturbations to create dynamic information. The achievement of these techniques requires close adherence to standards of trial outline and model fitting. Generally, there are two identifications techniques, the first technique is the, graphical procedure which is termed as process reaction curve which is a graphical procedure, and second is a Two-point method which is the common method to determine the parameters. In this paper we have went through only process reaction curve. The first method is to determine the empirical model which can be determined by either by PRC method or using the Two-point method. In this paper, we went through two different procedures which had been enlisted below:

- Determination of the empirical model.
- Tuning with the help of different PID tuning procedure.

2 Methodology

2.1 Determination of the Empirical Model

Determining the model is a method to obtain the numerical relation of the system dynamics with the help of the system input–output information [2]. For the given SISO process identification of the process follows a straight path, therefore, several different techniques have been mentioned in many literatures [3]. For determining the model having the system which is linear, and does not depend on the time that is the system which is linear time-invariant system (LTI) we went through the technique that is mentioned in [4, 5]. For determining the model that are nonlinear Hammerstein and Weiner [6, 7] are well acknowledgeable because of its widely use in day-to-day life. Here, we are going through two different procedures that had been enlisted above that is:

- PRC method.
- Two-point method.
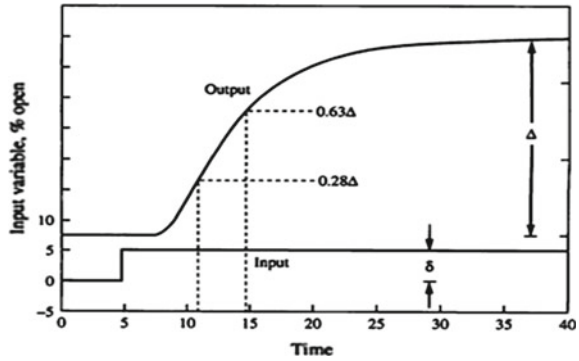
2.1.1 Process Reaction Curve (PRC).

The process reaction curve is likely the most generally utilized technique for distinguishing dynamic models. This method is easy to perform, and it provides an accurate model [8].

The process reaction curve in this paper was done in a four-step common procedure that had been enlisted below

- Allow the system to stabilize at the specified set point.
- After the system stabilizes, provide a small perturbation in the input.
- Allow the system to stabilize at the given assigned perturbation.
- Gather all the input and output data after the system gets stabilized and perform the graphical process reaction curve calculations.

The model that we had obtained is in the form of X(s) as an input variable and Y(s) as an output variable.

$$\frac{Y(s)}{X(s)} = \frac{K_p e^{-\theta s}}{\tau s + 1}$$

2.1.2 Two-point Method

This method is completely a graphical approach where all the parameters for calculating the parameter of the FOPDT model will be obtained using the graphical approach. Here, we must plot two points one is at 63.2% of the total time and other is 28.3% of the total time [8]. Figure 1 shows the actual plot.

In this method, the intermediate parameters that we obtained from Fig. 2 is

δ = Magnitude of the input change.

Δ = Magnitude of steady state change in the output.

And, we have to monitor the time at which our output reaches the 28.3 and 63.2% of its final value. Now, to obtain the time at which the output reaches 28.3 and 63.2%

Fig. 2 Schematic of level control
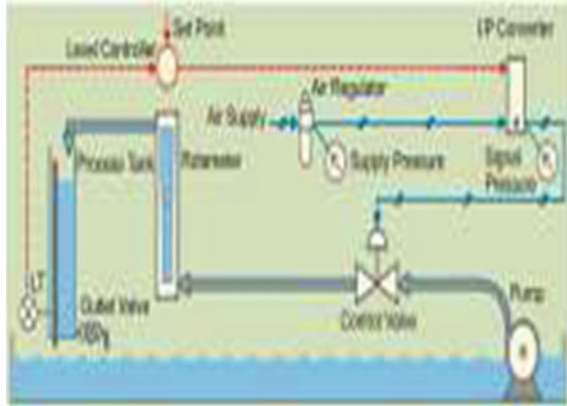
of the given final value. Let the time at 28.3% denote as t_1 and the time at 63.2% denote as t_2, so the respective time we can obtain is:

$$\tau = 1.5(t_2 - t_1), \theta = t_2 - \tau. K_p = \frac{\Delta}{\delta}$$

So, now with the help of t_1 and t_2, we can obtain the value of τ, θ and K_p.

2.2 Tuning of PID Controller

In the process control field, there are so many tuning procedures exit to tune the given controller, but all comes with a great height of disadvantages and with some drawbacks [9]. As of now, we have to find the optimal gain value of PID and this is termed as tuning of the PID parameter. It means that using this tuned parameter, we can easily get our desired output in a perfect manner.

Here, in this paper, we had discussed about six controller tuning methods and at last, we have compared all the tuning methods and we will see which tuning method suits our level control trainer the best. The different types of tuning methods are listed below.

(1) Ziegler–Nichols tuning.
(2) Cohen–Coon tuning.
(3) Wang–Juang–Chan tuning.
(4) Chien–Horns–Reswick tuning.
(5) AMIGO (Approximated M-Constrained Integral Gain Optimization).
(6) Real-time tuning.

2.2.1 Ziegler–Nichols Tuning

This tuning method can only be done in closed-loop response only. This method is proposed by Ziegler and Nichols in 1942 [10]. The above method will not work for unstable process. It is done by keeping the integral and derivative gain to zero. After keeping both the parameter to zero, the value of proportional gain is increased from zero to until the system reaches the ultimate gain value and output keep o oscillating without having undershoot or overshoot. After performing this step, we can calculate the value of ultimate proportional gain that is K_u, oscillation time period that is T_u. Using this two parameters, we can easily obtain the value of proportional gain K_p, integral gain K_i and derivative gain K_d. The value of this parameter is listed below:

$$K_p = 0.588 K_u, \ K_i = 0.5 T_u, \ K_d = 0.125 T_u$$

2.2.2 Cohen–Coon Tuning

As the working of Ziegler–Nichols is not perfect for the process like flow, pressure because this method requires a very fast settlement. The Cohen–Coon method works well for all self-regulating process and the system which provides fast response. The parameter of PID obtained using Cohen–Coon method is listed below

$$K_p = \left(\frac{1}{k}\right)\left(\frac{\tau}{t_d}\right)\left(\frac{4}{3} + \frac{t_d}{4\tau}\right) \quad T_i = t_d \left(\frac{32 + 6\frac{t_d}{\tau}}{13 + 8\frac{t_d}{\tau}}\right) \quad T_d = t_d \left(\frac{4}{11 + 2\frac{t_d}{\tau}}\right)$$

2.2.3 Wang–Juang–Chan Tuning

According to optimum ITAE criterion, the tuning procedure which is obtained by Wang, Juang and Chan provides efficient parameter of the PID. If we have the value of k, L and T parameters of the given plant, then

$$K_p = \frac{\left(0.7303 + \frac{0.5307T}{L}\right)T + 0.5L}{k(T + L)} \quad T_i = T + 0.5L \quad T_d = \left(\frac{0.5LT}{T + 0.5L}\right)$$

2.2.4 The Chien–Horns–Reswick

This method gives priority to the set point responses and to the development of the disturbances process. To use this tuning method, we should know the values of τ, k and L and using this parameter, we can easily obtain the PID parameter

$$K_p = \left(\frac{0.6\tau}{kL}\right) \quad T_i = \tau \quad T_d = 0.5L$$

Fig. 3 Real-time level
trainer kit
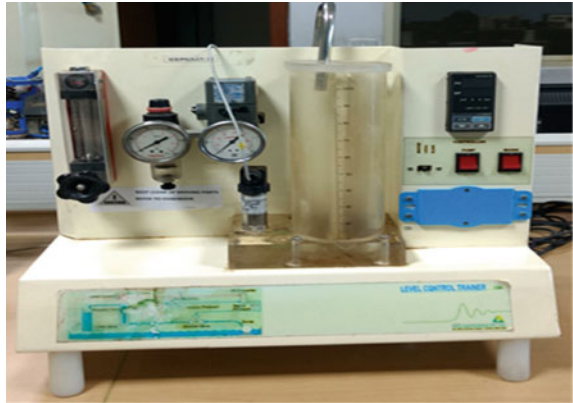
2.2.5 AMIGO (Approximated M-Constrained Integral Gain Optimization)

This method will be proceeded when and only when the given system can be implemented as a first-order plus time-delay (FOPDT). Now, to use the AMIGO method first, we should know the values of τ, L and with the help of this, we can get the parameter of the PID.

$$K_p = 0.2 + 0.45\left(\frac{\tau}{L}\right) \quad T_i = L\left(\frac{0.4L + 0.8\tau}{L + 0.1\tau}\right) \quad T_d = \left(\frac{0.5L\tau}{0.3L + \tau}\right)$$

2.2.6 Real-Time Tuning

As we went more briefly to calculate precisely the value of PID parameter, we can calculate effectively using a real-time MATLAB command, and we took the help of the "PID TUNER".

PID Tuner provides a fast and widely applicable single-loop PID tuning method for the Simulink PID Controller blocks. With this method, you can tune PID controller parameters to achieve a robust design with the desired response time.

3 Real Model Implementation

Figure 2 provides a small glimpse of the process that will be happening in the level control trainer and Fig. 3 represent the actual system.

The given setup consists of supply water tank fitted with pump for the circulation of the water. The level transmitter that we are using is fitted on transparent process tank. A capacitance type, two wire range, 0–300 mm, head-mounted transmitter transmits
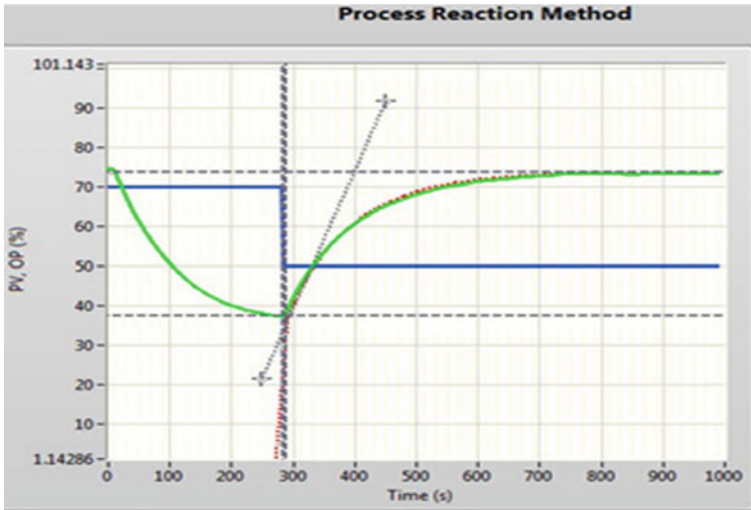
Fig. 4 Process reaction curve

signal in the form of 4–20 mA to interfacing unit. The process parameter that is level controlled using microprocessor-based indicating controller, which manipulates pneumatic control valve through I/P converter. The pneumatic control valve adjusts the flow of water to the tank. The controller is connected to the computer through USB port.

3.1 Real Plant Identification

The implementation of the level control has been done in open-loop process and we obtain the corresponding curve, which is termed as process reaction curve which has been shown in Fig. 4 (Fig. 5).

The parameters that we should obtain with the help of the given graph are enlisted below:

- At 70% output, the value of PV% = 47.3
- At 50% outpu,t the value of PV% = 73.5
- Slope = 0.34882
- Time Delay = 4.90573 s

Fig. 5 Step response of the
FOPDT model
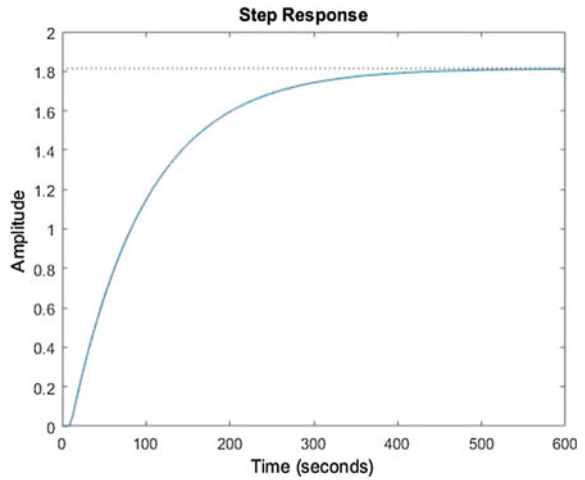
Thus, the FOPDT model using the process reaction curve is show below

$$e^{-4.905732\,s}\left(\frac{1.81429}{104.023s + 1}\right)$$

When we try to implement it through MATLAB by providing a step input, we
gather some information like time-domain specification and the step response of the
delay process.

- Rise Time: 228.5408 s, Settling Time: 411.8490 s
- Settling Minimum: 1.6408, Settling Maximum: 1.8135
- Overshoot: 0%, Undershoot: 0%
- Peak: 1.8135, Peak Time: 809.5837 s

The step response of the given FOPDT model in MATLAB is shown in Fig. 6. As we had to find the value of slope and time delay, now it is time to find the optimized value of all the controller parameter using the different tuning procedures.

3.2 Implementing the Tuning to the Given PID Controller Parameters

The tuning of PID controller means that we must find the corresponding optimized values K_p, K_i, and K_d that will suit best for our level control system and provide the better result of our desired set point. K_p, K_i, and K_d values are found using different tuning methods and are tabulated in Table 1. Figures 6, 7, 8, 9, 10 and 11 shows the step response for each tuning methods.

The whole process was implemented in both in MATLAB and SIMULINK and we observed the different time specifications of different tuning procedures by providing a small step input change to the system.

The given Table 2 compares all the tuning methods with the help of time specification. From the given Table 1, we can predict which tuning method will be best suited for our given level control trainer. For the best optimizing process, the value of rise time and settling time should be less and for the % overshoot, its value should lie below 12%.
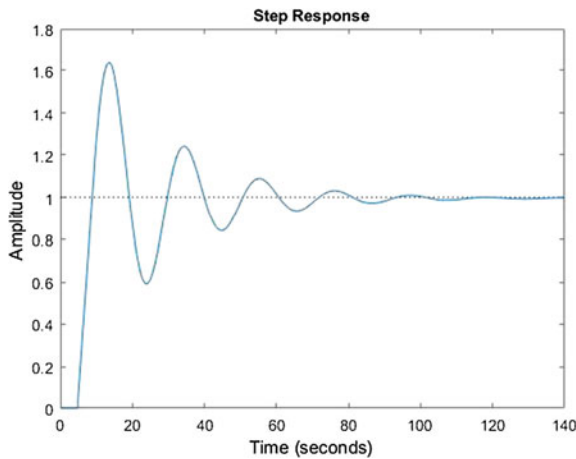
Fig. 6 Ziegler–Nichols tuning response

Table 1 PID controller parameters

Tuning method	K_p	T_i	T_d
Ziegler–Nichols	14.0249	9.81164	2.452866
Cohen–Coon	15.721	11.83884	4.4829
Wang–uang–Chan	2.88387	95	4.7368
Chien–Horns–Reswick	2.976	90	5
AMIGO	4.25	40	4.8387
PID tuner	0.5417	0.0081343	0
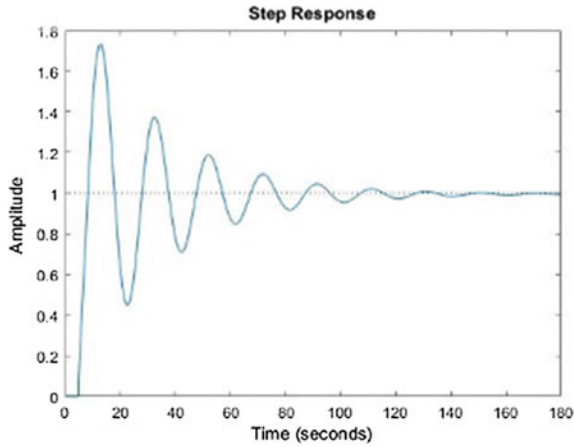
Fig. 7 Cohen–Coon tuning response
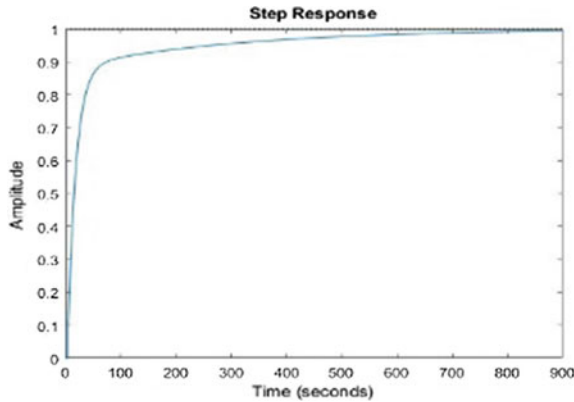
Fig. 8 Wang–Juan–Chan tuning response
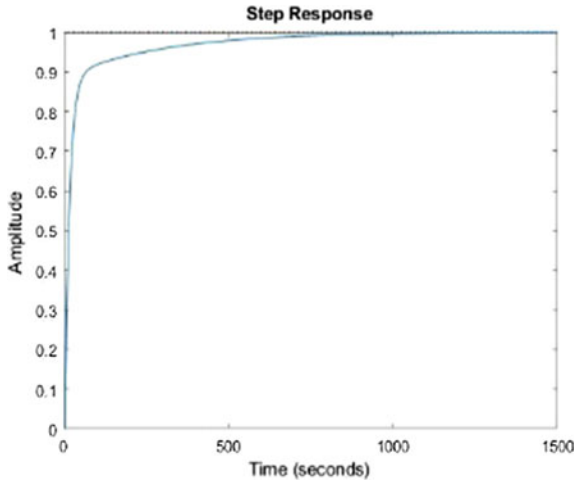
Fig. 9 Chien–Horns–Reswick tuning
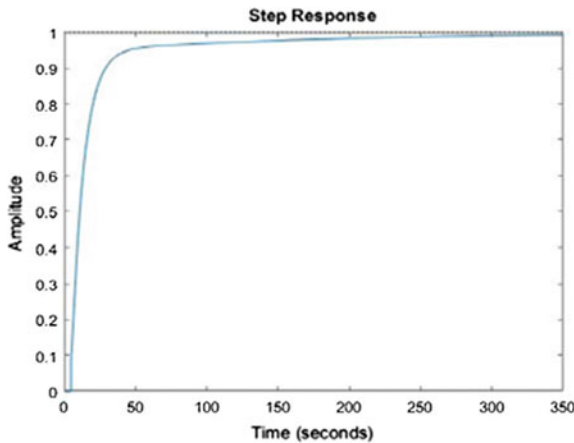
Fig. 10 AMIGO tuning response

4 Conclusion

The maintaining of the level process is the one of the important parameters that should be taken care of for better running of the plant. In this paper, we had to differentiate the different tuning method and studying their step response in the given real-time level control trainer. So, after going through all the MATLAB responses and scope result and at last, we can conclude that the tuning which will be best suited for the real-time level control trainer is the Ziegler–Nichols tuning method. As this tuning method is providing satisfactory rise time that is 3.2902 s but a better settling time

Fig. 11 Real-time tuning (PID tuner)
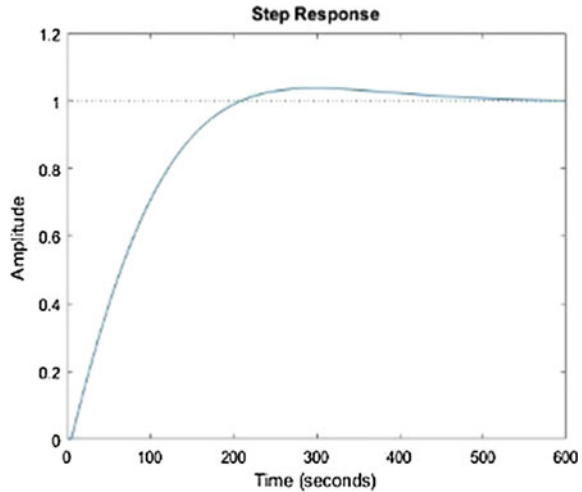
Table 2 Time-domain specification of given FOPDT model in MATLAB

Tuning method	Rise time (s)	Settling time (s)	% overshoot	Peak time (s)
Ziegler–Nichols	3.2902	89.4497	63.9093	13.5319
Cohen–Coon	2.9451	123.926	73.2336	1.7323
Wang–Juang–Chan	69.2385	545.4891	3.5836	0.9978
Chien–Horns–Reswick	62.9658	518.8147	0	0.9999
AMIGO	23.5937	178.0595	0	0.9979
PID tuner	137.2688	411.4076	3.6767	297.6244

that is 89.447 s and satisfactory % overshoot that is 63.093% as compared to any other tuning methods. The real-time tuning that we had used is PID tuner is providing the satisfactory % overshoot which is very less as compared to any other PID tunings that had been done but have not good rise time and settling time. So, if we tune our level control trainer using Ziegler–Nichols method, we can obtain satisfactory desired output. The whole paper is about to design an empirical model for our level control trainer using either using the process reaction curve (PRC) or using Two-point method which is completely a statically procedure. Most of the engineers try to find the empirical model using the Two-point method this because finding the slope in the process reaction curve is little bit tricky. So therefore, for this reason, most of the engineer recommends Two-point method where we must find two values of the time that is in 63.2 and 28.3% of the total time.

References

1. Donald, C.R., LeBlanc, E.S.: Process Systems Analysis and Control. McGraw-Hill, University of Michigan (1999)
2. Seborg, D.E., Edgar, T.F., Mellichamp, F.T.: Process Dynamic and Control. John Wiley & Son, New York (2004)
3. Ljung, L.: System identification-theory for the user. Prentice-Hall, Upper Saddle River, NJ (1994)
4. Viberg, M.: Subspace-based methods for the identification of linear time-invariant systems. Automatica **31**, 1835–1851 (1995)
5. Ljung, L.: Estimating linear time-invariant models of nonlinear time varying systems. Eur. J. Control **7**(2–3), 203–219 (2001)
6. Bai, E.W.: A blind approach to the Hammerstein–Wiener model identification. Automatica **38**, 967–979 (2002)
7. Wang, D., Ding, F.: Extended stochastic gradient identification algorithms for Hammerstein-Wiener ARMAX systems. Comput. Math. Appl. **56**, 3157–3164 (2008)
8. Marlin, E.T.: Process Control-Designing Processes and Control System for Dynamic Performance, pp. 175–206. McGraw-Hill, Ontario (1995–2014)
9. Stephanopoulos, G: Chemical Process Control-An Introduction to Theory and Practice. Prentice Hall India Learning Private Limited (2008)
10. Ziegler, J.G., Nichols, N.B., Rochester, N.Y.: Optimum settings for automatic controllers. Trans. ASME **64**, 759–768 (1942)

Comparative Analysis of Energy Consumption in Sensor Node Scheduling Heuristics in Wireless Sensor Network

Sunita Gupta and Sakar Gupta

Abstract Wireless Sensor Networks (WSNs) are commonly used in many wireless applications like battlefield, environmental monitoring, etc. WSN is consisting of an excess of sensors that operates for many months to years to complete their assigned tasks. Due to the small size of a sensor node, the power supply attached to the sensor node is very limited in size. Thus, energy conservation becomes a challenging issue in WSN design and researchers face problem to get long operating hours without affecting the system performance. In this paper, a hardware and battery models are surveyed that can affect battery life and cause the difference between the simulation and application results. Varieties of mathematical models have been studied to serve as analytical tools in quantifying battery utilization and discharge characteristics. However, batteries are the primary power supply source. They fail earlier in some applications than their projected working time. So Energy Harvesting WSN (EHWSN) can be used if possible. This paper compares and shows a pattern of battery utilization in lifetime maximizing heuristics for WSN. In (Gupta and Roy Global J Comput Sci Technol E Netw Web Secur 15(6), 2015 [1]), proposed a heuristic called Q-Coverage Maximum Connected Set Cover (QC-MCSC) and it is used for energy minimization. It schedules the activities of nodes having Q-Coverage and Connectivity constraints. QC-MCSC is compared with existing heuristics, High Energy Small Lifetime (HESL), Triple Phase Iterative Connected Set Cover (TPICSC) and Maximum Set Cover (MSC). In this paper, a comparison of performance of QC-MCSC heuristic is done with existing heuristics over battery utilization pattern in WSN.

Keywords Wireless sensor network · Energy consumption and battery model
Coverage and connectivity

S. Gupta (✉)
Swami Keshvanand Institute of Technology Management & Gramothan (SKIT),
Jaipur, India
e-mail: drsunitagupta2016@gmail.com

S. Gupta
Poornima College of Engineering, Jaipur, India
e-mail: sakargupta@gmail.com

1 Introduction

There are various key design issues to make WSN to perform as desired like WSN dynamics and node deployment method, the order of Coverage and Connectivity, mobility, clustering, transmission media and communication, power consumption, data aggregation, fault tolerance, scalability and network topology [2]. A major challenge for WSNs, however, lies in the node energy constraint. The energy source for sensor nodes is typically two AA batteries. On the other hand, expected lifetimes of batteries should range from months to years. It is undesirable to recharge or replace the batteries of thousands of sensor nodes frequently, when a node is depleted of energy. It cannot fulfil its role further, unless the source of energy is replaced. So the usefulness of a Wireless Sensor Node finishes when its battery runs out. The energy exhaustion of any node indicates that the network needs to be reorganized.

In simple WSN, a sensor is a key element of a WSN node. A node is also equipped with some other main components like Microcontroller, Transceiver, Memory and Power unit. Sensed information by a sensor is stored in a buffer and forwarded to remaining nodes or fusion centre. The power supply is used to supply energy to storage, Sensing and Transceiver unit. A simple model of WSN is given in Fig. 1.

In Energy Harvesting WSN (EHWSN), ambient energy is provided from energy harvesting device as shown in Fig. 2. In energy harvesting process, energy is generated from some outside sources like Sun, wind, etc. Thus, there is no problem of battery capacity limitation. Hence, lifetime using EHWSN is increased.

Next section includes related work about battery models and mathematical models. A battery model is introduced in Sect. 3. In Sect. 4, a hardware model for battery utilization and discharge behaviour of a wireless sensor's battery is discussed. In Sect. 5, a comparison of the performance of QC-MCSC Heuristic with existing heuristics is done over the pattern of battery utilization in Wireless Sensor Network. Finally, Sect. 6 concludes this paper.

Fig. 1 Model of wireless sensor node
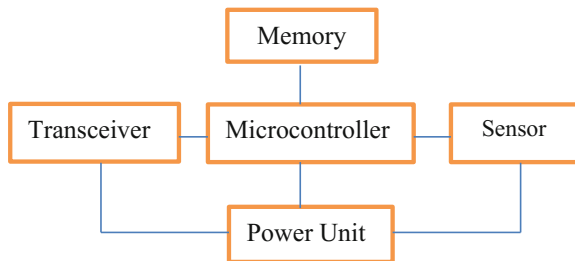
Fig. 2 Model of energy harvesting wireless sensor node
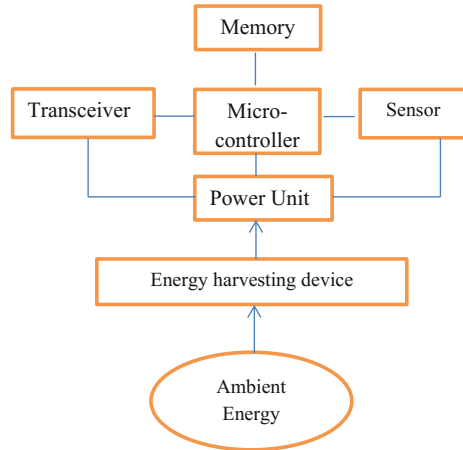
2   Related Work

The majority of work on Wireless Sensor research relies on generic battery models, which place little emphasis on actual hardware platforms and batteries. The battery discharge characteristic of a sensor node was first examined experimentally by authors in [3]. The experiment was conducted with a commercial lithium-ion battery. The sensitivity of it was examined to the load profile of a sensor node with limited WSN parameters. The capacity extracted from the battery is gradually down by voltage converter which is used for power supply to a base node.

Many models of low-power wireless communication transceivers have been proposed in recent years to lower sensor energy usage. For example energy-efficient modulation scaling [4, 5], delay controlled transmission schemes [6, 7], energy-efficient routing [8, 9], power management-based task scheduling for digital communication processors [10], etc. A survey of these different energy minimization techniques is given in [11]. A comparison of some energy-efficient sensor node scheduling algorithms are also given in [12].

Measuring the battery life through real experiments is time consuming. Thus, numerous simulation tools like NS-2 [13] have been proposed in recent years to estimate it. A variety of simulation strategies have been suggested to exploit battery characteristics for designing more "battery friendly" systems and communication protocols [14]. Many mathematical models have been proposed that provides upper bounds on lifetime [15–18].

3   Battery Model

In WSNs, batteries are the only source of energy supply to sensor nodes. Because of very small size and capacity of batteries, WSN lifetime is very limited. In order to increase overall lifetime of a WSN, energy efficiency of every node is to be improved or energy of each and every node is to be utilized properly. Each Wireless Sensor Network node radio can be in one of the three states. It is in an active state when the transmitter is transmitting or receiver is receiving. In sleep state, its radio is off. It does not consume energy in a sleep state. The transceiver is neither transmitting nor receiving. In idle state, it consumes less energy as compared to the active state. Thus the largest amount of energy is consumed in the active state. A large quantity of energy is lost during communication because of many reasons like Collision, Overhearing, Control Packet Overhead, Idle Listening, Over emitting, Interference etc.

A simplest model for power consumption of a sensor node (P_{node}) is given below in Eq. 1 [19].

$$P_i = \left(T_{iactive} * P_{iactive} + T_{isleep} * P_{isleep}\right) / T \tag{1}$$

$$P_{node} = \sum_i Pi \tag{2}$$

The node can be divided into i different units (like processing unit, transceiver unit, sensors, etc.). Each unit can be in two states: active and sleep state. Sleep state consumes very low power and $P_{sleep} \ll P_{active}$. In time T, A unit i can change between two states. $T_{iactive}$ is the total time a component passes in active state. T_{isleep} is the total time that a sensor node spends in sleep mode. For all i units, total time $T = T_{iactive} + T_{isleep}$. P_{node} is the sum of power consumption of all components. Power consumption in active time ($P_{iactive}$) and the time that a component spends in active state (T_{active}) are to be reduced, to minimize P_{node} in total.

4   Hardware Model

For computing and calculating the power consumption of the MAC protocols for different transceivers, a set of variables are used and are given in Table 1. For converting the power consumed by a node into the lifetime of a node, a battery model given in [20] is used. Lifetime of Ni–Mg batteries is long and has a continuous output voltage. But ,alkaline batteries are used as their price is lower. The power leakage is 10% of the total energy E in 1 year. $p_{leak} = (0.1 * E)/((24) * (365))$, where E is in Wh. For P power consumption, the battery will runs off after time T as given follows:

$$T = E/(P + P_{leak}) = E/(24 * 365) * p + 0.1 * E \text{ (In Years)} \tag{3}$$

Table 1 Parameters used in the hardware model for energy consumption

Variables	Parameter
C_{sleep}	Sleep current in mA
C_{batt}	Capacity of battery in mAh
V	Voltage
P_{leak}	Power leakage of battery
$L_{preamble}$	Preamble length in bytes
L_{packet}	Packet length in bytes
t_i	Radio sampling interval in seconds
R	Sample rate in packets per second
P_{tx}	Power consumption when transmitting
P_{rx}	Power consumption when receiving
T_{SeRx}	Settling time for receiving mode
T_{SeTx}	Settling time for transmitting mode
L	Expected lifetime (s)
B	Raw bit rate
θ	Clock drift rate

The energy consumed by a node decides the overall lifetime of the nodes as in [21]. If lifetime of the node is improved, then the total energy consumption decreases. Energy used by a sensor is the addition of energy used in transmitting (E_{tx}), receiving (E_{rx}), listening on the radio channel (E_{listen}), sleeping (E_{sleep}) and sampling data (E_d). The notations listed in Table 1 are used in this paper. The average energy consumption E_{consp} by a node is given by

$$E_{consp} = E_{rx} + E_{tx} + E_{listen} + E_d + E_{sleep} \tag{4}$$

The total energy E in measured in Wh. The energy E is resultant of product of capacity in Ah and voltage. Starting voltage is 1.5 V and ending voltage is 0.9 V. The mean of the voltages taken which is 1.2 V. In LR6 battery of capacity $C_{batt} = 2.6$ Ah, the energy for the lifetime T is

$$E = V * I * T = ((1.2) * (2.6) * (T))/T = 3.12 \, \text{Wh} \tag{5}$$

Using this simplest model, for zero-power depletion, the battery lifetime is 10 years. Here, the complexity of the battery is not mixed with the complexity of the protocols.

5   Comparison of QC-MCSC Heuristic with HESL, TPICSC and MSC

Comparative Performance of QC-MCSC with Existing Heuristics HESL [22], TPICSC [23] and MSC [24] over the pattern of battery utilization for a given battery Model is given below. QC-MCSC is compared with existing heuristics techniques HESL, TPICSC and MSC over the network lifetime. As seen from the table, QC-MCSC achieves the lifetime higher than existing heuristics Techniques HESL, TPICSC and MSC.

From the graph, it is proved that the energy consumed by nodes in heuristics QC-MCSC is fewer as compared to heuristics HESL, TPICSC and MSC (Fig. 3 and Table 2).
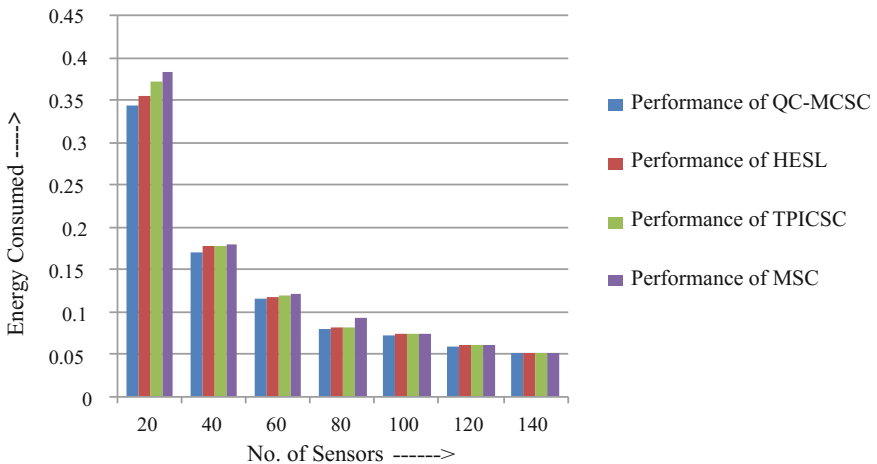
Fig. 3 Comparison of the performance of QC-MCSC heuristic with HESL, TPICSC and MSC over pattern of battery utilization for a given battery model

Table 2 Comparison of performance of QC-MCSC heuristic with HESL, TPICSC and MSC over pattern of battery utilization for a given battery model

No. of Sensors	Performance of QC-MCSC [1]		Performance of HESL [22]		Performance of TPICSC [23]		Performance of MSC [24]	
	Lifetime of sensor	Energy consumed	Lifetime of sensor	Energy consumed	Lifetime of sensor	Energy consumed	Lifetime of sensor	Energy consumed
20	9.087	0.3433476	8.763	0.3560424	8.384	0.3721374	8.143	0.3831511
40	18.267	0.1707998	17.463	0.1786634	17.565	0.1776259	17.376	0.1795580
60	26.738	0.1166878	26.374	0.1182983	26.038	0.1198248	25.894	0.1204912
80	34.373	0.0803357	38.537	0.0809611	33.572	0.0818103	33.423	0.09334889
100	42.829	0.0728478	42.459	0.0734826	42.127	0.0740617	41.727	0.0747717
120	52.283	0.0596752	51.637	0.0604217	51.582	0.0604862	51.125	0.0610268
140	61.016	0.0511341	59.765	0.0522044	60.303	0.0517387	59.782	0.0521896

The implementation of QC-MCSC heuristic for real-time applications will help in realizing highly efficient and low-cost Wireless Sensor Networks.

6 Conclusion

In this paper, battery of a sensor node, its characteristics, different battery models and hardware models are discussed. Different mathematical models for battery utilization are also discussed. A Comparison of performance of QC-MCSC Heuristic is done with HESL, TPICSC and MSC over a pattern of battery utilization for a given battery model. There are many other methods that can also be used for maximizing the lifetime of a WSN like clustering, routing, etc. The mathematical model can also be used for these approaches to analyse the pattern of battery utilization. The order of coverage and connectivity can also be increased to get the more reliable WSN, but battery utilization increases in that case.

References

1. Gupta, S., Roy, K.C.: Q-Coverage maximum connected set cover (QC-MCSC) heuristic for connected target problem in wireless sensor network. Global J. Comput. Sci. Technol. E Netw. Web Secur **15**(6) (2015), ISSN: 0975-4172
2. Gupta, S., Roy, K.C.: Energy efficient target coverage issues in wireless sensor network. Int. J. Comput. Organ. Trends **1**(3) (2011), ISSN: 2249-2593
3. Park, S., Savvides, A., Srivastava, M.B.: Battery capacity measurement and analysis using lithium coin cell battery. In: Proceedings of International Symposium on Low Power Electronics and Design, pp. 382–387 (2001)
4. Schurgers, C., Aberthorne, O., Srivastava, M.B.: Modulation scaling for energy aware communication systems. In: Proceedings of International Symposium on Low Power Electronics and Design, pp. 96–99 (2001)
5. Yang, Z., Yuan, Y., He, J.: Energy aware data gathering based on adaptive modulation scaling in wireless sensor networks. In: Proceedings of IEEE 60th Conference on Vehicular Technology, vol. 4, pp. 2794–2798 (2004)
6. Woo, A., Culler, D.: A transmission control scheme for media access in sensor networks. In: Proceedings of 7th Annual International Conference on Mobile Computing Network, pp. 221–235 (2001)
7. Wang, X., Ren, Y., Zhao, J., Guo, Z., Yao, R.: Energy efficient transmission protocol for UWB WPAN. In: Proceedings of IEEE 60th Conference on Vehicular Technology, pp. 5292–5296 (2004)
8. Ramanathan, R., Rosales-Hain, R.: Topology control of multihop wireless networks using transmit power adjustment. In: Proceedings of Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies, pp. 404–413 (2000)
9. Li, D., Jia, X., Liu, H.: Energy efficient broadcast routing in static ad hoc wireless networks. IEEE Trans. Mob. Comput. **3**(2), 144–151 (2004)
10. Lahiri, K., Raghunathan, A., Dey, S., Panigrahi, D.: Battery-driven system design: a new frontier in low power design. In: Proceedings of the 7th Asia and South Pacific and the 15th International Conference on VLSI Design, pp. 261–267 (2002)

11. Gupta, S., Roy, K.C.: Comparison of different energy minimization techniques in wireless sensor network. Int. J. Comput. Appl. (0975–8887) **75**(18) (2013)
12. Gupta, S., Roy, K.C.: Comparison of sensor node scheduling algorithms in wireless sensor networks. Int. Res. J. Eng. Technol. (IRJET), **02**(06) (2015), e-ISSN: 2395-0056. www.irjet.net
13. The Network Simulator—ns-2. http://www.isi.edu/nsnam/ns/ (2011). Accessed 31 Jan 2011
14. Luo, J., Jha, N.K., Battery-aware static scheduling for distributed real-time embedded systems. In: Proceedings of Design Automation Conference, pp. 444–449 (2001)
15. Bhardwaj, M., Chandrakasan, A.P.: Bounding the lifetime of sensor networks via optimal role assignments. In: Proceedings of Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies, vol. 3, pp. 1587–1596 (2002)
16. Bhardwaj, M., Garnett, T., Chandrakasan, A.P.: Upper bounds on the lifetime of sensor networks. In: Proceedings of IEEE International Conference on Communications, vol. 3, pp. 785–790 (2001)
17. Coleri, S., Ergen, M., Koo, T.J.: Lifetime analysis of a sensor network with hybrid automata modeling. In: The International Workshop on Wireless Sensor Networks and Applications, Atlanta, Georgia (2002)
18. Dasgupta, K., Kalpakis, K., Namjoshi,P.: An efficient clustering-based heuristic for data gathering and aggregation in sensor networks. In: Proceedings of IEEE Wireless Communications and Networking, pp. 1948–1953 (2003)
19. Kulau, U., Bušching, F., Wolf, L.: A node's life: increasing WSN lifetime by dynamic voltage scaling. In: 2013 IEEE International Conference on Distributed Computing in Sensor Systems
20. El-Hoiydi, A.: Spatial TDMA and CSMA with preamble sampling for low power ad hoc wireless sensor networks. In: Proceedings of the Seventh International Symposium on Computers and Communications (ISCC'02) 1530-1346/02 $17.00 © 2002 IEEE
21. Polastre, J., Hill, J., Culler, D.: Versatile Low Power Media Access for Wireless Sensor Networks, 3–5 Nov 2004
22. Chaudhary, M., Pujari, A.K.: Q-coverage problem in wireless sensor networks. Int. Conf. Distrib. Comput. Netw. (ICDCN), Springer publications (2009)
23. Jamali, M.A., Bakhshivand, N., Easmaeilpour, M., Salami, D.: An energy-efficient algorithm for connected target coverage problem in wireless sensor networks. In: 2010 3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT), vol. 9, pp. 249–254. IEEE Conference Publications (2010)
24. Cardei, M., Thai, M.T., Li, Y., Wu, W.: Energy efficient target coverage problems in wireless sensors networks. Proceedings of IEEE INFOCOM 2005, 24th Annual Joint Conference of the IEEE Computer and Communications Societies, vol. 3, pp. 1976–1984. IEEE

Medical Diagnosis of Parkinson Disease Driven by Multiple Preprocessing Technique with Scarce Lee Silverman Voice Treatment Data

Alok Kumar Shukla, Pradeep Singh and Manu Vardhan

Abstract Parkinson's disease is a chronic neurodegenerative disorder characterized by the progressive deterioration of motor function that affects vocal performance and can result in noticeable disruption of vocal performance degradation. Early diagnosis of Parkinson's disease is very crucial in preventing the disease's progression. However, it is a complicated task for specialists or clinicians due to a wide scale of symptoms and progressive changes in disease's symptoms over time. No standard framework exists to determine what percentage of diseases arises. This research work aims at early identification of patients with Parkinson's disease using multiple preprocessing techniques known as Multi-Preprocessing System (MPS). The main aim of the proposed diagnostic framework is to investigate the potential of sustained vowel phonations of Lee Silverman Voice Treatment dataset which improve the classification performance. The proposed framework depends on three stages of preprocessing: (a) novel ensemble method for feature selection, (b) Discretization, and (c) Principal Component Analysis (PCA). The experimental results emphasize that the proposed MPS method provides additional support to significant reduction of cardinality and outperforms the state-of-the-art feature selection methods in terms of Accuracy, Sensitivity, Precision, and F-measure, using seven learning algorithms. The evaluation results show that proposed method using Random Forest (RF) achieves the highest performance in terms of accuracy as 94.98%, sensitivity as 93.18%, precision as 94.96%, and F-measure as 94.7%.

Keywords Discretization · Feature selection · Lee Silverman voice treatment Preprocessing · Principal component analysis

A. K. Shukla (✉) · P. Singh · M. Vardhan
Department of Computer Science & Engineering,
National Institute of Technology, Raipur, Raipur, India
e-mail: akshukla.phd2015.cs@nitrr.ac.in

P. Singh
e-mail: psingh.cs@nitrr.ac.in

M. Vardhan
e-mail: mvardhan.cs@nitrr.ac.in

1 Introduction

Data mining in the restorative area gathers information from the past experiences and examines them to recognize patterns and solutions for the current circumstances. It is an effective diagnostic procedure for recognizing obscure and significant data from the expansive volume of therapeutic data taken from medical science, medication procedures, and clinical tests [1]. Data mining-based techniques can be used for the change of prescient models that empowers classification and prediction. In the fields of data mining, the core problem is to select the meaningful features and ignore extraneous features from the medical dataset. Therefore, most of the researchers have used efficient preprocessing techniques such as feature selection, discretization, and principal component analysis on the medical dataset to reduce the computational cost of any learning algorithm.

Over the past few years, numerous machine learning (ML) methods have used an ensemble model for better performance. It is the way toward building a logical model in the wake of finding information from the data which conveys us to the idea of machine learning (ML). It is formally characterized as the perplexing computation procedure of automatic data recognition and intelligent decision-making in view of sample data [2, 3]. Ensemble learning method has been successfully employed in classification complications, but is also a means for improving other ML functions, i.e., feature selection (FS). The ensemble can be formed in many ways, here, we use a heterogeneous ensemble method (based on ranking of feature selector), namely Correlation-based Feature Selection (CFS), ReliefF, Information Gain (IG), and Chi-Square (CS), respectively.

At an abstract level, Parkinson's disease (PD) is a prolonged mental disorder as depression or neurosis categorized by the progressive worsening of motor use and the rising of significant non-motor problems [4]. According to worldwide statistical analysis report, diseases estimated at least 0.001% of the population [5, 6]. The specified study on vocal deficiency, approximate 29% of people suffer in the huge majority of Parkinson's disease and crucial problematic disease [7]. The most usual symptoms of vocal impairment symptoms are quietness, monotone, harshness, noise, vague articulation, and vocal quiver [8]. The vocal damage range can be evaluated using persistent vowel phonations or run the gauntlet. The cause of Parkinson's disease, the vocal deficiencies in running speech may not arise in sustained vowels, and quantity of speech is more intricate due to articulatory and other linguistics [9, 10]. Moreover, the measurement of voice deficiency can be considered by sustained vowel structure or gauntlet. The fact is voice quality and running voice affects due to the PD (e.g., sequences of vowels and consonants). Due to voice and other linguistic rattle, the measure the performance of running voice is very difficult [11].

The goal of this exploration work is to demonstrate that the choice of the huge features from the accessible crude medicinal dataset helps the physician to arrive at an accurate diagnosis. The features are subjected to a three-stage filtration process, which forms the subset with the lowest cardinality, i.e., according to the corresponding rank of the feature, the minimum number of features is selected that expands

the accuracy. The technique utilizes an efficient strategy of combining preprocessing with ranking method. The observational outcomes demonstrate that the proposed Multi Preprocessing System (MPS) framework achieves remarkable dimensionality reduction in the medical dataset.

The key contribution of this paper developed a novel ensemble method with the help of the heterogeneous filter-based feature selector (based on ranking of features), namely CFS, ReliefF, IG, and CS. Then, applied discretization and principal component analysis for better classification of medical data. The method is established to perform satisfactorily in terms of classification performance for a huge number of features-based medical dataset. The rest of the paper is planned as follows. Sections 2 and 3 provides the related work on ensemble of feature selection method, discretization, and dimension reduction technique. In Sect. 4, introduced the proposed framework to take the salient features from the dataset and distinguish the two cohorts (accept versus reject). Sections 5 and 6 discuss the different classification techniques and shows the experimental result on the LSVT dataset. Section 7 discusses the conclusion.

2 Related Work

In the current literature, there are still a few weaknesses in analytic strategies for Parkinson's disease. For example, (i) interrelationships among determination factors have not been considered, (ii) dynamic alters in disease direction after sometime have been disregarded while dynamic finding strategy could cause better control of the circumstance and make better choices over the long time. In machine learning [12], all the computational methods prescribe a couple of critical features for malady determination. At that point, these prescribed critical features may enable specialists or clinicians the biomedical mechanism to better with a more pre-learning about the reason for illness and give the quickest conclusion to recovery the tainted patients as right on time as could be expected under the circumstances. Following our related work focus on the feature selection methods and highlights the limitations of previous studied by features selection methods with the aim of picking a features subset that maximize the performance (for example, accuracy) of a assumed classifier. In most of the cases, the random forest is a base classifier to discriminate vigorous subjects from people with Parkinson's (PWP) and measurement of the diagnostic system is measured with ROC [13]. To the best our knowledge, detection of Parkinson's disease and classification based on evolutionary approaches have been introduced by only a few researchers to overcome diagnosis problem. The discrimination of normal or abnormal cases, numerous research papers have been tried to solve this problem by acoustic measurements of dysphonia [14]. According to Little et al. [15] tried to finding the novel dataset and focus on the effectiveness of nonstandard measurement and serene sustained vowel phonation's 23 were PD patients out of 31 subjects, maximum accuracy reached 91.4%. In fact, as a recent work indicated by Sakar et al. [16] offered a patient dataset with 20 PD out of 40 subjects in order to the inven-

tion, the disorder terms unremitting vowels, words, numbers, and petite sentences to each 26 distinct individuals set by summarized leave-one-out validation technique with accuracy 77.5%. Gillespie and Gartner-Schmidt [17] determined what factors of speech-language pathologists (SLPs) deem most important when discharging a patient from voice therapy. Sakar and Kursun [3] attempted to choose the relevant and correlated features by the most popular technique maximum relevance -minimum redundancy and support vector machine classification using leave-one-subject-out (LOSO) technique to avoid bias. Das [18] has shown a relative study of Decision Tree (DT), Neural Network (NN), and Regression to efficiently distinguish healthy individuals from PD dataset. Zuo et al. [19] have proposed a novel hybrid particle swarm optimization (HPSO) and fuzzy k-nearest neighbor method with achieved accuracy at 97.47%.

3   Material and Existing Method

3.1   Dataset Description

In this study, we use LSVT dataset taken from UCI ML-repository [20]. The aim of dataset is to classify dysphonia features of Parkinson's dataset. The LSVT dataset consists of 126 supported vowel phonations and 310 dysphonia features.

3.2   Feature Selection

Feature Selection (FS) is an approach by which reduction is done on available dataset which has increased explosively in both numbers of samples and dimensionality. It is generally used in the fields of bioinformatics, text mining, and data mining problems. It is divided into four categories: Filter [21], Wrapper [22], Hybrid [23], and Ensemble [24]. In this paper, our focus is on the filter-based FS method, to choose the informative features from the dataset for better diagnosis of diseases. Here, we use most common feature selection techniques, i.e., ReliefF [25], Correlation-based Feature Selection (CFS) [26], Information Gain (IG) [27], and Chi-Square(CS) [28].

3.3   Discretization

Discretization is a vital preprocessing method that is basically used in data mining problems which reduces the noise and accelerates the learning process [29]. The essential goal of discretization is to mapping a continuous feature into m intervals $U = \{(d_0, d_1) \ldots (d_{m-1}, d_m)\}$, where d_0 and d_m are the minimal and the maximal

values of features. In the past literature, lots of method of discretization has been proposed for un-supervised and supervised methods [30].

3.4 Dimensionally Reduction

In machine learning problems, dimension reduction is an effective tool to improve the performance and reduced the number of random variables under consideration via obtaining a set of principal components so decreases the training time and may increase the classification accuracy. In the past literature, lots of methods have been proposed for dimensionality reduction in the context of data mining fields. In this correspondence, we used the most potent and prominent data reduction technique, namely Principal Component Analysis (PCA) [31].

4 Proposed Methodology

Each feature selection technique (FS) has its own qualities and shortcomings, and the execution is based on the type of dataset, but also on the constraints related to the scenario (accuracy, time and cost). In any case, in spite of the accessibility of developing quantities of strategies, analysts by and large concur that no perfect FS method exists. Some information of existing FS algorithms is required in order to be able to have the capacity to pick a strategy that is fitting to the issue. One possible solution to this issue is to use an ensemble of heterogeneous FS methods as shown in Fig. 1.
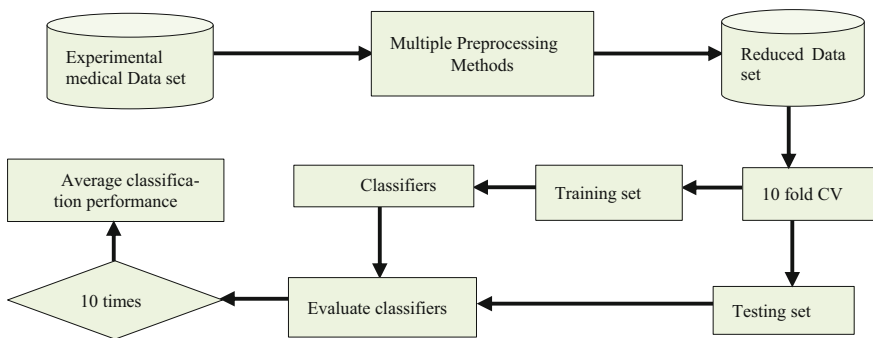
Fig. 1 Overall process of multi-preprocessing system (MPS)

4.1   First Stage of Preprocessing—Ensemble Method

In this paper, we generated four strategies using different feature selection methods on medical dataset (see pseudo-code 1). This approach assesses the qualities and shortcomings of the individual techniques. The several FS methods (i.e., CFS, IG, ReliefF, and CS) are measure rank on the medical dataset, and the output is combined using an amalgamation method.

Pseudo-Code 1 : **Heterogeneous Ensemble Method**
Data: M – number of ranker methods
Value : δ – threshold of the number of features to be selected
Outcome : P – Best ranking feature set
For each n from 1 to M do
Obtaining ranking A_M using feature selection method M
End for
A = combining ranking A_m with a ranking combination method
A_t = Select δ top attributes from A
Obtain top ranked features
Return P

4.1.1   Ranking Combination

Ensemble Method is a rich strategy to create enhanced machine learning (ML) results. We present the ensemble method that obtains a final positioning of highlights by combining outputs (rankings) from singular rankers utilizing a combination technique also known as aggregator to create a surprising final output. Several different combination methods are available in the literature, here, we used ranking vote method.

4.1.2   Ranking Vote

Ranking Voting is a group decision-making scheme and has been found to be just as effective as other more complex schemes [32]. For an input sample, individual feature selection (FS) method produces an exclusive decision regarding the identity of the sample. Then, the identity is assigned according to what the frequency of voters agree (output prediction is the one that receives more than half of the votes). In the context of feature selection, the input samples are the features ranking and the identity each features ranking is ensemble or not in combination. If none of the output ranking get more than one of the votes, we may say that the ensemble method could not make a stable ranking for this instance, then we pickup smallest vote value (in terms of score or weight). Although this is a widely used technique, you may try the most voted ranking as the final ranking according to Eq. 1.

$$\sum_{n=1}^{M} d_{n,j} = argmax_{j \in \{1,2,...,L\}} \sum_{n=1}^{M} d_{n,j} \qquad (1)$$

Here, M represents the number of feature selection method and L is selected number of attributes. For attribute j, the sum $\sum_{n=1}^{M} d_{n,j}$ tabulates the number of votes for j. Plurality chooses the attribute j which maximizes the sum.

4.2 Second and Third Stage of Preprocessing: Discretization and Dimensionally Reduction

In the second stage, we present the efficient preprocessing technique, namely discretization on medical dataset. We transform the data features using a standard discretization algorithm with a finite number of intervals. The data attributes have discretization values with a finite number of intervals (set d=3) as $d = \{d_1, d_2, d_3, \ldots, d_n\}$ for better average accuracy with low computation cost. In order to select the relevant features from the medical dataset and then discretized the selected features by the method. While the subsequent stage knowing as Dimensionally Reduction, which consists of the feature space of reduced dataset with subset of features F' containing n features. Principal component analysis technique is applied to each class to produce the covariance matrix which is specifically suitable to target class.

Pseudo-Code 2 : Multi-Preprocessing System (MPS)

Input: Dataset D has $F = \{f_i, i = 1, 2, \ldots, n\}$ with n features, Ɛ: number of classifiers, δ: threshold of the number of features to be selected, D': reduced medical data set with X training set and Y testing set, M: number of ranker methods.

Output: Accuracy, Precision, Sensitivity, F-measure.

Initialization: $F' = \emptyset$ and $At = \{\}$:

begin

    For each $f_m \in M$

       Evaluate At for f_m according **to Pseudo-Code 1.**

     End For

     ensemble ($At, 'vote'$)

    select $(f_n) \leftarrow f (At, \delta)$

 select the feature F' based on ensemble feature ranking from f_n ; $F' = F \setminus f_n$

    Return D' with feature subset F'

 Perform discretization process on reduced dataset with the discretize level of each features

Perform Dimensionally Reduction Process on discretize dataset

 For i=1: Ɛ do

  randomly split using tenfold Cross Validation

   training set X with F'

    test an Ɛ classifier from testing set Y

     evaluate performance

End For

End

Return Accuracy, Precision, Sensitivity, and F-measure

Abovementioned pseudo-code 2 describes the structure of proposed (MPS) methodology. The reduced medical dataset randomly splits in training datasets and testing datasets using renowned method for evaluation called as tenfold Cross-Validation (CV). The building a classifier, much of the work is concentrated on selected relevant features by MPS methodology, which gives maximum performance in terms of Accuracy, Precision, Sensitivity, and F-measure, respectively.

5 Classification Technique

To measure the acceptability of feature subset for classification, different error estimation strategies have been suggested. In the past literature, most of the researchers have been applied the popular validation method, i.e., CV or bootstrap techniques [33], respectively. In this paper, we use ten-fold CV method on the LSVT dataset, which randomly splits the dataset into training and testing data subgroups: In training dataset that consists of 0.90 of the data samples and other testing subset consisting of 0.10 of the data samples to estimate the performance based on confusion matrix.

5.1 Classification

In the pattern recognition and data mining fields, classification is the process of given previously unknown sample, decide to which group it belongs that is, a function that assigns a class label to the instances described by a set of attributes. In this paper, we use seven different classifiers, i.e., Naïve Bayes (NB) [34], Support Vector Machine (SVM) [35], Random Forest (RF) [36], K-Nearest Neighbor (KNN) [37], Multi-layer Perceptron (MLP) [38], Decision Tree (J48) [39], and NBTREE [40], respectively.

6 Result and Discussion

6.1 Performance Measures

We measure the classification performance with the help of seven classifiers, i.e., Support Vector Machine (SVM), Naïve Bayes (NB), Random Forest (RF), J48 (DT), K-Nearest Neighbor (KNN), NBTREE, and Multi-Layer Perceptron (MLP) with four performance parameters, i.e., accuracy, sensitivity, precision, and F-measure, respectively, using WEKA toolbox [41]. These performance measures are defined as:

(a) **Accuracy**: To predict the percentage of correctly classified samples, it is formulated as: Accuracy $(\%) = \frac{TN+TP}{TP+TN+FN+FP} * 100$

(b) **Sensitivity**: Percentage of positive instances that are predicted as positive. It is also called TPR or Recall. It is formulated as: Sensitivity $=$ Recall (Re) $(\%) = \frac{TP}{TP+FN} * 100$

(c) **Precision**: It is the percentage of positive predictions that are correct. This is also called PPV (positive predicted value). It is formulated as: Precision (Pr) $(\%) = \frac{TP}{TP+FP} * 100$

(d) **F-measure**: F-measure has a parameter that sets the tradeoff between recall and precision. The standard F-measure is F1, which gives equal importance to recall and precision which define as: F-measure $= \frac{2 * Pr * Re}{Pr + Re}$

Here, TP, TN, FP, and FN are true positive, true negative, false positive, and false negative in the independent datasets, respectively. All the experiments of classification and performance evaluation are performed using MATLAB and WEKA data mining tool.

6.2 Experimental Results

This section illustrates the performances of proposed method against individual preprocessing method with respect to seven classifiers. Here, the maximum threshold is set to be 0.30 and as a preprocessing step select the top-ranked features by vote scheme for LSVT datasets. The classification performance of the proposed method on the LSVT datasets is evaluated using the tenfold cross-validation (10-CV) approach as shown in Table 1. The proposed method average accuracy is computed as the average of the ten runs.

In our experiment, Fig. 2 shows the assessment of the performance of the algorithms. Figure 2 shows that the proposed method achieves the highest accuracy (94.98%) with RF classification with respect to other existing methods. The maximum classification accuracy for preprocessing method is reported in Table 1 at bold value. The classification accuracies of the existing classifiers using different preprocessing techniques are between 56.34 and 94.98%. Accuracies achieved by J48, KNN, MLP, NB, NBTREE, and SVM classifier using the proposed MPS method is 92.18%, 91.28%, 94.17%, 93.89%, 92.68%, and 93.08%, respectively.

The ultimate goal of the researcher is to measure the predictive classification accuracy and also other performance measures like precision, sensitivity, and F-measure to distinguish the effect of feature selection. As shown in Figs. 3 and 4, the proposed method with RF classifier shows the highest performance in terms of precision 94.96% and sensitivity 93.18%, respectively. As reported in Table 2, the proposed MPS method demonstrates the effectiveness of comparatively other classification method and existing feature selection method. The proposed MPS methods performance is effectively better than the remaining preprocessing methods shown as bold in Table 2. Furthermore, for LSVT dataset, the sensitivity and precision versus the number of the present classifier for the proposed method and the best 14 comparison preprocessing methods are shown in Figs. 3 and 4. The range of precision value is spread 91.40%, 90.71%, 92.93%, 92.93%, 93.08%, 94.96%, and 91.90% for

Table 1 Classification accuracies (%) with J48, KNN, MLP, NB, NBTREE, RF, and SVM

Method	% of accuracy						
	J48	KNN	MLP	NB	NBTREE	RF	SVM
ALL FEATURES	75.42	75.43	86.50	56.34	75.47	83.30	66.64
DI	86.51	88.88	89.68	91.26	88.09	81.74	90.48
PCA	72.22	68.25	77.78	74.60	74.60	80.95	79.36
CFS	77.87	83.31	85.75	66.77	82.45	85.76	67.08
RELIEFF	73.15	75.47	86.56	56.39	80.08	84.14	68.07
IG	82.54	69.84	79.37	51.97	73.02	84.92	65.60
Chi-Square	76.19	69.84	80.19	53.98	80.18	82.57	66.68
CFS + DI	85.74	85.77	91.30	91.71	92.15	90.53	92.04
RELIEFF + DI	84.91	84.13	90.52	84.91	86.58	78.69	89.01
IG + DI	85.71	81.29	83.74	91.29	89.68	93.66	91.27
Chi-Square + DI	85.72	86.29	89.69	91.27	89.69	90.47	91.30
CFS + DI + PCA	89.64	88.12	92.45	92.15	92.06	91.27	92.83
RELIEFF + DI + PCA	80.18	86.57	84.94	88.12	85.77	88.86	88.89
IG + DI + PCA	72.22	68.26	76.19	74.60	73.81	78.57	79.37
Chi-Square + DI + PCA	90.49	88.09	85.71	86.51	85.72	90.47	92.06
Proposed MPS	**92.18**	**91.28**	**94.17**	**93.89**	**92.68**	**94.98**	**93.08**
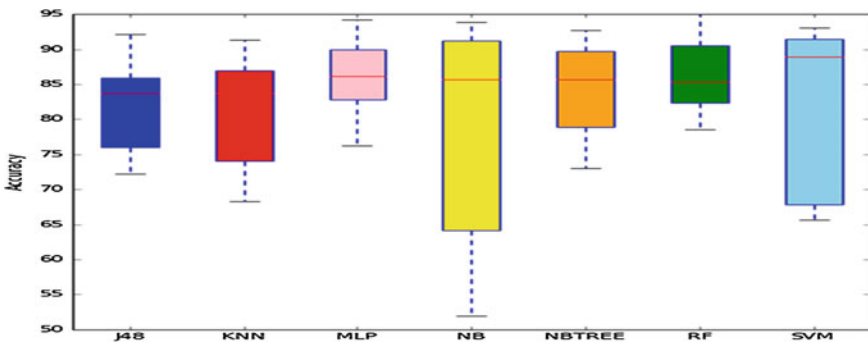
Fig. 2 The performance of respective machine learning methods

the J48, KNN, MLP, NB, NBTREE, RF, and SVM, respectively. The sensitivity is getting the following performance 91.90% for the J48, 91.66% for the KNN, and 93.03% for the MLP, 92.10%, 93.12%, 93.18%, 92.19% for the NB, NBTREE, RF, and SVM, respectively. As shown in Fig. 5, F-measure [42] is estimated by the proposed methods using RF as 94.7%. The RF classification method on LSVT data with an accuracy 94.98%. The sensitivity and precision value of RF classifier are 93.18% and 94.96%, respectively. We investigate the highest F-measure as 94.7% of RF amongst the other classifiers by the proposed method.
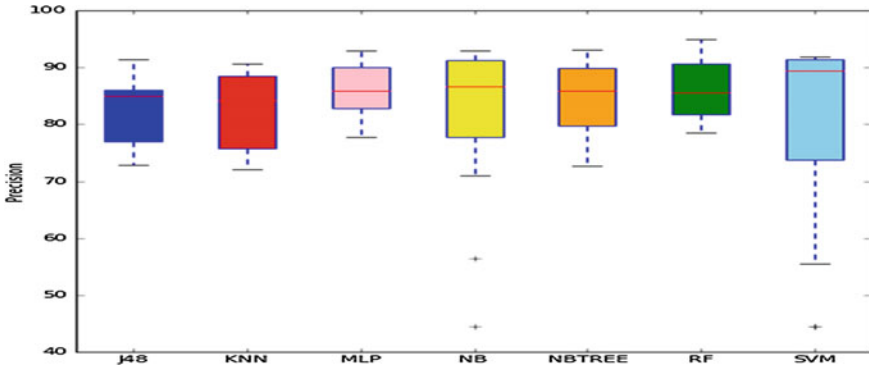
Fig. 3 The performance of precision respective machine learning methods
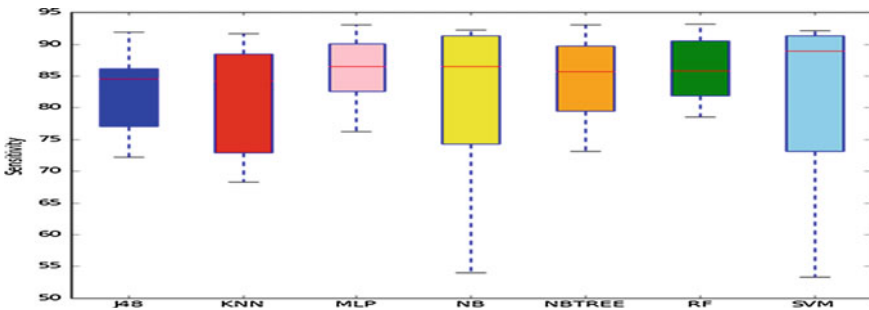
Fig. 4 The performance of sensitivity respective machine learning methods

Table 2 Performances of different classifiers in percentage

Method		Performances of different classifiers in percentage (%)						
		J48	KNN	MLP	NB	NBTREE	RF	SVM
PCA	Precision	72.81	75.80	79.23	77.74	73.62	81.55	80.77
	Sensitivity	72.23	68.31	77.87	74.64	74.65	81.07	79.48
DI	Precision	86.48	89.35	89.64	91.33	88.27	81.93	90.61
	Sensitivity	86.50	88.91	89.72	91.31	88.14	81.79	90.53
CFS	Precision	78.15	83.46	85.74	44.43	82.31	85.60	44.44
	Sensitivity	77.85	83.35	85.77	66.74	82.51	85.79	66.72
RELIEFF	Precision	73.01	75.43	86.52	56.34	79.49	84.16	66.74
	Sensitivity	72.82	76.01	86.53	55.77	78.75	83.52	53.33
Chi-Square	Precision	75.74	72.06	80.91	84.06	79.95	81.04	55.43
	Sensitivity	76.21	69.82	80.25	74.07	80.22	82.08	66.76
IG	Precision	82.34	72.03	80.32	71.05	73.03	85.27	44.44
	Sensitivity	82.50	69.81	79.43	54.04	73.09	84.91	66.76

Table 2 (continued)

Method		Performances of different classifiers in percentage (%)						
		J48	KNN	MLP	NB	NBTREE	RF	SVM
CFS+DI	Precision	85.73	85.72	91.37	91.70	92.14	90.51	91.32
	Sensitivity	85.74	85.72	91.21	91.73	92.02	90.31	91.19
RELIEFF+DI	Precision	84.93	84.14	90.53	84.92	86.50	78.67	88.95
	Sensitivity	84.52	84.23	90.44	85.21	86.73	78.50	88.67
IG+DI	Precision	85.56	88.41	84.70	91.35	89.86	92.90	91.54
	Sensitivity	85.73	88.72	87.10	91.35	89.73	92.70	91.31
CHI-SQUARE+DI	Precision	85.65	81.78	89.67	91.33	89.84	90.40	91.51
	Sensitivity	85.70	82.91	89.7	91.3	89.7	90.5	91.3
CFS+DI+PCA	Precision	89.70	88.72	92.63	92.10	92.10	91.31	91.50
	Sensitivity	89.7	89.1	92.41	92.31	92.10	91.31	91.37
RELIEFF+DI+PCA	Precision	80.31	86.45	84.80	88.08	85.91	89.98	89.46
	Sensitivity	80.22	86.57	84.91	88.16	85.72	88.94	88.95
IG+DI+PCA	Precision	72.84	75.85	77.70	77.72	72.76	78.52	80.70
	Sensitivity	72.21	68.33	76.21	74.61	73.83	78.65	79.43
CHI-SQUARE+DI+PCA	Precision	90.43	88.72	85.91	86.60	85.70	90.89	90.21
	Sensitivity	90.51	88.13	85.73	86.55	85.71	90.57	92.10
Proposed MPS	Precision	**91.40**	**90.71**	**92.93**	**92.93**	**93.08**	**94.96**	**91.90**
	Sensitivity	**91.90**	**91.66**	**93.03**	**92.10**	**93.12**	**93.18**	**92.19**
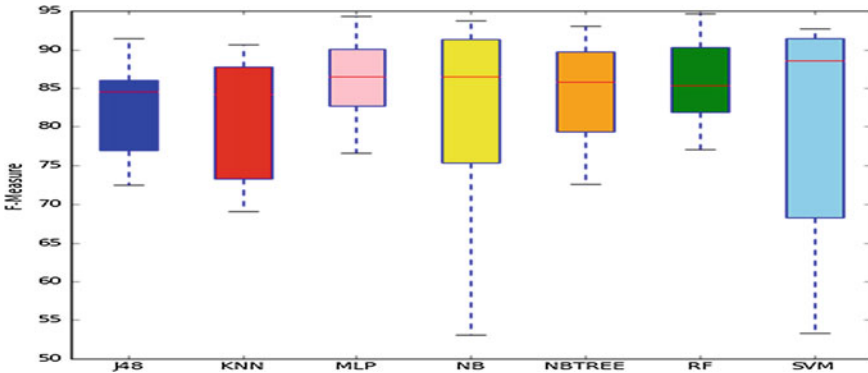
Fig. 5 The performance of F-measure respective machine learning methods

7 Conclusions

Although the individual preprocessing method in the medical dataset is a thoughtful approach but they have few drawbacks. To overcome the drawback of individual preprocessing method, we have developed a multistage preprocessing model, which effectively selects the meaningful features of diseases, to provide efficient prediction. The proposed framework depends on three stages of preprocessing: (a) novel ensemble method for feature selection, (b) discretization, and (c) principal component analysis known as Multi-Preprocessing System (MPS). The experimental results emphasize that proposed (MPS) method provides additional support to significant reduction of features in LSVT dataset and outperforms the state-of-the-art feature selection method in terms of Accuracy, Sensitivity, Precision, and F-measure using seven classification algorithms. The experimental results show that the proposed method using RF achieves the highest performance in terms of accuracy as 94.98%, sensitivity as 93.18%, precision as 94.96%, and F-measure as 94.7%.

References

1. Bashir, S., Qamar, U., Khan, F.H.: IntelliHealth: a medical decision support application using a novel weighted multi-layer classifier ensemble framework. J. Biomed. Inform. **59**, 185–200 (2016)
2. Liu, K.H., Zeng, Z.H., Ng, V.T.Y.: A hierarchical ensemble of ECOC for cancer classification based on multi-class microarray data. Inf. Sci. (Ny) **349–350**, 102–118 (2016)
3. Sakar, C.O., Kursun, O.: Telediagnosis of Parkinson's disease using measurements of dysphonia. J. Med. Syst. **34**(4), 591–599 (2010)
4. Olanow, C.W., Stern, M.B., Sethi, K.: The scientific and clinical basis for the treatment of Parkinson disease. Neurology **72**(4), S1–S36 (2009)
5. Von Campenhausen, S., Bornschein, B., Wick, R., Bötzel, K., Sampaio, C., Poewe, W., Oertel, W., Siebert, U., Berger, K., Dodel, R.: Prevalence and incidence of Parkinson's disease in Europe. Eur. Neuropsychopharmacol. **15**(4), 473–490 (2005)
6. Tsanas, A., Little, M.A., Mcsharry, P.E., Member, S., Spielman, J., Ramig, L.O.: How valid is the clinical diagnosis of Parkinson's disease in the community? J. Neurol. Neurosurg. Psychiatr. **73**, 529–535 (2002)
7. Hartelius, L., Svensson, P.: Speech and swallowing symptoms associated with Parkinson's disease and multiple sclerosis. Folia Phoniatr. Logop. **46**, 9–17 (1994)
8. Tsanas, A., Little, M.A., Fox, C., Ramig, L.O.: Objective automatic assessment of rehabilitative speech treatment in Parkinson's disease. IEEE Trans. Neural Syst. Rehabil. Eng. **22**(1), 181–190 (2014)
9. Remacle, M., Eckel, H.E.: Surgery of larynx and trachea. Surg. Larynx Trachea, 1–308 (2010)
10. Redford, M., Story, B.H.: Mechanisms of voice production. Handb. Speech Prod. 34–58 (2015)
11. Parkinsion disease. National Institutes of Health
12. Holmes, G., Donkin, A., Witten, I.H.: WEKA: a machine learning workbench. Intell. Inf. Syst. 357–361 (1994)
13. Tsanas, A., Little, M.A., McSharry, P.E., Spielman, J., Ramig, L.O.: Novel speech signal processing algorithms for high-accuracy classification of Parkinsons disease. IEEE Trans. Biomed. Eng. **59**(5), 1264–1271 (2012)

14. Tsanas, A., Little, M.A., McSharry, P.E., Ramig, L.O.: Accurate telemonitoring of Parkinsons disease progression by noninvasive speech tests. IEEE Trans. Biomed. Eng. **57**(4), 884–893 (2010)

15. Little, M.A., McSharry, P.E., Hunter, E.J., Spielman, J., Ramig, L.O.: Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. IEEE Trans. Biomed. Eng. **56**(4), 1015–1022 (2009)

16. Sakar, B.E., Isenkul, M.E., Sakar, C.O., Sertbas, A., Gurgen, F., Delil, S., Apaydin, H., Kursun, O.: Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. IEEE J. Biomed. Heal. Informatics **17**(4), 828–834 (2013)

17. Gillespie, A.I., Gartner-Schmidt, J.: Voice-specialized speech-language pathologist's criteria for discharge from voice therapy. J. Voice (2017)

18. Das, R.: A comparison of multiple classification methods for diagnosis of Parkinson disease. Expert Syst. Appl. **37**(2), 1568–1572 (2010)

19. Zuo, W.-L., Wang, Z.-Y., Liu, T., Chen, H.-L.: Effective detection of Parkinson's disease using an adaptive fuzzy k-nearest neighbor approach. Biomed. Signal Process. Control **8**(4), 364–373 (2013)

20. UCI DataSet (Online). https://archive.ics.uci.edu/ml/datasets.html

21. Ferreira, A.J., Figueiredo, M.A.T.: Efficient feature selection filters for high-dimensional data. Pattern Recognit. Lett. **33**(13), 1794–1804 (2012)

22. Wang, A., An, N., Yang, J., Chen, G., Li, L., Alterovitz, G.: Wrapper-based gene selection with Markov blanket. Comput. Biol. Med. **81**, 11–23 (2017)

23. Moradi, P., Gholampour, M.: A hybrid particle swarm optimization for feature subset selection by integrating a novel local search strategy. Appl. Soft Comput. J. **43**, 117–130 (2016)

24. Ang, J.C., Mirzal, A., Haron, H., Nuzly, H., Hamed, A.: Supervised, unsupervised, and semi-supervised feature selection: a review on gene selection. IEEE/ACM Trans. Comput. Biol. Bioinforma. **13**(5), 971–989 (2016)

25. Sun, Y.: Iterative RELIEF for feature weighting: algorithms, theories, and applications. IEEE Trans. Pattern Anal. Mach. Intell. **29**(6), 1035–1051 (2007)

26. Michalak, K., Kwasnicka, H.: Correlation based feature selection method. Int. J. Bio-Inspired Comput. **2**(5), 319–332 (2010)

27. Lai, C., Yeh, W., Chang, C.: Gene selection using information gain and improved simplified swarm optimization. Neurocomputing **218**, 331–338 (2016)

28. Seri, R.: A tight bound on the distance between a noncentral Chi Square and a normal distribution. IEEE Commun. Lett. **19**(11), 1877–1880 (2015)

29. García, S., Luengo, J., Sáez, J.A., López, V., Herrera, F.: A survey of discretization techniques: taxonomy and empirical analysis in supervised learning. IEEE Trans. Knowl. Data Eng. **25**(4), 734–750 (2013)

30. Dash, R., Paramguru, R.L., Dash, R.: Comparative analysis of supervised and unsupervised discretization techniques. Int. J. Adv. Sci. Technol. **2**(3), 29–37 (2011)

31. Jolliffe, I.T.: Principal component analysis. Springer, New York. **2**, 37–52 (1986)

32. Rankawat, S.A., Dubey, R.: Biomedical signal processing and control robust heart rate estimation from multimodal physiological signals using beat signal quality index based majority voting fusion method. Biomed. Signal Process. Control **33**, 201–212 (2017)

33. Hastie, T., Tibshirani, R., Friedman, J.: The elements of statistical learning. Elements **1**, 337–387 (2009)

34. Friedman, N., Geiger, D., Goldszmidt, M.: Bayesian network classifiers. Mach. Learn. **29**, 131–163 (1997)

35. Vapnik, V.N.: An overview of statistical learning theory. IEEE Trans. Neural Netw. **10**(5), 988–999 (1999)

36. Pang, H., Goerge, S.L., Hui, K., Tong, T., George, S.L., Hui, K., Tong, T.: Gene selection using iterative feature elimination random forests for survival outcomes. IEEE/ACM Trans. Comput. Biol. Bioinforma. **9**(5), 997–1003 (2012)

37. Weinberger, K.Q., Blitzer, J., Saul, L.K.: Distance metric learning for large margin nearest neighbor classification. Adv. Neural Inf. Process. Syst. (2006)

38. Twiddle, H.: Neural networks: a comprehensive foundation, vol. 2, p. 41. Prentice Hall (2004)
39. Wu, X., Kumar, V., Ross, Q.J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Yu, P.S., Zhou, Z.H., Steinbach, M., Hand, D.J., Steinberg, D.: Top 10 algorithms in data mining. Knowl. Inf. Syst. **14**(1), 1–37 (2008)
40. Farid, D.M., Zhang, L., Rahman, C.M., Hossain, M.A., Strachan, R.: Hybrid decision tree and naïve Bayes classifiers for multi-class classification tasks. Expert Syst. Appl. **41**(4 PART 2), 1937–1946 (2014)
41. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA data mining software. SIGKDD Explor. Newsl. **11**(1), 10 (2009)
42. Powers, D.M.W.: Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation (2011)

Minimum Shared-Link-Count Forwarding for Alleviating Congestion in Wireless Sensor Networks

Sanu Thomas and Thomaskutty Mathew

Abstract In a wireless sensor network, when two or more forwarding paths share a common link that link is overloaded compared to an unshared link. The overload may cause congestion at the terminal nodes of that link. Therefore, link-disjoint paths are preferred to reduce the eventual congestion. When the topology does not support full link-disjoint paths, the primary objective is to maximize the link-disjoint paths or to minimize the shared-link paths to alleviate congestion. When several minimum shared-link paths are available, we select the least-cost route. Since the data forwarding task is basically, "many to one" communication, the least-cost minimum shared-link paths form a tree. Thus, the goal is to determine the optimal tree which has minimum shared-link paths with least cost. A new iterative algorithm is presented to determine such an optimal tree. The number of iterations is equal to the number of sources. In the first iteration, we construct the shortest path for one of the sources and assign virtual weights to the edges (links) of that path. Then, calculate the next minimum edge weight path from the next source and so on. The virtual weights are manipulated from iteration to iteration in such a way that the least cost and the minimum shared-link criteria are satisfied. The proposed algorithm adjusts the virtual weights and determines the shortest path. This is a centralized algorithm which can be easily converted into a distributed one.

Keywords Minimum shared-link paths · Minimum edge weight paths
Convergecast · Virtual weights · Shortest path tree

S. Thomas (✉)
School of Technology and Applied Sciences, Pullarikunnu, Kottayam, Kerala, India
e-mail: thomas.sanu@gmail.com

T. Mathew
School of Technology and Applied Sciences, Edappally, Kochi, Kerala, India
e-mail: drtkmathew@gmail.com

1 Introduction

In a typical Wireless Sensor Network (WSN), the data sensed by the sensor nodes
are sent to the sink (or the Base Station) by convergecast [1–3] which is same as
many-to-one communication. In general, the converge-cast uses Shortest Path Tree
(SPT) for data forwarding [4, 5] from sensors to the sink. SPT provides the shortest
path from the sensors to the root node (sink). In a simple scenario, the path length is
taken as the sum of the hop counts along the path. In general, the path length is the
sum of edge weights along the path. Therefore, essentially, SPT is the minimum edge
weight path tree. Let the nodes in the WSN be identified by 1, 2, …, N. In a multi-hop
WSN, node i can communicate with node j in a single hop, if they are within the
communication range of each other. Then, node j and i are one-hop neighbors. The
edge weight $w(i, j)$ is 1 (in terms of hop count) when no additional attributes are
specified. With additional edge attributes like bandwidth, reliability, delay, cost, etc.,
the weight can be different from 1. In this paper, we assume that the initial edge
weights are known. In this scheme, the sink is the root of the tree. Thus, this is a
rooted tree.

1.1 Conventional Convergecast and the SPT

The SPT provides the shortest path (path length measured in terms of edge weights)
from the specified source nodes to the root (sink) of the tree. In the conventional
convergecast (CCC), the source nodes forward their data to the sink (root) through the
paths formed by the SPT. The CCC uses the SPT to reduce the overall communication
cost and delay. Consider a simple example of a tree of two source nodes as shown
in Fig. 1a.

The source nodes are **a** and **c**. The sink is **t**. The weight of each edge is taken as
1. In Fig. 1a, the length (or weight) of the path **abt** as well as **cbt** is 2 (or 2 hops).
The edges **ab**, **cb**, and **bt** form an SPT. The total length (or weight) of the SPT is 3.
Source **a** sends its data along **abt** and source **c** sends its data along **cbt**.

Figure 1a is an example of CCC-SPT. The disadvantages of CCC-SPT are conges-
tion at the junctions, unbalanced traffic load along the links, and unequal utilization
of neighborhood nodes. In Fig. 1a, congestion at junction node **b** is higher because it
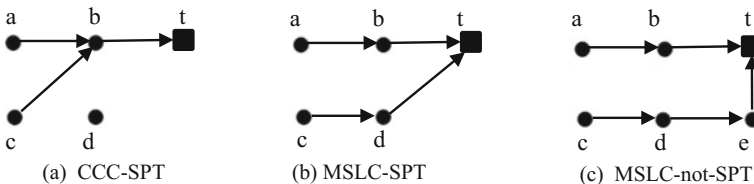
Fig. 1 CCC-SPT, MSLC-SPT, and MSLC-not-SPT

is receiving inputs from the two links **ab** and **cb**. The traffic load on link **bt** is twice that on either **ab** or **cb**. From the utilization point of view, node **b** is overused compared to node **d** which is unused. We mitigate these disadvantages by constructing a special type of spanning tree called Minimum Shared-Link-Count Shortest Path Tree (MSLC-SPT).

1.2 Minimum Shared-Link-Count Shortest Path Tree

In a tree, one or more paths may pass through an edge. In Fig. 1a, edge **bt** accommodates two paths, **abt** and **cbt**, while edge **ab** and **cb** each accommodate only one path. Thus, the share of edge **bt** is higher than that of **ab**. To quantify the *share value* of an edge, we introduce the term Shared-Link-count (SLC) of an edge. The SLC of an edge is the number of distinct paths accommodated by minus 1. This definition is logically similar (but not arithmetically) to the definition of *edge sharability* as given in [6].

The significance of minus 1. We take the normal quota of an edge as one path. Additional paths accommodated are its SLC. When an edge accommodates only one path, that edge is not sharing any other path. Therefore, its SLC (share value) is 0. If an edge belongs to two paths, then its SLC = 1, because it is sharing one extra above its normal quota. In general, SLC of an edge is the number of paths passing through minus 1. The algebraic expression for SLC is given in Eq. (4), Sect. 3.1. In Fig. 1a, the SLCs of edge **ab** and **cb** are 0's while that of edge **bt** is 1. In Fig. 1b, the SLCs of all the edges are 0's.

Minimum Shared-Link-Count Shortest Path Tree (MSLC-SPT) is that tree having minimum total edge weight among those trees having minimum Total Shared-Link-Count (TSLC). In Fig. 1b, the two paths **abt** and **cdt** are edge disjoint. Each edge accommodates only one path. Therefore, the SLC of all edges in Fig. 1b is 0. Therefore, the TSLC is also zero. On the other hand, in Fig. 1a, the TSLC is 1. Therefore, the degree of congestion in Fig. 1a is higher than that in Fig. 1b. Now, let us compare Fig. 1b with Fig. 1c. In Fig. 1c also, the TSLC = 0 as in Fig. 1b. But the path length of the tree in Fig. 1c is 5, whereas the path length of the tree in Fig. 1b is 4. Therefore, the tree of Fig. 1b is the MSLC-SPT while that of Fig. 1c is not an SPT even though it has MSLC. Hence, the tree of Fig. 1b is better than that in Fig. 1c. The significance of SLC is, lower the SLC of an edge, lower is the congestion in that edge. Minimum SLC tree is also a minimum in-degree tree.

Let L be the number of specific leaf nodes to be connected by a tree in a given graph with a given root. There will be L number of paths in that tree. Let T_1 be the set of all possible such trees connecting the L nodes to the root. From T_1, we consider only those trees which have minimum TSLC. Let us denote this set as T_2. Then, from T_2, we choose only those trees which have minimum total edge weight and designate them by the set T_3. Thus, we have $T_3 \subseteq T_2 \subseteq T_1$. Set T_3 may contain more than one member. Trees belonging to T_3 which have minimum TSLC as well as the

lowest total edge cost are defined as the MSLC-SPTs. Our objective is to construct the MSLC-SPT for a given set of source nodes in a given WSN.

1.3 Widespread Convergecast

In Widespread Convergecast, we use MSLC-SPT, whereas CCC uses a Conventional SPT which is hereafter called CSPT. An MSLC-SPT occupies more nodes and hence geographically spread over a wider region compared to a CSPT. A CSPT and an MSLC-SPT are shown in Fig. 2.

In Fig. 2, the source nodes are {3, 4, 5, 7, 8} and the root node is 66. In Fig. 2a, the CSPT is narrowspread and in Fig. 2b, the MSLC-SPT is widespread. The convex hull of the MSLC is bigger than that of CSPT. It is an edge-disjoint tree and the total SLC value is 0. In Fig. 2a, SLC values of certain edges are nonzeros. The SLC values of edges of CSPT of Fig. 2a are shown in Table 1.
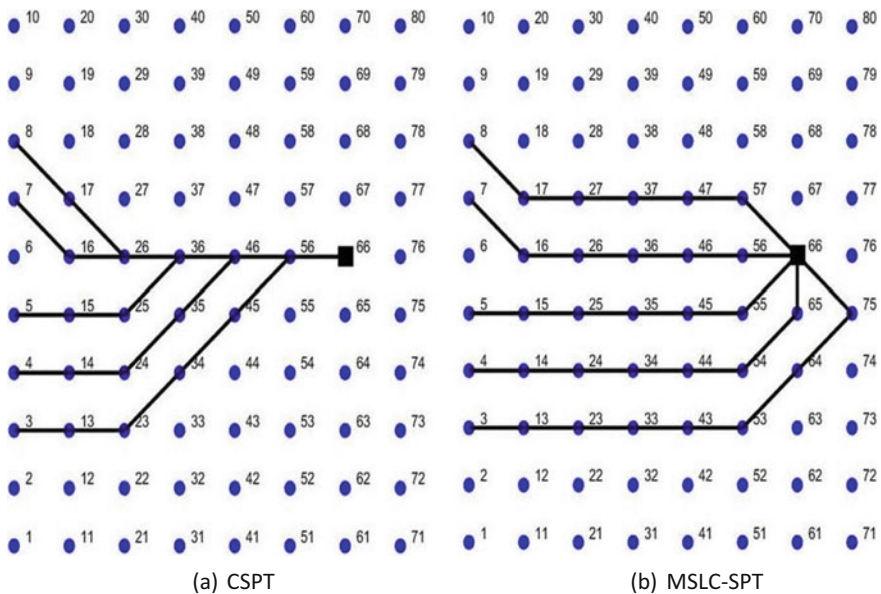
Fig. 2 Narrowspread and widespread trees

Table 1 SLC values at different edges in Fig. 2a

Edges	(26–36)	(36–46)	(46–56)	(56–66)	Others
SLC	1	2	3	4	0

2 Related Works

Considering the network congestion, shortest paths alone are not sufficient. To reduce the congestion, sharing of links should be reduced. In [6], the minimum sharability constraint is discussed in detail. Trees with Minimum Shared Edges are described in [7] and it is shown that the construction of such a tree is an NP-hard problem. Minimum in-degree trees naturally decrease the congestion. Construction of such trees is discussed in [8, 9]. NRBP (Near Receiver Branching Point) is presented in [10] where the branch points are forced to be nearer to the root. Degree constrained minimum spanning trees are studied in [11, 12] using the genetic algorithm and an improved multi-colony ant algorithm is used in [13]. When several objectives are to be optimized simultaneously, it is a multi-criteria optimization. Such multi-criteria shortest path tree problems are solved in [14, 15]. When the number of shared edges is zeros that becomes an edge-disjoint tree. In [16], the authors have given the "binary blocking flow" algorithm for constructing the edge-disjoint paths. Vertex disjoint shortest paths and their complexities are described in [17, 18].

3 Characteristics of an MSLC-SPT

Let the given WSN be represented by a graph $G(V, E)$ where V is the set of vertices (nodes) and E is the set of edges (links) of the WSN. In that graph, MSLC-SPT is a tree that connects a given set of sources (sensor nodes) to the root (sink) through intermediate nodes. The tree is rooted at the sink and the data flow is from the source nodes towards the sink. The MSLC-SPT is a rooted tree rooted at the sink. Data flows from the leaf nodes to the sink. (The tree does not contain any cycles). The out-degree of all the nodes except the sink node is 1. The out-degree of the sink node is 0 (see Fig. 2).

3.1 SLC and TSLC

In our scheme, SLC is an important parameter of an edge in MSLC-SPT. Let S be the given set of K distinct source nodes as, $S = \{s_1, s_2, …, s_K\}$ and t be the destination node (sink). Then, we have K distinct paths as

$$\mathbf{P_i} = path(s_i, t) \tag{1}$$

for $i = 1$ to K. Here, P_i contains a sequence of edges spanning the nodes from s_i to t. We define the edge-path membership [6] index $m(e, P_i)$ for $e \in E$ as

$$m(e, \boldsymbol{P_i}) = \begin{cases} 1 \text{ if } e \in \boldsymbol{P}_i \\ 0 \text{ otherwise} \end{cases} \tag{2}$$

Let $\boldsymbol{P}$ be the collection of all P_i's as $\boldsymbol{P} = \{P_1, P_2, ..., P_k\}$. The sum of edge indices over all paths represented by $m(e, \boldsymbol{P})$ is given by

$$m(e, \boldsymbol{P}) = \sum_{i=1}^{K} m(e, \boldsymbol{P_i}) \tag{3}$$

$m(e, \boldsymbol{P})$ gives the total number of paths passing through edge e. If $m(e, \boldsymbol{P}) = 0$, then e does not belong to the tree formed by $\boldsymbol{P}$. In defining SLC, we consider only those edges ($e \in \boldsymbol{P}$) belonging to the tree or those edges satisfying the condition $m(e, \boldsymbol{P}) > 0$. Then, the Shared-Link-Count [6] of edge e is defined as

$$\text{SLC}(e) = m(e, \boldsymbol{P}) - 1 \tag{4}$$

Here, $e \in \boldsymbol{P}$. If only one path passes through edge e, then there is no sharing and $\text{SLC}(e) = 0$. The total SLC over all edges in the tree is given by

$$\text{TSLC} = \sum_{e \in \boldsymbol{P}} \text{SLC}(e) \tag{5}$$

For low link congestion TSLC has to be kept low.

3.2 Average Shared-Link-Count (ASLC)

When overall link congestion of a tree is to be assessed, we use ASLC which is defined as

$$\text{ASLC} = \frac{\text{TSLC}}{\text{Number of Shared Links}} \tag{6}$$

A Shared Link is one whose SLC is greater than 0.

Example 1. Consider the trees as shown in Fig. 3. In Fig. 3a, SLC(**bg**) = 2 and SLC(**gt**) = 2. The shared links are **bg** and **gt**. The Number of Shared Links (NSL) = 2. The TSLC = 4 and the value of ASLC = 4/2 = 2.

In Fig. 3b, SLC(**bg**) = 1 and SLC(**gt**) = 2. The shared links are **bg** and **gS**. The NSL = 2. The TSLC = 3 and the value of ASLC = 3/2 = 1.5.

In Fig. 3c, SLC(**bg**) = 0 and SLC(**gt**) = 2. The shared link is **gS**. The NSL = 1. The TSLC = 2 and the value of ASLC = 2/1 = 1.

From Fig. 3, we see that the overall congestion is lower in Fig. 3b compared to Fig. 3a, c and the ASLC value is also lower for Fig. 3b. From this example, we can

see that lower the ALSC, lower is the overall congestion. Lowering ALSC is same as minimizing the maximum in-degree of a tree [7–9].

3.3 Nearness of Branch Nodes to the Root

The branch nodes of the MSLC-SPT should be as nearer to the root (sink) as possible [10]. Consider the trees in Fig. 3a, c. In Fig. 3a, the branch node (junction node) **b** is two hops away from **t**, while in Fig. 3c, the branch node **g** is one hop away from t. Therefore, the TSLC is less in Fig. 3c than in Fig. 3a. Therefore, the Distance of the Branch Node from the Root (DBNR) should be as low as possible so as to reduce the link congestion. Thus, while constructing the MSLC-SPT, the values of DBNRs should be as low as possible.

Fig. 3 Trees with different TSLCs and ASLCs
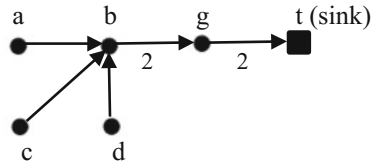
(a) SLC(**bg**) = 2 and SLC(**gt**) = 2.   TSLC = 4. ASLC =2
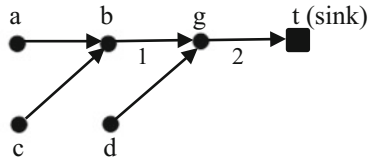
(b) SLC(**bg**) = 1 and SLC(**gt**) = 2. TSLC = 3. ASLC = 1.5
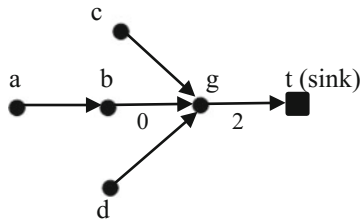
(c) SLC(**bg**) = 0 and SLC(**gt**) = 2.   TSLC = 2. ASLC = 2

3.4 Comprehensive Objective

The objective is to construct an MSLC-SPT for the purpose of Widespread Converge-cast, which minimizes the link and node congestion with the following requirements.

1. Keep out-degree at 1 for all active nodes to prevent cycles.
2. Minimize Shared-Link-Count (MSLC) by

   a. Minimizing Total Shared-Link-Count.
   b. Minimizing Average Shared-Link-Count.

3. Minimize the distance between branch nodes and the Root.
4. Minimize total Edge Weight (or SPT) while satisfying the above requirements.

   A rooted tree that fulfills these requirements is called the optimal MSLC-SPT. To get clear visual graphs and for demonstrating the various characteristics of MSLC-SPT easily, we use *unit grid graphs*.

4   Unit Grid Graph

We use a grid-based WSN. The WSN is represented by a square mesh grid which covers the entire sensing region as shown in Fig. 4. The sensor nodes are placed at grid points. The planar graph $G(V, E)$ is formed from the grid. The grid points are the vertices (nodes) of the graph that are represented by V. The edges of the graph are represented by E.
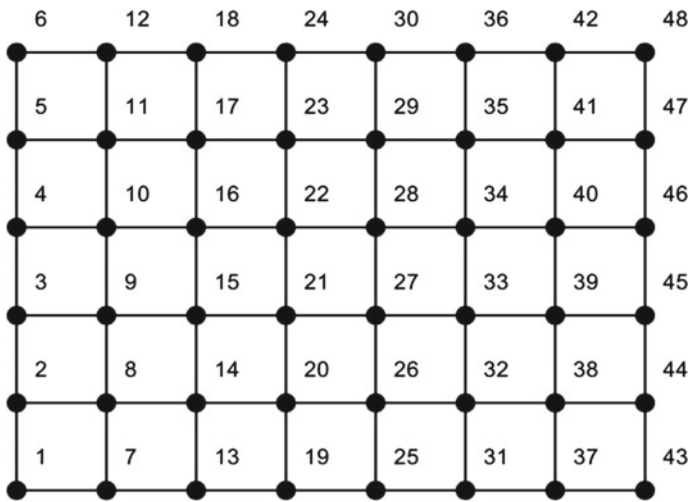
Fig. 4 Grid graph numbering scheme (width = 8 and height = 6 in terms of grid points)

4.1 Node Connectivity

For convenience and good visualization, we use 4-connectivity for all the nodes (vertices). In 4-connectivity, each non-border node is connected to its four immediate neighbors, north, south, east, and west (see Fig. 4). The connections are bidirectional. Thus, G is an undirected graph. The nodes which are more than one hop are not directly connected. The Euclidean distances for the connecting edges are normalized and taken as 1. The distances for the edges between unconnected nodes are set at Infinity. Thus, an edge weight of ∞ means no direct connection. Therefore, the elements of the *Adjacency Matrix* are either 1 or ∞.

4.2 Edge Weights

In the proposed method, the edge weights play a very important role. We represent the weight of the edge which joins node i to node j by $w(i, j)$. Here the indices i and j belong to the range $\{1: N\}$ where N is the total number of nodes.

Basic Edge weights: The basic (initial) edge weight is designated by $w_1(i, j)$. In the square grid graph,

$$w_1(i, j) = w_1(j, i) = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are connected} \\ \infty & \text{otherwise} \end{cases} \tag{7}$$

for $i, j \in \{1: N\}$. Thus, the graph is an undirected (bidirectional) graph. To prevent self-loops, $w_1(i, i)$ is set to ∞.

Rectilinear Paths: The graph is a square grid graph and all the edges are either parallel to x-axis or y-axis. Since a path is a chain of edges, the edges making up the path are parallel to the Cartesian coordinates. Therefore, the paths in the graph are rectilinear.

5 Realization of MSLC-SPT

As mentioned earlier, there are K distinct source nodes and K distinct paths from the sources to the root. The MSLC-SPT is built by constructing one path at a time. It is an iterative process.

5.1 Shortest Path First

Initially, determine the shortest paths and the path lengths using the Dikjstra or any other suitable algorithm with only the basic edge weights without imposing any other constraints of Sect. 3.4. Then, sort the paths in the ascending order of their lengths and reorder the source node set S accordingly as $S = \{s_1, s_2, \ldots, s_K\}$. We start with the

construction of the shortest path $P_1 = P(s_1, t)$ first, then the next shortest path P_2 and so on. The shorter path occupies a lesser number of nodes and edges. Hence, more number of nodes and links are available for choosing the next path which is longer and needs more nodes and edges to satisfy the constraints as given in Sect. 3.4. The first path (the shortest one) is constructed without any consideration of constraints of Sect. 3.4 except the edge weights. This single shortest path automatically satisfies all the constraints.

5.2 Construction of the Second Path

Let the first path P_1 be along s_1 a_1 a_2 a_3 a_4 a_5t as shown in Fig. 5. The length of P_1 is 6. Let the starting point of the second path be s_2. With basic edge weights being 1, there are seven equal shortest paths of length 7, from s_2 to t as shown in Table 2.

The shared links of each possible second path and its TSLC and other parameters are given in Table 2. The branch nodes are highlighted in bold in column 2. In terms TSLC, the best choice is s_2 b_1 b_2 b_3 b_4 b_5 b_6t which has zero TSLC. But when we use a standard Shortest Path Function (SPF) (using popular algorithms like Dijkstra, Bellman—Ford, etc.) to determine the shortest path, it may pick up any one of the seven paths. To inhibit the SPF from selecting the undesirable paths, we assign *virtual edge weights* to the edges of the first path, so that these undesirable (to P_2) edges attain higher weights than the edges of the desired path. Then, the SPF automatically picks the desired path which now has the shortest total weight (length). For example, after establishing P_1, let us assign new *virtual weights* of say 10 each to the edges

Fig. 5 Layout of P_1 and available shortest paths for P_2
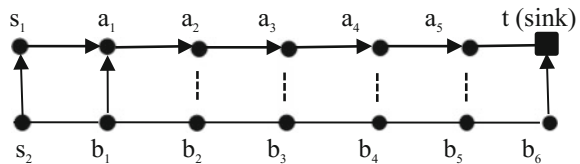
Table 2 Shortest second paths and their TSLC values

Sl. no.	Available paths for forming P_2	Shared links having SLC = 1	TS LC	DBNR	Total weight (TW)
1	s_2**s_1**a_1 a_2 a_3 a_4 a_5t	s_1-a_1-a_2-a_3-a_4-a_5-t	6	6	$6 \times 10 + 1$
2	$s_2$$b_1$**$a_1$** a_2 a_3 a_4 a_5t	a_1-a_2-a_3-a_4-a_5-t	5	5	$5 \times 10 + 2$
3	$s_2$$b_1$$b_2$ **a_2** a_3 a_4 a_5t	a_2-a_3-a_4-a_5-t	4	4	$4 \times 10 + 3$
4	$s_2$$b_1$$b_2$ b_3 **a_3** a_4 a_5t	a_3-a_4-a_5-t	3	3	$3 \times 10 + 4$
5	$s_2$$b_1$$b_2$ b_3 b_4 **a_4** a_5t	a_4-a_5-t	2	2	$2 \times 10 + 5$
6	$s_2$$b_1$$b_2$ b_3 b_4 b_5 **a_5**t	a_5-t	1	1	$1 \times 10 + 6$
7	$s_2$$b_1$$b_2$ b_3 b_4 b_5 b_6t	none	0	0	$0 \times 10 + 7$

of P_1 which is (s_1-a_1-a_2-a_3-a_4-a_5-t). Then the Total Weights (TWs) of the different available paths for P_2 are shown in the last column of Table 2. From this, we see that the SPF picks the path at serial number, Sl. No. 7, whose TSLC $= 0$ and TW $= 7$. Thus, the best path gets selected for P_2 because of virtual weights.

5.3   Shortest Path Function (SPF)

Let the SPF for a graph G with edge weights matrix (adjacency matrix) W, source node s, and destination node t as parameters, be expressed as

$$P = \text{SPF}(W, s, t) \tag{8}$$

where $s =$ source and $t =$ destination (root). P is the output vector of the nodes of the selected path. We use Dijkstra algorithm to get SPF. In the proposed scheme, we successively assign different values to W to get the desired paths.

5.4   Construction of Successive Paths

First path P_1 is calculated as

$$P_1 = \text{SPF}(W_1, s_1, t) \tag{9}$$

Here, W_1 is the initial weight matrix for the entire graph. Then, the weight matrix is modified to get W_2 which depends on P_1, W_1 and the constraints given by Sect. 3.4. The relationship between W_2 and W_1 is written as

$$W_2 = \phi(P_1, W_1) \tag{10}$$

Here, $\phi(..)$ simply represents the mapping to be used to get W_2 from P_1 and W_1. In general, at the jth iteration

$$W_j = \phi(P_{j-1}, W_{j-1}) \tag{11}$$

for $j = 2$ to K. The successively modified weights represented by the matrices, W_2, W_3, …, W_K are called the *virtual weights*.

5.5   Minimization of the Number of Shared Links

Path Weight Vector: At the end of jth iteration, W_j represents the modified edge weights of the entire graph. But when a particular path P_j is of interest, we represent

Fig. 6 Layout that
demonstrates the
significance of q

(a) TSLC=1   (b) TSLC = 0

the edge weights of that path by the vector $W_j(P_j)$. Let $W_j(P_j)$ be represented in terms of its individual edge weights as

$$W_j(P_j) = [w_j(1), w_j(2), \ldots, w_j(u), \ldots, w_j(L)] \tag{12}$$

Here, L is the total number of edges in path P_j and $w_j(u)$ is the weight of the uth edge, for $u = 1$ to L. The edge weight vector $W_j(P_j)$ is a subset of W_j. Now, to prevent P_{j+1} from sharing any links of P_j, we can add ∞ to every element of $W_j(P_j)$ to update it and to get $W_{j+1}(P_j)$ as,

$$w_{j+1}(u) = w_j(u) + \infty \tag{13}$$

for $u = 1$ to L. Here, $w_{j+1}(u)$ is the uth element of $W_{j+1}(P_j)$ and $W_{j+1}(P_j)$ is the updated weight vector of path P_j. Now, the weight matrix of the entire graph W_j is updated to W_{j+1}.

Then, if any edge of P_j is included in determining P_{j+1} by the function SPF (W_{j+1}, s_{j+1}, t), the total weight would be ∞. Therefore, every edge of P_j would be excluded by the function SPF (W_{j+1}, s_{j+1}, t) while generating the path P_{j+1}. But, in some cases the connectivity to the sink may exist only through one or more links of P_j. Then, we have to allow those links of P_j in P_{j+1} also. Therefore, instead of adding ∞ as in (13), we add a large number q, called the edge weight increment parameter, to the edges of $W_j(P_j)$ to get the edge weights of $W_{j+1}(P_j)$ as

$$w_{j+1}(u) = w_j(u) + q \tag{14}$$

for $u = 1$ to L.

Choice of q: There are L edges in $W_j(P_j)$ and each one of them is incremented by q. Therefore, the cumulative increment is $L \times q$. This value $L \times q$ should be greater than the largest estimated roundabout path from s_{j+1} to t through the previously unused (virgin) links. The largest round about path can be taken as the perimeter of the graph (see Fig. 4) in terms of the hop count. In In general q should be greater than the perimeter of the grid graph in terms of the edge weights.

Example 1 Consider the layout as shown in Fig. 6.

In Fig. 6, $s_1 = 7$ and $s_2 = 2$. Path P_1 is constructed along 7-8-t. The weight vector for P_1 is $W_1(P_1) = [w_1(7, 8), w_1(8, t)] = [1, 1]$. Let q chosen be, $q = 2$. Then, from

(a) Out-degree at 8 is 2.   (b) Edges of ∞ weight

(c) Path chosen by SPF(…) after setting the weights of unused
out-degree edges of nodes to ∞

Fig. 7 Layout that demonstrates SWUODEI

(14), $W_2(P_1) = [3, 3]$. Therefore, $w_2(8, t) = 3$. Now, path P_2 that has to start from node 2 has two shortest paths [2, 5, 8, t] and [2, 5, 6, 9, 12, t] both have an equal total weight of 5. Therefore the shortest path function SPF(…) may pick the path [1, 2, 5, 8] which results in link (8-t) getting shared by P_1 and P_2. The SLC of link (8-t) is one and the TSLC$=1$ as labeled in Fig. 6a. Instead of $q = 2$, take $q = 3$. Then $w_2(8, t) = 4$. Then the path weight of [2, 5, 8, t] would be 6 and that of [2, 5, 6, 9, 12, t] would be 5. Therefore, the SPF(…) now picks the path [2, 5, 6, 9, 12, t] which results in TSLC$=0$ as labeled in Fig. 6b. This method of reducing TSLC by incrementing edge weights as given by (14) is named as the Edge Weight Increment (EWI) technique.

5.6 To Limit the Out-Degree of an Active Node to One

Consider the trees shown in Fig. 7.

Here, $s_1 = 5$ and $s_2 = 7$. Path P_1 is [5-8-t]. After applying the EWI method, let the weights of links (5-8) and (8-t) be increased by 10 each. But the weights of link (7-8) and (8-9) remain at the basic value one. Therefore, P_2 selected by the SPF(…) will be along link (7-8) as shown in Fig. 7a. Now, the out-degree of node 8 is 2. To exclude node 8 from P_2, we set the edge weights of the unused (virgin) out-degree edges of node 8 are set to ∞. Thus, the weights of edges {(8-9), (8-5), (8-7)} are set to ∞. In general, the unused out-degree edges of all the nodes along P_1, including the source and destination, are set to ∞ as marked in red in Fig. 7b. Now, out-degree edges (5-2), (5-4), (5-6) of node 5 and (t-12), (t-8) of node t are set to ∞ and are marked in red. Now, P_2 starting from node 7 cannot use edges (5-2), (5-6), (8-9)

because the weights of these edges are ∞, and cannot use (5-8) and (8-t) because of the increased edge weights due to the addition of q which has been set to 10. Therefore, the SPF(…) picks the path [7-4-1-2-3-6-9-12-t] as shown in Fig. 7c. This process of setting the Weights of the unused out-degree edges to Infinity is called SWUODEI operation which is carried out in successive iterations.

SWUODEI operation: At the end of iteration j, path P_j is ready and available. Let $P_j = [v_1, v_2, …, v_L, t]$. Here, L is the path length, $v_1 = s_j$ is the source node and $v_2, …, v_L$ are the intermediate nodes. Let ODE(P_j) represent the out-degree edges of the nodes of P_j. Then, SWUODEI operation is carried out as

(a) Get ODE(P_j), the set of out-degree edges of all the nodes of P_j.
(b) From ODE(P_j) delete the edges of the path P_j to get F as,
F = ODE(P_j) − edges(P_j). // edges(P_j) = Edges of path P_j.
// F gives the unused out-degree edges of nodes along P_j
(c) Set the weights of the edges in F to ∞.
//Other weights of the graph remain same.

SWUODEI operation eliminates the edges of F from the subsequent paths.

5.7 Algorithm MSLC-SPT

INPUTS: Directed graph **G(V, E)** with connectivity matrix W_1 of normalized basic weights. Suitably chosen EWI parameter q. Source list $S = \{s_1, s_2, …, s_K\}$ and the sink node t, where K is the total number of sources.
OUTPUT: MSLC-SPT
1.   Pre-process S to get all the shortest paths from S to t using SPF(W_1, S, t). Arrange them in the ascending order of the path lengths to get the modified source list S. Rename the modified source list as,
$S = \{s_1, s_2,..., s_K\}$.   //Refer section 5. 1.
2.   Get P_1 using the SPF(W_1, s_1, t) and the path length L of P_1.
3.   Set $j =1$.
4. Do EWI operation by calculating the virtual weight matrix W_{j+1} using (14) as, $w_{j+1}(u) = w_j(u) + q$ for $u = 1$ to L.
5.   Do SWUODEI operation to update W_j. //Refer section 5.6.
6.   Increment j as, $j = j +1$. // Now j points to the virtual weight matrix
// calculated in step 4.
7.   Get P_j with this modified weight W_j as, P_j = SPF(W_j, s_j, t).
8.   If $j < K$ // Iterations not over
Go to step 4. //Go to the next iteration
Else
Go to step 9. // Iterations over. MSLC-SPT is ready.
Endif
9. Exit

(a) MSLC-SPT   (b) CSPT

Fig. 8 MSLC-SPT (widespread tree) and CSPT (conventional SPT)

Table 3 Comparison between MSLC-SPT and CSPT

Type	TSLC	No. of shared links	ASLC	Tree length	No. of tree nodes
MSLC-SPT	22	10	2.20	129	130
CSPT	105	19	5.53	28	29

Time complexity of "**MSLC-SPT algorithm**": In the preprocessing step, the SPF(…) is used once and later in algorithm, it used K times. The SPF algorithm by Dijkstra has the time complexity of $O(|E|\log(|V|))$. Therefore, **MSLC-SPT** algorithm has the time complexity of $O(K.|E|\log(|V|))$. The EWI and SWUODEI operations are less time complexity than Dijkstra algorithm and hence do not contribute to the overall complexity.

Example 2 The grid size is 12×14. Number of nodes $N = 168$. The source node set $S = [1, 11, 12, 13, 14, 30, 31, 32, 33, 34, 35]$ with $K = |S| = 11$. The sink node (root) $= 134$. The result of MSLC-SPT from our algorithm is shown in Fig. 8a.

The Conventional SPT [6] (CSPT) is shown in Fig. 8b. Comparison between the two trees is shown in Table 3. Shared-Link parameters are higher in CSPT compared to MSLC-SPT. In Fig. 8b, most of the active nodes and links are highly congested and the traffic is severely unbalanced. The energy of the non-rechargeable batteries of these nodes gets depleted quickly and this causes the early death of these sensors. This, in turn, seriously affects connectivity of the entire WSN. The whole undesirable process is cumulative and the expected life of the WSN will be reduced. In MSLC-SPT the energy consumption is spread out widely and hence prevents the early death of active sensors.

Example 3 MSLC-SPT with 8-connectivity (4-connectivity plus four additional diagonal connectivity, that is, Northeast, Southeast, Southwest, and Northwest).

Fig. 9 MSLC-SPT with 8-connectivity

Fig. 10 Tree construction time versus number of nodes

Here, $t = 153$. Source node set $S = [1, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35]$ with $K = |S| = 11$. Our algorithm works well for an 8-connected graph also. The MSLC-SPT obtained for such a graph is shown in Fig. 9. The tree generated is very nearly edge disjoint compared to the tree of Fig. 8a.

In Fig. 9, we can see that the in-degree nodes are as near to the root as possible. Thus, a low DBNR is achieved.

Example 4 Here, the time taken to construct CSPT and MSLC-SPT are experimentally determined for $N = 210$–364 in steps of 14 and K is kept constant at 8. The time versus N bar graph is shown in Fig. 10.

From Fig. 10, we see that the construction times of CSPT and MSLC-SPT are nearly equal. This shows that the EWI and SWUODEI operations of MSLC-SPT consume negligibly small time compared to that of SPF(…). This is a very important advantage of MSLC-SPT algorithm. Therefore, MSLC-SPT provides low-congestion paths with less time over head.

6   Comparison with Other Methods

In [6] and [7], the minimum shared link SPT is realized using the Minimum Cost Network Flow (MCNF) method which is also an iterative method. In MCNF, a series of graph transformation takes place with edges being added, deleted, and swapped. The time complexity of MCNF is O(K.(|E|logK + |V|log|V|)), whereas our MSLC_SPT has a lesser complexity of O(K. |E|log(|V|). In [8], MD-RPL a distributive algorithm is given which also builds a series of trees each one better than the next by successively reducing the degree of the tree. In [9] the process is similar to that in [8]. On the other hand, in MSLC-SPT, the in-degree of the tree is automatically limited while the paths are generated. Here, once an edge is selected, it is not changed at all in successive iterations. Therefore, MSLC-SPT is faster and more efficient than MD-RPL. Another algorithm NRBP (Near Receiver Branching Point) is presented in [10]. This is a distributed algorithm which requires several types of message exchange. In NRBP, the tree construction time required is large compared to that of MSLC-SPT. Genetic Algorithm is used in [11] which has auxiliary complex operations and may give approximate results for large N. MSLC-MST scales up easily and gives correct results for large N and K.

7   Conclusions

A new algorithm to construct minimum shared-link-count shortest path rooted tree is presented. In this tree, the convergecast is spread wide and that reduces the node and link congestion. This is an iterative algorithm which uses Dijkstra shortest path method in each iteration. In between iterations, the effective edge weights are modified such that the resulting shortest path tree provides minimum shared links and minimum in-degree nodes. The method provides a low shared-link-count. The spreading of the data flow over a wider area offers better load balancing. When a new node joins as a source, the iterative path selection process makes the construction of the new path very easy. Compared to other algorithms, the proposed algorithm has a moderate time complexity and is easy to implement. It can be extended to generate multiple minimum shared-link trees when multiple sinks are present. The method can be extended to generate multipath routes which reduce the overall congestion. This algorithm is a centralized one. But it can be modified into a distributed one.

References

1. Incel, O.D., Gosh, A., Krishnamachari, B., Chintalapudi, K.: Fast data collection in tree-based wireless sensor networks. IEEE Trans. Mob. Comput. **1**, 86–99 (2012)
2. Incel, O.D., Ghosh, A., Krishnamachari, B., Chintalapudi, K.: Fast data collection in tree-based wireless sensor networks **11**(1) (2012)
3. Gandham, S., Zhang, Y., Huang, Q.: Distributed time-optimal scheduling for convergecast in wireless sensor networks. Comput. Netw. **52**(3), 610–629 (2008)
4. Yilmaz, O., Demirci, S.: Shortest hop multipath algorithm for wireless sensor networks. Comput. Math Appl. **63**(1), 48–59 (2012)
5. Guo, W., Park, J.H., Yang, L.T., Vasilakos, A.V., Xiong, N., Chen, G.: Design and analysis of a SPT-based topology control scheme with PSO for wireless sensor networks. In: 2011 IEEE Asia-Pacific Services Computing Conference, Jeju Island, pp. 360–367 (2011)
6. Zheng, S.Q., Yang, B., Yang, M., Wang, J.: Finding minimum-cost paths with minimum sharability. In: IEEE INFOCOM 2007—26th IEEE International Conference on Computer Communications, Anchorage, AK, pp. 1532–1540 (2007)
7. Omran, M.T., Sack, J.-R., Zarrabi-Zadeh, H.: Finding paths with minimum shared edges. In: Proceedings of 17th Annual International Conference on Computing and Combinatorics, Volume 6842 of Lecture Notes on Computer Science, pp. 567–578 (2011)
8. Mamdouh, M., Elsayed, K., Khattab, A.: RPL load balancing via minimum degree spanning tree. In: 2016 IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), New York, NY, pp. 1–8 (2016)
9. Christian, L., Mario. V.-P.: A distributed approximation algorithm for the minimum degree minimum weight spanning trees. J. Parallel Distrib. Comput. **68**(2), 200–208 (2008)
10. Bag-Mohammadi, M., Samadian-Barzoki, S., Yazdani, N.: Near receiver branching point multicast protocol. In: 9th Asia-Pacific Conference on Communications (IEEE Cat. No. 03EX732), vol. 3, pp. 1167–1171 (2003)
11. Shi, K., Song, Q., Lin, S., Xu, G., Cao, Z.: An improved genetic algorithm for degree-constrained minimum spanning trees. In: 2016 Chinese Control and Decision Conference (CCDC), Yinchuan, pp. 4603–4607 (2016)
12. Zhou, G., Gen, M.: Genetic algorithm approach on multi-criteria minimum spanning tree problem. Eur. J. Oper. Res. **114**(1) (1999)
13. Sun, X., Chang, C., Su, H., Rong, C.: Novel degree constrained minimum spanning tree algorithm based on an improved multicolony ant algorithm. Math. Probl. Eng. **2015**(4), 1–13 (2015)
14. Párraga-Álava, J., Dorn, M., Inostroza-Ponta, M.: Using local search strategies to improve the performance of NSGA-II for the multi-C minimum spanning tree problem. In: 2017 IEEE Congress on Evolutionary Computation (CEC), San Sebastian, pp. 1119–1126 (2017)
15. Lu, Y.: The improvement of multi-criterion minimum spanning tree based on the GA. In: 2011 International Conference of Information Technology, Computer Engineering and Management Sciences, Nanjing, Jiangsu, pp. 107–109 (2011)
16. Goldberg, A.V., Rao, S.: Beyond the flow decomposition barrier. J. ACM **45**(5), 783–797 (1998)
17. Itai, A., Perl, Y., Shiloach, Y.: The complexity of finding maximum disjoint paths with length constraints. Networks **12**(3), 277–286 (1982)
18. Kobayashi, Y., Sommer, C.: On shortest disjoint paths in planar graphs. Discrete Optim. **7**(4), 234–245 (2010)

Spatio-temporal Characterization of Axoplasmic Fluid Pressure with Respect to Ionic Diffusivities

Suman Bhatia, Phool Singh and Prabha Sharma

Abstract In this paper, spatio-temporal characterization of axoplasmic fluid pressure has been performed with respect to ionic diffusivities. It has been observed that the propagation speed of axoplasmic fluid pressure is 19.5 m/s when longitudinal ionic diffusivities are considered along with ionic conductances at a temperature of 18.5 °C. However, this propagation speed of axoplasmic fluid pressure increases to 19.7 m/s when longitudinal ionic diffusivities are considered along with temperature-dependent diffusivities across the membrane at the same temperature of 18.5 °C. This is an important result where it has been possible to obtain axoplasmic pressure propagation velocities with respect to temperature-dependent ionic diffusivities. Also, based on the fact that increased intracellular pressure may lead to a number of neuronal disorders, temperature-dependent ionic diffusivities can be further fine-tuned to reduce the intracellular pressure and hence avoiding neuronal disorders.

1 Introduction

Importance of ionic diffusivities for information transmission along the axonal length has been well defined in Hodgkin–Huxley model [11]. Considering their important role, number of experiments have been performed to obtain the diffusion rate of different ions resulting in neuronal excitation and hence information transmission for both myelinated and unmyelinated nerves [2, 12, 14, 16]. As far as, conductance-based neuronal models are concerned, impact of ionic diffusivities has been analyzed for generation and propagation of action potentials [13, 17]. However, there are

S. Bhatia (✉) · P. Sharma
The NorthCap University, Gurugram, India
e-mail: ersuman80@gmail.com

P. Sharma
e-mail: prabhasharma@ncuindia.edu

P. Singh
AGCW, Pondicherry University, Puducherry, India
e-mail: drphoolsinghmaths@gmail.com

other neuronal models based on thermodynamic properties [1] and is also based on axoplasmic fluid properties [6] for describing action potential phenomenon. Neuronal model based on thermodynamics does not take into account the role of ion channel dynamics, whereas the neuronal model based on axoplasmic fluid properties also considers ion channel dynamics defined by H-H model. This model provides a good platform for understanding the behavior of axoplasmic fluid properties with respect to the ion channel dynamics. However, considering the importance of pressure as one of the thermodynamic variables responsible for neuronal excitation [5, 8, 15] and its associated role with anesthesia [4, 7, 9, 10] and number of other neuronal disorders [3, 5], it is quite important to characterize the pressure waves with respect to ionic diffusivities responsible for neuronal excitation. With this view, in this paper, spatio-temporal characterization of pressure waves has been performed with respect to ionic diffusivities. This has been done by including pressure equation in the model given in [6]. The speed at which the pressure wave propagates along the axoplasmic fluid has also been computed with respect to ionic diffusivities at temperature of 18.5 °C.

Apart from abstract and introduction given above, the remaining paper has been organized as follows: Sect. 2 presents the description of variables used in the model followed by the description of our model. Section 3 gives implementation details for solving the model followed by the results given in the Sect. 4. Section 5 concludes the paper and also provides a description of directions for future work followed by references given at the end.

2 Model Description

Table 1 is the description corresponding to different variables/parameters used in the model.

Subscript i in the Table 1 is representative of sodium, potassium, and chlorine ions.

Equation for mass added per unit volume per unit time

$$\frac{\partial \rho}{\partial t} = S_c \tag{1}$$

Pressure equation

$$\frac{\partial P}{\partial x} = -\left(F \times Valency_{Na} \times \frac{\rho Y_{Na}}{M_{Na}} + F \times Valency_K \times \frac{\rho Y_K}{M_K} + F \times Valency_L \times \frac{\rho Y_L}{M_L} \right) \frac{\partial V}{\partial x} \tag{2}$$

Equations for ionic species transport: Sodium Ions, Potassium Ions, and Chlorine Ions Sodium Ions (Na^+) Transport Equation

Table 1 Nomenclature and parametric description for the model

Variables/Parameters	Description	Units
c_m	Capacitance per unit membrane area	F/m^2
D_i	Longitudinal ionic diffusivities	m^2/s
$D_{i,T}$	Trans-membrane ionic diffusivities	m^2/s
F	Faraday constant	C/mol
$\bar{g}_i$	Ionic conductances per unit membrane area	S/m^2
i_{app}	Current applied per unit membrane area	A/m^2
k_B	Boltzman constant	J/K
μ	Dynamic viscosity	Pa s
M_i	Molar mass of ions i	kg/mol
P	Pressure per unit volume	Pa
r	Axonal radius	m
rad_i	Ionic radii i	m
R_a	Resistance per unit axial length	Ω/m
R_u	Universal gas constant	J/K/mol
ρ	Mass density	kg/m^3
S_c	Source term	$kg/m^3/s$
T	Temperature in Kelvin	K
V	Membrane voltage	V
V_i	Equilibrium potential of ions i	V
Y_i	Ionic mass fraction i	Dimensionless
v_F	Axoplasmic fluid viscosity	Pa s
v_W	Water viscosity	Pa s
$\dot{w}_i'''$	Ionic rate of addition per unit volume per unit time	$kg/m^3/s$

$$\frac{\partial(\rho Y_{Na})}{\partial t} = \frac{\partial}{\partial x}\left(\rho D_{Na}\frac{\partial Y_{Na}}{\partial x}\right) + \dot{w}_{Na}''' + \frac{\partial}{\partial x}\left(\frac{\rho D_{Na} Valency_{Na} F Y_{Na}}{R_u T}\frac{\partial V}{\partial x}\right)$$

(3)

Potassium Ions (K^+) Transport equation

$$\frac{\partial(\rho Y_K)}{\partial t} = \frac{\partial}{\partial x}\left(\rho D_K\frac{\partial Y_K}{\partial x}\right) + \dot{w}_K''' + \frac{\partial}{\partial x}\left(\frac{\rho D_K Valency_K F Y_K}{R_u T}\frac{\partial V}{\partial x}\right)$$

(4)

Chlorine Ions (Cl^-) Transport equation

$$\frac{\partial(\rho Y_L)}{\partial t} = \frac{\partial}{\partial x}\left(\rho D_L \frac{\partial Y_L}{\partial x}\right) + \dot{w}_L''' + \frac{\partial}{\partial x}\left(\frac{\rho D_L \, Valency_L \, F Y_L}{R_u T}\frac{\partial V}{\partial x}\right) \quad (5)$$

Equations for ionic species addition: Na^+, K^+ and Cl^- ions

Rate at which Na^+ ions are added per unit volume per unit time

$$\dot{w}_{Na}''' = -\frac{M_{Na}}{Valency_{Na} F}\frac{2}{r}\bar{g}_{Na}m^3 h(V - V_{Na}) \quad (6)$$

Rate at which K^+ ions are added per unit volume per unit time

$$\dot{w}_K''' = -\frac{M_K}{Valency_K F}\frac{2}{r}\bar{g}_K n^4(V - V_K) \quad (7)$$

Rate at which Cl^- ions are added per unit volume per unit time

$$\dot{w}_L''' = -\frac{M_L}{Valency_L F}\frac{2}{r}\bar{g}_L(V - V_L) \quad (8)$$

Equations for gating variables representing their variation with respect to time

$$\frac{\partial n}{\partial t} = \alpha_n(1 - n) - \beta_n n \quad (9)$$

$$\frac{\partial m}{\partial t} = \alpha_m(1 - m) - \beta_m m \quad (10)$$

$$\frac{\partial h}{\partial t} = \alpha_h(1 - h) - \beta_h h \quad (11)$$

Equations for opening and closing of ionic channels representing their dependency on instant membrane voltage V

$$\alpha_n = \frac{0.01(V + 10)}{(\exp\frac{V+10}{10} - 1)} \quad (12)$$

$$\beta_n = 0.125 \exp\left(\frac{V}{80}\right) \quad (13)$$

$$\alpha_m = \frac{0.1(V + 25)}{(\exp\frac{V+25}{10} - 1)} \quad (14)$$

$$\beta_m = 4 \exp\left(\frac{V}{18}\right) \quad (15)$$

$$\alpha_h = 0.07 \exp \left(\frac{V}{20} \right) \tag{16}$$

$$\beta_h = \frac{1}{(\exp \frac{V+30}{10} + 1)} \tag{17}$$

Equation for membrane voltage

$$\frac{\partial(\rho Y_{Na})}{M_{Na}\partial t} Valency_{Na} + \frac{\partial(\rho Y_K)}{M_K \partial t} Valency_K + \frac{\partial(\rho Y_L)}{M_L \partial t} Valency_L = \left(\frac{2}{rF} \right) c_m \frac{dV}{dt} \tag{18}$$

Equation representing the sum of mass added by each ionic species per unit volume per unit time

$$S_c = \dot{w_{Na}}''' + \dot{w_K}''' + \dot{w_L}''' \tag{19}$$

Equation for axoplasmic fluid viscosity

$$vis_W = 2.414 * 10^{-5} * 10^{\frac{247.8}{T-140}} \tag{20}$$

$$vis_F = K * vis_W \tag{21}$$

Here, K is the constant multiplication factor used to obtain axoplasmic fluid viscosity from water viscosity.

Equations for longitudinal diffusivities of ionic species: Na^+, K^+ and Cl^- ions
Equation for longitudinal diffusivity of Na^+ ions

$$D_{Na} = \frac{k_B T}{6\Pi vis_F r_{Na}} \tag{22}$$

Equation for longitudinal diffusivity of K^+ ions

$$D_K = \frac{k_B T}{6\Pi vis_F r_K} \tag{23}$$

Equation for longitudinal diffusivity of Cl^- ions

$$D_L = \frac{k_B T}{6\Pi vis_F r_{Cl}} \tag{24}$$

All the above equations starting from (1) to (24) define the model given in Bhatia et al. [6] incorporating pressure equation which has not been earlier considered in the abovesaid model. The model comprises of parameters namely, density, pressure, ionic mass fraction, ionic addition rate, membrane voltage, axoplasmic viscosity, and temperature-dependent ionic diffusivities defining axoplasmic fluid properties for neuronal excitation.

3 Implementation

The proposed model for characterizing axoplasmic fluid pressure with respect to ionic diffusivities has been implemented using C++. Given below are the important implementation concerns:

- Axonal considerations: For performing the simulations of the model described in Sect. 2, axonal length of 1 m and radius of 0.000238 has been considered. Length of 1 m has been further divided into 10,000 nodes and finite differencing has been utilized for performing discretization.
- External stimulus applied: An external stimulus of 0.3325 A/sq m has been applied at the starting node of an axonal length for action potential generation.

The resting values of different variables have been computed at time $t = 0$ and the forward differencing method has been used for computing boundary values at the upstream end, whereas at the downstream end, all the parameters have been set to resting state values.

4 Results

In the present work, spatio-temporal characterization of pressure along the axonal length has been done with respect to ionic diffusivities when both longitudinal and trans-membrane diffusivities have been made temperature dependent and in case, when only longitudinal diffusivities have been made temperature dependent. Although the results for pressure waves have been plotted at different timings, here we have shown the plots only at two times: time $t1 = 20$ ms and time $t2 = 25$ ms while characterizing the pressure with respect to temperature-dependent longitudinal and while characterizing the axoplasmic pressure with respect to ionic diffusivities when both longitudinal and trans-membrane diffusivities are temperature dependent. However, while comparing the pressure behavior with respect to ionic diffusivities, pressure behavior along the axonal length has been plotted only at a single time $t = 20$ ms. Also, axoplasmic fluid pressure waves are to be interpreted while moving from right to left in the graph and considering the wave to be stationary.

Given below are the important observations from the graphical results:

- Spatio-temporal characterization of axoplasmic fluid pressure along the axonal length with respect to temperature dependent longitudinal diffusivities [Fig. 1]: As depicted in the figure, the speed at which pressure propagates along the axonal length at a temperature of 18.5 °C is 19.5 m/s. This is the case, when the impact of temperature- dependent longitudinal diffusivities have been analyzed on axoplasmic pressure wave. This confirms the pressure wave propagation along with the action potential in the axoplasmic fluid [5].
- Spatio-temporal characterization of axoplasmic fluid pressure along the axonal length with respect to ionic diffusivities when both longitudinal and across the
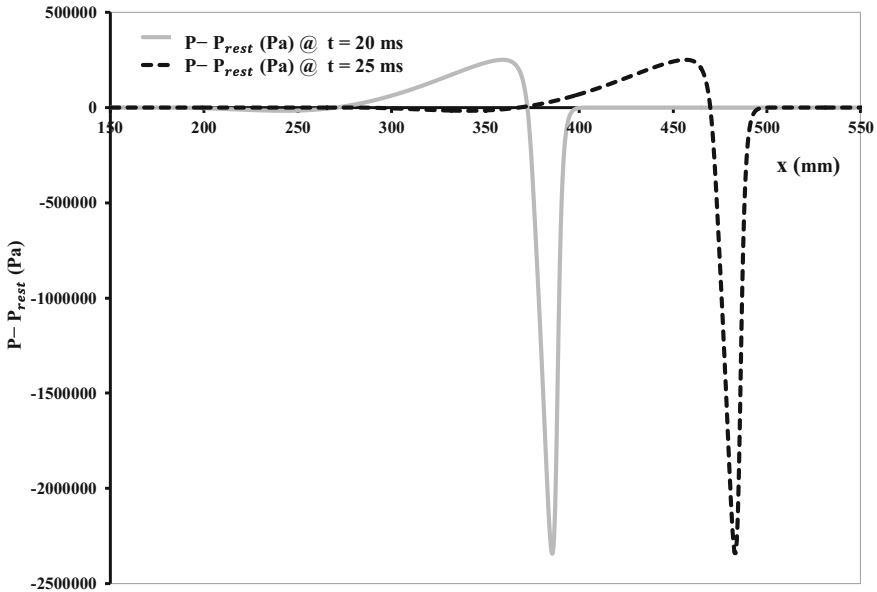
Fig. 1 Here, x-axis represents the axonal length, x(mm) along which the pressure behavior has been plotted at two different times $t1 = 20$ ms and $t2 = 25$ ms. The graph depicts the pressure behavior for longitudinal diffusivities $D_{Na} = 1.849 \times 10^{-9}\,\mathrm{m^2/s}$, $D_K = 1.366 \times 10^{-9}\,\mathrm{m^2/s}$ and $D_L = 1.042 \times 10^{-9}\,\mathrm{m^2/s}$ obtained at temperature of 18.5 °C

membrane diffusivities are temperature dependent [Fig. 2]: The behavior of axo-plasmic fluid pressure has also been analyzed in regard to temperature-dependent longitudinal and trans-membrane ionic diffusivities at two different times $t1 = 20$ ms and $t2 = 25$ ms. It has been observed that when both the ionic diffusivities are temperature dependent, the speed at which the axoplasmic fluid pressure propagates increases to 19.7 m/s at the same temperature of 18.5 °C.

- Comparitive analysis of axoplasmic fluid pressure wave with respect to ionic diffusivities (in case, when both the ionic diffusivities have been made temperature dependent and in case, when only longitudinal diffusivities are temperature dependent [Fig. 3]: As shown in the figure, it can be easily understood that increased propagation of axoplasmic pressure results when both the ionic diffusivities have been made temperature dependent. Inward movement of ions due to temperature-dependent trans-membrane diffusion increases the positive charged density leading to increased interionic repulsive forces and hence an earlier decrease in pressure, whereas reduced negative peak followed by an earlier rise in pressure wave is associated with earlier removal of potassium ions from the membrane.

- Utilizing ionic diffusivities for controlling axoplasmic fluid pressure [Fig. 3]: As it has been observed that when both the ionic diffusivities are temperature dependent, axoplasmic pressure decreases earlier as compared to when we have only longitudinal diffusivities that have been made temperature dependent. Hence to
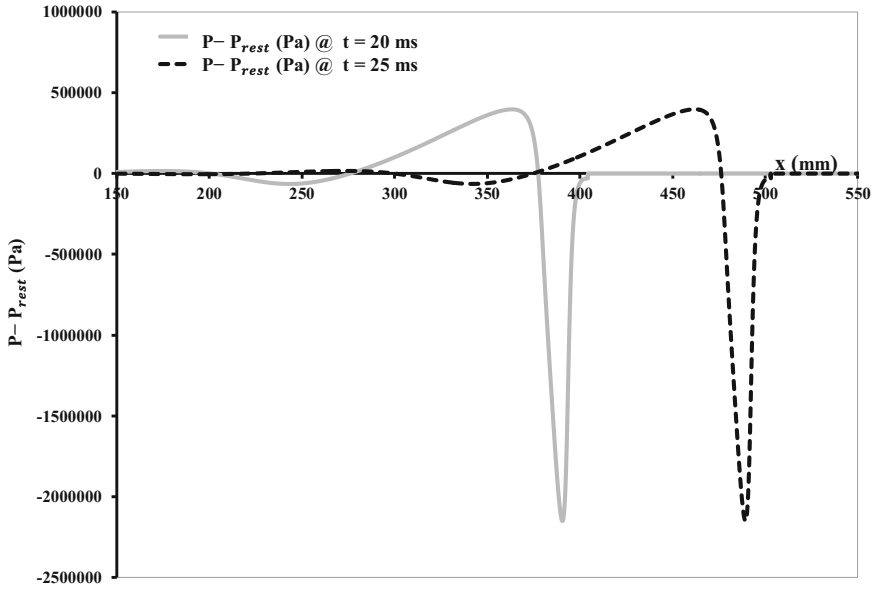
Fig. 2 Here, x-axis represents the axonal length, x(mm) along which the pressure behavior has been plotted at two different times time $t1 = 20$ ms and time $t2 = 25$ ms. The graph depicts the pressure behavior for trans-membrane diffusivities $D_{Na,T} = 1.68 \times 10^{-14}$ m^2/s, $D_{K,T} = 1.24 \times 10^{-14}$ m^2/s and $D_{L,T} = 3.76 \times 10^{-18}$ m^2/s obtained at temperature of 18.5 °C

ensure that axoplasmic fluid pressure is in controlled state as per the requirement for normal neuronal functionality and not resulting in disorders like Alzheimer, threads, tangles, plaques, etc., as mentioned in [3], we can further fine-tune ionic diffusivities to control axoplasmic fluid pressure.

5 Conclusion and Future Work

In the present work, the behavior of axoplasmic fluid pressure has been analyzed in regard to temperature-dependent ionic diffusivities. Although the impact of ionic diffusivities has been analyzed at different temperatures, here the results have been presented only for the temperature of 18.5 °C. Pressure being one of the thermo-dynamical parameters of the model under consideration, can also be analyzed with respect to temperature for providing insights into the thermodynamical perspective of the model. Also based on the fact given in [3], that reduced intracellular pressure waves may lead to a number of diseases like Alzheimer, threads, tangles, plaques, etc., the work can be further extended in a direction to balance the axoplasmic fluid
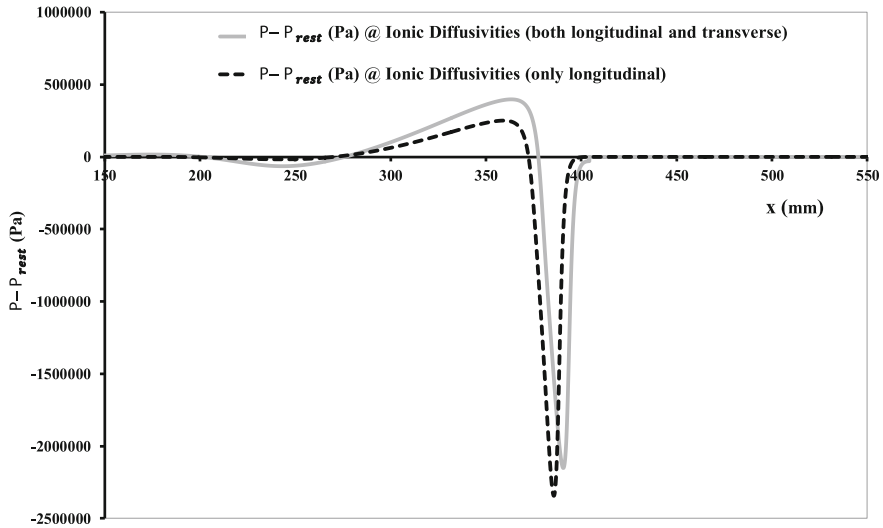
Fig. 3 In this graph, the solid line represents the pressure behavior with respect to ionic diffusivities (both longitudinal and trans-membrane diffusivities): $D_{Na} = 1.849 \times 10^{-9} \text{ m}^2/\text{s}$, $D_K = 1.366 \times 10^{-9} \text{ m}^2/\text{s}$ and $D_L = 1.042 \times 10^{-9} \text{ m}^2/\text{s}$ and $D_{Na,T} = 1.68 \times 10^{-14} \text{ m}^2/\text{s}$, $D_{K,T} = 1.24 \times 10^{-14} \text{ m}^2/\text{s}$ and $D_{L,T} = 3.76 \times 10^{-18} \text{ m}^2/\text{s}$, whereas dashed line depicts pressure behavior with respect to only longitudinal diffusivities: $D_{Na} = 1.849 \times 10^{-9} \text{ m}^2/\text{s}$, $D_K = 1.366 \times 10^{-9} \text{ m}^2/\text{s}$ and $D_L = 1.042 \times 10^{-9} \text{ m}^2/\text{s}$. Both these ionic diffusivities have been computed at $18.5\,^{\circ}\text{C}$

pressure by varying ionic diffusivities. Hence, it also provides a direction for clinical research where ensuring appropriate axoplasmic fluid pressure can help in avoiding abovesaid disorders.

References

1. Andersen, S.S.L., Jackson, A.D., Heimburg, T.: Towards a thermodynamic theory of nerve pulse propagation. Prog. Neurobiol. **88**(2), 104–113 (2009). https://doi.org/10.1016/j.pneurobio.2009.03.002

2. Arhem, P.: Diffusion of sodium in axoplasm of myelinated nerve fibre. Potential clamp analysis. Acta Physiol. Scand. **97**(4), 415–425 (1976). https://doi.org/10.1111/j.1748-1716.1976.tb10282.x

3. Barz, H., Barz, U.: Pressure waves in neurons and their relationship to tangled neurons and plaques. Med. Hypotheses **82**(5), 563–566 (2014). https://doi.org/10.1016/j.mehy.2014.02.012

4. Barz, H., Barz, U., Schreiber, A.: Neuronal Impulse Theory and Alzheimer's Disease. J. Alzheimers Dis. Park. **4**(1), 1–3 (2014). https://doi.org/10.4172/2161-0460.1000134

5. Barz, H., Schreiber, A., Barz, U.: Impulses and pressure waves cause excitement and conduction in the nervous system. Med. Hypotheses **81**(5), 768–772 (2013). https://doi.org/10.1016/j.mehy.2013.07.049

6. Bhatia, S., Singh, P., Sharma, P.: HodgkinHuxley model based on ionic transport in axoplasmic fluid. J. Integr. Neurosci. **16**(4), 401–417 (2017). https://doi.org/10.3233/JIN-170029

7. Causey, G.: The effect of pressure on nerve fibres. J. Anat. **82**(4), 262–270-1 (1948)
8. El Hady, A., Machta, B.B.: Mechanical surface waves accompany action potential propagation. Nat. Commun. **6**, 6697 (2015). https://doi.org/10.1038/ncomms7697
9. Graesboll, K., Sasse-Middelhoff, H., Heimburg, T.: The Thermodynamics of General and Local Anesthesia. Biophys. J. **106**(10), 2143–2156 (2014). https://doi.org/10.1016/j.bpj.2014.04.014
10. Heimburg, T., Jackson, A.D.: The thermodynamics of general Anesthesia. Biophys. J. **92**(9), 3159–3165 (2007). https://doi.org/10.1529/biophysj.106.099754
11. Hodgkin, A.L., Huxley, A.F.: A quantitative description of membrane current and its application to conduction and excitation in nerve. J. Physiol. **117**(4), 500–544 (1952)
12. Hodgkin, A.L., Keynes, R.D.: The mobility and diffusion coefficient of potassium in giant axons from Sepia. J. Physiol. **119**(4), 513–528 (1953). https://doi.org/10.1113/jphysiol.1953.sp004863
13. Liu, Y., Zhu, F.: Collective diffusion model for ion conduction through microscopic channels. Biophys. J. **104**(2), 368–376 (2013). https://doi.org/10.1016/j.bpj.2012.11.3826
14. Qian, N., Sejnowski, T.J.: Electro diffusion model of electrical conduction in neuronal processes. In: Woody, C.D., Alkon, D.L., McGaugh, J.L. (eds.) Cellular Mechanisms of Conditioning and Behavioral Plasticity, pp. 237–244. Springer, US (1988). https://doi.org/10.1007/978-1-4757-9610-0_23
15. Rvachev, M.M.: On axoplasmic pressure waves and their possible role in nerve impulse propagation. Biophys. Rev. Lett. **05**(02), 73–88 (2010). https://doi.org/10.1142/S1793048010001147
16. Tasaki, I., Singer, I., Watanabe, A.: Cation interdiffusion in squid giant axons. J. Gener. Physiol. **50**(4), 989–1007 (1967)
17. Xiang, Z.X., Liu, G.Z., Tang, C.X., Yan, L.X.: A model of ion transport processes along and across the neuronal membrane. J. Integr. Neurosci. **16**(1), 33–55 (2017). https://doi.org/10.3233/JIN-160002

A Simple Reconfigurable Printed Antenna for C-Band Applications

Aakanksha and Bidisha Dasgupta

Abstract A simple reconfigurable printed antenna is reported in this paper. The structure comprises of two semicircular slots separated symmetrically by one thin metallic strip and two PIN diodes, and is excited by a U-shaped microstrip feed line. The PIN diodes are placed at both ends of the strip. The frequency tuning can be achieved in two different operating frequency bands within the C-band corresponding to the two different states of PIN diodes. This frequency-hopping property of the proposed radiator provides 21.88% and 26.3% ($S_{11} < -10$ dB and VSWR < 2) impedance bandwidths (IBW) during the ON and the OFF states of the PIN diodes, respectively. Moreover, the proposed design offers peak gain of about 4.2 and 2.5 dBi during the OFF and ON states of the diodes with low cross-polarization (XP) level in the two principal planes. The antenna provides broadside and monopole-like radiation patterns in the ON and OFF states of the PIN diodes, respectively. Ansys High-Frequency Structural Simulator (HFSS) version 16.2 (Ansys High-Frequency Structural Simulator (HFSS) Ver.16.2, [1]) has been used to carry out the simulations for the presented antenna. The proposed antenna has the potential to be used for wireless and space applications.

Keywords Frequency reconfigurable antenna · Pattern diversity
PIN diode switch · Semicircular microstrip slot antenna

Aakanksha (✉) · B. Dasgupta
Department of Electronics and Communication Engineering,
Indian Institute of Information Technology, Guwahati, Guwahati, India
e-mail: aakaankshaa@gmail.com

B. Dasgupta
e-mail: bidisha@iiitg.ac.in
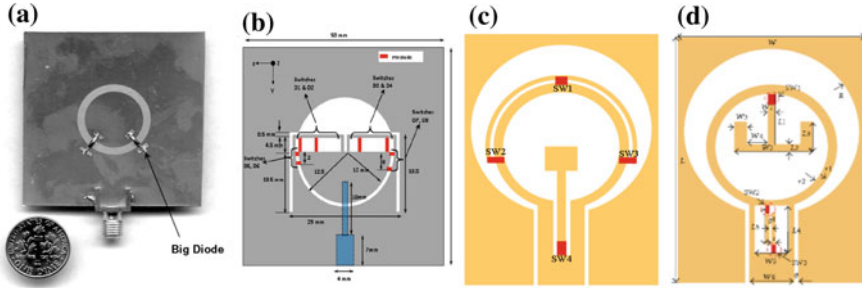
Fig. 1 Some past studies on reconfigurable semicircular/circular slot antennas **a** pattern reconfigurable [5], **b** frequency reconfigurable [6], **c** frequency reconfigurable (dual/single band-notched UWB antenna) [7], and **d** frequency reconfigurable (tuneable dual band-notched antenna) [8]

1 Introduction

Due to greater consumer demand for high-speed wireless communications, and the resulting increase in spectrum use, the use of reconfigurable antennas is being considered as one of methods for optimizing spectrum allocation. Reconfigurability in an antenna can be achieved by changing its operating frequency, radiation pattern, or polarization using different techniques like (i) electrical (by integrating electronic switching components), (ii) optical (by incorporating photoconductive switches), (iii) physical (by altering radiating structure of an antenna), and (iv) using smart materials (like ferrite, liquid crystal, etc.) [2]. Electrical switching techniques like the integration of some radio frequency micromechanical system (RF-MEMS) [3], varicap devices [4], and PIN diodes [5–8] in the antenna structure can easily be used to achieve reconfigurability. Such electrical switching components redistribute the surface current or alter the radiating structure/edges of the antennas to achieve reconfigurability [2]. With the advancement of planar microwave technology, compact printed antennas can be fabricated at low costs. Also, the integration of such electrical switching components is easier with the planar antenna [9]. So, considering the merits of printed and reconfigurable antennas, a frequency reconfigurable radiator is designed and reported in this paper. The parametric optimizations and the study of performance parameters have been done using HFSS, and some of the results are presented in the subsequent sections. Some past studies on reconfigurable semicircular/circular slot antennas are presented in Fig. 1.

2 Configuration

Figure 2 reveals the geometry of the radiator. There are two semicircular slots, separated symmetrically by a thin metallic rectangular strip. Two PIN diodes are incorporated into the structure and are placed at each end of the strip just between the
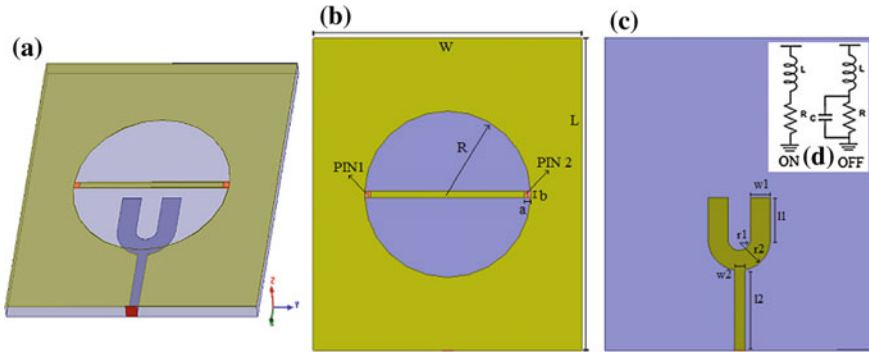
Fig. 2 **a** Perspective view of the proposed antenna, **b** top view (radiator), **c** bottom view (feed), **d** lumped circuit model of PIN diode during forward bias or ON state and reverse bias or OFF state

edge of the rectangular strip and the edge of the circular slot. The antenna is excited by a U-shaped microstrip feed line supported by FR4 substrate (relative permittivity ε = 4.4) of dimensions 50 mm × 43 mm along "x" and "y" axes, respectively, and a thickness of t = 1.6 mm. The ground plane is located at the top of the FR4 substrate. First, a circular slot of radius 13.3 mm with center at the exact center of the substrate is etched out of the ground plane leaving a copper strip of dimensions 1 mm × 26.6 mm along the diameter. Two additional small squares of 1 mm × 1 mm are etched out at the ends of this thin metallic strip and PIN diodes are connected across them as shown in Fig. 2b. The structure in Fig. 2c shows the U-shaped feed. The PIN diodes perform switching function in the design. During the forward biased condition of PIN diodes, the switches are acting in ON state while in the reverse bias, the switches are in OFF state. The equivalent lumped circuit model of the diode in the two states (ON and OFF) is shown in Fig. 2d (inset top right) and the same has been used in simulation [9]. The inductor value is almost negligible in both states (0.09 nH) and does not affect the radiations significantly. The resistor values are 4.7 Ω for ON state and 2 kΩ for OFF state, respectively, with capacitance of 0.017 pF in OFF state [10]. The detailed parameters of the proposed structure are given below: *Ground Plane (GP)*: Length (L) = 50 mm, Width (W) = 43 mm; *Radiator*: Slot radius (R) = 13.3 mm, a = 1 mm, b = 1 mm; *Feed*: $l1$ = 6.7 mm, $w1$ = 3.2 mm, $r1$ = 1.8 mm, $r2$ = 5 mm, $l2$ = 12.9 mm, $w2$ = 1.66 mm.

3 Simulated Results

First, one circular slot antenna with radius R is designed and excited by a simple 50 Ω microstrip line as shown in Fig. 3 (inset left bottom corner). The return loss and radiation characteristics reveal the fact that with increment in R the return loss improves but the peak gain deteriorates. So, the radius R is chosen as 13 mm considering
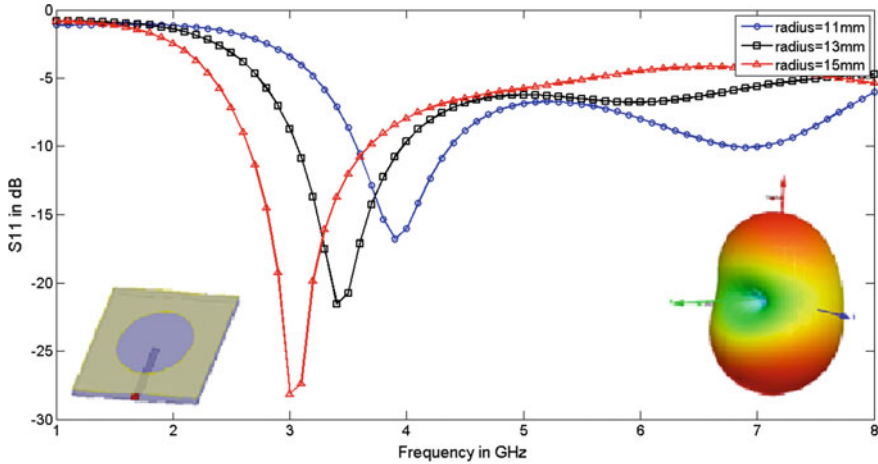
Fig. 3 Simulated S_{11} versus frequency characteristics of the structure shown in inset (left) for different slot radii, along with the far-field 3D plot for R = 13 mm at 3.4 GHz

the trade-off. After that, the feed is modified to be U-shaped. The modification of the feed leads to an improvement in the impedance bandwidth and we get IBW of 106.5% ($S_{11} < -10$ dB). Next, a horizontal metallic strip is incorporated across the diameter (referred to as *antenna I*, Fig. 4.) and then two small rectangular boxes of dimension 1 mm × 1 mm × 0.035 mm were etched out at the ends of diameter (referred to as *antenna II*, Fig. 5.). The return loss characteristics shown in Figs. 4 and 5 reveals the fact that the semicircular slots operate over the frequency band 6–7.4 GHz, whereas the circular slot antenna with thin metal strip operates over the band 3.4–4.5 GHz. The three-dimensional (3D) radiation characteristics are also shown in Figs. 4 and 5 and indicates that at 6.6 GHz, it provides broadside radiation pattern with peak gain 4 dBi, whereas at 4 GHz, it offers monopole-like radiation characteristics with peak gain 2.7 dBi, respectively.

Now, one PIN diode each is connected at both ends of the strip keeping the benefit in mind that when the diodes are ON, the effective geometry is similar to *antenna I* shown in Fig. 4, but when the diodes are OFF, it is like *antenna II* in Fig. 5. The return loss characteristics of the proposed antenna shown in Fig. 6a also verify the fact that at ON and OFF states of PIN diodes, the antenna operates over 5.8–6.2 GHz and 3.4–4.4 GHz, respectively. The proposed antenna geometry is again fed by a simple microstrip line and it reveals the fact that the performance of the U-shaped feed is superior compared to the simple conventional feedline. Sometimes, the microstrip feed causes spurious radiation, but for a good design, the feedline should not radiate. To clarify this issue in present design, the return loss characteristic of only the feedline has been studied and the minimum S_{11} value is about -2.7 dB, which reveals the fact that the feed is not taking part in radiation. So, there is no spurious radiation from the feed. The 3D radiation pattern of the proposed antenna has been shown in Fig. 6b which indicates the pattern diversity of the antenna. The radiation pattern of
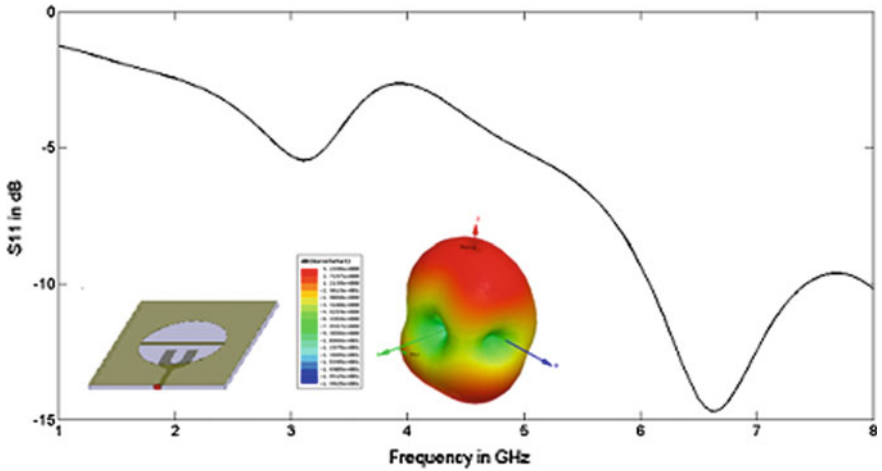
Fig. 4 Simulated S_{11} versus frequency characteristics and far-field 3D plot at 6.6 GHz of the *antenna I* shown as inset (left)
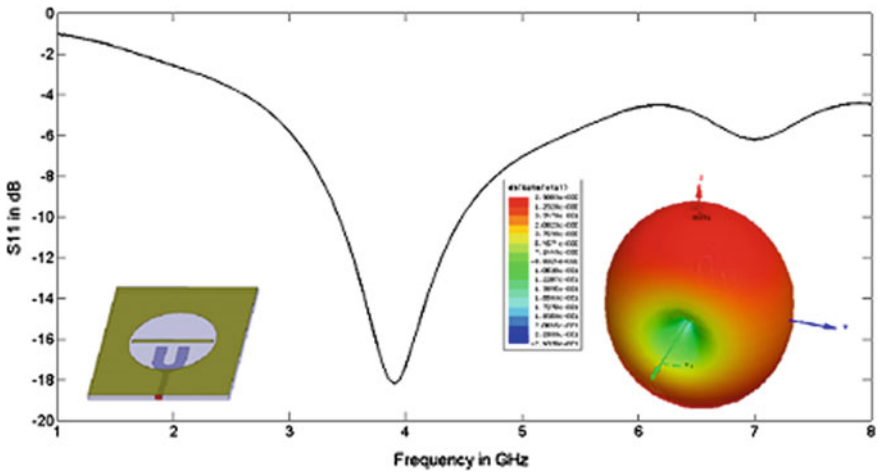
Fig. 5 Simulated S_{11} versus frequency characteristics of the structure and 3D far-field pattern at 4 GHz of the *antenna II* shown as inset (left)

the antenna in two principal planes at two different operating frequencies are shown in Fig. 7 and they indicate low cross-polarization levels.

The surface current densities of the proposed antenna during ON and OFF states are similar to the surface current densities of *antenna I* and *antenna II*, respectively, as shown in Fig. 8a, b. When the PIN diodes are ON, i.e., two similar semicircular slots are radiating. At resonant frequency 6.4 GHz, the current distribution of the lower semicircular slot is prominent and three half-wavelength variations along

(a)

(b)

PIN diode OFF at 3.8 GHz   PIN diode ON at 6.4GHz
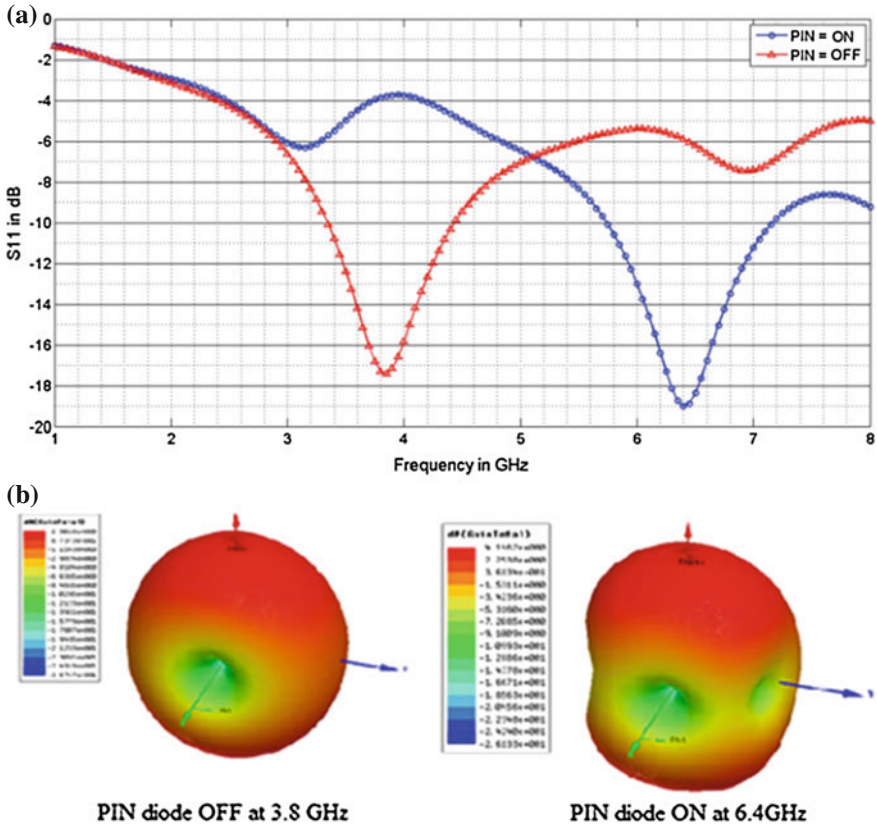
Fig. 6 a Simulated S_{11} versus frequency characteristics of the proposed antenna for PIN ON and OFF states. **b** Far-field radiation patterns at their resonant frequencies

(a)   **(b)**
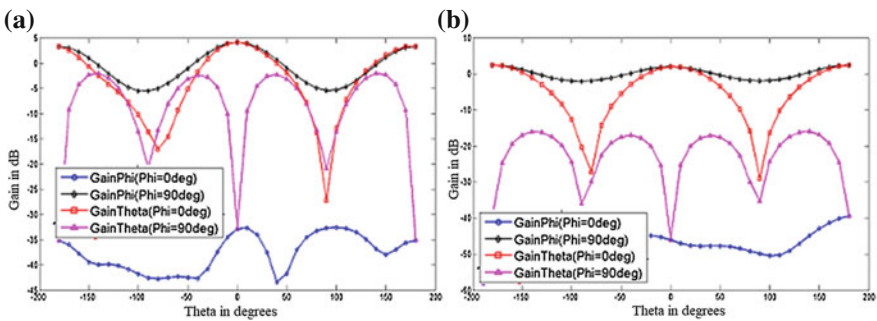
Fig. 7 Field patterns in two principal planes at **a** 6.4 GHz (PIN diodes ON) and **b** 3.8 GHz (PIN diodes OFF)

(a)

PIN ON

No Slots for PIN diodes

(Active Antenna at 6.4 GHz)

(Equivalent Passive Antenna at 6.4 GHz)

(b)

PIN OFF

Slots for PIN diode present

(Active antenna at 3.8GHz)

(Equivalent Passive antenna at 4GHz)
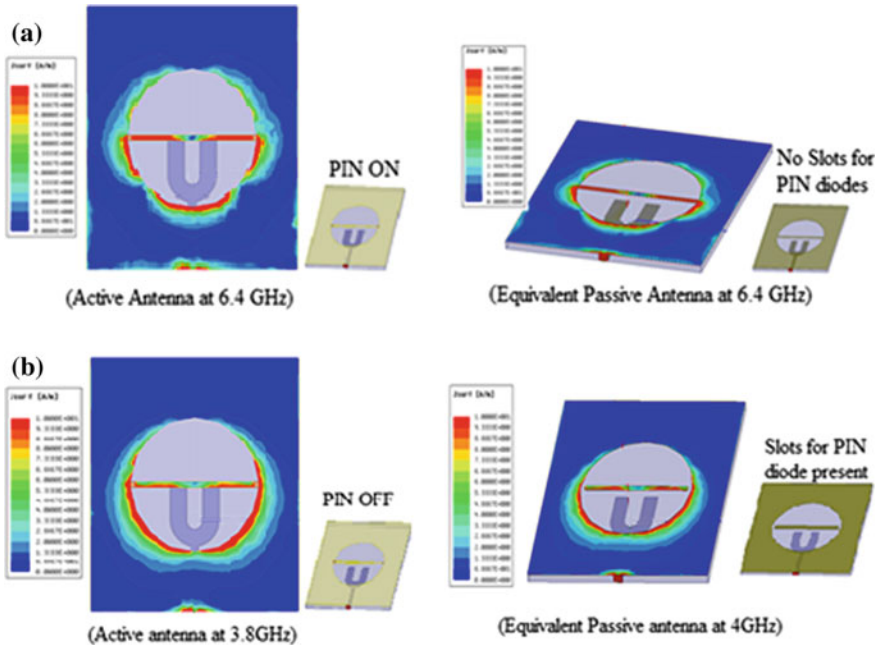
Fig. 8 **a** Surface current density pattern of the active (PIN diodes ON) and equivalent passive antennas at frequencies corresponds to S_{11} minima. **b** Surface current density pattern of the active (PIN diodes OFF) and equivalent passive antennas at frequencies corresponds to S_{11} minima

the perimeter of the semicircular slot are prominent. If λ is the wavelength corresponding to the frequency of S_{11} minimum, then $n(=3)$ times of $(\lambda/2) = $ perimeter of the semicircular slot $= (\pi + 2) R$. Similarly, when the diodes are OFF, i.e., circular slot with strip-line is radiating, at resonant frequency of 3.8 GHz, current distribution is mostly along the perimeter of the circular slot and it has two times of the half-wavelength variation. At resonant frequency 3.8 GHz, $n(=2)$ times of $(\lambda/2) = $ perimeter of the circular slot $= 2\pi R$. So, in the present design, the frequency reconfigurability is obtained by changing the current distribution as well as the antenna geometry by introducing PIN diode switching elements.

4 Conclusion

A simple low-profile, low-cost, and frequency reconfigurable antenna with diversified patterns in far-field is reported in this paper. The frequency tuning is achieved by integrating PIN diodes in the structure. To verify the results of the proposed active antenna in the two states, an equivalent passive antenna design analysis was carried out and was found to be in accordance. The biasing circuitry of pin diodes in

the prototype plays a significant role in the performance of the antenna. So, in the future, before fabrication, further analysis on biasing circuitry will be carried out using some electromagnetic circuit designer tool. Also, many different methods for antenna miniaturization can be employed to make the antenna more compact like using a thicker substrate, introducing matching circuitry, and using artificial magnetic conductor (AMC) surface among others. One potential application of the proposed structure can be in RF front ends for wireless and space communication. Moreover, frequency hopping in the useful bands for WLAN and Wi-Fi applications is also achievable by changing the dimensions of the proposed radiator.

References

1. Ansys High Frequency Structural Simulator (HFSS) Ver.16.2
2. Christodoulou, C., Tawk, Y., Lane, S., Erwin, S.: Reconfigurable antennas for wireless and space applications. Proc. IEEE **100**, 2250–2261 (2012)
3. Balcells, J., Damgaci, Y., Cetiner, B., Romeu, J., Jofre, L.: Polarization reconfigurable MEMS-CPW antenna for mm-wave applications. In: 2010 Proceedings of the Fourth European Conference on Antennas and Propagation (EuCAP), pp. 1–5. IEEE, Barcelona (2010)
4. Korošec, T., Ritoša, P., Vidmar, M.: Varactor-tuned microstrip-patch antenna with frequency and polarisation agility. Electron. Lett. **42**, 1015 (2006)
5. Nikolaou, S., Ponchak, G., Papapolymerou, J., Tentzeris, M.: Design and development of an annular slot antenna (ASA) with a reconfigurable radiation pattern. In: 2005 Asia-Pacific Microwave Conference Proceedings (2005)
6. Sharma, S., Kumar, R., Vijay, R.: A frequency re-configurable semi-circular slot antenna for cognitive radio. In: 2015 1st International Conference on Next Generation Computing Technologies (NGCT) (2015)
7. Li, Y., Li, W., Yu, W., Wan, J.: A small multi-function circular slot antenna for reconfigurable UWB communication applications. In: 2014 IEEE Antennas and Propagation Society International Symposium (APSURSI) (2014)
8. Li, Y., Li, W., Ye, Q.: A CPW-fed circular wide-slot UWB antenna with wide tunable and flexible reconfigurable dual notch bands. Sci. World J. **2013**, 1–10 (2013)
9. Bhartia, P., Bahl, I., Garg, R., Ittipiboon, A.: Microstrip antenna design handbook. Artech House, Boston (2001)
10. Garver, R.V.: Microwave diode control devices, Artech House Publication (1977)

Design and Simulation of Nature-Inspired Patch Antenna with CPW Feed for Dedicated Short-Range Communication Technology Using HFSS EM Simulation Software

Kanika Joshi, Vivekanand Tiwari and Dheeraj Bhardwaj

Abstract Microstrip patch antennas are one of the most widely used antenna structures in communication because of their small size but they also have a disadvantage of having a narrow bandwidth. Here, we have used the superformula given by Gielies, which is used to represent any natural shape mathematically. In this paper, the performance of a triple modified superformula-inspired patch antenna is investigated and all the results are systematically presented. The antenna is designed using Rogers RO4003 substrate ($\varepsilon r = 3.55$) and is excited using a CPW feed. The first bandwidth is of 53% at central frequency 1.23 GHz and the second bandwidth is of 52% at central frequency 5.93 GHz. The 5.9 GHz frequency is used by the dedicated short-range communication technology (DSRC). It is used to enhance the safety of transportation. The bandwidth, directivity, and gain have improved drastically than a circular patch antenna.

Keywords Superformula · CPW · Broadband · Dual band, etc.

1 Introduction

Microstrip patch antennas are the most widely studied antenna structures perhaps due to their inherent properties like compact size, lightweight, and their capability to get integrated with other circuit components. However, their narrow bandwidth and low gain performance restricts their possible applications in modern commu-

K. Joshi (✉)
Department of Electronics and Communication, BIT Mesra, Extension Centre, Jaipur, India
e-mail: kanikajoshi@outlook.in

V. Tiwari
Department of Electronics and Communication, Manipal University Jaipur, Jaipur, India
e-mail: vivekanand.tiwari@jaipur.manipal.edu

D. Bhardwaj
Department of Physics, BIT-Mesra, Jaipur Campus, Jaipur, Rajasthan, India
e-mail: dbhardwaj.bit@gmail.com

nication systems. Broadband antennas are generally preferred in modern wireless communication systems while circularly polarized antennas are preferred in space communication systems. The coplanar waveguide (CPW) feeds are used to increase the compatibility with uniform microwave integrated circuits. In active and passive elements the CPW feed can be easily integrated. By the help of changing the slot and strip widths, one can easily improve the impedance matching and bandwidth [1, 2].

Form in plants and other organisms have interested scientists for a long time. Many geometric shapes and forms like spherical, circular, cylindrical, etc., are commonly observed in nature. Various mathematical tools and equations like Fourier analysis, superellipse equation, etc., were used to mathematically represent the natural shapes. But these equations had the disadvantage of having a limited symmetry. In the year 2003, Gielis presented a paper in which he explained a geometrical approach to define many abstract, naturally occurring, and man-made geometrical shapes and forms using one common formula called the superformula. It considered the number of edges and their depths in the design. This highly increased the symmetry and the correctness of the design. We can produce many types of natural shapes by the variation of different parameters of the formula used [3].

In this paper, we have proposed and simulated a design of nature-inspired shaped antenna using the superformula with CPW feed. All the simulations are performed using the HFSS simulator. The High-Frequency Structure Simulator (HFSS) can solve electromagnetic structures. It is a based on finite element method. The complex RF electronic circuit elements design and simulate by the help of HFSS. Professor Zoltan Cendes developed HFSS at Carnegie Mellon University in 1989 [4]. Its features includes computation of s-parameters and full-wave fields for any 3D passive structures, powerful drawing tools for design entry, powerful post-processor for understanding electrical performance, and model advanced material library including spiral inductors, calculation of far-field patterns, wideband fast frequency sweep, etc.

2 Design of the Antenna

2.1 Circular Patch Antenna

In this research work, we have simulated the radiation performance of a single layered, CPW feed superformula-inspired antenna designed using Rogers RO4003 substrate ($\varepsilon r = 3.55$, $\tan\delta = 0.0027$, substrate thickness "h" = 1.6 mm) [5]. As first step, the dimensions of a circular patch antenna is calculated on the basis of operating frequency band (C band) using the following equations [1]

$$a = \frac{F}{\{1 + \frac{2h}{\pi \varepsilon r F}[1\text{n}\left(\frac{\pi F}{2h}\right) + 1.7726]\}^{1/2}} \tag{1}$$

where $F = \frac{8.791 \times 10^9}{fr\sqrt{\epsilon r}}$.

This antenna is then excited using a CPW feed. The dimensions of the feed are calculated using the following formula

$$f_L = \frac{c}{\lambda} = \frac{7.2}{(L + r + p)} \text{ GHz} \qquad (2)$$

where the height of the planar antenna and radius of the equivalent cylindrical antenna are L and r, respectively, in cm, and p is the length of the 50 Ω feed line in cm. For a circular patch antenna, the value of L = $2a_0$ and r = $a_0/4$. Here, a_0 = a, i.e., radius of the circular patch antenna [2].

After calculations, the final dimensions of the circular patch are radius "Rp" = 18 mm, with ground dimensions "Lg" = 33 mm and "Wg" = 14 mm are designed as shown in Fig. 1 and its radiation performance are simulated by applying HFSS simulator.

The thought process involved in designing the antenna is depicted in the flowchart.

2.2 Results of the Nature-Inspired Circular Patch Antenna

It is detected that the circular patch antenna acts as a narrow band antenna with a central resonating frequency at 4.23 GHz. The impedance bandwidth of this antenna is 0.4394 GHz or 43.97%.

The variation of return loss parameter is shown in Fig. 2 and the VSWR plot of the antenna is shown Fig. 3.

3 The Nature-Inspired Patch Antenna Using Superformula

The superformula given by Gielis [3] is shown in below Eq. (3).

$$r \left(\left| \frac{\cos\left(\frac{m\theta}{4}\right)}{a} \right|^{n2} + \left| \frac{\sin\left(\frac{m\theta}{4}\right)}{b} \right|^{n3} \right)^{\frac{-1}{n1}} \qquad (3)$$

The parameters n1, n2, and n3 and m are positive real numbers in superformula. The number of corners, sectors, or hollows fixed on the shape and the distance between them are determined by value of parameter m. The shape is inscribed or circumscribed on the unit circle which used to determine by the values of n2 and n3. The symmetry of the design is decided by the parameters a and b. Several figures can be created using the superformula, but only some of them can be used for antenna design due to symmetry requirement and limited edge diffraction, etc. Therefore, we chose a design which was symmetric in nature. So, the value of a = b = 1. The
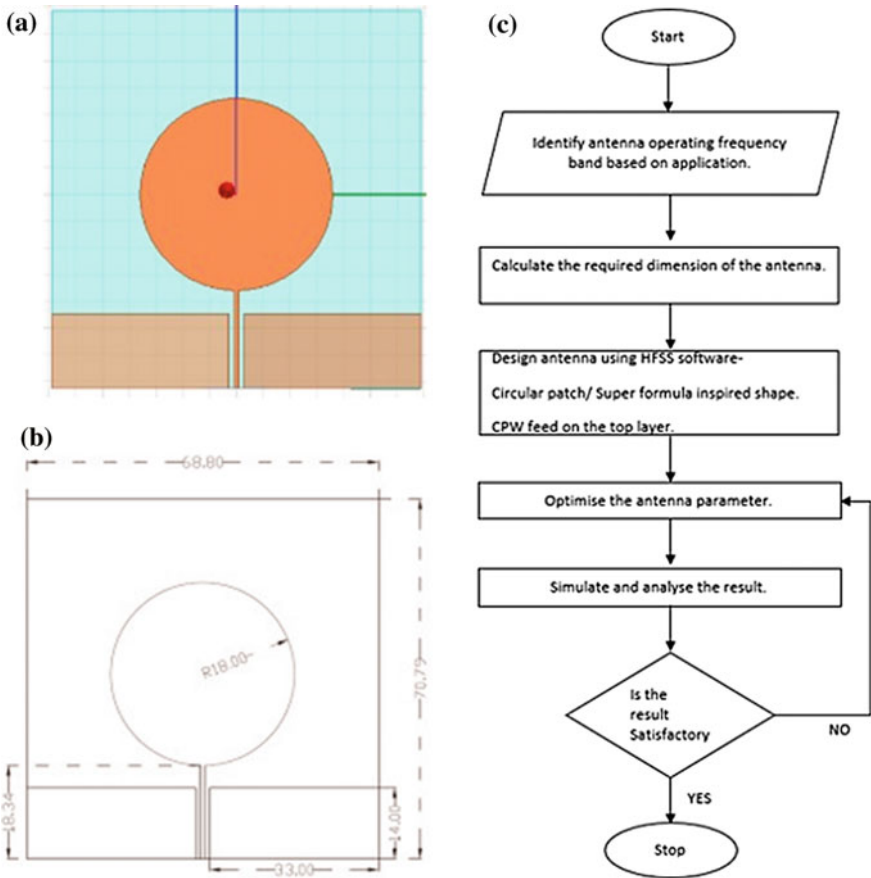
Fig. 1 **a** Simulated design of the circular patch antenna with parameters. **b** Theoretical design of the circular patch antenna. **c** Flowchart of circular patch antenna design

modification will be circumscribed on the edges of a circular antenna. So, value of $n1 = n2 = n3 = 5$. Finally, total number of edges on the on the patch are taken as $m = 30$.

3.1 First Modification for Nature-Inspired Antenna

Keeping the above considerations in mind, the circular patch antenna is modified by adding a polygon modification along the circumference. After optimization, the size of each side of polygon $P1 = 2$ mm. The antenna now resonates at 5.3 GHz frequency but a very narrow band is obtained. The CPW feed design formula is based on the ratio of W/d, i.e., width of the antenna ground and thickness of the feed. Hence, we
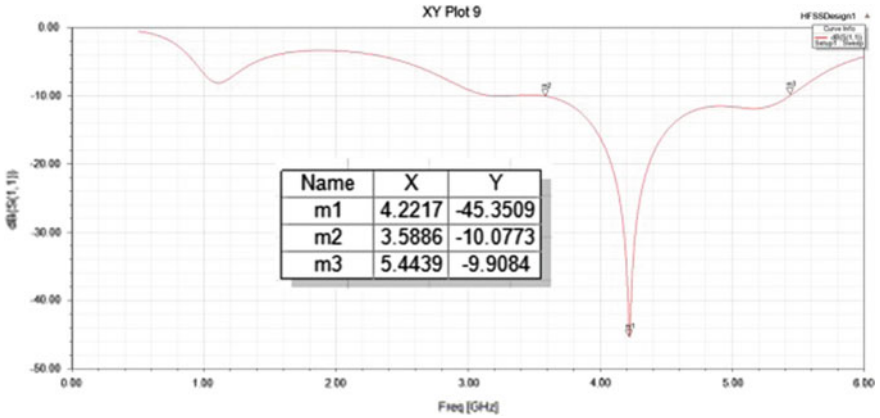
Fig. 2 Variation of return loss parameter as a function of frequency
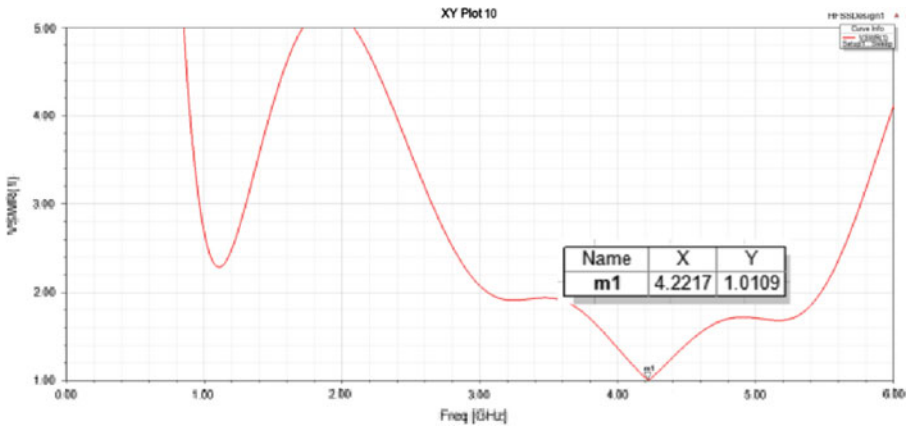
Fig. 3 The VSWR plot of the antenna

are free to modify the length of the ground plane to improve matching and bandwidth. The length of the ground plane and dielectric is then changed to improve bandwidth and the final value of dielectric length is 59.69 mm.

It is observed that the superformula-inspired patch antenna now acts as a dual-band broadband antenna with two central resonating frequencies at 5.035 and 1.034 GHz. The measured impedance bandwidth of this antenna is 0.3074 GHz or 30.74% and 0.3088 GHz or 30.88% (Fig. 4).

The variation of return loss parameter is shown in Fig. 5 and the VSWR plot of the antenna is shown Fig. 6.
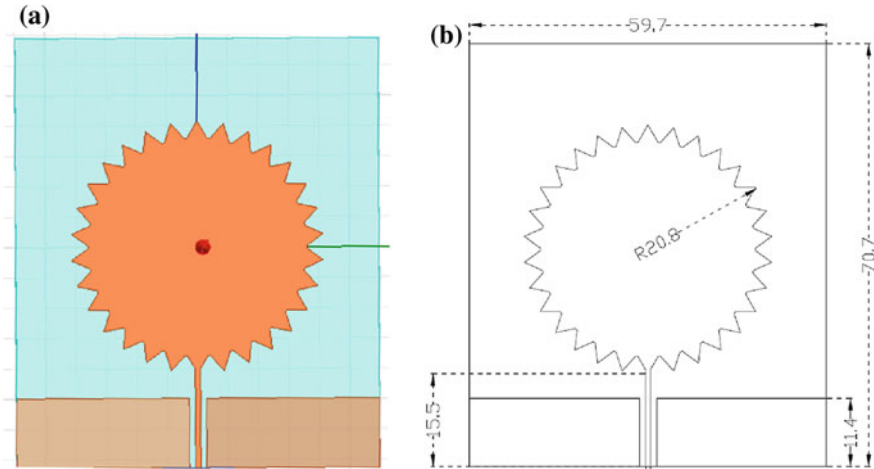
Fig. 4 a Simulated design of the nature-inspired patch antenna with parameters. **b** Theoretical design of the superformula-inspired patch antenna
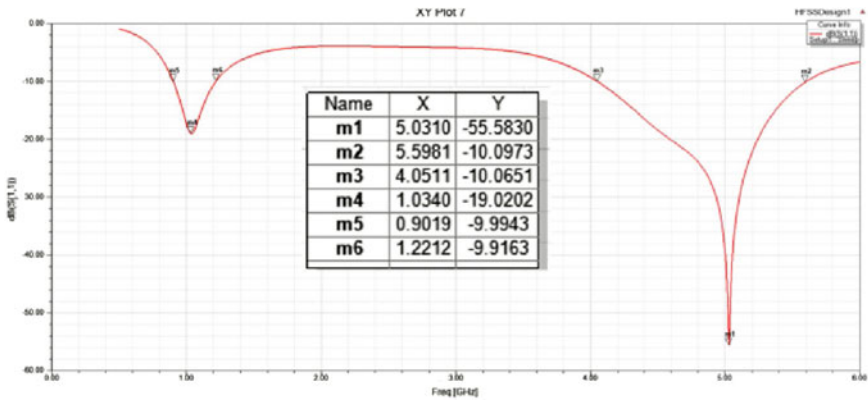
Fig. 5 Variation of return loss parameter as a function of frequency

3.2 Triple Modification for Nature-Inspired Antenna Using Superformula

To further increase the antenna parameters, different geometries with single, double, and triple modifications are studied and shown in Fig. 7. It is observed that after adding the second outer modification, the bandwidth of the antenna increases but the resonant frequency shifts in the forward direction (8 GHz) and the matching also improves.
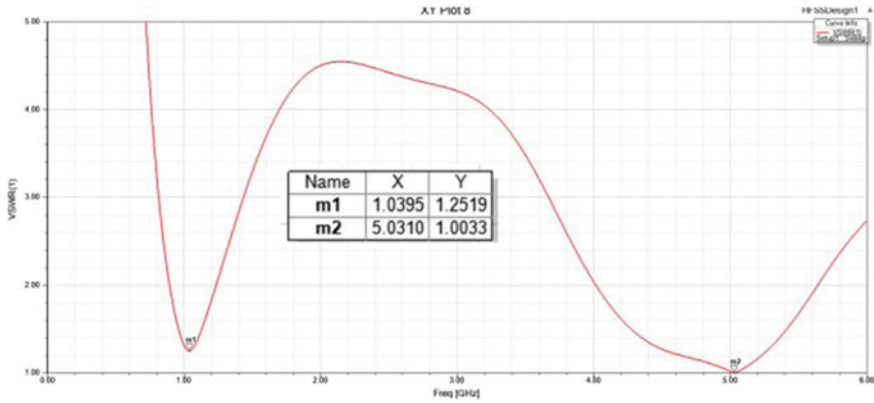
Fig. 6 The VSWR plot of the antenna
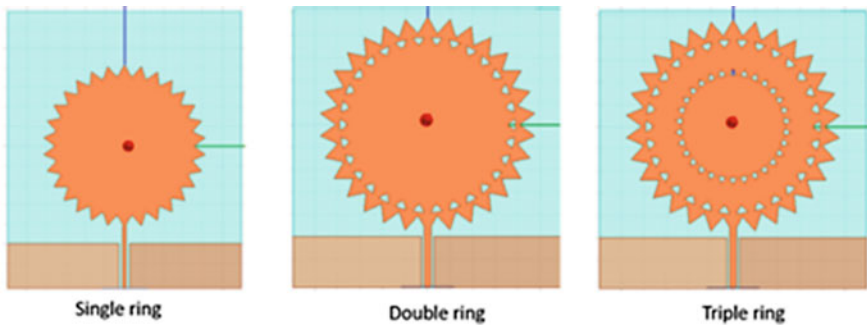
Fig. 7 Different geometries for superformula-inspired patch antenna
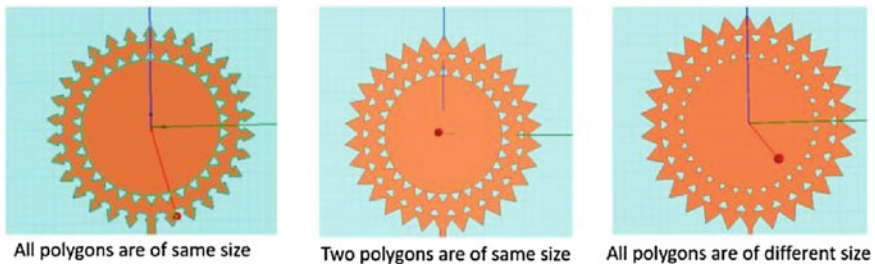
Fig. 8 Different geometries with variable polygon size for superformula-inspired patch antenna

So, a third inner modification is added to the design which shifts the resonant frequency back in the desired range. The size of these three polygons is also varied and studied and shown in Fig. 8.

In the end, the size of each side of polygon P1 on the circumference is 2 mm. The size of each side of polygon P2 on the outer modification is 3 mm and on the inner
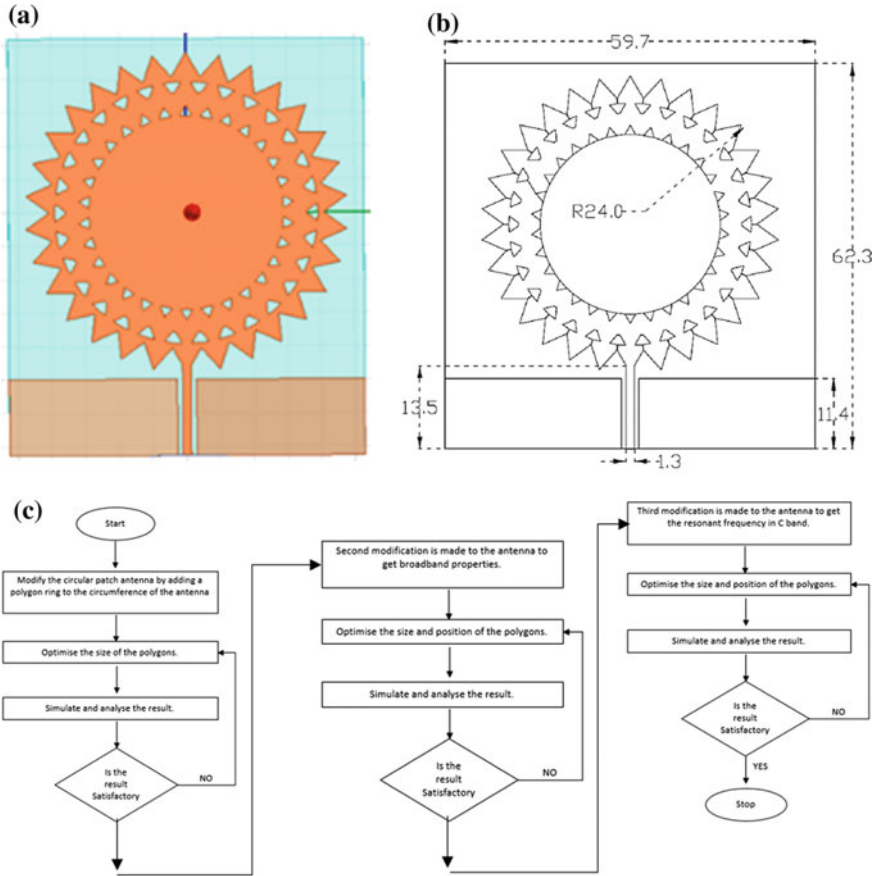
Fig. 9 a Simulated design of the rectangular patch antenna. **b** Design of the superformula-inspired patch antenna with dimensions in mm. **c** Flowchart of superformula-inspired patch antenna design

modification is 0.5 mm. The length of the ground plane and dielectric is reduced by 9.01 mm. Therefore, the new length of ground "lg" $= 28.44$ mm and the new length of dielectric "ld" $= 59.69$ mm and the width "wd" $= 60.4$ mm. The design of the antenna is shown in Fig. 9.

The thought process involved in designing the antenna is depicted in the flowchart.

Name	X	Y
m1	1.2157	-16.4758
m2	1.0140	-9.9724
m3	1.5475	-9.8703
m4	5.9394	-37.4572
m5	5.3604	-28.8958
m6	3.7402	-9.8723
m7	6.3949	-10.0202

Fig. 10 Variation of return loss parameter as a function of frequency

Name	X	Y
m1	1.2157	1.3531
m2	5.9394	1.0272
m3	5.3408	1.0738
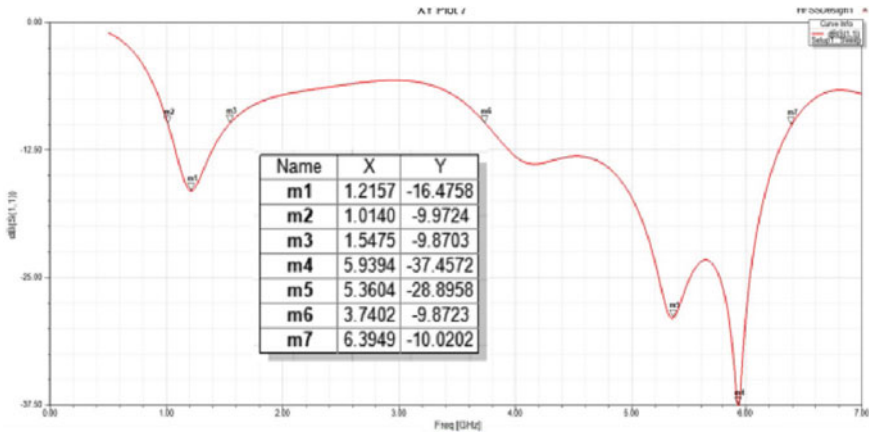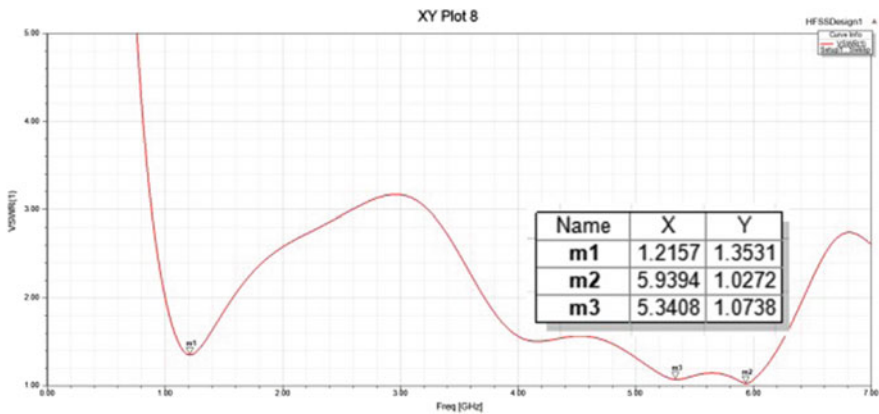
Fig. 11 The VSWR plot of the antenna

3.3 Result of the Nature-Inspired Patch Antenna Based on Superformula

It is observed that the superformula-inspired patch antenna acts as a dual-band broad-band antenna with two central resonating frequencies at 5.93 and 1.23 GHz. The impedance bandwidth of this antenna is 0.5255 GHz or 52.56% and 0.4166 GHz or 41.66%. The variation of return loss parameter is shown in Fig. 10 and the VSWR plot of the antenna is shown Fig. 11.

Table 1 Table of parameters for different antennas

S.no.	Name of antenna	Resonant freq. (GHz)	Bandwidth (GHz)	Ground dimension (Mm)	Peak gain (dBm)	Peak directivity (dBm)
1	Circular patch antenna	4.23	0.43	33 × 11.42	27	26.9
2	Single modification for nature-inspired antenna	5.03, 1.03	0.30, 0.30	28.45 × 11.42	22.23, 32.28	22.26, 28.31
3	Triple modification for nature-inspired antenna	5.93, 1.23	0.52, 0.41	28.45 × 11.42	35, 33	29, 11

4 Conclusion

In this research work, we presented the design and a comparative analysis of a Nature-Inspired microstrip patch antenna based on superformula. We observed that after inserting the three superformula modifications along the circumference and centre gives us a broadband antenna. The antenna now resonated in the C band. A comparison of the various antenna results and parameters is given in Table 1.

References

1. Balanis, A.: Antenna Theory Analysis and Design. Wiley, New York (1997)
2. Ray, K.P.: Design aspects of printed monopole antennas for ultra-wide band applications. Int. J. Antennas Propag. 2008, Article ID 713858, 8 pages
3. Gielis, J.: A generic geometric transformation that unifies a wide range of natural and abstract shapes. Am. J. Bot. **90**, 333–338 (2003)
4. HFSS by Ansys electronics desktop (2015)
5. Omar, A., Rashad, M., Al-Mulla, M., Attia, H., Naser, S., Dib, N., Shubair, R.M.: Compact design of UWB CPW-fed-patch antenna using the superformula. In: 2016 5th International Conference on Electronic Devices, Systems and Applications (ICEDSA), Ras Al Khaimah, United Arab Emirates, 6–8 Dec 2016
6. Pozar, D.M.: Microwave Engineering, Reading, MA, Addition-Wesley (1996)
7. Garg, R., Bhartia, P., Bahl, I., Ittipiboon, A.: Microstrip Antenna Design Hand Book. Artech House, Norwood (2001)
8. Naser, S., Dib, N.: Design and analysis of super-formula-based UWB monopole antenna and its MIMO configuration. Wireless Pers. Commun. **94**(4), 3389–3401 (2017)

Cardiac Arrhythmia Classification Using Machine Learning Techniques

Namrata Singh and Pradeep Singh

Abstract Cardiac arrhythmia refers to the medical condition during which the heart beats irregularly. Effective monitoring of cardiac patients can save enormous amount of lives. During the past few years, much importance has been gained by cardiac disease classification and prediction. This paper presents a model for diagnosis of cardiac arrhythmias. It works by selecting best features with the help of three filter-based feature selection methods on three different machine learning methods applied over cardiac arrhythmia dataset. Feature selection is a crucial preprocessing step in determining factors responsible for patients suffering from arrhythmia. In particular, we want to examine the underlying health factors of patients that could potentially be a powerful predictor for deaths that are related to heart. Three types of machine learning methods, namely, linear SVM, random forest, and JRip, were employed for analyzing the performance of the feature selection methods. Experimental analysis shows that highest accuracy of 85.58% was obtained with random forest classifier using gain ratio feature selection method with a subset of 30 features.

Keywords Cardiac · Arrhythmia · Disease prediction · Heartbeat classification
Feature selection

1 Introduction

Clinical biomedical datasets are getting accumulated over the years owing to the advancements in data collection and storage technology methods. The past clinical studies have shown the identification of disease and its likely consequence performed through analysis of electronic health records. These records contain valuable infor-

N. Singh (✉) · P. Singh
Department of Computer Science & Engineering,
National Institute of Technology, Raipur, Chhattisgarh, Raipur, India
e-mail: nsingh.phd2016.cs@nitrr.ac.in

P. Singh
e-mail: psingh.cs@nitrr.ac.in

mation about the features which are the potential risk factors and other information related to the causation of the disease [1, 2]. The advancement in hardware and software technology has helped in improving biomedical research investigations and numerous clinical studies. The applications of various computational intelligence and machine learning algorithms on the cardiac arrhythmia dataset would help in accurate diagnosis, prognosis, and making clinical decisions.

Cardiac arrhythmia or irregular heartbeat refers to the conditions where the heartbeat becomes irregular, either too fast or too slow. If the heart rate is too fast, i.e., above 100 beats per minute, then the situation is known as tachycardia, and if the heart rate is too slow, i.e., below 60 beats per minute, then it is known as bradycardia. These conditions occur due to problems with the electrical conduction system of the heart. Statistics reveal that a significant number of persons are suffering from heart attack, stroke, or other cardiac-related diseases. These cardiovascular diseases are one of the major causes of death worldwide and are considered as one of the most widespread and costly health problems faced by the nations [3, 4].

The electrical activity of the heart is represented by the electrocardiogram (ECG) and is the main component in disease diagnosis. It is considered as one of the basic signals for detecting the abnormalities of heart and diagnosing the patient's condition. The important features of ECG are extracted and are applied for making decisions [5]. Hence, for diagnosing the cardiac problem, proper feature description and its extraction or selection become very crucial before classification. ECG consists of basic three waves, i.e., P, QRS, and T. A normal cardiac cycle is composed of a P wave, the QRS complex, and finally the T wave. It is mainly shown by three segments, viz., the TP segment, the ST segment, and the PR segment, and the corresponding time periods (intervals) are given as R-R interval, RT interval, and the PR interval [6, 7].

Machine learning plays a vital role in disease classification and prediction in the medical domain. It provides immense help to the physicians to manage the large amount of captured medical data. Various machine learning and data mining available provide deeper and complete understanding of biomedical data and also solutions to the complex nonlinear problems [8]. With the help of these methods, early and better diagnosis of diseases can be done that can save the costs and expensive medical tests [9]. The different data mining techniques employed by the various researchers can help in treatment of diseases such as diabetes [10], heart failure [11] and cardiovascular disease such as myocardial infarction risk prediction.

The main objective of the research paper is to prove that a minimum subset of features can provide diagnostic models which are economical, accurate, and reasonable. The work performed in this paper consists of performance evaluation of three filter-based feature selection methods, viz., χ^2-statistic, symmetric uncertainty, and gain ratio on the prediction of patients suffering from cardiac arrhythmia. Moreover, for evaluation of the performance of the feature selection methods three classification methods, namely, linear SVM, random forest, and JRip, are used.

The remaining sections of the paper are arranged as follows. Section 2 exhibits the review of the recent work done over arrhythmia dataset and applying various techniques of feature selection and machine learning with conventional and many

recent classifiers. Section 3 gives an introduction about the class distribution of the arrhythmia disease dataset, and the detailed illustration of the three feature selection techniques and three classification methods. Section 4 illustrates the experimental results and corresponding discussions. Finally, the conclusion of the paper is provided in Sect. 5.

2 Related Work

Medical data mining discovers unknown patterns from biomedical datasets efficiently. These patterns can then be utilized for disease diagnosis and treatment. Past researches prove that artificial intelligence techniques such as an artificial neural network (ANN), support vector machine (SVM), naïve Bayes (NB), decision support, and regression have shown good performance in classifying and predicting biological and medical data. Following research papers focus on using different feature selection and machine learning techniques on arrhythmia dataset for medical data analysis and prediction.

In this paper [12], the authors have used various data mining schemes such as naïve Bayes, J48, and OneR for classification of arrhythmia from ECG medical datasets. The aim of the work done was to classify cardiac arrhythmia and to analyze the performance of different machine learning algorithms. OneR and naïve Bayes showed the most stable accuracy while J48 algorithm was unable to perform well.

The research paper [13] showed an implementation of SVM, random forests, naïve Bayes, neural network, and feature selection techniques as utilized by the existing papers on the cardiac arrhythmia dataset. Also, an implementation of a learner using a blend of RF and linear kernel SVM was performed which gave a classification error of 77.4%. Thus, the work done provided some slight improvement over the generalization errors reported before.

In this paper [14], after application of data preprocessing and feature selection techniques, the classification is done using five different machine learning algorithms, namely, logistic regression, SVM, random forest, decision trees, and KNN. The best accuracy obtained with SVM was 73%. During data preprocessing, z-score normalization and some of the attribute selection methods were used to identify the most important attributes for determining the class of arrhythmia.

The paper [15] has proposed a classification technique using combination of machine learning methods and ECG diagnostic criteria which improved the accuracy of detecting arrhythmia disease using electrocardiogram (ECG) data. After performing data preprocessing and feature selection procedures, different machine learning techniques such as neural networks, decision trees, random forest, gradient boosting, and support vector machines were applied. An accuracy of 84.82% was achieved with the combination of SVM and Gradient Boosting. The classification performance was evaluated using measures such as confusion matrix, kappa score, confidence interval, area under ROC curve (AUC), and overall accuracy.

This study [16] focuses on finding the attributes that describe the arrhythmia dataset more better in 278 attributes using genetic algorithms. A classification accuracy of 90% is achieved by using the attributes from the outcomes of Genetic Algorithm (GA). The fitness function was to minimize the classification error. It was found that the classification accuracy that was achieved from 278 features can also be acquired from the seven features. These features were heart rate, linear existence of ragged R wave, and other five measurements that were taken from ECG.

The paper [17] describes a new approach to ECG cardiac arrhythmia analysis using Hidden Markov Models (HMM). Most of the ventricular arrhythmias can be predicted by the detection of QRS complexes and R-R intervals. This approach of HMM is a combination of both structural and statistical knowledge of the ECG signal.

The paper [18] denotes the survey of ECG-based heartbeat classification for arrhythmia detection. It presents a review of various techniques of ECG-based automated abnormalities of heartbeat classification by presenting the ECG signal processing, the heartbeat segmentation methods, feature description techniques, and the data mining algorithms used. Additionally, the description of some of the repositories, limitations, and future challenges are also proposed in the literature.

Modular neural network (MNN) model has been used [19] to classify arrhythmia into normal and abnormal groups. The neural network model has been constructed by varying the number of hidden layers from one to three and evaluating the classification performance using six different measures. The results have shown the testing classification accuracy up to 82.22%.

3 Materials and Methods

The techniques of feature selection and classification are being used widely for disease diagnosis and prognosis. The results for managing and interpreting the high dimension biomedical datasets are very encouraging and have provided significant achievement for the classification accuracy and other performance parameters. The details of filter-based feature selection approaches and various types of learners such as individual, ensemble, and rule-based learners are described below.

3.1 Feature Selection Methods

The techniques of feature selection applied in this paper consist of three filter-based chi-square, symmetrical uncertainty, and gain ratio feature selection algorithms. The filter-based feature selection methods are independent of the classifier, and evaluation of the characteristics of a feature subset uses the inherent properties of the training data; therefore, these prove to be flexible when combined with various other learners and also utilize less computational complexity as well as time.

Chi-square Statistic. The value of χ^2-statistic is calculated for each feature individually with respect to the classes. Each numeric attribute is discretized before the computation of χ^2-statistic. For each feature, this statistic is computed as

$$\chi^2 = \sum_{x \in X_i} \sum_{c \in C} \frac{(n_{(x \in X_i \& c \in C)} - e_{(x \in X_i \& c \in C)})^2}{e_{(x \in X_i \& c \in C)}} \tag{1}$$

where $n_{(x \in X_i \& c \in C)}$ is the number of instances in X_i for class c whose value is x. The expected frequency $e_{(x \in X_i \& c \in C)}$ is given as

$$e_{(x \in X_i \& c \in C)} = \frac{n_{x \in X_i} \times n_{c \in C}}{n} \tag{2}$$

where $n_{x \in X_i}$ is the number of instances in X_i with value x, and $n_{c \in C}$ denotes the number of instances having class c. Here, n is the total number of samples in the dataset. After calculation of the values of χ^2-statistic for each feature, sorting of the values is done in descending order for selection of the most important feature.

Symmetric Uncertainty (SU). The Symmetric Uncertainty (SU) is a filter feature selection method based on the concept of normalized mutual information. It measures the relevance among two variables f and C using the following formula [20]:

$$SU(f, C) = \frac{2 * MI(f, c)}{H(C) + H(f)} = \frac{2 * (H(C) + H(C/f))}{H(C) + H(f)} \tag{3}$$

where $MI(f, c)$ represents the mutual information between f and C, $H(f)$ represents the entropy of f, $H(c/f)$ computes the conditional entropy estimating the remaining uncertainty of C given the value of f. This feature selection method normalizes the value of $MI(f, c)$ to the range [0, 1].

Gain Ratio (GR). Gain ratio makes use of the split information of attributes into Information Gain (IG) statistic [21]. It was given to improve the bias of IG toward attributes which possess large diversity value. The inherent information of a given attribute can be computed by the entropy distribution of the instance value. The gain ratio of a given feature f with the feature value v can be computed using Eqs. 4 and 5.

$$GainRatio(v, f) = \frac{Information\_Gain(v, f)}{Intrinsic\_Value(f)} \tag{4}$$

where:

$$Intrinsic\_Value(f) = -\sum \frac{|S_i|}{|S|} * \log_2 \frac{|S_i|}{|S|} \tag{5}$$

where $|S|$ is the total number of possible values feature f can take, while $|S_i|$ is the number of actual values of the feature f.

3.2   Classification Methods

The classifiers utilized in this paper are divided into three parts, namely, the individual learners: linear SVM, ensemble learners: random forest, and rule-based learners: JRip. A total of three different types of learners are used for classification. A detailed description about the learners is given below.

Linear SVM. Linear Support Vector Machine (SVM) [22] are supervised classification algorithms that analyze the classification and regression problems of the data. Suppose the training data has n points $x_i \in R^d$ with corresponding labels as $y_i \in \{-1, +1\} \, for \, i = 1 \ldots n$. In case if the data is linearly separable, the primal SVM learning problem creates a hyperplane w^* by maximizing the geometric margin between the two classes, on the one hand, and separating the data, on the other hand. Thus, the "soft" 1-norm margin is maximized for non-separable data. The dual Lagrangian formulation of the soft 1-norm SVM reduces to the quadratic program as shown below in Eq. 6:

$$\max_{\alpha_i} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j \tag{6}$$

Subject to:

$$\sum_{i=1}^{n} y_i \alpha_i = 0; 0 \leq \alpha \leq C \ i = 1 \ldots n$$

Random Forest. Random Forest (RF) [23] is one of the ensemble-based classification algorithms built upon the bagging concept of L. Breiman. It performs random selection of features for construction of decision trees with controlled variance. The random forest algorithm is constructed based upon the combination of tree predictors such that every tree depends upon the values of a random vector independently. RF is used in various applications such as accident analysis, mechanical engineering, financial engineering, and biology. In random forest, each individual tree is constructed in a particular way:

(a)   Given a set of training data N here n random samples with repetitions, i.e., bootstrapping are used as training set.
(b) M input variables are determined for every node of the tree such that $m \ll M$ variables are selected for every node. The variable with the greatest importance chosen randomly is selected as the node. The value of m remains constant.
(c)   The development of each tree is done up to its maximum expansion.

JRip. JRip (RIPPER) [24] is one of the most popular and basic rule-based algorithms. As an optimized version of IREP, Repeated Incremental Pruning to Produce Error Reduction (RIPPER), a propositional rule learner was proposed by William W. Cohen. This RIPPER algorithm extracts the rules directly from data. It works in four stages, viz., growth, pruning, optimization, and selection. Initially, using some

of the heuristic methods, a rule set is formed upon the growing set. This rule set is then simplified repeatedly with the application of a set of pruning operators. The selection of pruning operators is done such that it provides the highest minimization of error over the pruning set.

4 Experimental Results and Discussions

For performing our experiments, we have used arrhythmia dataset from the UCI machine learning repository [25]. All the simulations were performed on Intel(R) Core(TM) i7-4770 CPU 3.40 GHz system. The experimentation was performed using Weka [26] data mining tool. The biomedical data used is arrhythmia dataset which is composed of 279 features, of which 206 are linear valued and the others are categorical. The total numbers of instances in the dataset are 452. The purpose is to distinguish between the existence and nonexistence of cardiac arrhythmia and to classify between one of the 16 groups. The class 1 is referred to "normal" or "healthy" ECG, classes 2–15 belong to distinct classes of arrhythmia, and class 16 belongs to the remaining of the unclassified samples. The total numbers of samples in healthy classes are 245. The rest 15 classes represent the different types of arrhythmia. Since the numbers of samples in these classes are highly imbalance, we have combined them into one class as "abnormal". Thus, the "abnormal" class contains 185 samples. The 22 "unclassified" samples are removed from the dataset. Furthermore, since the data contains missing values, therefore, numeric attributes are replaced their means, and nominal attributes are replaced by their modes. After performing all the data cleaning steps, three filter-based feature selection techniques, namely, χ^2-statistic, symmetric uncertainty, and gain ratio with threshold values as 10, 20, 30, 40, and 50 are utilized. For performance analysis of feature selection methods, three classification methods are also employed. These classification methods are divided into three categories, i.e., among individual learner: linear SVM, ensemble learner: random forest and rule-based learner: JRip. Figure 1 displays the flowchart of the proposed system.

In arrhythmia dataset, overall, there are 279 features and 452 instances among which some are redundant and many of them are irrelevant. An important preprocessing step before classification is feature selection. It is important to identify important risk factors which are the responsible for causing the disease. The performance metric used in overall experimental study was accuracy. Experimental study shows that the highest classification accuracy of 85.58% was obtained with 30 features (gain ratio) using random forest classifier as shown in Fig. 2 and Table 1 in bold. Linear SVM achieved 79.3% accuracy with 50 features by utilizing χ^2-statistic as shown in Fig. 3 and Table 2 in bold. Again with χ^2-statistic based feature selection, rule-based learner JRip achieved 81.86% accuracy by selection of 30 features as shown in Fig. 4 and Table 3 in bold. Moreover, this learner also generated a comprehensive set of eight rules for classification between normal and abnormal classes.

The classifier involved in the experiment through which the rules are extracted is JRip. The rule set is generated with a total of 10 features. The attributes utilized

Fig. 1 The proposed system model for arrhythmia classification
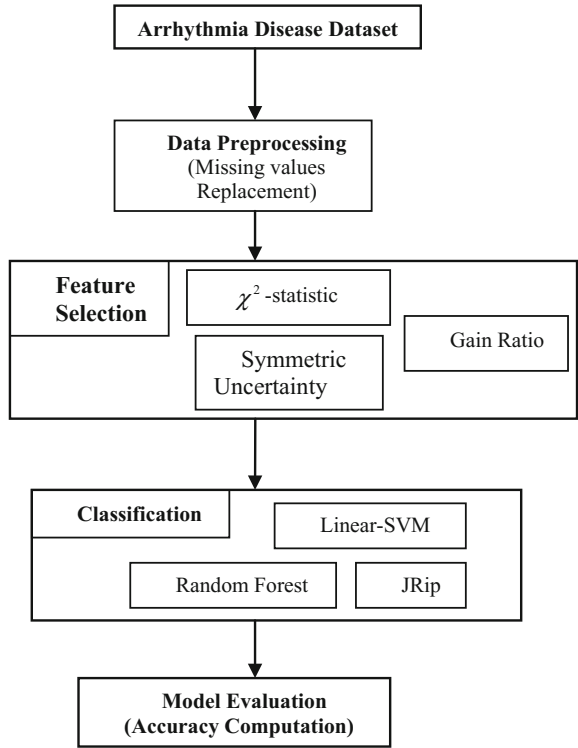
Fig. 2 Classification accuracies obtained with chi-square, symmetric uncertainty, and gain ratio feature selection methods using random forest classifier
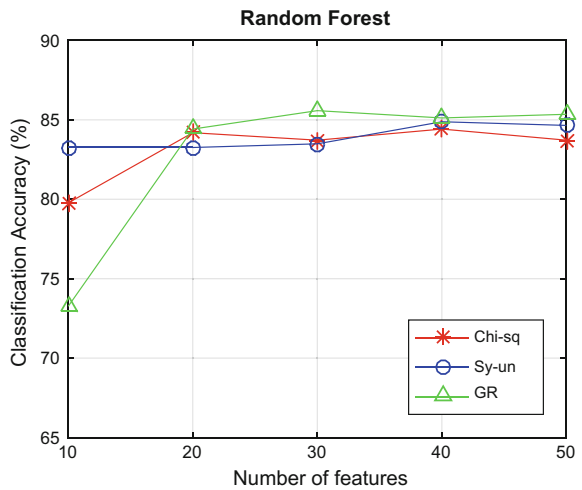
during rule-based classification are chDI_TwaveAmp, chDII_QRSTA, QRSduration, chV1_Swave, chDII_TwaveAmp, chAVR_QRSTA, heart rate, chV1_RPwaveAmp,

Table 1 Classification accuracy by random forest classifier

Method	10	20	30	40	50
Chi-square	79.77	84.19	83.72	84.42	83.72
Symmetrical uncertainty	83.26	83.26	83.49	84.88	84.65
Gain ratio	73.26	84.42	**85.58**	85.12	85.35

Fig. 3 Classification accuracies obtained with chi-square, symmetric uncertainty, and gain ratio feature selection methods using linear SVM classifier
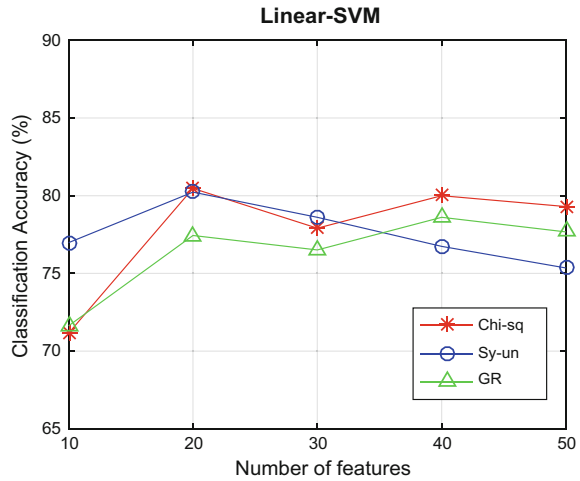
Table 2 Classification accuracy by linear SVM classifier

Method	10	20	30	40	50
Chi-square	71.16	80.47	77.91	80	**79.30**
Symmetrical uncertainty	76.98	80.23	78.61	76.74	75.35
Gain ratio	71.63	77.44	76.51	78.61	77.67

Table 3 Classification accuracy by JRip classifier

Method	10	20	30	40	50
Chi-square	78.61	80.93	**81.86**	78.84	78.14
Symmetrical uncertainty	77.67	78.37	79.30	79.54	74.42
Gain ratio	70.93	79.54	77.67	76.51	77.44

chV5_JJwaveAmp, and chV3_JJwaveAmp. These are some of the crucial risk factors which if diagnosed early can lead to early identification of arrhythmia and other cardiac-related diseases. The generated rules can help in effective and timely diagnosis of the disease.

Fig. 4 Classification
accuracies obtained with
chi-square, symmetric
uncertainty, and gain ratio
feature selection methods
using JRip rule-based
classifier
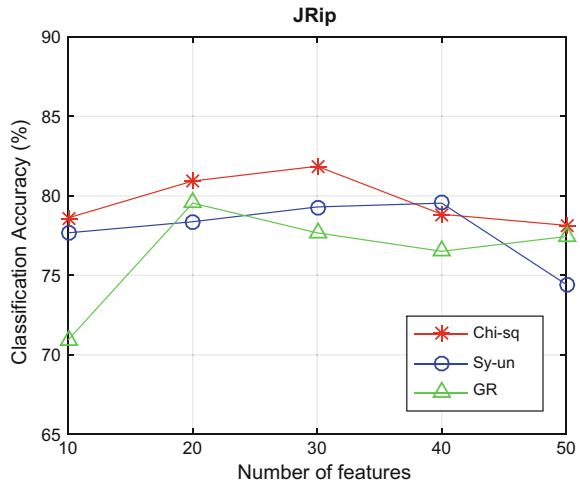
A total set of eight rules were obtained with JRip classifier using statistic with 30 features and are specified as under:

1. (chDI_TwaveAmp<=0.7) and (chDII_QRSTA<=9.8) → **class = abnormal**
2. (QRSduration>=88) and (chV1_Swave<=40) → **class = abnormal**
3. (chDII_TwaveAmp<=1.5)   and   (chAVR_QRSTA>=−9.7)   →   **class = abnormal**
4. (heart rate<=58) → **class = abnormal**
5. (chV1_RPwaveAmp>=0.7) → **class = abnormal**
6. (QRSduration>=92) and (chV5_JJwaveAmp<=0.6) → **class = abnormal**
7. (heart rate>=87) and (chV3_JJwaveAmp<=−0.2) → **class = abnormal**
8. else **class → normal**

5   Conclusion

The paper demonstrated the performance of three filter-based feature selection techniques in the diagnostic process of arrhythmia biomedical dataset. In the results, it was found that among ensemble classifiers, random forest outperformed with an accuracy of 85.58% in comparison to the other types of learners. Also, there is an improvement in the classification accuracy with reduced number of features, i.e., 30 features using gain ratio as the feature selection technique. This proves that all features are not required for disease diagnosis and prediction. We can also predict the disease very accurately with the small feature subset. Furthermore, as future work, we can apply other feature selection techniques and obtain much better accuracy with reduced number of features. Application of other filter, wrapper, and hybrid

feature selection methods can also be applied in classification of cardiac arrhythmia and other heart diseases.

References

1. Esteban, S., Tablado, M.R., Peper, F.E., Mahumud, Y.S., Ricci, R.I., Kopitowski, K.S., Terrasa, S.A.: Development and validation of various phenotyping algorithms for Diabetes Mellitus using data from electronic health records. Comput. Methods Programs Biomed. **152**, 53–70 (2017)
2. Wang, Y., Li, P.-F., Tian, Y., Ren, J.-J., Li, J.-S.: A shared decision-making system for diabetes medication choice utilizing electronic health record data. IEEE J. Biomed. Heal. Informatics. **21**, 1280–1287 (2017)
3. Heart Disease and Stroke | At A Glance Reports | Publications | Chronic Disease Prevention and Health Promotion | CDC, https://www.cdc.gov/chronicdisease/resources/publications/aag/heart-disease-stroke.htm
4. Andersson, C., Vasan, R.S.: Epidemiology of cardiovascular disease in young individuals. Nat. Rev. Cardiol. (2017)
5. Pławiak, P.: Novel methodology of cardiac health recognition based on ecg signals and evolutionary-neural system. Expert Syst. Appl. **92**, 334–349 (2018)
6. Mehta, S.S., Lingayat, N.S.: Detection of P and T-waves in electrocardiogram. In: Proceedings of the World Congress on Engineering and Computer Science. pp. 22–24 (2008)
7. Bsoul, A.A.R., Ji, S.-Y., Ward, K., Najarian, K.: Detection of P, QRS, and T components of ECG using Wavelet Transformation. In: ICME International Conference on Complex Medical Engineering, 2009. CME, pp. 1–6 (2009)
8. Chariatis, A.: Very fast online learning of highly non linear problems. J. Mach. Learn. Res. **8**, 2017–2045 (2007)
9. Canlas Jr., R.D.: Data Mining in Healthcare: Current Applications and Issues (2009)
10. Songthung, P., Sripanidkulchai, K.: Improving type 2 diabetes mellitus risk prediction using classification. In: 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE), pp. 1–6 (2016)
11. Mohan, K.R., Paramasivam, I., Narayan, S.S.: Prediction and diagnosis of cardio vascular disease–a critical survey. In: 2014 World Congress on Computing and Communication Technologies (WCCCT), pp. 246–251 (2014)
12. Soman, T., Bobbie, P.O.: Classification of arrhythmia using machine learning techniques. WSEAS Trans. Comput. **4**, 548–552 (2005)
13. Gupta, V., Srinivasan, S., Kudli, S.S.: Prediction and Classification of Cardiac Arrhythmia
14. Fazel, A., Algharbi, F., Haider, B.: Classification of Cardiac Arrhythmias Patients
15. Batra, A., Jawa, V.: Classification of arrhythmia using conjunction of machine learning algorithms and ECG diagnostic criteria. Int. J. Biol. Biomed. **1**, 1–7 (2016)
16. Yeniterzi, S., Yeniterzi, R., Kücükural, A., Sezerman, U.: Feature selection with genetic algorithms on cardiac arrhythmia database. In: The 2nd International Symposium on Health Informatics and Bioinformatics (HIBIT) (2007)
17. Coast, D.A., Stern, R.M., Cano, G.G., Briller, S.A.: An approach to cardiac arrhythmia analysis using hidden Markov models. IEEE Trans. Biomed. Eng. **37**, 826–836 (1990)
18. Luz, E.J.D.S., Schwartz, W.R., Cámara-Chávez, G., Menotti, D.: ECG-based heartbeat classification for arrhythmia detection: a survey. Comput. Methods Programs Biomed. **127**, 144–164 (2016)
19. Jadhav, S.M., Nalbalwar, S.L., Ghatol, A.A.: ECG arrhythmia classification using modular neural network model. In: 2010 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES), pp. 62–66 (2010)

20. Hall, M.A., Smith, L.A.: Feature selection for machine learning: comparing a correlation-based filter approach to the wrapper. In: FLAIRS conference, AAAI. pp. 235–239., Florida, USA (1999)
21. Mitchell, T.M.: Machine Learning. McGraw Hill (1997)
22. Vapnik, V.: The Nature of Statistical Learning Theory. Springer Science & Business Media (2013)
23. Breiman, L.: Random forests. Mach. Learn. **45**, 5–32 (2001)
24. Cohen, W.W.: Fast effective rule induction. In: Proceedings of the Twelfth International Conference on Machine Learning, pp. 115–123 (1995)
25. Bache K., Lichman M.: UCI Machine Learning Repository. http://archive.ics.uci.edu/ml/
26. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA data mining software. SIGKDD Explor. Newsl. **11**, 10 (2009)

Design of an Energy Harvesting System for Wireless Power Transmission Using Microstrip Antenna

Tejaswee Triyambak Das, Shubhendu Vinayak and Smitha N. Pai

Abstract Power supply drives technological devices. As the wired world looks long antiquated, it is imperative transition towards wireless power transmission (WPT). WPT has been lately enticing in a varying range of applications in numerous fields and has also become a very alluring area of research activity because of the potential it has in providing solutions to our daily requirement, so the need arises to delve into it to explore new, innovative and efficient means of energy transmission. WPT will be indispensable in the imminent future, as this technology assists in the transmission of electrical energy to any device from the power source using air as the transmission medium without the use of any conductors. In this paper, an energy harvesting device is proposed to convert ambient electromagnetic waves, predominantly from the 2.4 GHz ISM radio frequency (RF) channel into direct current (DC) electricity. A microstrip antenna is designed along with a booster circuit to help charge devices wirelessly. This novel technology could be used for charging phone batteries *on-the-go*, thereby eliminating the hassles of wired connections.

Keywords Wireless power transmission · Energy harvesting · Rectenna
Wireless charging · Microstrip patch antenna

T. T. Das · S. Vinayak · S. N. Pai (✉)
Information and Communication Technology Department, M.I.T. Manipal
Academy of Higher Education, Manipal, India
e-mail: smitha.pai@manipal.edu

T. T. Das
e-mail: tejaswee.das@gmail.com

S. Vinayak
e-mail: svinayak57@gmail.com

1 Introduction

The twentieth century has seen a rapid development in wireless technologies. The wired world now looks obsolete, and so the transition towards a wireless world is imminent. Nikola Tesla was the first to propose and demonstrate a wireless transmission system [1] based on Heinrich Hertz's electromagnetic wave propagation theory [2, 3]. In [2], the author proposes the idea of wireless transmission of electrical energy as a solution for the world's energy crisis. Although wireless communication has been a great success, WPT has much to achieve and is a fast evolving, promising field for extensive research. WPT is the process of transmitting electrical energy to a load from a power source using air as the media without any interconnecting wires. A wireless transmitter conveys the field energy across an intervening space to a receiver, where it is converted back to electrical energy and used.

With the advent of new wireless devices, wireless power density is growing since there is an increase in the number of various electromagnetic (EM) power sources like mobile stations, wireless fidelity (Wi-Fi) routers and other similar radio transmitters. RF energy is broadcasted from billions of radio transmitters all around, including mobile phones, radios, base stations and TV broadcast stations. An evolving concept of energy harvesting aims to harvest this ambient surplus RF energy, convert it into useable form and use it to power low-energy devices. With the development of Internet of things (IoT), the importance of WPT and energy harvesting has increased manifold.

Energy harvesting is achieved using a special type of rectifier antenna, namely, rectenna, which converts electromagnetic energy into DC electricity. These antennas are used in WPT systems to transmit power by using radio waves.

In this paper, an energy harvesting device is proposed which operates at 2.45 GHz, with the aim to charge low-power devices wirelessly.

Section 2 elaborates on the related work carried out in the field out so far. The methodology shows the design and fabrication of the device in Sect. 3. Section 4 discusses the simulation and results. Finally, the proposed design is concluded in Sect. 5.

2 Related Work

WPT involves the conversion from DC-to-AC power, followed by transmission of electromagnetic (EM) radiation of RF, laser, light or microwave. At the receiving end, the AC energy is rectified, and DC energy obtained is used as the power supply.

Technology used in WPT is used in many applications like microwave power transmission in space, solar-powered satellite systems (SPSS) and radio frequency identification (RFID) [4–9]. Energy harvesting in RF was applied to miniature devices like sensors and biomedical implants to increase their battery life and avoid replacements [9, 10]. Brown [11] designed rectenna which is a combination of an antenna and

a rectifying circuit to convert microwave power to DC power. The antenna combined with the rectifying circuit forms the rectenna. Reference [12] proposed a formula to calculate the resonant frequency values within the given ratio of inset width. This is used to obtain the best possible match of a microstrip line with a feeding of 50 Ω. The maximum deviation that is obtained from the proposed formula is 0.2%. An investigation of microstrip patch antenna characteristics using different substrates concludes RT Duroid 5880 as the best substrate to be used, giving a gain of 8 dBm [13]. A microstrip patch antenna on RT Duroid 5880, designed and simulated using computer simulation technology (CST) achieved a bandwidth of 100 MHz. Simulated result parameters were as follows: return loss of -15.76 dB, voltage standing wave ratio (VSWR) of 1.38, directivity of 8.591 dBi and a gain of 7.752 dBi [14]. The antenna performance is improved using parameters like gain, radiation pattern and return loss. This design provides a gain of 8.99 dB with a return loss of -43 dB, used for WiMAX communication [15]. The energy received using the 2.5 GHz resonant frequency in a wireless energy transmission system using microstrip antenna was higher than the resonant frequency of 1.94 GHz. The highest power received was 1.062 mW for a resonant frequency of 2.5 GHz and with a received power of 0.95 mW for a resonant frequency of 1.94 GHz [16]. Of the various slots shapes in microstrip patch antenna, E-shaped antenna shows the highest bandwidth followed by U-shaped and H-shaped. E-shaped antenna shows remarkable bandwidth improvement of about 57.36% and VSWR is less than 2 [17].

Reference [18] discusses in depth about electromagnetic radiation energy harvesting using rectenna-based approach. Rectenna stands out to be one of the most promising methods for harvesting wireless energy. For higher power density incident on receiving antenna, a higher rectenna conversion efficiency is measured [19]. The planar rectennas were designed using three different substrates: FR4, RT 6010 and RT 5870 with dielectric constants (ε_r) 4.5, 10.2 and 2.35, respectively. The output voltage at 1 m transfer distance (V_{DC}) was measured as 0.2 V, 0.3 V and 0.35 V, respectively, for three different substrates [20]. At a single ISM band with a frequency of 2.4 GHz, the performance of rectenna which collects low incident power density level is designed along with its performance. The rectifier gives a conversion efficiency of 34%, at low input power level of -10 dBm [21]. An adaptive rectenna with a monopole designed in the 900 MHz ISM band achieved a 40% conversion efficiency over RF input varying from -17 to 27 dBm with a maximum peak conversion efficiency of 82% at 12 dBm [22].

A dual-band rectenna designed for an operation in GSM 900 and 1800 MHz, when tested under ambient conditions and well-matched impedance gives a maximum of 415 mV [23]. The most acclaimed work in this field has been by [24]. The authors could prototype the first camera and temperature sensor which were battery-free and powered from Wi-Fi without compromising the communication performance of the network.

In this paper, it is proposed to design an energy harvesting device operating in the ISM band at 2.4 GHz frequency, using rectenna and booster circuits to generate a voltage of 5 V which can be used to charge low-power devices like mobile phones wirelessly. The work in this paper is focused on devising a mechanism to
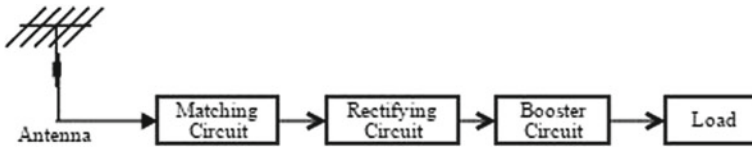
Fig. 1 Block diagram representation of proposed device

use the normal 2.4 GHz Wi-Fi band to power devices with appreciable voltage outputs and efficiency which was achieved using state-of-the-art components and better techniques as compared to others.

3 Methodology

The proposed device in this paper consists mainly of four parts as shown in Fig. 1, namely, antenna, matching circuit, rectifying circuit and booster.

A block diagram representation of the design proposed is shown in Fig. 1. The functionality of various components of the block diagram is explained as follows.

3.1 Antenna

Antenna is an essential component of air transmission. Radio waves are obtained from the electric power and vice versa [25] using this device. Although many types of radio antennas are available, the one used in this paper is a microstrip patch antenna. Microstrip patch antenna is low cost, low profile, low volume and lightweight, conformable to both planar and non-planar surfaces. These are printed antennas and so there is simplicity in their manufacture and fabrication. Most importantly, they are robust mechanically and versatile in terms of resonant frequencies and supports both linear and circular polarization. There is a distinct and immediate need for a compact, low-cost, reliable microstrip antenna which can operate over multiple bands [26] as used in mobile phones.

Microstrip antenna consists basically of three main parts as shown in Fig. 2: conducting patch, dielectric substrate and ground plane.

Conducting patch is responsible for generation of surface current. The dielectric substrate radiates the incident RF waves and a ground plane prevents the antenna from radiating both ways, hence helps in better directivity.

Microstrip antenna is normally formed by etching a radiating patch which is made up of a conducting material on a dielectric substrate and with a ground plate at the bottom of the substrate which acts as a reflector. The microstrip antenna designed in this paper is made of copper, and RT Duroid 5880, with a low dielectric constant,

Fig. 2 Microstrip patch
antenna
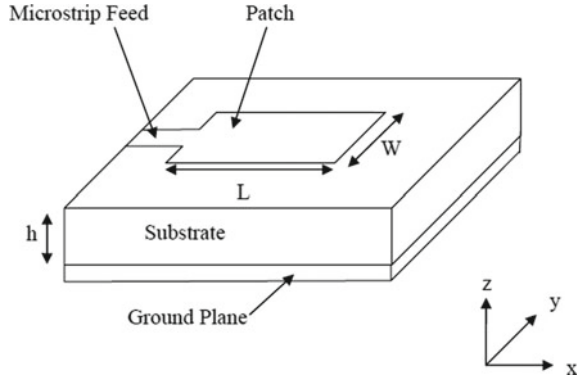
$\varepsilon_r = 2.2$, loss tangent, $\tan \delta = 0.0004$ and height $h = 1.57$ mm, is used as substrate. RT Duroid 5880 gives maximum radiation due to its least dielectric constant.

The various antenna parameters of the microstrip antenna used in this paper were calculated using the following equations:

$$W = \frac{c}{2 f_o \sqrt{\frac{(\varepsilon_r + 1)}{2}}} \tag{1}$$

$$L_e = \frac{c}{2 f_o \sqrt{\varepsilon_e}} \tag{2}$$

$$\varepsilon_e = \frac{\varepsilon_r + 1}{2} + \frac{\varepsilon_r - 1}{2} \left[1 + 12 \frac{h}{W} \right]^{\frac{-1}{2}} \tag{3}$$

$$\Delta L = 0.412 h \frac{(\varepsilon_e + 0.3)\left(\frac{W}{h} + 0.264\right)}{(\varepsilon_e - 0.258)\left(\frac{W}{h} + 0.8\right)} \tag{4}$$

$$L = L_e - 2\Delta L \tag{5}$$

where

W   patch width;
c   speed of electromagnetic wave in free space;
f_o frequency of incident wave;
ε_r dielectric constant of substrate;
L_e effective patch length;
ε_e effective dielectric constant of substrate;
ΔL length extension;
L_e effective patch length;
L   patch length;
h   height of substrate.

Figure 3a shows the antenna dimensions of designing and Fig. 3b shows the CST Student Suite. Table 1 shows the dimensions of the antenna.

(a)

(b)
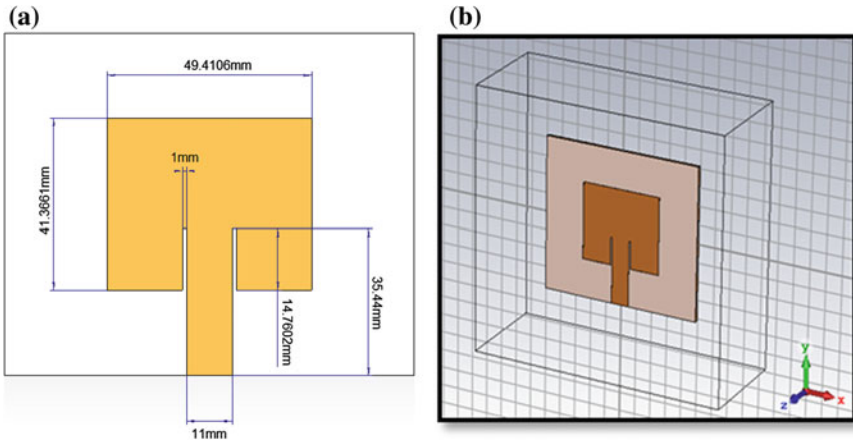
Fig. 3 Microstrip patch **a** antenna dimensions and **b** design in CST studio

Table 1 Antenna dimensions

Parameters	Measurements (mm)
F_i	14.7602
G_{pf}	1
L	41.3661
W	49.4106
W_f	11
H	1.57
h_t (copper cladding)	0.035

Friis equation for calculation of power at receiving antenna.

$$P_r = P_t \cdot \frac{G_t G_r \lambda^2}{(4\pi R)^2} \tag{6}$$

where

P_r power at receiver;
P_t power at transmitter;
G_t gain of transmitting antenna;
λ wavelength;
G_r gain of receiving antenna;
R   distance between transmitter and receiver.

3.2 Matching Circuit

The matching circuit is an integral part of any RF circuit. It helps in matching the impedance. In accordance with maximum power theorem, to obtain maximum power, the input impedance of load and the output impedance of its corresponding signal source need to be equal. Impedance matching also helps in avoiding signal reflection and hence prevents its weakening. The inset feed used in this microstrip antenna acts like an impedance matching circuit which is an added advantage; hence, no other special arrangements are needed for matching impedance.

3.3 Rectifying Circuit

The next block is the rectifier, which is an electrical device to convert the alternating current (AC) to DC allowing current flow through only one direction. The rectifier consists of a Schottky diode, two tapered microstrip lines, a matching microstrip line, a $\lambda/4$ microstrip line, and an output low pass filter. The Schottky HSMS-2820 diode has a low built-in voltage with fast-switching resistance, good RF characteristics and low series resistance [27].

However, in this paper, a separate rectifying circuit has not been used. The microstrip antenna and booster circuit have been designed, constructed and tested separately, and finally, results have been combined assuming substantial rectifying loss.

3.4 Booster Circuit

The output from the rectifier is then fed to a DC–DC booster converter. A booster is a DC-to-DC voltage converter that steps up the voltage. It is a class of switched mode power supply (SMPS); it steps up the input power to a level which can be used to provide sufficient power source to the load, which can be any electrical device of appropriation to the apparatus.

An ultralow voltage step-up converter LTC 3108 along with a complete power management system is used in the booster circuitry. A DC/DC converter, LTC®3108, is ideal for harvesting and managing surplus energy obtained from low input voltage source. The input can step up from as low as 20 mV. V_{OUT} of 2.35 V, 3.3 V, 4.1 V or 5 V can be obtained. 2.2 V at 3 mA is the low dropout linear regulator (LDO) for this booster. Power management solution is obtained by using a small step-up transformer for the LTC3108 for data acquisition and wireless sensing [28].

The transformer used is a 1:100 WE-EHPI Energy Harvesting Coupled Inductor 74,488,540,250. For 1:20 transformer, the input voltage must be 60 mV, and for a 1:100 WE-EHPI, the input voltage can be as low as 20 mV.

4 CST Design and Simulation

CST Studio Student Edition was used to design the microstrip antenna before the fabrication process. Figure 3b shows the design obtained from studio and Fig. 6 shows the fabricated view.

Figure 4a shows the S-parameter plot variation of return loss (in dB) over a range of frequencies. The antenna is having the best impedance match at resonance with minimum return loss. The antenna resonates at 2.45 GHz.

The efficiency of RF power transmission from the power source is determined by the VSWR. Figure 4b shows the VSWR plot attaining a minimum at a resonant frequency of 2.45 GHz. Figure 5a shows the power radiated plot with a peak at 2.45 GHz. Figure 5b shows the surface current flow of the antenna. A good directivity is seen in the antenna at both the lower and upper resonant frequency.
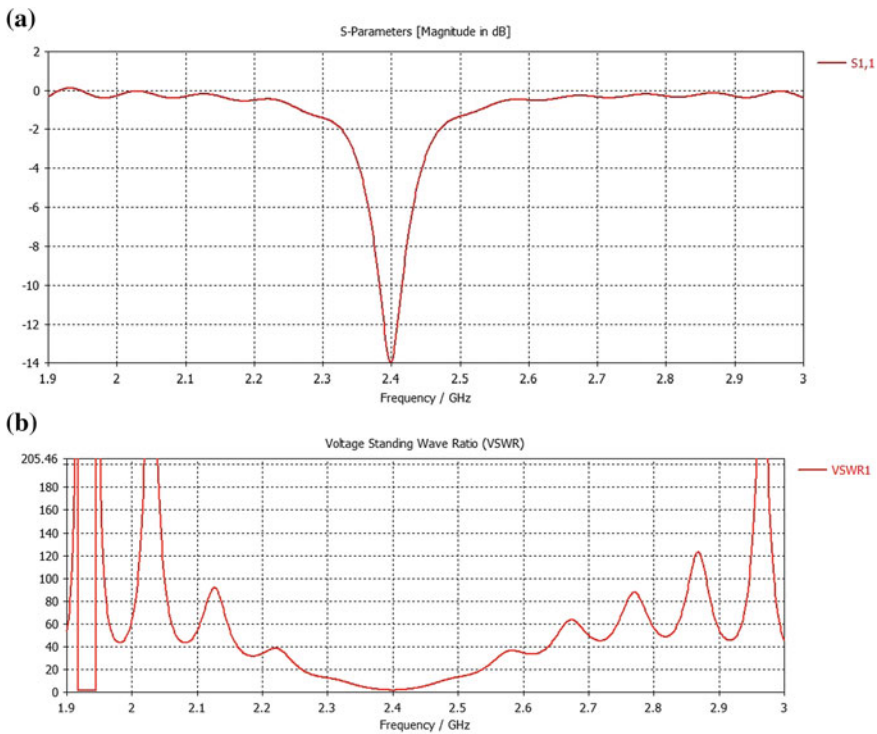
Fig. 4 **a** S-parameter and **b** voltage standing wave ratio (VSWR)
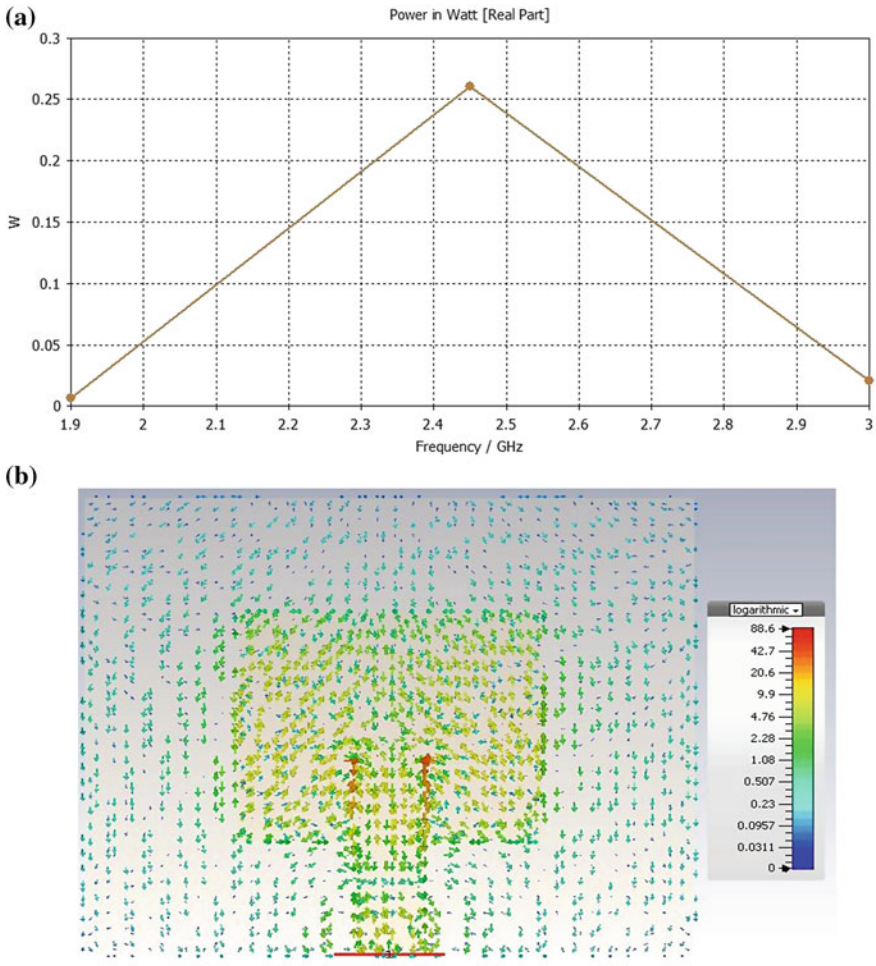
Fig. 5 a Power radiated and **b** surface current

5   Fabrication

The substrate laminate was cut precisely according to the above measurements to obtain the patch. Figure 6 shows the fabricated antenna.

Figure 7 shows the booster circuit design. Figure 8 shows the LTC 3108, soldered into an SSOP16 adapter board with WE-EHPI coupled Inductor 1:100 and Ceramic capacitor 1 µF, 330 pF. This prototype has been designed for 3.3 V, i.e. VS1 is connected to Vaux and VS2 connected to GND. The pins can be configured to get outputs of 2.35 V, 4.1 V or 5 V as per requirements [29].
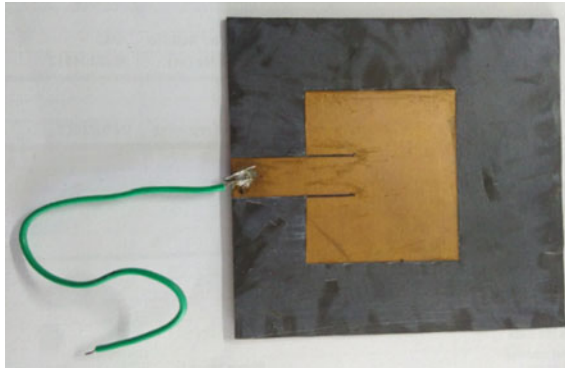
Fig. 6 Front view of the patch antenna made from RT Duroid 5880
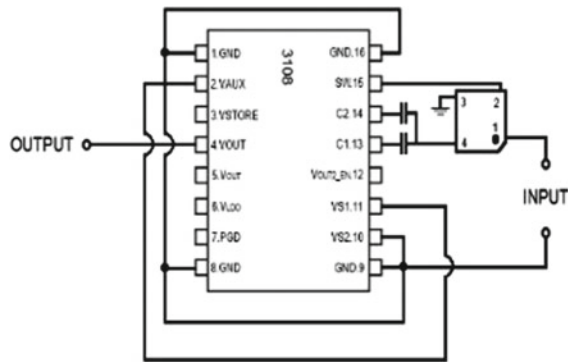
Fig. 7 Booster circuit design
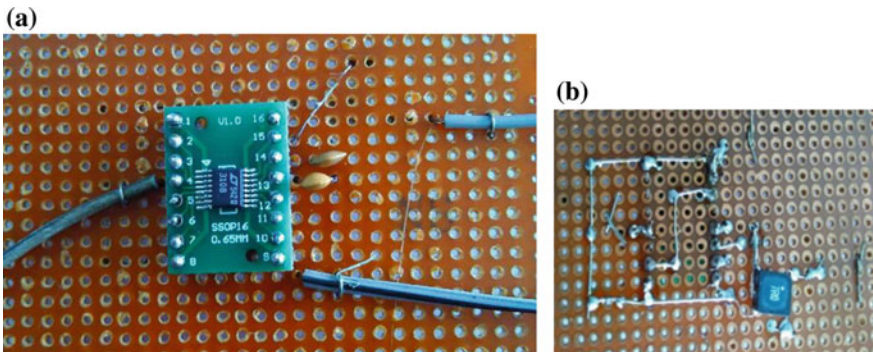
Fig. 8 **a** Front and **b** rear view of designed circuit
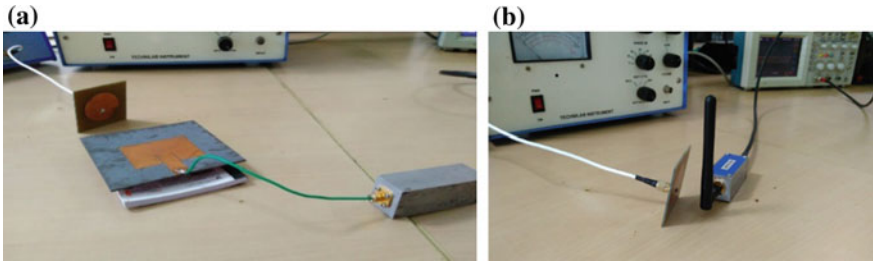
Fig. 9 **a** Setup to test the designed patch antenna and **b** setup to test GSM-022 stub antenna
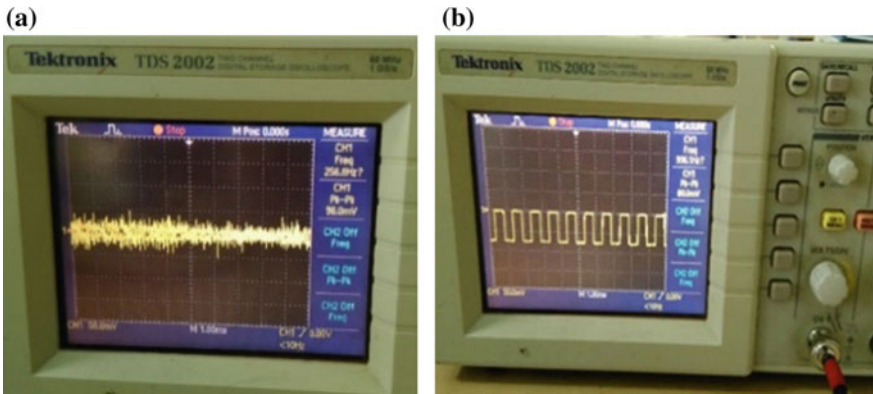
Fig. 10 Output of **a** designed patch antenna and **b** stub antenna

6 Results

The designed microstrip antenna is tested using Microstrip Antenna Trainer Kit. The kit consists of voltage controlled oscilloscope (VCO) and VSWR Meter. The VCO is adjusted to a frequency of about 2.448–2.453 GHz to test the response of the designed antenna around the designed resonant frequency of 2.45 GHz. Microstrip antenna is placed orthogonally to the transmitter and connected to the detector as shown in Fig. 9. The detector is connected to an oscilloscope to record waves and voltage.

A GSM-022 Stub antenna is connected to the detector. The detector is connected to oscilloscope. Figure 10 shows the output obtained from the designed patch antenna and stub antenna.

A peak–peak voltage of 96.0–114.2 mV was obtained for the designed patch antenna. The voltage plot showed uneven spikes because of excessive noise due to unprofessional physical hand fabrication of the antenna. The GSM-022 Stub antenna showed a peak–peak voltage of 80.0 mV for the same frequency. The voltage plot was however precise in this case and proper square wave was observed due to proper noise cancellation. A Voltage Source is used to provide input to the booster circuit.

Fig. 11 Output of booster circuit

It has an input rating of 240 V at 50 Hz and can provide an output voltage in the range of 0–270 V. The input voltage is varied between 0 and 1 V and readings noted. Figure 11 shows the output obtained from the booster circuit. The input was kept stable at 30 mV. A peak output voltage of 1.8078 V was recorded at the output.

7   Conclusion

In this paper, an energy harvesting device has been proposed which could be used to charge low-power devices. The microstrip patch antenna developed is small in size and easy to fabricate. It is operational in the Wi-Fi band of 2.45 GHz. This proposed device can be used to power low-powered devices using Wi-Fi. The antenna was tested alongside a GSM-022 Stub antenna. The designed antenna gave an output of 96.0 mV whereas GSM-022 antenna gave an output of 80.0 mV when tested at a frequency of 2.45 GHz.

The booster circuit produced promising results for a design of 3.3 V. At an input as low as 30.18 mV, it was able to give an output of 1.8078 V with an efficiency of 59.9%. Assuming a considerable loss for a design of 5 V, antenna and rectification loss, we can expect an efficiency of almost 70%.

The novel technique in the design of this device is its simple design, ease of fabrication and the use of state-of-the-art components. The microstrip antenna is compact, reliable and can be printed and fabricated easily. The antenna in this paper is designed using high-frequency laminates of RT Duroid 5880 which gives maximum radiation as compared to other substrates available. Although the antenna and boosting circuit modules have been tested separately in this paper, the inset feed of the microstrip antenna matches the impedance doing away with the need of a matching circuit. The inset can be calculated accurately and precisely cut for better impedance matching. The booster circuit is designed using LTC 3108. The advantages of using this IC is that it needs no other external power supply. It also eliminates the need for a separate

rectification circuit to convert to DC. The most notable feature of LTC 3108 is its manual pin configuration to obtain different outputs. Power management solution is achieved by using a small step-up transformer. The transformer used—1:100 WE-EHPI Energy Harvesting Coupled Inductor—is very small in size and can operate at a very low voltage. The input to the circuit can be boosted up from voltage levels as low as 20 mV.

The future work in this project relates to better rectification circuit designs and hence better conversion efficiency. Final product fabrication aims to assemble this whole prototype and make it compact so that it can be used easily.

References

1. Tesla, N.: My Inventions: The Autobiography of Nikola Tesla. Hart Brothers. http://www.teslasautobiography.com (1982)
2. Tesla, N.: Experiments with Alternate Currents of High Potential and High Frequency. McGraw Publishing Company. http://www.tfcbooks.com/tesla/1892-02-03.htm (1904)
3. Tesla, N.: The Transmission of Electrical Energy without Wires, The Electrical World and Engineer. http://www.tfcbooks.com/tesla/1904-03-05.htm (1904)
4. Yang, Y., Zhang, X., Ye, Samii Y.R.: Wideband e-patch shaped antenna for wireless telecommunications. IEEE Trans. Antennas Propag. **AP-49**, 1100 (2001)
5. Ge, Y., Esselle, K., Bird, T.: E-patch shaped antennas for fast wireless networks, IEEE Trans. Antennas Propag. (2004)
6. Balanis, C.A.: Antenna Theory: Analysis and Design. Wiley (2005)
7. Tong, K.-F., Wong, T.-P.: Circularly polarized U-slot antenna. IEEE Trans. Antennas Propag. **55**(8), 2383–2385 (2007)
8. Tong, K.F., Lee: Microstrip patch antenna—characteristics and recent advances. IEEE Proc. **100**(7), 2169 (2012)
9. Shackelford, A.K., Lee, K.L., Luk, K.M.: Design of small size wide bandwidth micro strip patch antenna‖, Department of Electrical Engineering, IEEE Antennas Propag. **45**(1) (2003)
10. Abu Tarboush, H.F.A., Al-Raweshidy, H.S., Nilavlan, R.: Triple-band U-slots patch antenna for WiMAX applications. In: Proceedings of 14th Asia–Pacific Conference on Communications, pp. 1–3 (2008)
11. Brown, W.C.: The history of the development of the rectenna. In: Proceedings of Space Power Symposium, 271–280 (1980)
12. Matin, M.A., Sayeed, A.I.: Design rule for inset-fed rectangular patch antenna. WSEAS Trans. Commun. **9**(1) (2010)
13. Mukherjee, T., Rao, P.V., Bhavani, M.V.L., Sinha, A.K.: Micro strip patch antenna characteristics using different dielectric substrates. Int. J. Adv. Res. Electr. Electron. Instrum. Eng. **4**(5) (2015)
14. Patel, A.V., Kumar, M., Pawar, R.P., Dev, V.M., Mahalakshmi, M.N.: Design and simulation of microstrip patch antenna on RT DUROID 5880 substrate for object detection using RADAR. Int. J. Sci. Eng. Res. **6**(4) (2015)
15. Lal, K.N., Singh, A.K.: Modified design of microstrip patch antenna for WiMAX communication system. In: Students' Technology Symposium (TechSym), IEEE, 386–389 (2014)
16. Meng, L.K., Swee, T.T, Yun, L.K., Kassim, J.: Wireless energy transfer using microstrip antenna. Am. J. Appl. Sci. **11**, 195–201 (online). http://thescipub.com/PDF/ajassp.2014.195.201.pdf. Accessed 05 Jan 2018

17. Mehta, S., Jain, M.: An overview: various slots shapes of micro-strip patch antenna. Int Conf. on Multidiscip. Res. Pract. **1**(7), 175–178 (2014)
18. Abadal, G., Alda, J., Agustí, J.: Electromagnetic radiation energy harvesting the rectenna based approach. In: ICT-Energy—Nanoscale Energy Management Concepts Towards Zero-Power Information and Communication Technology. Intech Ed, Feb 2014
19. Zhang, J., Huang, Y.: Rectennas for Wireless Energy Harvesting (online). https://pdfs.semanti cscholar.org/5481/1c983816396b805f497570cd9112a5e68c51.pdf. Accessed 05 Jan 2018
20. Heikkinen, J., Salonen, P., Kivikoski, M.: Planar rectennas for 2.45 GHz wireless transfer. IEEE Radio Wirel. Conf. 63–66 (2000)
21. Alicalapa, F., Douyere, A., Riviere, S., Lan Sun Luk, J.D.: A compact rectenna device at low power. Prog. Electromagn. Res. C **16**, 137–146 (2010)
22. Yagoub, M.C.E., Almohaimeed, A.M., Amaya, R.E.: Efficient rectenna with wide dynamic input power range for 900 MHz wireless power transfer applications. IEEE Elec. Power Energy Conf. 12–14 (2016)
23. Ho, D.K., Kharrat, I., Ngo, V.D., Vuong, T.P., Nguyen, Q.C., Le, M.T.: Dual-band rectenna for ambient RF energy harvesting at GSM 900 MHz and 1800 MHz. In: Proceedings of IEEE International Conference on Sustainable Energy Technology (ICSET), pp. 306–310 (2016)
24. Talla, V., Kellogg, B., Ransford, B., Naderiparizi, S., Gollakota, S., Smith, J.R.: Powering the next billion devices with Wi-Fi (online). http://arxiv.org/abs/1505.06815 (2015)
25. Graf, R.F.: Modern Dictionary of Electronics, p. 29. Elsevier Science Publishing. ISBN 9780080511986
26. Paul, B., Mridula, S., Aanandan, C.K., Mohanan, P.: A new microstrip patch antenna for mobile communications and bluetooth applications. Microw. Opt. Technol. Lett. **33**, 285–286 (2002)
27. Zulkifli F.Y., Leza, Y.M., Basari, Rahardjo, E.T.: Design of rectifier for rectenna application. In: Proceeding of Asia Pacific Microwave Conference, 6–9 Dec 2015
28. Lee, K.F, Tong, K.F.: Microstrip patch antennas-basic characteristics and some recent advances. IEEE Proc. **100**(7), 2169 (2012)
29. Linear Technology, LTC3108-Ultralow Voltage Step-Up Converter and Power Manager, LTC 3108 datasheet (online). http://www.linear.com/product/LTC3108 (2010)

Design and Implementation of a Wearable Real-Time ECG Monitoring System Based on Smartphone

R. P. Tripathi, Ankita Tiwari, G. R. Mishra and Dinesh Bhatia

Abstract This research paper presents a smart approach for design and implementation of a wearable system that is able to monitor the electrical activity of the heart of a person using a smartphone. The real-time ECG monitoring system is implemented using a FPGA IC, a smartphone, and a front-end amplifier circuit for amplifying the ECG signal, and other component includes a capacitor, comparator, and a register. A USB link of PHY/LINK 1.1 and delta modulator is implemented in the digital circuit block of the FPGA IC. The digital block for the delta modulator consist of a sinc filter, a sampler basically a D flip-flop, and a FIR filter. Clock frequency at which sampler works is 6 MHz, delta modulator takes the input voltage in the 0.25–1.25 V @ up to 1.75 kHz frequencies, and it generates a corresponding 8-bit code @ 23.5 kS/s. The FPGA IC and all other circuits are powered by a smartphone through a micro USB link and provide a power supply of DC 5 V. In our system, we have used the screen of the smartphone as a display device and the captured ECG is displayed on the smartphone. For the calculation of the heart rate of the captured ECG signal, we used the smartphone and the result with waveform is stored in the memory of smartphone.

Keywords FPGA · Delta modulator · USB 1.1 PHY/LINK · CVD

R. P. Tripathi (✉) · A. Tiwari · G. R. Mishra
Department of Electrical & Electronics Engineering, ASET, Amity University,
Lucknow 226001, India
e-mail: raghvendra5936@gmail.com

A. Tiwari
e-mail: ankitaiit15@gmail.com

G. R. Mishra
e-mail: grmishra@gmail.com

D. Bhatia
Department of Biomedical Engineering, North Eastern Hill University, Shillong 793022,
Meghalaya, India
e-mail: bhatiadinesh@rediffmail.com
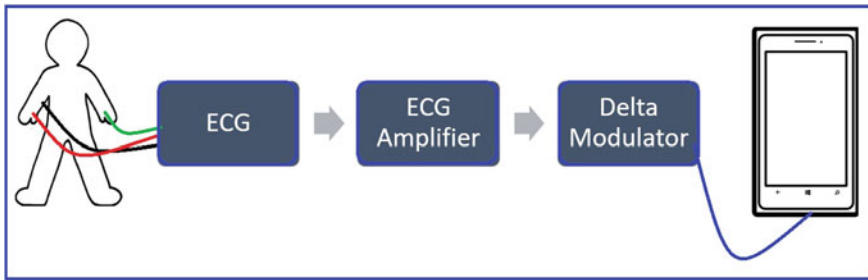
Fig. 1 Architecture of the proposed system

1   Introduction

Nowadays, cardiovascular diseases (CVD) [1] are one of the major critical diseases among all and having a large number of patients in each and every country of the world. As the statistics provided by the various health magazines, CVDs having a significant portion in the overall deaths is encountered due to health-related problems. To avoid these deaths, a simple method that can be very effective is the early diagnosis of the CVDs. In order to perform the early diagnosis of the cardiovascular disease, we need to monitor the functioning of the heart round the clock. The functioning of the heart can be measured by the Electrocardiography (ECG) [2].

For this reason, we require a portable device that is able to monitor the ECG of the person. For better and pleasant monitoring, it is desired that the device should not interfere with the daily activities of the person. So, keeping these things in our mind, we are working towards the development of a system that is wearable and also able to fulfill the commercial needs and accuracy level for the monitoring of cardiac activity.

Therefore, in this paper, we have presented the first phase of work, and in this work, we have collected the raw ECG data using three electrodes and displayed this collected signal on the screen of a smartphone. Further, we have compared the performance of our system by the presenting same signal on the Matlab software using the data stored in the memory of the smartphone. The whole idea depends upon an observation, as we observed from our analysis that the use of the smartphone among the people is very popular, so if we become able to display the ECG signal in real time like many other things for example, Live cricket score, live weather reports, live train status, etc. Then, it will become very easy to monitor the cardiovascular system [3]. And, this monitoring will definitely help us in the early detection of the CVDs. That is why in the first phase, we have designed a monitoring system for ECG using smartphone, in order to interface the electrodes with smartphone, we have used a FPGA IC, the architecture of the system is presented in the Fig. 1. Section 2 of the paper presents a detailed explanation of the architecture, Sect. 3 explains the experimental setup, and in the Sect. 4, we have presented the result, and in Sect. 5, we have discussed the future scope and conclusion of the system.
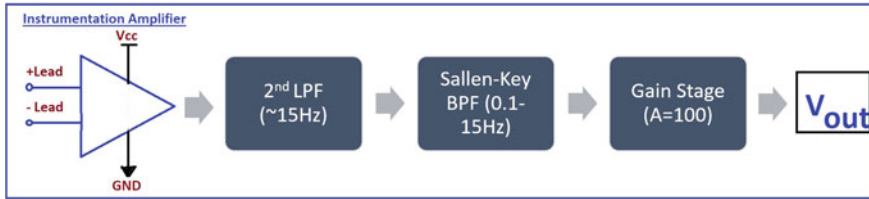
Fig. 2 Architecture of the ECG amplifier

2 Architecture of the System

We have implemented the ECG monitoring system without giving any extra power supply to the circuit, and in this proposed system, we have given the power using the USB interface that draws the power from a smartphone delivers to the FPGA board [4] and ECG amplifier [5].The detailed architecture of the system is shown in Fig. 1. Three electrode leads are attached to the human body to acquire the ECG signal waveform; the +ve (red) electrode lead is attached from the right wrist and the –ve (yellow) electrode lead is attached to the left wrist, of the person. The ground (black) electrode lead is attached from the right elbow of the human body. The architecture of the proposed system includes an ECG amplifier for amplifying the low amplitude signals and low-pass filter is used for removing the high-frequency noise from the collected ECG waves [6]. Sallen–key BPF [7] is used to remove very low-frequency noise (~0.1 Hz).

2.1 ECG Amplifier

Since the amplitude of acquired ECG signals through the attached ECG electrodes is around the 0.5 mV that is a very low value. Therefore, in order to acquire and analyze the ECG signals, we need to amplify these signals the. Figures 2 and 3 present the block diagram of the system and the schematic diagram of the ECG amplifier. To amplify the ECG signal, we have used an instrumentation amplifier, a second-order LPF is used for removing the high-frequency noise, to remove the extremely low-frequency noise [8] a Sallen–key BPF is used further to provide the gain we have used a gain amplifier at the last stage. At the last stage of the system, the gain amplifier is implemented to amplify the coming ECG signal and it also removes the noise.
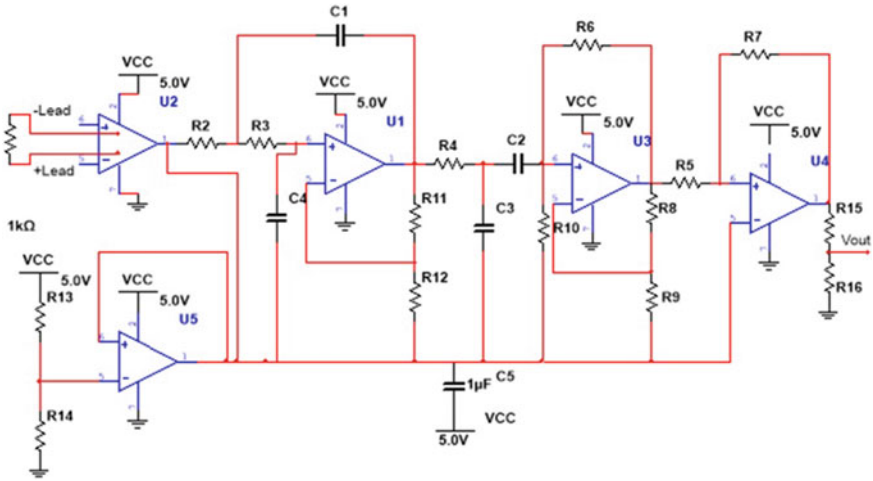
Fig. 3 Detailed schematic diagram of the used ECG amplifier

2.2 Delta Modulator

In the transfer on information where the quality of information is not an important parameter, we use the technique known as delta modulation [9]. The device that performs the delta modulation is termed as the delta modulator and it converts the analog information into digital and then performs the transmission. In our system, we have implemented a delta modulator FPGA IC [10], a comparator, and a capacitor [11] with resistor. Figure 4 shows a block diagram representation of the implemented delta modulator circuit. The output generated through the delta modulator has the form of DPCM (differential pulse code modulation). The DPCM [9] is coded into n-bit digital data stream samples. The comparator circuit compares the present value of the input signal with the latest sampled value and accordingly generates one output data bit. The clock frequency at which the sampler works is 6 MHz and an analog input voltage in the range of 0.25–1.25 V @ 1.75 kHz can be supplied to the delta modulator. The delta modulator circuit generates the output at the data rate of 23.4 kS/s and the output is of 8-bit.

3 Experimental Setup

In the implementation of the proposed system, we have placed the electrodes on the skin of a person as shown in Fig. 5. The red lead of the electrode (+ve) is attached to the right hand's wrist, whereas the blue (−ve) lead electrode is attached through left hand's wrist. And, the black lead that is ground electrode is attached to the right hand
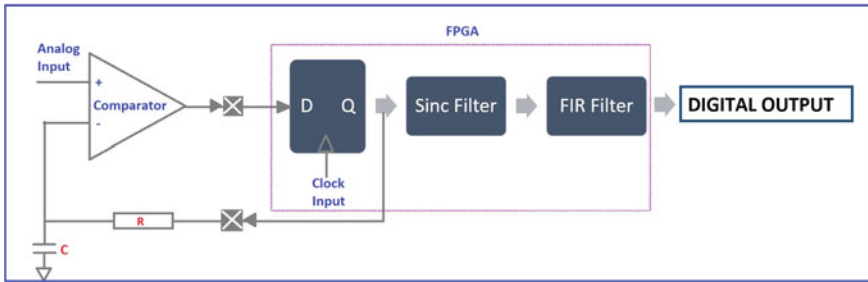
Fig. 4 Block diagram of delta modulator

Fig. 5 Electrode position in ECG acquisition

above the red electrode and near to the right elbow of the person. The acquisition of the data is performed in the no body movement situation.

After collecting the ECG signals, we have given the ECG signals to the ECG amplifier who amplify these low amplitude signals. A DC 5 V is supplied from the smartphone to all other circuits through a micro USB port. After the ECG signal is received, to calculate a heart rate, we have used linear regression theory. Using the linear regression, we obtain an approximate relation for the linear functions. The equation used in the calculation of slope for the linear functions has been shown in the following equation. x and y values are two variables used and the notation- and- are the average of the variables x and y. Here, variable n is the value which indicates the number of data in the signal. After this, we have removed the noise associated with the raw signal using the filter techniques [11]. The calculation of slope value for the whole data set is obtained by comparing present slope from the previous slope value. Using the slope calculation, we have determined the Q, R, and S points associated with the ECG signal. We have used these Q, R, and S values in calculating the heart rate. Flowchart of the followed algorithm is shown in Fig. 6.

$$SLOPE = \frac{\sum_{i=1}^{n} (x_i y_i - n x \bar{y})}{\sum_{i=1}^{n} (x_i^2 - n x^2)}$$

Fig. 6 Flowchart of the
proposed algorithm
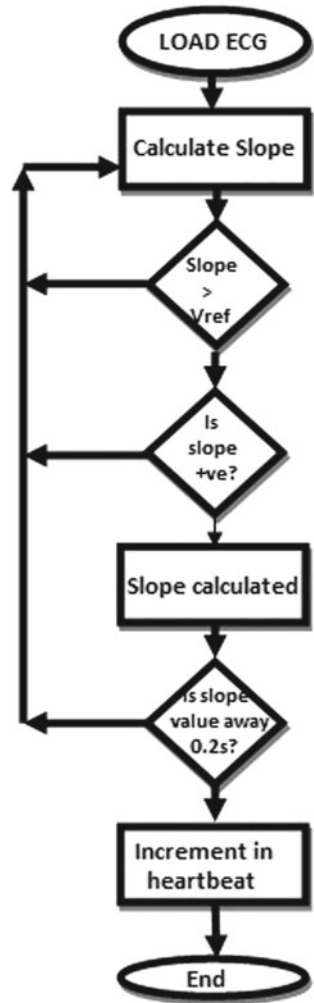
In order to transmit the recorded signal, we have implemented the delta modulator using FPGA IC. The implemented circuit of the delta modulator takes the analog input voltage from the instrumentation ECG amplifier in the range of 0.25–1.25 V @ 1.75 kHz frequency, and changes an analog input signal into an 8-bit digital output signal the data rate in the conversion is obtained as 23.4 kS/s. This 8-bit of the data is transmitted to the smartphone using the USB link [12], and the USB link takes the 8-bit digital output and serializes the data, then it transfers into the smartphone. The smartphone performs the processing of the data received by the USB link, and after the processing, it plots the received ECG waveform using the display of smartphone, smartphone calculates the heart rate using the received signal, after the plotting the waveform and determining the heart rate the results are stored into the flash memory

Fig. 7 Waveform obtained
in the smartphone
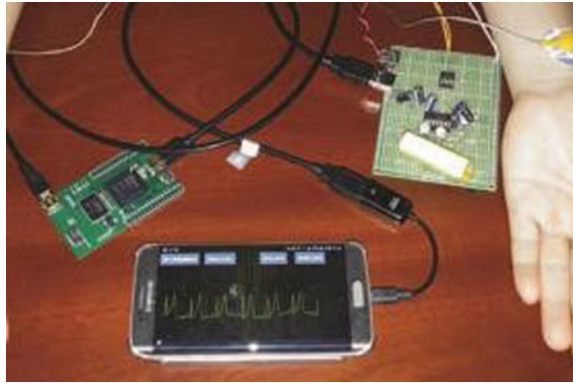
Table 1 Comparative results
between the smartphone and
Matlab

Method	Heart rate
Smartphone	79
Matlab	81
Error	2

organized in the smartphone. Further, we accessed this stored data into the Matlab
software and here, we have plotted the waveform. After this, we have performed the
comparative analysis that explain us the difference between the heart rate calculated
by smartphone and Matlab [13]. We also compared the power requirements of the
proposed system and other existing systems.

4 Result and Comparison

Figure 7 shows the waveform that we have obtained through our implemented sys-
tem. We have displayed the real-time ECG signal on the screen of the smartphone.
Whereas, Fig. 8 presents the waveform that we plotted using the data stored in the
flash memory of the smartphone. This waveform has been plotted after removing the
noise in the Matlab software.

After obtaining the waveform through both the ways, we have performed the com-
parison between the results that we have obtained through smartphone and Matlab.
We have compared the heart rate values obtained using the smartphone with the heart
rate values those we have obtained through Matlab software. Table 1 summarizes the
comparative results of the overall experiment, in terms of the heart rate. The results
presented in Table 1 are the average values of the heart rates calculated by the eight
sets of ECG signal.

We have also presented the comparison between the power requirements of the
monitoring system. Table 2 present the record of the power consumed by the individ-
ual sections of the implemented monitoring system for ECG monitoring. The values
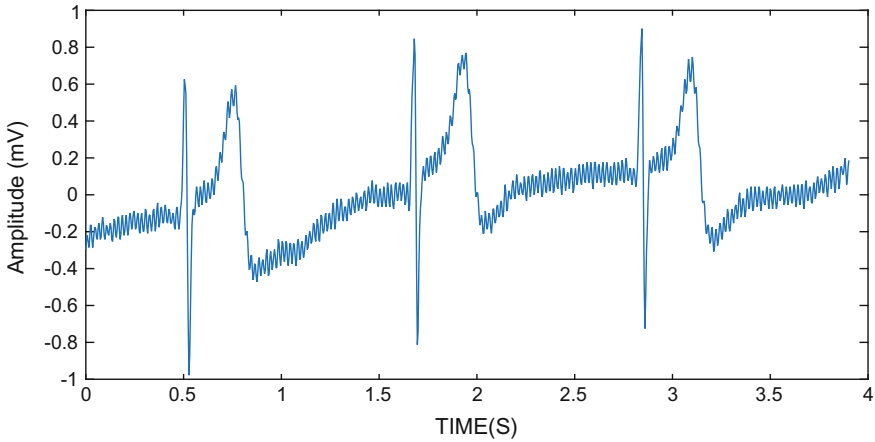
Fig. 8 ECG waveform in the Matlab

Table 2 Power consumption analysis

	Current (mA)	Power (mW)
ECG AMP	4	20
FPGA board	66	330
Total	70	350

presented in Table 2 are the average values calculated for the eight sets of the ECG data.

5 Conclusion and Future Scope

We have successfully implemented the monitoring system for the electrical activity generated by the heart. We have displayed the real-time ECG waveform on the screen of a smartphone and further, we have stored the collected ECG values in the flash memory of the smartphone. After that, we have compared the waveform generated by the same ECG values in the Matlab software. These waveforms are almost similar to the smartphone's waveform. Hence, we have concluded that using smartphone, we can monitor the ECG. Hence, the design of a portable ECG monitoring system has been implemented. The future work related to this system will include a cloud-based access of the data. In that, we will store the ECG data on the cloud database using nano wireless electrodes and then, we will try to access that data in real time like cricket score, running train status, etc. That will give a new direction for the ECG tele-monitoring systems. This will enable the cardiologist to monitor their patients anywhere and at any point of time without any extra circuitry; just a smartphone will work as the ECG machine for all of their patients. And the patient will also be able

to view there ECG report without any extra requirement. That will result in a very effective monitoring and will reduce the effects of CVDs.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the Institutional Ethics Committee (IEC) of the North Eastern Hill University Shillong-793022 and North Eastern Indira Gandhi Regional Institute of Health and Medical Sciences with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent Informed consent was obtained from all individual participants included in the study.

References

1. Gemperle, F., Kasabach, C., Stivoric, J., Bauer, M., Martin, R.: Design for wearability. In: International Symposium on Wearable Computers, Pittsburgh, PA, pp. 116–122 (1998)
2. Jimenez, L.O., Landgrebe, D.A.: Supervised classification in high dimensional space: geometrical, statistical, and asymptotical properties of multivariate data. IEEE Trans. Syst. Man Cybern. Part C **28**(1), 39–54 (1998)
3. Pantelopoulos, A., Bourbakis, N.: survey on wearable sensor-based systems for health monitoring and prognosis. IEEE Trans. Syst. Man, Cybern. C Appl. Rev. **40**(1), 112 (2010)
4. Selye, H.: The Stress of Life, 2nd edn. The McGraw-Hill Companies, New York (1978)
5. Sufi, F., Fang, Q., Khalil, I., Mahmoud, S.: Novel methods of faster cardiovascular diagnosis in wireless telecardiology. IEEE J. Sel. Areas Commun. **27**(4), 537–552 (2009)
6. Abenstein, J.P.: Algorithms for real-time ambulatory ECG monitoring. Biomed. Sci. Instrument. **14**, 73–79 (1978)
7. www.ecglibrary.com
8. Mehta, D., Zaartu, M., Feng, S.W., Cheyne II, H.A., Hillman, R.E.: Mobile voice health monitoring using a wearable accelerometer sensor and a smartphone platform. IEEE Trans. Biomed. Eng. **59**(11), 3090–3096 (2012)
9. Bishop, C.M.: Novelty detection and neural network validation. Proc. IEE Conf. Vision Image Signal Process. **141**(4), 217–222 (1994)
10. Orphanidou, C., Clifton, D., Smith, M., Feldmar, J., Tarassenko, L.: Telemetry-based vital sign monitoring for ambulatory hospital patients. In: Proceedings of the IEEE Engineering in Medicine and Biology Society Annual Conference, Minneapolis, MN, USA, pp. 4650–4653 (2009)
11. Wu, W.H., Bui, A.A.T., Batalin, M.A., Liu, D., Kaiser, W.J.: Incremental diagnosis method for intelligent wearable sensor systems. IEEE Trans. Inf Technol. Biomed. **11**(5), 553–562 (2007)
12. Clifton, L., Clifton, D.A., Pimentel, M.A.F., Peter, J.: Predictive monitoring of mobile patients by combining clinical observations with data from wearable sensors. IEEE J. Biomed. Health Inf. **18**(3) (2014)
13. Mukhopadhyay, S.C.: Fellow wearable sensors for human activity monitoring: a review. IEEE Sens. J. **15**(3) (2015)

A Comparative Study Between the Two Cases for the Effect of a Single Toxicant on a Biological Species in Case of Deformity

A. K. Agrawal, Anuj Kumar Agarwal, A. W. Khan and Piyush Kumar Tripathi

Abstract In this paper, a comparative study has been carried out to study the effect of a single toxicant on a biological species. The comparison study is based on two cases, in the first case, the toxicant is being constantly emitted in the environment from some external sources and in the second case, the toxicant is assumed to be discharged in the environment by the biological species itself. In both cases, we have considered a situation that aftereffect of this toxicant, some members of biological species show deformity as incapable in reproduction. This comparative study shows that toxicant is lethal in both cases. But toxicant affects more severe and fast when emitted by biological species itself.

A. K. Agrawal (✉) · P. K. Tripathi
Department of Mathematics, Amity School of Applied Sciences,
Amity University Uttar Pradesh, Lucknow, India
e-mail: akagrawal@lko.amity.edu

P. K. Tripathi
e-mail: pktripathi@lko.amity.edu

A. K. Agarwal
Department of Mathematics, School of Management Sciences, Lucknow, India
e-mail: anuj.kumaragarwal20@gmail.com

A. W. Khan
Department of Mathematics, Integral University, Lucknow, India
e-mail: awkhan@iul.ac.in

1 Introduction

To better understand real world, we create miniatures of real-world phenomenon in a set of mathematical equations, which is known as mathematical model. Using these mathematical models, we easily get a proper picture of real-world phenomenon such as environmental health, diseases, biological species growth, economic process, etc., [2, 7, 8, 12].

There are many studies existing in which toxicants' cause of morphological deformities, genetic defects, impotence, decrease libido, and reproductive failure in biological species [1–4, 6–8, 11, 13–16]. Via mathematical models, some studies show that a subclass of biological species is critically affected by environmental toxicants and shows deformity as incapable in reproduction [2, 7, 8]. In 2016, Kumar et al. [7, 8] studied the effect of a single toxicant on reproduction of a biological species in different cases such as toxicant emitted by biological species itself (e.g., industrial and domestic activities, vehicular exhaust, radioactive wastes, use of pesticides, etc.,) and from some external sources (e.g., volcanic eruptions, forest fires, etc.,), respectively. But some reproductive failures due to uptake of toxicants have been observed, such as carbon disulfide cause low sperm count and decreased libido in men and menstrual disturbance in women [1]; Lead effects the male reproductive system by disrupting hormonal regulations, less sperm production in seminiferous tubules of the testes [16]; Inorganic Arsenic causes developmental and reproductive toxicity as affects fetal development, reductions of testosterone and gonadotrophins [6]. These toxicants are emitted in the environment from both ways biological species itself and external sources [1, 10, 17]. So, we give a comparative study between two cases of same kind toxicant's effect on biological species. In case 1, toxicant is constantly emitted in the environment from some external sources. In case 2, toxicant is emitted in the environment by biological species itself. In both cases, aftereffect of same kind toxicant some members of biological species show deformity as incapable in reproduction.

In this comparative study, the models and their results are based on some particular facts such that a logistically growing biological species of population density $N(t)$ at time t, is surviving in a polluted environment having a toxicant with environmental concentration $T(t)$ at time t. This toxicant emitted in the environment from some external sources at a constant rate Q or discharged in the environment by biological species itself at the rate λ. $U(t)$ is the concentration of toxicant $T(t)$, taken up by the biological species $N(t)$ at time t, this toxicant harmfully affected the biological species $N(t)$ and generate a subclass of species $N_D(t)$ having those members which are incapable in reproduction, the subclass of remaining members of species are assumed as $N_A(t)$.

2 Toxicant Constantly Emitted from Some External Sources

The proposed models are

$$\frac{dN_A}{dt} = (b - d)N_A - r_1 N_A U - \frac{r N_A N}{K(T)}$$

$$\frac{dN_D}{dt} = r_1 U N_A - \frac{r N_D N}{K(T)} - (\alpha + d)N_D$$

$$\frac{dT}{dt} = Q - \delta T - \gamma TN + \pi \nu NU$$

$$\frac{dU}{dt} = \gamma TN - \beta U - \nu NU \tag{2.1}$$

$$N(t) = N_A(t) + N_D(t), N_A(0), N_D(0), T(0) \geq 0, U(0) \geq cT(0), 0 \leq \pi \leq 1$$

Since $N(t) = N_A(t) + N_D(t)$, the reduced form of model (2.1) is as follows:

$$\frac{dN}{dt} = rN - \frac{r N^2}{K(T)} - (\alpha + b)N_D$$

$$\frac{dN_D}{dt} = r_1 U(N - N_D) - \frac{r N_D N}{K(T)} - (\alpha + d)N_D$$

$$\frac{dT}{dt} = Q - \delta T - \gamma TN + \pi \nu NU$$

$$\frac{dU}{dt} = \gamma TN - \beta U - \nu NU \tag{2.2}$$

2.1 Equilibrium Points and Dynamical Behavior

The model (2.2) has two non negative equilibrium points $E_1\left(0, 0, \frac{Q}{\delta}, 0\right)$ and $E_2(N^*, N_D^*, T^*, U^*)$, where E_1 is a saddle point. E_2 is locally asymptotically stable under Routh–Hurwitz conditions of model (2.2) and becomes unstable and model corresponding to the parameter Q, under the Liu's criterion [5, 9]. The model (2.2) shows a supercritical Hopf bifurcation and the stable bifurcating periodic solutions

exist for $Q > Q^*$, where Q^* is the critical value of bifurcation parameter at which model system (2.2) undergoes a Hopf bifurcation.

3 Toxicant Discharged in the Environment by Biological Species Itself

The proposed models are

$$\frac{dN_A}{dt} = (b - d)N_A - r_1 N_A U - \frac{r N_A N}{K(T)}$$

$$\frac{dN_D}{dt} = r_1 U N_A - \frac{r N_D N}{K(T)} - (\alpha + d)N_D$$

$$\frac{dT}{dt} = \lambda N - \delta T - \gamma T N + \pi \nu N U$$

$$\frac{dU}{dt} = \gamma T N - \beta U - \nu N U \tag{3.1}$$

$$N(t) = N_A(t) + N_D(t), \, N_A(0), \, N_D(0), \, T(0) \geq 0, \, U(0) \geq cT(0), \, 0 \leq \pi \leq 1$$

Again, the reduced form of model (3.1) using the fact that $N(t) = N_A(t) + N_D(t)$.

$$\frac{dN}{dt} = rN - \frac{rN^2}{K(T)} - (\alpha + b)N_D$$

$$\frac{dN_D}{dt} = r_1 U(N - N_D) - \frac{r N_D N}{K(T)} - (\alpha + d)N_D$$

$$\frac{dT}{dt} = \lambda N - \delta T - \gamma T N + \pi \nu N U$$

$$\frac{dU}{dt} = \gamma T N - \beta U - \nu N U \tag{3.2}$$

3.1 Equilibrium Points and Dynamical Behavior

The model (3.2) have two non negative equilibrium points $E_3(0, 0, 0, 0)$ and $E_4(N^*, N_D^*, T^*, U^*)$, where E_3 is a saddle point. E_4 is locally asymptotically stable under Routh–Hurwitz conditions of model system (3.2) and becomes unstable corresponding to the parameter λ under the Liu's criterion [5, 9]. The model (3.2) shows a supercritical Hopf bifurcation and the stable bifurcating periodic solutions exist for $\lambda > \lambda^*$, where λ^* is the critical value of bifurcation parameter at which model system (3.2) undergoes a Hopf bifurcation.

4 Comparative Study

We executed this comparative study with the help of carrying capacity function

$$K(T) = K_0 - \frac{b_1 T}{1 + b_2 T}$$

and a set of parameters is as follows:

$$b = 0.044, K_0 = 10.0, d = 0.0001, b_1 = 0.01, b_2 = 1.0,$$

$$\alpha = 0.0001, r_1 = 0.8, \delta = 0.025, \gamma = 0.0001, \pi = 0.5,$$

$$\nu = 0.0001, \beta = 0.01, \mathbf{Q} = 0.0001, \boldsymbol{\lambda} = \mathbf{0.0001}$$

Also, the initial values for both cases is

$$N(0) = 0.50, N_D(0) = 0.00, T(0) = 0.00, U(0) = 0.00.$$

Here, we set the initial emission of single toxicant at equal rates $Q = \lambda = 0.0001$ for both cases. The model (2.2) and the model (3.2) become locally stable at equilibrium point $E_2(9.9365, 0.0628, 0.0039, 0.0004)$ and $E_4(9.4282, 0.5364, 0.0364, 0.0031)$, respectively. It shows that environmental/uptake concentration of toxicant emitted by species itself (with value **0.0364/0.0031**) is higher than that of some external sources (with value **0.0039/0.0004**). The density of total population (with value **9.9365**) when toxicant emitted by species itself is lower than density of total population (with value **9.9482**) when toxicant emitted from some external sources. The density of deformed subclass (with value **0.5364**) when toxicant emitted by species itself is higher than density of deformed subclass (with value **0.0628**) when toxicant emitted from some external sources.
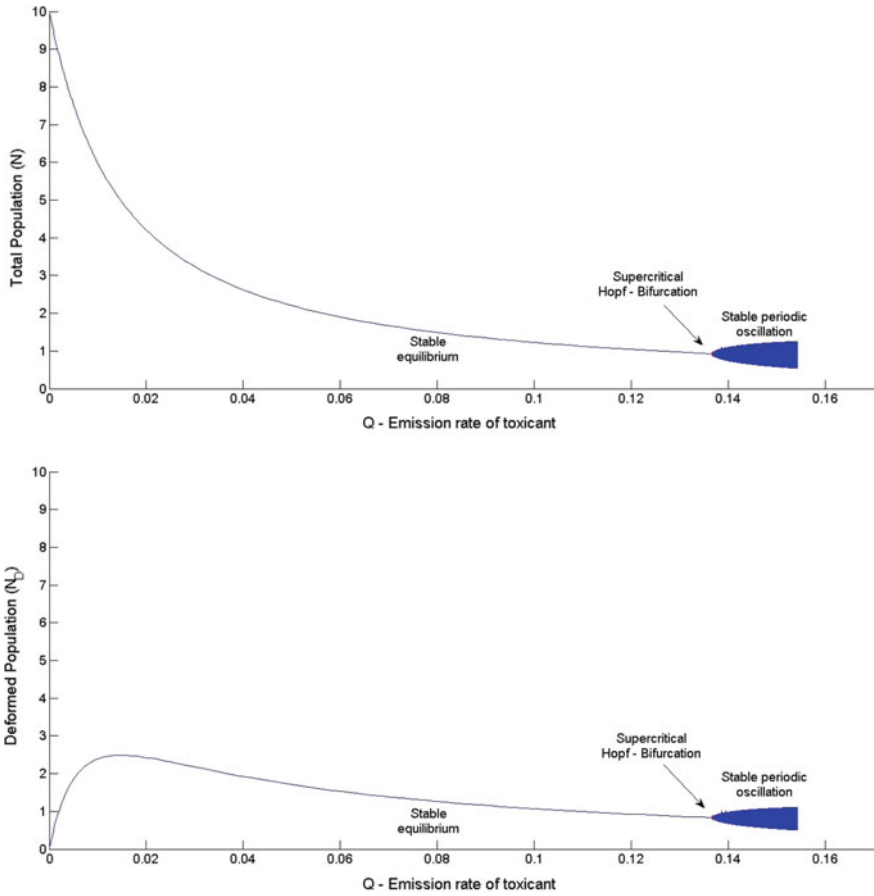
Fig. 1 Dynamic behavior of densities of total and deformed populations (N and N_D) for emission rate Q

As we increase the emission of single toxicant rates Q and λ in each case, the model (2.2) and the model (3.2) becomes unstable and shows Hopf bifurcation at **$Q=0.13674$** and **$\lambda=0.01766$**, respectively. It shows that the model (3.2) becomes unstable faster than the model (2.2) corresponding to the emission rate of toxicant (see Figs. 1 and 2).
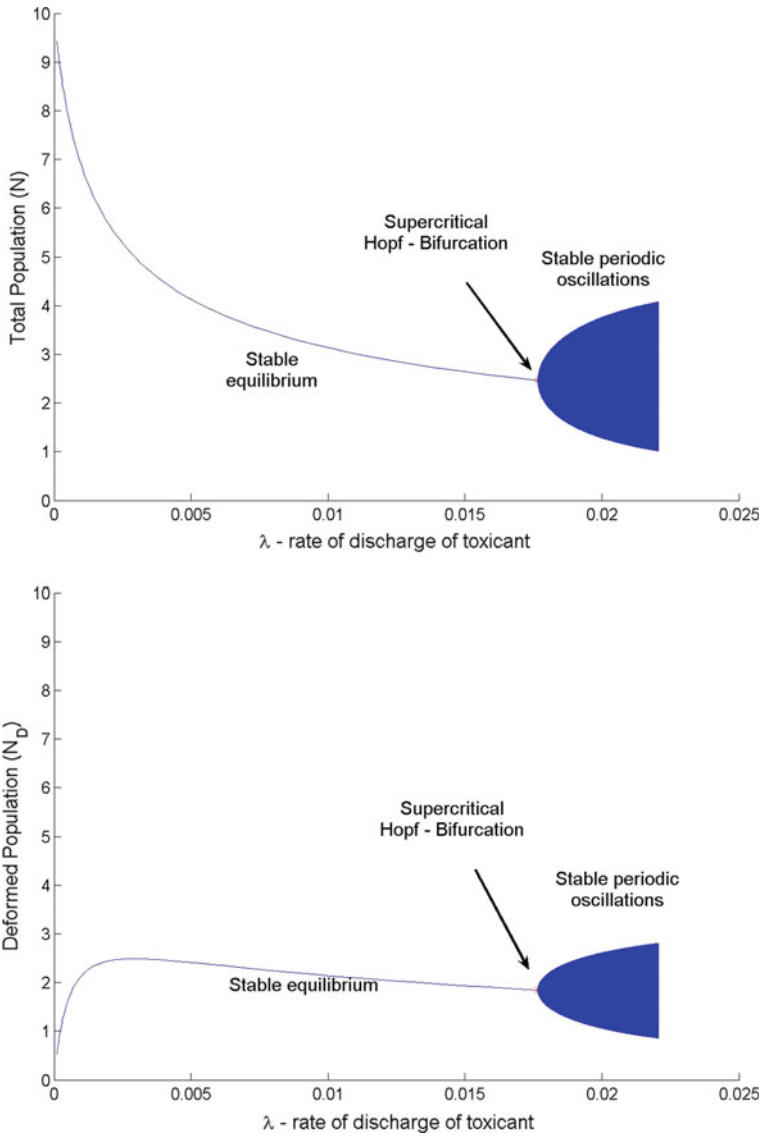
Fig. 2 Dynamic behavior of densities of total population N and deformed population N_D with respect to the parameter λ

5　Conclusion

This paper describes a comparative study between the two cases of effect of a single toxicant on a biological species. The two cases are defined as the toxicant is being constantly emitted in the environment from some external sources and discharged in the environment by the biological species itself. Aftermath of this comparative study, we found that either the toxicant emitted by biological species itself or toxicant emitted by some external sources both are lethal for species. Consequently, when toxicant emitted by species itself, the environmental concentration of toxicant increases faster and affected the biological species sooner than other case.

References

1. Agency for Toxic Substances and Disease Registry (ATSDR).: Toxicological Profile for Carbon disulfide (Update). Public Health Service, U.S. Department of Health and Human Services, Atlanta (1996)
2. Agrawal, A.K., Shukla, J.B.: Effect of a toxicant on a biological population causing severe symptoms on a subclass. S. Pac. J. Pure Appl. Math. **1**(1), 12–27 (2012)
3. Cushman, R.E.: Chironomid deformities as indicators of pollution from synthetic, coal—derived oil. Freshw. Biol. **14**, 179–182 (1984)
4. David, M., Marigoudar, S.R., Patil, V.K., Halappa, R.: Behavioral, morphological deformities and biomarkers of oxidative damage as indicators of sublethal cypermethrin intoxication on the tadpoles of D. melanostictus (Schneider, 1799). Pestic. Biochem. Physiol. **103**(2), 127–134 (2012)
5. Hassard, B.D., Kazarinoff N.D., Wan, Y.H.: Theory and application of Hopf bifurcation. London Mathematical Society Lecture Note Series, vol. 41 (1981)
6. Kim, Y.J., Kim, J.M.: Arsenic toxicity in male reproduction and development. Dev. Reprod. **19**(4), 167–180 (2015)
7. Kumar, A., Agrawal, A.K., Hasan, A., Misra, A.K.: Modeling the effect of toxicant on the deformity in a subclass of a biological species. Model. Earth Syst. Environ. **2**(1), 1–14 (2016)
8. Kumar, Anuj, Khan, A.W., Agrawal, A.K.: The effect of an external toxicant on a biological species in case of deformity: a model. Model. Earth Syst. Environ. **2**(3), 1–8 (2016)
9. Liu, W.M.: Criterion of Hopf bifurcation without using eigenvalues. J. Math. Anal. Appl. **182**(1), 250–256 (1994)
10. Rasmussen, R.A., Khalil, M.A.K., Dalluge, R.W., Penkett, S.A., Jones, B.: Carbonyl sulfide and carbon disulfide from the eruptions of Mount St. Helens. Science **215**(4533), 665–667 (1982)
11. Saquib, M., Ahmad, A., Ansari, K.: Morphological and physiological responses of Croton Bonplandianum baill to air pollution. Ecoprint **17**, 35–41 (2010)
12. Shukla, J.B., Agrawal, A.K., Sinha, P., Dubey, B.: Modeling effects of primary and secondary toxicants on renewable resources. Nat. Res. Model. **16**(1), 99–120 (2003)
13. Sun, P.L., Hawkins, W.E., Overstreet, R.M., Brown-Peterson, N.J.: Morphological deformities as biomarkers in fish from contaminated rivers in Taiwan. Int. J. Environ. Res. Publ. Health **6**, 2307–2331 (2009)

14. Veeramachaneni, D.N.R., Palmer, J.S., Amann, R.P.: Long term effects on male reproduction of early exposure to common chemical contaminants in drinking water. Hum. Reprod. **16**(5), 979–987 (2001)
15. Veroli, A.D., Selvaggi, R., Goretti, E.: Chironomid mouthpart deformities as indicator of environmental quality: a case study in Lake Trasimeno (Italy). J. Environ. Monit. **14**, 1473–1478 (2012)
16. Vigeh, M., Smith, D.R., Hsu, P.-C.: How does lead induce male infertility? Iran. J. Reprod. Med. **9**(1), 1–8 (2011)
17. World Health Organization.: Air Quality Guidelines for Europe, 2nd edn. WHO Regional Office for Europe, Copenhagen, Denmark (2000)

Nanoswimmer Energy Transduction System: Influence of Branching

Shivani Nain, Jitendra Singh Rathore and Niti Nipun Sharma

Abstract Nanoswimmers are of interest among researchers for their utility in propelling nanorobots to specific target for drug delivery, nanosurgery, in vivo biomedical applications such as in treatment of brain tumor and Alzheimer's disease and similar applications. On-board powering is the major concern for locomotion of nanoswimmer and is being considered to be addressed by energy transduction mechanism to harness energy from surrounding using energy of stochastic vibrations by electrostatic, electromagnetic, and piezoelectric means. Among all, piezoelectric is emerging as a promising conversion transduction mechanism of energy harnessing for artificial nanoswimmer. In this context, in present work, an elastic flagellum of a nanoswimmer is modeled as a cantilever beam and a simulation study is done in COMSOL. The novel design of branched flagellum is conceived, modeled, and simulated. COMSOL simulation studies have been performed to compare the effect of primary and secondary branching in flagellum design in terms of stress and electric potential. Enhancement in stress and electric potential is observed approximately 20 and 15% on increasing secondary branching uniformly on the main structure of cantilever beam towards free end and keeping primary branches constant. An enhanced stress allows for larger efficiency of conversion mechanism and, therefore, it is concluded that branching of flagellum can be pivotal in increasing on-board harnessing of energy for propulsion of nanorobots.

Keywords Nanoswimmer · PVDF · Piezoelectric · Branches

S. Nain (✉) · J. S. Rathore
Mechanical Engineering Department, Birla Institute
of Technology & Science, Pilani 333031, India
e-mail: shivaninain511@gmail.com

N. N. Sharma
School of Automobiles, Mechanical & Mechatronics,
Manipal University, Jaipur 303007, India

1   Introduction

Advancement in the area of nanotechnology has drawn the attention of researchers to design locomotion of nanorobots over the past six decades. Locomotion design is combined efforts of chemist, biologist, physicists, pharmacists, engineers, and mathematicians. Scientists are imitating natural bacteria such as *E.coli* and *Paramecium* at the nanoscale and investigated their modes of propulsion such as planar and helical theoretically [1–5]. Various types of actuation mechanisms have been explored till now such as chemical actuation [6], bacterial actuation [7], magnetic actuation [8], and thermal actuation mechanism [9] for artificial nanoswimmer and it possesses some limitations [10]. To overcome those limitations of different actuation mechanisms, piezoelectric-based on-board energy harnessing scheme is an alternative to artificial nanoswimmer.

Energy transduction mechanism implies conversion of mechanical or vibrational energy from the surrounding environment into electrical form by electrostatic [11], electromagnetic [12] and piezoelectric approach [13]. Piezoelectric conversion technique is most suitable for energy harnessing purpose at micron scale because it does not require any external means in comparison with electrostatic and electromagnetic mechanism. Selection of material becomes an essential parameter for fabricating and navigating artificial nanoswimmer inside the human body environment. Piezoelectric material such as lead–zirconate–titanate (PZT), lead–titanate ($PbTiO_2$), lead––zirconate ($PbZrO_3$), and barium–titanate ($BaTiO_3$) are being employed in energy harnessing process in the past few decades. Due to some constraints of piezo-ceramics [14], piezoelectric polymers, polyvinylidenefluoride (PVDF) has been synthesized and can be used for energy-harvesting application [15]. PVDF seems most appropriate piezoelectric material because of flexibility, biocompatible, lightweight, and inexpensive [16] to design on-board powering scheme for nanoswimmer.

In nature, cilia and flagella exhibit rotation and beating as a mode of propulsion for eukaryotic and prokaryotic bacteria. In the present design, cantilever beam sculpted as cilia of PVDF piezoelectric material fixed at one end. The aim of the present study is to amplify energy harnessing to actuate nanoswimmer. The design is instigated by amplification of stressing on the trunk of tree due to branching (static load) and fluidic pressure (dynamics load). The dynamic model of tree considered mass (mass of beam and branches), spring (Young's modulus of PVDF), and damping (viscosity of fluid, placement of branches). The conception of dynamic model is presented in Fig. 1, which comprises of primary branch as mass m1 attached to the beam (M) and secondary branch of mass (m2) to primary branch. Different designs of secondary branching is simulated in COMSOL and optimized for enhanced stress and electric potential generation for on-board powering of nanoswimmer. The resulting data is examined to determine the effect of fluid pressure load and the response of the tree branch structure of energy transduction system for artificial nanoswimmer.
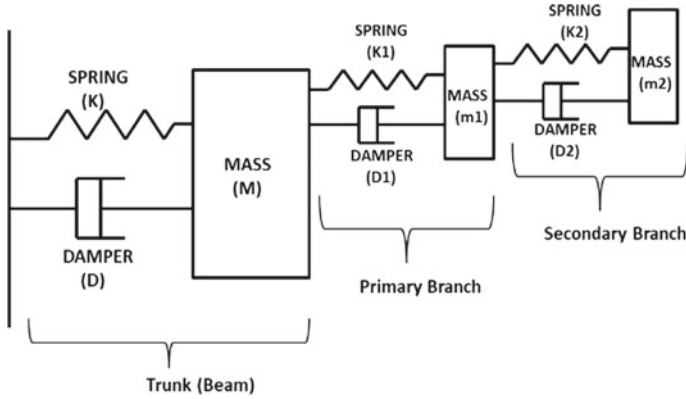
Fig. 1 Dynamic model of tree branch concept [17]

2 Design and Simulation

Investigation is being performed on different designs of energy transduction system. In the previous analysis [18], four designs were explored to study the effect of number of primary branches and their placement. The spaces between branches are kept constant. The design shown in Fig. 2 branches toward the distal end are showing maximum amplification of stress by 68% and leads to the generation of electric potential approximately 60% through simulation [18]. It needs to investigate further to study the effect of secondary branches (static load) and fluidic pressure (dynamic load) for stress development on the main structure of beam cilia and is attempted in the subsequent section.

In Fig. 3, design 1 secondary branches are increased from 4 to 8 towards the fixed end of beam while primary branches are kept same in number as 4. In design 1 (a), design 1(b) and design 1 (c) (Fig. 3) 4 and 8 secondary branches are positioned uniformly on 4 primary branches. In design 2, the number of secondary branches is

Fig. 2 Model of unbranched and branched cilia toward free end

4 Primary Branches towards fixed end

4 Primary Branches- 4 Secondary Branches towards fixed end

4 Primary Branches- 8 Secondary Branches towards fixed end

Design 1 (a)   Design 1 (b)   Design 1 (c)

4 Primary Branches towards distal end from fixed end

4 Primary Branches- 4 Secondary Branches towards distal end from fixed end

4 Primary Branches- 8 Secondary Branches towards distal end from fixed end

Design 2 (a)   Design 2 (b)   Design 2 (c)

Fig. 3 Designs of secondary branches (Primary branches constant in number as 4 toward the fixed end and free end of beam as cilia)

increased toward the free end of beam. The gaps between branches are consistent. Modeling and designs of energy transduction mechanism have been carried out in COMSOL multiphysics finite element software.

Three physics namely, Fluid–Structure Interaction (FSI), piezoelectric and electrostatic are combined to envisage the effect of fluid velocity in the form of stress and electric potential on the branched beam structure representing cilia toward the fixed end of the beam. In FSI module, a parabolic velocity profile is introduced as an inlet boundary condition represented by Eq. (1) mentioned below [19]:

Inlet Velocity:

$$[U \times 16(H - Y) \times (H - X) \times Y \times X] \div H^4 \tag{1}$$

Where U is the mean velocity of water defined in the range from 5 to 8 μm/s to maintain Reynolds number less than 1 and to sustain laminar flow regime. H is the height of the outer domain in which fluid is flowing, i.e., 300 μm to avoid wall effects. X and Y are planes in which fluid is entering.

Fluid velocity leads to the generation of stress on the piezoelectric beam, which is considered as a boundary load condition in the piezoelectric module. In electrostatic physics, upper surface of beam is taken as floating potential, where the developed electric potential is measured toward the fixed end and the lower surface is grounded.

Table 1 Geometric and material properties of the piezoelectric energy transduction system

Geometric parameters	Length (μm)	Width (μm)	Thickness (μm)
Beam	50	0.25	0.25
Branches	5	0.025	0.025
Material properties	Young's modulus (GPa)	Density (kg m^{-3})	Poisson's ratio
PVDF	2	1789	0.3

The geometric parameters and material properties are shown in Table 1. PVDF is chosen for piezoelectric beam because of biocompatibility and suitability for in vivo application.

3 Result and Discussion

A design of the piezoelectric energy transduction system is simulated in COMSOL. The aim of the design is to amplify the stressing on the main structure of beam by doing branching of beam as tree branch structure. The pruning concept (cutting of branches) is required to reduce the stress on the trunk of tree, but amplification of stress necessitates enhancing the electric potential for movement of nanoswimmer. Both static and dynamic loads are considered for evaluation of stresses developed on the trunk (beam) of the tree. Energy transduction scheme for artificial nanoswimmer consists of four primary branches and the number of secondary branches varies from 4 to 8. The results have been taken toward the fixed end of beam. The stress is developed on the surface of beam due to fluidic velocity applied to the inlet of the outer domain to maintain low Reynolds number. This developed stress is converted into electric potential through the electrostatic module coupled to the piezoelectric module. The simulation results toward the fixed end are plotted in Fig. 4 for design 1 and design 2 (refer Fig. 3) by varying the number of secondary branches toward the fixed end and free end of the beam, shows von Mises stress on y-axis and fluid velocity on x-axis. It is observed from the plotted results that as the number of secondary branches increases stress also increases, which is being transformed into electric potential by electrostatic physics. The primary branches are kept constant but their placement varies toward the fixed end and distal end of the beam. The von Mises stress is found to be increased up to approximately 2515 N/m^2 and percentage increase in stress is 20% on increasing number of secondary branches from 4 to 8 with respect to 4 primary branches toward distal end of the beam. The von Mises stress is transformed into electric potential by electrostatic physics and is increased 0.55 μV approximately 15%, which seems to be favorable for energy harnessing scheme for nanoswimmer.

Design 1: Von Mises Stress v/s Velocity	Design 2: Von Mises Stress v/s Velocity
% increase in stress- 15.84% (from 4 primary branches to 4 primary branches and 8 secondary branches).	% increase in stress- 20.09% (from 4 primary branches to 4 primary branches and 8 secondary branches).
Design 1: Electric Potential v/s Velocity	Design 2: Electric Potential v/s Velocity
% increase in electric potential- 11.95 % (from 4 primary branches to 4 primary branches and 8 secondary branches).	% increase in electric potential- 15.23% (from 4 primary branches to 4 primary branches and 8 secondary branches).

Fig. 4 Variation in stress and electric potential on increasing secondary branches

4 Conclusion

The energy transduction design for artificial nanoswimmer simulated in COMSOL is based on the concept of tree branching structure. Here, static load and dynamic load such as effect of secondary branches and fluid velocity has been investigated and compared by incorporating fluid structure interaction and piezoelectric module. The objective is to design the energy transduction scheme using piezoelectric material to propel artificial nanoswimmer for in vivo biomedical application. In this paper, analysis of secondary branches keeping primary branches same in number as 4, leads to an increase in stress and electric potential. Two designs of the energy transduction

system have been compared on the basis of their placement of primary branches and by increasing secondary branches on it. The stress is within the allowable limit of PVDF material. An increase in stress and electric potential is observed approximately 20 and 15% for design 2 in which branches are toward free end of the beam with respect to the fixed end. Tertiary branches need to be explored further for optimization of energy transduction system for locomotion of nanoswimmer.

References

1. Taylor, G.: Analysis of the swimming of microscopic organisms. Proc. R. Soc. London. Ser. A. Math. Phys. Sci. **209**, 447–461 (1951)
2. Lauga, E., Powers, T.R.,: The hydrodynamics of swimming microorganisms. Rep. Prog. Phys. **72**, 96601–96636 (2009)
3. Rathore, J.S., Sharma, N.N.: Engineering nanorobots: chronology of modeling flagellar propulsion. J. Nanotechnol. Eng. Med. **1**, 31001–31007 (2010)
4. Deepak, K., Rathore, J.S., Sharma, N.N.: Nanorobot propulsion using helical elastic filaments at low Reynolds numbers. J. Nanotechnol. Eng. Med. **2**, 110091–110096 (2011)
5. Kotesa, R.S., Rathore, J.S., Sharma, N.N.: Tapered flagellated nanoswimmer: comparison of helical wave and planar wave propulsion. Bionanoscience. **3**, 343–347 (2013)
6. Loget, G., Kuhn, A.: Electric field-induced chemical locomotion of conducting objects. Nat. Commun. **2**, 535–540 (2011)
7. Zhang, L., Abbott, J.J., Dong, L., Peyer, K.E., Kratochvil, B.E., Zang, H., Bergeles, C., Nelson, B.J.: Characterizing the swimming properties of artificial bacterial flagella. Nano Lett. **9**, 3663–3667 (2009)
8. Wang, L., Xu, H., Zhai, W., Huang, B., Rong, W.: Design and characterization of magnetically actuated helical swimmers at submillimeter-scale. J. Bionic Eng. **14**, 26–33 (2017)
9. Sul, O.J., Falvo, M.R., Taylor II, R.M., Washburn, S., Superfine, R.: Thermally actuated untethered impact-driven locomotive microdevices. Appl. Phys. Lett. **89**, 203512–203515 (2006)
10. Nain, S., Sharma, N.N.: Propulsion of an artificial nanoswimmer: a comprehensive review. Front. Life Sci. **8**, 2–17 (2015)
11. Mitcheson, P.D., Miao, P., Stark, B.H., Yeatman, E.M., Holmes, A.S., Green, T.C.: MEMS electrostatic micropower generator for low frequency operation. Sens. Actuat. Phys. **115**, 523–529 (2004)
12. Arnold, D.P.: Review of microscale magnetic power generation. IEEE Trans. Magn. **43**, 3940–3951 (2007)
13. Truitt, A., Mahmoodi, S.N.: A review on active wind energy harvesting designs. Int. J. Precis. Eng. Manuf. **14**, 1667–1675 (2013)
14. Anton, S.R., Sodano, H.A.: A review of power harvesting using piezoelectric materials (2003–2006). Smart Mater. Struct. **16**, 1–24 (2007)
15. Dakua, I., Afzulpurkar, N.: Piezoelectric energy generation and harvesting at the nano-scale: materials and devices. Nanomater. Nanotechnol. **3**, 1–16 (2013)

16. Majumdar, R., Singh, N., Rathore, J.S., Sharma, N.N.: In search of materials for artificial flagella of nanoswimmers. J. Mater. Sci. **48**, 240–250 (2013)
17. James, K.: Dynamic loading of trees. J. Arboric. **29**, 165–171 (2003)
18. Nain, S., Rathore, J.S., Sharma, N.N.: Harness enough energy for locomotion of an artificial nanoswimmer: design and simulation. In: Proceeding in AIP on 3rd International Conference on Emerging Technologies: Micro to Nano (ETMN-2017), pp. 1–4 (2017)
19. Miller, S.T., Campbell, R.L., Elsworth, C.W., Pitt, J.S., Boger, D.A.: An overset grid method for fluid-structure interaction. World J. Mech. **4**, 217–237 (2014)

Design of an Adaptive Soft Sensor for Measurement of Liquid Level Independent of Liquid

K. V. Santhosh and Sneha Nayak

Abstract The proposed paper discusses about a sensing technique for measurement of the liquid level. The objective of the proposed technique is it should be able to measure the liquid level even with impurities added to it. Usually, a measurement technique is designed and calibrated to particular conditions. In case of the liquid level measurement, the constraints would be type of liquid, temperature of liquid, and so on. This paper concentrates on designing a system which when used along with the traditional signal conversion circuit should be able to produce output, which would be accurate even when some impurities are added to the liquid. The designed system consists of an observer in the form of Kalman filter which would estimate the disturbance parameter, and suppress its effect from the measured signal. The designed technique is implemented and tested using the practical system. Results show successful implementation of the proposed technique.

Keywords Adaptation · Observer · Kalman filter · Liquid level

1 Introduction

Level is an important parameter which is measured and controlled in most of the process control industries. Level measurement may be that of water, petroleum, sugar, or any other form of solid or liquid. In applications such as wastewater treatment plant, fuel monitoring, food and dairy, dye making, pharmaceutical industries, etc., level measurement is of primary concern. Accuracy plays a vital role in level measurement. Inappropriate measurements can cause the levels to be excessively higher or lower than their measured value and thus leading to defective process/product.

K. V. Santhosh · S. Nayak (✉)
Department of Instrumentation and Control Engineering, Manipal Institute
of Technology, Manipal Academy of Higher Education, Manipal 576104, India
e-mail: snehanayak31@gmail.com

K. V. Santhosh
e-mail: kv.santhu@gmail.com

Liquid level measurement is basically classified into two subdivisions, namely contact and noncontact type. In the proposed work, basic study is made on the available measurement techniques to understand its behavior, a few of them are reported.

Paper [1] reports a novel technique for measuring the level and volume of liquids in a cylindrical container. The important feature of the proposed technique is that it is nonintrusive which avoids unwanted contamination of liquid samples, and it has larger dynamic range of measurement and good sensitivity. Paper [2] discusses a novel level measurement technique using three parallel plate capacitive structures. This technique can be directly applied for measuring the level of any kind of nonconductive liquid without calibration and further it is unaffected by factors such as temperature, humidity, dust, etc. Paper [3] describes a microwave multilevel gaging system for measurement and control of liquid levels in storage tanks which is usually employed in industrial applications. Paper [4] discusses a theoretical study to determine the flow rate and level of poorly conducting liquid using electromagnetic method. In [5], a discussion is made on the properties of humidity sensor that can be used for measuring level in cryogenic liquids such as nitrogen, oxygen, argon under various temperature conditions. In [6], a low-cost level sensing technique has been proposed in which the optical fiber has been employed as a level probe. Further experimental results have proved that different fiber cladding thickness can be used for sensing in different operating ranges. Paper [7] describes an experimental study in which liquid level and specific gravity have been measured simultaneously using a dual optic fiber sensing system. In [8], a liquid level measurement technique has been discussed which uses pulsed laser has several advantages over antenna array technique. Paper [9] discusses a method proposed for continuous level measurement using a fiber optic sensor and it is experimentally validated to be robust against changing environmental conditions, also it is stable and easily reconfigurable. In [10], a liquid level sensor has been demonstrated which is used for discriminating high refractive index. This sensor finds application in biochemical industries because of its features such as low cost, simple in structure, and can be integrated in biochemical probes easily.

Paper [11] reports a technique for measuring the level of liquid using image processing technology. This is a noncontact measurement technique which comprises of a camera, laser light sources, and a PC. In [12], a novel method of liquid level sensing with the knowledge of reflection time of a pulse train is discussed. In addition to this, the acoustic properties such as temperature, liquid viscidity, and density can be estimated. Paper [13] reports a liquid level sensing device with capacitive sensing. In this device, comb electrode is used which does not need additional calibration and is mainly used in biomedical applications. In [14], liquid level measurement is demonstrated based on multimode interference effect. An optical fiber laser sensor is used for level measurement of three different liquids with different refractive index. Paper [15] proposes a technique in which the sensor can continuously measure the dynamic change in the water level and also obtain the speed of the water flow. Paper [16] proposes a noncontact measurement technique in which inductive sensor is used to measure level and conductivity simultaneously with good amount of accuracy. In [17], ultrasonic technology is used to detect and analyze the level of raw oil in an oil

storage tank. Paper [18] discusses time-domain reflectometry technique for monitoring the level of liquids. Also this technique can be used for real-time estimation and to determine the nature of various objects. In [19], a design of calibration circuit for Capacitance Level Sensor (CLS) using artificial neural networks is reported. From the survey carried on, it is evident that level measurement is one of the critical processes in industries. Survey discusses several techniques used for level measurement. Of the above reported techniques, CLS is found to be most widely used sensor. In the proposed paper, we have reported the study on input–output characteristics of the CLS, and discussion is extended to understand its behavioral changes when impurities are present in liquid whose level is to be measured. In the paper, an observer is designed to compute the amount/quantity of such impurities added. The output from the capacitive level sensor is extracted to analyze and a suitable observer is designed to make the measurement robust and also quantify the impurities.

Section 1 gives a brief introduction about the importance of liquid level measurement and surveys are carried out to identify various techniques which is used for level measurement. Section 2 briefs about the basic working principle of capacitive level sensor. Problem statement is discussed in Sect. 3. In Sect. 4, methodology is reported. Section 5 reports the conclusion of the work reported.

2 Capacitive Level Sensor

Capacitive level sensor consists of two parallel plate/electrodes submerged in the liquid whose liquid level needs to be measured as shown in Fig. 1. Across these two plates, capacitance is formed when subjected to a potential and whose capacitance is given by

$$C = \frac{\varepsilon K}{a} \tag{1}$$

where C is the capacitance

ε is permittivity of medium

K is surface area of plates

a is distance between the plates.

Now, as the liquid level varies two regions are formed, one with liquid as a medium and the other with air as medium. These two regions will exhibit variations in its capacitance when the liquid level varies because of change in the dielectric medium [20]. So, the total capacitance will be the summation of these two capacitances which is given by

$$C = C_1 + C_2 \tag{2}$$

where C_1 is the capacitance across plates dipped in liquid

C_2 is the capacitance having air as medium.
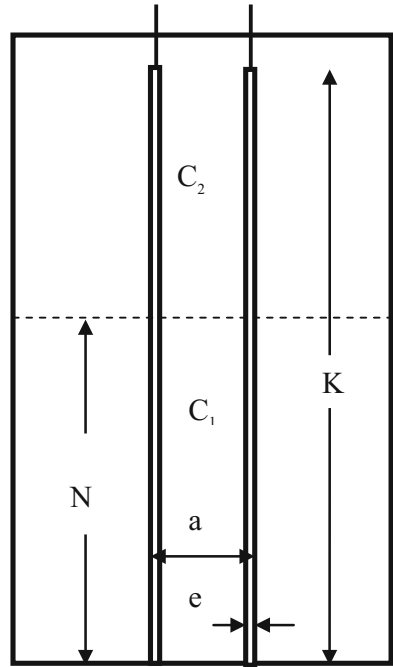
Now, "*C*" can be rewritten as

$$C = \frac{\in_0 \in_r eN}{a} + \frac{\in_0 \in_r e(K-N)}{a} \tag{3}$$

where e is cross-sectional breadth of the plate

K is length of the electrodes/parallel plates

N is length of the electrodes/parallel plate submerged in liquid

$\in_0$ is the permittivity of free space

$\in_r$ is the relative permittivity of liquid.

In Eq. (3), if the parallel plates are replaced by the electrode whose diameter is "e", then the Eq. (3) will be represented as

$$C = \frac{2\pi \in_0 \in_r}{\ln\left[\frac{a}{e} - \sqrt{\left[\frac{a}{e}\right]^2 - 1}\right]} N + \frac{2\pi \varepsilon_0}{\ln\left[\frac{a}{e} - \sqrt{\left[\frac{a}{e}\right]^2 - 1}\right]} (K-N) \tag{4}$$

From Eq. (4), it is clear that the capacitive output "C" is a function of "N" which is level of liquid and also $\in_r$ permittivity of liquid (which in turn is a function of liquid composition).

3　Problem Statement

A study is carried out to understand the input–output behavior of capacitance level sensor using the process set up as seen in Fig. 2. The output capacitance from CLS is converted to frequency first with the help of timer circuit, and then followed by frequency to voltage converter, so as to acquire the signal to computer using a data acquisition card. The acquired voltage corresponding to change in liquid level is plotted in Fig. 3. It is found that input–output characteristic of CLS is almost nonlinear.

Fig. 2 Experimental set up used

Fig. 3 Output obtained from
CLS signal conversion
circuit for variation in liquid
level
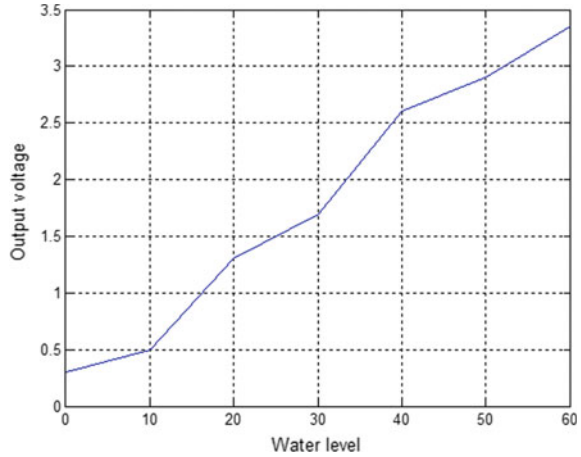
Now to understand the behavior of measurement for variations in parameter, it is decided to change the liquid or manipulate the liquid behavior by adding some additives/impurities. To test, the water solution is mixed with impurities in the form of chalk powder as can be seen in Fig. 4.

For analysis, the experiment is conducted by varying the amount of chalk powder mixed like 5 g/L (specimen 2) of water, 10 g/L (specimen 3). The output obtained for variation of liquid level is plotted in Fig. 5. Figure 5 indicates the nonlinear relations between liquid level and CLS output. Also, it can be seen that output voltage also varies with addition of impurities.

From the discussion, it is clear that the objective of the proposed work is to design a system which would be able to measure the level of liquid linearity and also with variations in added impurities.

4   Problem Solution

In order to achieve the objectives discussed in the earlier section, a design is proposed consisting of an observer. An observer is designed using a Kalman filter. For designing the filter, the first stage would be able to obtain the function's corresponding to liquid level and CLS output. For this, we use the curve-fitting technique as seen in Fig. 6, Fig. 7, and Fig. 8, respectively.

The equation obtained from the curve fitting of specimen 1 can be given by Eq. (5)

$$Y_1 = 2.8 * 10^{-7} x^4 - 4.9 * 10^{-5} x^3 + 2.6 * 10^{-3} x^2 + 0.01x + 0.28 \qquad (5)$$

The equation obtained from the curve fitting of specimen 2 can be given by Eq. (6)

Fig. 4 Experimental set up when the liquid is added with chalk powder
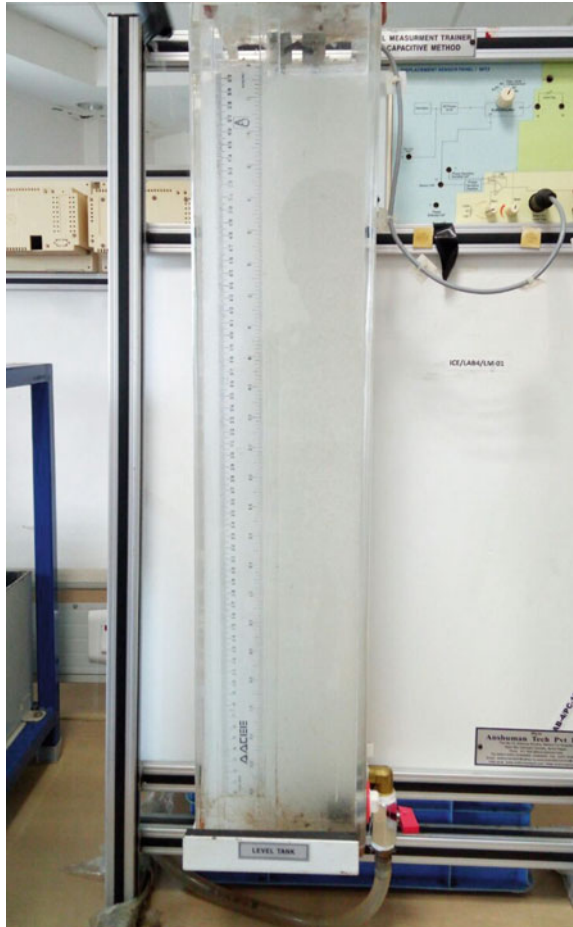
$$Y_2 = 2.5 * 10^{-7}x^4 - 4.4 * 10^{-5}x^3 + 2.3 * 10^{-3}x^2 + 0.006x + 0.68 \qquad (6)$$

The equation obtained from the curve fitting of specimen 1 can be given by Eq. (7)

$$Y_3 = 1.2 * 10^{-6}x^4 - 0.16 * 10^{-3}x^3 + 6.2 * 10^{-3}x^2 - 0.01x + 1.02 \qquad (7)$$

From the above equations and response, it is clear that the output of CLS depends on the liquid level along with the liquid, or in other case, we can consider that the impurities act like the noise for the liquid level measurement. To make the measurement independent of liquid impurities, a filter is designed. In the proposed example, a Kalman filter is considered, the reason for considering Kalman is that it is recursive algorithm which takes into account the history of measurement, and evaluates the linear model, in steps from the current state to the next state by a linear transfor-
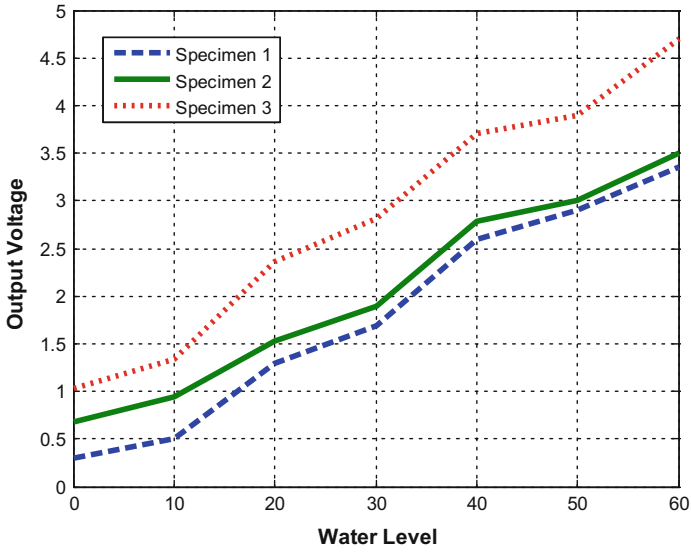
Fig. 5 Output of CLS for variation of liquid level and additives
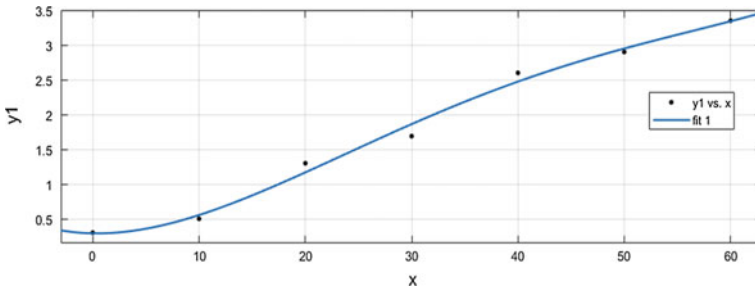
Fig. 6 Curve-fitting function for specimen 1
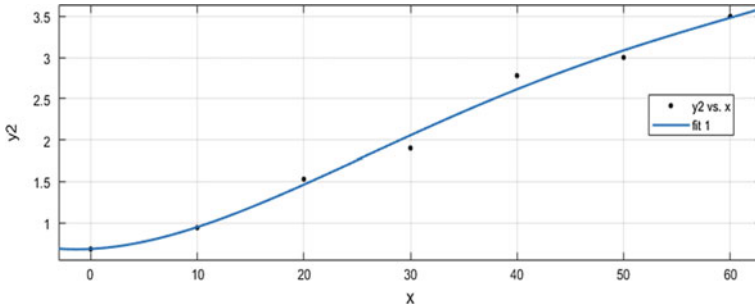
Fig. 7 Curve-fitting function for specimen 2
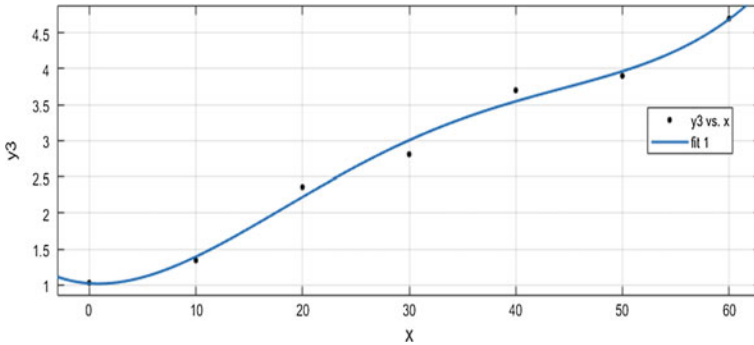
Fig. 8 Curve-fitting function for specimen 3

mation. Or in other words, it functions as a state estimator to estimate unobserved variable based on noisy measurement, in this case chalk powder.

In general, the model used for Kalman is derived from state model given by

$$x(t + 1) = A \cdot x(t) + B \cdot u(t) + e(t) \tag{8}$$

where $x(t+1)$ is considered as current state which is a combination of previous states $x(t)$ and control input $u(t)$ and noise $e(t)$. Assuming A to an identity, and ignoring the input u, Eq. (8) can be rewritten as

$$x(t + 1) = x(t) + e(t) \tag{9}$$

Observer model for the same can be given by measurement vector, y(t) which is given by Eq. (10)

$$y(t) = x(t) + r(t) \tag{10}$$

where r(t) is the measurement noise or noise induced by noisy measurement. The prediction equation can be given by Eq. (11) with P as certainty of prediction, and Q being certainty of measurement.

$$P'(t + 1) = P(t) + R(t) \tag{11}$$

R is noise caused by system, using this prediction equation Kalman gain K(t) can be computed by Eq. (12)

$$K(t + 1) = P'(t + 1)\left(P'(t + 1) \cdot Q(t + 1)\right)^{-1} \tag{12}$$

Using the above equation, the Kalman filter is designed and is used in the proposed technique to make the measurement linear and independent of impurities added.

5   Results and Analysis

Kalman filter is designed to produce an output which is linear and also independent of impurities added. For the case of testing the proposed technique to evaluate the performance, the output is analyzed for variation in liquid level with addition of impurities. For testing, we have considered the liquid as water and impurities as chalk powder. The concentration of chalk powder is varied to test the adaptation of designed algorithm. Table 1 shows the output obtained from designed filter for variation in liquid level and chalk powder concentration.

From the above Table 1, input–output characteristics plot as in Fig. 9, and the error analysis plot as in Fig. 10 clearly indicates that the designed Kalman filter-based observer is able to condition the CLS signal to produce a linear output and also output independent of impurities.

Table 1 Results obtained for measurement

Actual liquid level in cm	Concentration of chalk in g/L	Measured liquid level in cm	Percentage error
5	5	4.86	2.80
8	5	8.01	−0.12
12	10	12.03	−0.25
15	5	14.87	0.87
15	15	14.84	1.07
18	10	17.92	0.44
20	0	19.98	0.10
23	5	22.89	0.48
27	15	27.04	−0.15
30	10	30.11	−0.37
30	15	30.06	−0.20
32	0	31.91	0.28
35	10	35.08	−0.23
40	15	40.91	−2.27
40	20	39.92	0.04
46	20	45.87	0.28
51	10	51.12	−0.24
54	5	54.14	−0.26
58	10	58.11	−0.19
58	0	57.92	0.14
58	15	58.27	−0.47
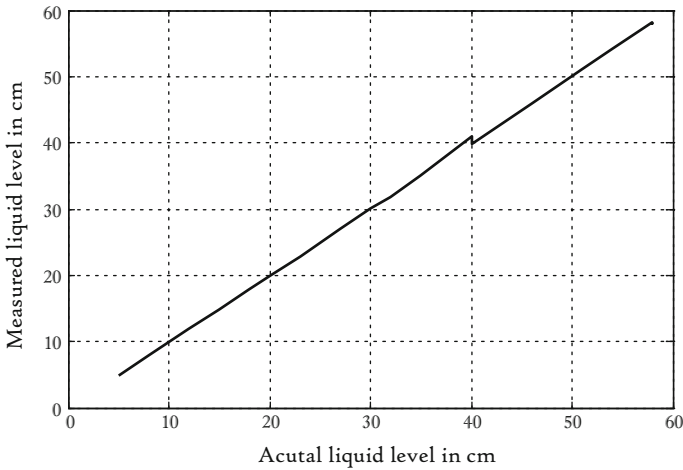
Fig. 9 Input–output characteristics of the proposed system
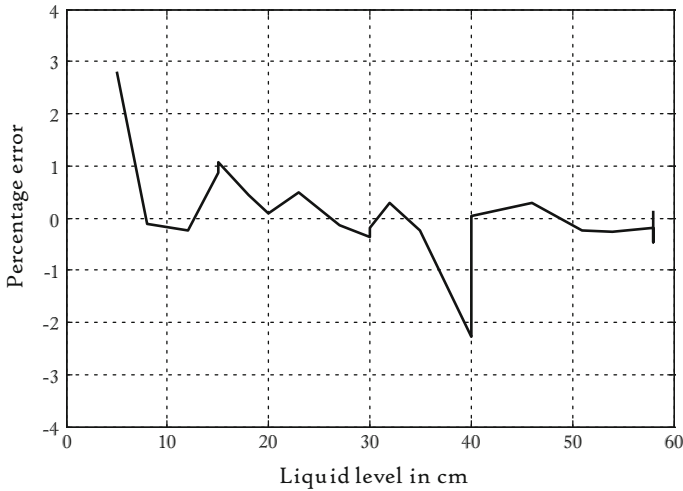
Fig. 10 Percentage error for measured output

6 Conclusion

Design of an adaptive liquid level measurement system, level process being one of critical measurement process in many industries was considered here in the proposed system. The objective of the reported work was to design a system which can produce a linear measurement, even with changes in liquid. For measurement, a CLS was considered, because of the fact that its output varies with liquid level and also liquid. Output of CLS was acquired on the system to design the Kalman filter, so as to nullify

the noise (or effect of impurities) on measurement. The designed system was tested with different liquid levels and varying concentration of impurities, for which the system produced accurate and adaptive results as can be seen in Table 1, Figs. 9, and 10. The root mean square of percentage error produced by the reported technique is 0.88%, making it practically usable process.

References

1. Singh, H.K., Chakroborty, S.K., Talukdar, H., Singh, N.M., Bezboruah, T.: A new non-intrusive optical technique to measure transparent liquid level and volume. IEEE Sens. J. **11**(2), 391–398 (2011)
2. Canbolat, H.: A novel level measurement technique using three capacitive sensors for liquids. IEEE Trans. Instrument. Meas. **58**(10), 3762–3768 (2009)
3. Weiß, M., Knochel, R.: A sub-millimeter accurate microwave multilevel gauging system for liquids in tanks. IEEE Trans. Microw. Theor. Techniq. **49**(2), 381–384 (2001)
4. Velt, I.D., Mikhailova, Y.D.: An electromagnetic method of measurement of the flow rate and level of a liquid in free-flow pipelines. Meas. Tech. **56**(7), 797–804 (2013)
5. Arkharov, I.A., Emel'yanov, V.Y.: An HTS electronic unit for cryogenic liquid level measurement. Chem. Pet. Eng. **38**(1), 24–27(2002)
6. Betta, G.: An optical fiber-based technique for continuous-level sensing. IEEE Trans. Instrument. Meas. **44**(3), 686–689 (1995)
7. Lai, C.W., Lo, Y.L., Yur, J.P., Chuang, C.H.: Application of fiber Bragg grating level sensor and fabry-pérot pressure sensor to simultaneous measurement of liquid level and specific gravity. IEEE Sens. J. **12**(4), 827–831 (2012)
8. Maunder, A., Taheri, O., Fard, M.R.G., Mousavi, P.: Calibrated layer-stripping technique for level and permittivity measurement with UWB radar in metallic tanks. IEEE Trans. Microw. Theor. Techniq. **63**(7), 2322–2334 (2015)
9. Ricchiuti, A.L., Barrera, D., Urrutia, A., Goicoechea, J., Arregui, F.J., Sales, S.: Continuous liquid-level sensor based on a long-period grating and microwave photonics filtering techniques. IEEE Sens. J. **16**(6), 1652–1658 (2016)
10. Liu, Y., Li, Y., Yan, X., Li, W.: High refractive index liquid level measurement via coreless multimode fiber. IEEE Photon. Technol. Lett. **27**(20), 2111–2114 (2015)
11. Yu, C.: Liquid level measurement by using an image method. Signal Proc. 2320–2323 (2014)
12. Liu, B., Wang, D.Y., Wang, A.: Liquid-surface-level sensing based on transverse pulse train technique. IEEE Sens. J. **16**(8), 2317–2321 (2016)
13. Gong, C.S.A., Chiu, H.K., Huang, L.R., Lin, C.H., Hsu, Z.D., Tu, P.H.: Low-cost comb-electrode capacitive sensing device for liquid-level measurement. IEEE Sens. J. **16**(9), 2896–2897 (2016)
14. Guo, Z., Meng, Q., Zhou, M., Dong, X.: Optical fiber laser sensor for liquid level measurement based on multimode interference effect. Wirel. Opt. Commun. Conf. 1–4 (2016)
15. Zhang, H.X., Hou, Y.L., Feng, L.S., Su, S., Zhang, J.W., Liu, J., Liu, W.L., Liu, J., Xiong, J.J.: Polymer optical fiber continuous liquid level sensor for dynamic measurement. IEEE Sens. J. **15**(9), 5238–5242 (2015)
16. Yin, W., Peyton, A.J., Zysko, G., Denno, R.: Simultaneous noncontact measurement of water level and conductivity. IEEE Trans. Instrum. Meas. **57**(11), 2665–2669 (2008)

17. Ren, G.C., Guo, H.C., Yang, Y., Yang, Z.W., Tan, H.: Informatics in control, automation and robotics, LNEE 132, pp. 25–30. Springer, Berlin, Heidelberg (2011)
18. Cataldo, A., Tarricone, L., Vallone, M., Attivissimo, F., Trotta, A.: Uncertainty estimation in simultaneous measurements of levels and permittivities of liquids using TDR technique. IEEE Trans. Instrument. Meas. **57**(3), 454–466 (2008)
19. Santhosh, K.V., Roy, B.K.: Design of an adaptive calibration technique based on LSSVM for liquid level measurement. Emerg. Trends Appl. Comput. Sci. India (2013)
20. Sawhney, A.K.: A course in Electrical and Electronics Measurement and Instrumentation. Dhanpat Rai & Company (1997)

A Facile Synthesis of Graphene Oxide (GO) and Reduced Graphene Oxide (RGO) by Electrochemical Exfoliation of Battery Electrode

Rajdeep Vartak, Adarsh Rag, Shounak De and Somashekara Bhat

Abstract Graphene is a two-dimensional allotrope of carbon, in which carbon atoms are arranged in a hexagonal structure. All carbon atoms in graphene are sp^2 hybridized due to which three atoms are bonded to neighboring atoms and one electron is unbounded. Here, we have demonstrated a facile and low-cost technique to synthesize the graphene oxide and reduced graphene oxide from readily available battery graphite by means of an electrochemical method using liquid electrolyte. The graphite in the battery electrode has been expanded using acid for effective exfoliation. This is used to increase the interlayer spacing of graphite causes the van der Waals forces to cease. Graphene oxide and reduced graphene oxide (reduction using ascorbic acid) have been synthesized using this technique. The optical microscope image has shown few layer flakes and X-ray powder diffraction (XRD) demonstrated the reduction of oxygen content. The proposed setup is cheap and environmentally benign in nature.

Keywords Graphene oxide · Reduced graphene oxide · Electrochemical technique · XRD · Optical microscope

1 Introduction

GRAPHENE is a two-dimensional hexagonal network extracted from sp^2-hybridized Carbon. In short, it is a stack of graphene layers, which form graphite. The oxide version of graphene is graphene oxide. The graphene, graphene oxide and reduced graphene oxide has attracted the significant attention of people all over the world. Andre Geim and Konstantin S. Novoselov of University of Manchester in 2004 have done the pioneering work to yield the individual sheets of graphene [1]. These films have some interesting properties of high current density (10^6 Ω^{-1} cm^{-1}) [2], thermal conductivity (3000 W mK^{-1}) [3], visible transmittance (97.7%) [2], sur-

R. Vartak · A. Rag · S. De (✉) · S. Bhat
Department of Electronics and Communication Engineering, Manipal Institute of Technology,
Manipal Academy of Higher Education (MAHE), Manipal, India
e-mail: shounak.de@manipal.edu

face area (2630 m^2 g^{-1}) [4] and mechanical strength (Young's modulus >1 TPa) [5]. These promising properties have found applications in electronics, photonics, the field of transistors, solar cells, flexible displays as sensors or as a superconductive membrane in battery and several other fields [6–9]. Graphene oxide (GO) sheets are hydrophilic oxygenated graphene sheets containing oxygen functional groups on their basal planes and edges [10–12]. Graphene oxide (GO) has found prominence after discovery of graphene. GO is reduced to reduced graphene oxide (RGO). This resembles the graphene that contains some oxygen atom clung to the structure. Recently, GO has attracted much attention as a potential originator for the large-scale production of graphene-based materials leading to low cost of synthesis; and the tunability of the electronic properties as a semimetal, a semiconductor, or an insulator. These materials have also found the applications in a number of electronic devices [13, 14] and as a component in nanocomposite [15], energy storage [16], biomedical applications [17], catalyst [18], and as surfactant [19].

Typically, GO is prepared by top-down and bottom-up approaches. The top-down approach is mechanical exfoliations and chemical exfoliations [20]. Reliability of large area and homogeneity is larger issues in this technique. In the bottom-up technique, epitaxial growth [21] and chemical vapor deposition [22] is used. Defect generations and high-quality graphene synthesis over large area is a big concern.

Here, we propose to study the facile technique of electrochemical exfoliation of battery electrodes (graphite electrodes) for the synthesis of GO [23]. Generally, hydrazine Hydrate is used for reduction of GO; but hydrazine Hydrate is toxic and functionalizes the GO with nitrogen heteroatoms [24]. Electrochemical exfoliation is a greener process, comparatively cheaper, capable of mass production of few-layer graphene (FLG) by mere control of potential operations at ambient conditions of temperature and pressure.

2 Experimental Details

2.1 Chemicals Used

All chemicals used in this study were analytical reagent grade and used without further purification. Potassium hydroxide (KOH) pellets, Sodium hydroxide (NaOH) pellets, Acetone, and Isopropyl alcohol (IPA) have been brought from Merck. Deionized water (DI) (Millipore) of resistivity 18.2 MΩ-cm has been used. A pH meter (ELCO) was calibrated using pH 4 and pH 7 calibration buffer solutions. Battery electrodes have been extracted from the local battery.

(a)   **(b)**
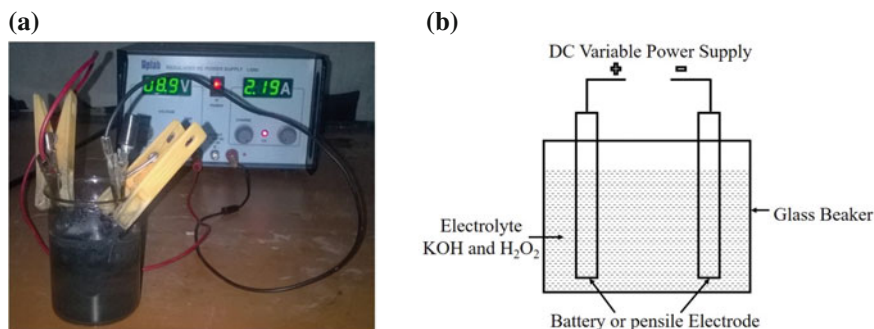
Fig. 1 **a** Schematic of Electrochemical experiment set up, **b** photograph of actual setup

2.2 Setup for Electrochemical Exfoliation

There are 3 main components in the setup as shown in Fig. 1, which are listed below along with their description.

Electrolyte: The electrolyte is an ionic species which will facilitate the flow of current through the setup and allow graphene sheets to be exfoliated from the electrode. The concentration of the electrolyte is varied to obtain different conditions for each iteration of the reaction. The electrolytes used in this study were Sodium hydroxide, Potassium hydroxide, Ammonium sulfate, and Hydrogen peroxide as shown in Table 1. The concentration is varied to investigate the effect on the product. Hydrogen peroxide is used to enable efficient exfoliation.

Electrodes: The electrodes are graphite rods from which graphene sheets are exfoliated. Anode and cathode are interchangeable, being identical in nature. It can be a battery electrode or pencil electrode. The electrode is obtained from dry cells.

Power Supply: This is the source which drives the reaction by the flow of current through the electrodes and electrolyte. It enables variation in voltage and polarity so that different conditions can be applied to the reaction. A regulated DC power supply is used. Power is supplied to the electrodes via crocodile clips. The electrode polarity is reversed from time to time. Voltage can be varied from 0 to 35 V.

Table 1 Specification of electrolyte used in this experiment

Sl. No.	Reagent	Chemical formula	Concentration (mol/L or M)
1	Potassium hydroxide	KOH	0.6 M
2	Sodium hydroxide	NaOH	0.6 M
3	Hydrogen peroxide	H_2O_2	65–60 mM
4	Ammonium sulfate	$(NH_4)_2SO_4$	0.1 M

Table 2 List of reaction conditions

Sample name	Electrolyte	Electrode	Voltage (V)	Time (min)
S1	KOH, 0.6M	Battery electrode	5	30
S2	KOH, 0.6M	Battery electrode	5	60
S3	KOH, 0.6M	Battery electrode	5, 8	30, 30
S4	NaOH, 3M	Battery electrode	1, 3	10, 10
S5	KOH, 0.6M	Graphite foil	5	30
S6	$(NH_4)_2SO_4$, 0.1M	Expanded battery electrode	10	10

2.3   Synthesis of Graphene Oxide (GO) Flakes Using Electrochemical Exfoliation

In a typical synthesis, the graphite electrodes (battery) were connected to the anode and cathodes. The electrolysis cell filled with 0.6 M potassium hydroxide (KOH) aqueous solution at room temperature. A constant potential of 3 V was applied to the anode with respect to the cathode for 1 min to start the process. Thereafter, 1.5 A current density through the graphite anode was maintained at about 10 min.

The polarity was reversed every two minutes to etch every electrode by the same amount throughout the experiment. GO layers were gradually exfoliated from the expanded graphite electrodes and dispersed in the aqueous solution. The process was carried out for 10 min. The solution with dispersed GO layers was subsequently taken from the electrolysis cell. This solution was centrifuged at 6000 rpm for 10 min to remove large agglomerates. The black thin layer could be seen on the top of the solution. This layer was transferred onto the glass substrate by immersing and pulling out slowly for further characterization. After this, top of the solution was then poured out. It was observed that the GO suspension is stable in nature. The experimental condition is enumerated in Table 2. The basic setup is shown in Fig. 2.

2.4   Filtration of GO Using Vacuum Filtration Pump

Vacuum filtration is the process of separating solid particles from the liquid. Figure 3 shows the vacuum filtration setup in which, this process is carried out with the help of filter flask, sample collector, cellulose acetate filter paper, vacuum pump, tubing and distilled water. The cellulose acetate filter paper was of 0.2 µm was brought from commercially available Sartorius brand. Initially sample collector with a vacuum suction hole was kept on the filter flask. After this, the tube was connected between the vacuum pump and the vacuum suction hole of the sample collector. Then, the filter paper was kept in the sample holder with the help of tweezer. It was rinsed with a distilled water to start the filtration process. The water was vacuumed down
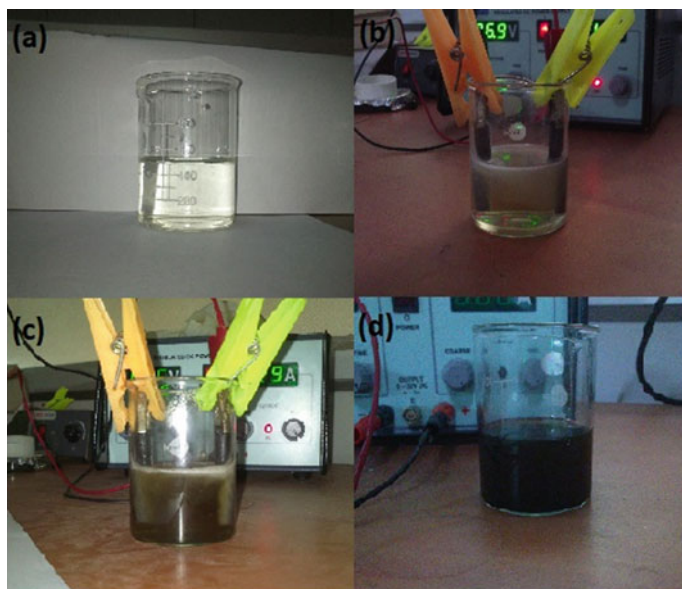
Fig. 2 Electrochemical exfoliation reaction at different stages. **a** Electrolyte solution; **b** electrodes are attached to power supply, **c** time evolution of the reaction; **d** final product

Fig. 3 Vacuum filtration unit at Department of ECE, MIT, Manipal. The filter flask and sample collector are in right side while the vacuum pump is in left. Cellulose acetate filter after filtration (inset)

through the membrane of the sample holder and filter paper was properly stuck in the sample holder. The electrolyte solution with sample particles then poured in the sample holder. Slowly, the liquid electrolyte was vacuumed into the filter flask. After 5 min, only the black particles were seen on the filter paper.

To remove the remaining electrolyte contents in the sample, it was rinsed with distilled water and vacuum suction was done again. In this way, washing of the sample was done for two–three times with distilled water and let the electrolyte to be separated from the sample. Some black particles were observed on the filter paper. The filter paper was then removed with the help of a tweezer.
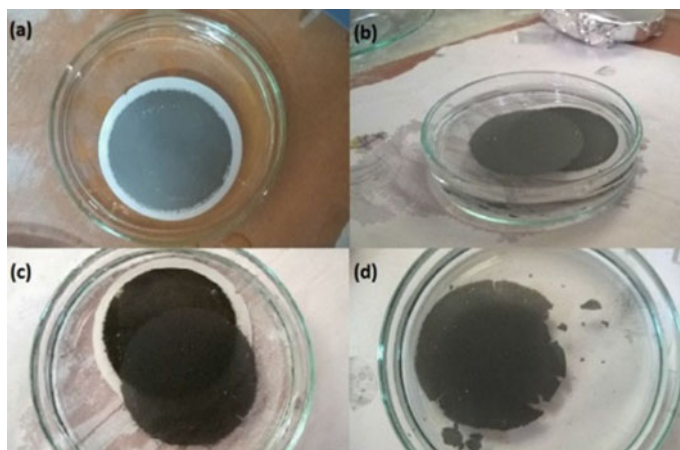
Fig. 4 Removal of graphene oxide sheet from cellulose acetate membrane. **a** Cellulose Acetate membrane with black precipitate floating on the water surface; **b**, **c** the separated thin film floating on water while the filter paper is submerged; **d** film floating after removal of the membrane

GO thin film on the filter paper has been kept on top of the water surface in a petri dish filled with DI water and the membrane was carefully placed on top of the water surface, making sure that no part of it submerges. The membrane was allowed to rest for 10 min, and then gently pushed down by applying force evenly from all directions. As a result, a layer of flakes was separated from the membrane and remained floating on the surface, while the cellulose acetate membrane sank to the bottom (Fig. 4). The floating layer was transferred to a glass substrate by immersing the substrate below the layer, and gently dragging it up [25]. The substrate was allowed to dry for 24 h under ambient conditions. The same result has been achieved by Thema et al. [26].

2.5 Synthesis of Reduced Graphene Oxide (RGO) Using Ascorbic Acid

After exfoliating GO from the electrochemical process, we reduced the obtained graphene oxide with Ascorbic Acid (AA) as reducing agent to lessen the oxygen contents from it. The dispersed GO was first heated in the oven to get it into a powder form for 3 h. The GO powder and 0.1 M AA was mixed with a volume ratio of 1:1. The mixture was then sonicated for 30 min. The reduction products were centrifuged at 14,000 rpm to separate big flacks. After this, excess 30% of H_2O_2 were added to the mixture to remove remaining AA by sonicating it for 30 min. The product was then again centrifuged at 14,000 rpm and washed three–four times with ethanol. It was then dried in the oven at 120° for half an hour. After cooling the final
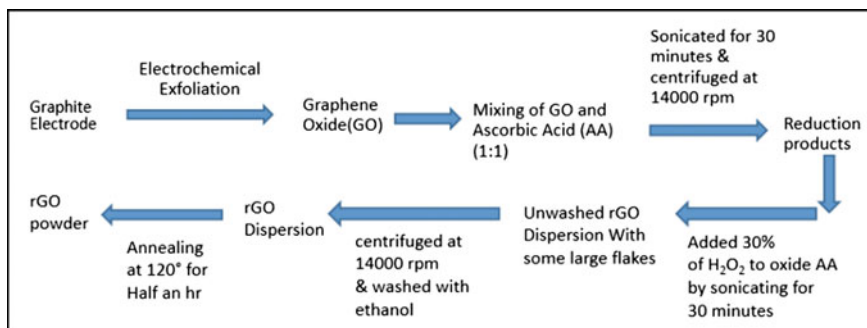
Fig. 5 Flow diagram for synthesizing RGO from GO

product, we went for X-ray diffraction which is the most promising characterization technique. The schematic of the process is shown in Fig. 5.

2.6 Material Characterization

The primary investigation of the morphology has been done using optical microscopy NIKON eclipse LV100 microscope. The images have been captured using Clemex Captiva software. The absorbance of sample solutions was detected by UV-Vis spectroscopy using a Shimadzu UV-1800 UV-visible Spectrophotometer. The aqueous suspensions of graphene oxide or reduced graphene oxide were used as the UV-Vis samples, and the pure water was used as a reference. The X-ray powder/thin film diffraction patterns have been recorded on a RIGAKU- Miniplex 600 diffractometer using Cu Kα radiation source ($\lambda = 0.154$ nm). The SEM measurement has been done on CARL ZEISS EVO MA18 series SEM. Electrical characterization has been done using a two-probe system with metal contact fabricated using thermal evaporation systems.

3 Results and Discussion

The dispersion obtained on the filter paper by electrochemical exfoliation can be visually observed to find information regarding the presence of GO. It has been observed that GO dispersions have a color ranging from yellow to dark black as shown in Fig. 6.

The samples obtained on the glass substrate were analyzed by optical microscopy. This enabled us to observe the size of flakes which were deposited. Transmission mode and reflectance mode were used. The analysis shows a lot of variation in size. The average size of flakes obtained in similar studies is in the range of a few microns
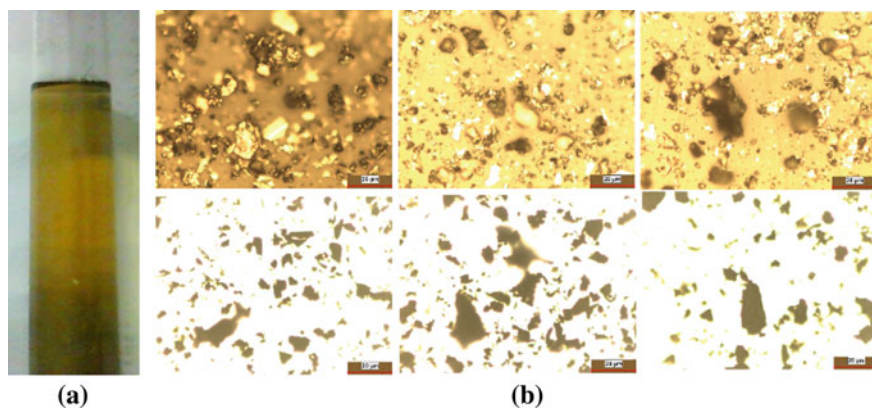
Fig. 6 **a** Reaction products of the electrochemical exfoliation technique in a test tube. **b** Optical microscopy of flakes showing various sizes for sample 15 showing a different area of exposure

Table 3 Obtained peaks from UV-Vis spectroscopy analysis

Sample name	Electrolyte	Electrode	Voltage (V)	Time (min)	Peak (nm)
S1	KOH, 0.6M	Battery electrode	5	30	224
S2	KOH, 0.6M	Battery electrode	5	60	227
S3	KOH, 0.6M	Battery electrode	5, 8	30, 30	227
S4	NaOH, 3M	Battery electrode	1, 3	10, 10	226
S5	KOH, 0.6M	Graphite foil	5	30	224
S6	$(NH_4)_2SO_4$, 0.1M	Expanded battery electrode	10	10	224

[8]. Here, we got flakes ranging from a few microns to tens of microns. It is shown in Fig. 6b. The obtained peaks for each sample are summarized in Table 3.

The peak positions as shown in Fig. 7 varies from 224 to 227 nm showing few-layer graphene oxide. Peak position for Sample S2 and S3 are near to 230 nm, which corresponds to $\pi-\pi^*$ Plasmon peak. The peaks are sharp in nature. The peak near to 226 nm indicates the presence of double-bond conjugations (C = C and C = O) exist in samples.

The GO solutions are reduced using ascorbic acid. Common practice is the use of hydrazine hydrate which is toxic and not environment friendly [24]. It seems ascorbic acid is one of the facile technique to reduce the GO suspensions. These samples have been characterized using X-ray diffraction (compositional characterization) as shown in Fig. 8. The broad peak around 24° shows the presence of the RGO which is near to 26.6° peak found in (002) graphite (d-spacing 3.35 Å at $2\theta = 26.61$) [27].

Fig. 7 UV spectra of GO thin films of sample S1–S6. Experimental conditions are listed in Table 3
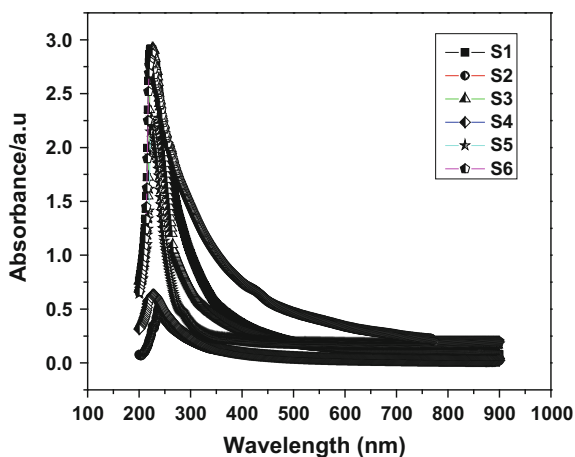
Fig. 8 XRD spectrum of RGO after reducing GO with ascorbic acid
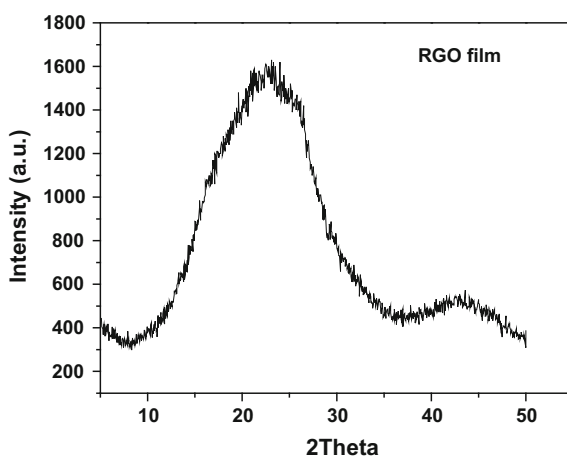
These results verify the reduction of GO to RGO which reveals that functional group containing the oxygen has been reduced.

Basic current–voltage measurement has been done in the coplanar configuration of the sample to check the electrical conductivity. This has been done using a Keithley Instruments 4200 SCS (Semiconductor Characterization System). RGO films have been deposited on glass and silicon substrates. Silver electrodes were deposited on them to enable connection of the probe. The metallization has been done using a Thermal evaporation. Base pressure is maintained at a range of 10^{-5} mbar. As it is evident from Fig. 9, contact is ohmic in nature and no rectification is found. One sample is found to have hysteresis because of some defects trap in the film. The maximum current is 1 mA which is well within order.

Fig. 9 I–V characteristics of RGO film
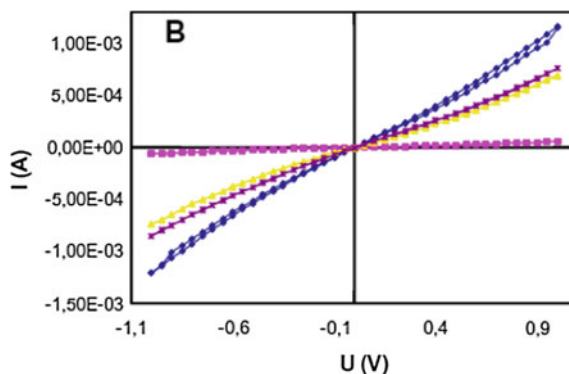
4 Conclusion

The graphene oxide and reduced graphene oxide has been synthesized using facile and low-cost technique. The use of environmentally benign reducing agent has led to the reduction of the oxygen content in the GO samples. The optical microscope shows the flake size of various samples. The best films have been found corresponding to optimal parameters of the electrochemical exfoliation of battery electrode. UV–visible spectroscopy elucidates the graphene oxide film with a peak at 227 nm showing π–π^* transition. The GO film has been reduced by ascorbic acid which is not toxic. XRD study reveals that peak formation at 24° near to graphite peak. The film shows good conductivity value showing no contact cut dropping.

References

1. Novoselov, K.S., Geim, A.K., Morozov, S.V., Jiang, D., Zhang, Y., Dubonos, S.V., Grigorieva, I.V., Firsov, A.A.: Electric field effect in atomically thin carbon films. Science **306**, 666–669 (2004)
2. Li, X.S., Zhu, Y.W., Cai, W.W., Borysiak, M., Han, B.Y., Chen, D., Piner, R.D., Colombo, L., Ruoff, R.S.: Transfer of large-area graphene films for high-performance transparent conductive electrodes. Nano Lett. **9**, 4359–4363 (2009)
3. Balandin, A.A.: Thermal properties of graphene and nanostructured carbon materials. Nat. Mater. **10**, 569–581 (2011)
4. Stoller, M.D., Park, S., Zhu, Y., An, J., Ruoff, R.S.: Graphene-based ultracapacitors. Nano Lett. **8**, 3498–3502 (2008)

5. Lee, C., Wei, X.D., Kysar, J.W., Hone, J.: Measurement of the elastic properties and intrinsic strength of monolayer graphene. Science **321** 385–388 (2008)
6. Dreyer, D.R., Ruoff, R.S., Bielawski, C.W.: From conception to realization: an historial account of graphene and some perspectives for its future. Angew. Chem. **49**, 9336–9344 (2010)
7. Yin, B., Liu, Q., Yang, L., Wu, X., Liu, Z., Hua, Y., Yin, S., Chen, Y.: Buffer layer of PEDOT:PSS/graphene composite for polymer solar cells. J. Nanosci. Nanotechnol. **10**, 1934–1938 (2010)
8. Zhao, T., Liu, Y., Li, T., Zhao, X.: Electrochemical performance of amorphous carbon nanotube as anode materials for lithium ion battery. J. Nanosci. Nanotechnol. **10**, 3873–3877 (2010)
9. Katkov, V.L., Osipov, V.A.: Energy distributions of field emitted electrons from carbon nanosheets: manifestation of the quantum size effect. J. Nanosci. Nanotechnol. **90**, 308 (2009)
10. Compton, O.C., Nguyen, S.T.: Graphene oxide, highly reduced graphene oxide, and graphene: versatile building blocks for carbon-based material. Small **6**, 711 (2010)
11. Mao, S., Pu, H., Chen, J.: Graphene oxide and its reduction: modeling and experimental progress. RSC Adv. **2**, 2643 (2012)
12. Dreyer, D.R., Park, S., Bielawski, C.W., Ruoff, R.S.: The chemistry of graphene oxide. Chem. Soc. Rev. **39**, 228 (2010)
13. Eda, G., Chhowalla, M.: Chemically derived graphene oxide: towards large-area thin-film electronics and optoelectronics. Adv. Mater. **22**, 2392 (2010)
14. Wan, X., Long, G., Huang, L., Chen, Y.: Graphene—a promising material for organic photovoltaic cells. Adv. Mater. **23**, 5342–5358 (2011)
15. Huang, X., Qi, X., Boey, F., Zhang, H.: Graphene-based composites. Chem. Soc. Rev. **41**, 666–686 (2012)
16. Lightcap, I., Kamat, P.V.: Graphitic design: prospects of graphene-based nanocomposites for solar energy conversion, storage, and sensing. Acc. Chem. Res. **46**, 2235–2243 (2013)
17. Sun, X., Liu, Z., Welsher, K., Robinson, J.T., Goodwin, A., Zaric, S., Dai, H.: Nano-graphene oxide for cellular imaging and drug delivery. Nano. Res. **1**, 203–212 (2008)
18. Pyun, J.: Graphene oxide as catalyst: application of carbon materials beyond nanotechnology. Angew. Chem. **50**, 46–48 (2011)
19. Kim, J., Cote, L.J., Franklin Kim, F., Yuan, W., Shull, K.R., Huang, J.: Graphene oxide sheets at interfaces. J. Am. Chem. Soc. **132**, 8180–8186 (2010)
20. Stankovich, S., Dikin, D.A., Piner, R.D., Kohlhaas, K.A., Kleinhammes, A., Jia, Y., Wu, Y., Nguyen, S.T., Ruoff, R.S.: Synthesis of graphene-based nanosheets via chemical reduction of exfoliated graphite oxide. Carbon **45**, 1558–1565 (2007)
21. Chen, H., Zhu, W., Zhang, Z.: Contrasting behavior of carbon nucleation in the initial stages of graphene epitaxial growth on stepped metal surfaces. Phys. Rev. Lett. **104**, 186101–186105 (2010)
22. Campos-Delgado, J., Kim, Y.A., Hayashi, T., Morelos-Gómez, A., Hofmann, M., Muramatsu, H., Endo, M., Terrones, H., Shull, R.D., Dresselhaus, M.S., Terrones, M.: Thermal stability studies of CVD-grown graphene nanoribbons: defect annealing and loop formation. Chem. Phys. Lett. **469**, 177–182 (2009)
23. Fernández-Merino, M.J., et al.: Vitamin C is an ideal substitute for hydrazine in the reduction of graphene oxide suspensions. J. Phys. Chem. C **114**, 6426–6432 (2010)
24. Shin, H.-J., et al.: Efficient reduction of graphite oxide by sodium borohydride and its effect on electrical conductance. Adv. Funct. Mater. **19**, 1987–1992 (2009)
25. Xueqiu, Y., Jong-Hyeon, C., et al.: An electrochemical route to graphene oxide **11**, 5965–5968 (2011)
26. Thema, F., Moloto, M., Dikio, E., Nyangiwe, N., Kotsedi, L., Maaza, M., Khenfouch, M.: Synthesis and characterization of graphene thin films by chemical reduction of exfoliated and intercalated graphite oxide. J. Chem. **2013**, 1–6 (2013)
27. Zhang, J., Yang, H., Shen, G., Cheng, P., Zhang, J., Guo, S.: Reduction of graphene oxide viaL-ascorbic acid. Chem. Commun. **46**, 1112–1114 (2010)

Smart Calibration Technique for Auto-ranging of LVDT Using Support Vector Machine

K. V. Santhosh and Preeti Mohanty

Abstract Design of a calibration circuit for linear variable differential transformer (LVDT) used in the measurement of thickness. The objective of the proposed work is to design a calibration technique, which is adaptive to variation in the range of measurement. Sensitivity in measurement is one of the important parameters, which is always expected to be higher for an ideal instrument. Sensitivity of an instrument is fixed during the process of calibration for an instrument and it depends on the minimum and maximum values of measurement. Whenever there exists a condition, involving only a part of measurement range the sensitivity remains constant, in general the sensitivity should have been increased. For varying the sensitivity, there will be a need to recalibrate the instrument which is time consuming and tedious. In the proposed work, a Support Vector Machine (SVM)-based learning algorithm is used in place of a conventional calibration circuit, which will calibrate automatically based on the specified range.

Keywords Adaptation · Calibration · LVDT · Smart · SVM

1 Introduction

Thickness analysis is one of the widely used parameters, followed to access the quality of a given material. Thickness measurement is also associated with quantitative measurement, like in case of Automated Teller Machine (ATM) the desired number of notes for the denomination specified is measured with the help of thickness mea-

K. V. Santhosh · P. Mohanty (✉)
Department of Instrumentation and Control Engineeirng, Manipal Institute of Technology,
Manipal Academy of Higher Education, Manipal 576104, India
e-mail: mohanty.preeti@gmail.com

K. V. Santhosh
e-mail: kv.santhu@gmail.com

surement. In either case, it becomes very essential for the measurement to be accurate. Measurement of thickness can be classified into techniques like online and offline. In online mode, the measurement is carried out without disturbance to the process functionality, and on the other hand offline measurement involves halting the process functionality for measurement. Most of the measurements are online in nature.

Different kinds of displacement sensors are available, a study is carried out to know the available measurement techniques like in [1], a self-sensing method is introduced to measure displacement based on charge amplifier. Paper [2] discusses a high-sensitivity Fiber Bragg Grating (FBG) displacement sensor for structural health monitoring. In [3], a non-contact measurement system is designed based on the principle of neural network function approximation, which can be used for high-precision measurement of linear travel in extreme environment. Paper [4], discusses interferometer-based optical fiber displacement sensor. A capacitive sensor that is suitable for measuring both linear and angular displacement of a shaft is discussed in [5]. In [6], a database of speckle patterns used for producing a low cost, non-contact, high-resolution displacement measurement is discussed. Paper [7] discusses the performance of a dual-wavelength technique devised to compensate power fluctuations in intensity-modulated plastic optical fiber sensors, which were specifically considered for the measurement of displacement in industrial and civil applications. Paper [8] discusses how an experimental device measures the displacement of an outer ring of a roller bearing subjected to two loads: a stationary load and a rotational load. Paper [9] discusses an experimental approach to measure linear displacement using RGB color coding algorithm. A fiber optic displacement measurement model based on the finite reflective plate is discussed in paper [10]. Paper [11] discusses a design of a heterogeneous integrated wireless sensor for displacement measurement. In [12], a high-resolution and wide range three-axis grating encoder was demonstrated. Paper [13] discusses its aim to link the measured displacement data to the quantification of the structural health condition, by validating the feasibility of simultaneous identification of structural stiffness and unknown excitation forces in the time domain using output-only vision-based displacement measurement. In [14], a lateral displacement measurement method based on fringing field capacitors is discussed. Ultrasonic sensor based on multimode interference in single-mode–multimode–single-mode optic fiber structure for non-contact displacement measurement is discussed in [15]. Paper [16] discusses the design and validation of an online intelligent displacement measurement technique with a linear variable differential transformer (LVDT) using Artificial Neural Network (ANN). Paper [17] discusses the circular grating Talbot interferometer to sense and measure the in-plane displacement. An apparatus capable of angular rotation measurements needed to calculate torque applied to a horizontal directional drilling drill-bit is discussed in [18]. In [19] a method for optimizing the performance of a real-time, long-term, and accurate acceleration-based displacement measurement technique, with no physical reference point is discussed. A combinatorial-code grating eddy-current sensor for the shortage of the low coarse

localization accuracy of phase-difference grating eddy-current sensor (PDGECS) is discussed in [20]. In [21] based on the time grating approach, the sensing mechanism of a capacitive nanometer sensor is discussed. Paper [22] discusses the characterization and implementation of a methodology for the measurement of large displacements using artificial vision techniques, for structural applications. In [23], the inherent error as well as the robustness of the phase unwrapping method is discussed.

From the study of reported work, it is clear that LVDT is one of the most widely used displacement sensors. LVDT is considered for displacement measurement because of its property to measure smaller dimension value, and linear characteristics. For accurate measurement of displacement, it is necessary to calibrate the output of LVDT. The process of calibration needs to be repeated every time measurement is carried on. In the proposed paper, the drawback of repeated calibration is avoided with the help of an adaptive calibration technique.

2　Linear Variable Differential Transformer

Linear variable differential transformer, commonly abbreviated as LVDT, consists of a transformer with a single primary winding and two secondary windings. As shown in Fig. 1, the windings are connected in series opposition. The translational displacement of the object to be measured is physically attached to the central iron core of the transformer. This is done so as to transfer all the motions of the body to the core [24, 25].
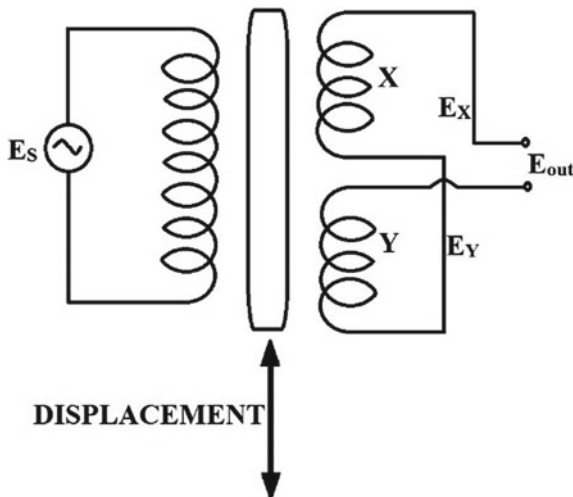
Fig. 1　Schematic of LVDT

The excitation voltage E_S is given by

$$E_S = E_P \sin \omega t \tag{1}$$

For an excitation voltage E_S, the voltage induced in the secondary windings E_X and E_y are given by

$$E_X = K_X \sin(\omega t - \varphi) \tag{2}$$
$$E_Y = K_Y \sin(\omega t - \varphi) \tag{3}$$

The parameters K_X and K_Y depend on the amount of coupling between the respective secondary and primary windings and therefore on the position of the iron core. When the core is in the center, K_X becomes equal to K_Y. Hence,

$$E_X = E_Y = K \sin(\omega t - \varphi) \tag{4}$$

As the secondary windings are connected in the series opposition mode,

$$E_{out} = E_X - E_Y \tag{5}$$

Hence with the core in the central position, $E_{out} = 0$. If the core is displaced toward winding X by a distance 'x', and if $K_X = K_1$ and $K_Y = K_2$ then,

$$E_{out} = (K_1 - K_2) \sin(\omega t - \varphi) \tag{6}$$

Alternatively E_{out} changes with the movement of the core towards the winding Y.

The output voltage of an LVDT is a linear function. The variation of output voltage with displacement at different positions of the core is shown in Fig. 2.

For a small displacement starting from the null point the characteristic curve is linear. Beyond a particular range of displacement, the curve starts changing and deviates from being a straight line.

Fig. 2 Characteristics of LVDT
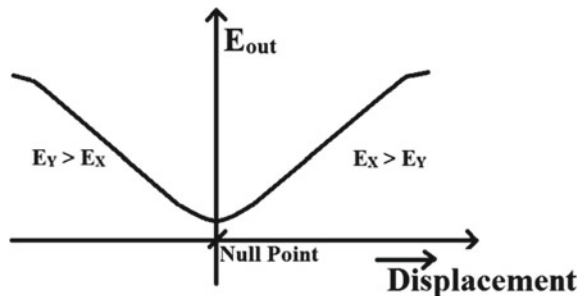
3   Problem Statement

In the proposed example, LVDT is used as a displacement sensor to measure the thickness of a particular object. The raw output acquired from LVDT is converted to DC voltage with the help of signal conditioning circuit as shown in Fig. 3. The signal conditioning circuit for LVDT consists of two parts, the first stage is the differential amplifier, followed by rectifier circuit. The differential amplifier stage consists of single stage opamp circuit with input from both the secondary windings. A DC voltage (+V) is also provided to nullify any offset in the signal. The single-sided signal obtained from the amplifier is rectified with the help of an active full wave rectifier. The output from the signal conditioner is given by

$$V_{out} = -\frac{R_2}{R_1}\left\{+V\left[\left(1 + \frac{R_B}{R_A}\right) + I_{(+)}\left(1 + \frac{R_B}{R_A}\right)\right] - I_{(-)}\left(\frac{R_B}{R_A}\right)\right\} \qquad (7)$$

In the proposed work, a setup as shown in Fig. 4 is used. It consists of a LVDT with a range of 0–20 mm. Signal from the above circuit is acquired to the system with the help of ELVIS. When the setup is subjected to measurement of thickness between 0 and 18 mm, it produced the characteristics as shown in Fig. 5. The output signal from the signal conditioner is calibrated to the range of 1–5 V (with the help of +V offset voltage) and gain resistance (R_2).

From the characteristics, it is seen that the output of the LVDT is almost linear. But now if the same device is used to measure thickness from 0 to 12 mm or 0 to 14 mm, etc., the output obtained would be between varying from 0.5 V to 3.6 V and 0.5 to 4.4 V, respectively. In both the cases, the sensitivity of measurement will reduce. On recalibration the results obtained are for 0 to 12 mm thickness measurement is shown in Fig. 6. Similarly, 0–14 mm in Fig. 7 and 0–10 mm in Fig. 8.

Fig. 3 Signal conditioning circuit for LVDT
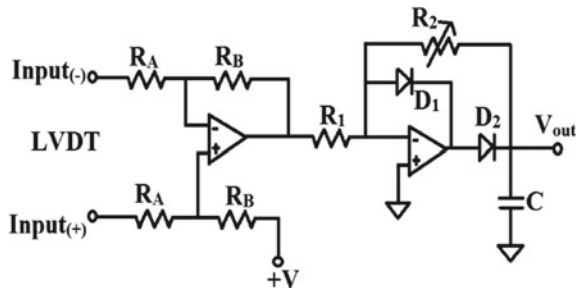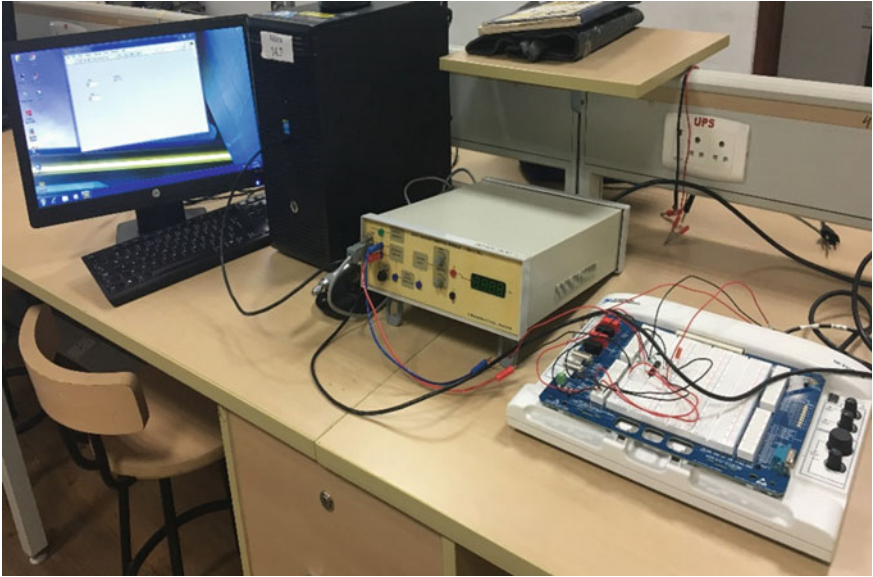
Fig. 4 Experimental setup
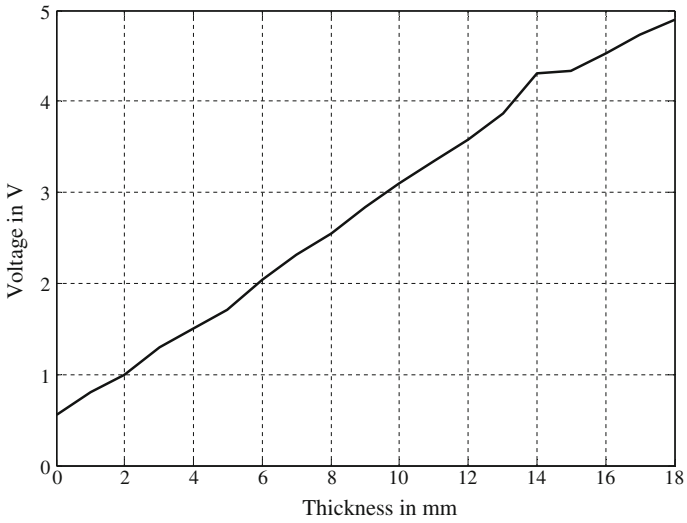
Fig. 5 Input output characteristics for measurement range of 0–18 mm
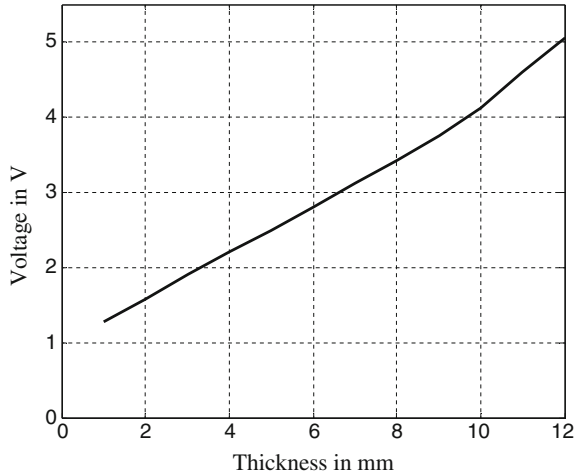
Fig. 6 Input output characteristics for measurement range of 0–12 mm
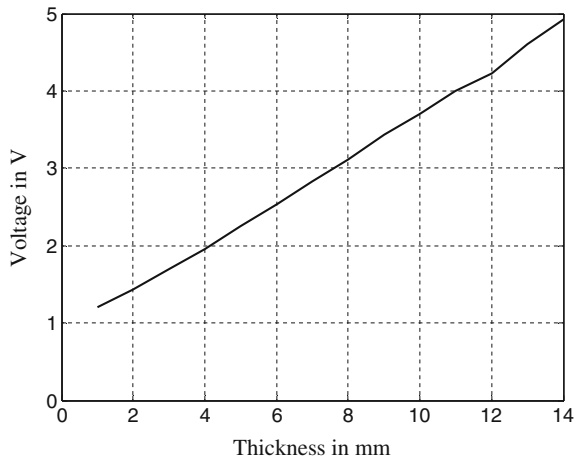
Fig. 7 Input output characteristics for measurement range of 0–14 mm

4 Problem Solution

From the discussion of available displacement measurement technique, it is seen that measurement system has a drawback of repeated calibration so as to achieve ideal sensitivity. If the range of measurement is varied the sensitivity of the measurement system varies as the output signal is always kept to be standard, either 1–5 V or 4–20 mA. In this section, a system is designed to make the measurement independent of range so as to obtain the highest sensitivity in measurement. Support vector machine is used to design the calibration circuit.
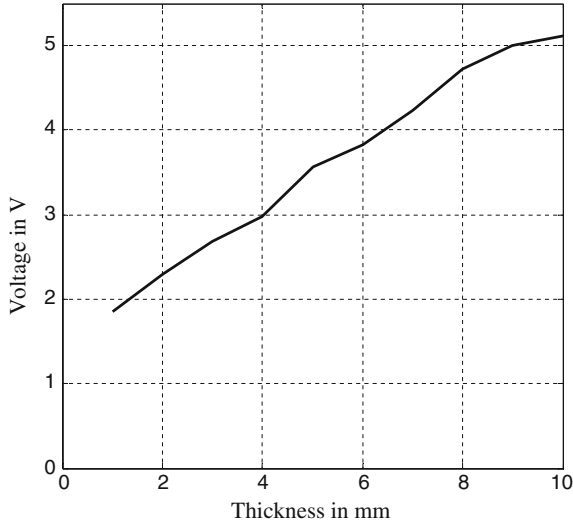
Fig. 8 Input output characteristics for measurement range of 0–10 mm

Support Vector Machine (SVM) is a supervised learning algorithm, which is used most commonly for classification and regression problems [26]. In the current work, we are solving the regression problem, the objective function is to train the output of the LVDT signal conversion circuit to project the target output, which is the desired output having maximum linearity and sensitivity. SVM regression analysis can be represented by

$$g(y) = f(x) + \mathcal{E} \tag{8}$$

where

g(y) Target vector
f   Predictor
$\mathcal{E}$ error/noise.

The objective of the training function would be to predict a new case for 'f' such the SVM was not presented earlier. Training is achieved by carrying out sequential optimization of the error function which is given by

$$\frac{1}{2}v^T v + K \sum_{j=1}^{N} \beta_j + K \sum_{j=1}^{N} \beta_j^* \tag{9}$$

Epsilon-SVM regression analysis is considered here, 'v' being the update vector, 'β' represents the data handling parameter, 'j' in the index varying from 1 to N training

classes, 'y' is the target class with 'x' being independent variable. For optimization of Eq. (9), a minimization function is considered which is given by

$$\left(v^T \partial(x_j) + b\right) - y_j \le \mu + \beta_j \tag{10}$$

$$y_j - \left(v^T \partial(x_j) + b_j\right) \le \mu + \beta_j^* \tag{11}$$

With '∂' being kernel function, 'μ' is threshold and is always assumed to be greater than or equal to 0. Linear kernel function is considered in the current problem.

To compute the target basic relation is considered, where the minimum and maximum range is provided by the user. Minimum range is substituted for 1 V and maximum range for 5 V and a linear relation between them is considered as a target. Training is carried over to map the output of LVDT acquired through data acquisition card of Elvis.

5 Results and Discussion

Once the signal captured from LVDT for various ranges are trained using the SVM-based adaptive calibration technique, it is subjected to test. For testing various ranges are considered, the user enters the minimum thickness and maximum thickness measured. The instrument auto adapts and calibrates by itself to produce a linear and highly sensitive output. Three different ranges are considered like test-1 having measurement range of 0–15 mm, test-2 having measurement range of 5–18 mm, test-3 having measurement range of 3–13 mm, and test-4 having measurement range of 10–18 mm, the test cases are considered different from those used for training the SVM.

Analysis of Table 1 and Fig. 9, gives the clear information about the displacement system performance. It is seen that the proposed displacement system was able to measure displacement accurately, producing linear characteristics and higher sensitivity without recalibration.

6 Conclusion

Displacement measurement being a widely used measurement phenomenon in many processes, accuracy and linearity are an essentially expected characteristics. Sensitivity is also equally looked parameter as an instrument. The sensitivity of an instrument is invariably a variable related to the range of measurement. In the reported work, a technique is designed which will be adaptive to the range of measurement and produce the output with the highest sensitivity and characteristics. An experiment

Table 1 Results obtained by the proposed technique

Actual thickness in mm	Output voltage in V	Sensitivity (V/mm)
Test-1		
0	0.989	0.268
2	1.51	
3	1.83	
4	2.11	
5	2.36	
7	2.94	
10	3.74	
12	4.31	
13	4.62	
15	5.02	
Test-2		
5	1.01	0.305
7	1.65	
9	2.28	
12	3.09	
13	3.38	
14	3.70	
15	4.10	
17	4.65	
18	4.98	
Test-3		
3	0.97	0.412
4	1.42	
6	2.19	
7	2.64	
8	3.12	
9	3.44	
10	3.83	
11	4.28	
13	5.09	
Test-4		
10	1.07	0.490
11	1.55	
13	2.46	
15	3.55	
16	4.10	
17	4.55	
18	4.99	
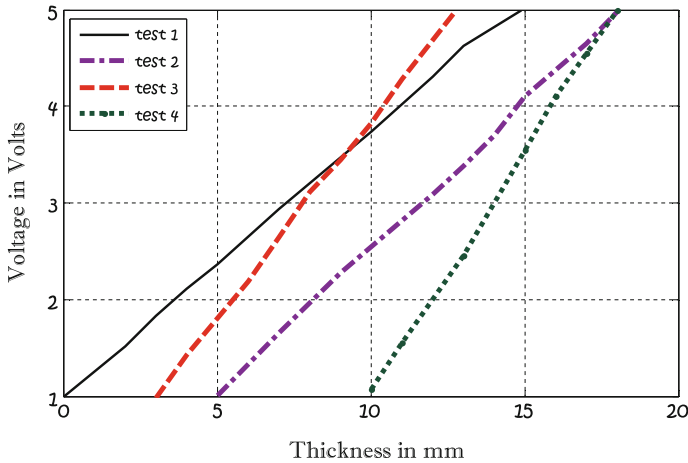
Fig. 9 Input output characteristics for the proposed system

was carried out on thickness measurement using LVDT, where the output of LVDT was processed using the reported technique and the results produced by the system shows that it was of higher linearity and sensitivity.

References

1. Du, Z., Zhang, T., Deng, L., Zhou, C., Cao, Z., Wang, S.: A charge-amplifier based self-sensing method for measurement of piezoelectric displacement. In: International Conference on Mechatronics and Automation (2017)
2. Li, T., Tan, Y., Chaoyang S., Guo, Y., Najdovski, Z., Hongliang R., Zude, Z..: A high-sensitivity fiber bragg grating displacement sensor based on transverse property of a tensioned optical fiber configuration and its dynamic performance improvement. Sens. J. **17**(18), 5840–5848 (2017)
3. Liu, R., Zhang, K.: A novel design of non-contact GMR displacement measurement system based on neural network. In: International Conference on Information and Automation (2017)
4. Zhu, C., Chen, Y., Du, Y., Zhuang, Y., Liu, F., Gerald, R.E., Huang, J.A.: Displacement sensor with centimeter dynamic range and submicrometer resolution based on an optical interferometer. Sens. J. **17**(17), 5523–5528 (2017)
5. Anandan, N., George, B.: A wide-range capacitive sensor for linear and angular displacement measurement. Indust. Electron. **64**(7), 5728–5737 (2017)
6. Farsad, M., Goch, G., Evans, C.: Application of correlation curve fitting to improve the absolute displacement measurement using speckle correlation. Precis. Eng. 129–134 (2016)
7. Vallan, A., Casalicchio, M.L., Olivero, M., Perrone, G.: Assessment of a dual-wavelength compensation technique for displacement sensors using plastic optical fibers. IEEE Trans. Instrument. Meas. **61**(5), 1377–1383 (2012)
8. Rasolofondraibe, L., Pottier, B., Marconnet, P., Chiementin, X.: Capacitive sensor device for measuring loads on bearings. Sens. J. **12**(6) (2012)
9. Murthy, A., Rao, S.S., Herbert, M.A., Karanth, N.: Experimental study on linear displacement measurement sensor using RGB color variation technique with PID controller. In: International conference on Computer, Communications and Electronics (2017)

10. Li, Y., Guan, K., Hu, Z.: Fiber optic displacement measurement model based on finite reflective surface. Opt. Laser Technol. **84**, 32–39 (2016)
11. Kisic, M.G., Blaz, N.V., Zivanov, L.D., Damnjanovic, M.S.: Heterogeneous integrated wireless displacement sensor. IEEE Trans. Mag. **53**(11), 4001–4104 (2017)
12. Jie, L., Jian, G., Feng, W., Tan, J.: High-resolution and wide range displacement measurement based on planar grating. Opt. Commun. **404**, 132–138 (2017)
13. Feng, D., Feng, M.Q.: Identification of structural stiffness and excitation forces in time domain using noncontact vision based displacement measurement. J. Sound Vib. **406**, 15–28 (2017)
14. Wenlian, W., Wenxia, W., Liu, Z.: Investigation of displacement measurement method based on fringing field capacitor. IET Sci. Meas. Technol. **11**(1), 63–66 (2017)
15. Sun, A., Wu, Z., Fang, D., Zhang, J., Wang, W.: Multimode interference-based fiber-optic ultrasonic sensor for non-contact displacement measurement. Sens. J. **16**(14), 5632–5635 (2016)
16. Santhosh, K.V., Roy, B.K.: Online implementation of an adaptive calibration technique for displacement measurement using LVDT. Appl. Soft Comput. **53**, 19–26 (2017)
17. Agarwal, S., Shakher, C.: In-plane displacement measurement by using circular grating Talbot interferometer. Opt. Lasers Eng. **75**, 63–71 (2015)
18. Gard, M.F.: Optical measurement of angular deformation and torque inside a working drillstring. IEEE Trans. Instrument. Meas. **65**(8), 1895–1901 (2016)
19. Abir, J., Longo, S., Morantz, P., Shore, P.: Optimized estimator for real-time dynamic displacement measurement using accelerometers. Mechatronics **39**, 1–11 (2016)
20. Weiwen, L., Hui, Z., Wei, T., Chunfeng, L.V.: Research on combinatorial-code grating Eddycurrent absolute-position sensor. IEEE Trans. Instrum. Meas. **61**(4), 1113–1124 (2012)
21. Peng, K., Liu, X., Chen, Z., Yu, Z., Pu, H.: Sensing mechanism and error analysis of a capacitive long range displacement nanometer sensor based on time grating. Sens. J. **17**(6), 1596–1607 (2017)
22. Franco, J.M., Mayag, B.M., Marulanda, J., Thomson, P.: Static and dynamic displacement measurements of structural elements using low cost RGB-D cameras. Eng. Struct. **153**, 97–105 (2017)
23. Bernal, O.D., Zabit, U., Bosch, T.: Study of laser feedback phase under self-mixing leading to improved phase unwrapping for vibration sensing. Sens. J. **13**(12), 4962–4971 (2013)
24. Morris, A.: Measurement and Instrumentation Principles. Butterworth Heinemann, Woburn (2001)
25. Sawhney, A.K.: Puneet Sawhney Mechanical Measurements and Instrumentation & Control. Dhanpat Rai & Co., Delhi (2008)
26. Cristianini, N., Shawe, J.: An Introduction to Support Vector Machines and other Kernel-Based Learning Methods, Cambridge University Press (2000)

Analysis of a Flow Process for Variation of Orifice Dimensions with Design of Adaptive Instrumentation

V. Sravani and K. V. Santhosh

Abstract Design of an instrumentation system for a flow measurement process using orifice as sensor is proposed in the present work. The objective of the proposed work is to analyze the behavior of orifice when its dimensions are varied, and propose a mechanism which will be able to produce accurate measurement even with those changes. Orifice being an element which is induced in the flow path whose flow rate is to be measured, undergoes a high level of shear stress, because of which there is always a possibility that the dimension would exhibit a change. Also, the wear and tear would be of high level, causing the replacement of orifice plate, and practically it is not possible to achieve a similar plate with 100% match in dimensions. These changes in dimensions will reflect in measurement errors. The proposed work tries to incorporate a calibration system which can adapt to the dynamics and will produce output independent of these changes in dynamics. For the design of calibration techniques, neural network algorithms are considered. The designed system is subjected to test and results show achievement in the proposed objective.

Keywords Flow process · Instrumentation · Neural network · Orifice

1   Introduction

Process industries hold a major portion of the existing industrial sector and contribute to a large extent to the economic growth. There are different measurements taking place in process industries to know the current status of the process as well as to maintain it at a certain desired set point. The different measurements which are taking place are pressure, temperature, flow, density, viscosity, pH, level, etc.

V. Sravani (✉) · K. V. Santhosh
Department of Instrumentation and Control Engineeirng, Manipal Institute of Technology,
Manipal Academy of Higher Education, Manipal 576104, India
e-mail: sravani.vemulapalli@gmail.com

K. V. Santhosh
e-mail: kv.santhu@gmail.com

Flow is one of the complex process variables measured in industries, there are many processes which can be maintained by controlling the flow rate like maintaining of temperature in a stirred tank heater by regulating a flow of a steam, removal of heat in an exothermic reaction of continuous stirred tank reactor by regulating the flow of the coolant, etc. The flow is also measured in industries like petroleum, pharmaceutical, power generation, food processing, etc. Hence, there is a need of accurate and precise measurement of flow, otherwise, it may lead to generation of defective products, revenue loss or sometimes physical damage to the system as well. There are various flow measuring devices available in market, selection of a particular device depends upon many factors like size, principle of operation, working environmental conditions, cost effectiveness, etc. One of the most popular flow measuring devices is orifice flow meter due to its ruggedness, simple construction, easy installation and replacement, no moving parts, high rangeability, applicability to all kinds of fluids, adaptable to extreme weather conditions, and cost-effective [1]. Orifice flow meter makes use of differential pressures, and it measured upstream and downstream of orifice plate to estimate the actual flow rate.

Many researchers have previously worked on orifice flowmeter either experimentally or through simulation. Due to hardware constraints, simulating the characteristics of the flow has become primary choice for the investigators. Computational fluid dynamics (CFD) is widely used by researchers to develop a simulation model to analyze the characteristics of flow across the orifice. In paper [2], the characteristics of pressure drop and flow variables downstream of orifice plate using CFD with beta ratio of 0.7 at Reynolds number = 75,000 have been studied. Numerical simulations to analyze the behavior of Newtonian and non-Newtonian fluids flowing through orifice with respect to Trouton ratio (extensional viscosity to shear viscosity) at low Reynolds numbers were carried out in paper [3]. The paper [4] has reported the effect of different beta ratios on discharge coefficient. They have experimentally found that sensitivity of discharge coefficient to swirl, decreases with increasing beta ratio. In paper [5], the authors have compared the pressure loss in standard orifice and perforated orifice in a square duct for measurement of gas flow rate. The pressure loss coefficient depends on free area ratio (total cross-sectional area within the orifices/cross-sectional area of the duct), thickness/diameter ratio, and Reynolds number. The effect of Reynolds number on the pressure loss coefficient is negligible at high Reynolds number. By increasing the plate thickness to 1.5 times, the orifice diameter, the pressure loss through an orifice plate can be reduced. The Computational Fluid Dynamics (CFD) simulations were done to obtain above stated results. The performance of slotted orifices with varying length and width of rectangular perforations and one slotted orifice with a circular perforation and a β ratio of 0.40 were simulated by the authors in paper [6]. The shape of the perforation has no effect on differential pressure but rectangular perforations showed better pressure recovery where length/width of perforations is 3. They also found that slotted orifice generates smaller permanent pressure loss when compared to the standard orifice with the same beta ratio. In paper [7], the authors have investigated the effect of orifice geometry on single- and two-phase pressure losses. Orifice is considered to be thick when its thickness-to-diameter ratio is greater than 0.5. The authors have reported that for
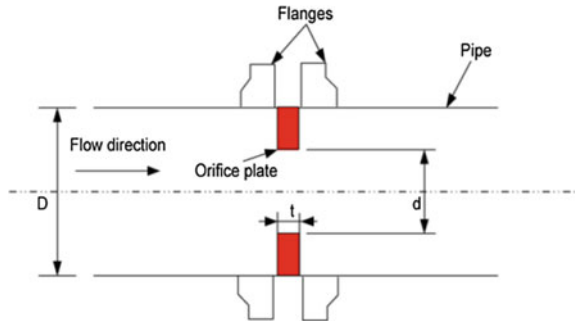
single-phase flow, vena contracta is outside the restriction in thin orifice, whereas in thick orifices, it is within the restriction. For a given thickness-to-diameter ratio, pressure drop increases with increasing Reynolds number and decreases with increasing thickness-to-diameter ratio under same Reynolds number in single-phase flow. In paper [8], the authors have reported that flow through clean and contaminated plate has difference in size of vena contracta and discharge coefficient. Orifice plates are contaminated by pipeline sludge, oil, grease, etc. Discharge coefficient increase in case of contaminated plates. Simulation results are obtained by sticking metal disks of known physical dimensions to show the effect of contamination. The authors have performed simulations with different beta ratio, contaminations, and tappings.

From the previous work, it is summarized in Table 1 that output of the orifice flow meter is effected by the various factors like beta ratio, thickness, Reynolds number, and working fluid.

2 Orifice Flow Meter

The orifice flow meter consists of a plate, with a circular or semicircular hole drilled in it, which is kept in a path of fluid flowing in a pipe as seen in Fig. 1. This restriction provided by the plate develops a differential pressure upstream and downstream of an orifice plate. Hence, the orifice flowmeter consists of an orifice plate as primary element and the transduction is performed by suitable pressure transmitters. The equation for flow rate for obstruction-type flowmeters can be derived using law of mass conservation, law of energy conservation, law of continuity, and Bernoulli's law.

$$Q_v = \frac{C_d}{\sqrt{1 - \beta^4}} \frac{\pi}{4} d^2 \sqrt{\frac{2(P_1 - P_2)}{\rho}} \tag{1}$$

where

Q_v volumetric flow rate

Table 1 Summary of previous work

Paper no.	Remarks
[9]	• Planar particle image velocimetry for determining effect of beta ratio on different regions of flow field **Working conditions**: Working fluid: water; Reynolds number: 25,000 Beta ratio: 0.41, 0.5, 0.62; Pipe diameter: 46 mm
[10]	• Thickness effect the location of vena contracta and pressure drop which also depends on Reynolds number **Working conditions**: Working fluid: water; Reynolds number: 25,000, 6000 Beta ratio: 0.06, 0.03, 0.015; Pipe diameter: 10 mm Thickness/diameter: 1.87, 6.93
[11]	• Single hole orifice has low discharge coefficient compared to multi-hole orifice • Manufacturing of multi-hole orifice is a tedious task **Working conditions**: Working fluid: water; Reynolds number: 500–20,000 Beta ratio: 0.5; Pipe diameter: 21.2 mm Thickness: 3.3 mm
[12]	• Reduction in pressure loss with increase in beta ratio • Location of vena contracta depends on Reynolds number and beta ratio. **Working conditions**: Working fluid: water; velocity: 0.376–28.087 m/s Beta ratio: 0.4065, 0.162, 0.081, 0.5; Pipe diameter: 12.3 mm Thickness: 2 mm
[13]	• Experimentally showed that discharge coefficient is dependent on Reynolds number and temperature **Working conditions**: Working fluid: water; Reynolds number: 440,000–3,600,000 Beta ratio: 0.742; Pipe diameter: 203.3 mm
[14]	• As temperature increases, viscosity decreases as well as the flow rate • The velocity profile depends on the temperature of the medium due to density, viscosity and installation effects **Working conditions**: Working fluid: water; Reynolds number: up to 1,000,000 Beta ratio: 0.5; Pipe diameter: 102.26 mm

C_d discharge coefficient
β beta ratio (d/D)
P_1 upstream pressure of the orifice plate
P_2 downstream pressure of the orifice plate
ρ density of the flowing fluid.

The above equation is valid for incompressible fluids and also, it assumed the elevation of orifice is constant throughout. The flow rate for compressible fluids can be calculated by simply multiplying the above equation by "ξ" which is expansion factor of compressible fluid. The mass flow rate is calculated by multiplying the above equation by density of the particular fluid under measurement. It can be noted

from the equation that the relation between pressure drop and flow rate is nonlinear, hence, there is a need of square root device after differential pressure transmitter. The flow rate through an orifice is effected by variations in various parameters like thickness of orifice plate, Reynold's number, viscosity, temperature, contaminations of plate, beta ratio, and multi-hole orifice plates [16].

3 Problem Statement

From the summary of the previous work, it is seen that output of the orifice flow meter is effected due to unknown changes in Reynolds number, thickness of the orifice plate, the beta ratio, and working fluid. In the present paper, analysis of orifice flow meter is carried by varying thickness of the plate and beta ratio at different Reynolds numbers using ANSYS 17.0. Further attempt is made to overcome these variations by designing a suitable observer. The work carried out is divided into two parts:

(1) Simulation of orifice flowmeter with required parameters using Computational Fluid Dynamics.
(2) Design a neural network based observer to estimate the output independent of beta ratio and plate thickness.

4 CFD Modeling

The steady-state flow through an orifice meter can be simulated by solving continuity and momentum equation at appropriate values. In the present work, standard k-e turbulence model is used due to its simplicity, better performance, and requires less computational power. In all the simulations, velocity is set at the inlet of the orifice meter, to vary the Reynolds number between 2000 and 15,000, and pressure is set at outlet and no-slip condition is set at wall. All the simulations are carried out in ANSYS 17.0. All the discretized equations were solved with SIMPLE algorithm. The solutions were converged at 10–5 residuals. The temperature was kept constant throughout the simulation.

4.1 Geometry

The orifice plate of different beta ratios 0.4, 0.5, and 0.6 are kept in a circular pipe of diameter 12.3 mm in accordance with paper [12] to validate the simulation model. The required upstream and downstream pipe length is provided for maintaining fully developed flow. The thickness-to-diameter ratio is varied in order to analyze its

Fig. 2 Orifice flow meter
with beta ratio$= 0.4$ at
different thickness values
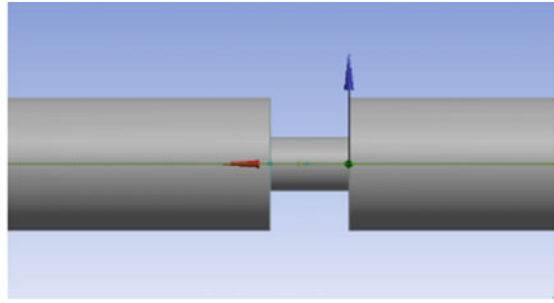
(a) Thickness=6mm
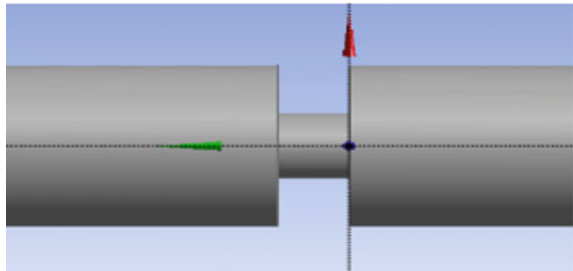
(b) Thickness=4mm
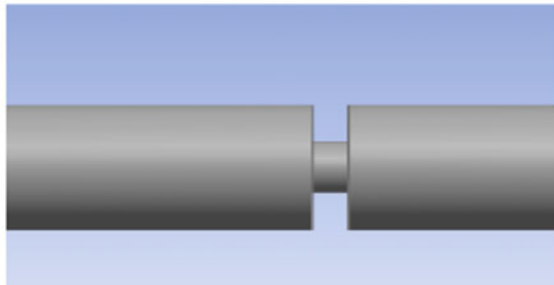
(c) Thickness=2mm

effect on pressure drop. The pressure tappings are taken at 1D and 1D, upstream and downstream of orifice, respectively. The different geometries with beta ratio$= 0.4$ is shown in Fig. 2 for different thickness values.

4.2 Computational Mesh

Mesh generation plays a vital role in CFD analysis for accurate results. Grid independence study was carried out by varying the element size of full body. There are six edges defined in total geometry with number of divisions equal to 50. Good quality
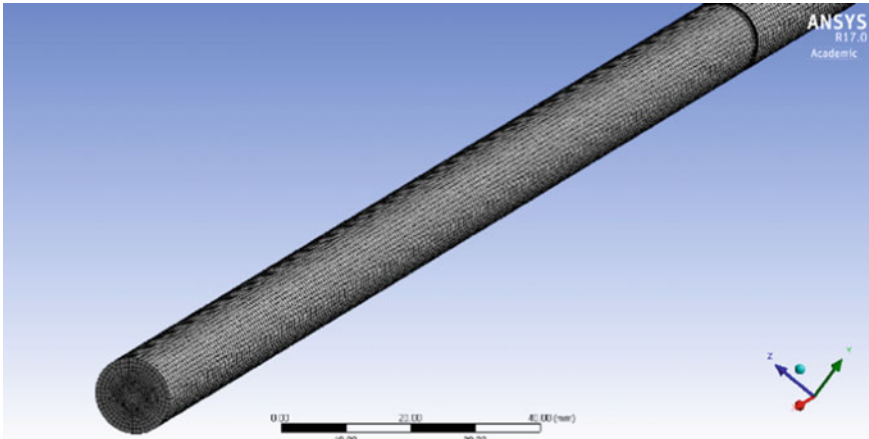
Fig. 3 Hexahedral mesh for beta ratio = 0.4, thickness = 2 mm, number of elements = 372,376
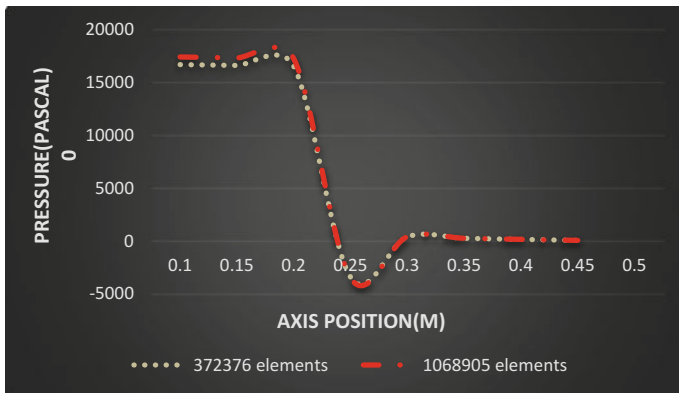
Fig. 4 Effect of grid size with pressure profiles at 372,376 and 1,068,905 elements

mesh has been ensured with orthogonal quality and aspect ratio at optimum value. Hexahedral mesh of small section of 12.3 mm pipe diameter is shown in Fig. 3.

In order to find the suitable grid size after which there is no considerable change in output, grid independence study was carried out for beta = 0.4 and thickness = 2 mm at 372,376 hexahedral elements and 1,068,905 elements. Figure 4 shows the pressure profile for two grids, and it is quite evident that there is no much difference in the output with respect to axial position. Hence, the further analysis was carried out using 372,376 hexahedral elements. These results were validated with the work of [12] for the same geometry. The same procedure was repeated to find the exact mesh size for different geometries.

Table 2 Details of geometry and mesh

Sl. no.	Beta ratio (β)	Thickness (mm)	Number of hexahedral elements
1	0.4	2	372,376
2		4	374,367
3		6	377,362
4	0.5	2	341,088
5		4	348,954
6		6	344,569
7	0.6	2	304,860
8		4	306,111
9		6	302,740

Table 2 gives the details about the number of hexahedral elements used for different geometries after performing grid independence study for a pipe upstream length of 246 mm and pipe diameter of 12.3 mm.

4.3 Analysis from CFD

Pressure profile of orifice flow meter at various beta ratios and thickness was analyzed using CFD simulation. One of the pressure profiles is shown in Fig. 5. From the CFD simulation, it is found that as beta ratio increases pressure drop decreases irrespective of thickness of orifice plate, which is shown in Figs. 6, 7, and 8. It is also seen that as thickness of the plate increases, pressure drop decreases irrespective of beta ratio. The pressure drop is also a function of Reynolds number. It is also found that at higher values of Reynolds number, there is significant decline in pressure drop as the thickness of the plate increases.

From the characteristics plot of orifice, it is evident that the pressure drop across the orifice plates depends on the dimension of the orifice plate-like hole to pipe diameter and thickness of plates. Though the effect of plate thickness is not predominant at lower Reynolds number, it has a considerable effect at higher Reynolds number, also as the thickness varies the characteristics tends to move towards nonlinearity. The effect of hole-to-pipe diameter ratio is very large. The objective of the proposed work will be to produce output which is independent of hole to pipe diameter and plate thickness.
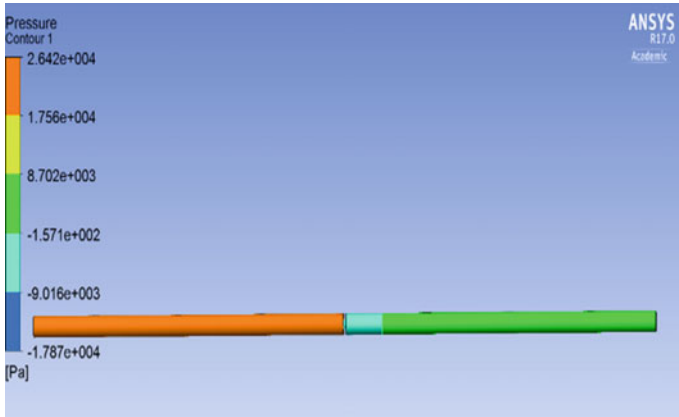
Fig. 5 Pressure contour at 13,800 Reynolds number for beta ratio = 0.4, thickness = 2 mm

Fig. 6 Differential pressure versus Reynolds number at beta ratio of 0.4 and varying thickness
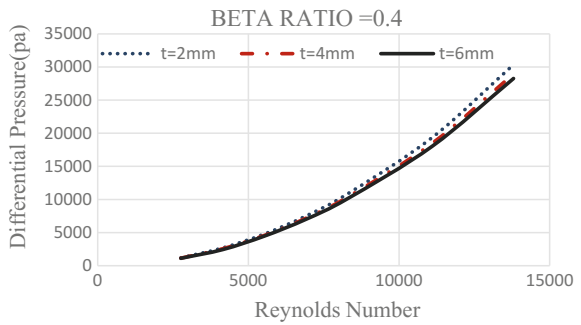
Fig. 7 Differential pressure versus Reynolds number at beta ratio of 0.5 and varying thickness
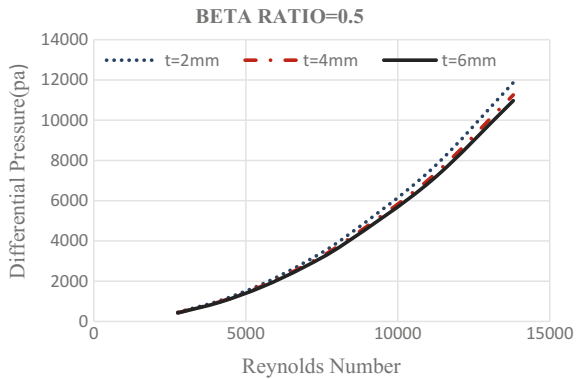
Fig. 8 Differential pressure versus Reynolds number at beta ratio of 0.6 and varying thickness
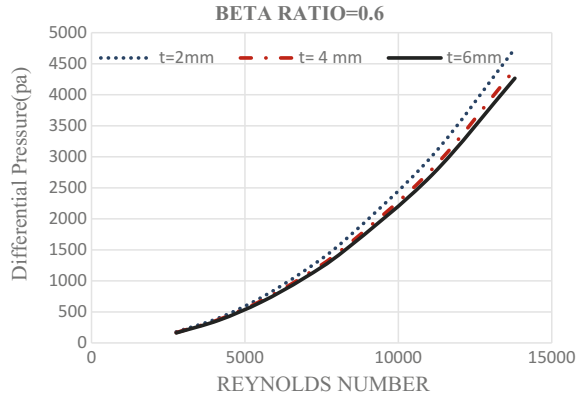
Fig. 9 Generalized block diagram of neural network model
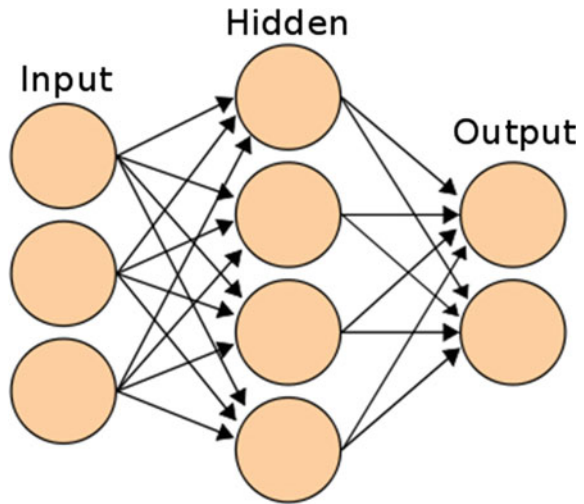
5   Problem Solution

To achieve the objectives mentioned in the earlier stage, a neural network algorithm is used. Neural network is a type of nonparametric model, using the concept of black box design. Neural network has three main components in it, first being the input matrix, second being the output or target matrix, and the final is the neural network block. Learning is the key functionality of neural network block, the block consists of perceptron's, which receive input, these inputs are operated on weighted functions and transfer function of neurons. Each of these neurons are interconnected across different layers. Figure 9 shows the generalized block diagram of neural network block.

Different structure, algorithms, and transfer functions of neural network are available in literature, in the proposed we use a back propagation based neural network structure with artificial bee colony algorithm.

Basic computation of neural network can be represented as

$$Y = X \cdot W \tag{2}$$

where Y is the target function, X being the input function, and W is weighted function of neural network block. Objective of the neural network training is to find out "W" such that when operated on "X" will yield "Y". For backpropagation scheme, the function is given by two paths. One is the forward path and other is the feedback path whose equations are shown in Eq. (3) and Eq. (4), respectively.

$$Y = S\left(\sum_{i=1}^{n+1} W_i \cdot X'\right) \tag{3}$$

$$E = X''\left(1 - X'\right)(Y - T) \tag{4}$$

Update function of artificial bee colony is given by

$$a^{t+1} = a^t + \delta * \left(a^t \cdot E - a \cdot \lambda\right) \tag{5}$$

where

a is the weights of neuron
δ is the weight update vector
E is the error vector
λ is the damping factor.

Neural network is trained to achieve the Regression "R" to be 1, in the current work we were able to achieve of R equal to 0.989 and Mean Square Error (MSE) of 0.324E−6, using two hidden layers each having seven neurons. For training, 120 sets of data are considered of which 70% is considered for training, and 15% each for validation and testing.

6 Results and Discussion

The trained neural network is tested for checking its performance. For testing, flow rate are measured for varying hole-to-pipe ratio and plate thickness, and the results obtained are shown in Table 3. In the proposed work, flow is with varying Reynolds number in the range 0–15,000, with the hole-to-pipe ratio in the range 0.3–0.7, and plate thickness ranging from 2 to 8 mm.

Table 3 Details of geometry and corresponding differential pressure

Reynolds number	Beta ratio (β)	Thickness (mm)	Differential pressure (Pa)
3000	0.7	2	6890
3500	0.6	3	8071
3500	0.3	8	8044
4000	0.5	4	9170
4500	0.4	8	10,480
5000	0.6	3	11,860
5300	0.5	7	12,260
6200	0.7	2	14,310
7300	0.4	4	16,700
7700	0.3	6	17,580
8200	0.6	3	18,810
8700	0.7	2	20,181
9100	0.3	7	21,010
9900	0.6	2	22,798
11,000	0.5	6	25,470
12,600	0.3	8	29,110
13,200	0.4	3	30,200
13,800	0.5	5	31,600
14,400	0.3	7	33,180
15,000	0.4	3	34,700

Fig. 10 Input and output characteristics of the measured system
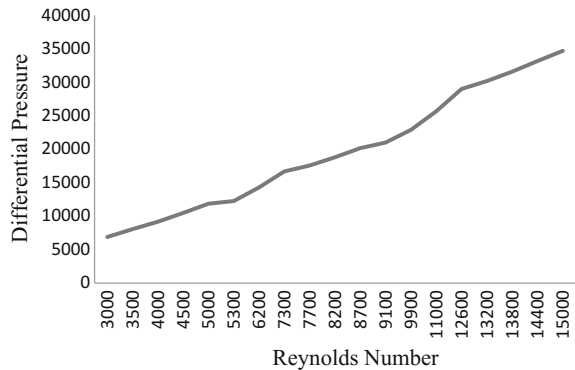
The input–output characteristics of the measured system are plotted in Fig. 10. From the understanding of the results shown in Table 1 and Fig. 10, it is clear that the proposed neural network algorithm is able to linearize the system, along with making the adaptive to plate thickness and hole-to-pipe ratio.

References

1. Husain, Z.D.: Theoretical Uncertainty of Orifice Flow Measurement. Daniel Measurement and Control White Papers. www.daniel.com (2010)
2. Erdal, A., Andersson, H.I.: Numerical aspects of flow computation through orifices. Flow Meas. Instrum. **8**, 27–37 (1997)
3. Muñoz-Díaz, E., Solorio-Ordaz, F.J., Ascanio, G.: A numerical study of an orifice flowmeter. Flow Meas. Instrum. **26**, 85–92 (2012)
4. Morrison, G.I., Hauglie, J., DeOtte Jr., R.E.: Beta ratio, axisymmetric flow distortion and swirl effects upon orifice flow meters. Flow Mem. Instrum. **6**, 207–216 (1995)
5. Gan, G., Riffat, S.B.: Pressure loss characteristics of orifice and perforated plates. Exp. Therm. Fluid Sci. **14**, 160–165 (1997)
6. Kumar, P., Bing, M.W.M.: A CFD study of low pressure wet gas metering using slotted orifice meters. Flow Meas. Instrum. **22**, 33–42 (2011)
7. Roul, M.K., Dash, S.K.: Single-phase and two-phase flow through thin and thick orifices in horizontal pipes. J. Fluids Eng. **134**(09), 091301 (2012)
8. Reader-Harris, M., Barton, N., Hodges, D.: The effect of contaminated orifice plates on the discharge coefficient. Flow Meas. Instrum. **25**, 2–7 (2012)
9. Shan, F., Liu, Z., Liu, W., Tsuji, Y.: Effects of the orifice to pipe diameter ratio on orifice flows. Chem. Eng. Sci. **152**, 497–506 (2016)
10. Cioncolini, A., Scenini, F., Duff, J.: Micro-orifice single-phase liquid flow: pressure drop measurements and prediction. Exp. Therm. Fluid Sci. **65**, 33–40 (2015)
11. Singh, V.K., Tharakan, T.J.: Numerical simulations for multi-hole orifice flow meters. Flow Meas. Instrum. **45**, 375–383 (2015)
12. Shah, M.S., Joshi, J.B., Kalsi, A.S., Prasad, C.S.R., Shukla, D.S.: Analysis of flow through an orifice meter: CFD simulation. Chem. Eng. Sci. **71**, 300–309 (2012)
13. Furuichia, N., Cordova, L., Lederer, T., Terao, Y.: Comparison of high temperature and high Reynolds number water flows between PTB and NMIJ. Flow Meas. Instrum. **52**, 157–162 (2016)
14. Buker, O., Lau, P., Tawackolian, K.: Reynolds number dependence of an orifice plate. Flow Meas. Instrum. **30**, 123–132 (2013)
15. Ntamba, B.N., Fester, V.: Pressure losses and limiting reynolds numbers for non-Newtonian fluids in short square-edged orifice plates. J. Fluids Eng. **134**, 091204 (2012)
16. Reader-Harris, M.: Orifice Plates and Venturi Tubes, Experimental Fluid Mechanics. Springer International Publishing Switzerland (2015). https://doi.org/10.1007/978-3-319-16880-7_2

Secure and Verifiable Outsourcing Algorithm for Large-Scale Matrix Multiplication on Public Cloud Server

Malay Kumar and Manu Vardhan

Abstract Cloud computing is poised to equip every computing node over the internet with the sheer computing power of data centers. It provides unlimited computing resources economically, conveniently, and ubiquitously in pay-per-use manner. It enables computationally weak client to execute large computations by outsourcing their computation load to the cloud servers. However, outsourcing of data and computation to the third-party cloud servers brings multifarious security and privacy challenges that needed to be understood and address before the development of outsourcing algorithm. The goal of this article is to show that the monomial matrix-based constructions are unable to protect the input and output privacy. Therefore, it becomes imperative to propose an improved algorithm to remove the shortcomings of the existing algorithm while retaining all the merits of previous algorithm.

1 Introduction

The outsourcing paradigm involves two entities, the client and the cloud server. The client wants to execute some complex polynomial time computation, but lacks the necessary resources required to complete the execution. Therefore, it outsources the large computations to the cloud servers, who have the necessary capability to complete the job [5]. The main security concern of this arrangement arises due to the lack of trust between the client and cloud server. The cloud server might record all the computation and data that came into its custody during the execution of the problem, and later behaves maliciously [8]. Due to the sensitivity of data, it became extremely

M. Kumar (✉) · M. Vardhan
Department of Computer Science and Engineering, National Institute
of Technology, Raipur, Chhattisgarh 492010, India
e-mail: mkumar.phd2014.cs@nitrr.ac.in

M. Vardhan
e-mail: mvardhan.cs@nitrr.ac.in

important to protect and preserve the privacy of data before outsourcing to the cloud. Moreover, the client also has no reliance on the computed result performed by the cloud. Additionally, the cloud also has financial advantages of executing a computation extremely fast, behave lazily (lazy computation), terminate a problem arbitrary, and produce incorrect results. Consequently, releasing up the valuable computing resources for other computations to increase the financial output [6]. That is why it has become extremely important to implement an efficient verifiable outsource computation, where the client can verify the integrity of result with minimum effort and interaction (noninteractive verifiable computation) [4].

The main contributions of the article are mentioned in the below points:

1. Many recent algorithms in the literature uses disguising transformations to hide the data from the cloud server but allows them to perform operation without seeing the secret data. These approaches are very appealing from the point of view of simplicity, efficiency, and flexibility, but we will demonstrate that these existing transformations have critical security flaw.
2. The paper discussed a novel way to perform the privacy-preserving matrix multiplication outsourcing to the cloud server. The privacy-preserving algorithm is very efficient and protects the confidentiality of the data set.
3. Later, the proposal will be analyzed and experimental evaluation proves the proposal.

The remainder section of this paper is organized into following sections: Sect. 2 discusses the security analysis of existing methods and mention the critical flaws. Section 3 presents an outline, framework, and design goals of secure outsourcing algorithm. Section 4, formulates the problem. Also, it proposed the privacy-preserving method for matrix multiplication problem. Section 5 covers the evaluation and result analysis. Finally, Sect. 6 have conclusion and future direction.

2 Privacy Analysis of Matrix-Based Transformation Operation

In this section, we have performed a security analysis for various likewise algorithms such as matrix multiplication [5], matrix inversion [9], and system of linear equation [1]. These algorithms are able to protect the information only if the data matrix is dense else, they are exposed to security vulnerabilities. The matrix-based transformation techniques transform a data matrix $M \in \mathbb{R}^{n \times n}$ into some disguised domain by multiplying randomized invertible matrices. Let the invertible matrices be P_1 and $P_2 \in \mathbb{R}^{n \times n}$, where P_1 and P_2 satisfies $\exists! P(i, j) \neq 0 \, \forall i \in [1, \ldots, n] \, or \, \forall j \in [1, \ldots, n]$. The transformation technique disguised the elements of data matrix, since the matrices P_1 and P_2 is sparse. Therefore, multiplication of such matrices, i.e., $M' = P_1 \times M \times P_2$ causes only $O(n^2)$ to the client. Efficiency is the main reason that this method of transformation is appealing in the various state-of-the-art algorithms. The privacy analysis of matrix-based transformation operation, which

illustrates that this method of transformation is not indistinguishable for secure and verifiable outsourcing algorithms under the ciphertext-only attack (COA).

The matrix-based transformation technique transforms a data matrix $M \in \mathbb{R}^{n \times n}$ into some disguised domain by multiplying randomized invertible matrix. Let the invertible matrices be P_1 and $P_2 \in \mathbb{R}^{n \times n}$, where P_1 and P_2 satisfies $\exists! M(i, j) \neq 0 \forall i \in [1, \ldots, n] \, or \, \forall j \in [1, \ldots, n]$. The transformation technique disguised the elements of data matrix, since the matrices P_1 and P_2 are sparse where the *nonzero* elements in $\{NZ_{P_1}, NZ_{P_2} \ll n\}$. Therefore, multiplication of such matrices, i.e., $\widehat{M} = P_1 \times M \times P_2$ only cost $O(n^2)$ to the client. Efficiency is the main reason that this method of transformation is appealing in the various state-of-the-art algorithm.

In this section, we perform the privacy analysis of matrix-based transformation operation, which illustrate that this method of transformation is not indistinguishable for secure and verifiable outsourcing algorithms under the ciphertext-only attack (COA). Let the data matrix $M \in \mathbb{R}^{n \times n}$ and the key matrix $P \in \mathbb{R}^{n \times n}$, where P satisfies $\exists! M(i, j) \neq 0 \forall i \in [1, \ldots, n] \, or \, \forall j \in [1, \ldots, n]$.

Theorem *The transformation operation, i.e., $M_L = P \times M$ and $M_R = M \times P$. Then, the number of zero entries in $M = M_L = M_R$.*

Proof Here, $M_L = P \times M$, where $P \exists! M(i, j) \neq 0 \forall i \in [1, \ldots, n] \, or \, \forall j \in [1, \ldots, n]$. The matrix $P = \alpha \times \delta_{\pi(i), j}$ α is a nonzero random number, δ is Kronecker delta function written as $\delta_{i,j} = 0 \, if \, i \neq j, \, 1 \, if \, i = j$, and π is permutation function. Therefore, the matrix P is an identity matrix pre-multiplied by a series of elementary matrices for row transformation R_i^{π} and multiplication R_{α_i} such as

Row Transformation

$$P = \alpha \times \delta_{\pi(i), j}$$

Here, $\delta_{i,j} = \begin{Bmatrix} 1 \, if \, i = j \\ 0 \, if \, i \neq j \end{Bmatrix} = I$

So,

$$P = R_1^{\pi} \times R_2^{\pi} \ldots \times R_n^{\pi} \times R_{\alpha_1} \times R_{\alpha_2} \ldots \times R_{\alpha_n} \times I \qquad (1)$$

Then $M_L = P \times M$,

$$M_L = (R_1^{\pi} \times R_2^{\pi} \ldots \times R_n^{\pi}) \times (R_{\alpha_1} \times R_{\alpha_2} \ldots \times R_{\alpha_n}) \times M \qquad (2)$$

Further, the product of matrix set $\{R_{\alpha_1}, R_{\alpha_2} \ldots, R_{\alpha_n}\}$ can be denoted as diagonal matrix

$$R_{\alpha}^{diag} = (R_{\alpha_1} \times R_{\alpha_2} \ldots \times R_{\alpha_n}) \qquad (3)$$

where the matrix $R_{\alpha}^{diag} = \{r_1, r_2, \ldots, r_n\}, \{0 \nexists r_i\}, \forall i \in \{1, 2, \ldots n\}$. Further, putting the value of Eq. 3, $M_L = (R_1^{\pi} \times R_2^{\pi} \ldots \times R_n^{\pi}) \times R_{\alpha}^{diag} \times M$, The matrix generated from

the multiplication of matrices $R_\alpha^{diag} \times M$ has same number of nonzero element as in matrix M. Further, the multiplication of elementary matrices $(R_1^\pi \times R_2^\pi \ldots \times R_n^\pi)$ only changes the row positions of the data matrix M. However, the nonzero elements in the matrix M_L will remains same as M. Hence, number of zeros entries elements in $M_L = M$.

Further, the proof is similar for the multiplication from the right-hand side for the transformation operation, i.e.,

Column Transformation

$$M_R = M \times P \tag{4}$$

Similar to the previous proof for M_L, the elementary matrices $R_\alpha^{diag} = (R_{\alpha_1} \times R_{\alpha_2} \ldots \times R_{\alpha_n})$ only scale the nonzero values of data matrix M. However, the multiplication by the elementary matrices $(C_1^\pi \times C_2^\pi \ldots \times C_n^\pi)$ performs column-wise traversal. Therefore, nonzero entries in matrix $M = M_R$. Hence, proved $M = M_R = M_L$. This proof shows that the transformation operations perform on the data matrix able to hide the nonzero entries but revels information about the zero entries. Therefore, these transformation techniques fail the equality for privacy. Gennaro et al., [4] states that,

The secure and verifiable algorithm (VC) is said to be secure, if for any adversary mathematical {A} running in probabilistic time, there exists a negligible function *negli*(.) such that

$$Adv_A^{Priv}(VC, f, \lambda) \le negli(.) \tag{5}$$

We have shown in the proof that the transformation schemes reveal information about the data-set. Therefore, such transformation schemes are fails to meet the privacy commitment stated in the Eq. (5). The secure algorithms matrix multiplication [5], matrix inversion [9], system of linear equation [3], regression analysis [3], and linear programming [7] are able to protect the information only if the data matrix is dense else, they disclose meaningful information. Therefore, an immediate attention is required for these state-of-the-art algorithms, which address their security concerns.

3 Problem Formulation

This section presents the proposed system model, security challenges, design goal, and the framework of the proposed algorithm.

3.1 Secure and Verifiable Outsourcing Algorithm Framework

Gennaro et al. [4] has given general definition of the secure and verifiable delegation of computation to the cloud server.

- $(pk, sk) \leftarrow \textbf{\textit{Setup}}(\lambda, f)$: The client C computes the key pairs (pk, sk) for the target function f on the basis of some randomized key generation algorithm. Further, these keys are utilized in the subsequent algorithms for the problem transformation and verification.
- $k \leftarrow \textbf{\textit{KeyGen}}(1^\lambda)$: On the input of security parameter λ, the randomized key generation algorithms generates the security keys. If $k = (\tau_x^{pv}, \tau_x^{sv})$, τ_x^{pv}, and τ_x^{sv} are the public values and the secret value of key k.
- $(\sigma_x, \tau_x) \leftarrow \textbf{\textit{ProbTrans}}_{SK}(x, k)$: The problem transformation algorithm uses the secret key SK to perform the disguise of the input x as a public value σ_x. This transformed value is outsourced to the cloud server and the secret value τ_x is kept secret with the client.
- $\sigma_y \leftarrow \textbf{\textit{Solve}}_{PK}(f, \sigma_x)$: The cloud server performs the desired computation of the function on the transformed value σ_x and produces the output σ_y, which is the disguised value of the real output of y, where $y = f(x)$.
- $y \cup \perp \leftarrow \textbf{\textit{Verify}}(\sigma_y, \tau_x^{sv})$: The client C executes the verification algorithm to check the integrity of result whether the server has been producing correct result. The verification algorithm outputs $y = f(x)$ if the algorithm found the result is correct, else $\perp$.

3.2 Threat Model

The threat model considered in this section closely follows the previous work on secure outsourcing in [3, 5, 9]. The cloud server might be "curious", "lazy" and dishonest simultaneously.

- **Semi-Trusted Model**: In "*honest but curious model*", the cloud server might record all the information such that input/output data and the computation primitives, which later is used to access significant information.
- **Untrusted Model**: The untrusted model, also known as malicious model, where the cloud server might be "curious", "lazy" and dishonest simultaneously.

3.3 Design Goals of Secure Algorithm

A secure and efficient outsourcing algorithm should satisfy the following four properties of correctness, privacy, efficiency, and verifiability simultaneously: **Correctness**: The client and the cloud server should follow the outsourcing algorithm instructions

correctly. **Privacy**: The MM outsourcing algorithm should able to maintain the privacy of input and the output. **Efficiency**: The computation cost for (transformation, verification, and retransformation) should be substantially lesser than executing the MM problem. **Verification**: The verification algorithm should detect cheating and server misbehavior of the cloud server. Simultaneously, it should be efficient and not involved in any expensive step.

4 Secure Outsourcing Algorithm for Matrix Multiplication

In this section, we present the proposed algorithm for secure outsourcing matrix multiplication algorithm.

4.1 *Mathematical Prelude*

- **Uniform Distribution**

    In probability theory and statistics, the uniformly distributed random variable set X is a family of random variable, and each element of the set is equally probable in a given interval. The uniform distribution is denoted as, $X \sim u(a, b)$ with $a \leq b$, where a and b are the limits of the interval over which the uniform distribution is defined.

- **Probability Density Function**

    Further, the probability density function of the uniform distribution is defined as

$$f_X(x) = \left\{ \begin{array}{ll} \frac{1}{b-a} & if \ a \leq x \leq b \\ 0 & otherwise \end{array} \right\} \tag{6}$$

The equation indicates that the values in the random set X is uniformly distributed with equal probability of $\frac{1}{b-a}$.

Where $= 2^p$ $(j > 0)$, p is a positive constant. The vector multiplication of two vectors U and V are resultant generations of key matrix P, which performs the transformation of data matrix.

$$P_{i,j} = u_i \times v_j^T \tag{7}$$

The vector multiplication of uniformly distributed random variable U and positive constant vector V generates uniformly distributed matrix. Therefore, the elements of matrix $P_{i,j}$ is also uniformly distribute with the probability density function defined as

$$f\left(P_{i,j}\right) = \begin{cases} \frac{1}{2K} & -K < P_{i,j} < K \\ 0 & otherwise \end{cases} \qquad (8)$$

where $P_{i,j} = u_i \times v_j^T$. Therefore, K is ranging between 2^{p+q} and 2^{p+q+r}.

4.2 Proposed Privacy-Preserving Matrix Transformation

Recalling the outsourcing algorithm discussed by [5]. The authors have performed privacy preserving by applying the monomial matrices. When analyzing this algorithm, we have found that this algorithm is unable to preserve the privacy of input/output pairs when the matrices are sparse.

After certain investigations, we have found that, if we perform the transformations by adding the random values to the input matrices. It will hide the original values of the input matrices and also the method is very efficient. The method is disused as follows:

$$M_1' = M_1 + P_1 \qquad (9)$$

$$M_2' = M_2 + P_2$$
$$R' = (M_1 + P_1) \times (M_2 + P_2)$$
$$R' = (M_1 \times M_2 + M_1 \times P_2 + P_1 \times M_2 + P_1 \times P_2)$$
$$T = M_1 \times P_2 + P_1 \times M_2 + P_1 \times P_2$$
$$R' = (M_1 \times M_2 + T) \qquad (10)$$

where M_1 and M_2 are input matrices such that $M_1, M_2 \in \mathbb{R}^{n \times n}$ and the key matrices $P_1, P_2 \in \mathbb{R}^{n \times n}$ the values of key matrices are uniformly distributed. However, this method transformed the input matrix M_1 and M_2 unconditionally and resulting transformation becomes a dense structure. This is against the motivation of outsourcing computation, since introduction of dense matrices required polynomial time computation and memory requirement. Therefore, in the pursuit of better solution, we have found that if we introduce some conditions, the process of transformation will maintain the sparsity of the input matrix. In this way, the proposal not only successfully maintains the privacy but also retains the input structure of input matrix that is dense matrix remains as dense and sparse will remain sparse matrix. Therefore, the client performs division of matrices in smaller parts and masks only those subpart, which has at least one nonzero element. In particular, the client divides the input matrix $M \in \mathbb{R}^{m \times n}$ into smaller matrices $\mathcal{M}_{\gamma,\delta} \in \mathbb{R}^{(\frac{m}{\alpha} \times \frac{n}{\beta})}$ for $\gamma \in [1, m/\alpha]$ and $\delta \in [1, n/\beta]$, where m/α and n/β are integers. As mentioned into the Eq. (9), the client perform the transformation is as follows: $M' = M + P$.

Where M input matrix $M \in \mathbb{R}^{m \times n}$ and P is secure key matrices as defined in the Eq. (7) $P_{i,j} = u_i \times v_j^T$. Further, the matrix P can be written as the summation of

sub-matrices which is divided into the order of $\gamma \in [1, m/\alpha]$ and $\delta \in [1, n/\beta]$. That is

$$P = \sum_{\gamma \in [1, \frac{m}{\alpha}], \delta \in [1, \frac{n}{\beta}]} p_{\gamma, \delta} \tag{11}$$

$$p_{\gamma, \delta} = u_\gamma \times v_\delta^T \tag{12}$$

Further, the data matrix M is divided into the sub-matrices $\mathcal{M}_{\gamma, \delta}$ such that $\{(\gamma, \delta | \ni \mathcal{M}_{i,j} \exists \mathcal{M}_{\gamma, \delta} : \mathcal{M}_{i,j} \neq 0\}$. Similarly, the key matrix P is also divided into the sub-matrices $p_{\gamma, \delta}$. And, the uniformly distributed number u_γ where $\gamma \in [1, m/\alpha]$ and positive vector constant v_δ $\delta \in [1, n/\beta]$ is generated as mentioned in Sect. 4.1. And, the other element of the key matrix P is padded with zero. Therefore, the final transformation of sub-matrix is $\mathcal{M}'_{\gamma, \delta} = \mathcal{M}_{\gamma, \delta} + p_{\gamma, \delta}$. The choice of α and β effects the dimension of the sub-matrices and therefore, the number of zero's are superimposed due to the transformation. However, this method still reveals some zero elements and their positions. Therefore, to hide this information, we randomly permute the rows and the column of the transformed matrix M'. To perform the permutation of rows and column, the following two functions are introduced.

The Cauchy's **two-line notation** is a notation which is used to describe a permutation on a finite set. Let S be a finite set and $\pi : S \to S$ be a permutation. In the two-line notations, the first row presents the list of elements of set S and the second row presents the list of permutations under the permutation function π.

$$\begin{pmatrix} 1 \dots n \\ \pi(1) \dots \pi(n) \end{pmatrix} \tag{13}$$

where $\pi(i)$ is the permutation function, $i = 1, \dots n$. π^{-1} is the inverse of permutation function π. An example presenting the Cauchy's two-line notation is illustrated as follows:

Let $S = \{1, 2, 3, 4, 5\}$ and the permutation function π is defined as $\pi(1) = 2$, $\pi(2) = 5, \pi(3) = 4, \pi(4) = 3, \pi(5) = 1$. The two-line notations for π is, $S = \begin{Bmatrix} 1, 2, 3, 4, 5 \\ 2, 5, 4, 3, 1 \end{Bmatrix}$, also known as transposition notation. Further, the Kronecker delta function is presented as,

Next, these two primitive functions are used to generate the security keys in the key generation algorithm.

$$\delta_{(x,y)} = \begin{Bmatrix} 1, & if \ x = y \\ 0, & if \ x \neq y \end{Bmatrix} \tag{14}$$

Using the Eqs. (13) and (14), the client generates two matrices which also have orthogonal property, i.e., $O \times O^T = 1$. Then, the client multiplies from L.H.S. for row-wise permutation, then R.H.S for column-wise permutation.

$$\widehat{M} = O_1 \times M' \times O_2 \tag{15}$$

As we know, the multiplication of orthogonal matrix, i.e., $O_1 \times O_1^T = 1$ and similarly $O_2 \times O_2^T = 1$. Applying the security mechanism for matrix multiplication operation as mentioned below

$$R' = (O_1(M_1 + P_1)O_2^T) \times (O_2(M_2 + P_2)O_3^T)$$
$$(O_1(M_1 + P_1) \times (M_2 + P_2)O_3^T \tag{16}$$

Multiplying the Eq. (5) with O_1^T from the left-hand side and O_3^T from the right-hand side will yield as

$$(O_1(M_1 + P_1) \times (M_2 + P_2)O_3^T$$
$$R' = (M_1 + P_1) \times (M_2 + P_2)$$
$$R' = (M_1 \times M_2 + M_1 \times P_2 + P_1 \times M_2 + P_1 \times P_2)$$
$$T = M_1 \times P_2 + P_1 \times M_2 + P_1 \times P_2$$
$$R' = (M_1 \times M_2 + T)$$

We are leaving the proof of correctness, security, verifiability, and efficiency for the proposed algorithm due to space constrained, that will be incorporated in the extended version.

5 Experiment Analysis

The client generates the uniform random number u and vector constant v^T in order to develop the random matrix P. The client generates the random sequence of u using some pseudo-random number generator which takes On for the generation of n numbers. The client performs the vector multiplication of $u \times v^T$ for the generation of matrix P which takes $O(n^2)$. Further, the client performs the transformation operation by summing the secret key matrix P with the input matrix M which takes another $O(n^2)$. Further, the matrices are randomizes row-wise and column-wise and such operation also takes $O(n^2)$. Therefore, effectively, the computation at the client side has $O(n^2)$ time complexity.

To conduct the experimental analysis, random instances of problems have been generated where the dimension of the problem $(n \times n)$ ranging from 1000 to 20,000, (n) is size of problem. The proposed algorithm will workout for any dimension of matrix, rectangular, or square until all multiplication conditions are met. The algorithms are started with performing multiplication on the various instances of problem. Although, this algorithm generates random outputs but the dimension of the output remains same as input. Therefore, when the matrix multiplication is performed on this transformed output, the time of execution would nearby be same as the non-encrypted problem. Further, for a resource-constrained client, memory is

Fig. 1 Comparison of
running time between
server-side computation and
client-side computation
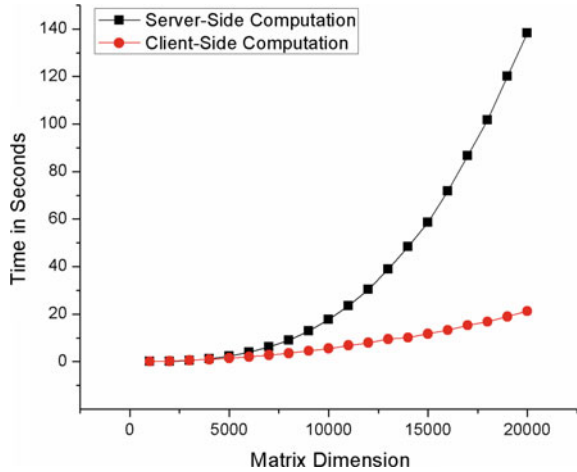
another bottleneck, but due to the special arrangement of key matrices, the algorithm required only sub-quadratic time complexity for I/O operations to the size of input. Therefore, the proposed algorithm provides ample opportunity for the client with limited computing resources to perform large complex computation. The same is visible in the comparison of client-side computation and cloud-side computation of the proposed matrix multiplication algorithm (Fig. 1).

Further, for the sake of completeness, we have compared the proposed algorithms with the previous algorithm [5]. It is worth noting here that the cloud-side computation complexity in both the algorithms is similar but the complexity difference on the client side makes the proposed algorithm more efficient in terms of execution time and memory requirements. Therefore, it is also visible in the experimental evaluation. The execution time at cloud side is almost similar, while at the performance degraded on the client system due to complex phases in the previous algorithm. Therefore, in the comparison, we are considering the running time at the client-side system. Figure 2 shows the comparison of running time of the algorithms, it is visible that the previous algorithm took more time than the proposed algorithm starting with smaller difference in the computation size, as the problem size increases the proposed algorithm perform better.

In the experiment, the largest size problem 20 K dimension matrix takes 21.11 s when outsources to the cloud server while 139.88 s when computed itself. In this way, the client archives an astonishing time saving 84.9% than the direct method and 23.91 times performance gain over the existing algorithm [5] . Finally, the performance of the algorithm depends on the underlying algorithm but as long as there is a computational gap between the client and cloud server of $O(n)$. The client is able to achieve the performance gain.

Fig. 2 Comparison of running time between proposed versus the existing algorithm
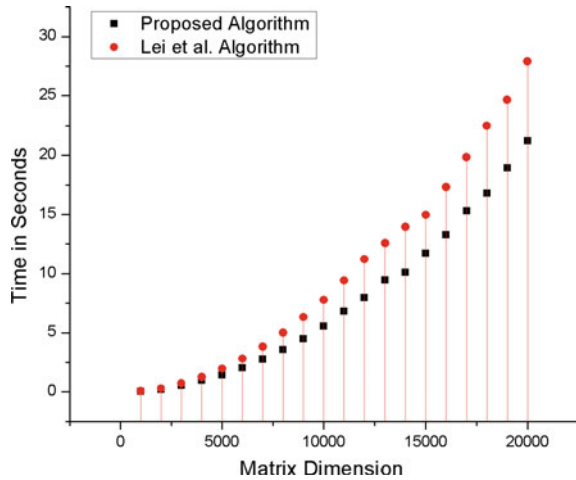
6 Conclusion

In this paper, we have proposed an improved algorithm for secure outsourcing of matrix multiplication problem. The previous algorithm is improved in two directions, first, the security and second, the efficiency. In the analytical analysis, we have found that the previous algorithm has marginalized security arrangement, if the input matrix is a sparse matrix. Later, we have improved the algorithm by imposing a stringent security arrangement in the proposal. Finally, the experimental analysis has been conducted and the results analysis demonstrates that the proposed work is meeting the algorithm design goals and proves the superiority over the previous algorithm.

References

1. Chen, F., Xiang, T., Lei, X., Chen, J.: Highly efficient linear regression outsourcing to a cloud. IEEE Trans. Cloud Comput. **2**(4), 499–508 (2014)
2. Chen, F., Xiang, T., Yang, Y.: Privacy-preserving and verifiable protocols for scientific computation outsourcing to the cloud. J. Parallel Distrib. Comput. **74**, 2141–2151 (2014). https://doi.org/10.1016/j.jpdc.2013.11.007
3. Chen, X., Huang, X., Li, J., Ma, J., Lou, W., Wong, D.S.: New algorithms for secure outsourcing of large-scale systems of linear equations. IEEE Trans. Inf. Forensics Sec. **10**(1), 69–78 (2015)
4. Gennaro, R., Gentry, C., Parno, B.: Non-interactive verifiable computing: outsourcing computation to untrusted workers. In: Lecture Notes Computer Science (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics), vol. 6223, pp. 465–482. LNCS (2010). https://doi.org/10.1007/978-3-642-14623-7_25
5. Lei, X., Liao, X., Huang, T., Heriniaina, F.: Achieving security, robust cheating resistance, and high-efficiency for outsourcing large matrix multiplication computation to a malicious cloud. Inf. Sci. (Ny) **280**, 205–217 (2014). https://doi.org/10.1016/j.ins.2014.05.014

6. Shan, Z., Ren, K.U.I., Blanton, M., Wang, C.: Pract. Sec. Comput. Outsourc. A Surv. **1**, 1–40 (2017)
7. Vaidya, J.: Privacy-preserving linear programming. In: Proceedings of the 2009 ACM Symposium on Applied Computing, pp. 2002–2007. ACM, March 2009
8. Wang, C., Ren, K., Wang, J.: Secure optimization computation outsourcing in cloud computing: a case study of linear programming. IEEE Trans. Comput. **65**, 216–229 (2016). https://doi.org/10.1109/TC.2015.2417542
9. Lei, X., Liao, X., Huang, T., et al.: Outsourcing large matrix inversion computation to a public cloud. IEEE Trans. Cloud Comput. **1**, 78–87 (2013). https://doi.org/10.1109/TCC.2013

Performance Optimization of Self-exited Piezoelectric Vibration Sensor

K. V. Santhosh and Noronha E'silva Nathan

Abstract Performance evaluation of self-exited piezoelectric vibration sensor with respect to dimensions, shape and material are carried out in the present paper, with an objective to find the design and dimension for which we get optimum performance in terms of frequency response and sensitivity. A new design is proposed based on the parameters discussed, for design COMSOL tool is used. Response for the designed vibration sensor for variation in frequency, with varying dimensions is performed for both differential and lateral loads. Results obtained give us information regarding the operating load, bandwidth, power, voltage and current ranges for best performance.

Keywords Frequency response · Optimization · Piezoelectric · Vibration

1 Introduction

Structural health monitoring is one of the key areas, which is considered to be of higher importance in the present day. Vibration analysis is widely used approach in structural monitoring. Detailed evaluation of sensor to match the objective of measurement is essential, along with it being a self-exited device which would not consume power. Self-excitation is essential as these sensors are placed in a random fashion and would be expected to be wireless device.

A study is carried out to understand the present works in the area mentioned. Many researchers have reported work in the area of power harvesting and power management, some of them have been discussed here. Energy harvesting from pulse waves of human artery is proposed in [1] by placing the piezoelectric structure on skin as patches. For harvesting energy, a design of coil is reported in [2] by placing the coil near equipment with high power-like transformers, overhead lines. In [3], a power managing system for a benthic microbial fuel cell with multi anode is reported.

K. V. Santhosh (✉) · N. E'silva Nathan
Department of Instrumentation and Control Engineering, Manipal Institute of Technology,
Manipal Academy of Higher Education, Manipal 576104, India
e-mail: kv.santhu@gmail.com

587

Focus of this paper is to address the loss of energy due to single anode in underwater sensors. In [4], four methods of electric circuits is used in wind energy harvesting system and the optimal selection of electric circuit to increase the efficiency of energy harvesting system is reported. Energy harvesting using piezoelectric system from the ocean is reported in [5]. This harvesting system is basically for low-frequency variations like ocean waves. Energy harvesting using a nano coil in measurement of masses small in size is reported in [6]. Energy is harvested from vibrations based on electromagnetism. In [7], most recent methods for harvesting kinetic energy are reported using piezoelectric and electromagnetic transduction. A comparison of both the methods is presented in this paper. Application of piezoelectric sensor as touch sensor is reported in [8], this paper also emphasizes on the advantages of using piezo sensors as touch sensors in terms of speed, self-excitation and durability. Room temperature detection using a self-powered humidity sensor is reported in [9], in this paper, excitation to the sensor is provided by zinc oxide nanosheets, and also it is stated that the sensitivity of sensor is more.

Many researchers have worked in the area vibration measurement using different techniques. Some of the vibration measurement sensors and techniques are reported here. Vibration measurement of gas turbine engine by non-contact method is reported in [10] using laser and scanning vibrometer. Measurement of vibration using position sensor is reported in [11]. Reported vibration sensing system measures 20–800 Hz range of vibration. In [12], vibration measurement of a rotating object through image-based tracking system is reported. A laser scanning vibrometer is used for getting the detailed vibration information for tracking the object. In [13], a scheme is proposed for measuring low-frequency vibration using a double frequency distributed Bragg reflector laser based on Doppler effect. Analysis and measurement of vibrational displacement by non-contact method using imaging techniques is reported in [14]. In [15], a method for nano-vibration measurement is reported that works based on semiconductor crystal photo electromotive force effect. Source used here is He–Ne laser and photo detector used is the crystal. Measurement of civil structure ambient Vibration using accelerometer is reported in [16]. Data collected was logged and also compared with the simulation model for analysis. Finding of the stiffness of sandy soil through measurement of soil surface vibration by analyzing the spectral changes of surface waves is reported in [17]. Measurement of humidity and vibration using a micro wire is reported in [18]. In this paper, relative humidity and vibration is measured based on interaction with the evanescent field of microwave. From the understanding of reported work, it is clear that vibration measurement using piezo devices is most widely used approach, due to its self-excitation capabilities, second, it is also seen that optimization in terms to structure and dimension is quite essential to attain best performance. In view of these, the proposed work tries to identify a novel model for vibration sensor, and there by analyzing its functionality to evaluate the performance.

2 Sensor Modelling

In this section, a detailed study of vibration sensor design is discussed. The main objective of the design is to achieve maximum strain for a given load. The design considered in the work has two cylindrical rings, adjacent to each other as shown in the Fig. 1. Two rings are held together with a rigid base structure, which is considered as a reference point grounded.

The smaller inner holes are rigidly fixed, and fixed to ground. Inside these two circles, a thin layer of silver electrodes of 0.025 µm thickness are placed. The two electrodes are shorted and they act as one plate of a capacitance. The dimension of the structure considered here is breadth X = 4.05 µm, height Y = 8.05 µm and thickness Z = 1.0 µm.

3 Methodology

Functionality of the sensor is analyzed by subjecting it with load as shown in Fig. 2. Structural deformation caused due to the load applied is shown in the Fig. 3. On analysis, it seem that the change in width (m) due to force acting is given as

Change in width

$$\nabla = 7.3 * 10^{-7} * F \tag{1}$$

Similarly, analysis is carried out for differential force and the deformation pattern for the load applied as shown in Fig. 4. This change in deformation can be related to

Fig. 1 Lateral image of proposed vibration sensor
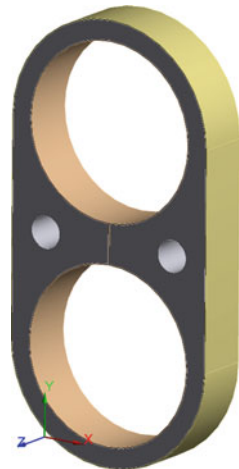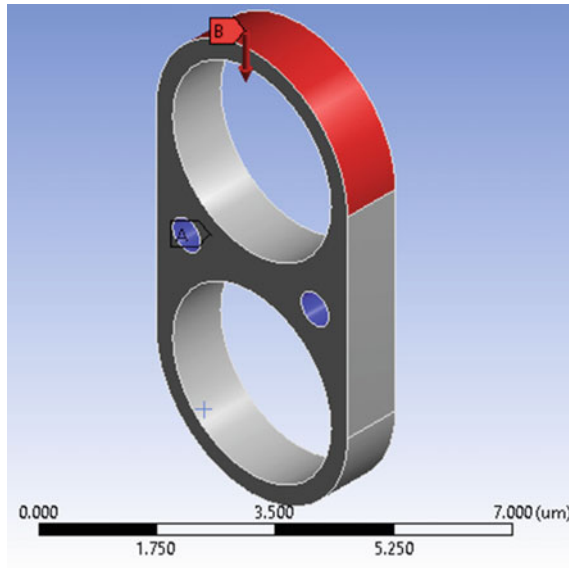
Fig. 2 Vibration sensor subjected with vertical load

analysis of change in capacitance and the mathematically analysis is derived in the next section.

Assuming that the structure is installed with silver electrodes one set being placed in the inner circle and the other on the outer circle, the device would behave like a capacitor. Deformation in sensor structure would induce the change in capacitance. The detailed derivation for capacitance is shown (Fig. 5).

Circumferential distance

$$v = \int_{r_1}^{r_2} E \cdot dA \tag{1}$$

$$E(2\pi Rl) = \frac{Q}{\varepsilon} \tag{2}$$

$$E = \frac{Q}{2\pi Rl\varepsilon} = \frac{\lambda l}{2\pi Rl\varepsilon} = \frac{\lambda}{2\pi R\varepsilon} \tag{3}$$

$$v = \int_{r_1}^{r_2} \frac{\lambda}{2\pi R\varepsilon} \cdot dR = \frac{\lambda}{2\pi\varepsilon} \int_{r_1}^{r_2} \frac{1}{R} \cdot dR$$

$$= \frac{\lambda}{2\pi\varepsilon} [\ln R]_{r_1}^{r_2} \ dR$$

$$= \frac{\lambda}{2\pi\varepsilon} \ln\left(\frac{r_2}{r_1}\right) \tag{4}$$

Capacitance is calculated by considering the two inner circles as two coaxial cables, and is given in terms of charge 'Q' and voltage 'V'.
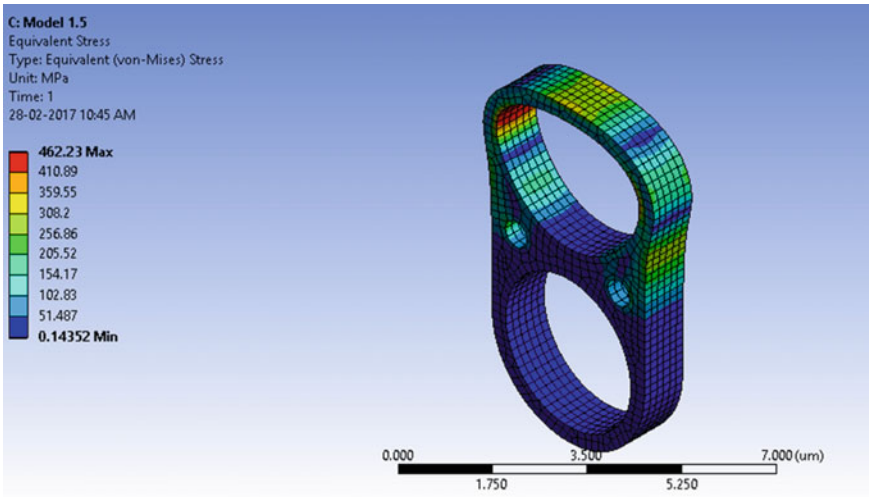
Fig. 3 Deformation analysis for singular load
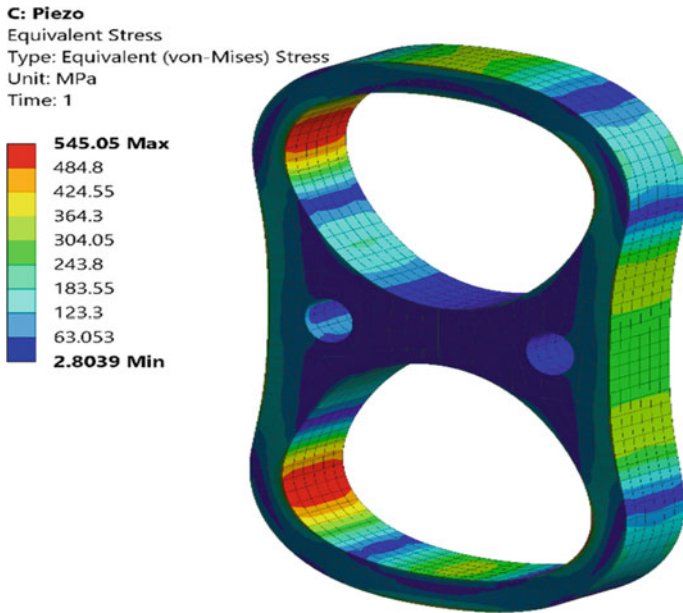
Fig. 4 Deformation analysis for differential load

Fig. 5 Detailed schematic
of vibration sensor
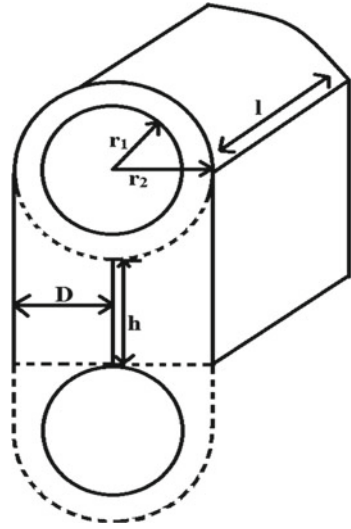
$$C_1 = \frac{Q}{V}$$

$$= \frac{Q}{\frac{\lambda}{2\pi\varepsilon}\ln\left(\frac{r_2}{r_1}\right)} = \frac{Q}{\frac{(Q/l)}{2\pi\varepsilon}\ln\left(\frac{r_2}{r_1}\right)}$$

$$C_1 = \frac{2\pi\varepsilon l}{\ln\left(\frac{r_2}{r_1}\right)} \tag{5}$$

Total capacitance

$$C = \left[\frac{2\pi\varepsilon l}{\ln\left(\frac{r_2}{r_1}\right)}\right] \times 2 + \frac{\varepsilon A}{D}$$

$$= \frac{4\pi\varepsilon l}{\ln\left(\frac{r_2}{r_1}\right)} + \frac{\varepsilon h l}{D} \tag{6}$$

$$= \varepsilon l\left[\frac{4\pi}{\ln\left(\frac{r_2}{r_1}\right)} + \frac{h}{D}\right] \tag{7}$$

Fig. 6 Physical model representation of sensing system
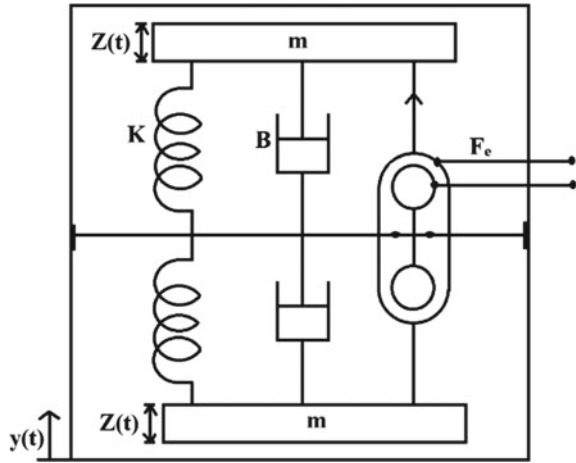
From the above relations it is clear that changes in stain/load applied on the sensing body would cause changes in capacitance formed. The next stage of would be to analyze the frequency response of the sensor. For analysis a Williams and Yates [19] model is considered as shown in Fig. 6. The model is derived such that it is self-exited.

Applied force,

$$y(t) = Y \sin(\omega t) \tag{8}$$

Reactive force,

$$z(t) = Z \sin(\omega t + \varphi) \tag{9}$$

$$F_e = B_e \dot{z} \tag{10}$$

where

B   mechanical clamping due to friction
F_e Restoring Force.

Natural frequency of the system can be given by

$$\omega_n = \sqrt{\frac{K}{m}} \tag{11}$$

$$S_d = \frac{B}{2m\omega_n} \tag{12}$$

$$S_e = \frac{B_e}{2m\omega_n} \tag{13}$$

From the above relation, the equation for generated power can be given by

$$P(w) = \frac{m\left(\frac{\omega}{\omega_n}\right)^3 \omega^3 Y^2 S_e \times 2}{\left[1 - \left(\frac{\omega}{\omega_n}\right)^2\right]^2 + \left[2(S_d + S_e)\frac{\omega}{\omega_n}\right]^2} \tag{14}$$

For deriving the maximum condition, it is expected that, $\omega = \omega_n$ and $S_e = S_d$

$$P_{max} = \left[\frac{m\omega_n^3 Y^2}{16 S_d}\right] \times 2 = \frac{m\omega_n^3 Y^2}{8 S_d} \tag{15}$$

For the desired function, a set of functionalities are considered:

- The mass should be as large as possible within the available volume of the device.
- The maximum displacement of the mass should be as large as possible in the space available.
- The spring should be designed, so that the resonant frequency of the device matches the vibration frequency of the application.
- The electrical load impedance should be chosen to give a low enough damping factor to increase the movement of the mass.

4   Results and Discussion

Considering the above discussed methodology, a piezoelectric vibration sensor is designed. To evaluate its performance a set of tests are conducted. Preliminarily, it is essential to consider a set of values, inside of the casing is coated by PTFE, so that the coefficient of friction is under 0.2. Zirconium dioxide is used as accelerating mass as it is higher in density in comparison to silicon. Also, a mass of 5.68e−14 Kg is added at one of the ends. Test is conducted for varying frequency of vibration for different magnitude of load. The response achieved is shown in Figs. 7, 8, 9 and 10.

From the responses as seen in Figs. 7, 8, 9 and 10, it is clear that in the reported work, a detailed analysis of vibration sensors was carried out starting from the design. The characteristics graph shows that the bandwidth attained is similar and is independent of load magnitude.
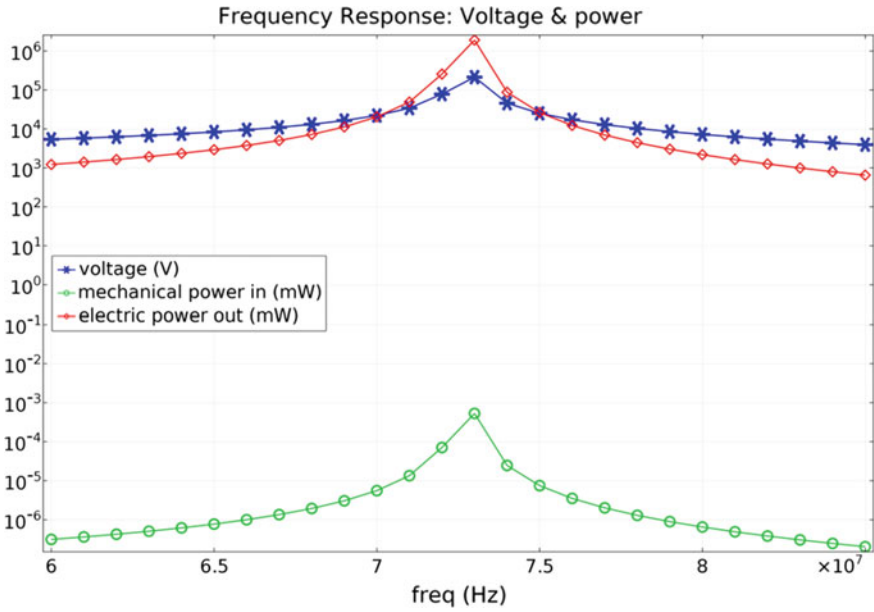
Fig. 7 Response for the load of 1 N
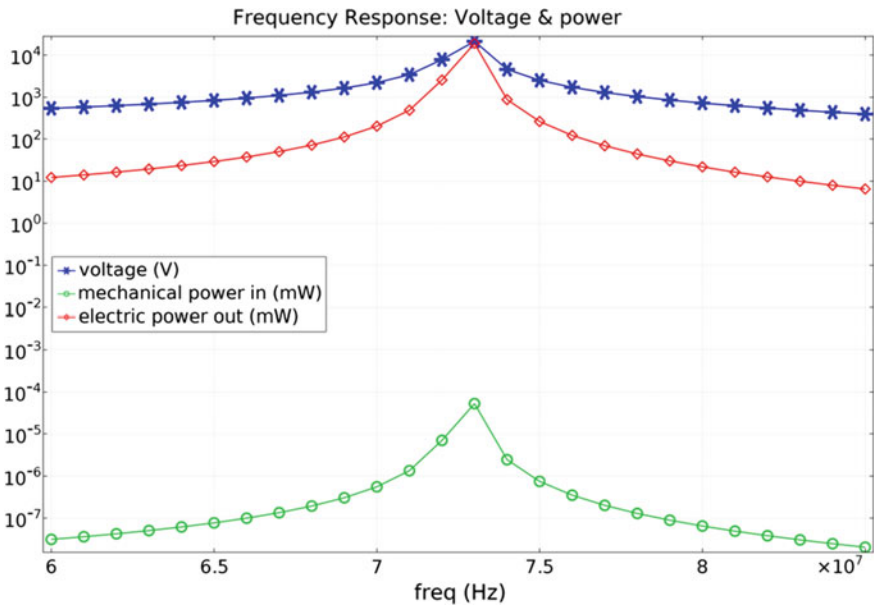
Fig. 8 Response for the load of 0.1 N
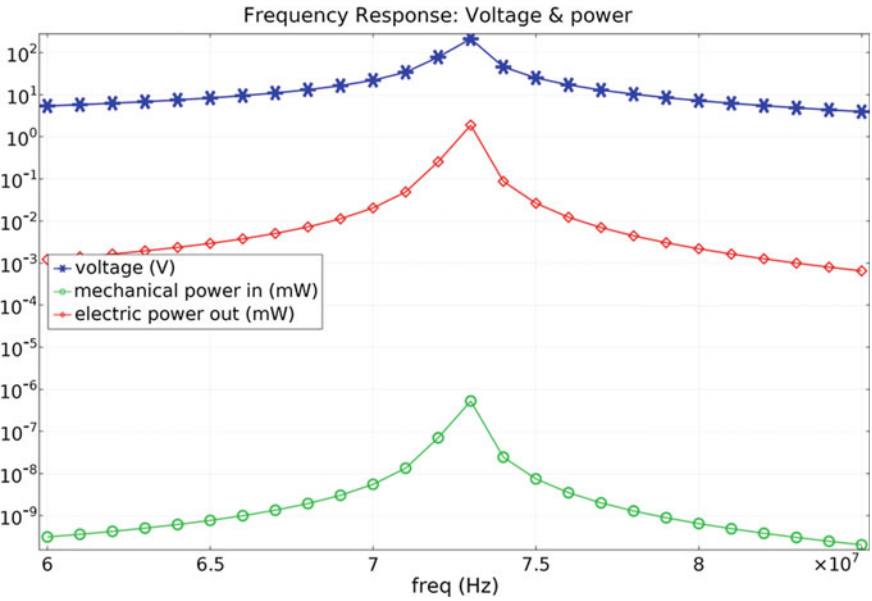
Fig. 9 Response for the load of 1 mN
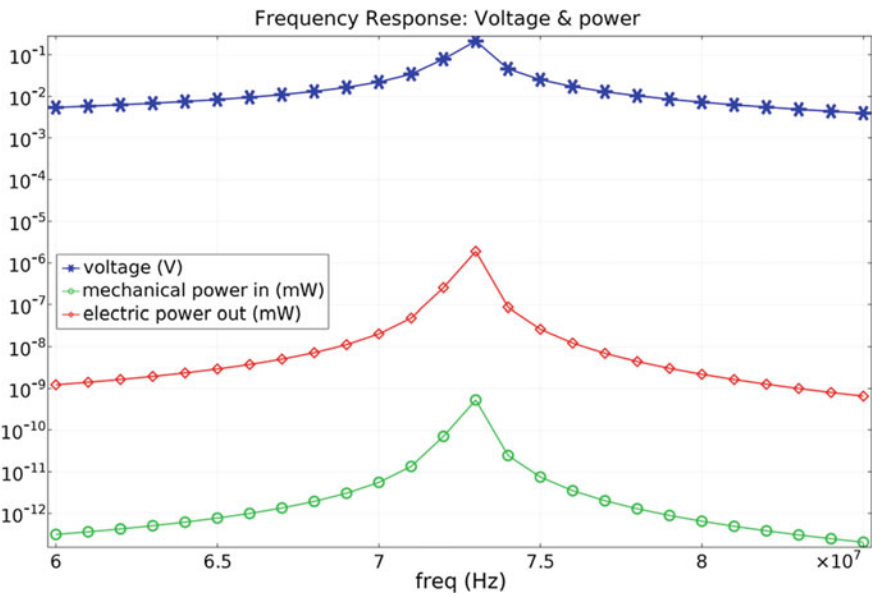
Fig. 10 Response for the load of 1 μN

References

1. Yoon, S., Sim, J.K., Cho, Y.H.: A flexible piezoelectric pulsewave energy harvester for application to high-efficiency multi-functional skin patches. J. Microelectromech. Syst. **25**(2), 388–393 (2016)
2. Yuan, S., Huang, Y., Zhou, J., Xu, Q., Song, C., Yuan, G.: A high-efficiency helical core for magnetic field energy harvesting. IEEE Trans. Power Electron. **32**(7), 5365–5376 (2017)
3. Umaz, R., Garrett, C., Qian, F., Li, B., Wang, L.: A power management system for multianode benthic microbial fuel cells. IEEE Trans. Power Electron. **32**(5), 3562–3570 (2017)
4. Zhao, L., Yang, Y.: Comparison of four electrical interfacing circuits in wind energy harvesting. Sens. Actuat. A **261**, 117–129 (2017)
5. Hwang, W.S., Ahn, J.H., Jeong, S.Y., Jung, H.J., Hong, S.K., Choi, J.Y., Cho, J.Y., Kim, J.H., Sung, T.H.: Design of piezoelectric ocean-wave energy harvester using sway movement. Sens. Actuat. A **260**, 191–197 (2017)
6. Asadi, E., Askari, H., Khamesee, M.B., Khajepour, A.: High frequency nano electromagnetic self-powered sensor: concept, modelling and analysis. Measurement **107**, 31–40 (2017)
7. Khaligh, A., Zeng, P., Zheng, C.: Kinetic energy harvesting using piezoelectric and electromagnetic technologies—state of the art. IEEE Trans. Industr. Electron. **57**(3), 850–860 (2010)
8. Noh, M.S., Kim, S., Hwang, D.K., Kang, C.Y.: Self-powered flexible touch sensors based on PZT thin films using laser lift-off. Sens. Actuat. A: Phys. (2017)
9. Modaresinezhad, E., Darbari, S.: Realization of a room-temperature/self-powered humidity sensor, based on ZnO nanosheets. Sens. Actuat. B: Chem. **237**, 358–366 (2016)
10. Lezhin, D.S., Falaleev, S.V., Safin, A.I., Ulanov, A.M., Vergnano, D.: Comparison of different methods of non-contact vibration measurement. Proc. Eng. **176**, 175–183 (2017)
11. Saha, A., Das, S., Suresh, M., Kiran, V.R., Dey, N.: FPGA based self-vibration compensated two dimensional non-contact vibration measurement using 2D position sensitive detector with remote monitoring. Measurement **111**, 271–278 (2017)
12. Kim, D., Khalil, H., Nam, J., Park, K.: Image-based tracking system for rotating object vibration measurement using laser scanning vibrometer. Int. J. Prec. Eng. Manuf. **16**(8), 1717–1721 (2015)
13. Zhang, B., Cheng, L., Liang, Y., Jin, L., Guo, T., Guan, B.O.: Low-frequency vibration measurement by a dual-frequency DBR fiber laser. Photon. Sen. 1–5 (2017)
14. Grigor'ev, A.V., Zatylkin, A.V., Yurkov, N.K.: Method for contactless three-component vibration measurement. Measure. Tech. **59**(12), 1291–1296 (2017)
15. Gao, X., Zhang, B., Feng, Q., Xie, X., Yang, L.: Nano-vibration measurements using the photoelectromotive force effect in the GaAs crystal. Instr. Exp. Tech. **59**(3), 470–475 (2016)
16. Kutanis, M., Boru, E.O., Işık, E.: Alternative instrumentation schemes for the structural identification of the reinforced concrete field test structure by ambient vibration measurements. KSCE J. Civil Eng. **21**(5), 1793–1801 (2017)
17. Gosk, W., Czech, K.R.: Determination of sandy subsoil stiffness on the basis of surface vibration measurement. Proc. Eng. **189**, 105–210 (2017)
18. Rota-Rodrigo, S., López-Aldaba, A., Pérez-Herrera, R.A., Bautista, M.D., Esteban, Ó., López-Amo, M.: Simultaneous measurement of humidity and vibration based on a microwire sensor system using fast Fourier transform technique. J. Lightwave Technol. **34**(19), 4525–4530 (2016)
19. Williams, C.B., Yates, R.B.: Analysis of micro-electric generator for Microsystems. Sens. Actuat. A **52**, 1–3 (1996)

Effects of Different Shapes of Piezo-Acoustics-Based Adaptive Glucose Sensing System (PABAGS) on Generated Pressure and Its Analysis

Bhupendra Sindhal, Varshali Sharma and Ritu Sharma

Abstract Blood glucose can be monitored using the miniaturized Piezo-Acoustic sensor. The piezoelectric material can be used to generate acoustic pressure in blood by applying electric potential on the Piezoelectric material, and that pressure depends on blood density. This paper describes designing of Piezo-Acoustic transducer using lead-free piezoelectric materials to calculate its performance with glucose level in blood of human. The finite element method-based software is used for the design of model. The size of the piezoelectric material optimized in this paper with respect to its width and thickness (height) and shape acting as a channel to produce maximum transmitting pressure. The transmitted pressure is agitated by "2 MHz" frequency tremble at "1.5 V". "Barium Titanate" lead-free piezoelectric material is used as material. The pressure generated is found to be varying with the change in shape of the Piezo-Acoustic transducer.

Keywords Finite element method software · Barium titanate lead-free piezoelectric material · Blood glucose (sugar) · Piezo-acoustic transducer · MEMS PABAGS system

B. Sindhal (✉) · R. Sharma
Department of Electronics and Communication, Malaviya National
Institute of Technology, Jaipur, Jaipur, India
e-mail: 2016peb5310@mnit.ac.in

R. Sharma
e-mail: rsharma.ece@mnit.ac.in

V. Sharma
Department of Electronics and Communication,
Manipal University Jaipur, Jaipur, India
e-mail: varshali.manipal.electronics@gmail.com

1 Introduction

Key facts: Data published by World Health Organization (WHO) claims that the number of people with diabetes had increased by 314 million as compared to 108 million in 1980 to 422 million in 2014. Moreover, there is an alarming increase in the number of diabetic patients over 18 years of age (adults). It had risen from 4.7% in 1980 to 8.5% in 2014. In 2015, around 1.6 million deaths were caused due to diabetes. WHO also projected that in 2030, diabetes will be the leading reason for most of the deaths in the world [1]. One in three or 33% of Indians over 30 years of age are suffering from diabetes [2].

Hyperglycemia is a metabolic disorder in which sugar level of a person is high for the long term. Hyperglycemia leads to increased thirst, increased hunger, and frequent urination. It can cause many complications. Such as cardiovascular disease, chronic kidney disease, stroke, damage to the eyes and foot ulcers. **Hypoglycemia** is also a metabolic disorder in which low blood sugar level for a short span of time. Hypoglycemia leads to sweating, dizziness, shaking, headache, weakness, rapid heartbeat, irritability, fatigue, impaired vision, and anxiety. The diabetic patients are having blood glucose level less than "70 mg/dl" is classified as hypoglycemia. The patients having blood glucose level more the "180 mg/dl" classified as hyperglycemia. Thus, it becomes essential to monitor regularly the glucose (Sugar) level in blood [3, 4].

The traditional method of monitoring blood glucose level, nowadays, is minimal invasive blood glucose monitoring method which works on glucose oxidase principle. This method can be carried out by penetrating a needle in the skin usually fingertip to take blood and put that blood on a reagent test strip than that test strip is inserted into the traditional glucometer. By this method, the glucose concentration is converted into electric signals [5] which are used to determine the glucose level. However, this method is expensive, painful and may cause infection and also not very accurate as compared to blood taken directly from veins. Therefore, this method cannot be used further for the blood glucose measurement purpose.

Lot of research is going on to develop noninvasive glucose measurement methods which include NIR Spectroscopy, Raman Spectroscopy, Optical Coherence Tomography, Photoacoustic Spectroscopy, and more. By all the above methods, glucose concentration is measured without extracting blood from the body and without penetrating any needle to take the blood from the human body. All the above methods have their advantages and challenges [6].

In our proposed model, piezo-acoustic transducer (ultrasonic transceiver) for blood glucose prediction is a noninvasive approach using ultrasonic Microelectronics Mechanical Systems (MEMS) miniaturized system. Ultrasonic MEMS technology has many advantages such as flexible geometry, reduced voltage requirement, and mixing of different resonant frequencies for integration with supporting microelectronics circuits [7]. The piezoelectric technology can be used to measure the nature and the properties of the passing ultrasonic wave from the blood medium of various densities. Any increase in the glucose level in blood increases its density. The advantage of using piezoelectric material is that it offers high pressure per density

ratio for the actuator and highly chemically stable in the hostile environment. To make this model, it is needed to have high electromechanical coupling coefficient, relatively large dielectric constant, and sizeable piezoelectric coefficient [8]. Most commonly used piezoelectric material is lead zirconate titanate (PZT), but it cannot be used as it is not environment-friendly, it contains more than 60% of lead by weight which is harmful to the human body [9]. Many lead-free piezoelectric materials are present, nowadays, such as Barium Titanate ($BaTiO_3$) (BT), Barium Titanate Poled, Barium Sodium Niobate ($Ba_2NaNb_5O_{15}$) (BNN), Lithium Tantalate ($LiTaO_3$), and Lithium Niobate ($LiNbO_3$) (LN). In which, Barium titanate (BT) is most suitable lead-free piezoelectric material because it generates high pressure as compared to other lead-free piezoelectric materials and which is best suitable for the human body [10, 11].

The goal of this paper is to compare different shapes and size of lead-free piezo-electric material which is capable of generating high acoustic pressure. In this paper we have mentioned only simulated results and no sample prototype have been developed, so we have not utilized any human blood sample for this purpose.

2 Proposed PABAGS System

To measure the blood glucose level by the proposed system, proper target location needs to be chosen such as the earlobe, finger webbing, and lips of a human. The earlobe is used as a target medium in this experiment because it does not contain cartilage, blood supply is high, and the device will easily reach to the earlobe [11–13]. The block diagram of a suggested model of the system with ultrasonic MEMS sensor is shown in Fig. 1.

2.1 Working Principle

Ultrasonic wave is transmitted through the medium at the frequency of 2 MHz which is above the human's normal hearing range, so that it will not destroy eardrums. The piezoelectric transducer generates these ultrasonic waves, and when these waves are transmitted through the biological tissues, it generates vibration in the medium. Standing waves have maximum values called nodal values (stand still) and minimum

Fig. 1 Block diagram of PABAGS system

values called antinodal values (undergo maximum displacement) are created. This wave generates acoustic energy in the medium due to the presence of the ultrasonic field [12].

2.2 Equations

If the volume of molecules (v) and its path length (L) from the pressure node, the radiation force (F) applied on that molecule are represented from the gradient of acoustic energy by Eq. (2) provided that the diameter of the molecule should be smaller than the ultrasonic wavelength.

Compressibility factor (Z) is expressed by Eq. 1

$$f(Z, \rho) = [\{(5\rho_c - 2\rho_w)/(2\rho_c + \rho_w)\} - (Z_c/Z_w)] \tag{1}$$

where Z_w is a compressibility factor of suspending medium, Z_c is compressibility of molecule present in the medium, ρ_c is densities of molecules present in the medium, and ρ_w is densities of suspending medium [12].

The radiation force is defined by Eq. (2)

$$F = -[\pi\rho_o 2v\, Z_w/(2\,\lambda)] \cdot f(Z, \rho) \cdot \sin(4\pi L/\lambda) \tag{2}$$

where ρ_o is the amplitude of the ultrasonic waves, Z_w is a compressibility factor of suspending medium, λ is the wavelength of ultrasonic standing waves, Z is compressibility factor, and ρ is the density of the molecule [12].

Any increase in glucose concentration in the blood increases its density, and thus the pressure generated by transducer also increases.

3 The Model Geometry of Piezo-Acoustic Transducer

For simulation, Finite Element Method software is used which provides 3D partial differential equations and the 2D axis-symmetric configurations to solve acoustic piezoelectric device and to imitate the device in the blood sample medium [7].

The blood medium considered as a cylindrical structure of height 7 mm and radius 0.5 mm. Piezoelectric material is used have different shapes.

3.1 The Square of Side is 0.3 mm as Shown in Fig. 2

See Fig. 2.

3.2 T Shape as Shown in Fig. 3

See Fig. 3.
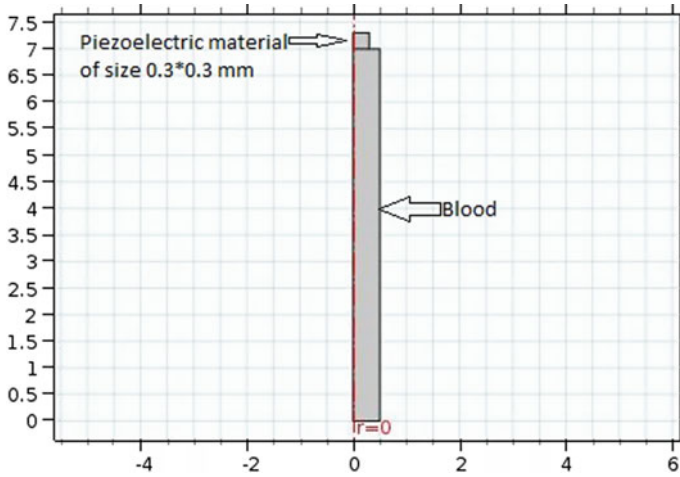
Fig. 2 Geometry of 3 * 3 square piezo-acoustic transducer and blood
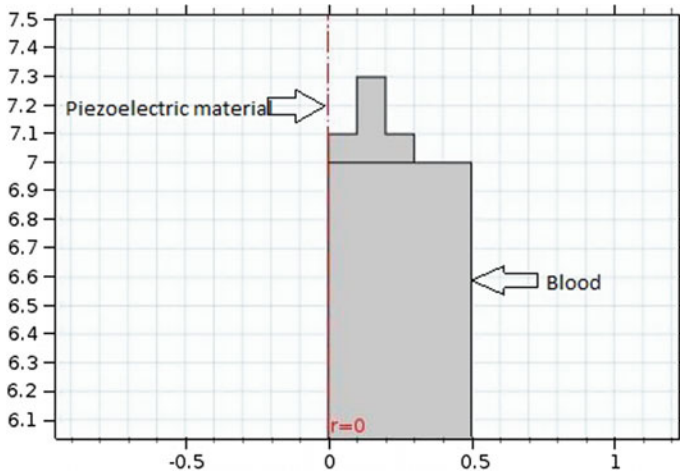
Fig. 3 Geometry of T shape piezo-acoustic transducer and blood

3.3 L Shape as Shown in Fig. 4

See Fig. 4.

3.4 Staircase Shape as Shown in Fig. 5

From the above models, the geometry of ultrasonic piezoelectric MEMS device using Barium Titanate (BaTiO$_3$) (BT) is designed and simulated.
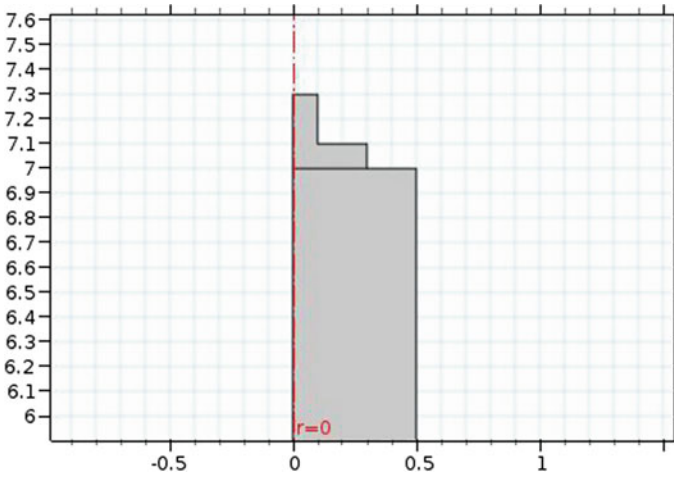
Fig. 4 Geometry of L shape piezo-acoustic transducer and blood
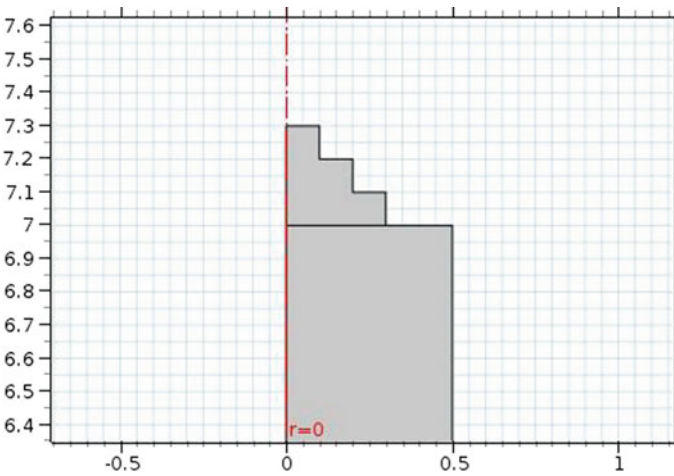
Fig. 5 Geometry of staircase shape piezo-acoustic transducer and blood

4 Results and Discussion

4.1 Effect of Various Frequencies

The frequency is varied in the range of 1–3 MHz as shown in Fig. 6; It can be seen that piezoelectric material produces maximum sustainable pressure at 2 MHz and after 2 MHz, the generated pressure is high which may harm humans.

4.2 Effect of Distinct Shapes of Piezoelectric Material on Generated Acoustic Pressure

The applied pressure is thus taken as 2 MHz from the observations of Fig. 6. As the size of the piezoelectric material is increased, the pressure generated in medium also increases, because piezoelectric coefficient depends on size due to the strong orientation of axis [13, 14].

When the shape of piezoelectric material is changed, the acoustic pressure generated in the blood medium also changes as depicted in Fig. 7.

Simulation results of different shapes of the Piezo-Acoustic transducer and also justified by Eq. 2 are as shown in Figs. 8, 9, 10, and 11, respectively.

Thus, as seen from all the above figures that when ultrasonic waves pass throw the biological tissues, a vibration is produced in the medium. This tissue undergoes maximum displacement near the piezoelectric electrode and thus generates a pressure; the waveform is shown in Figs. 8, 9, 10, and 11 and obeys the Eq. 2.

	1 MHz	1.5 MHz	2 MHz	2.5 MHz	3 MHz
Series1	-290.14	-1110.2	-3240.9	-9971.2	-59050
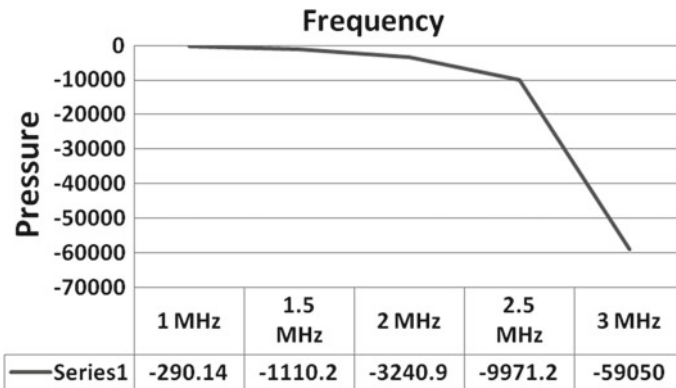
Fig. 6 Pressure versus frequency graph to obtain optimized frequency

Effects of Different Shapes of Piezoelectric Material

	L shape	T shape	Stairecase Shape	3*3 Square Shape
——Series1	-248.33	-341.74	-452.27	-3065
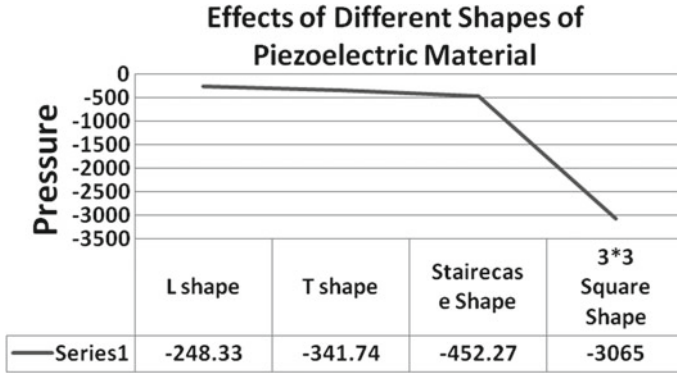
Fig. 7 Effect of different shapes of piezo-acoustic transducer at 2 MHz frequency
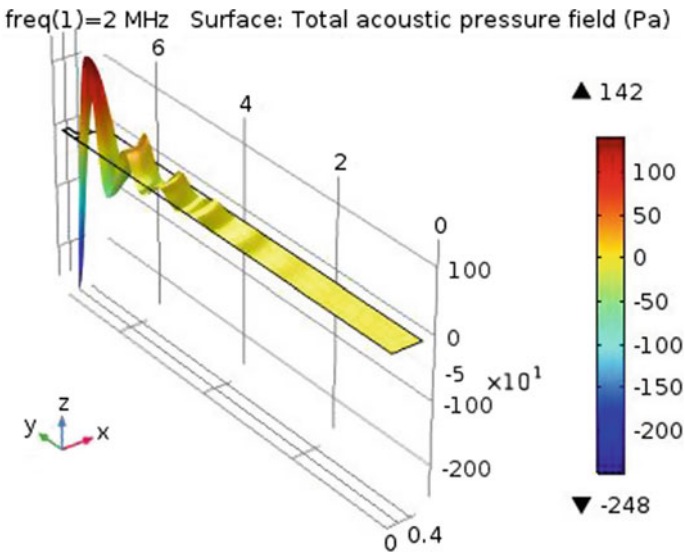
Fig. 8 Simulation result of L shape piezo-acoustic transducer

5 Conclusion

The acoustic pressure generated in blood medium by distinct shapes of piezoelectric material simulation by Finite Element based Software; it is found that as the cross section area of the piezo-acoustic transducer (piezoelectric material) increases, the acoustic pressure generated in blood medium also gets increases. Hence, that different shape can be used to obtain further accuracy. However, that square shapes is best suitable for generating a pressure $|Pa| = 3065$ (kg/m s^2) which can be sustained by a human and can further be used to measure blood glucose level.
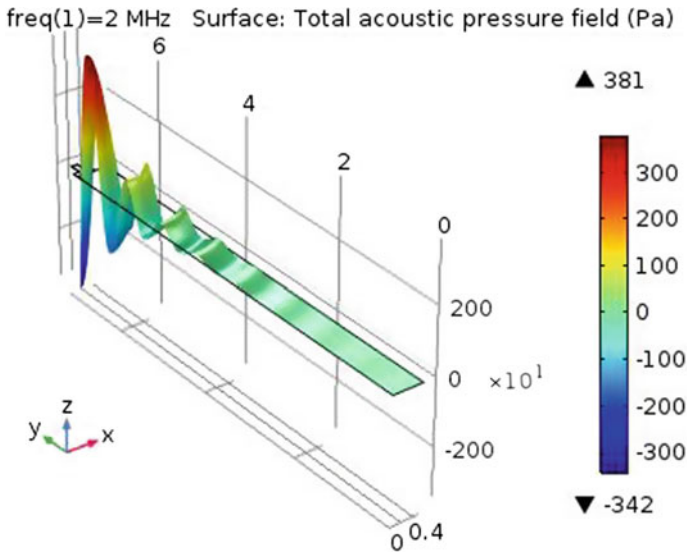
Fig. 9 Simulation result of T shape piezo-acoustic transducer
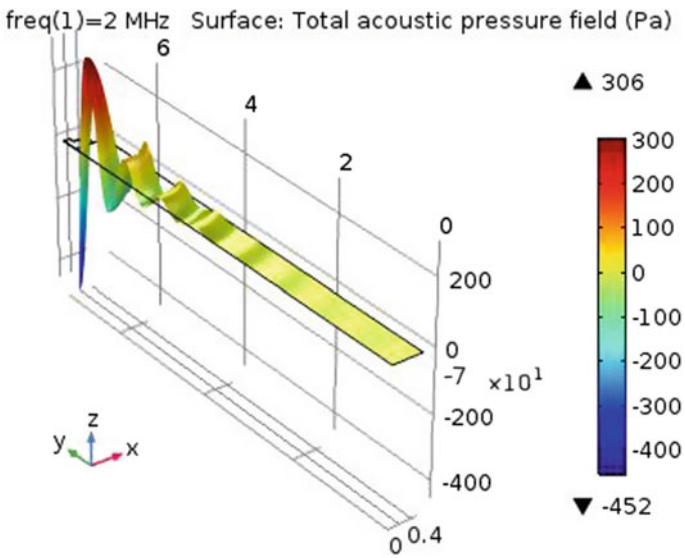
Fig. 10 Simulation result of Staircase shape piezo-acoustic transducer
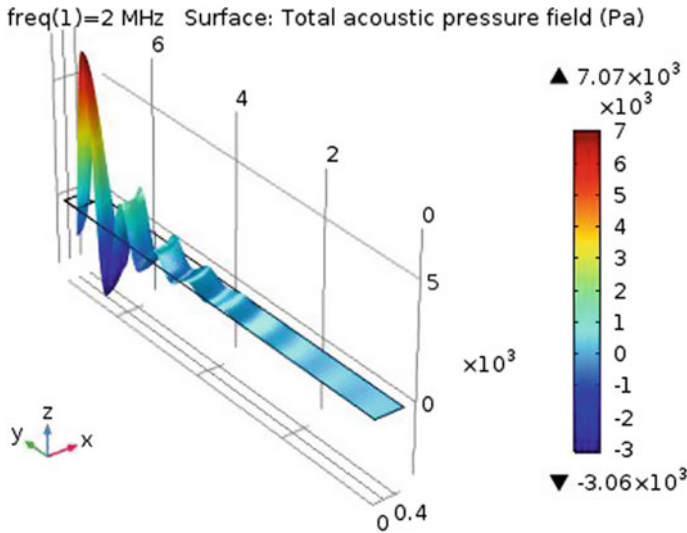
Fig. 11 Simulation result of 3 * 3 square shape piezo-acoustic transducer

References

1. Diabetes fact sheet (Updated: Jan 2017). http://www.who.int/mediacentre/factsheets/fs312/en/. Accessed 20 Feb 2017
2. International Diabetes Federation. http://www.idf.org/our-network/regions-members/south-east-asia/south-east-asia-news.html. Accessed 20 Feb 2017
3. Davis, S.N., Lastra-Gonzalez, G.: Diabetes and low blood sugar (Hypoglycemia). J. Clin. Endocrinol. Metab. **93**(8), (2008)
4. Phillips, P.J., Flynn, P.J., Scruggs, T., Bowyer, K.W., Chang, J., Hoffman, K., Marques, J., Min, J., Warek, W.: Overview of facial recognition grand challenge. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2005)
5. Yanez, M.G.: Glucose Meter Fundamentals and Design, Freescale Semiconductor (2013)
6. So, C.-F., Choi, K.-S., Wong, T.K.S., Chung, J.W.Y.: Recent advances in noninvasive glucose monitoring. In: Medical Devices: Evidence and Research (2012)
7. Pattnaik, P., Kamilla, S.K., Das, D.B.: Alternate glucometer bio-sensor model based on ultrasonic MEMS transceivers. In: Proceeding of COMSOL Conference, Bangalore (2013)
8. Greve, D.W., Oppenheim, I.J.: Coupling of MEMS ultrasonic transducer. In: Proceedings of IEEE, Sensors, vol. 2, pp. 814–819 (2003)
9. Aksel, E., Jones, J.L.: Advances in Lead-Free Piezoelectric Materials for Sensors and Actuators, Sensors, vol. 10, pp. 1935–1954 (2010)
10. Pattnaik, P., Pradhan, S.K., Kamilla, S.K., Das, D.P.: Studies of lead-free piezo-electric materials based ultrasonic MEMS model for bio sensor. In: Proceeding of COMSOL Conference 2012, Bangalore (dated 3–4 Nov 2012)

11. Sindhal, B., Sharma, R.: Piezo-acoustics based adaptive glucose sensing system. In: National Conference on Advanced Communication Technology and Network, MNIT, Jaipur (Dated 24–25 Oct 2017)
12. Asaduzzaman, A., Samadarsinee, S., Chidella, K.K.: Simulating Multisensor Noninvasive Blood Glucose Monitoring Systems, pp. 1–7. IEEE (2016)
13. GlucoTrack. http://www.healthline.com/diabetesmine/glucose-testingviaearlobe#5. Accessed 20 Feb 2017
14. Belgacem, B., Calame, F., Muralt, P.: Design, modeling with process fabrication of piezoelectric micromachined ultrasonic transducers. In: IEEE Ultrasonics Symposium, pp. 483–486. Rotterdam, The Netherlands (2005)

Proposed Modifications in the Excitation Codebook Structure of ITU-T CS-ACELP Speech Codec and Its Overall Comparative Performance Analysis with CELP-Based AMR-NB Speech Codec

Nikunj Tahilramani and Ninad Bhatt

Abstract The paper addresses the tactic of using forward mode excitation codebook structure of extended G.729 (ITU-T. Recommendation, 2007 [1]) consuming two non-zero pulses in each track in ITU-T 8 Kbps CS-ACELP speech codec (ITU-T. Recommendation, 2007 [1]). The proposed modification in legacy speech codec for determining optimized excitation codebook codevector using least significant search engine avoids two different codebook structures to minimize the complexity of search engine. Modification in the excitation codebook structure of the legacy speech codec triggers the bit rate of 11.6 Kbps. Comparative performance analysis is performed between the proposed speech codec and CELP-based speech codec using different subjective and objective quality assessment parameters.

Keywords G.729E · Least significant pulse replacement · CS-ACELP speech codec · Subjective and objective quality assessment

1 Introduction

Hybrid class of speech coders always acts as an appropriate intermediate solution in terms of output decoded speech quality between source coders and waveform coders. Analysis-by-synthesis (ABS)-based speech coders are classified under the roof of

N. Tahilramani (✉)
ECC Deparment, Uka Tarsadia University, Surat, Gujarat, India
e-mail: nikunjvtec@gmail.com

N. Bhatt
ECC Deparment, C. K. Pithawala College of Engineering, Surat, Gujarat, India
e-mail: bhattninad@gmail.com

hybrid class of speech coders, which works with the medium bit rate and produces quite fair quality of speech at the output. CS-ACELP (Conjugate Structure-Algebraic Code Excited Linear Prediction) is one of the classic examples of speech coders which are categorized under the classification of ABS speech coders where extracted parameters of speech signal are tested in ABS feedback loop before transmission to the decoder. The chief benefit of ABS speech codecs is that the decoder is embedded in encoder itself which actually reduces the residual error of excitation every time by comparing synthesized speech and original speech at the encoder itself.

Here, ITU-T 8 Kbps CS-ACELP speech codec is considered as base speech codec for modification purposes. The input speech is partitioned into the slots of 10 ms for the parameter extraction and analysis purposes in legacy speech codec. Segment partition of 10 ms of input speech incorporates 80 samples in one frame of speech signal with the sampling frequency of 8000 samples/s. Modification is proposed in the excitation codebook structure of the legacy speech codec which actually is responsible for the production of accurate excitation of the speech signal, which is further utilized for parameter extraction and transmission to the decoder to generate synthesized output speech signal. Excitation of the speech codec is represented in terms of optimized excitation code vector which are selected per frame in a way that the energy of the residual noise is minimized every time in ABS loop.

The proposed speech codec is compared with CELP-based Adaptive multi-rate-narrow band (AMR-NB) speech codec which actually works with eight different bit rates. Switching of the bit rate in AMR codec is dependent on voice traffic, transmission errors, and background noise. The main goal of adaptation in AMR speech codec is to provide good speech quality under fluctuating channel conditions. The compared speech codec comes under the full-rate codec mode of AMR having highest bit rate of 12.2 Kbps.

The flow of the paper is narrated as follows. Section 2 deals with the introduction of 12.2 Kbps GSM AMR speech coder. In Sect. 3, the proposed modification in the excitation code structure **of ITU-T 8 Kbps** speech codec is explained. Section 4 describes the steps for searching of optimum excitation code vector from modified excitation codebook structure. In Sect. 5, simulation results and comparative performance analysis are shown using different objective and subjective parameters. Decision remarks are given in Sect. 6.

In the following subsequent sections, **ITU-T 8 Kbps** CS-ACELP speech codec, 12.2 Kbps CELP-based **12.2 Kbps** speech codec, and 11.6 Kbps speech codec will be notified as legacy speech codec, AMR-NB speech codec, and proposed speech codec.

2   12.2 KBPS GSM AMR Full-Rate Speech Codec

In AMR-NB speech codec, excitation and short-term prediction parameters are scalar quantized for obtaining simplicity in implementations. Here, excitation parameters like pitch lag (P), codebook index (k), gain (G), and pitch filter coefficient (b) are
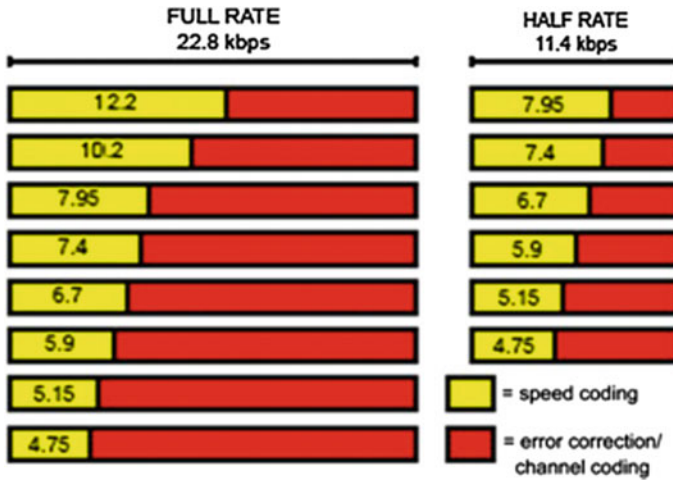
Fig. 1 GSM AMR speech codec in different modes [2]

computed and transmitted per subframe, whereas LP coefficient (a) are computed and transmitted per frame. Furthermore, having assumed that the lag of the filter (P) is searched in the range 16–160 (equivalent to 50–500 Hz), there exist a requirement of 8 bits to encode and represent it exactly on the basis of subframe [2]. Bits of all other excitation parameters are tuned and set in order to map the total effective bitrates of designed CELP coder in line with bitrates of different modes of AMR-NB coder [2] as shown in Fig. 1. AMR also provides the facility of voice activity detection (VAD) and discontinuous transmission (DTX) where the link is checked and flag of VAD is considered as "0" in the silent passage of speech signal which actually indicates discontinuous transmission to effectively utilize the bandwidth of a channel. AMR-NB full-rate speech codec is mainly categorized into eight diverse bit rates of 12.2, 10.2, 7.95, 7.4, 6.7, 5.9, 5.15, and 4.75 Kbps.

3 Proposed Modification in Excitation Codebook Structure of Legacy Speech Codec

The legacy speech codec works with the excitation codebook structure having 40 pulse position is divided into a total of 4 different tracks. 40 pulse positions are positioned in the excitation codebook on the principle of interleaved single-pulse permutation (ISPP) with 16 pulse positions in last track and 8 pulse positions each in first three tracks. Determination of optimized excitation code vector having four non-zero pulse contribution is achieved by utilizing least significant pulse replacement algorithm and the best pulse combination which reduces the residual error is declared as final excitation code vector. Apart from the determination of an optimum

Table 1 Modified excitation codebook structure

Track	Pulses	Signs	Positions
1	m_1, m_0	s_1, s_0: ± 1	0, 5, 10, 15, 20, 25, 30, 35
2	m_3, m_2	s_3, s_2: ± 1	1, 6, 11, 16, 21, 26, 31, 36
3	m_5, m_4	s_5, s_4: ± 1	2, 7, 12, 17, 22, 27, 32, 37
4	m_7, m_6	s_7, s_6: ± 1	3, 8, 13, 18, 23, 28, 33, 38
5	m_9, m_8	s_9, s_8: ± 1	4, 9, 14, 19, 24, 29, 34, 39

combination of pulse positions from four different tracks, degree of complexity of search engine also plays a crucial role. The requirement is to determine the finest code vector works with a reduced degree of complexity of search engine which deals with a limited number of searches. The study of search engine starts with the full search approach applied on the legacy speech codec which requires 8192 searches to determine excitation code vector. Afterwards research in the area of determination of best search engine which takes a less number of searches with reduced complexity started enormously in the area of speech coding. Many search engines were invented, but few of them which were having moderate complexity levels were focused search approach (1440 searches) and depth first search approach (320 searches) [3].

Complexity in any of the search engine applied to the excitation codebook structure of the legacy speech codec always increases because of the 16 pulse position in final track instead of 8. Due to double the number of pulses in last track compared to first 3, recursive iteration is carried out for two times with first having non-zero pulse from first eight pulse positions of the last track, and the second combination comprises of last pulse position from next eight pulse positions of the last track.

The proposed approach avoids the complexity due to 16 pulse positions in the last track by dividing it into 2 different tracks to make 5 track **excitation code structures** as shown in Table 1.

On the same note, contribution of two non-zero pulse positions is considered to represent final excitation code vector instead of 1 to increase the quality of excitation which ultimately increase quality of synthesized or decoded output speech signal. The proposed modification actuates the bit rate of 11.6 Kbps in which 35 bits are required for the indices and sign of the of the final excitation code vector as shown in Table 2.

4   Searching of Optimum Excitation Codevector from Modified Excitation Codebook Structure

Initial codevector comprises of ten pulse position from five different tracks. First ten non-zero pulses of initial code vector are selected by finding first two largest of correlation vector d(n) from individual track. Searching of the next stage excitation code vector with respect to previous initial code vector starts by replacing one pulse

Table 2 Bit allocation structure for different parameters of speech signal in proposed 11.6 Kbps speech codec

Parameter	Subframe 1	Subframe 2	No. of bits/frame
Long-term predictor codebook delay	8	5	13
Line spectrum pair	–	–	18
Indices of fixed excitation codebook	30	30	60
Long-term predictor-delay parity	1	–	1
Gain of codebook (stage 1)	3	3	6
Sign of fixed excitation codebook	5	5	10
Gain of codebook (stage 2)	4	4	8
Total			116

among two from each track with all other pulses from that respective track to find out the least important pulse. The best ten pulse combination which maximizes the ratio of Eq. 1 (Q_k) will be acknowledged as final optimized excitation code vector combination. Search engine is used to determine the least significant pulse in each track which requires 60 number of searches as each track is having 8 pulse positions with 1 non-zero pulse position is kept constant [1].

$$\max_k Q_k = \max_k \frac{C_k^2}{E_k} = \max_k \frac{(d^t c_k)^2}{c_k^t \varnothing_{c_k}}$$

$$= \frac{(\sum_{j=0}^{M-1} s_j d(m_j))^2}{\sum_{j=0}^{M-1} \varnothing(m_j, m_j) + 2 \sum_{i=0}^{M-2} \sum_{j=i+1}^{M-1} s_i s_j \varnothing(m_i - m_j)} \tag{1}$$

Here,

M   Total number of tracks in Excitation codebook structure
C_k Kth Excitation codebook vector
D   Correlation vector.

$$d(n) = \sum_{i=n}^{K-1} x_2(i)h(i-n), i = 0, \ldots, K \tag{2}$$

$$\varnothing(i, j) = \sum_{n=j}^{K-1} h(n-i)h(n-j), j = i, \ldots K \tag{3}$$

Here, K is the entire pulse locations in excitation codebook structure, h(n) is an impulse response of a linear predictive synthesizing filter, and $x_2(n)$ is a target signal which is updated for every next stage of determination of excitation codevector [1].

Only 1 sign bit is transmitted instead of two from each respective track using the concept of pivot pulse position [7]. The smallest pulse position codeword among two from each track is declared as pivot pulse if the sign of both non-zero pulses are the same which are either -1 or $+1$ [9]. In that case, sign bit of the pivot pulse position is only transmitted. At the decoder, if the first decoded 3-bit positional codeword is smaller than the second one, then receiver will automatically determine the second pulse position sign bit which is same as the sign of pivot pulse position which is having smaller pulse position codeword. While in another case, the largest pulse position is declared as pivot pulse position and the sign of it is transmitted to the decoder. Decoder will determine the sign of other pulse position which is exactly opposite to the received one because the decoded pivot pulse position is larger among the two.

5 Comparitive Performance Analysis Between Proposed Speech Codec and AMR-NB Speech Codec

Here, comparative analysis is performed between proposed 11.6 Kbps speech codec and 12.2 Kbps CELP-based AMR-NB speech codec with different subjective and objective analysis parameters. Here, the input is four different speech wave files taken from NOIZEUS corpus database [4], where two are female voice and other two are male voice speakers. The wave files are played at 128 Kbps bitrate having 8 kHz sampling frequency and each sample of wave file is represented by 16 bits.

5.1 Subjective Analysis Results

Subjective analysis [5, 6] is performed with ten dissimilar subjects to give MOS ratings with five men and five women. The final MOS ratings are calculated from the judgement of all ten subjects as rating is averaged. The comparative MOS ratings of the proposed speech codec and AMR-NB speech codec are shown in Fig. 2. The MOS ratings of the proposed-based speech codec are quite better than the AMR-NB speech codec. The naturalness in terms of intelligibility of the proposed speech codec for all wave files is fairly satisfactory even though the bit rate of the proposed speech codec is less compared to AMR-NB speech codec.

5.2 Objective Analysis Results

The results of the objective analysis parameters are shown in Tables 3, 4, and 5.

Table 3 Waveform-based assessment

Wave file	ABS	MSE	RMSE	SNR	SNRseg
GSM AMR-NB speech codec (12.2 kbps)					
Sp18.wav	101.63	0.000112	0.0106	11.5163	7.6934
Sp04.wav	99.203	0.000113	0.0106	12.6828	8.4478
Sp26.wav	136.53	0.000146	0.0121	10.8590	8.2569
Sp24.wav	144.67	0.000083	0.0091	12.6091	10.2304
Proposed Speech codec (11.6 kbps)					
Sp18.wav	1.5637	0.000102	0.0105	11.7989	7.8541
Sp04.wav	0.6131	0.000100	0.0104	12.6685	8.5519
Sp26.wav	0.3827	0.000138	0.0118	10.9513	8.2591
Sp24.wav	0.4278	0.000079	0.0087	0.0087	10.4882

Table 4 Perceptual-based assessment and composite measures

Speech codecs	Wave file	Gender	No. of samples	Covl	Csig	Cbak	PESQ
AMR 12.2 kbps	Sp04.wav	M	16,928	4.19	4.75	3.71	3.6500
	Sp18.wav	F	18,269	3.67	4.39	3.32	2.9971
	Sp24.wav	M	27,768	3.85	4.35	3.68	3.3640
	Sp26.wav	F	19,934	3.84	4.53	3.48	3.1568
Proposed CS-ACELP (11.6 kbps)	Sp04.wav	M	16,928	4.22	4.76	3.74	3.8600
	Sp18.wav	F	18,269	3.70	4.44	3.42	3.0121
	Sp24.wav	M	27,768	3.88	4.38	3.54	3.5487
	Sp26.wav	F	19,934	3.86	4.56	3.50	3.2658

Table 5 Spectral-based objective evaluation

Algorithm	Wave file	Gender	No. of samples	LLR	WSS	fwSNRseg	ISD	CEP
AMR 12.2 kbps	Sp04.wav	M	16,928	0.2783	27.506	27.506	0.3563	2.9008
	Sp18.wav	F	18,269	0.2032	32.681	11.7186	0.2968	2.4508
	Sp24.wav	M	27,768	0.4941	28.431	14.0519	0.5601	4.8662
	Sp26.wav	F	19,934	0.2258	25.203	13.9092	0.2954	2.7915
Proposed CS-ACELP (11.6 kbps)	Sp04.wav	M	16,928	0.3689	27.617	14.8489	0.4080	3.8094
	Sp18.wav	F	18,269	0.3108	36.276	10.7227	0.2975	3.4630
	Sp24.wav	M	27,768	0.6328	29.837	14.2603	0.9431	5.7174
	Sp26.wav	F	19,934	0.3239	26.235	13.9509	2.2962	3.6816
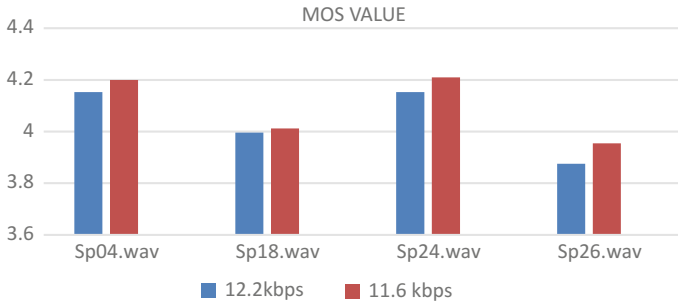
Fig. 2 MOS scores for different wave files

5.3   *Waveform-Based Assessment*

The results of waveform-based analysis for proposed speech codec are shown in Table 3 which is fairly analogous to AMR-NB speech codec. Even though there is a shortfall of 6 bits in proposed speech codec, the results of the premium objective quality assessment parameter called as PESQ (Perceptual Evaluation of Speech Quality) for the proposed speech codec are more comparable to AMR-NB speech codec.

5.4   *Spectral-Based Analysis*

From the Table 5, it can be witnessed that the values of LLR, WSS, ISD, and ISD increases with reduction in the bit rate [2], whereas the values of fwSNRseg in the case of the proposed speech codec is quite comparable and it is more for some of the speech wave files which actually decreases with the reduction in the bit rate.

6   Concluding Remarks

The motive of developing speech coder is to denote the speech with fewer bits as possible while preserving quality, intelligibility, and pleasantness of the decoded speech. Since the last few years, an assortment of speech coding techniques has been developed and they all have been recognized by various reputed organizations like ITU-T, ETSI, Research and Development Centre for Radio Systems of Japan (RCR), Telecommunications Industry Association (TIA), and United States Department of Defense (US DoD) [1]. Here, the proposed speech codec act as a trade-off between legacy speech codec and AMR-NB 12.2 Kbps speech codec. The reason behind comparative analysis of proposed speech codec with 12.2 Kbps AMR-NB speech

codec only is because it is the nearest bit rate with respect to the proposed speech codec. The performance of the proposed speech codec is analyzed with different objective and subjective analysis parameters and from the derived result, it can be concluded that the quality of the proposed speech codec is quite consistent and comparable with the AMR-NB speech codec. The remaining 6 bits of proposed speech codec can be utilized for the purpose of information hiding [8] or ownership proof.

References

1. ITU-T. Recommendation 2007: coding of speech at 8 kbit/s using conjugate-structure algebraic-cod-excited linear prediction (CS-ACELP)
2. Bhatt, N., Kosta, Y.: Overall performance evaluation of adaptive multi rate 06.90 speech codec based on code excited linear prediction algorithm using MATLAB. Int. J. Speech Technol. **15**, 119–129 (2012)
3. Tahilramani, N., Bhatt, N.: Steganography in speech signal with Enhanced multi-pulse excitation codevector with reduced number of bits. In: IEEE International Conference on Electrical, Electronics, Signals, Communication and Optimization (EESCO), pp. 1–4, 24–25 Jan 2015
4. The NOIZEUS database: http://www.utdallas.edu/~loizou/speech/noize (2009)
5. Bhatt, N., Kosta, Y.: Proposed modifications in ETSI GSM 06.10 full rate speech codec and its overall evaluation of performance using MATLAB. Int. J. Speech Techol. **14**(3), 157. Springer (2011)
6. ITU-T. Recommendation 2003: Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm (ITU-T Rec. P. 835, 2003)
7. Bernard, A.: Algebraic codebook system and Method. U.S. Patent US6847929B2, 25 Jan 2005
8. Bhatt, N., Kosta, Y.: Implementation of variable bit rate data hiding techniques on standard and proposed GSM 06.10 full rate coder and its overall comparative evaluation of performance. Int. J. Speech Technol. **16**(3), 285–293 (2012)
9. Tahilramani, N., Bhatt, N.: Proposed modifications in ITU-T G.729 8 Kbps CS-ACELP speech codec and its overall performance analysis. Int. J. Speech Technol. **20**(3), 615–628 (2017), https://doi.org/10.1007/s10772-017-9431-3

A Review of Internet of Things from Indian Perspective

Kartik Upadhyay, Ashwani Kumar Yadav and Palak Gandhi

Abstract Internet of things (IoT) refers to the technology of interconnected devices communicating among themselves with the help of the internet. Internet of Things has created a buzz all around the world. When we talk about India in particular, the masses lack the basic knowledge of IoT and it may take a while for them to get used to the idea of the emerging technology. Knowledge of IoT needs to be spread among the masses in India, so that IoT can expand in India. India has created its IoT policy aiming to occupy 5–6% of global market. In this paper, we have discussed about the government policies regarding IoT, opportunities, challenges, and applications developed in this sector. The paper intends to provide a detailed review of the development of IoT in India and the challenges involved.

Keywords IoT · M2M · Smart city · Smart agriculture

1 Introduction

1.1 A Subsection Sample

Internet of Things (IoT) can be described in some basic terms as an interconnection of things which are connected to each other using the internet. The idea behind this is to connect the things, so that they can exchange information and communicate among each other using certain information exchanging devices, and the internet. It attains the objective of intelligent recognizing, finding, tracing, monitoring, and

K. Upadhyay (✉) · A. K. Yadav · P. Gandhi
ASET, Amity University Rajasthan, Jaipur, India
e-mail: upadhyaykartik@hotmail.com

A. K. Yadav
e-mail: ashwaniy2@gmail.com

P. Gandhi
e-mail: palakgandhi.1@hotmail.com
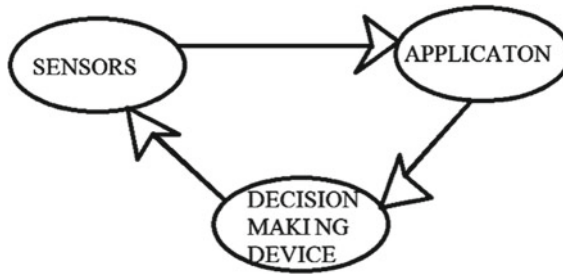
Fig. 1 Stages of IoT

dealing things [1]. The growth of IoT industry has helped in bringing the physical world and the computer-based systems close to each other. With the advancement of the VLSI technology and decreasing size of transistors present on a single chip of a microcontroller, the IoT paves its way towards its growth in the industry. There is a fine difference between IoT and the internet. As the name suggests the internet is a part of the Internet of things while IoT is the extension of the primary Internet [2].

The basic difference between the indigenous Internet and the internet of things is, that the internet of things does not just restricts the connectivity between computers, tablets, and smartphones but also connects multiplicity of sensors and actuators. India is still fighting ways to accept IoT [3].

Currently, India has an IoT industry making revenue of $130 million annually and there are 400 companies currently working in the IoT sector. India has been estimated to touch $15 billion in the IoT industry by 2020. The rate at which the tech-savvy consumers are increasing, it is not a dream difficult to achieve. In this paper, we discuss about the opportunity, status and capability of Internet of Things in India in Sect. 2. In Sect. 3, we have discussed Internet of Things in India. In Sect. 4, typical applications of IoT in India. In Sect. 5, challenges of Internet of Things in India, and in the last Sect. 6, we have given the conclusion (Fig. 1).

2 Opportunity, Status, and Capability of IoT in India

The opportunity, status, and capability of Internet of Things in India are discussed as follows.

2.1 Opportunity of IoT in India

Internet of Things is not a new concept but it is an extension of the old concepts along with the addition of new technologies. IoT can create huge opportunities in India.

It can provide ways that can reduce the work load of people by assigning certain specific tasks to mobile phones that can control a device over cloud and check their status. For example, a refrigerator that can inform the owner that it has run out of vegetables by sending a text message on the mobile phone.

IoT will also ensure increase in productivity and optimal utilization of resources and time. IoT can define a fine interaction between the physical and the virtual world by making them reliable so that the interaction becomes smooth and errorless. The development of many intelligent devices will generate a grid that will help the supply chains to interconnect in a better way [4].

India has an IoT policy document that enlists the IoT policy that India needs to adopt in order to compete in the race of digitization. Trillions of sensors work with billions of intelligent systems, which involve millions of applications attracting millions of consumers giving the direction to business. In the coming years, the digital highway will be analogous to water [5].

The Indian Government plans to achieve the establishment of 100 smart cities in the near future which will create many opportunities for the IT developers working in this field. The demonetization in the recent past with the aim of making India digitized in transactions is the first step towards the upliftment of IoT industry in India.

2.2 Status of IoT

IoT is going to create a new wave in the country and it has its use in every sector ranging from manufacturing, agriculture, electronics of daily use to home appliances, etc. India currently faces most of the socioeconomic challenges and lacks proper infrastructure for the development of IoT. By 2011, the count of devices connected by Internet (12.5 billion) has exceeded the count of human beings (7 billion) on the earth [3]. The Telecom sector of India accounts for 36% of the total revenue of the IoT industry. The electronics sectors accounts for 29% and the oil sector accounts for 23% of it. Hence, India owes 88% of the revenue to all of these industries mentioned above [6]. In India, there are professionals who have been working in this field since 8 years. Every year, a large number of fresh professionals are added to the IoT industry which contributes to its development. The companies like Tata in India have started to track the status and the apparent life of the truck with the help of sensors. There is scope for IoT in the insurance sector as well. The only problem that India needs to deal with is the internet, which is not a common phenomenon in India. The people in the rural areas still find it hard to handle and work with smart phones.

2.3 Capability of IoT

Internet of Things is relatively a new phenomenon for the Indians. IoT can be used to safeguard the data from the attack by the hackers. The IoT aims at providing smart cities, smart automobiles, and smart architectures and much more which is discussed later. There is so much that the IT-BPM companies need to do in order to compete with the world class companies like Google, Microsoft, and Apple Inc., etc. The current IT-BPM companies of India are working under the TCSs IoT center of excellence while collaborating with Intel, and taking up research initiatives on IoT under Infosys Laboratories and HCLs IoT incubation center with Microsoft. The Indian IT-BPM need to boast themselves up and move forward from just being mere partners with the global companies that work in the field of IoT [7].

3 India and IoT

The growth, building blocks, and the R&D plans of IoT in India have been discussed below.

3.1 Growth of IoT in India

The Indian Government aims at developing 100 smart cities in future and had allocated Rs. 7,060 crores in the Union Budget of 2014 for the massive expansion of IoT industry in the country. The introduction of the Digital India Program by the Government which aims at making the country digitized will also help in the progress of IoT industry in India. The Department of Electronics and Information Technology (DeitY) born from the Ministry of Communication and Information Technology makes it a separate government entity working in the expansion of IoT.

The DeitY intention is to transform India into a digitally vested society and knowledge economy [8]. The three key areas which come under the Digital India vision are as follows:

1. Digital Infrastructure as a Utility to Every Citizen—some of the key aspects of this vision are making high speed Internet, which is the most important requirement, private, and shareable space over public cloud space, ensuring the cybersecurity and enabling participation of people in financial space [8]
2. Governance and Services on Demand—some of the key aspects of this vision are making transactions digital and fast, services transformed digitally to ensure fast governance and easy business, real-time availability of services from platforms like mobile and online [8]
3. Digital Empowerment of Citizens—some of the key aspects of this vision are ensuring digital literacy and easy access of digital resources, the digital conve-

Broadband Highways Broadband throughout India and better delivery of same	e-Governance-Reforming Technology Better transaction using IT and readily availabile information	Electronic Manufacturing Zero imports is the goal of the government
Universal access to phones Aims at expanding network to make phones accessible	e-Kranti-Electronic Delivery Comprises of e-Healthcare, e-Police, e-Prosecution etc.	IT for Jobs Training and Teching people of rural areas to be job ready in IT sectors
Public Internet Access Program Aim to make post offices multiservice centres	Information for all Open source for all citizens to interact using portals like MyGov	Early Harvest Program To provide Wi-Fi to all universities and make email secure from hackers

Fig. 2 Components of digital India [8]

nience of certificates and documents, availability of digital services in Indian languages, and readiness of entitlements through cloud [8] (Fig. 2).

The Ministry of Electronics and Information Technology (MeitYs) vision is to develop connected and smart IoT based system for our country's economic, social, environmental and global needs [9]. The ministry also released a policy called the IoT policy which was revised in the year 2015. According to the policy, the key participants of the IoT enterprise will be the Citizens, the Government, and the Industry. Active participation of each of the enterprises will be required for the overall development of the IoT industry. There are some of the basic aspects of the MeitY that it needs to focus on in order to build a strong IoT environment in India.

The MeitY outlines some of the objectives which are listed below:

1. To create IoT Industry in the country of USD 15 billion by 2020. The connected devices will increase from 200 million to 2.7 billion by 2020. Assumption is that India will share 5−6% of the global IoT industry [9].
2. IoT specific skill set needs to be undertaken to attract global as well as national markets [9].
3. To start R&D for all the technologies required to give impetus to IoT [9].
4. To cultivate products for the IoT market that would be specifically for India [9].

3.2 Building Blocks of IoT

Here, are a few building blocks of IoT that lays the basic foundation of IoT in any network in any kind of environment (Fig. 3).

The building blocks mentioned above can be summed up in a nutshell as a step towards the betterment of IoT status in India. These blocks when used properly can
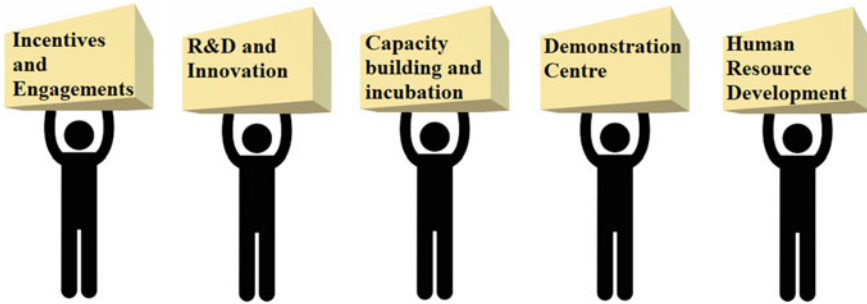
Fig. 3 Building blocks of IoT

be a great way to enhance the IoT quality and create a huge market for the same in the country.

3.3 R&D Plans

Government's IoT policy has enlisted some of the points for the development of R&D plans in the field of IoT and to fund them.

IIRC scheme has been initiated by the DeitY. Under the IIRC (International IoT Research Collaboration) scheme of the government, it will pledge agreements with other countries to invite projects for R&D. IIRC scheme will fund the IoT industry in the form of loans and will contribute 50% of the total cost. The organization appointed for the project will handle the whole process from starting till the end.

The main objectives of this plan are:

1. To create labs that would give impetus to R&D plans.
2. To recognize the R&D members in every field which aim at developing IoT.
3. To pledgee cloud-based source in order to help in the work of R&D.

This scheme came into picture with the objective of attracting investments from private sectors and has international partners. India needs to join hands with the international partners to ensure fast growth in this direction and for the betterment of the scheme.

In the light of growth of IoT industry in India, many startups have taken up the onus of ensuring proper development. Some of them are:

CARIQ: SMART DRIVING—It specializes in recording both mileage and speed, and also the driving pattern of the car.
SENSEGIZ—This sensor can be hooked to the thing that the owner is most likely to miss and using it can forget about losing it.
ENTRIB SHOPWORX—This makes the floor of a manufacturing industry smarter by giving a view about the manufacturing machines.

Table 1 Application of IoT

Fields	Typical applications
Smart water	It includes checking the quality of water in public places and timely check of the rise of the water level [11]
Smart environment	Projects aiming at raising alarms when the CO_2 level exceeds a certain value and raise alarms about seismic tremors [12]
Smart health	It is a smart network between the doctor and the patients [13]
Smart waste management	To setup project that would help in the "SWACH BHARAT" initiative [14]
Smart safety	Projects to build wearables that can ensure safety of people [15]
Smart logistics and supply chain	Ensuring timely availability of ambulance in case of emergency, and reduce in food wastage [16]
Smart manufacturing and industrial IoT	Its objective is to timely check the equipment in order to prevent leakages of any form and prevent from fire and gas leakages [17]

TEEWE—A USB device which one connected to the any device like laptop and smart phone can make it a remote control of the Television.

LIFEPLOT—This is the first handheld ECG devices that can do all the work without the requirement of a paper.

Lately, Andhra Pradesh becomes the first state in the country to pass an IoT policy of its own and has decided to become IoT hub of the country. NASSCOM has established a center for excellence to bring together all the nurturing talents in the field of IoT and help it grow.

Recently, Luminous Company that specializes in making inverters has collaborated with Airtel to make the customer experience better than before. Airtel helped them realize this dream by making an IoT solution that can help the customer in the optimization of the inverter and they can track its performance in real time [10].

4 Applications of IoT

In India, IoT applications development includes are shown in the Table 1.

5   Challenges of IoT in India

A standard structure is needed to deploy IoT in a sophisticated manner. The wireless network should be capable enough to encompass the traffic that is a mixture of both machine-type and human-type [18]. IoT helps to transfer data from devices to Internet to enable it to be analyzed but this poses a threat to the security of data [19].

The IoT is bringing several opportunities for research scholars, entrepreneurs, and manufacturers to grow but presently IoT requires some standard theory and architecture to amalgamate real and virtual worlds and to solve privacy and security issues. Some challenges are:

5.1   Architecture Challenge

IoT encompasses of escalating amount of interconnected objects or things which are embedded and autonomous. Communication among objects are anticipated to work anytime, everywhere for any purpose, these communication is expected to be wireless, automatic, uninterrupted, portable, distributed, and multifarious. Data integration of various systems is sturdy and will be handled by modular interoperable components. To improve decision-making and make useful sense of Big Data, deduce and draw relation between large volumes of data, a system solution with sophisticated infrastructure is required, that is why one architecture cannot be applicable to other applications. Flexible architecture is required to include diverse applications of IoT. While laying the architecture for IoT the system, designers have to take under consideration all the functional and nonfunctional requirements which might be a tough task, but the changing requirements make a mountain of the molehill when the complexity increases [20].

An architecture must be unwarped, unbounded, and having high standards, it must not bound user to use specific solutions. IoT architectures must be open to provide for cases of identifying, communicating, and processing through smart objects [21].

5.2   Technical Challenge

IoT technology is complicated for several reasons. To make IoT successful, it is important to connect as many things as possible. Size of IoT network becomes enormous due to all the connected things, and to interconnect different applications and networking technologies heterogeneous architecture is required. Different kinds of application will require different set of protocols, different security measures [22]. An economical network with reliable communication is required for connecting all the objects. All the connected objects will produce data in huge amount, first of all, a network which can handle the huge data is needed, then to process the huge

data and extract useful information we need new big data processing techniques, as conventional data processing techniques are unable to deal with the such amount of data and at last to store enormous data generated by the objects or things we require very large space. Cloud computing systems provides solution till some extant to store data out somewhere else [21].

5.3 Privacy and Security Challenge

The IoT will be interconnection of billions of devices. Its architecture can be divided in five layers. Large number of devices is out there communicating with each other wirelessly, which put a very large and important amount of information in open. Its architecture gives rise to many security issues that are discussed here and their solution will be new area of research [23].

Main security issues with Perception Layer is that sensors may get damaged, broken, and stolen. That is why it is beneficial to deploy sensors out of human reach and also ideally sensor should also be wireless, anti-radiation, and highly resistive from physical damage and have ability to work under temperature variations. Network layer executes the essential conception behind the Internet of Things, i.e., M2M communication, because exchange of information among the machines takes place in this layer, this is the reason why it is important to make this layer secure. There are possibility of illegal data access, confidential information leakage, and eavesdropping to manipulation of the data and other attacks in this layer. Highly secure communication is required to prevent unauthorized access and other possible attacks. To prevent unsafe data or virus coming from physical layer data filtration is required in this layer. Encryption of the data can be done to make this layer secure and safe. Hacking in application layer may lead to leak of personal information as well as confidential data which can lead to data tampering. Highly secure protocols are needed to solve or minimize such incidents. Smart access management system may help into stop unauthorized person to access the information. Encryption and decryption are the key to handle such issues.

In general, the wireless network depends on the router in order to communicate among themselves. While on the other hand, M2M communication does not as it is the one-to-one straight communication between two devices [24].

Software-Defined Networks (SDN), and more specifically, Software-Defined network Perimeters (SDP) are potent tool for the M2M IoT, and especially for emergency responders [25]. The IoT architecture would be made possible due to IPv6, which will provide global addressees to billions of devices taking part in M2M communication. Current approach for the security in networks is physical firewall. But in this approach, these devices will be left out in the open exposing them to the threats of internet hackers. But SDP can change all that. It will help in deployment of a network on-the-fly that can be reconfigured dynamically anytime and would be secure and Ad hoc in nature. Within an SDP logical network, M2M devices with global IP addresses can remain secure and hidden from the plain sight. It means a mobile and

secure M2M network can be created with SDP without investing too much in the physical network infrastructure. Minimum requirement would be Global IP addresses and Internet connectivity. Another advantage of creating this type of network by SDP is that it does not matter which technology is used for network communications [26].

5.4 Digital Literacy Challenge

Digital Literacy refers to the fact of evaluating, creating, and navigating information using the digital technologies [27]. In India, 40% of the total population lives below the poverty line and literacy rate is 25–30%, so talking about digital literacy stands nowhere in the frame. More than 90% of the population is below the digital literacy line. While the country is becoming the second largest in the world in terms of number of mobile users, it still lacks proper internet connectivity. Fast and high speed Internet in the country is the need of the hour, so that rate of digital progress in the country can increase.

NASSCOM and other private sectors have joined hands with the government to come up with National Digital Literacy Mission (NDLM) which aims at producing one person in every household who is illiterate. The Government has come up with projects like National Optical Fiber Network (NOFN) to extend the telecom infrastructure in the country. The four zones that need to focus are affordability, convenience, suitability, and acquaintance. In addition, it needs to be easily accessible and there needs to be some opportunities where this work can be put to display. Lastly, people need to be aware about the technology, so to ensure that there must be some awareness programs.

NDLM Program aims at spreading awareness about digital literacy in rural areas. NDLM will help the country to move a step forward in the digital economy.

However, the Ministry of Information Technology in India lays emphasis on the increase in production of M.Tech and Ph.D. level fellows in India [8]. While the need of the hour is to standardize the education of the engineering sectors to enhance the IoT development in India [28].

6 Conclusion

IoT technology is in toddler state in India, and the country has a long way to go. India needs to fasten its seat belt for the proper expansion of IoT. IoT will create a new generation of people that can lead a hassle-free life. IoT will generate a new field from the research scholars to the business class, which will lead to boast Indian economy and will contribute to the growth of its GDP. A proper utilization of resources is necessary to develop the proper infrastructure for the growth of IoT industry. This paper purpose is to provide the information about development of IoT in India that has been compiled with all the latest advancements in the country in this regard.

References

1. Stankovic, J.A.: Research directions for the internet of things. IEEE Internet Things J. 3–9 (2014)
2. Upadhyay, K., Dhanda, S.S., Rawat, S.: IoT: pillars and technology, in Indian, vol. 9, pp: 1–8 (2016)
3. Sarin, G.: Developing smart cities using internet of things: an empirical study. In: IEEE International Conference on Computing for Sustainable Global Development (INDIACom), pp. 315–320 (2016)
4. Singh, N., Singh, S.: Internet of Things (IoT): Security challenges, business opportunities & reference architecture for E-commerce. In: IEEE Green Computing and Internet of Things (ICGCIoT), 2015 International Conference
5. IoT reshaping 'The Society' (2016)
6. http://IoTindiamag.com/2016/09/IoT-india-industry-study-2016/, IoT India Industry Study (2016)
7. http://blogs.nasscom.in/3-things-indian-it-companies-needs-to-do-to-make-money-from-IoT/, Achyuta Ghosh, 12 Aug 2016
8. Digital India: Department of Electronics & Information Technology [DeitY]
9. IoT Policy: Ministry of Electronics and Information Technology [MeitY]
10. Mashable India Homepage. https://mashable.com/2017/07/28/internet-of-things-india/#qewp2Xfd4aq3. Accessed 08 Feb 2018
11. Cherukota, N., Jadhav, S.: Architectural framework of smart water meter reading system in IoT environment. In: IEEE International Conference on Communication and Signal Processing, 6–8 April 2016
12. Islam, T., Mukhopadhyay, S.C., Suryadevara, N.K.: IEEE "smart sensors and internet of things: a postgraduate paper". IEEE Sens. J. **17**(3), 578–584 (2017)
13. Thaduangta, B., Choomjit, P., Mongkolveswith, S., Supasitthimethee, U., Funilkul, S., Triyason, T.: Smart healthcare: basic health check-up and monitoring system for elderly. In: 2016 International IEEE Computer Science and Engineering Conference (ICSEC)
14. Yeow, W.L., Low, Y.S., Folianto, F.: Smartbin: smart waste management system. In: IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), pp. 1–2 (2015)
15. Sarishma, R.T., Singhal, A.: Intelligent accident management system using IoT and cloud computing. In: IEEE 2nd International Conference on Next Generation Computing Technologies (NGCT-2016), pp. 89–92, Oct 2016
16. Blirwald, W., Baumann, S., Keil, R., Leitzke, C., Richter, K.: "Smart logistics"—usage of innovative information and communication technologies in production logistics in IEEE RFID systems and technologies (RFID SysTech) (2007)
17. Lin, Y., Ieromonachou, P., Sun, W.: Smart manufacturing and supply chain management. In: IEEE logistics, Informatics and Service Sciences (LISS), 2016 International Conference (2016)
18. Bader, A., ElSawy, H., Gharbieh, M., Alouini, M.-S., Adinoyi, A., Alshaalan, F.: First mile challenges for large-scale IoT. In: IEEE Communications Magazine, pp. 138–144, March 2017
19. Singh, N., Singh, S.: Internet of things (IoT): security challenges, business opportunities & reference architecture for E-commerce. In: IEEE International Conference on Green Computing and Internet of Things (ICGCIoT), pp. 1577–1581 (2015)
20. Banakar, R.M., Malagund, K.B., Mahalank, S.N.: Non functional requirement analysis in IoT based smart traffic management system. In: IEEE (2016)
21. Chen, S., Xu, H., Liu, D., Hu, B., Wang, H.: A vision of IoT: applications, challenges, and opportunities with China perspective. IEEE Internet Things J. **1**(4), 349–359 (2014)
22. Serbanati, A., Medaglia, C.M., Ceipidor, U.B.: Building blocks of the internet of things: state of the art and beyond, deploying RFID—challenges, solutions, and open issues. In: InTech, pp. 351–366 (2011)

23. Gou, Q., Liu, Y., Yan, L., Li, Y.: Construction and strategies in IoT security system. In: IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing, pp. 1129–1131 (2013)
24. Chen, H.C., Al Faruque, M.A., Chou, P.H.: Security and privacy challenges in IoT-based machine-to-machine collaborative scenarios. IEEE 16 Oct 2016
25. Aijaz, A., Aghvami, A.-H.: Cognitive machine-to-machine communications for internet-of-things: a protocol stack perspective. In: IEEE Internet Things J. **2**(2), 103–112 (2015)
26. Balfour, R.E.: Building the "internet of everything" (IoE) for first responders. In: 2015 Long Island Systems, Applications and Technology, pp. 1–6. Farmingdale, NY (2015)
27. Dutta, S., Mathur, R.: EPICS high: digital literacy project in India. In: IEEE Integrated STEM Education Conference, March 8 2014
28. Wani, P.W., Patki, A.B., Kulkarni, C.D., Joshi, R.D.: Emerging trends in IoT standards and legislation. In: IEEE International Conference on Internet of Things and Applications (IOTA), pp. 373–378, 22–24 Jan 2016

Numerical Study of Water-Based Carbon Nanotubes' Nanofluid Flow over a Nonlinear Inclined 3-D Stretching Sheet for Homogeneous–Heterogeneous Reactions with Porous Media

Shalini Jain and Preeti Gupta

Abstract A numerical investigation of water-based CNTs' nanofluid flow over an inclined 3-D nonlinear stretching sheet with homogeneous–heterogeneous reactions in the presence of porous media subject to convective boundary condition has been done. In this paper, we have considered two types of CNTs; one is single-walled carbon nanotube and the second is multi-walled carbon nanotube. By using suitable similarity governing equations, the equations are converted into ordinary differential equations. The reduced equations are solved numerically by Runge–Kutta fourth-order method with shooting technique. The influence of the pertinent parameters on different profiles like velocity, concentration, and temperature are discussed and presented graphically.

Keywords CNTs · Nonlinear inclined stretching sheet · Porous media
Homogeneous–heterogeneous reaction

1 Introduction

Among the engineers and scientists, the hot topic of research is new energy resources in order to achieve the energy demand. In many industrial processes, nanofluids are frequently used due to its low thermal conductivity. Nanofluids were introduced by Choi et al. [1] and he investigated solid nano-sized particles dropped in a carrier fluid, and hence we get a new type of complex fluid, which is called nanofluid. The structure of regular fluid is simple compared to nanofluid. It is due to the shape of the nanoparticle. Many researchers Hayat et al. [2] and Jain and Choudhary [3] studied the impact of Marangoni convection in viscous fluid flow of carbon water nanofluid.

S. Jain (✉)
Department of Mathematics, University of Rajasthan, JLN Marg, Jaipur, India
e-mail: drshalinijainshah@gmail.com

P. Gupta
Department of Mathematics & Statistics, Manipal University Jaipur, Jaipur 303007, India
e-mail: pritimbgupta@gmail.com

633

Aman et al. [4] studied four different types of molecular liquids, which are taken with CNTs' Maxwell nanofluids in free-convection flow. Hayat et al. [5] discussed three-dimensional flow for homogeneous–heterogeneous reactions of carbon nanotubes in the presence of porous media.

Most of the chemical reacting systems include homogeneous–heterogeneous reactions, for instance, in combustion, biochemical system, and catalysis. The connection between homogeneous–heterogeneous reactions is exceptionally complex. The chemical reaction is used in fog formation and dispersion, ceramics and polymer production, food processing, and is also involved in the consumption and creation of reactant species. Merkin [6] analyzed the catalyst surface, heterogeneous reaction, and homogeneous reaction for cubic autocatalysis, and introduced that the surface reaction is dominant near the surface. Kameswaran et al. [7] and Hayat et al. [8] discussed boundary-layer flow of homogeneous–heterogeneous reactions over a shrinking and stretching sheet in a nanofluid flow with a porous medium. Many researchers Sheikh and Abbas [9], Bachok et al. [10] studied homogeneous–heterogeneous reactions over a sheet for the stagnation point flow.

A numerical study of boundary-layer flow through a stretching surface has attracted many researchers due to its wide applications in industry and technology. This type of flow has applicability in extrusion of plastic sheets, glass fiber production, wire drawing, paper production, and many more. In technological and industrial processes, various situations are obtained where the stretching surface may not be unavoidably linear. Hayat et al. [11] and Jain and Bohra [12] studied three-dimensional magnetohydrodynamic nanofluid flow with a convective boundary condition over a stretching nonlinear surface. Mustafa et al. [13] discussed water-based magnetite nanofluid's rotating flow over a stretching surface by nonlinear thermal radiation. Gopal et al. [14] studied Joule's and viscous dissipation on Casson fluid flow with inclined magnetic field over a chemical reacting stretching sheet. Kandasamy et al. [15] discussed MHD SWCNT-type nanofluid flow through the base fluid, water, and seawater due to thermal radiation energy.

Convective boundary condition denotes the heat transfer rate through the surface. These conditions are directly proportional to local differences in temperature. We can say that a convective boundary condition increases the temperature and the thermal conductivity of nanofluids. So, the convective boundary condition is more suitable for nanofluid compared to constant temperature conditions. Resultant convective boundary condition appeared in terms of Biot number. Mahanthesh et al. [16] studied radiative heat transfer in three-dimensional MHD nanofluid flow subject to convective boundary condition over a nonlinear stretching sheet. Nayak [17] studied heat transfer investigation of nanofluid flow by MHD shrinking surface with convective condition. Matin et al. [18] studied entropy in mixed convection MHD nanofluid flow over a sheet. Das et al. [19] studied entropy analysis of unsteady nanofluid flow past stretching sheet with convective condition. Jain and Parmar [20] studied MHD Williamson fluid flow over stretching cylinder through porous media.

The aim of this investigation is to explore the characteristics of homogeneous–heterogeneous reactions in the three-dimensional inclined stretching sheet of water-based nanofluid CNTs embedded in porous medium. Here, nonlinear inclined
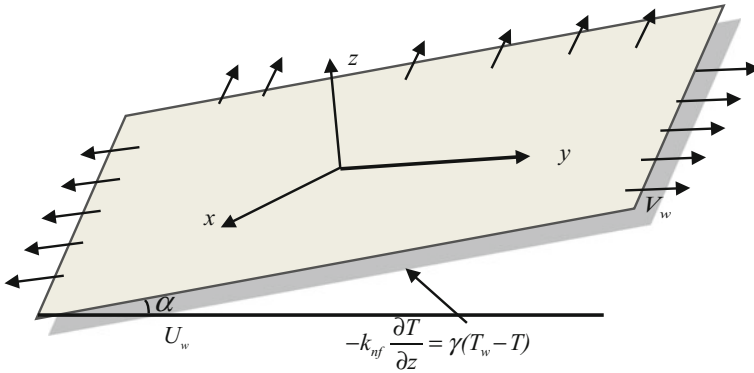
Fig. 1 Schematic diagram

stretching sheet is modeled and both SWCNTs- and MWCNTs-type nanoparticles are considered. Diffusion coefficients of both species are considered equal. Convergent series solution is obtained by shooting method. The effect of pertinent parameters on velocity, concentration, temperature, Nusselt number, and skin friction is discussed and analyzed through figures and tables.

2 Formulation of the Problem

Consider water-based carbon nanotubes over a nonlinear inclined 3-D stretching sheet with homogeneous–heterogeneous reactions through porous media under the convective boundary condition. Velocities along x- and y-directions are $U_w(x, y) = c(x + y)^n$, $V_w(x, y) = d(x + y)^n$, where $n > 0$, and c and d are positive constants. Homogeneous and heterogeneous equations with two chemical species A and B have been considered. The cubic catalysis in homogeneous reaction is (Fig. 1)

$$A + 2B \rightarrow 3B, \quad \text{rate} = k_c ab^2, \tag{1}$$
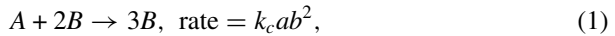
while the catalyst surface on heterogeneous reaction is

$$A \rightarrow B, \quad \text{rate} = k_s a, \tag{2}$$

where k_c and k_s are rate constants, and a and b are concentrations of chemical species A and B. The governing equations are given as follows:

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0, \tag{3}$$

Table 1 Thermophysical properties of base fluid and nanoparticles

Physical characteristics	Water (base fluid)	SWCNT (nanoparticle)	MWCNT (nanoparticle)
ρ (kg/m^3)	997.1	2600	1600
C_p (J/kg K)	4179	425	796
k (W/mK)	0.613	6600	3000

$$u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + w\frac{\partial u}{\partial z} = \upsilon_{nf}\frac{\partial^2 u}{\partial z^2} - \frac{\upsilon_{nf}}{k_p}u + g[\beta_T(T - T_\infty)]\cos\alpha, \qquad (4)$$

$$u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + w\frac{\partial v}{\partial z} = \upsilon_{nf}\frac{\partial^2 v}{\partial z^2} - \frac{\upsilon_{nf}}{k_p}v + g[\beta_T(T - T_\infty)]\sin\alpha, \qquad (5)$$

$$u\frac{\partial a}{\partial x} + v\frac{\partial a}{\partial y} + w\frac{\partial a}{\partial z} = D_A\frac{\partial^2 a}{\partial z^2} - k_cab^2, \qquad (6)$$

$$u\frac{\partial b}{\partial x} + v\frac{\partial b}{\partial y} + w\frac{\partial b}{\partial z} = D_B\frac{\partial^2 b}{\partial z^2} + k_cab^2, \qquad (7)$$

$$u\frac{\partial T}{\partial x} + v\frac{\partial T}{\partial y} + w\frac{\partial T}{\partial z} = \frac{k_{nf}}{(\rho c_p)_{nf}}\frac{\partial^2 T}{\partial z^2} - \frac{1}{(\rho c_p)_{nf}}\frac{\partial q_r}{\partial Z}, \qquad (8)$$

where u, v, w are fluid velocities along the directions x, y, z, respectively; D_A and D_B are diffusion coefficients; k_p is the permeability of porous medium; β_T is the thermal expansion.

Under the boundary conditions

$$z = 0, \ u = c(x + y)^n, \ v = d(x + y)^n, \ w = 0, \ D_A\frac{\partial a}{\partial z} = k_sa, \ D_B\frac{\partial b}{\partial z} = -k_sa,$$

$$-k_{nf}\frac{\partial T}{\partial z} = \gamma(T_w - T), \qquad (9)$$

at

$$z \to \infty, \ u \to 0, \ v \to 0, \ a \to a_0, \ b \to 0, \ T \to T_\infty, \qquad (10)$$

where μ_{nf}, υ_{nf}, ϕ, ρ_{nf}, ρ_{CNT} are the viscosity of nanofluid, kinematic viscosity of nanofluid, volume fraction of nanoparticle, density of nanofluid, and density of carbon nanotube, respectively. Thermophysical properties of fluid and nanoparticles are given in Table 1.

$$\mu_{nf} = \frac{\mu_f}{(1 - \phi)^{2.5}}, \ \upsilon_{nf} = \frac{\mu_{nf}}{\rho_{nf}}, \ \rho_{nf} = (1 - \phi)\rho_f + \phi\rho_{CNT},$$

$$(\rho c_p)_{nf} = (1 - \phi)(\rho c_p)_f + \phi(\rho c_p)_{CNT}. \qquad (11)$$

The effective thermal conductivity of nanofluid is expressed as

$$\frac{k_{nf}}{k_f} = \left(\frac{1 - \phi + 2\phi \left(\frac{k_{CNT}}{k_{CNT} - k_f} \right) \ln \left(\frac{k_{CNT} + k_f}{2k_f} \right)}{1 - \phi + 2\phi \left(\frac{k_f}{k_{CNT} - k_f} \right) \ln \left(\frac{k_{CNT} + k_f}{2k_f} \right)} \right). \tag{12}$$

The Rosseland approximation is expressed as

$$q_r = -\frac{4\sigma^*}{3k^*} \frac{\partial T^4}{\partial z}, \tag{13}$$

where σ^* is the Stefen–Boltzmann constant and k^* is the mean absorption coefficient. The temperature difference has been considered very small, so that T^4 may be expressed as a linear function of temperature.

$$T^4 \approx 4T_\infty^3 T - 3T_\infty^4 \tag{14}$$

Similarity transformations are

$$u = c(x + y)^n f'(\eta), \quad v = c(x + y)^n g'(\eta), \quad b = a_0 h(\eta), \quad a = a_0 \varphi(\eta),$$

$$T - T_\infty = (T_w - T_\infty)\theta, \quad \eta = \left(\frac{c(n+1)}{2\upsilon_f} \right)^{1/2} (x + y)^{\frac{n-1}{2}} z$$

$$w = -\left(\frac{c\upsilon_f(n+1)}{2} \right)^{1/2} (x + y)^{\frac{n-1}{2}} \left\{ (f + g) + \frac{n-1}{n+1} \eta (f' + g') \right\}, \tag{15}$$

on substituting Eq. (15) in Eqs. (3)–(8), Eq. (3) identical satisfies and Eqs. (4)–(8) are transformed into the following form:

$$f''' + (1 - \phi)^{2.5} A_1 \left\{ (f + g)f'' - \frac{2n}{n+1}(f' + g')f' + \frac{2}{n+1} \delta \cos \alpha_1 \theta \right\} - \frac{2K}{n+1} f' = 0, \tag{16}$$

$$g''' + (1 - \phi)^{2.5} A_1 \left\{ (f + g)g'' - \frac{2n}{n+1}(f' + g')g' + \frac{2}{n+1} \delta \sin \alpha_1 \theta \right\} - \frac{2K}{n+1} g' = 0, \tag{17}$$

$$\frac{1}{Sc} \varphi'' + (f + g)\varphi' - \frac{2}{n+1} k_1 \varphi h^2 = 0, \tag{18}$$

$$\frac{\delta}{Sc} h'' + (f + g)h' + \frac{2}{n+1} k_1 \varphi h^2 = 0, \tag{19}$$

$$\frac{1}{A_2 \Pr} \left(A_3 + \frac{4}{3} R \right) \theta'' + (f + g)\theta' = 0, \tag{20}$$

under the boundary condition at

$$\eta = 0 \, f(0) = g(0) = 0, \, f'(0) = 1, \, g'(0) = \alpha, \, \varphi'(0) = k_2 \varphi(0),$$

$$\delta_1 h'(0) = -k_2 \varphi(0), \, \theta(0) = 1 + \frac{A_3}{B_i} \theta'(0),$$

at

$$\eta \to \infty \quad f'(\infty) \to 0, \quad g'(\infty) \to 0, \quad \varphi(\infty) \to 1 \quad h(\infty) \to 0 \ \theta(\infty) \to 0,$$
$$\tag{21}$$

where K is the local porosity parameter, α is the ratio parameter, Sc is the Schmidt number, δ is the ratio of diffusion coefficient, k_1 and k_2 are the homogeneous and heterogeneous strength, R is the radiation parameter, and Pr is the Prandtl number. These nondimensional variables are defined by

$$K = \frac{\upsilon_f}{k_p c(x+y)^{n-1}}, \quad \alpha = \frac{d}{c}, \quad Sc = \frac{\upsilon_f}{D_A}, \quad \delta_1 = \frac{D_B}{D_A}, \quad Pr = \frac{\mu_f (c_p)_f}{k_f}, \quad R = \frac{4\sigma^* T_\infty^3}{k^* k_f}$$

$$k_1 = \frac{k_c a_0^2}{c(x+y)^{n-1}}, \quad k_2 = \frac{k_s}{D_A} \sqrt{\frac{2}{n+1}} \sqrt{\frac{\upsilon_f}{c(x+y)^{n-1}}}. \tag{22}$$

Assuming,

$$A_1 = (1-\phi) + \phi \frac{\rho_{CNT}}{\rho_f}, \quad A_2 = (1-\phi) + \phi \frac{(\rho C_p)_{CNT}}{(\rho C_p)_f}, \quad A_3 = \frac{k_{nf}}{k_f}.$$

Also, assuming that both the chemical species have equal diffusion coefficients D_A and D_B, i.e., $\delta_1 = 1$, thus

$$\varphi(\eta) + h(\eta) = 1. \tag{23}$$

Now, Eqs. (18) and (19) become

$$\frac{1}{Sc} \varphi'' + (f+g)\varphi' - \frac{2}{n+1} k_1 \varphi (1-\varphi)^2 = 0, \tag{24}$$

under the boundary conditions

$$\varphi'(0) = k_2 \varphi(0), \quad \varphi(\infty) \to 1. \tag{25}$$

Skin friction coefficients in dimensionless forms are given by

$$C_{fx} Re_x^{1/2} = \left(\frac{n+1}{2}\right)^{1/2} \left(\frac{1}{A_1 (1-\phi)^{2.5}}\right) f''(0) \tag{26}$$

$$C_{fy} Re_y^{1/2} = \left(\frac{n+1}{2}\right)^{1/2} \left(\frac{\alpha^{-3/2}}{A_1 (1-\phi)^{2.5}}\right) g''(0), \tag{27}$$

where $Re_x = \frac{U_w(x+y)}{\upsilon_f}$ and $Re_y = \frac{V_w(x+y)}{\upsilon_f}$ depict the local Reynolds numbers.

Nusselt number in dimensionless form is given by

$$Nu = -A_3 (Re_x)^{1/2} \left(\frac{n+1}{2}\right) \theta'(0)$$

3 Numerical Method

Equations (16), (17), (20), and (24) under the boundary conditions (21) and (25) have been solved numerically by fourth-order Runge–Kutta method with shooting technique. Runge–Kutta fourth-order method needs a finite domain $0 \leq \eta \leq \eta_\infty$. In this study, we have chosen $\eta_\infty = 10$. The boundary value problem is changed into initial value problem, which is defined as

$$f_3' = -\left\{(1-\phi)^{2.5}A_1\left\{(f_1+f_4)f_3 - \frac{2n}{n+1}(f_2+f_5)f_2 + \frac{2}{n+1}\delta\cos\alpha_1 f_{10}\right\} - \frac{2K}{n+1}f_2\right\}$$

$$f_6' = -\left\{(1-\phi)^{2.5}A_1\left\{(f_1+f_4)f_6 - \frac{2n}{n+1}(f_2+f_5)f_5 + \frac{2}{n+1}\delta\sin\alpha_1 f_{10}\right\} - \frac{2K}{n+1}f_5\right\}$$

$$f_8' = -Sc\left\{(f_1+f_4)f_8 - \frac{2}{n+1}k_1 f_7(1-f_7)^2\right\}$$

$$f_{10}' = \frac{A_2\,\mathrm{Pr}}{\left(A_3 + \frac{4}{3}R\right)}\{-(f_1+f_4)f_{10}\}.$$

Under the boundary condition

$$f_1(0) = f_4(0) = 0,\ f_2(0) = 1,\ f_5(0) = \alpha,\ f_9(0) = 1 + \frac{A_3}{B_i}f_{10}(0),\ f_8(0) = k_2 f_7(0),$$

$$f_3(0) = r_1, f_6(0) = r_2, f_8(0) = r_3, f_{10}(0) = r_4,$$

where r_1, r_2, r_3, and r_4 are the initial guesses.

4 Results and Discussion

The numerical results have been obtained for SWCNTs and MWCNTs. The effects of pertinent parameters like ratio parameter α, power law index n, nanoparticle volume fraction ϕ, local porosity parameter K, heterogeneous reaction parameter k_2, homogeneous reaction parameter k_1, Schmidt number Sc, radiation parameter R, Prandtl number Pr, Biot number Bi, and Brinkman number Br on the velocity, concentration, and temperature are calculated for fixed values which are as follows:

$\alpha_1 = \frac{\pi}{4}$, $\phi = 0.2$, $K = 0.1$, $\alpha = 0.3$, $Sc = 0.7$, $k_1 = 0.2$, $k_2 = 0.1$, $R = 0.2$, $Bi = 0.5$, $\delta = 1$, $n = 1$, and are presented through graphs and Tables 2 and 3.

Figure 2 depicts that an increment in local porosity parameter K shows decreasing trend for velocity profile and boundary layer thickness for both CNTs. Figure 3 illustrates that how the velocity profile is affected with the variation of volume fraction ϕ. Both momentum layer thickness and velocity are enhanced for larger volume fraction nanoparticles in cases of CNTs.

Table 2 The evaluation of the values of $f''(0)$ and $g''(0)$ with that of Junaid et al. and Mahanthesh et al. on taking $n = 1$, $\phi = 0$, $\alpha_1 = 0$, and $K = 0$ in this study

α	ϕ	Junaid et al. [21]		Mahanthesh et al. [16]		Present study	
		$f''(0)$	$g''(0)$	$f''(0)$	$g''(0)$	$f''(0)$	$g''(0)$
0	0	-1	0	-1	0	-1	0
0.5	0	-1.2247	-0.6123	-1.22474	-0.6123	-1.224745	-0.612372
1.0	0	-1.4142	-1.4142	-1.41421	-1.4142	-1.414213	-1.414213

Table 3 The numerical values of skin friction and Nusselt number at $n = 3$

α	ϕ	C_{fx}		C_{fy}		Nu	
		(SWCNT) (Water)	(SWCNT) (Water)	(SWCNT) (Water)	(SWCNT) (Water)	(SWCNT) (Water)	(SWCNT) (Water)
0.3	0.2	-6.06335	-6.612145	-10.59195	-11.612281	0.6151086	0.614676
0.7		-6.92344	-7.543509	-8.229895	-8.9728997	0.6251615	0.624852
1.5		-8.38560	-9.126955	-6.863624	-7.4681948	0.6380758	0.637859
2.0		-9.18206	-9.989833	-6.511089	-7.0814460	0.6435497	0.643360
	0.0	-5.24228	-5.242282	-9.183285	-9.1832854	0.5491658	0.549165
	0.2	-6.06335	-6.612145	-10.59195	-11.612281	0.6151086	0.614676
	0.4	-7.89321	-9.182844	-10.59195	-11.612281	0.6334274	0.635063

Fig. 2 Velocity for different values of K
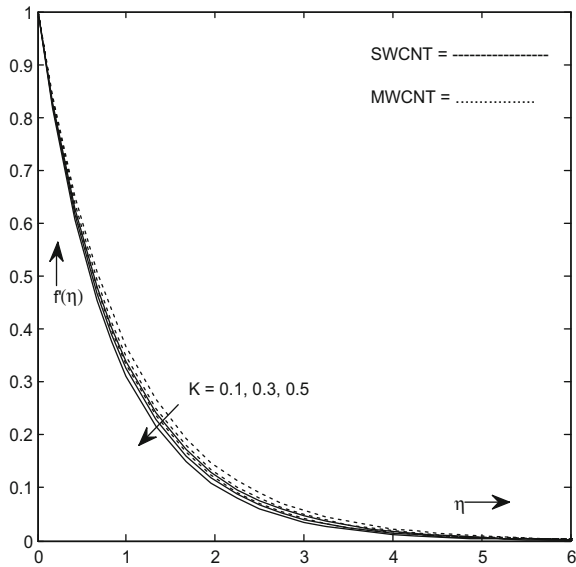
Figures 4 and 5 show the variation of stretching ratio parameter on the axial velocity and transverse velocity. Figure 4 shows that an enhance in stretching ratio parameter leads to the reduction of the boundary layer thickness along the x-direction (axial), whereas the reverse effect has been obtained for the y-direction (transverse).

Fig. 3 Velocity for different values of ϕ
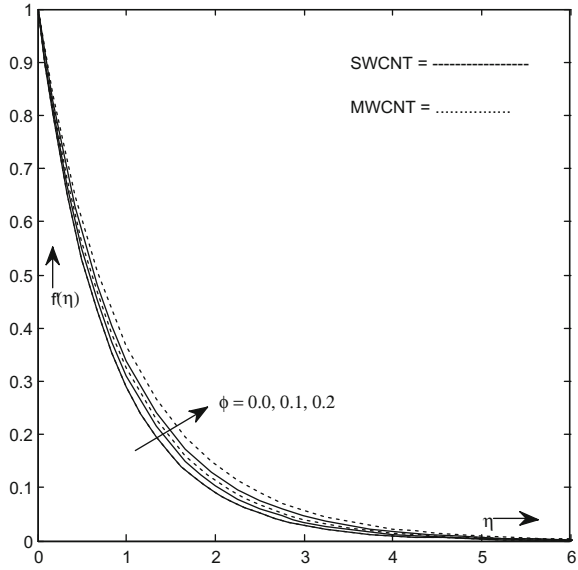
Fig. 4 Velocity for different values of α
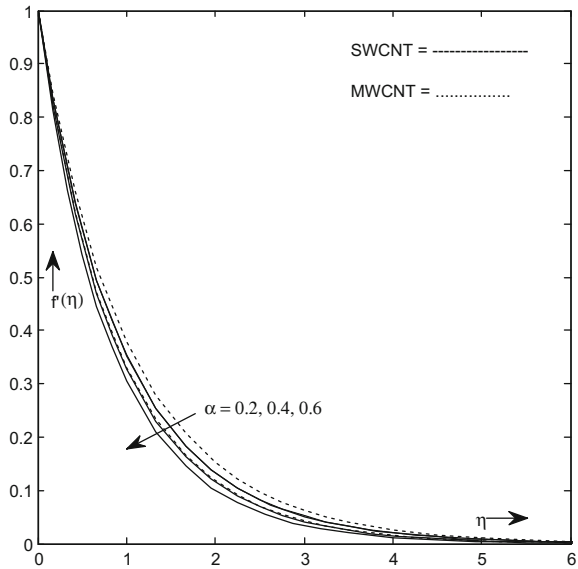
Figure 6 presents the variation of velocity profile $g'(\eta)$ affected with the volume fraction nanoparticle. Both velocity profile and related layer thickness are increased when nanoparticle volume fraction increases for both CNTs. Figure 7 shows $g'(\eta)$ via K. Larger K presents decreasing trend for velocity profile and layer thickness for both CNTs.

Fig. 5 Velocity for different values of α
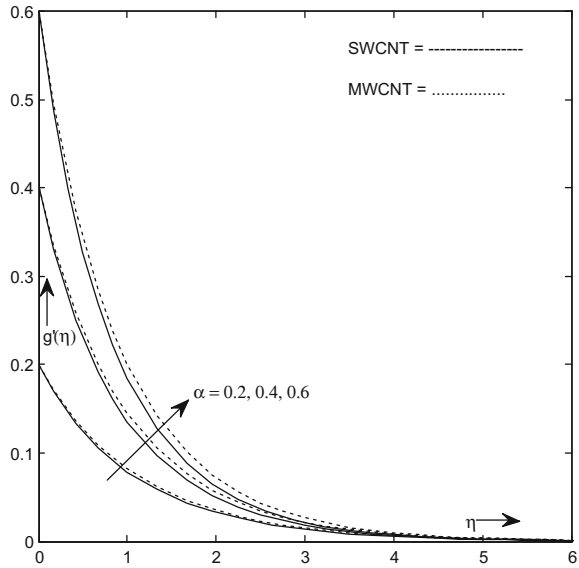
Fig. 6 Velocity for different values of ϕ
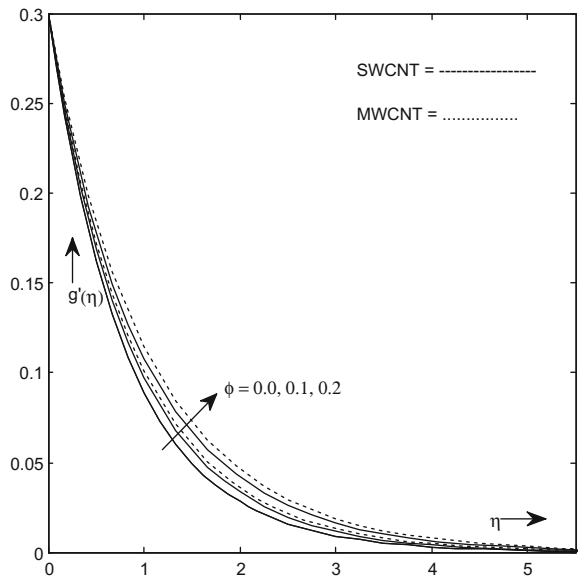
Figures 8 and 9 demonstrate the variation of homogeneous reaction parameter k_1 and heterogeneous reaction parameter k_2 on the concentration profile. It is noted that the increase in homogeneous reaction parameter k_1 and heterogeneous reaction parameter k_2 decays the concentration profile for both CNTs.

Fig. 7 Velocity for different values of K
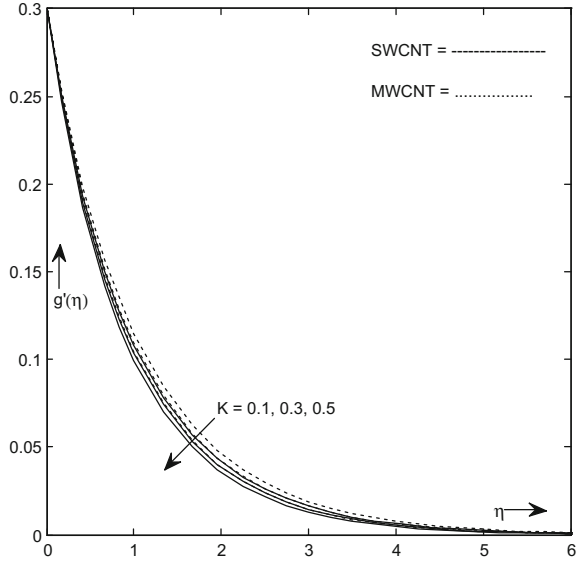
Fig. 8 Concentration for different values of k1
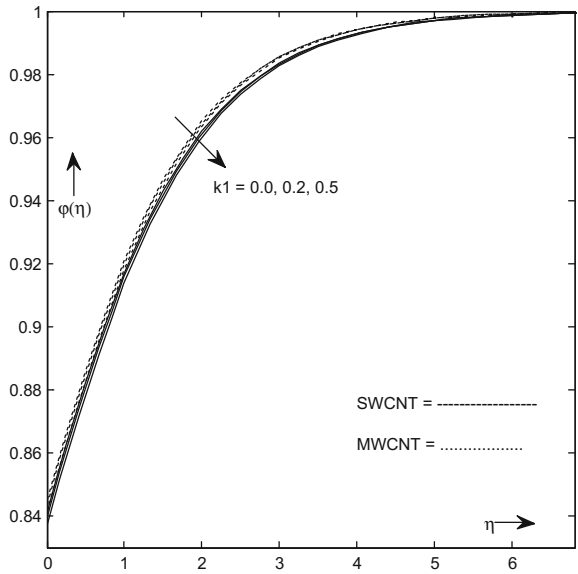
Figures 10 and 11 plot the concentration profile $\varphi(\eta)$ for the Schmidt number Sc and porosity parameter K. Higher values of porosity parameter and Schmidt number Sc constitute a larger concentration profile and for both CNTs more concentration layer thickness.

Fig. 9 Concentration for
different values of k2
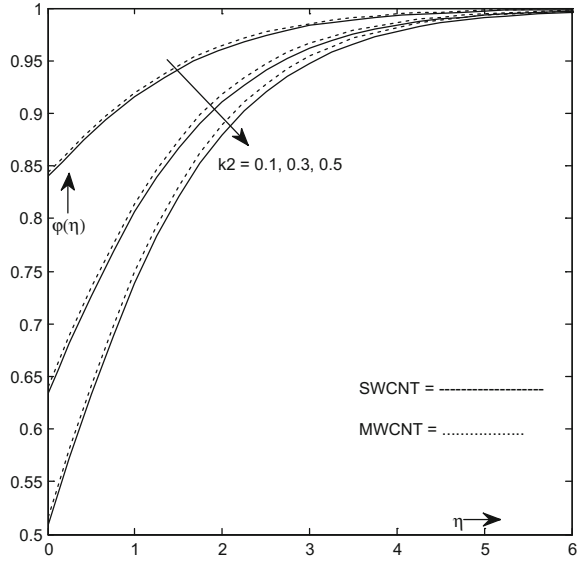
Fig. 10 Concentration for
different values of Sc
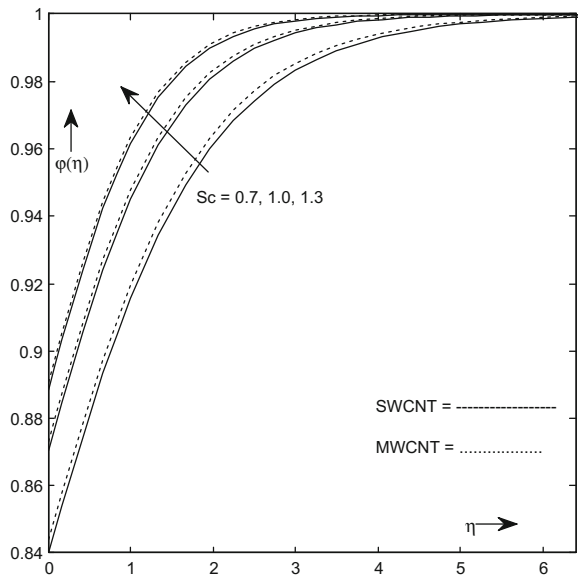
Figure 12 plots the concentration profile of the ratio parameter α. High value of
ratio parameter α constitutes a large concentration profile and more concentration
layer thickness for both CNTs. Figure 13 shows the impact of Biot number Bi on
the temperature distribution. An increment in the Biot number relates to stronger

Fig. 11 Concentration for different values of K
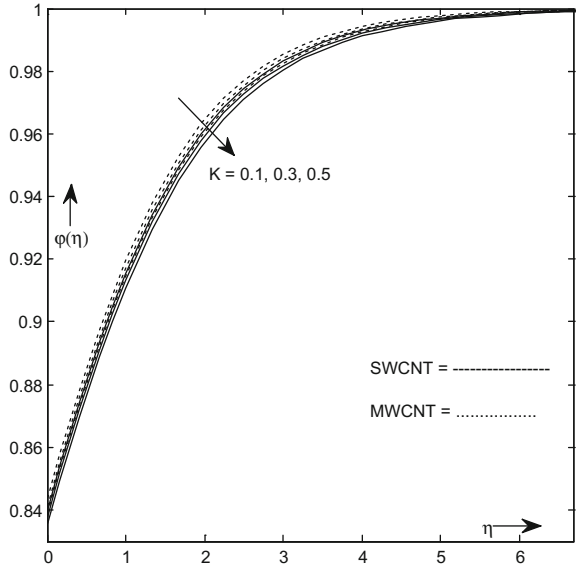
Fig. 12 Concentration for different values of α
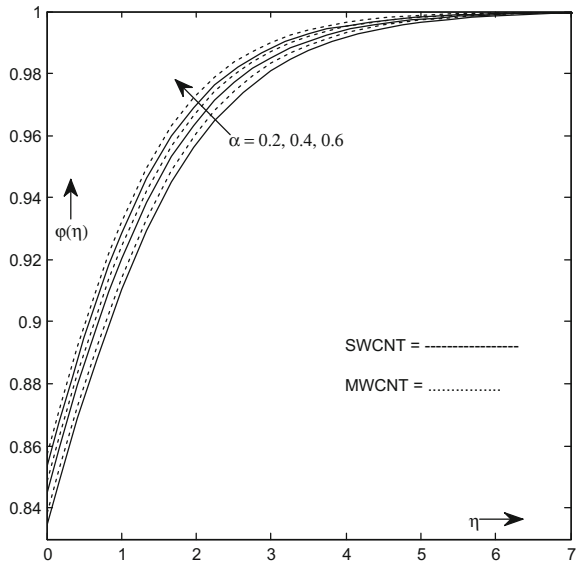
convection, which illustrates the higher temperature distribution and more thermal layer thickness for both CNTs.

Figure 14 is designed to compare both nonlinear and linear thermal radiation influences on thermal boundary layer. The black lines relate to linear thermal radiation and blue lines correspond to nonlinear thermal radiation. From this figure, the temperature in the boundary layer rises with an increase in radiation parameter. Further,

Fig. 13 Temperature for different values of Bi
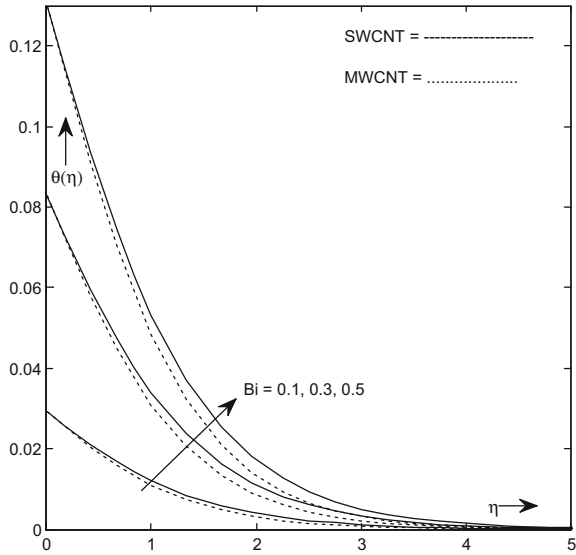
Fig. 14 Temperature for different values of ∝
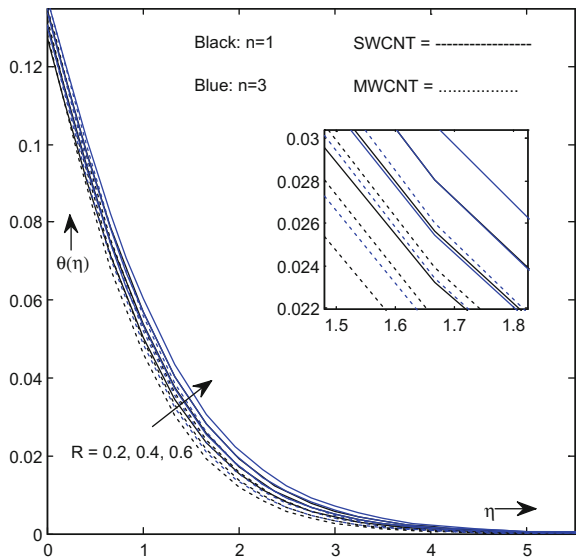
the temperature profile is lower for linear thermal radiation as compared with non-linear thermal radiation. Thus, we conclude that the nonlinear thermal radiation is more suitable for heating processes. Figure 15 describes that the temperature profile in the boundary layer is an increasing function of volume fraction parameter φ.

Figures 16 and 17 are plotted to inspect that how the temperature profile is affected by the ratio parameter and the porosity parameter. Figure 16 shows that the temper-

Fig. 15 Temperature for different values of φ
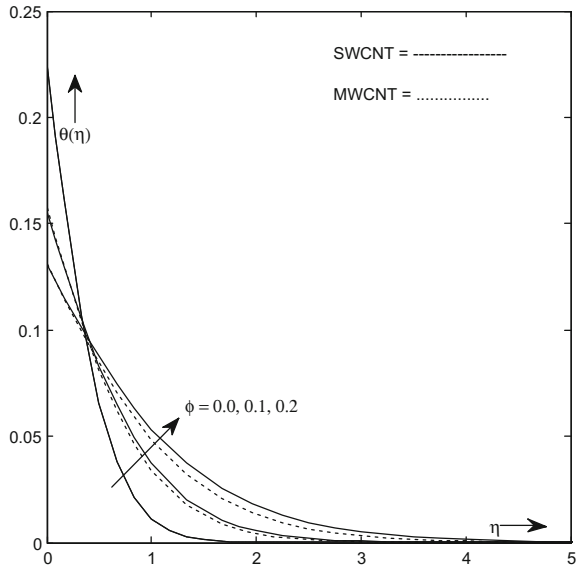
Fig. 16 Temperature for different values of α
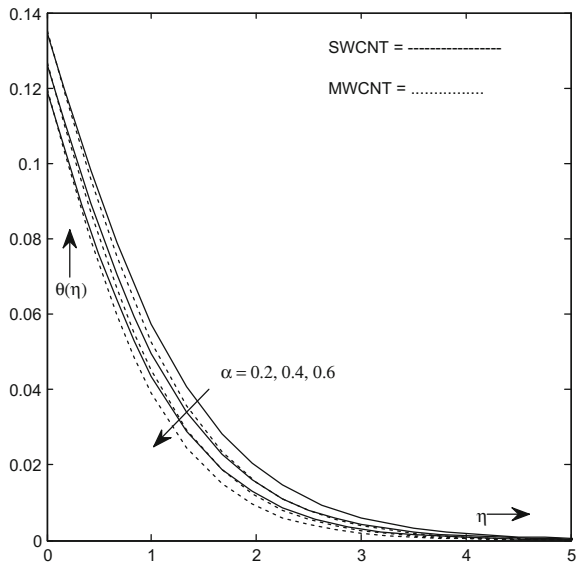
ature decreases when the ratio parameter increases. In Fig. 17, both the temperature profile and related layer thickness are increased when the porosity parameter increases for both CNTs.
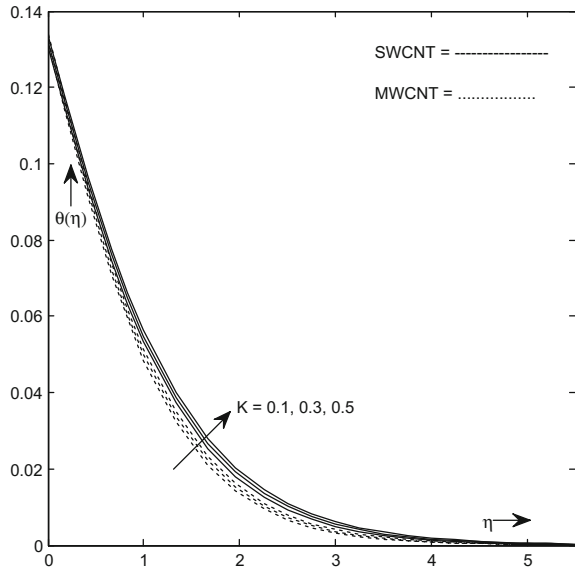
5 Conclusion

Homogeneous–heterogeneous reactions of water-based carbon nanotubes' nanofluid
flow over a nonlinear inclined 3-D stretching sheet embedded in porous media are
studied. The effects of different physical parameters in different flow fields were
examined. The major points are given as follows:

- Effects on concentration profile field of homogeneous and heterogeneous reaction
  parameters are quite the same.
- Velocity components decrease for higher values of local porosity parameter K.
- The nonlinear thermal radiation has high impact on flow fields compared with
  linear thermal radiation.
- By varying the Biot number and the nanoparticle volume fraction, could be con-
  trolled easily.
- Both the velocity components are higher for MWCNTs as compared with SWC-
  NTs.

References

1. Choi, S.U.S., Zhang, Z.G., Yu, W.: Anomalous thermal conductivity enhancement in nanotube
   suspension. Appl. Phys. Lett. **231**, 99–105 (1995)
2. Hayat, T., Khan, M.I., Farooq, M., Alsedi, A., Yasmeen, T.: Impact of Margoni convection
   in the flow of carbon-water nanofluid with thermal radiation. Int. J. Heat Mass Transf. **106**,
   810–815 (2017)

3. Jain, S., Choudhary, R.: Combined effects of suction/injection on MHD boundary layer flow of nanofluid over horizontal permeable cylinder with radiation. J. Adv. Res. Dyn. Control Syst. **11**, 88–98 (2017)

4. Aman, S., Khan, I., Ismail, Z., Salleh, M.Z., Al-Mdallal, Q.M.: Heat transfer enhancement in free convection flow of CNTs Maxwell nanofluids with four different types of molecular liquids. Sci. Rep. https://doi.org/10.1038/s41598-017-01358-3

5. Hayat, T., Ahmed, S., Muhammad, T., Alsedi, A., Ayub, M., Computational modeling for homogenous-heterogeneous reactions in three-dimensional flow of carbon nanotubes. Res. Phys. **7**, 2651–2657 (2017)

6. Merkin, J.H.: A model for isothermal homogenous-heterogeneous reactions in boundary layer flow. Math. Comput. Modell. **24**, 125–136 (1996)

7. Kameswaran, P.K., Shaw, S., Sibanda, S., Murthy, P.V.S.N.: Homogenous-heterogeneous reactions in a nanofluid flow due to a porous stretching sheet. Int. J. Heat Mass Transf. **57**, 465–472 (2013)

8. Hayat, T., Hussain, Z., Muhammad, T., Alsaedi, A.: Effect of homogenous-heterogeneous reactions in flow of nanofluids over a nonlinear stretching surface with variable surface thickness. J. Mol. Liq. **221**, 1121–1127 (2016)

9. Sheikh, M., Abbas, Z.: Homogenous-homogenous reactions in stagnation point flow of Casson fluid due to a stretching/shrinking sheet with uniform suction and slip effects. Ain Shams Eng. J. **8**, 467–474 (2017)

10. Bachok, N., Ishak, A., Pop, I.: On the stagnation-point flow towards a stretching sheet with homogenous-heterogeneous reactions effects. Commun. Nonlinear Sci. Num. Simulat. **16**, 4296–4302 (2011)

11. Hayat, T., Aziz, A., Muhammad, T., Alsedi, A.: On magnetohydrodynamic three-dimensional flow of nanofluid over a convectively heated nonlinear stretching surface. Int. J. Heat Mass Transf. **100**, 566–572 (2016)

12. Jain, S., Bohra, S.: Heat and mass transfer over a three-dimensional inclined non-linear stretching sheet with convective boundary conditions. Ind. J. Pure Appl. Phys. **55**, 847–856 (2017)

13. Mustafa, M., Mustaq, A., Hayat, T., Alsedi, A.: Rotating flow of magnetite-water nanofluid over a stretching surface inspired by non-linear thermal radiation. PLOS ONE **11**(2), e0149304. https://doi.org/10.1371/journal.pone.0149304

14. Gopal, D., Kishan, N., Raju, C.S.K.: Viscous and joule's dissipation on Casson fluid over a chemically reacting stretching sheet with inclined magnetic field and multiple slips. Inform. Med. Unlocked **9**, 154–160 (2017)

15. Kandasamy, R., Vignesh, V., Kumar, A., Hasan, S.H., Isa, N.M.: Thermal radiation energy due to SWCNTs on MHD nanofluid flow in the presence of seawater/water: lie group transformation. Ain Shams Eng. J. (2016)

16. Mahanthesh, B., Gireesha, B.J., Gorla, R.S.R.: Nonlinear radiative heat transfer in MHD 3-D flow of water based nanofluid over a non-linearly stretching sheet with convective boundary condition. J. Niger. Math. Soc. **35**, 178–198 (2016)

17. Nayak, M.K.: MHD 3D flow and heat transfer analysis of nanofluid by shrinking surface inspired by thermal radiation and viscous dissipation. Int. J. Mech. Sci. **124–125**, 185–193 (2017)

18. Matin, M.H., Nobari, M.R.H., Jahangiri, P.: Entropy analysis in mixed convection MHD flow of nanofluid over a non-linear stretching sheet. J. Theory. Sci. Tech. **7**(1) (2012)

19. Das, S., Chakraborty, S., Jana, R.N., Makinde, O.D.: Entropy analysis of unsteady magneto-nanofluid flow past accelerating stretching sheet with convective boundary condition. Appl. Math. Mech. Engl. Ed. **36**(12), 1593–1610 (2015)

20. Jain, S., Parmar, A.: Radiation effect on MHD Williamson fluid flow over stretching cylinder through porous medium with heat source. In: Applications of Fluid Dynamics Lecture Notes in Mechanical Engineering, pp. 61–78 (2018)

21. Junaid, A.K., Mustafa, M., Hayat, T., Alseadi, A.: Three-dimensional flow of nanofluid over a non-linearly stretching sheet: An application to solar energy. Int. J. of Heat and Mass Trans. **86**, 158–164 (2015)

Bioconvection Flow and Heat Transfer over a Stretching Sheet in the Presence of Both Gyrotactic Microorganism and Nanoparticle Under Convective Boundary Conditions and Induced Magnetic Field

Shalini Jain and Rakesh Choudhary

Abstract The present analysis considered the effects of heat and mass transfer with an induced magnetic field in both gyrotactic microorganisms and nanoparticle over a stretching sheet. The influences of convective boundary conditions have also been taken into account. PDEs are converted into nonlinear coupled ODEs by using suitable similarity transformation. These equations are evaluated numerically by fourth- and fifth- order Runge–Kutta–Fehlberg method with shooting technique in MAT-LAB. The evaluations are carried out to know the influence of various parameters on velocity, induced magnetic, temperature, concentration, and motile microorganism density profile and the results are shown through graphical representation.

Keywords Induced magnetic field · Stretching sheet · Bioconvection

1 Introduction

The bioconvection occurs when water-based nanofluids are having the motile microorganism. These self-propelled microorganisms are used to enhance the density of base fluid in a particular system. Bioconvection is used to define the phenomenon of macroscopic convection motion of the fluid invented due to the density gradient generated by combining swimming of microorganisms (i.e., bacteria and algae). The nanoparticles are not self-propelled, unlike the motile microorganisms, and the motion of nanoparticles is motivated by Brownian motion and thermophoresis, which

S. Jain (✉) · R. Choudhary
Department of Mathematics & Statistics, Manipal University Jaipur, Jaipur 303007, India
e-mail: shalini.jain@jaipur.manipal.edu

are arising in the nanofluid. Thus, we can say that the motile microorganism's motion is not dependent on the motion of nanoparticles. Bioconvection has many applications due to enhancement in mass transfer and mixing in bio-microsystems, enzyme biosensors, and biotechnology. Initially, Childress et al. [1] obtained the model for collective movement of microorganism in the occurrence of nanoparticles. He studied that the motility of the organism is defined by a typical upward swimming speed of microorganism with the velocity U. The obtained outcomes are compared with remarks of patterns formed by the ciliated protozoan "Tetrahymena pyriformis". The mechanism of up-swimming varies among bottom-heavy algae and oxytactic bacteria. Rational continuum models expressed and investigated in each of these cases for short-cell volume fraction [2, 3].

Kuznetsov [4] described the Galerkin method to acquire an analytical explanation for the critical Rayleigh number for the non-oscillatory condition in the presence of gyrotactic microorganism. Xu and Pop [5] investigated the HAM solution of mixed convection fluid flow through a stretching surface with the existence of nanoparticles. Raees et al. [6] extended the problem of Xu and Pop and analyzed the influence of gravity-driven nanoliquid in the occurrence of gyrotactic microorganism. Several authors such as Jain and Parmar [7], Jain and Bohra [8], Siddiqa et al. [9], Alsaedi et al. [10], and Govind et al. [11] have explored the same phenomenon.

Recently, some authors have paid their attention to account the effects of induced magnetic field in the phenomenon of boundary layer flow. The study of behavior of electrically conducting fluid in the existence of an electromagnetic field is called Magnetohydrodynamics (MHD) with many applications in various fields of engineering as well as geophysics, astrophysics, manufacturing, etc. The subject of MHD has been functional, for example, in a problem related with the incarceration of plasma by magnetic fields and in projects containing thermonuclear generation of energy. Chauhan and Rastogi [12] illustrated the effects of heat transfer and MHD flow past a stretching sheet with a porous medium. The authors such as Chauhan and Agarwal [13], Ali et al. [14], Jafar et al. [15], and Jain and Choudhary [16] analyzed the MHD and induced magnetic field phenomenon in boundary layer flow.

Several authors have studied the influence of convective boundary conditions with different aspects. Gireesha et al. [17] presented the effects of the induced magnetic field for boundary layer stagnation point flow over a stretching sheet with convective boundary conditions. The unsteady MHD flow of a bio-nanofluid with convective boundary conditions was studied by Faisal et al. [18]. Makinde and Aziz [19] gave a model on stretching sheet non-Newtonian fluid with convective boundary conditions. Khan and Makinde [20] studied the heat and mass transfer effects of gyrotactic microorganism in the existence of nanoparticles. Recently, Chakraborty et al. [21] and Ramzan et al. [22] also investigated the impacts of gyrotactic microorganism with nanoparticle for different pertinent parameters in boundary layer flow.

In the current study, we have examined the influence of the induced magnetic field in the occurrence of both gyrotactic microorganisms and nanoparticle over a stretching sheet with convective boundary conditions. The resulting governing equations are solved by fourth- and fifth- order Runge–Kutta–Fehlberg method with shooting technique.

2 Mathematical Formulation

Consider a two-dimensional stagnation point incompressible viscous electrically conducting fluid flow of a nanofluid over a stretching sheet with microorganism under convective boundary conditions. Assuming that $u_w(x) = cx^m$ is the stretching velocity and $u_e(x) = ax^m$ is the ambient fluid velocity, where a, c, and m are the constants and $c>0$ shows the case for stretching sheet. The impact of the induced magnetic field is also taken into account. For induced magnetic field, the electric field $\vec{E}$ is supposed to be negligible. It is well-known that the electrical currents which flow in the fluid will provide growth to an induced magnetic field. Here, the strength of the induced magnetic field $H_e(x)$ *is* applied in the normal direction to the sheet.

The nanofluid having gyrotactic microorganisms and the nanoparticle suspension is constant and the direction of microorganisms' swimming is independent. The temperature T_f, nanoparticle volume fraction C_f, and the density of motile microorganisms n_w are assumed at the boundary wall, whereas, T_∞, C_∞, and n_∞ are supposed as their ambient values. We ignore the viscous dissipation in the energy equation.

The equations represent the conservation of continuity, momentum, induced magnetic field, thermal energy, nanoparticles, and microorganism, respectively, which are as follows:

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0 \tag{1}$$

$$\frac{\partial H_1}{\partial x} + \frac{\partial H_2}{\partial y} = 0 \tag{2}$$

$$u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} = u_e\frac{du_e}{dx} + v\frac{\partial^2 u}{\partial y^2} + \frac{\mu_0}{4\pi\rho}\left(H_1\frac{\partial H_1}{\partial x} + H_2\frac{\partial H_1}{\partial y}\right) - \frac{\mu_0}{4\pi\rho}H_e\frac{dH_e}{dx} \tag{3}$$

$$u\frac{\partial H_1}{\partial x} + v\frac{\partial H_1}{\partial y} - H_1\frac{\partial u}{\partial x} - H_2\frac{\partial u}{\partial y} = \mu_e\frac{\partial^2 H_1}{\partial y^2} \tag{4}$$

$$u\frac{\partial T}{\partial x} + v\frac{\partial T}{\partial y} = \alpha\frac{\partial^2 T}{\partial y^2} + \tau\left[D_B\frac{\partial C}{\partial y}\frac{\partial T}{\partial y} + \frac{D_T}{T_\infty}\left(\frac{\partial T}{\partial y}\right)^2\right] \tag{5}$$

$$u\frac{\partial C}{\partial x} + v\frac{\partial C}{\partial y} = D_B\frac{\partial^2 T}{\partial y^2} + \frac{D_T}{T_\infty}\left(\frac{\partial^2 T}{\partial y}\right) \tag{6}$$

$$u\frac{\partial n}{\partial x} + v\frac{\partial n}{\partial y} + \frac{bW_c}{(C_f - C_\infty)}\left[\frac{\partial}{\partial y}\left(n\frac{\partial C}{\partial y}\right)\right] = D_m\frac{\partial^2 n}{\partial y^2}. \tag{7}$$

Subject to the boundary conditions

$$\text{at } y = 0, u = u_w(x), \ v = 0, \ H_1 = H_2 = 0, \ -\kappa\frac{\partial T}{\partial y} = h_f(T_f - T), n = n_w,$$

$$-D_B\frac{\partial C}{\partial y} = h_s(C_f - C),$$

$$\text{at } y \to \infty u \to u_e(x), \ H_1 \to H_e(x) = H_0x^m, \ T \to T_\infty, \ C \to C_\infty, n \to n_\infty, \tag{8}$$

where (u, v) and (H_1, H_2) are the velocity components and magnetic components in (x, y) directions, respectively, H is the induced magnetic field vector and H_0 is the value of $H_e(x)$ when $x = 0$, N and σ are the kinematic viscosity and electrical conductivity, respectively, μ_0 is the magnetic permeability, ρ is the fluid density, $\mu_e = 1/4\pi\sigma$ is the magnetic diffusivity, $\alpha = \kappa/\rho C_p$ is the thermal diffusivity, κ is the thermal conductivity, and C_p is specific heat at constant pressure. $\tau = (\rho c)_p/(\rho c)_f$ is the referred proportion between the effectual heat capability of nanoparticles and base fluid, D_B and D_T denote the Brownian diffusion coefficient and thermophoretic diffusion coefficient, respectively, n is the concentration of microorganism, b is chemotaxis constant, W_c is referred maximum cell swimming speed, D_m signifies the diffusivity of microorganism, h_f and h_s are heat and mass transfer coefficients, respectively.

Introducing similarity transformation

$$\psi = \sqrt{u_e(x)\nu x}\, f(\eta), \quad \eta = \sqrt{\frac{u_e(x)}{\nu x}}\, y, \quad \xi(\eta) = H_0\sqrt{\frac{\nu x^{m+1}}{a}}\, s(\eta),$$

$$\theta = \frac{T - T_\infty}{T_f - T_\infty}, \phi(\eta) = \frac{C - C_\infty}{C_f - C_\infty}, \ w(\eta) = \frac{n - n_\infty}{n_w - n_\infty}. \tag{9}$$

Equations (1)–(2) are satisfied with the following transformation

$$u = \frac{\partial\psi}{\partial y}, v = -\frac{\partial\psi}{\partial x}, \ H_1 = \frac{\partial\xi}{\partial y}, H_2 = -\frac{\partial\xi}{\partial x}. \tag{10}$$

Using similarity transformation (9)–(10), Eqs. (3)–(7) become

$$f''' + \left(\frac{m+1}{2}\right)ff'' - m(f'^2 - 1) + M^2\left[ms'^2 - \left(\frac{m+1}{2}\right)ss'' - m\right] = 0, \tag{11}$$

$$\lambda^* s''' + \left(\frac{m+1}{2}\right)fs'' - \left(\frac{m+1}{2}\right)f''s = 0, \tag{12}$$

$$\theta'' + Pr\left[\left(\frac{m+1}{2}\right)f\theta' + N_b\phi'\theta' + N_t\theta'^2\right] = 0, \tag{13}$$

$$\phi'' + Le\left(\frac{m+1}{2}\right)f\phi' + \frac{N_t}{N_b}\theta'' = 0, \tag{14}$$

$$w'' + Sc\left(\frac{m+1}{2}\right)fw' - Pe\left[(w + \Omega)\phi'' + \phi'w'\right] = 0. \tag{15}$$

With the boundary conditions

$$f(0) = 0, f'(0) = \lambda, \ s(0) = 0, \ s'(0) = 0, \ \theta'(0) = -Bi[1 - \theta(0)],$$
$$\phi'(0) = -N_d[1 - \phi(0)], \ w(0) = 1$$
$$f'(\infty) = 1, \ s'(\infty) = 1, \ \theta(\infty) = 0, \ \phi(\infty) = 0, \ w(\infty) = 0. \tag{16}$$

Here, $f'(\eta), s'(\eta), \theta(\eta), \phi(\eta)$ and $w(\eta)$ symbolize velocity without dimension close to the surface, magnetic profile without dimension, temperature without dimension, nanoparticle concentration without dimension, denseness of motile microorganism without dimension, respectively.

Here, $M^2 = \frac{\mu_0 H_0^2}{4\pi\rho a^2}$ is magnetic field parameter, $\lambda^* = \frac{1}{4\pi\sigma v}$ is the reciprocal of magnetic Prandtl number, $Pr = \frac{v}{\alpha}$ signifies Prandtl number, $N_b = \frac{\tau D_B}{\alpha}(C_f - C_\infty)$ denotes the Brownian motion parameter, $N_t = \frac{\tau D_T}{T_\infty \alpha}(T_f - T_\infty)$ is the thermophoresis parameter, $Le = \frac{v}{D_B}$ represents the Lewis number, $Sc = \frac{v}{D_m}$ is the Schmidt number, $Pe = \frac{bW_c}{D_m}$ is the bioconvection Peclet number, $\Omega = \frac{n_\infty}{n_w - n_\infty}$ is the dimensionless parameter, $\lambda = c/a$ is the stretching parameter $(\lambda > 0)$, $Bi = \frac{h_f}{\kappa}\sqrt{\frac{v}{ax^{m-1}}}$ denotes Biot number, and $N_d = \frac{h_s}{D_B}\sqrt{\frac{v}{ax^{m-1}}}$ is the convection–diffusion parameter.

3 Numerical Method

The coupled nonlinear Eqs. (11)–(15) with boundary conditions (16) are solved numerically using Runge–Kutta–Fehlberg fourth- and fifth- order (RKF45) with shooting technique. Initially, boundary value problems are converted into initial value problems.

In this method, the important thing is that to select the proper finite value of η, for current study we have chosen η_8 for η_∞. Later, we began with initial guess of $f''(0), s''(0), \theta'(0), \phi'(0)$, and $w'(0)$, which are found by secant method using step size $h = 0.00001$ and convergence criteria 10^{-6}.

Assume that

$$f = f_1, f' = f_2, f'' = f_3, s = f_4, s' = f_5, s'' = f_6, \theta = f_7, \theta' = f_8, \phi = f_9,$$

$$\phi' = f_{10}, w = f_{11} \text{ and } w' = f_{12}. \tag{17}$$

Using Eq. (17) and Eqs. (11)–(15) are converted into initial value problems

$$f_3' = -\left(\frac{m+1}{1}\right) f_1 f_3 - m(f_2^2 - 1) - M^2\left[mf_5^2 - \left(\frac{m+1}{2}\right)f_4 f_6 - m\right] \tag{18}$$

$$f_6' = \frac{1}{\lambda^*}\left[\left(\frac{m+1}{2}\right)(f_3 f_4 - f_1 f_6)\right] \tag{19}$$

$$f_8' = -\Pr\left[\left(\frac{m+1}{2}\right) f_1 f_8 + N_b f_8 f_{10} + N_t f_8^2\right] \tag{20}$$

$$f_{10}' = -Le\left(\frac{m+1}{2}\right) f_1 f_{10} - \frac{N_t}{N_b} f_8' \tag{21}$$

$$f'_{12} = Pe\left[(f_{11} + \Omega)f'_{10} + f_{10}f_{12}\right] - Sc\left(\frac{m+1}{2}\right)f_1 f_{12} \tag{22}$$

Subjected to the boundary conditions

$$
\begin{aligned}
&f_1(0) = 0, \ \ f_2(0) = \lambda, \ \ f_3(0) = r_1, \ \ f_4(0) = 0, \ f_5(0) = 0, \ f_6(0) = r_2, \\
&f_7(0) = (f_8(0)/Bi) + 1, \ \ f_8(0) = r_3, \ \ f_9(0) = (f_{10}(0)/N_d) + 1, \\
&f_{10}(0) = r_4, \ f_{11}(0) = 1, \ \ f_{12}(0) = r_5.
\end{aligned}
\tag{23}
$$

Here, r_1, r_2, r_3, r_4, and r_5 are the guesses for $f''(0), s''(0), \theta'(0), \phi'(0)$, and $w'(0)$, correspondingly.

4 Results and Discussion

Numerical results obtained by Eqs. (18)–(22) under the boundary conditions (23) are discussed through graphs. A detailed discussion of various parameters such as magnetic field parameter, reciprocal of magnetic Prandtl number, Brownian motion parameter, Biot number, thermophoresis parameter, convection–diffusion parameter, Peclet number, and Schmidt number is shown through graphs.

Figures 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10 illustrate the effects of magnetic field parameter and reciprocal of magnetic Prandtl number on velocity, induced magnetic field, temperature, concentration, and motile microorganism density profile. The velocity profile and magnetic field profile decrease with magnetic field parameter. This is due to magnetic field contrasting the transport phenomena, since the variation of magnetic parameter affects the variation of Lorentz force. Lorentz force is a drag-like force that generates more resistance to transport phenomena and that causes drops in the fluid velocity. For reciprocal of magnetic Prandtl number, the presence of H_1 and H_2 enhances the magnitude of the Lorentz force and delay the magnitude of the velocity. On the other hand, due to the above reason the temperature, concentration, and motile microorganism density profile enhances with magnetic field parameter and reciprocal of magnetic Prandtl number.

Figures 11 and 12 represent the effects of thermophoretic and Brownian motion parameters on temperature profile and concentration profile, respectively. The existence of nanoparticles (which accounts for the thermophoresis and Brownian motion parameters) in the base fluid enhances the thermal conductivity, and hence thermal and solute boundary layer width increases. It is also observed that in concentration profile near the wall, the profile decreases with thermophoresis parameter and Brownian motion parameter but after a midpoint, it shows enhancement. It is noticeable that the Brownian motion of nanoparticles plays a major role in nanofluids's heat transfer improvement.

Figures 13 and 14 display the impacts of Biot number and convection–diffusion parameter on temperature profile and concentration profile, respectively. Enhance-
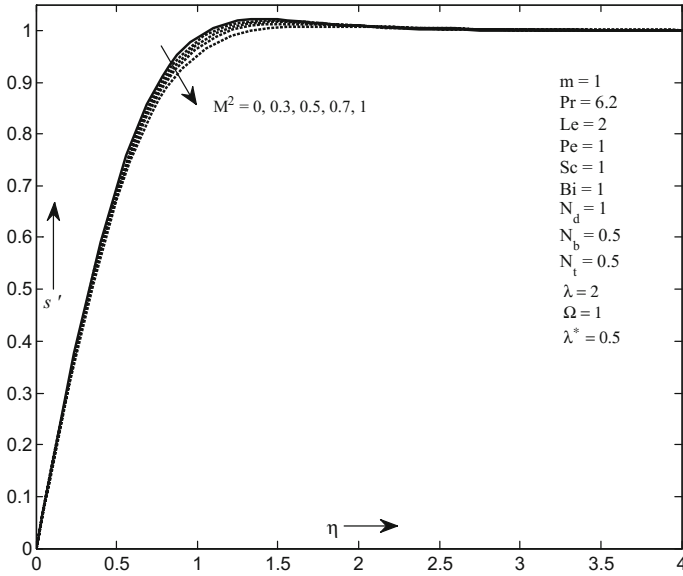
Fig. 1 Effects of M^2 on magnetic field profile
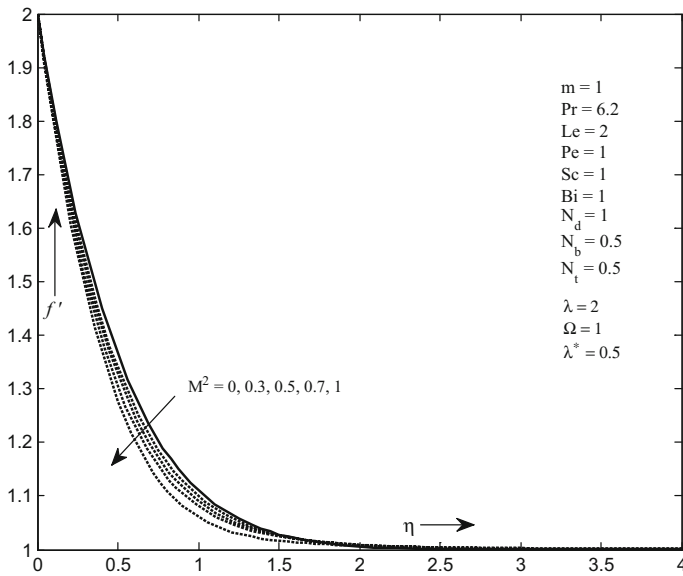
Fig. 2 Effects of M^2 on velocity profile
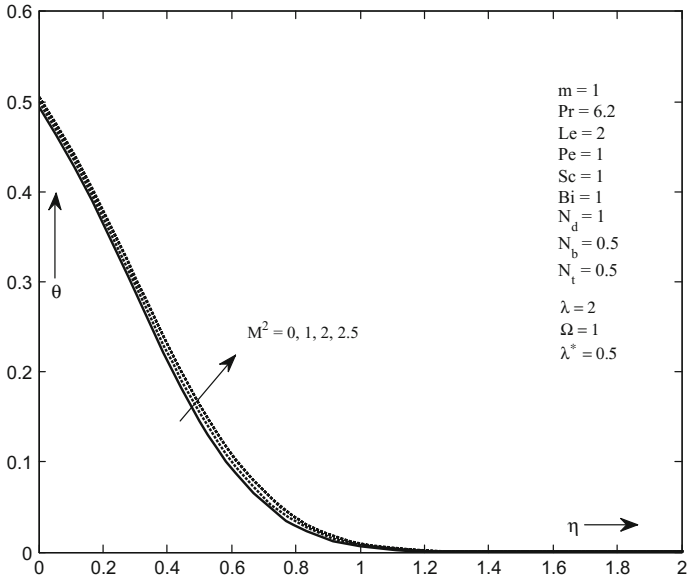
Fig. 3 Effects of M^2 on temperature profile
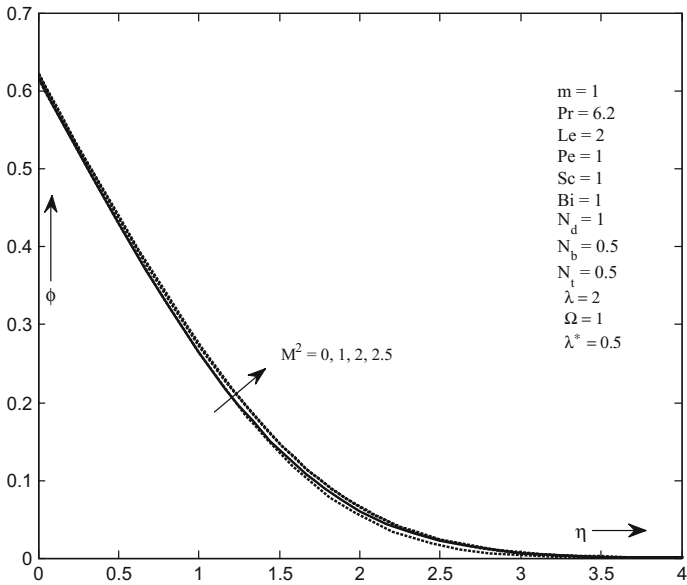
Fig. 4 Effects of M^2 on concentration profile
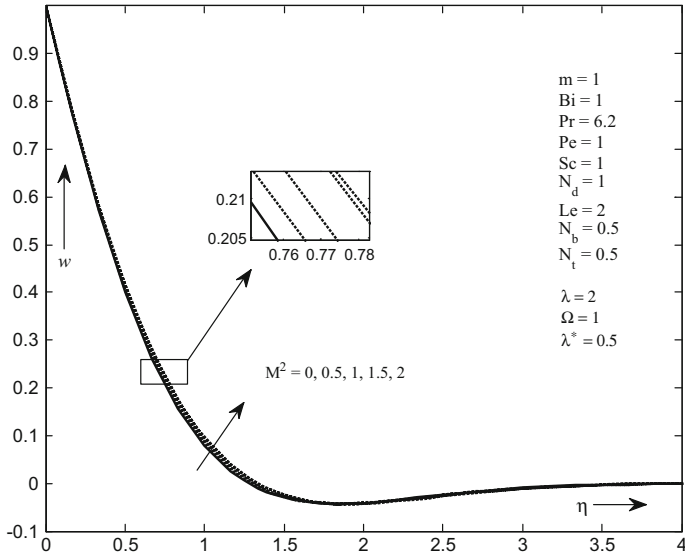
Fig. 5 Effects of M^2 on microorganism profile
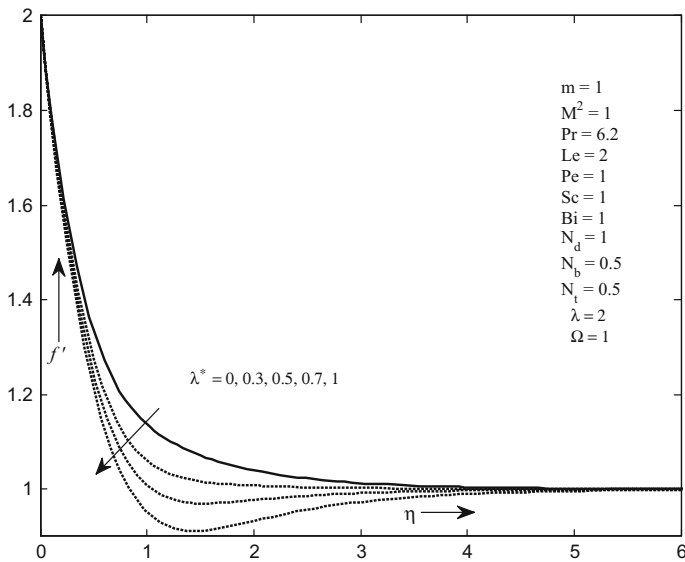
Fig. 6 Effects of λ^* on velocity profile
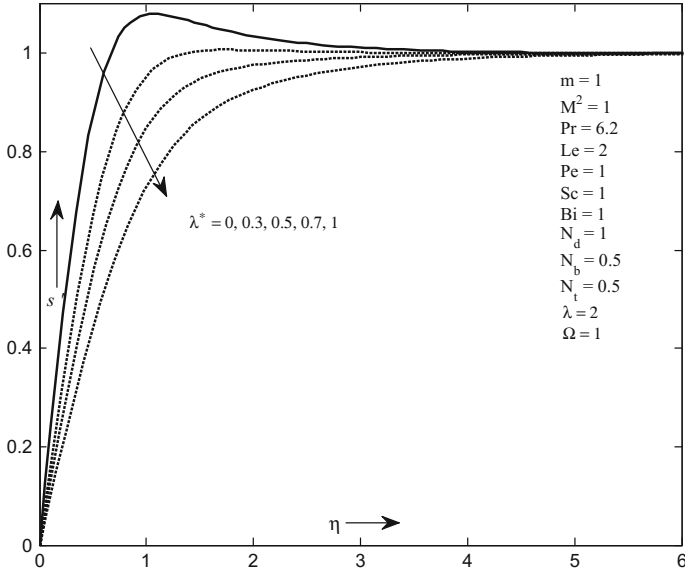
Fig. 7 Effects of λ* on magnetic profile
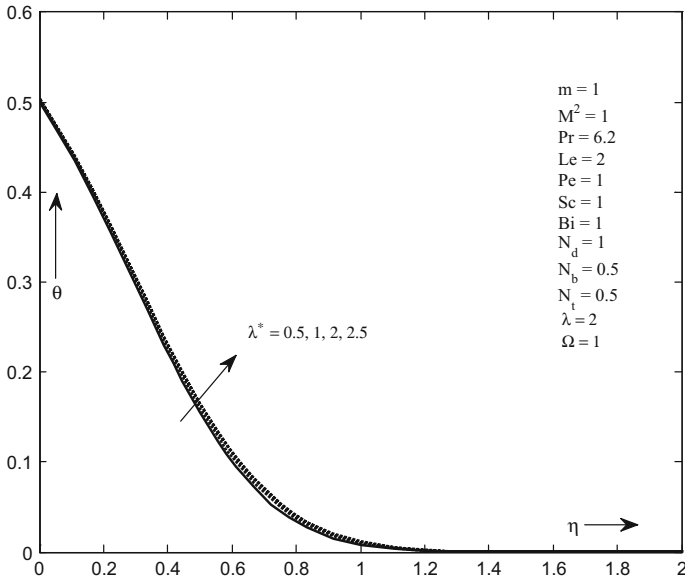
Fig. 8 Effects of λ* on temperature profile
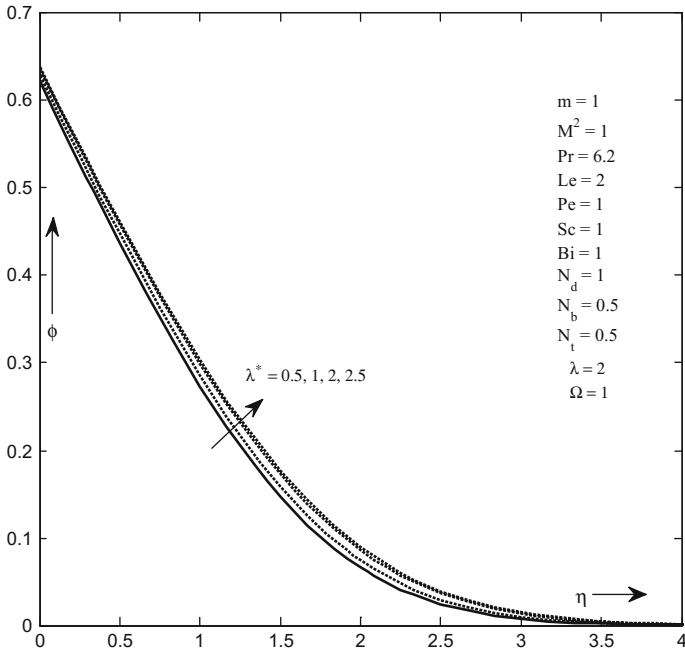
Fig. 9 Effects of λ* on concentration

ment in Biot number and convection–diffusion parameter causes an increase in the thickness of thermal boundary layer and solute boundary layer, correspondingly. Physically, the thermal resistance of the sheet decreases and convective heat transfer through the fluid on the right side of the sheet is enlarged for higher values of Biot number, and hence the thermal boundary layer thickness is increased.

In Fig. 14, due to the above reason in solute boundary layer profile, the solute boundary layer thickness increases with convection–diffusion parameter. The impact of Lewis number on concentration profile is plotted in Fig. 15. It is observed that for higher values of Lewis number, concentration profile shows reduction. This is due to that Lewis number is the ratio of kinematic viscosity and diffusion coefficient. Hence, enhancement in Lewis number leads to decrease in the diffusion coefficient in solute boundary layer thickness. Figures 16, 17, and 18 show that the density of motile microorganism is strongly affected by Schmidt number, dimensionless number Ω, and Peclet number. The foundation of the bioconvection is created from the interior energy of the microorganisms. Higher Peclet number and Schmidt number describe lower swimming speed of the motile microorganisms. The density of microorganisms decreases with the augment of Pe, which is the cause to the fact that the self-propelling (lower Pe) motile microorganisms outcome in high microorganism propulsion, which reduces the microorganism profile for Peclet number as well as Schmidt number and Ω.
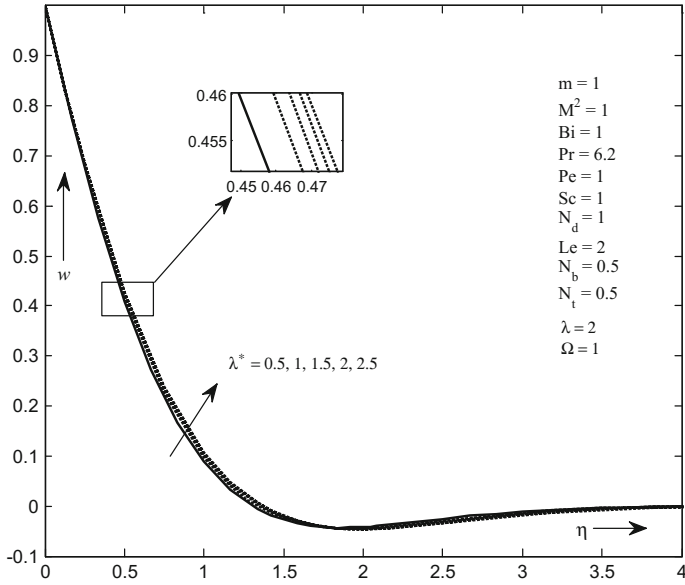
Fig. 10 Effects of λ^* on microorganism profile
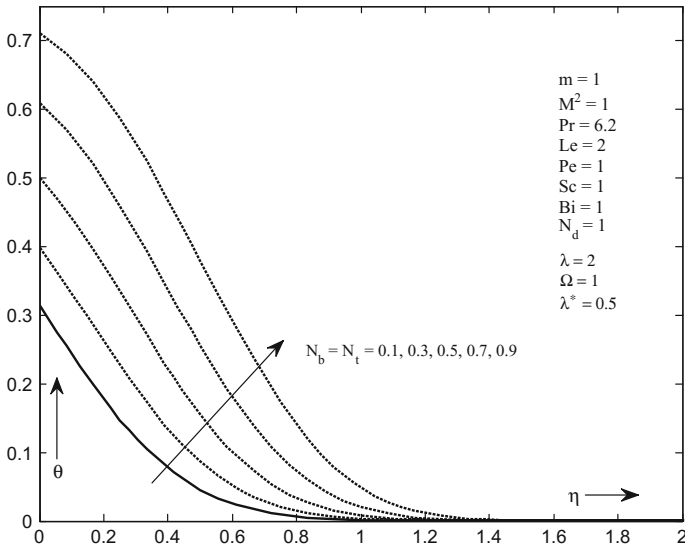
Fig. 11 Effects of N_b and N_t on temperature
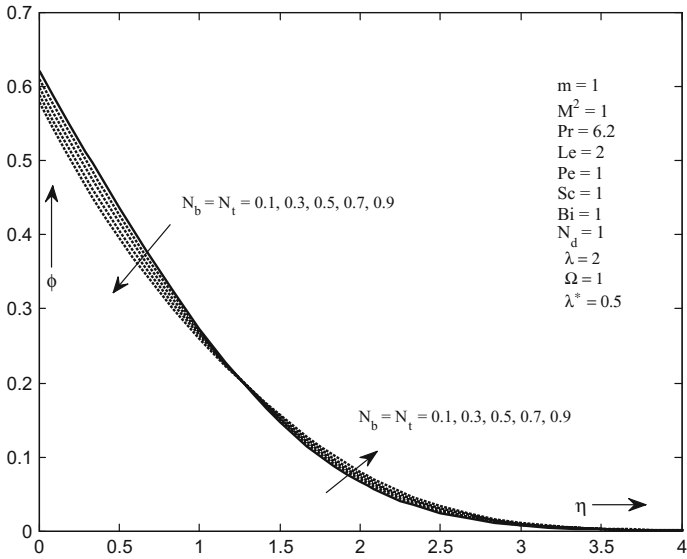
Fig. 12 Effects of N_b and N_t on concentration
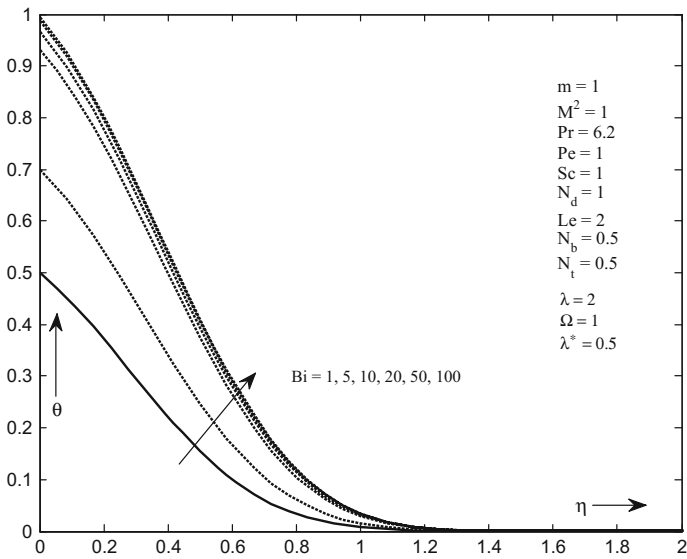
Fig. 13 Effects of Bi on temperature profile
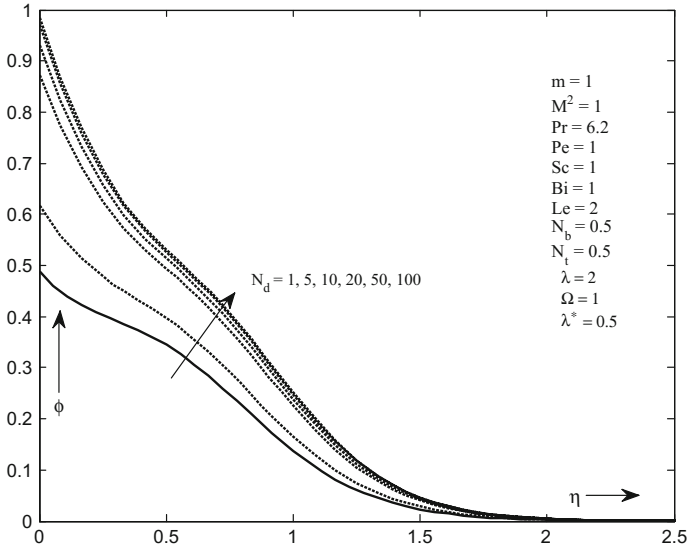
Fig. 14 Effects of N_d on concentration profile
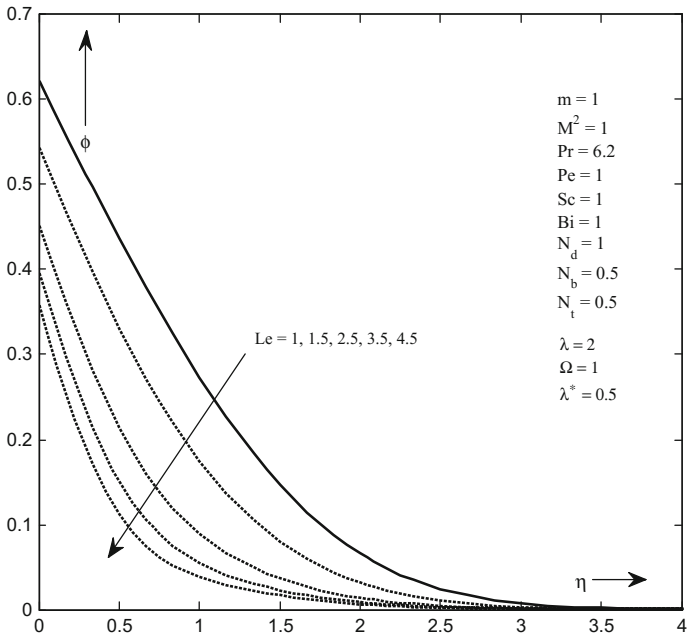
Fig. 15 Effects of Le on concentration profile
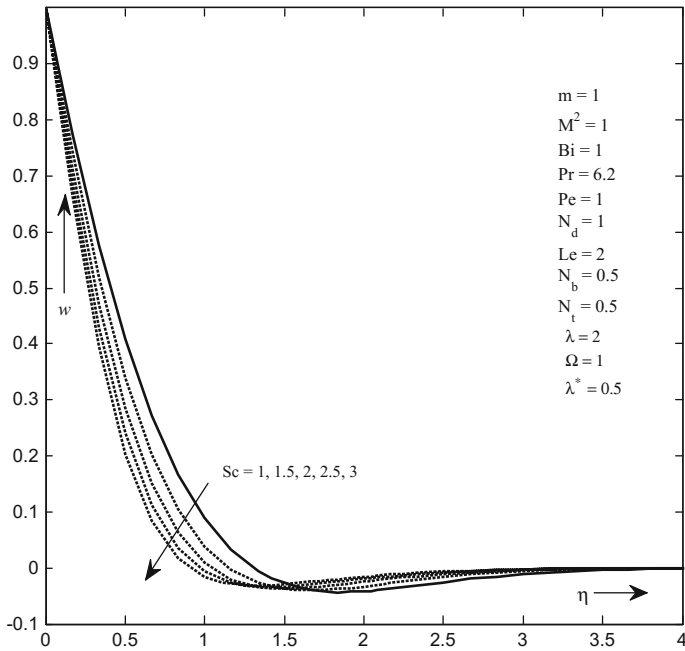
Fig. 16 Effects of *Sc* on microorganism profile
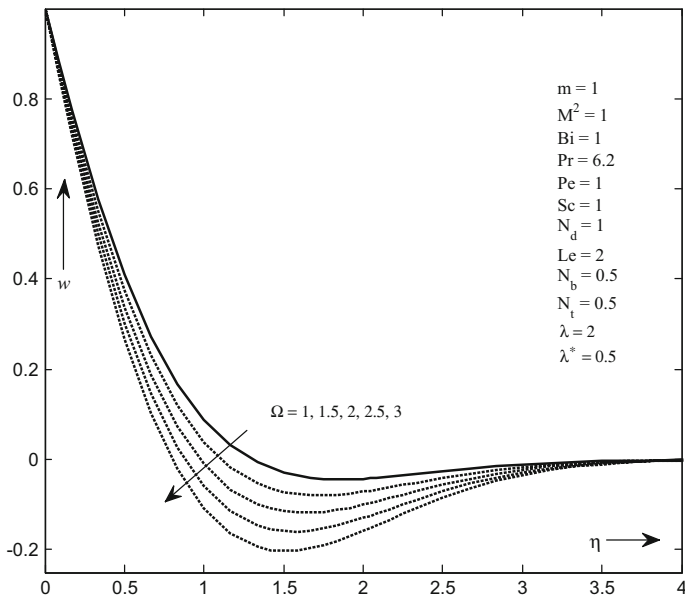
Fig. 17 Effects of *Ω* on microorganism profile
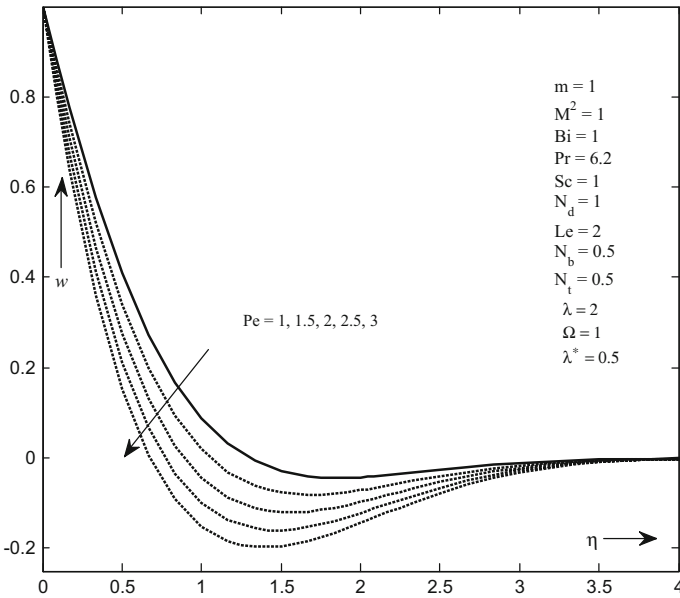
Fig. 18 Effects of *Pe* on microorganism profile

5 Conclusion

We have explored the impacts of heat and mass transfer with induced magnetic field in the occurrence of both gyrotactic microorganisms and nanoparticles over stretching sheet under convective boundary condition. Following are the important results comprehended from the current analysis:

- Velocity profile and magnetic profile show negative behavior when magnetic field parameter and reciprocal of magnetic Prandtl number increases, whereas temperature profile, concentration profile, and motile microorganism density profile shows enhancement for it.
- There is an increment in temperature profile, when the Brownian motion and thermophoresis parameters are enhanced; however, mixed results are shown in concentration profile.
- Temperature profile and concentration profile are an increasing function of Biot number and convection–diffusion parameter. On the other hand, concentration boundary layer width reduces with Lewis number.
- Motile microorganism density profile decreases with Schmidt number, dimensionless parameter, and Peclet number.

References

1. Childress, S., Levandowsky, M., Spiegel, E.A.: Pattern formation in a suspension of swimming microorganisms—equations and stability theory. J. Fluid Mech. **69**, 591–613 (1975)
2. Hillesdon, A.J., Pedley, T.J.: Bioconvection in suspensions of oxytactic bacteria: linear theory. J. Fluid Mech. **324**, 223–259 (1996)
3. Hill, N.A., Pedley, T.J.: Bioconvection. Fluid Dyn. Res. **37**(1/2), 1–20 (2005)
4. Kuznetsov, A.V.: The onset of nanofluid bioconvection in a suspension containing both nanoparticles and gyrotactic microorganisms. Int. Commun. Heat Mass Transf. **37**, 1421–1425 (2010)
5. Xu, H., Pop, I.: Mixed convection flow of a nanofluid over a stretching surface with uniform free stream in the presence of both nanoparticles and gyrotactic microorganisms. Int. J. Heat Mass Transf. **75**, 610–623 (2014)
6. Raees, A., Xu, H., Sun, Q., Pop, I.: Mixed convection in gravity-driven nano-liquid film containing both nanoparticles and gyrotactic microorganisms. Appl. Math. Mech. **36**, 163–178 (2015)
7. Jain, S., Parmar, A.: Radiation effect on MHD williamson fluid flow over stretching cylinder through porous medium with heat source. In: Applications of Fluid Dynamics. Lecture Notes in Mechanical Engineering, pp. 127–141 (2018)
8. Jain, S., Bohra, S.: Hall current and radiation effects on unsteady MHD squeezing nanofluid flow in a rotating channel with lower stretching permeable wall. In: Applications of Fluid Dynamics. Lecture Notes in Mechanical Engineering, pp. 127–141 (2018)
9. Sadia, S, Gul-e-Hina, Begum, N., Saleem, S., Hossain, M.A., Gorla, R.S.R.: Numerical solutions of nanofluid bioconvection due to gyrotactic microorganisms along a vertical wavy cone. Int. J. Heat Mass Trans. **101**, 608–613 (2016)
10. Alsaedi, A., Khan, I.M., Farooq, M., Gull, N., Hayat, T.: Magnetohydrodynamic (MHD) stratified bioconvective flow of nanofluid due to gyrotactic microorganisms. Adv. Powder Tech. **28**(1), 288–298 (2017)
11. Govind, R.R., Patil, V.S., Krishna Prasad, J.S.V.R.: Hydromagnetic bioconvection flow in the region of stagnation-point flow and heat transfer in non-Newtonian nanofluid past a moving surface with suction: similarity analysis. Int. J. Heat Tech. **35**(1), 25–31 (2017)
12. Chauhan, D.S., Rastogi, P.: Heat transfer and entropy generation in MHD flow through a porous medium past a stretching sheet. Int. J. Energy Tech. **3**(15), 1–13 (2011)
13. Chauhan, D.S., Agrawal, R.: MHD flow and heat transfer in a channel bounded by a shrinking sheet and a plate with a porous substrate. J. Eng. Phys. Thermophys. **84**(5), 1034–1046 (2011)
14. Ali, F.M., Nazar, R., Arifin, N.M., Pop, I.: MHD mixed convection boundary layer flow toward a stagnation point on a vertical surface with induced magnetic field was studied. J. Heat Transf. **133**, 1–6 (2011). Article ID: 022502
15. Jafar, K., Nazar, R., Ishak, A., Pop, I.: MHD boundary layer flow due to a moving wedge in a parallel stream with the induced magnetic field. Boundary Value Problems, 1–14 (2013)
16. Jain, S., Choudhary, R.: Effects of MHD on boundary layer flow in porous medium due to exponentially shrinking sheet with slip. Procedia Eng. **127**, 1203–1210 (2015)
17. Gireesha, B.J., Mahanthesh, B., Shivakumara, I.S., Eshwarappa, K.M.: Melting heat transfer in boundary layer stagnation-point flow of nanofluid toward a stretching sheet with induced magnetic field. Eng. Sci. Tech. **19**, 313–321 (2016)
18. Faisal, M.B., Uddin, M.J., Ismail, A.I.M.: Unsteady magnetoconvective flow of bio nanofluid with zero mass flux boundary condition. Sains Malays. **46**(2), 327–333 (2017)
19. Makinde, O.D., Aziz, A.: Boundary layer flow of a nanofluid past a stretching sheet with a convective boundary condition. Int. J. Thermal. Sci. **50**(7), 1326–1332 (2011)

20. Khan, W.A., Makinde, O.D.: MHD nanofluid bioconvection due to gyrotactic microorganisms over a convectively heat stretching sheet. Int. J. Thermal. Sci. **81**, 118–124 (2014)
21. Chakraborty, T., Das, K., Kundu, P.K.: Framing the impact of external magnetic field on bio-convection of a nanofluid flow containing gyrotactic microorganisms with convective boundary conditions. Alex. Eng. J. (2016). In Print
22. Ramzan, M., Chung, J.D., Ullah, N.: Radiative magnetohydrodynamic nanofluid flow due to gyrotactic microorganisms with chemical reaction and non-linear thermal radiation. Int. J. Mech. Sci. (2017). In print

Color Histogram- and Smartphone-Based Diabetic Retinopathy Detection System

Nikita Kashyap, Dharmendra Kumar Singh and Girish Kumar Singh

Abstract Diabetic Retinopathy (DR) is diabetes-related eye disorder. An initial eye examination is the finest method to avoid DR. In this paper, a low-price DR detection algorithm using mobile phone-and color histogram-based image retrieval technique has been proposed. The mobile phone will take a picture of the patient's eye with the help of 20D condensing lens, and then implement a color histogram retrieval program to find the similar picture from the collected database. The presented system reduces the professional's work of DR identification. Our aim is to make an effective, easy, and low-cost eye examination program, which is ideal for underdeveloped regions and make it available to one and all.

Keywords Image retrieval · Histogram · Diabetic retinopathy · Smartphone

1 Introduction

Diabetes is a long-term organ disease that is familiar nowadays. It started when the blood sugar level increases because of the weak creation of hormone, which allows sugar to enter into human body [1, 2]. Because of diabetes, the amount of sugar in small blood vessels of eye increases, the vision becomes difficult and causes blindness. This is called Diabetic Retinopathy. It blocks and distorts the blood vessels of the retina, which is the reason for fluid liking and vision distorting. After 10–15 years of diabetes, about 30% of patients develop a serious visual impairment [3, 4].

N. Kashyap (✉)
Department of ET&T, Dr. C. V. Raman University, Bialspur, India
e-mail: nikita2010k@gmail.com

D. K. Singh
Department of EEE, Dr. C. V. Raman University, Bialspur, India
e-mail: dmsingh2001@rediffmail.com

G. K. Singh
Dr. Harisingh Gour University, Sagar, India
e-mail: gkrsingh@gmail.com
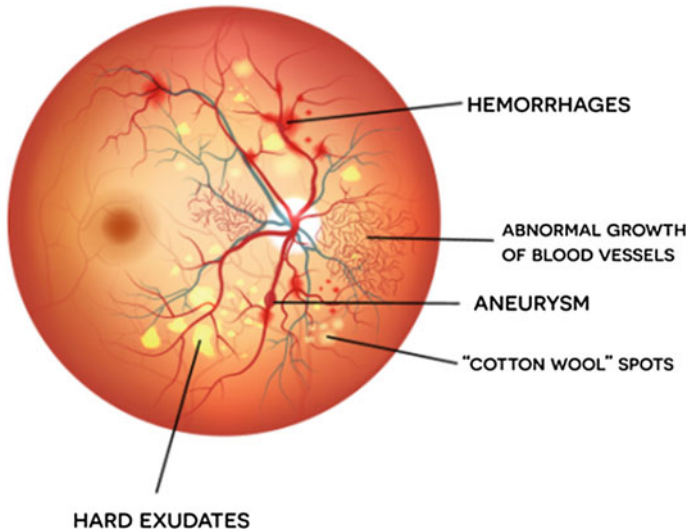
Fig. 1 Diabetic retinopathy image [7]

The diabetic retinopathy is basically divided into four phases. As the time goes, the situation of the disease increases from normal to first DR phase and then from first to further advance phases. The phases are defined below [5, 6] (Fig. 1)

(a) Mild nonproliferative DR: This is the first phase of retinopathy on which the eye's blood vessels are starting to swell. This swelling is known as an aneurysm.
(b) Moderate nonproliferative DR: On this phase, the blood vessels start to distort, swelling increases, and creates dot-and-blot hemorrhage.
(c) Severe nonproliferative DR: As the disease progresses, the blocking of blood vessels starts and yellow spots called cotton wool spots start to create on retina side.
(d) Proliferative diabetic retinopathy: This is the last and risky phase of DR that create new but fragile blood vessels on the eye side. These vessels leak fluid and growing of this vessel becomes the reason for serious visual impairment.

The identification of DR in the starting phase can excellently cut down the situation of visual impairment [8]. But the number of diabetic patients is very high as compared to the number of eye specialists, so there is a need of automated DR detection method so that the DR affected person alone can go to the hospital for treatment and analysis [9, 10].

For the analysis process, the most popular method is content-based image retrieval technique. It is a retrieval system that depends on particular features (text, shape, and color) of picture [11, 12]. This paper presents a color histogram-based CBIR detection system, which can easily and effectively identify the stages of Diabetic Retinopathy. The basic purpose of a color histogram is to create the feature points of the image. The feature points of the entire database and query image are compared to find out

the similarities between them by the formula of Euclidean Distance (ED) [13]. At the final stage of this method, the images with the list of possible distances from the query image are shown on screen. This paper shows the method by which mobile phones and 20D lens can be used to take the retinal image, so that the system can be available at any place at any time.

2 Proposed System

In this paper, mobile phone-and color histogram-based Diabetic Retinopathy identification system have been proposed. The introduction of both the methods is given below.

2.1 Color Histogram

The very effective data, which can be extracted from pictures for the retrieval work is the colors. They are the attractive and optical property of a picture and they are obvious to search and operate. In this paper, we analyze a method in which matching work of pictures depends on the color property explained by using the color histogram [14].

Histograms are usually calculated according to the intensity of RGB channels and provide independent histograms of channels [15]. They represent the pixel value, which contains the list of colors and that property is beneficial for fast examination of the number of digital images. A number of researchers are attracted to this method because it does not get affected by rotation, scaling, and translation of an image.

2.2 Smartphone as an Indirect Ophthalmoscope

Smartphones with valuable camera quality, 20D lens, and light-emitting diode (LED) can perform as an indirect ophthalmoscope. First, take the smartphone in one hand and the condensing lens in another as shown in Fig. 2 and then illuminate a patient's eye by using LED and capture the retinal image. This method is very simple and inexpensive [16, 17].

Fig. 2 Principle of smartphone indirect ophthalmology [18]

2.3 Algorithm and Flowchart

(a) Capture the retinal image of patient by using mobile phone and lens.
(b) Input the captured image.
(c) Change the size of the image into 256×256.
(d) Transform the RGB image into histogram form.
(e) Measure the properties: mean, skewness, and Std (Standard deviation) [14] as

Mean:

$$M_a = \sum_{b=1}^{N} \frac{1}{N} X_{ab}, \tag{1}$$

Standard deviation:

$$\sigma_a = \sqrt{\left(\frac{1}{N} \sum_{b=1}^{N} (X_{ab} - M_a)^2 \right)}, \tag{2}$$

Skewness:

$$S_a = \sqrt{\left(\frac{1}{N} \sum_{b=1}^{N} (X_{ab} - M_a)^3 \right)}. \tag{3}$$

(f) Make the image database of normal and diabetic retinopathy-affected eye.

(g) Read all the database images and repeat the steps 3–5 for each and every database image.

(h) Match the properties of patient's retinal image with the properties of each and every database image by calculating Euclidean Distance (ED) [14] as

$$E_d = \sum_{a=1}^{q} t_{a1} \left| M_a^1 - M_a^2 \right| + t_{a2} \left| \sigma_a^1 - \sigma_a^2 \right| + t_{a3} \left| S_a^1 - S_a^2 \right|, \tag{4}$$

where

X_{ab} is the a_{th} channel on the b_{th} pixel
A   is the current channel (e.g., $1 = H, 2 = S, 3 = V$)
Q   is the total number of channels
M_a^1, M_a^2 are the mean values of query image and database image, respectively
σ_a^1, σ_a^2 are the standard deviation (Std) of query image and database image, respectively
S_a^1, S_a^2 are the skewness of query image and database image, respectively
t_a is the weight at each channel point

(i) Organize all the values in an ascending order.
(j) Search the list of possible distances and present the matched images as result.
(k) Find the ability of the method by measuring the value of precision and recall as the following [14]:

$$\text{Precision} = \frac{\text{(No. of retrieved images that are relevant)}}{\text{Total no. of retrieved images}}, \tag{5}$$

$$\text{Recall} = \frac{\text{(No. of retrieved images that are relevant)}}{\text{Number of relevant images in the database}}. \tag{6}$$

Figure 3 shows the flowchart of the overall proposed system for DR detection.

3 Experimental Results

The number of images related to Diabetic Retinopathy (DR) is selected and analyzed. The retinal picture of the patient is taken by using a smartphone with lens, and then the image is processed by using color histogram displayed in Fig. 4.

After calculating all the feature points of query image and collected database, the points are compared by Euclidean distance. The value of Euclidean distances is arranged in descending structure and according to that we have obtained the following top five retrieval images (Fig. 5 and Table 1).
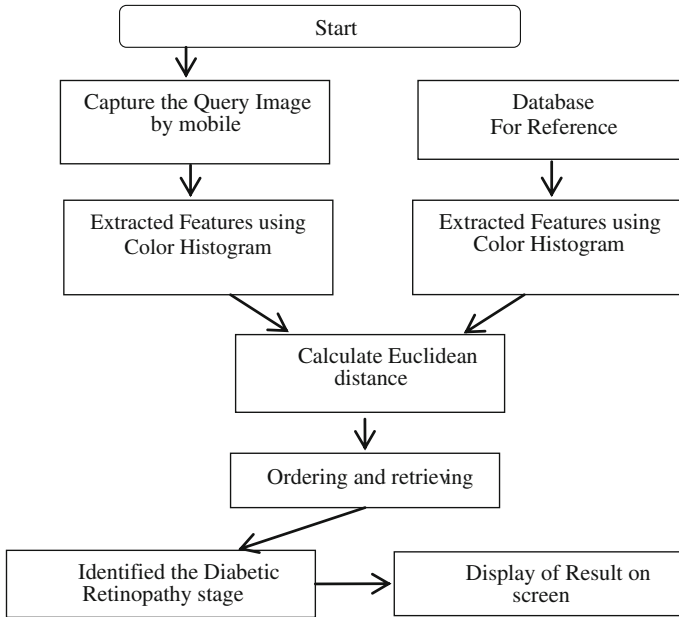
Fig. 3 Flowchart of the proposed technique
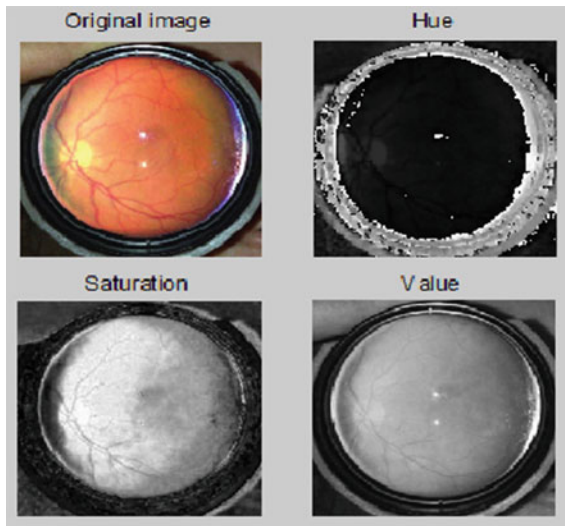
Fig. 4 The HSV color space in different planes (The image is of a patient's eye captured by mobile phone with condensing lens and subsequently three figures were generated by the program used for the study)
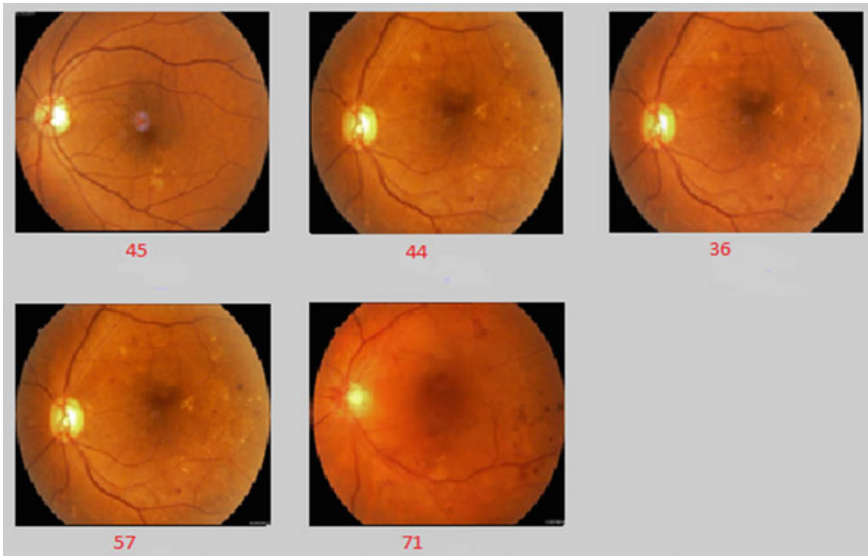
Fig. 5 The retrieved retinal image (The images are the retinal images from the database collected from Luthra Hospital, Bilaspur, which were obtained after comparing the retinal image of the patient to the database)

Table 1 Description of images

Image	Patient's name	Sex	Age	Diseases
QueryImg	Mrs. Sapnakaushik	Female	45	–
45.jpg	Mrs. Manjula Bose	Female	56	NPDR
44.jpg	Mr. Ajay Kurre	Male	47	NPDR
36.jpg	Mr. Raj Mahuriya	Male	55	NPDR
57.jpg	Mrs. Trivani Maniksha	Female	43	NPDR
71.jpg	Mrs. Saluchana Bai	Female	51	NPDR

Table 2 represents the extracted feature points of retinal images and Table 3 represents the value of Euclidean Distance (ED).

Precision: 62%

Recall: 55%

Displayed result: "Patient has nonproliferative diabetic retinopathy".

Table 2 Extracted feature points of images

Image	Hue plane				Saturation plane				Value plane			
	E_1	σ_1	S_1		E_2	σ_2	S_2		E_3	σ_3	S_3	
Query	0.05	0.054	8.56		0.56	0.42	−0.56		0.37	0.31	−0.06	
45.jpg	0.05	0.053	8.62		0.54	0.45	−0.58		0.38	0.37	−0.08	
44.jpg	0.04	0.058	8.84		0.46	0.46	−0.58		0.32	0.35	−0.05	
36.jpg	0.03	0.057	8.91		0.58	0.52	−0.51		0.41	0.33	−0.08	
57.jpg	0.06	0.19	4.16		0.48	0.34	−0.47		0.38	0.40	−0.18	
71.jpg	0.05	0.21	4.23		0.47	0.42	−0.59		0.38	0.32	−0.19	

Table 3 Euclidean Distance (ED)

ED				
Image	Hue plane	Saturation plane	Value plane	Complete image
45.jpg	0.021	0.086	0.053	0.160
44.jpg	0.026	0.085	0.057	0.168
36.jpg	0.022	0.088	0.061	0.171
57.jpg	0.028	0.089	0.062	0.179
71.jpg	0.027	0.090	0.067	0.184

4 Conclusion

The initial identification of Diabetic Retinopathy (DR) cuts down the improvement of this disease and prevents vision loss. The proposed method decreases the difficult computational task and enhances the identification work. The efficiency also rises because the pictures are compared by both pixel and color instruction. The result represents that the few number of images retrieved with 62% precision and 55% recall decreases the operating time. As in this system, the smartphone is used as an indirect ophthalmoscope, and it can be available at any place and at any time for the underdeveloped region.

References

1. Kauppi, T., Kalesnykiene, V., Kamarainen, J., Lensu, L., Sorri, I., Raninen, A., Voutilainen, R., Uusitalo, H., Kalviainen, H., Pietila, J.: Diaretdb1: diabetic retinopathy database and evaluation protocol. In: Proceedings of Medical Image Understanding and Analysis (MIUA), pp. 61–65 (2007)
2. Wanjale, K., Borawake, T., Shahideep: Content based image retrieval for medical image techniques and storage methods review paper. Int. J. Comput. Appl. **1**, 19 (2012)
3. Abràmoff, M., Reinhardt, J., Russell, S., Folk, J., Mahajan, J., Niemeijer, M., Quellec, G.: Automated early detection of diabetic retinopathy. Ophthalmology **117**, 1147–1154 (2012). https://doi.org/10.1016/j.ophtha.2010.03.046
4. Kayal, D., Banerjee, S.: A new dynamic thresholding based technique for detection of hard exudates in digital retinal fundus image. In: Proceedings of the 1st International Conference on Signal Processing and Integrated Networks (SPIN'14), pp. 141–144 (2014)
5. Mahendran, G., Dhanasekaran, R., Narmadha Devi, K.N.: Identification of exudates for diabetic retinopathy based on morphological process and PNN classifier. In: Proceedings of the 3rd International Conference on Communication and Signal Processing (ICCSP'14), pp. 1117–1121 (2014)
6. Aqeel, A.F., Ganesan, S.: Automated algorithm for retinal image exudates and drusens detection, segmentation, and measurement. In: Proceedings of the IEEE International Conference on Electro/Information Technology (EIT'14), pp. 206–215 (2014)

7. Rokade, P.M., Manza, R.R.: Automatic detection of hard exudates in retinal images using haar wavelet transform. Int. J. Appl. Innov. Eng. Manag. **4**, 402–410 (2015)
8. Eagle Eye Centre. Diabetic Retinopathy Treatment (As on Date: 20/10/2017). http://www.eag leeyecentre.com.sg/service/diabetic-retinopathy/
9. Ramasamy, K., Raman, R., Tandon, M.: Current State of Care for Diabetic Retinopathy in India. Springer Science and Business Media, New York (2013). https://doi.org/10.1007/s1189 2-013-0388-6
10. Desai, C., Gupta, S., Shirgaon, P.: Diagnosis of diabetic retinopathy using CBIR method. In: International Journal of Computer Applications Proceedings on National Conference on Role of Engineers in National Building. pp. 12–15 (2016)
11. Kashyap, N., Sinha, G.R.: Image watermarking using 2-level DWT. Adv. Comput. Res. **4**, 42–45 (2012)
12. Singh, J., Kaleka, J.S., Sharma, R.: Different approaches of CBIR techniques. Int. J. Comput. Distributed Syst. **1**, 76–78 (2012)
13. Habashy, S.: Identification of diabetic retinopathy stages using fuzzy C-means classifier. Int. J. Comput. Appl. **77**, 0975–8887 (2013)
14. Hussain, C.A., Venkata Rao, D., Praveen, T.: Color histogram based image retrieval. Int. J. Adv. Eng. 63–66 (2013)
15. Kashyap, N., Singh, D.K.: Color histogram based image retrieval technique for diabetic retinopathy detection. In: 2nd IEEE International Conference for Convergence in technology (I2CT) (2017). https://doi.org/10.1109/i2ct.2017.8226238
16. Prasanna, P., Jain, S., Madabhushi, N.: Decision support system for detection of diabetic retinopathy using smartphones. In: International Conference on Pervasive Computing Technologies for Healthcare and Workshops, pp. 176–179 (2013)
17. Kashyap, N., Singh, D.K., Singh, G.K.: Mobile phone based diabetic retinopathy detection system using ANN-DWT. In: 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON) (2017). https://doi.org/10.1109/upcon.2017.8 251092
18. Stryjewski, T.P., MD, MPP, Shizuo Mukai, MD.: Talk about a smart ophthalmoscope. Ophthalmology Management. https://www.ophthalmologymanagement.com/issues/2015/Septemb er-2015/talk-about-a-smart-ophthalmoscope

Prediction of Diabetes Using Artificial Neural Network Approach

Suyash Srivastava, Lokesh Sharma, Vijeta Sharma, Ajai Kumar and Hemant Darbari

Abstract Diabetes is one of the major diseases of the population across the world. Diabetes is a chronic disease that occurs either when the pancreas does not produce enough insulin or when the body cannot efficiently use the insulin it produces. In 2014, 8.5% of adults aged 18 years and older had diabetes. In 2012, diabetes was the direct cause of 1.5 million deaths and high blood glucose was the cause of another 2.2 million deaths [1]. Over the time, diabetes can damage the heart, blood vessels, eyes, kidneys, and nerves. Early diagnosis can be made through a relatively inexpensive method of computation. In this paper, Machine Learning, a branch of Artificial Intelligence is used to analyze and make the diabetes prediction model. Various researchers have also been done to predict the diabetes machine learning algorithm, but this is an additional effort in the research work based on a specific type of patient in a specific community. In this research work, a data sample of Pima Indians was taken to predict the possibility of diabetes. Among several algorithms of Machine learning, Artificial Neural Network (ANN) was chosen for building the model to predict diabetes. This model is ideal for predicting the possibility of diabetes with 92% accuracy while tested with the sample test data. This model can achieve more accuracy if it is trained with a large sample training data in the future.

Keywords Diabetes · Machine learning · Artificial neural network · Pima Indian · Classification

S. Srivastava · L. Sharma (✉)
Manipal University Jaipur, Jaipur 303007, Rajasthan, India
e-mail: lokesh.sharma@jaipur.manipal.edu.edu
URL: http://jaipur.manipal.edu

V. Sharma · A. Kumar · H. Darbari
Center for Development of Advanced Computing (C-DAC), Pune, India

1   Introduction

Diabetes mellitus: More commonly referred to as "diabetes"—a chronic disease associated with abnormally high levels of the sugar glucose in the blood. Diabetes is due to one of the two mechanisms: Inadequate production of insulin (which is made by the pancreas and lowers blood glucose), or Inadequate sensitivity of cells to the action of insulin [2]. Diabetes mellitus also may develop as a secondary condition linked to another disease, such as pancreatic disease, a genetic syndrome, such as myotonic dystrophy, or drugs, such as glucocorticoids. Gestational diabetes is a temporary condition associated with pregnancy. In this situation, blood glucose levels increase during pregnancy but usually returns to normal after delivery [3]. Based on the data from the 2011 National Diabetes Fact Sheet, diabetes affects an estimate of 25.8 million people in the US, which is about 8.3% of the population. Additionally, approximately 79 million people have been diagnosed with pre-diabetes [4]. Pre-diabetes refers to a group of people with higher blood glucose levels than average but not high enough for a diagnosis of diabetes. Increased awareness and treatment of diabetes should begin with prevention. Many studies regarding diabetes prediction have been conducted for several years. The primary objectives are to predict what variables are the causes, at high risk, for diabetes and to provide a preventive action toward an individual at increased risk for the disease. Several parameters are considered for the study, which is explained in the next section. A healthy diet, regular physical activity, maintaining a healthy body weight and avoiding the use of tobacco can prevent or delay the onset of type 2 Diabetes Mellitus [5].

Nowadays, medical healthcare systems are rich in information. Wise use of these data can produce some predictive outcome. Pima Indians in Arizona have participated in a longitudinal diabetes study that has provided data publicly, which is used by many researchers for the study of diabetes. Various tools and techniques of Artificial Intelligence have been devised for early detection of diabetes after extracting information from the vast data set. Pardha Repalli [6], in their research work, predicted how likely the people with different age groups are affected by diabetes based on their lifestyle activities. They also found out factors responsible for the individual to be diabetic. Various classification methods were also used to detect Diabetes. Among them, some of the frequently used classifiers and clustering techniques are— Random forest, K-Means, J.48 algorithm, fuzzy approaches [7, 8]. Artificial Neural Network has also been widely used in medical research and studies for the disease prediction like Malaria and Cancer [9, 10]. Some research has also proved that ANN is also suited for the early diagnosis of diabetes [11]. Classification and prediction of the patient's condition based on risk factors are an application of artificial neural networks. The predictive capability of each neural network within the fully trained dataset was analyzed as well as the predictive capabilities of the neural networks on unseen data.

2 Methodology

The following subsections show the process or the method through which diabetes prediction model has developed.

2.1 Data Collection

Data has been collected from Kaggle's website (platform for predictive modeling and analytics competitions in which companies and researchers post data for research purpose) Pima Indians Diabetes Dataset, Class variable (0 or 1) Which is the original source of Research Center of National Institute of Diabetes and Digestive and Kidney Diseases, RMI Group Leader Applied Physics Laboratory The Johns Hopkins University [12]. This data has already been used for forecasting the onset of diabetes mellitus using ADAP learning algorithm. In this dataset, all patients here are females at least 21 years old of Pima Indian heritage. The total Number of Instances is 768, which is completely used in this study. It contains 8 attributes plus one class (Label) column. Each attribute is numeric-valued; attributes of this dataset are as follows:

- Number of times pregnant
- Plasma glucose concentration at 2 h in an oral glucose tolerance test
- Diastolic blood pressure (mm Hg)
- Triceps skinfold thickness (mm)
- 2-hour serum insulin (mu U/ml)
- Body mass index (weight in kg/(height in m^2)
- Diabetes pedigree function
- Age (years)
- Class variable (0 or 1).

This dataset also contains the Missing Attribute Values, which is handled in the next step of methodology (Preprocessing) using some statistical techniques. In class values, distribution is like if there is 1, then it interpreted as "tested positive for diabetes", if the class value is 0, it means "tested negative for diabetes" training and test data are divided into a certain number: 688 for training and 80 for testing data. The mean of each attribute is shown in Table 1. The screenshot of Training Sample Data and Testing Sample Data is shown in Figs. 1 and 2, respectively.

2.2 Data Preprocessing

Data value in different attributes are having some missing values. These missing values can lead to inaccurate result; also it may reduce the model accuracy. So to handle these missing value the mean of column method is used to replace 0

Table 1 Mean value of attributes

Attribute name	Mean value
Number of times pregnant	3.8
Plasma glucose concentration	120.9
Diastolic blood pressure (mm Hg)	69.1
Triceps skin fold thickness (mm)	20.5
2-hour serum insulin (mu U/ml)	79.8
Body mass index	32.0
Diabetes pedigree function	0.5
Age (years)	33.2

	A	B	C	D	E	F	G	H
1	1	140	74	26	180	24.1	0.83	23
2	1	144	82	46	180	46.1	0.34	46
3	8	107	80	0	0	24.6	0.86	34
4	13	158	114	0	0	42.3	0.26	44
5	2	121	70	32	95	39.1	0.89	23
6	7	129	68	49	125	38.5	0.44	43
7	2	90	60	0	0	23.5	0.19	25
8	7	142	90	24	480	30.4	0.13	43
9	3	169	74	19	125	29.9	0.27	31
10	0	99	0	0	0	25	0.25	22
11	4	127	88	11	155	34.5	0.6	28
12	4	118	70	0	0	44.5	0.9	26
13	2	122	76	27	200	35.9	0.48	26

Fig. 1 Screenshot of training sample data

with appropriate calculation [13]. To handle programmatically this missing values, NumPy package of Python was used to get mean function and manipulate the existing column array value from 0 to calculated result [14]. One thing is also important to prioritize the attribute, so that Artificial Neural Network calculate weight of each neurons (attribute) as per the given priority. Prioritizing attribute is need to get more accuracy of diabetes detection, which shows that which cause affects the diabetes detection on which priority. Table 2 show the attribute priority.

2.3 Data Preprocessing

For the prediction of diabetes, the model is built in core python using Artificial Neural Network(ANN) Algorithm.

	A	B	C	D	E	F	G	H	I	
1	6	148	72	35	0	33.6	0.63	50	1	
2	1	85	66	29	0	26.6	0.35	31	0	
3	8	183	64	0	0	23.3	0.67	32	1	
4	1	89	66	23	94	28.1	0.17	21	0	
5	0	137	40	35	168	43.1	2.29	33	1	
6	5	116	74	0	0	25.6	0.2	30	0	
7	3	78	50	32	88	31	0.25	26	1	
8	10	115	0	0	0	35.3	0.13	29	0	
9	2	197	70	45	543	30.5	0.16	53	1	
10	8	125	96	0	0	0	0.23	54	1	
11	4	110	92	0	0	37.6	0.19	30	0	
12	10	168	74	0	0	38	0.54	34	1	
13	10	139	80	0	0	27.1	1.44	57	0	
14	1	189	60	23	846	30.1	0.4	59	1	
15	5	166	72	19	175	25.8	0.59	51	1	
16	7	100	0	0	0	30	0.48	32	1	

Fig. 2 Screenshot of testing sample data

Table 2 Attribute priority

Attribute name	Priority[a]
Diastolic blood pressure (mm Hg)	1
Number of times pregnant	2
Age (years)	3
Triceps skin fold thickness (mm)	4
Diabetes pedigree function	5
Body mass index (weight in kg/(height in m)^2)	6
2-hour serum insulin (mu U/ml)	7
Plasma glucose concentration a 2 h in an oral glucose tolerance test	8

[a]means lower weight, 8 means higher weight

Python: Python is a general-purpose language, most data analysis functionality is available in packages like NumPy and pandas [15]. It has efficient high-level data structures and a simple but effective approach to object-oriented programming.

Artificial Neural Networks (ANN): Artificial Neural Networks (as shown in Fig. 3) is a family of models inspired by the biological neural network (the central nervous systems of animals, in particular, the brain) and are used to estimate or approximate functions that can depend on a large number of inputs and are unknown [16]. Artificial neural networks are presented as systems of interconnected "neurons" which exchange messages between each other. The connections have numeric weights that can be tuned based on experience, making neural nets adaptive to inputs and capable of learning [17]. Three types of parameters typically define an ANN:

1. The interconnection pattern between the different layers of neurons
2. The learning process for updating the weights of the interconnections
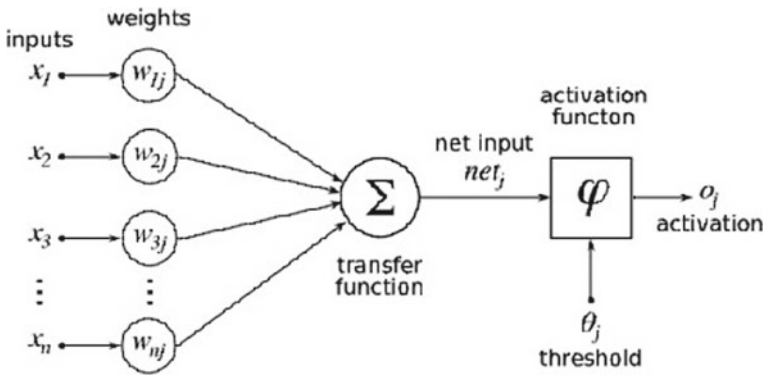
Fig. 3 Artificial neural networks (ANN) model

3. The activation function that converts a neuron's weighted input to its output activation.

Mathematically, a neuron's network function f(x) is defined as a composition of other functions gi(x), which can further be defined as a composition of other functions. This can be conveniently represented as a network structure, with arrows depicting the dependencies between variables. A widely used type of composition is the nonlinear weighted sum (given in Eq. 1),

$$f(x) = K \left(\sum_{i=1} w_i g_i(x) \right) \tag{1}$$

where K (commonly referred to as the activation function) is some predefined function, such as the hyperbolic tangent.

This ANN algorithm itself is having various components to simulate the values and learn using the history data for better prediction [15]. These components were written in python as a function to call and execute:

1. **Read_CSV**(): Training data file Diabetes_TrainingData.csv and converting in array to read by python. Using Pandas package [18] and related function array can be formed to easily supply as input Training values.
2. **Assigning Random weight**(): INPUT_NEURONS variables used to weight for input Hidden (WiH), Then HiD (Hidden input Neurons) to HIDDEN_NEURONS for assigning weight. Finally, transfer HIDDEN_NEURONS weights to OUTPUT_NEURONS.
3. **NeuralNetwork**(): First define number of epoch, which is epoch = 0 for initial and give training rate, which is TRAINING_REPS should always be greater than epoch. TrainInputs[ ] is an array which stores weight and input neurons values, trainOutput[ ] stores output hidden neurons values and learn for new values.

$$\frac{1}{1 + e^{-x}}$$
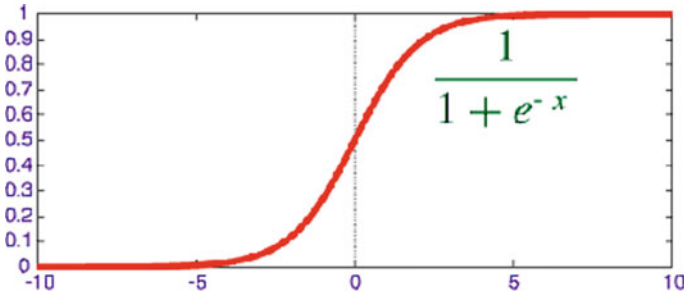
Fig. 4 Sigmoid function

4. **feedForward()**: First Neurons values are transferred to hidden layer neurons, where these values for each neurons are multiplied and stored in actual variable, which is the sum of all the multiplied neurons and weight value.
5. **backPropagate()**: Backpropagation is a method to calculate the gradient of the loss function with respect to the weights in an artificial neural network [19]. It is commonly used as a part of algorithms that optimize the performance of the network by adjusting the weights. Here, backpropagation call the sigmoidDerivative function and define LEARN_RATE (Initially lower value), then calculate error in each sigmoid layer.
6. **sigmoid(val)**: The sigmoid function is a type of activation function for artificial neurons. The most basic activation function is the heaviside (binary step, 0 or 1, high or low). The sigmoid function (a special case of the logistic function) and its formula looks as shown in Fig. 4.
7. **ErrorCal()**: Here, the final error shows the model accuracy and Actual and Predicted values, which is finally 8% at the end of building model and get prediction.
8. **Graph_Plot()**: This shows the result in graphical format. Package MatPlotLib used to plot the graph of actual and predicted values [20]. This graph shows under result section.

3 Accuracy Measurement of Model

Root Mean Squared Error (RMSE) and ROC (Receiver Operating Characteristic) performance parameters of ANN model considered for the analysis of accuracy [21]. ANN produced Root Mean Squared Error 0.39 and ROC area 0.88. As per performance guide for classifications accuracy, it shows that ROC > 0.80 is considered GOOD classifier and ROC 0.77 as FAIR. The classifier should achieve ROC value closer to 1 for higher accuracy of making prediction. The Screenshot of the output shown in Fig. 5.

Fig. 5 Output
prediction—actual versus
predicted


```
Actual: 1.0 Predicted: 0.87
Actual: 1.0 Predicted: 0.89
Actual: 1.0 Predicted: 0.78
Actual: 0.0 Predicted: 0.22
Actual: 0.0 Predicted: 0.21
Actual: 0.0 Predicted: 0.3
Actual: 0.0 Predicted: 0.26
Actual: 0.0 Predicted: 0.32
Actual: 0.0 Predicted: 0.27
Actual: 1.0 Predicted: 0.76
Actual: 1.0 Predicted: 0.69
Actual: 0.0 Predicted: 0.22
Actual: 0.0 Predicted: 0.31
Actual: 1.0 Predicted: 1.23
```

4   Result and Conclusion

Result of Diabetes Prediction model is shown as graphical format, where red line
shows the predicted values of Diabetes and blue lines shows the actual value. Here, it
is considered that predicted value which closer to 1 and above 0.5 is considered as 1
(Positive) whereas, closer to 0 and below 0.5 is considered as 0 (Negative) Prediction
capacity of ANN-based model and can predict the possibility of developing diabetes
in the community of Pima Indians. It is also observed that learning with more sample
dataset can improve the accuracy with reducing error rate. Using relative values for
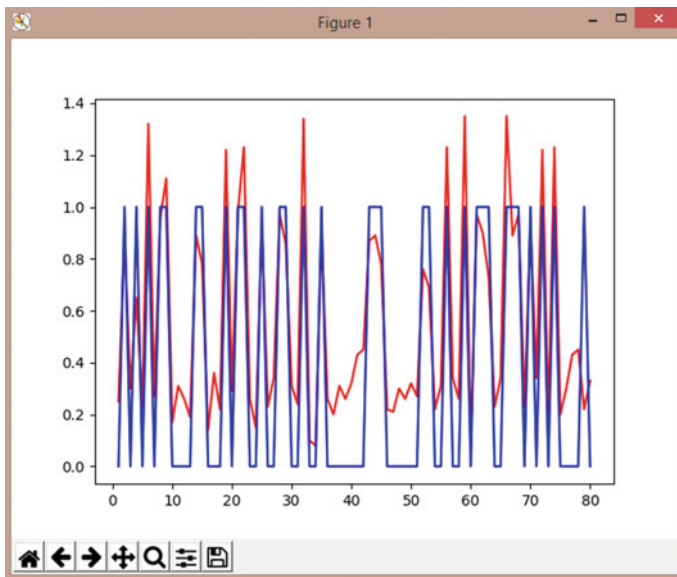
Fig. 6 Graph between predicted and actual value

the parameters for other demographic areas can be scaled up to the large areas. This model is useful for health policy makers, who can take preventive action before the occurrence of diabetes in large number (Fig. 6).

References

1. World Health Organization, Media Centre, Diabetes. http://www.who.int/mediacentre/factsheets/fs312/en/
2. http://www.medicinenet.com/script/main/art.asp?articlekey=2974
3. https://www.britannica.com/science/diabetes-mellitus
4. American Diabetes Association, Diabetes Statistics, June 2012. http://www.diabetes.org/diabetes-basics/diabetesstatistics/
5. http://www.who.int/topics/diabetesmellitus/en/
6. Repalli, P.: Prediction on Diabetes Using Data mining Approach. Oklahoma State University (2011)
7. Polat, K., Günes, S.: An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease. Digit. Signal Process. **17**, 702–710 (2007)
8. Karegowda A.G., Manjunath, A.S., Jayaram, M.A., Application of genetic algorithm optimized neural network connection weights for medical diagnosis of Pima Indians diabetes. Int. J. Soft Comput. (IJSC) **2**(2) (2011)
9. Sharma, V., Kumar, A.: Malaria outbreak prediction model using machine learning. Int. J.
10. Ganesan, N., Venkatesh, K., Rama, M.A.: Application of Neural Networks in Diagnosing Cancer Disease Using Demographic Data
11. Olaniyi, E.O., Adnan, K.: Onset diabetes diagnosis using artificial neural network. Int. J. Sci. Eng. Res. **5**(10) (2014)
12. Pima Indians Diabetes Database-Predict the onset of diabetes based on diagnostic measures. https://www.kaggle.com/uciml/pima-indians-diabetes-database
13. Saar-Tsechansky, M., Provost, F.: Handling missing values when applying classification models. J. Mach. Learn. Res. **8**, 1625–1657 (2007)
14. http://www.numpy.org/
15. The Python Tutorial. https://docs.python.org/3/tutorial/
16. https://en.wikipedia.org/wiki/Artificial_neural_network
17. http://machinelearningmastery.com/a-tour-of-machine-learning-algorithms
18. http://pandas.pydata.org/
19. http://neuralnetworksanddeeplearning.com/chap2.html
20. https://matplotlib.org/
21. https://www.cs.cornell.edu/courses/cs578/2003fa/performance_measures.pdf

Relative Stability Analysis of Two-Dimensional Linear Systems with Complex Coefficients

P. Ramesh

Abstract The paper presents two algebraic criteria for the relative stability analysis of two-dimensional systems, which are represented in the form of characteristics equation and further the equivalent single-dimensional characteristics equation is formed from the given two-dimensional characteristics equation. When the relative stability analysis is done based on the damped frequency of oscillation, the characteristic equations with complex coefficients arise. These complex coefficients are used in two different ways to form the modified Routh's tables for the two schemes named as sign pair criterion I and sign pair criterion II. It is found that the proposed algorithms offer computational simplicity compared to other algebraic methods and are illustrated with suitable examples, and the results were verified using MATLAB.

Keywords Complex coefficients · Relative stability · Routh's table

1 Introduction

The relative stability is important in obtaining an acceptable transient response of the linear continuous system. Mathematically, the roots of the system characteristics equation lie in a certain sector in the left half of the s-plane. Sreekala et al. posted a new procedure in which complex polynomial is used to form a modified Routh's table for the schemes named as Sign Pair Criterion I (SPC I) and Sign Pair Criterion II (SPC II) [1].

A new Routh-like algorithm for determining the number of Right-Half-Plane (RHP) roots of the polynomial with the complex case is presented by Agashe [2]. Benidir et al. had proposed a method to determine and form the Routh's table with the complex coefficient that is extended to the special case of hide leading elements [3].

P. Ramesh (✉)
Department of EEE, University College of Engineering, Anna University, Ramanathapuram, Tamil Nadu, India
e-mail: rameshucermd@gmail.com

Hwang et al. have revealed the damping margin and damped frequency of oscillation based on the relative stability analysis [4]. Usher had introduced the method to convert complex coefficient equation into the real coefficient equation, to analyze the stability of the system. The same method proposed by the analysis of damping margin is complicated, and hence complex coefficients are used in SPC I and SPC II [5].

Chen et al. had proposed the original Routh's table dealing with real polynomial, which is further investigated for complex polynomial [6]. Sreekala et al. had presented an extending Routh's criterion, a stability criterion is formulated which is directly applicable to handle the characteristic equation having complex coefficients [7].

Two-dimensional recursive digital filters do not satisfy the classical definitions for two-dimensional periodicity presented by Bauer et al. in [8]. Sivanandam et al. have proposed the application of 1-D stability criteria and extension of the Routh's stability test to the stability analysis of digital filters [9]. Serban et al. had proposed a BIBO stability algorithm that is based on a sufficient condition for BIBO stability of n-dimensional filters [10].

The beauty of the Routh's algorithm is finding the relative stability of the system without determining the roots of the system. The formation of Routh's table is done by retaining the "j" terms of the complex coefficients and the stability analysis is done using Sign Pair Criterion I (SPC I). The proof is given in [7]. In the second scheme, a geometrical procedure is presented which is named as Sign Pair Criterion II (SPC II) and is formulated with the help of "Modified Routh's table" after separating the real and imaginary parts of the characteristic equation by substituting s = "jω". Applying Routh–Hurwitz criterion, the number of the roots of F(s) = 0 having positive real part can be revealed. The proof for SPC II is given in [1]. Then, by the use of the proposed schemes relative stability is analyzed in a most simple way regardless of the order of the system. The computational simplicity is illustrated with examples.

2 Proposed Method

2.1 Sign Pair Criterion I [SPC I]

With the coefficient of s^n as positive, the characteristic equation $C(s)$ can be written as

$$C(s) = s^n + (a_1 + jb_1)s^{n-1} + (a_2 + jb_2)s^{n-2} + \cdots + (a_k + jb_k) = 0.$$

The first two rows of Routh-like table are written as shown below

1	jb_1	a_2	jb_3	a_4	...
a_1	jb_2	a_3	jb_4	a_5	...

Table 1 Routh-like table for SPC I

1	jb_1	a_2	jb_3	a_4	$\ldots$
a_1	jb_2	a_3	jb_4	a_5	$\ldots$
r_{31}	r_{32}	r_{34}	r_{35}	$\ldots$	
r_{41}	r_{42}	r_{43}	r_{44}	$\ldots$	
r_{51}	r_{52}	r_{53}	$\ldots$		
r_{61}	r_{62}	r_{63}	$\ldots$		
r_{71}	r_{72}	$\ldots$			
r_{81}	$\ldots$				
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	

Applying the standard Routh multiplication rule, the subsequent elements of Routh-like table are computed and the table is computed as given in Table 1.

To formulate (SPC I), the elements in the first column of Routh-like table are considered to form the pairs (P_i), $i = 1, 2, \ldots, n$ as depicted below

$$P_1 = (1, a_1) \quad P_2 = (r_{31}, r_{41}) \quad P_3 = (r_{51}, r_{61}) \quad P_4 = (r_{71}, r_{81}) \text{ and so on.}$$

2.2 Sign Pair Criterion II (SPC II)

The characteristic equation can be rewritten as

$$C(s) = s^n + (a_1 + jb_1)s^{n-1} + (a_2 + jb_2)s^{n-2} + \cdots + (a_k + jb_k) = 0. \tag{1}$$

Substituting $s = j\omega$ in the above equation

$$C(j\omega) = (j\omega)^n + (a_1 + jb_1)(j\omega)^{n-1} + (a_2 + jb_2)(j\omega)^{n-2} + \cdots + (a_k + jb_k) = 0,$$
$$= R(\omega) + jI(\omega) = 0 \tag{2}$$

where $R(\omega)$ is the real part and $I(\omega)$ is the imaginary part of $C(j\omega)$.

For the sake of simplicity, the polynomials $R(\omega)$ and $I(\omega)$ can be written is given as follows.

$$R(\omega) = \left(A_0\omega^n + A_1\omega^{n-1} + A_2\omega^{n-2} + \cdots + A_n\right), \tag{3}$$
$$I(\omega) = \left(B_0\omega^n + B_1\omega^{n-1} + B_2\omega^{n-2} + \cdots + B_n\right). \tag{4}$$

Using the coefficients of $R(\omega)$ and $I(\omega)$ polynomials, the second form of Routh-like table can be formulated as Table 2.

Table 2 Routh-like table for SPC II

A_0	A_1	A_2	$\ldots$	A_n
B_0	B_1	B_2	$\ldots$	B_n
c_0	c_1	c_2	$\ldots$	
d_0	d_1	d_2	$\ldots$	
e_0	e_1	e_2	$\ldots$	
f_0	f_1	f_2	$\ldots$	
g_0	g_1	$\ldots$	$\ldots$	
$\vdots$	$\vdots$	$\vdots$		

Algorithm for the Proposed Approach

1. Get the characteristic equation $C(s) = 0$ with complex coefficients.
2. With $s = j\omega$, form $C(j\omega) = R(\omega) + jI(\omega) = 0$.
3. Use the coefficients of $R(\omega)$ and $I(\omega)$, and form the first and second rows of Routh-like table.
4. If the first element in the first row is negative, multiply the full row elements by -1.
5. If the first element in the second row is zero, interchange first and second rows and multiply all elements in the second row by -1.
6. Follow the common Routh's multiplication rule to get the complete table with "$2n+1$" rows.
7. If any element of the first column starting from third, comes zero, it is replaced by a small value $+0.01$.
8. If all the elements in a row become zero, then the auxiliary polynomial is formed using the previous row elements and differentiated once; the coefficients of this modified polynomial are entered instead of zeros and the table is completed by applying the Routh's multiplication rule.
9. Get "n" sign pairs using the first column elements starting from second row.

For forming SPC II, the elements in the first column of Routh-like table are considered; but A_0 is not taken into account and with $2n$ elements the sign pairs are developed as

$$P_1 = (B_0, c_0) \quad P_2 = (d_0, e_0) \quad P_3 = (f_0, g_0), \ldots, P_n.$$

The given system represented by Eq. (1) is stable if the sign of each element of the pairs $P_1, P_2, P_3, \ldots, P_n$ remains the same.

3 Illustrative Examples

3.1 Illustration 1 [11]

$$C(s) = (4s_1 + 4)s_2^2 + (s_1^2 + 5s_1 + 1)s + (s_1^2 + s_1 + 4) = 0$$

$$C(s) = 4s_1 s_2^2 + 4s_2^2 + s s_1^2 + 5s_1 s + s + s_1^2 + s_1 + 4 = 0.$$

Convert the two-dimensional characteristics equation into one-dimensional characteristics equation [11].

$$C(s) = \frac{4}{s_1 s_2^2} + \frac{4}{s_2^2} + \frac{1}{s s_1^2} + \frac{5}{s_1 s} + \frac{1}{s} + \frac{1}{s_1^2} + \frac{1}{s_1} + 4 = 0$$

$$\frac{1}{s_1} = \frac{1}{s_2} = \frac{1}{s}$$

$$C(s) = 4s^3 + 2s^2 + 10s + 5 = 0.$$

For a choice of $\alpha = 1$ and with the substitution of $C(s) = C(j(S + \alpha))$

$$C'(s) = S^3 + (3 - j0.5)S^2 + (3 - j)S + (1.5 + j0.75) = 0.$$

Applying SPC I

The Routh table is formed as

+1	−j0.5	3	j0.75
3	−j	1.5	
−0.1666j	2.5	j0.75	
−46.01j	−2.33		
−69.04	0.75j		
−2.8328			

The sign pairs are formed as

$$P_1 = (+1, +3), \quad P_2 = (-0.1666j, -46.01j), \quad P_3 = (-69.04, -2.8328).$$

Since all the three pairs satisfy SPC I, the system is relatively stable.

Application of SPC II

$$C'(S) = \left(-3\omega^2 + \omega + 0.5\right) + j\left(-\omega^3 + 0.5\omega^2 + 3\omega + 0.75\right) = 0$$

The Routh table is formed as

0	−3	1	0.5
−1	0.5	3	0.75
−3	1	0.5	
0.1666	2.833	0.75	
52.0144	14.005		
2.788	0.75		
0.0126			

The sign pairs are formed as

$$P_1 = (-1, -3), \quad P_2 = (0.1666, 52.0144), \quad P_3 = (2.788, 0.0126)$$

Since all the three pairs satisfy SPC II, the system is stable.

Output verification using MATLAB

SPC I

```
Input coefficients of characteristic equation,i.e:[an an-1 an-2 ... a0]= [1 3 -0.5i -i 3 1.5 0.75i]


----------------------------------------
--------The Routh-Hurwitz array is:--------

m =

    1.0000                  0 - 0.5000i   3.0000              0 + 0.7500i
    3.0000              -3.0000           1.5000              0
    1.0000 - 0.5000i    2.5000            0 + 0.7500i         0
   -9.0000 - 3.0000i    2.4000 - 1.8000i  0                   0
    2.5500 - 0.3500i    0 + 0.7500i       0                   0
    1.1774 + 0.6792i    0                 0                   0
        0 + 0.7500i     0                 0                   0
```

Output verification using MATLAB

SPC II

```
Input coefficients of characteristic equation,i.e:[an an-1 an-2 ... a0]= [0 -1 -3 0.5 1 3 0.5 0.75]


----------------------------------------
--------The Routh-Hurwitz array is:--------

m =

       0   -3.0000    1.0000    0.5000
 -1.0000    0.5000    3.0000    0.7500
 -3.0000    1.0000    0.5000         0
  0.1667    2.8333    0.7500         0
 52.0000   14.0000         0         0
  2.7885    0.7500         0         0
  0.0138         0         0         0
  0.7500         0         0         0
```

3.2 Illustration 2 [4]

$$C(s) = 6(2 + z_1) + 5(2 + z_1)z_2 + (2 + z_1)z_2^2 = 0$$

$$C(s) = 12 + 6z_1 + 10z_2 + 5z_1z_2 + 2z_2^2 + z_1z_2^2 = 0.$$

Convert the two-dimensional characteristics equation into one-dimensional characteristics equation [11].

$$C(s) = 12 + \frac{6}{z_1} + \frac{10}{z_2} + \frac{5}{z_1z_2} + \frac{2}{z_2^2} + \frac{1}{z_1z_2^2} = 0$$

$$\frac{1}{z_1} = \frac{1}{z_2} = \frac{1}{z}$$

$$C(s) = 12z^3 + 16z^2 + 7z + 1 = 0.$$

For a choice of $\alpha = 1$ and with the substitution of $C(s) = C(j(S + \alpha))$

$$C'(s) = S^3 + (3 - j1.33)S^2 + (-0.5833 - j5.66)S + (0.4167 - j1.2461) = 0.$$

Applying SPC I

The Routh table is formed as

+1	$-j1.33$	$-j1.2461$	$j0.75$
3	$-j5.66$	0.4167	
$0.556j$	-0.72	$-j1.2461$	
$-9.54j$	7.140		
-0.3038	$-j1.2461$		
46.27			

The sign pairs are formed as

$$P_1 = (+1, +3), \quad P_2 = (0.556j, -9.54j), \quad P_3 = (-0.3038, 46.27).$$

Since $P_2 \& P_3$ fails to obey SPC I, then the system is unstable.

Application of SPC II

$$C(s) = \left(-3\omega^2 + 5.66\omega + 0.4167\right) + j\left(-\omega^3 + 1.33\omega^2 - 0.5833\omega - 1.2461\right) = 0$$

The Routh table is formed as

0	-3	5.66	0.4167
-1	1.33	-0.5833	-1.2461
-3	5.66	0.4167	
-0.5566	-0.72	-1.2467	
9.54	7.132		
-0.303	-1.2461		
-32.10			

The sign pairs are formed as

$$P_1 = (-1, -3), \quad P_2 = (-0.5566, 9.54), \quad P_3 = (-0.303, -32.10)$$

Since P_2 fails to obey SPC II, then the system is unstable.

Output verification using MATLAB

SPC I

```
Input coefficients of characteristic equation,i.e:[an an-1 an-2 ... a0]= [1 3 -1.33i -5.66i -0.5833 0.4167 -1.2461i]


-----------------------------------------
--------The Routh-Hurwitz array is:--------


m =

    1.0000              0 - 1.3300i  -0.5833            0 - 1.2461i
    3.0000              0 - 5.6600i   0.4167            0
           0 + 0.5567i  -0.7222                0 - 1.2461i  0
           0 - 9.5521i   7.1322                0           0
   -0.3066              0 - 1.2461i            0           0
  -31.6954              0                      0           0
           0 - 1.2461i  0                      0           0
```

Output verification using MATLAB

SPC II

```
Input coefficients of characteristic equation,i.e:[an an-1 an-2 ... a0]= [0 -1 -3 1.33 5.66 -0.5833 0.4167 -1.246


-----------------------------------------
--------The Routh-Hurwitz array is:--------


m =

        0   -3.0000    5.6600    0.4167
   -1.0000    1.3300   -0.5833   -1.2461
   -3.0000    5.6600    0.4167         0
   -0.5567   -0.7222   -1.2461         0
    9.5521    7.1322         0         0
   -0.3066   -1.2461         0         0
  -31.6954         0         0         0
   -1.2461         0         0         0
```

4   Conclusion

The relative stability analysis of a two-dimensional linear system represented in the form of their respective characteristics equations has been performed with the help of the proposed SPC I and SPC II schemes. The transient response behavior of a two-dimensional linear system having absolute stability has been carried out with the help of damped frequency of oscillations. For the analysis of relative stability, the

proposed Routh-like tables are formed and ascertained with the application of SPC I and SPC II. The proposed algebraic criteria are simple and direct in application compared to other schemes and the results were verified using MATLAB.

References

1. Sreekala, K., Sivanandam, S.N.: Relative stability analysis of linear systems based on damped frequency of oscillation. IOSR-JEEE 01–05 (2016)
2. Agashe, S.D.: A new general Routh-like algorithm to determine the number of RHP roots of a real or complex polynomial. IEEE Trans. Autom. Control **30**(4), 406–409 (1985)
3. Bendir, M., Picinbono, B.: The extended Routh's table with complex case. IEEE Trans. Autom. Control **36**(2), 253–256 (1991)
4. Hwang, H.H., Tripathi, P.C.: Generalization of the Routh-Hurwitz criterion and its applications. Electronicsletters **6**(13), 410–411 (1970)
5. Usher, I.A.: New application of the Hurwitz Routh stability criteria. Trans. AIEE Part-1. Commun. Electr. **76**(5), pp. 530–533 (1957)
6. Chen, S., Tsai, Jason S.H.: A new tabular form for determine root distribution of a complex polynomial with respect to the imaginary axis. IEEE Trans. Autom. Control **38**(10), 1536–1541 (1993)
7. Sreekala, K., Sivanandam, S.N.: An algebraic approach for stability analysis of linear system having complex coefficients polynomial. Int. Rev. Mech. Eng. **44**(3), 1213–1216 (2012)
8. Bauer, P., Jury, E.I.: Non-periodic modes in two dimensional (2-D) recursive digital filters under finite word length effects. IEEE Trans. Circuits Syst. **36**(7) (1989)
9. Sivanandam, S.N.: Sivakumar, D.: A new algebraic test procedure for stability analysis of multidimensional shift invariant digital filters. In: IEEE Electrical and Electronic Technology Conference, pp. 33–38. IEEE (2001)
10. Serban, I., Najim, M.: A new multidimensional Schur-Cohn type stability criterion. In: IEEE American Control Conference, NY, USA, pp. 5533–5538 (2007)
11. Ramesh, P., Manikandan, V.: Stability analysis of two-dimensional linear time invariant discrete systems. World Appl. Sci. J. **32**(8), 1506–1512 (2014)
12. Jury, E.I.: Inners approach to some problems of system theory. IEEE Trans. Autom. Control **16**(1971), 233–240 (1971)

An Algebraic Test for Analyzing Aperiodic Stability of Two-Dimensional Linear Systems with Complex Coefficients

P. Ramesh

Abstract To evaluate the performance of a linear time-invariant system various measures are available. In this work, it is proposed a two algebraic stability criteria using Routh-like table with complex coefficients to check whether the given two-dimensional linear system is aperiodically stable or not. To test the aperiodic stability of the given two-dimensional linear time-invariant discrete system represented in the form of its characteristic equation is converted into equivalent one-dimensional characteristics equation, whose characteristics equation having complex coefficients. Using this complex coefficient, a modified Routh table has been formed by two schemes named as Sign pair Criterion I (SPC-I) and Sign Pair Criterion II (SPC-II). These two criteria are very convenient compared to other available algebraic schemes for the analysis of aperiodic stability of two-dimensional linear systems.

Keywords Aperiodic stability · Routh's table · Sign pair criterion
Complex coefficients

1 Introduction

Information about the aperiodic stability of a two-dimensional linear system is of paramount importance for any design problem. This is generally used in communication systems; Audio systems and Instrumentation systems.

The stability of complex polynomials has been carried out by Routh–Hurwitz method was investigated in [1–6]. Frank [1] and Agashe [2] had proposed a new Routh-like algorithm to find the number of RHP roots in the complex case. Benidt and Picinbon [3] had presented an extended Routh table which deems singular cases of vanishing leading array element. By adding intermediate rows in the Routh array, Chen and Tsai [4] presents a tabular column, which is also a complicated one. Hashem

P. Ramesh (✉)
Department of EEE, University College of Engineering,
Anna University, Ramanathapuram, Tamil Nadu, India
e-mail: rameshucermd@gmail.com

[5] had proposed the stability analysis of complex polynomials using the J-fraction expansion, Hurwitz Matrix determinant and also generalized Routh's Array. Formation of Routh's Table by retaining the 'j' terms of the complex coefficients and the stability of one-dimensional system analysis using Sign Pair Criterion had proposed in [6].

To investigate the aperiodic stability, a generalized method had proposed in [8] by Fuller. Romonov [7] presents a new transformation to determine the aperiodic behavior of the linear system which results in complex coefficient polynomials. A favored but ponderous method for finding the number of real roots in a polynomial with real coefficients is by Sturm's theorem [9]. Bar-Itzhack and Calise had proposed [10] a three-steps transformation procedure which develops a polynomial whose number of right-hand plane poles equals the number of complex roots present in the original polynomial.

The additional significance of the two criteria is, it can be used to count the number of complex roots of a system having real coefficients which are not possible by the use of original Routh's table. These procedures can also be used for the design of linear systems. In the proposed methods, the characteristic equation having real coefficients are first converted to a complex coefficient equation using Romonov's transformation. These complex coefficients are used in two different ways to form the modified Routh's table for the two schemes named as sign pair criterion I (SPC-I) and sign pair criterion II (SPC-II). It is found that the proposed algorithms offer computational simplicity compared to other algebraic methods and is illustrated with suitable examples. The developed MATLAB program make the analysis most simple.

2 Aperiodic Stability

To check the aperiodic stability of a linear discrete system is represented by characteristics equation $F(s) = 0$, with real and complex coefficients is analyzed using Romonov transformation [7] suggested a transformed polynomial of $F(s)$ into a complex polynomial is given by

$$F'(s) = \left[F(s)_{s=js} + j \left\{ \frac{dF(js)}{d(js)} \right\} \right]$$

$$F'(s) = F(js) + j \left\{ \frac{dF(js)}{d(js)} \right\} \tag{1}$$

After this transformation, the real coefficient polynomial is converted to complex coefficient polynomial and the two proposed schemes SPC-I and SPC-II can be used for the aperiodic stability analysis of two-dimensional linear systems.

3   Proposed Method

3.1   Sign Pair Criterion I (SPC-I)

Let $F(S) = 0$ be the nth degree characteristic equation of a linear time-invariant system and written as

$$F(S) = S^n + (a_1 + jb_1)S^{n-1} + (a_2 + jb_2)S^{n-2} + \cdots + (a_n + jb_n) = 0 \quad (2)$$

where $(a_i + jb_i)$ are the complex coefficients. The first two rows of Modified Routh's Table for Eq. (2) are shown as follows:

1	jb_1	a_2	jb_3 …
a_1	jb_2	a_3	jb_4 …

Applying Routh multiplication rule, the complete table with '2n' number of rows are formed.

Using Table 1, pairs are formed using the first column and starting from the first row.

$$P_1 = (1, a_1), \ P_2 = (jc_1, jd_1), \ P_3 = (g_1, h_1)$$

It is ascertained that the two elements in each pair have to maintain same sign for the system to be stable. The proof is given in [6].

3.2   Sign Pair Criterion II (SPC-II)

In this approach, the characteristic equation is given in 's' domain is converted to the frequency domain by replacing $S = 'j\omega'$ and the real and imaginary parts are

Table 1 Routh-like table for SPC-I

1	jb_1	a_2	jb_3 …
a_1	jb_2	a_3	jb_4 …
jc_1	c_2	jc_3 …	
jd_1	d_2	jd_3 …	
⋮	⋮	⋮	
g_1			
h_1			

separated. The coefficients of real parts are used to form the first row of 'Modified Routh's table' and the coefficients of imaginary parts are the elements of the second row of the table. By applying the normal Routh multiplication rule, the complete Routh's Array is formed with '$2n + 1$' number of rows, where 'n' is the order of the system.

Algorithm for the proposed approach

1. Get $F(S) = 0$ with complex coefficients.
2. With $S = j\omega$, form $F(j\omega) = R(\omega) + jI(\omega)$.
3. Use the coefficients of $R(\omega)$ & $I(\omega)$, form the first and second rows of Routh's table.
4. If the first element in the first row is negative, multiply the full row elements by -1.
5. If the first element in the second row is zero, interchange first and second row.
6. Follow the Common Routh's multiplication rule to get the complete table with '$2n + 1$' rows.
7. Get 'n' sign pairs using the first column elements starting from the second row. Consider the nth degree characteristic equation $F(S) = 0$ of a linear time-invariant system with the complex. Coefficients,

$$F(S) = S^n + (A_1 + jB_1)S^{n-1} + (A_2 + jB_2)S^{n-2} + \cdots + (A_n + jB_n) = 0 \quad (3)$$

where $(A_i + jB_i)$ are the complex coefficients. By substituting $S = \text{'}j\omega\text{'}$ and separating real and imaginary parts, the characteristic equation can be written as follows:

$$F(j\omega) = R(\omega) + jI(\omega) = 0$$
$$= a_1\omega^n + a_2\omega^{n-1} + a_3\omega^{n-2} + \cdots + j\left(b_1\omega^n + b_2\omega^{n-1} + b_3\omega^{n-2} \cdots\right) \quad (4)$$

Using the coefficients of real and imaginary parts, the first two rows of Modified Routh's Table is formed

a_1	a_2	a_3	$a_4 \ldots$
b_1	b_2	b_3	$b_4 \ldots$

The direct Routh's multiplication rule is applied and the complete Modified Routh's Table with '$2n+1$' number of rows is formed as

From the elements of the first column of Table 2, starting from the second row, the following pairs may be grouped, respectively: $P_1 = (b_1, c_1)$, $P_2 = (d_1, e_1)$, $P_n = (g_1, h_1)$. It is ascertained that the two elements in each pair have

Table 2 Routh-like table for SPC-II

a_1	a_2	a_3	a_4 ...
b_1	b_2	b_3	b_4 ...
c_1	c_2	c_3 ...	
d_1	d_2	d_3 ...	
e_1	e_2	...	
$\vdots$	$\vdots$	$\vdots$	
g_1			
h_1			

to maintain a same sign for all the roots of $F(s)=0$ to be real and distinct. If one pair fails to satisfy this condition, it is inferred that there exist two numbers of complex roots (one complex conjugate pair) for $F(s)=0$ and the system is aperiodically unstable.

4 Illustrative Examples

4.1 Example 1 [11]

$$C(Z_1 Z_2) = 1 - 0.75Z_1 - 0.5Z_2 + 0.3Z_1 Z_2$$

Convert two-dimensional characteristics equation into one-dimensional characteristics equation [11]

$$C\left(\frac{1}{z_1}, \frac{1}{z_2}\right) = F(z) = 1 - \frac{0.7}{z_1} - \frac{0.5}{z_2} + \frac{0.3}{z_1 z_2} = 0$$

Substitute $z_1 = z_2 = z_3 = s$

$$F(s) = 1 - \frac{0.7}{s} - \frac{0.5}{s} + \frac{0.3}{s^2} = 0$$

$$F(s) = s^2 - 1.2s + 0.3 = 0$$

As per Eq. (1), substitute $s = js$

$$F'(js) = (js)^2 - (12)js + 0.3 + j[2(js) - 1.2] = 0$$

$$F'(s) = -s^2 - 1.2js + 0.3 - 2s - 1.2j = 0$$

Separate s^2, s terms also divide by -1

$$F'(s) = s^2 + (2 + 1.2j)s + (-0.3 + 1.2j) = 0$$

Application of SPC-I:

1	1.2J	-0.3
2	1.2J	0
0.6J	-0.3	0
2.2J	0	
-0.3	0	
0		

$$P_1 = (+1, +2), \quad P_2 = (+j0.6, +j2.2), \quad P_3 = (-0.3, 0)$$

The Sign Pairs are P$_1$ = (1, 2), P$_2$ = (j0.6, j2.2) and P$_3$ = (-0.3, 0). It is noted that the two elements in each pair have the same sign and obey SPC-I. Hence, the system is aperiodically stable.

Application of SPC-II:

Substitute $s = j\omega$ at $F'(s)$

$$F'(j\omega) = -\omega^2 + 2j\omega - 1.2\omega - 0.3 + 1.2j = 0$$

Separate real and imaginary terms

-1	-1.2	-0.3
0	2	1.2

Here, in the first row $(-)$ sign occur so multiply by (-1) on the first row and also second row having zero on the first column so interchange first and second rows

0	2	1.2
1	1.2	0.3
2	1.2	0
0.6	0.3	0
0.2	0	
0.3		
0		

$$P_1 = (+1, +2), \quad P_2 = (+0.6, +0.2), \quad P_3 = (+0.3, +0)$$

The Sign Pairs are formed as $P_1 = (+1, +2)$, $P_2 = (+0.6, +0.2)$ and $P_3 = (+0.3, +0)$. It is noted that the two elements in each pair have the same sign and obey SPC-II. Hence, the system is aperiodically stable.

Output Verification Using MATLAB

SPC-I

```
Input coefficients of characteristic equation,i.e:[an an-1 an-2 ... a0]= [1 2 1.2i 1.2i -0.3]

----------------------------------------
--------The Routh-Hurwitz array is:--------

m =

    1.0000              0 + 1.2000i   -0.3000
    2.0000              0 + 1.2000i      0
        0 + 0.6000i  -0.3000            0
        0 + 0.2000i      0              0
   -0.3000              0              0
```

SPC-II

```
Input coefficients of characteristic equation,i.e:[an an-1 an-2 ... a0]= [0 1 2 1.2 1.2 0.3]

----------------------------------------
--------The Routh-Hurwitz array is:--------

m =

        0    2.0000    1.2000
   1.0000    1.2000    0.3000
   2.0000    1.2000       0
   0.6000    0.3000       0
   0.2000       0         0
   0.3000       0         0

    |
```

4.2 Example 2 [11]

$$D(z_1, z_2) = 1 + 0.5z_1 + 0.01z_2 + 0.4z_1z_2$$

Convert two-dimensional characteristics equation into one-dimensional characteristics equation [11]

$$D\left(\frac{1}{z_1}, \frac{1}{z_2}\right) = F(z) = 1 + \frac{0.5}{z_1} + \frac{0.01}{z_2} + \frac{0.4}{z_1z_2} = 0$$

Substitute $z_1 = z_2 = z_3 = s$

$$F(s) = 1 + \frac{0.5}{s} + \frac{0.01}{s} + \frac{0.4}{s^2} = 0$$

$$F(s) = s^2 + 0.51s + 0.4 = 0$$

Put $s = js$ in the equation and also differentiate one time the equation,

$$F'(js) = (js)^2 + 0.51(js) + 0.4 + j[2(js) + 0.51] = 0$$

$$F'(s) = -s^2 + 0.51js + 0.4 - 2s + 0.51j = 0$$

Separate s^2, s and divide by -1 in the above equation,

$$F'(s) = s^2 + (2 - 0.51j)s + (-0.4 - 0.51j) = 0$$

Application of SPC-I:

1	$-0.51J$	-0.4
2	$-0.51J$	0
$-0.255J$	-0.4	0
$2.627J$	0	
-0.4	0	
0		

$$P_1 = (+1, +2), \quad P_2 = (-j0.255, +j2.627), \quad P_3 = (-0.4, 0)$$

The pairs are $P_1 = (+1, +2)$, $P_2 = (-j0.255, +j2.627)$ and $P_3 = (-0.4, 0)$ and P_2 fails to obey the
SPC-I. Hence, the system is aperiodically Unstable.

Application of SPC-II:

Substitute $s = j\omega$ at $F'(s)$ becomes

$$F'(j\omega) = \omega^2 + (2 - 0.51j)\omega + (-0.4 - 0.51j) = 0$$
$$F(j\omega) = -\omega^2 + 2j\omega + 0.51\omega - 0.4 - 0.51j = 0$$

Separate real and imaginary terms

$$F'(j\omega) = \left(-\omega^2 + 0.51\omega - 0.4\right) + j(2\omega - 0.51) = 0$$

−1	0.51	−0.4
0	2	−0.51

Here, in the first row (−) sign occurs so multiply by (−1) on first row and also second row having zero on the first column so interchange first and second rows.

0	2	−0.51
1	−0.51	0.4
2	−0.51	0
−0.255	0.4	0
2.627	0	0
0.4	0	0
0		

$$P_1 = (+1, +2), \quad P_2 = (-0.25, +2.627), \quad P_3 = (+0.4, 0)$$

The pairs are $P_1 = (+1, +2)$, $P_2 = (-0.25, +2.627)$ and $P_3 = (+0.4, 0)$ and P_2 fails to obey the SPC-II. Hence, the system is aperiodically Unstable.

Output Verification Using MATLAB:

SPC-I

```
Input coefficients of characteristic equation,i.e:[an an-1 an-2 ... a0]= [1 2 -0.51i -0.51i -0.4]

----------------------------------------
--------The Routh-Hurwitz array is:--------

m =

    1.0000              0 - 0.5100i  -0.4000
    2.0000              0 - 0.5100i       0
         0 - 0.2550i  -0.4000             0
         0 + 2.6273i        0             0
   -0.4000                  0             0

>>
```

SPC-II

```
Input coefficients of characteristic equation,i.e:[an an-1 an-2 ... a0]= [0 1 2 -0.51 -0.51 0.4]

----------------------------------------
--------The Routh-Hurwitz array is:--------

m =

         0    2.0000   -0.5100
    1.0000   -0.5100    0.4000
    2.0000   -0.5100         0
   -0.2550    0.4000         0
    2.6273         0         0
    0.4000         0         0
```

5 Conclusion

In many engineering applications, two-dimensional linear system has been proved very useful. Two methods for computing aperiodic stability analysis of two-dimensional linear discrete systems with complex coefficients has been proposed using a modified Routh criterion. The proposed procedure has been explained with illustrations and verified by using the MATLAB. From the results presented in this paper, it is hoped that one can get a complete understanding of the aperiodic stability of two-dimensional linear systems.

References

1. Frank, E.: On the zeros of polynomials with complex coefficients. Bull. Am. Math. Soc. 144–158 (1946)
2. Agashe, S.D.: A new general Routh like algorithm to determine the number of RHP roots of a real or complex polynomial. IEEE Trans. Autom. Control 406–409 (1985)
3. Benidt, M., Picinbon, B.: The extended Routh's table in the complex case. IEEE Trans. Autom. Control 253–256 (1991)
4. Chen, S.S., Tsai, J.S.H.: A new tabular form for determining root distribution of a complex polynomial with respect to the imaginary axis. IEEE Trans. Autom. Control 1536–154 (1993)
5. Hashem, A.M.K.: Network synthesis of complex impedance and complex reactance functions. Ph.D. thesis, Concordia University (1993)
6. Sivanandam, S.N., Sreekala, K.: An algebraic approach for stability analysis of linear systems with complex coefficients. Int. J. Comput. Appl. 13–16 (2012)
7. Jury, E.I.: Inners and Stability of Dynamics Systems. Wiley, New York (1974)
8. Fuller, A.T.: Conditions for aperiodicity in linear systems. Br. J. Appl. Phys. 450–451 (1955)
9. Van Valkenburg, M.E.: Modern Network Synthesis, pp. 100–103. Wiley, New York (1960)
10. Bar-Itzhack, I., Calise, A.J.: Counting complex roots in polynomials with real coefficients. Proc. IEEE 2024–2026 (1967)
11. Ramesh, P., Manikandan, V.: Stability analysis of two-dimensional linear time invariant discrete systems. World Appl. Sci. J. **32**(8), 1506–1512 (2014)

Some Results on *K*-Wright Type Hypergeometric Function

Sunil Joshi and Ekta Mittal

Abstract In the present paper, we have developed a new form of Wright hypergeometric function by introducing the *k*-parameter such that $k > 0$, $k \in N$. Our aim is to study the various properties of *k*-Wright type hypergeometric function, which include differentiation and different types of integration, Euler (Beta) transform, and Laplace transform have also been obtained.

Keywords Gauss's hypergeometric function · Generalized hypergeometric function · Wright type hypergeometric function · Euler *k*-Beta transform · Laplace transform

1 Introduction

The Gauss hypergeometric function is defined by

$$_2F_1(a, b; c; z) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{z^n}{n!}; \quad \{|z| < 1, \ c \neq 0, -1, -2, \ldots\}, \quad (1)$$

and the generalized hypergeometric function is defined as

$$_pF_q\big(a_1, \ldots, a_p; b_1, \ldots, b_q; z\big) = \sum_{n=0}^{\infty} \frac{(a_1)_n \ldots (a_p)_n}{(b_1)_n \ldots (b_q)_n} \frac{z^n}{n!};$$

$$\text{where } \{|z| < 1, \ p = q + 1\}, \quad (2)$$

S. Joshi (✉)
Manipal University Jaipur, Jaipur 303007, Rajasthan, India
e-mail: sunil.joshi@jaipur.manipal.edu

E. Mittal
The IIS University, Jaipur 302020, Rajasthan, India
e-mail: ekta.jaipur@gmail.com

where there is no denominator parameter equal to zero or negative integer. E.M. Wright further extended the generalization of the hypergeometric series in the following form:

$$_p\psi_q(z) = \sum_{n=0}^{\infty} \frac{\Gamma(\alpha_1 + \beta_1 n)\ldots\Gamma(\alpha_p + \beta_p n)}{\Gamma(\rho_1 + \mu_1 n)\ldots\Gamma(\rho_q + \mu_q n)} \frac{z^n}{n!},\tag{3}$$

where β_r and μ_t are real positive numbers such that

$$1 + \sum_{t=1}^{q} \mu_t - \sum_{r=1}^{p} \beta_r > 0.$$

When β_r and μ_t are equal to 1, then Eq. (3) differs from the generalized hypergeometric function $_pF_q$ by a constant multiplier only.

The generalized form of the hypergeometric function has been investigated by Dotsenko and Malovichko, and one of the special cases is considered by Dotsenko as

$$_2R_1^{\omega,\mu}(z) =_2 R_1(a, b; c; \omega, \mu; z)$$

$$= \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{n=0}^{\infty} \frac{\Gamma(a + n)\Gamma\left(b + \frac{\omega}{\mu}n\right)}{\Gamma\left(c + \frac{\omega}{\mu}n\right)} \frac{z^n}{n!}\tag{4}$$

And its integral representation can be expressed as

$$_2R_1^{\omega,\mu}(z) = \frac{\Gamma(c)\mu}{\Gamma(c-b)\Gamma(b)} \int_0^{\infty} t^{\mu b-1}(1 - t^{\mu})^{c-b-1}(1 - zt^{\omega})^{-a}dt,$$

where $Re(c) > Re(b) > 0$. This is the analogue of Euler's formula for the Gauss's hypergeometric function by Erdelyi [7]. In 2001, Virchenko et al. [9] defined the Wright type hypergeometric function by taking $\frac{\omega}{\mu} = \tau > 0$ in above Eq. (4), and thus we get

$$_2R_1(a, b; c; \tau; z) = \frac{\Gamma(c)}{\Gamma(b)} \sum_{n=0}^{\infty} \frac{(a)_k\Gamma(b + \tau n)}{\Gamma(c + \tau n)} \frac{z^n}{n!}, \tau > 0, |z| < 1.\tag{5}$$

The Euler Beta Transform is defined as

$$B(\alpha, \beta) = \int_0^1 u^{\alpha-1}(1 - u)^{\beta-1}du = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}, \quad (Re(\alpha) > 0, \ Re(\beta) > 0).\tag{6}$$

The Laplace transform of the function $f(z)$ is defined as

$$L(f(z); s) = \int_0^\infty e^{-sz} f(z) dz.$$ (7)

Here, we involve the *k*-parameter such that in the Wright hypergeometric function by taking $\tau > 0$

$$_2R_1((a, k), (b, k); (c, k); \tau; z) = \frac{\Gamma_k(c)}{\Gamma_k(b)} \sum_{n=0}^\infty \frac{(a)_{n,k} \Gamma_k(b + \tau nk)}{\Gamma_k(c + \tau nk)} \frac{z^n}{n!},$$

where $(|z| < 1, \tau > 0, \ k > 0, k \in N)$. (8)

If we are taking $\tau = 1$, then the *k*-Wright type hypergeometric function is converted into the *k*-Gauss hypergeometric function.

$$_2F_1((a, k), (b, k); (c, k); z) = \sum_{n=0}^\infty \frac{(a)_{n,k}(b)_{n,k}}{(c)_{n,k}} \frac{z^n}{n!}, \quad (|z| < 1, c \neq 0, -1, -2 \ldots).$$ (9)

If we take $k \to 1$, then Eqs. (8) and (9) are converted into the classical Gauss hypergeometric function and Wright type hypergeometric function. Diaz and Pariguan [2] introduced the generalized k-Gamma function as

$$\Gamma_k(z) = \lim_{n \to \infty} \frac{n! \, k^n (nk)^{\frac{z}{k} - 1}}{(z)_{n,k}}, k > 0, z \in C - kZ^-,$$

where $(z)_{n,k}$ is the k-pochhammer symbol and is given by

$$(z)_{n,k} = z(z + k)(z + 2k) \cdots (z + (n - 1)k); z \in C, \ k \in R, \ n \in N^+$$ (10)

When $k \to 1$, then $\Gamma_k(z) \to \Gamma(z)$, which is known as gamma function.
The integral representation of *k*-gamma function is defined as

$$\Gamma_k(z) = \int_0^\infty t^{z-1} e^{\frac{-t^k}{k}} dt; \ \text{Re}(z) > 0, k > 0.$$ (11)

The *k*-Beta function is defined as

$$B_k(x, y) = \frac{1}{k} \int_0^1 t^{\frac{x}{k}-1}(1 - t)^{\frac{y}{k}-1} dt; \ x > 0, y > 0$$

$$B_k(x, y) = \frac{\Gamma_k(x)\Gamma_k(y)}{\Gamma_k(x + y)}; \ \text{as } k \to 1 \text{ then } B_k(x, y) \to B(x, y).$$ (12)

2 Main Result

Theorem 1 *If $a, b, c, \tau \in C$; $Re(a) > 0, Re(b) > 0, Re(c) > 0, \ Re(\tau) > 0, k > 0$*
then

$$c \, _2R_1((a, k), (b, k); (c, k); \tau; z)$$

$$= c \, _2R_1((a, k), (b, k); (c + k, k); \tau; z) + \tau k z \frac{d}{dz} \, _2R_1((a, k), (b, k); (c, k); \tau; z)$$

(13)

Proof We start by taking the second term on the right-hand side of the above Eq. (13) and using Eq. (8), we get

$$= \tau k \left[\frac{\Gamma_k(c + k)}{\Gamma_k(b)} \sum_{n=0}^{\infty} \frac{(a)_{n,k} \Gamma_k(b + nk\tau)}{\Gamma_k(c + k + \tau kn)} \frac{n \, z^n}{n!} \right]$$

$$= \frac{c \Gamma_k(c)}{\Gamma_k(b)} \sum_{n=0}^{\infty} \frac{(a)_{n,k} \Gamma_k(b + nk\tau) \ (c + nk\tau)}{\Gamma_k(c + k + nk\tau)} \frac{z^n}{n!} - \frac{c \Gamma_k(c + k)}{\Gamma_k(b)} \sum_{n=0}^{\infty} \frac{(a)_{n,k} \Gamma_k(b + nk\tau)}{\Gamma_k(c + k + nk\tau)} \frac{z^n}{n!}$$

After simplifications, we will get the desired result.

Theorem 2 *If $a, b, c, \tau \in C$; $Re(a) > 0, Re(b) > 0, Re(c) > 0, Re(\tau) > 0$, and*
$k > 0$, then

$$_2R_1((a, k), (b, k); (c - \tau, k); \tau; z) - \, _2R_1((a, k), (b - k, k); (c - \tau, k); \tau; z)$$

$$= a\tau k z \frac{\Gamma_k(c - \tau)}{\Gamma_k(b)} \sum_{n=0}^{\infty} \frac{(a + k)_{n,k} \Gamma_k(b - k + \tau + nk\tau)}{\Gamma_k(c + nk\tau)} \frac{z^n}{n!}.$$

(14)

Proof Taking left-hand side of (14) and using Eq. (8), we have

$$= \frac{\Gamma_k(c - \tau)}{\Gamma_k(b)} \sum_{n=0}^{\infty} \frac{(a)_{n,k} \Gamma_k(b + nk\tau)}{\Gamma_k(c - \tau + nk\tau)} \frac{z^n}{n!} - \frac{\Gamma_k(c - \tau)}{\Gamma_k(b - k)} \sum_{n=0}^{\infty} \frac{(a)_{n,k} \Gamma_k(b - k + nk\tau)}{\Gamma_k(c - \tau + nk\tau)} \frac{z^n}{n!}$$

$$= \frac{\Gamma_k(c - \tau)}{\Gamma_k(b - k)} \sum_{n=0}^{\infty} \frac{(a)_{n,k} \Gamma_k(b - k + n\tau k)}{\Gamma_k(c - \tau + n\tau k)} \frac{z^n}{n!} \left[\frac{(b + n\tau k - k)}{(b - k)} - 1 \right].$$

(15)

Now, using the k-Pochhammer symbol in Eq. (15), we get

$$= \frac{a\tau k z}{(b - k)} \frac{\Gamma_k(c - \tau)}{\Gamma_k(b - k)} \sum_{n=1}^{\infty} \frac{(a + k)_{n-1,k} \ (b - k + (n - 1)\tau k + \tau k)}{\Gamma_k(c + (n - 1)k\tau)} \frac{z^{n-1}}{(n - 1)!}$$

$$= a\tau k z \frac{\Gamma_k(c - \tau)}{\Gamma_k(b)} \sum_{n=0}^{\infty} \frac{(a + k)_{n,k} \ (b - k + n\tau k + \tau k)}{\Gamma_k(c + nk\tau)} \frac{z^n}{n!}.$$

Theorem 3 *If $a, b, c, \delta, \tau \in C$; $Re(a) > 0, Re(b) > 0, Re(c) > 0, Re(\tau) > 0, Re(\delta) > 0; k > 0$, then*

$$\frac{\Gamma_k(c + \delta)}{\Gamma_k(\delta)} \int_0^1 u^{\frac{c}{k}-1}(1-u)^{\frac{\delta}{k}-1} {}_2R_1((a, k), (b, k); (c, k); \tau; zu^\tau)du$$

$$= \Gamma_k(c) {}_2R_1((a, k), (b, k); (c + \delta, k); \tau; z). \tag{16}$$

Proof We start with left-hand side of Eq. (16) and using (8), we have

$$= \frac{\Gamma_k(c + \delta)}{\Gamma_k(\delta)} \int_0^1 u^{\frac{c}{k}-1}(1-u)^{\frac{\delta}{k}-1}\left(\frac{\Gamma_k(c)}{\Gamma_k(b)}\sum_{n=0}^\infty \frac{(a)_{n,k}\Gamma_k(b + nk\tau)}{\Gamma_k(c + nk\tau)}\frac{(zu^\tau)^n}{n!}\right)du. \tag{17}$$

On changing the order of summation and integration in Eq. (17) and using the *k*-Beta function property, we get

$$= \frac{\Gamma_k(c + \delta)}{\Gamma_k(\delta)}\frac{\Gamma_k(c)}{\Gamma_k(b)}\sum_{n=0}^\infty \frac{(a)_{n,k}\Gamma_k(b + nk\tau)}{\Gamma_k(c + nk\tau)}\frac{(z)^n}{n!}\frac{\Gamma_k(c + \tau nk)\Gamma_k(\delta)}{\Gamma_k(c + \delta + \tau nk)}.$$

After a few steps, we get the desired result.

Theorem 4 *If $a, b, c, \delta, \lambda, \tau \in C$; $Re(a) > 0, Re(b) > 0, Re(c) > 0, Re(\delta) > 0$; $Re(\tau) > 0, Re(\lambda) > 0, k > 0$, then*

$$\frac{\Gamma_k(c + \delta)}{\Gamma_k(\delta)} \int_t^x (x - s)^{\frac{\delta}{k}-1}(s - t)^{\frac{c}{k}-1} {}_2R_1((a, k), (b, k); (c, k); \tau; \lambda(s - \tau)^\tau)ds$$

$$= (x - t)^{\frac{\delta+c}{k}-1}\Gamma_k(c) {}_2R_1((a, k), (b, k); (c + \delta, k); \tau; \lambda(x - t)^\tau). \tag{18}$$

Proof Taking left-hand side of (18), we have

$$\frac{\Gamma_k(c + \delta)}{\Gamma_k(\delta)} \int_t^x \left(\frac{(x - t) - (s - t)}{(x - t)}\right)^{\frac{\delta}{k}-1}(s - t)^{\frac{c}{k}-1} {}_2R_1((a, k), (b, k); (c, k); \tau; \lambda(s - \tau)^\tau)ds.$$

Applying the transformation formula by using $u = \frac{s-t}{x-t} \Rightarrow (x - t)du = ds$, and then further using Eq. (8), we get

$$= \frac{\Gamma_k(c)}{\Gamma_k(b)}\frac{\Gamma_k(c + \delta)}{\Gamma_k(\delta)} \int_0^1 (1 - u)^{\delta/k-1}u^{c/k-1}(x - t)^{c/k}\left(\sum_{n=0}^\infty \frac{(a)_{n,k}\Gamma_k(b + \tau nk)}{\Gamma_k(c + \tau nk)}\frac{\lambda^n u^{\tau n}(x - t)^{\tau n}}{n!}\right)du. \tag{19}$$

Finally, changing the order of integration and summation in (19) and applying k-Beta function, we get the right-hand side of the above equation.

Theorem 5 *If $a, b, c, \omega, \tau \in C; Re(a) > 0, Re(b) > 0, Re(c) > 0, Re(\omega) > 0;$
$Re(\tau) > 0, k > 0$, then*

$$\int_0^z t^{\frac{c}{k}-1} {}_2R_1\big((a, k), (b, k); (c, k); \tau; \omega t^\tau\big)dt = \frac{kz^{\frac{c}{k}}}{c} {}_2R_1\big((a, k), (b, k); (c + k, k); \tau; \omega z^\tau\big). \quad (20)$$

Proof Consider left-hand side of (20), and using (8), we have

$$\int_0^z t^{\frac{c}{k}-1} \frac{\Gamma_k(c)}{\Gamma_k(b)} \sum_{n=0}^\infty \frac{(a)_{n,k} \Gamma_k(b + \tau nk)}{\Gamma_k(c + \tau nk)} \frac{(\omega t^\tau)^n}{n!} dt.$$

On changing the order and solving the integration, we obtain

$$\frac{kz^{\frac{c}{k}}}{c} \frac{\Gamma_k(c + k)}{\Gamma_k(b)} \sum_{n=0}^\infty \frac{(a)_{n,k} \Gamma_k(b + \tau nk)}{\Gamma_k(c + k + \tau nk)} \frac{(\omega z^\tau)^n}{n!},$$

which is a required result.

Integral transform of $_2R_1((a,k), (b,k); (c,k);\tau;z)$

Theorem 6 *(Euler k-Beta Transform) If $a, b, c, \alpha, \beta, \tau \in C; Re(a) > 0, Re(b) >$
$0, Re(c) > 0, Re(\tau) > 0; k > 0$, then*

$$\frac{1}{k} \int_0^1 x^{\frac{\alpha}{k}-1}(1 - x)_2^{\frac{\beta}{k}-1} R_1((a, k), (b, k); (c, k); \tau; zx^\sigma)dz$$

$$= \frac{\Gamma_k(c)\Gamma_k(\beta)\Gamma_k(\alpha)}{\Gamma_k(b)\Gamma_k(\alpha + \beta)} \, {}_3\psi_2 \begin{pmatrix} (a, k), (b, \tau k), (\alpha, \sigma k) \\ \qquad\qquad\qquad\qquad ; z \\ (c, \tau k), (\alpha + \beta, \sigma k) \end{pmatrix}. \quad (21)$$

Proof Consider left-hand side of (21) and using Eq. (8), we have

$$\frac{1}{k} \int_0^1 x^{\frac{\alpha}{k}-1}(1 - x)^{\frac{\beta}{k}-1} \frac{\Gamma_k(c)}{\Gamma_k(b)} \sum_{n=0}^\infty \frac{(a)_{n,k} \Gamma_k(b + \tau nk)}{\Gamma_k(c + \tau nk)} \frac{(zx^\sigma)^n}{n!} dx.$$

After simplifying and using k-beta property, we get

$$= \frac{\Gamma_k(c)}{\Gamma_k(b)} \sum_{n=0}^\infty \frac{(a)_{n,k} \Gamma_k(b + \tau nk)}{\Gamma_k(c + \tau nk)} \frac{\Gamma_k(\alpha + \sigma nk)\Gamma_k(\beta)}{\Gamma_k(\alpha + \beta + \sigma nk)} \frac{(z)^n}{n!}$$

$$= \frac{\Gamma_k(c)\Gamma_k(\beta)\Gamma_k(\alpha)}{\Gamma_k(b)\Gamma_k(\alpha + \beta)} \, {}_3\psi_2 \begin{pmatrix} (a, k), (b, \tau k), (\alpha, \sigma k) \\ \qquad\qquad\qquad\qquad ; z \\ (c, \tau k), (\alpha + \beta, \sigma k) \end{pmatrix}.$$

Theorem 7 *(Laplace transform) If* $a, b, c, \alpha, \beta, \tau \in C;\ Re(a) > 0, Re(b) > 0, Re(c) > 0,\ Re(\tau) > 0; k > 0,$ *then*

$$
\int_0^\infty e^{-sx} (x)^{\frac{\alpha}{k}-1} \, {}_2R_1((a, k), (b, k); (c, k); \tau; zx^\sigma) dx
$$

$$
= \frac{\Gamma_k(c)\Gamma_k(\alpha)}{\Gamma_k(b)} \frac{k}{(ks)^{\frac{\alpha}{k}}} \, {}_3\psi_1 \left(\begin{array}{c} (a, k), (b, \tau k), (\alpha, \sigma k) \\ \hspace{2cm} ; \frac{z}{(ks)^\sigma} \\ (c, \tau k) \end{array} \right). \tag{22}
$$

Proof Let us start with right-hand side of Eq. (22), and we have

$$
\int_0^\infty e^{-sx}(x)^{\frac{\alpha}{k}-1} \frac{\Gamma_k(c)}{\Gamma_k(b)} \sum_{n=0}^\infty \frac{(a)_{n,k}\Gamma_k(b+\tau nk)}{\Gamma_k(c+\tau nk)} \frac{(zx^\sigma)^n}{n!} dx
$$

$$
= \frac{\Gamma_k(c)}{\Gamma_k(b)} \sum_{n=0}^\infty \frac{(a)_{n,k}\Gamma_k(b+\tau nk)}{\Gamma_k(c+\tau nk)} \int_0^\infty e^{-sx}(x)^{\frac{\alpha}{k}+\sigma n-1} dx.
$$

Putting $xs = \frac{u^k}{k}$ in the above equation, it will get converted into

$$
= \frac{\Gamma_k(c)}{\Gamma_k(b)} \sum_{n=0}^\infty \frac{(a)_{n,k}\Gamma_k(b+\tau nk)}{\Gamma_k(c+\tau nk)} \frac{k}{(ks)^{\frac{\alpha}{k}+\sigma n}} \int_0^\infty e^{-\frac{u^k}{k}}(u)^{\alpha+\sigma nk-1} dx.
$$

Now, using Eq. (10), we get

$$
= \frac{\Gamma_k(c)}{\Gamma_k(b)} \sum_{n=0}^\infty \frac{(a)_{n,k}\Gamma_k(b+\tau nk)}{\Gamma_k(c+\tau nk)} \frac{k\,\Gamma_k(\alpha+\sigma nk)}{(ks)^{\frac{\alpha}{k}+\sigma n}}.
$$

After further simplifications, we will get the desired result.

Theorem 8 *If* $a, b, c, \alpha, \beta, \tau \in C; Re(a) > 0, Re(b) > 0, Re(c) > 0,\ Re(\tau) > 0; k > 0,$ *then*

$$
k^m \frac{d^m}{dx^m}\left[x^{\frac{c}{k}-1} \, {}_2R_1((a, k), (b, k); (c, k); \tau; \omega x^\tau) \right]
$$

$$
= x^{\frac{(c-mk)}{k}-1} \frac{\Gamma_k(c)}{\Gamma_k(c-mk)} \, {}_2R_1((a, k), (b, k); (c, k); \tau; \omega x^\tau). \tag{23}
$$

Proof Starts with R.H.S. of (23), and we have

$$
k^m \frac{\Gamma_k(c)}{\Gamma_k(b)} \sum_{n=0}^\infty \frac{(a)_{n,k}\,\Gamma_k(b+\tau kn)}{\Gamma_k(c+\tau kn)} \frac{\omega^n}{n!} \frac{d^m}{dx^m}\left(x^{\frac{c+\tau nk}{k}-1} \right)
$$

$$= k^m \frac{\Gamma_k(c)}{\Gamma_k(b)} \sum_{n=0}^{\infty} \frac{(a)_{n,k}\,\Gamma_k(b+\tau kn)}{\Gamma_k(c+\tau kn)} \frac{\omega^n \left(\frac{c+\tau nk}{k}-1\right)\left(\frac{c+\tau nk}{k}-2\right)\cdots\left(\frac{c+\tau nk}{k}-m\right)}{n!} \left(x^{\frac{c+\tau nk}{k}-m-1}\right).$$

After solving, we get

$$= \frac{\Gamma_k(c)}{\Gamma_k(b)} \sum_{n=0}^{\infty} \frac{(a)_{n,k}\,\Gamma_k(b+\tau kn)}{\Gamma_k(c+\tau kn-mk)} \frac{\omega^n}{n!} \left(x^{\frac{c+\tau nk}{k}-m-1}\right)$$

$$= x^{\frac{(c-mk)}{k}-1} \frac{\Gamma_k(c)}{\Gamma_k(c-mk)} \left(\frac{\Gamma_k(c-mk)}{\Gamma_k(b)} \sum_{n=0}^{\infty} \frac{(a)_{n,k}\,\Gamma_k(b+\tau kn)}{\Gamma_k(c+\tau kn-mk)} \frac{(\omega x^\tau)^n}{n!} \right)$$

$$= x^{\frac{(c-mk)}{k}-1} \frac{\Gamma_k(c)}{\Gamma_k(c-mk)} \, {}_2R_1((a,k),(b,k);(c,k);\tau;\omega x^\tau).$$

References

1. Bateman, H., Erdélyi, A.: Higher Transcendental Functions, vol. 2, p. 149. McGraw-Hill, New York (1953)
2. Diaz, R., Pariguan, E.: On hypergeometric functions and pochhammer k-Symbol. Divulg. Mat. **15**(2), 179–192 (2007)
3. Dotsenko, M.R.: On some applications of Wrights hypergeometric function. Dokl. bolg. akad. naukite **44**(6), 13–16 (1991)
4. Kokologiannaki, C.G.: Properties and inequalities of generalized k-gamma, beta and zeta functions. Int. J. Contemp. Math. Sci. **5**(14), 653–660 (2010)
5. Krasniqi, V.: A limit for the k-gamma and k-beta function. In: International Mathematical Forum, vol. 5, no. 33, pp. 1613–1617 (2010)
6. Mubeen, S., Habibullah, G.M.: An integral representation of some k-hypergeometric function. Int. Math. Forums **7**(4), 203–207 (2012)
7. Rainville, E.D.: Special Functions. The Macmillan Company (1970)
8. Rao, S.B., Prajapati, J.C., Shukla, A.K.: Wright type hypergeometric function and its properties. Adv. Pure Math. **3**(3), 335–342 (2013)
9. Virchenko, N., Kalla, S.L., Al-Zamel, A.: Some results on a generalized hypergeometric function. Integral Trans. Spec. Funct. **12**(1), 89–100 (2001)
10. Virchenko, N.: On some generalizations of the functions of hypergeometric type. Fract. Calc. Appl. Anal. **2**(3), 233–244 (1999)

Complete Dielectric Resonator Model of Human Brain from MRI Data: A Journey from Connectome Neural Branching to Single Protein

Pushpendra Singh, Kanad Ray, D. Fujita and Anirban Bandyopadhyay

Abstract Using freely available MRI data of structural components mapping of human brain from different universities (primarily Rajat Jain, 25; 38-years-old lady from UK), we have built actual structural database of the human brain components, e.g., neural network connectome data, blood vessel map, ventricles, cavities for cerebral-spinal fluid, hippocampus regions of midbrain, etc. In previous studies, we have argued that every single element in the brain behaves as dielectric resonator. Here, we run rigorous dielectric resonance simulation to verify the hypothesis that the scale-free resonance does exist in the material architecture of the brain. From MRI-derived structures, we simulate the resonance frequencies, distribution of electric, and magnetic field of the brain components in CST and detect the phase response behavior, specially phase transition and symmetry breaking as a function of resonance frequency. We find that electric and magnetic fields distribute inhomogeneously in the dielectric structure, not just in the neural branches but also in the blood vessels and proteins like axon and microtubule bundles. The resonance frequencies show a characteristic topological pattern, specially, every single brain component is splitting electromagnetic field in such a way that at certain frequencies magnetic field dominates and at certain resonance frequency, electric field dominates. This distinct behavior of splitting fields at all spatial and time scale was never reported before. We speculate that there may exist a unified geometric pattern hidden in the vibra-

P. Singh · D. Fujita · A. Bandyopadhyay (✉)
Advanced Key Technologies Division, National Institute for Materials Science,
1-2-1 Sengen, Tsukuba, Ibaraki 3050047, Japan
e-mail: anirban.bandyo@gmail.com

P. Singh
e-mail: singhpushpendra548@gmail.com

P. Singh
Amity University Rajasthan Kant Kalwar, NH-11C, Jaipur Delhi Highway, Jaipur 303007, Rajasthan, India

K. Ray
Amity School of Applied Sciences, Amity University, Jaipur, Kant Kalwar, NH-11C, Jaipur 303007, Rajasthan, India

tional frequencies of the brain components, which hold important information for the brain's information processing.

1 Introduction

Human brain consumes 20% of energy of a human body, most of its ~30,000 proteins expression is observed in the brain. Since 1907, it is believed that neural membrane processes information via electric spike [1], This is known as the integrate-and-fire model. Later in 1952 it was shown by using a metallic electrode that the membrane does fire [2]. Ion channel responsible for this firing was demonstrated much later using patch clamp [3]. It is inherently considered that the neural circuitry of the brain does the information processing. The idea is inspired by the conventional integrated chip. This is the foundation of connectome project that intricately maps the structure of neural network [4]. Though molecular neuroscience is still the mainstream, which fundamentally rejects physical investigations of the brain as "unnecessary", there are challenges [5]. It is being argued even in the mainstream forums that there is something beyond [6]. Following the challenges, we have carried out intensive experiments earlier to experimentally establish that it is not the neural membrane but the microtubule or axon core is responsible for the neural firing. The ionic nerve impulse that we see is an effect of a large number of engineering processes happening inside [7].

However, researchers have been studying the topological roles in the information processing of the human brain. A direct correlation between cortical folding pattern and the size and shape of the lateral ventricle, spherical harmonics for lateral ventricle development [8] are being studied. The shape of the lateral ventricle can be approximated with a prolate spheroid. The poles of the anterior and inferior horns map to the north and south poles of the prolate spheroid, respectively [8]. Conformal geometry is a mathematical concept that suggests the preservation of a geometric shape in a scale-free manner. Conceptually, if there is a change in symmetry, still a geometric shape is preserved. Researchers modeling a human brain have argued that such conformal processing is fundamentally embedded in the brain structure [9]. Folding in the brain has been a topic of concern for several researchers [8] and for our dielectric resonator model of the human brain, folding is a key parameter that shapes the cavity [10].

Scale-free behavior in the brain has been observed by several researchers [11]. It has been argued that morphogenesis is fundamental to the compact wiring of the neural network, which is again a derivative of the topological symmetry breaking argued by other researchers [12]. Locking phase in the structural transformation and information processing appears to originate from a singular mechanism [13]. Resonance across the brain in a characteristic manner is thought to be the foundation

of information processing [14]. The spatiotemporal pattern of MEG pattern in the brain suggests the existence of the resonance behavior [15]. Fractal model of the brain is not new, but it has been argued that even a single ion channel responds in a fractal manner [16].

Recently the concept of resonance chain is proposed by combining the resonance band of every single component of the brain, the chain is a unique triplet of triplet ($3 \times 3 = 9$) groups of resonance frequencies generated by each of the 12 important components [17]. Scale-free topology integrating the entire brain is based on the finding that it is not the synapses but even a single microtubule residing inside a neuron could hold information and memorize it [18]. Electromagnetic resonance, that is, the foundation of information processing in proteins and their complexes, regulates the biological processes like dynamic instability [19]. We have argued earlier that the resonance behavior that originates in the molecular scale is fundamental to the brain.

Here, we have studied different components of the brain such as neuron, axon, connectome, cerebral hemisphere, blood vessel, microtubule, ventricle, midbrain, etc. Here, we have simulated several brain components based on MRI structural data file and we have observed interesting results related to the electric and magnetic field distribution through the brain components. Due to the space constraints, we have focused on the field distributions of the brain component that are discussed in the following sections.

2 Brain Components

2.1 Study Methodology

We have built the structure theoretically from the MRI or crystal data, and then edited or modified the structure reducing the noise that is generated during importing to CST. Finally, we have calculated the electromagnetic resonance band by searching for sharp peaks from kHz (10^3 Hz) to pHz (10^{15} Hz), and connecting the input energy supply port at various locations on the brain component under study. Following a thorough search for the resonance peaks, we have identified those peaks which provided the most significant phase-modulated electric and magnetic field distribution or a clocking of electromagnetic resonance frequencies in the ~THz frequency domain. The reason is that our background research on neuron and microtubules has shown that proteins transmit signal using thermal waves around THz domain. This article accounts for a brain, if that harvests thermal noise from noise. Here for each brain component, we are analyzing only the clocking phase phenomenon discovered homogeneously in every single brain component that we have studied. Magnetic field is depicted as M and an electric field is depicted as E.

2.2 Neurons

Human brain disorder could be understood by the structure and function of the neurons [20]. Here, to understand the dielectric property of a neuron structure, we have artificially created the entire neuron 3D topology using bits and pieces in the CST software. We have taken ideal geometric components only to reduce the structural defect and maximize the role of topology in the results we have obtained. The neuron structure is made empty replicating the membrane, and our model is in synchrony with the existing concept of neuroscience where a neuron membrane is considered as the fundamental location for processing information. Neuron's dendrite, cell body (soma) and axon, and all parts were created using such hollow geometric shapes. Figure 1a (top left) shows an ionic signal propagation pathway that is believed for nearly a century [1]. Here, we have used another very interesting component, we call it "axon". This is inspired by STORM analysis of neural core [21] and cryo-TEM studies of neurons by Hirokawa in the 1990s in Japan, which showed the internal structure of neurons. For us, we made axon in CST using basic geometric components, and replicated microtubule, actin, beta spectrin, and ankyrin G. Four molecules form a crystalline-ordered nanoporous structure inside an axon.

We have studied four combinations of neurons as shown in Fig. 1. Out of four panels, the panels a and d include the axon inside, while panels b and c do not have it. Comparison between ad and bc pairs provide an answer to the debate, which route is better. There are two choices, either the internal axon architecture is useful for the energy transmission or the existing faith on the membrane doctrine, that membrane alone is sufficient. Our electric and magnetic field distribution at the resonance frequency suggests that if the axon core is absent from the resonant vibration simulation, then the electric field distribution is drastically reduced in the axon region. It remains confined in the soma or in the dendritic region.

The second interesting finding explained by Fig. 1 comes from the location of the port or the location of energy supply unit. The port is in the axonal branch direction for the panels a and c while in the panels b and d, the port is in the dendritic branches. Thus, we find that contrary to what has been taught to us in the textbook, energy transmission from the axonal branches to the dendritic branches via soma enriches the dendritic branches more. It means if energy transmits from dendritic branches toward the axonal branches, then it broadly distributes with a very low energy density localization in the dendritic branches. This observation is just opposite to the direction of nerve impulse. During neuron firing, energy gets stored in the soma, and then it bursts from soma to the axonal branches via axon. Here, the reverse route is favored if we consider dielectric resonance to be one of the multiple energy transmission channels operating simultaneously with the neural membrane transmission.

Then, we have tried to excite soma by the following various means. In Fig. 1, in all panels, one could notice that soma does not absorb much electrical or magnetic energy and its dendritic branches are extremely sensitive in this context. After putting the artificial coaxial probe (designed in CST software) around the cell, a small amount of energy appears around the cell body but there is no energy transmitted through the
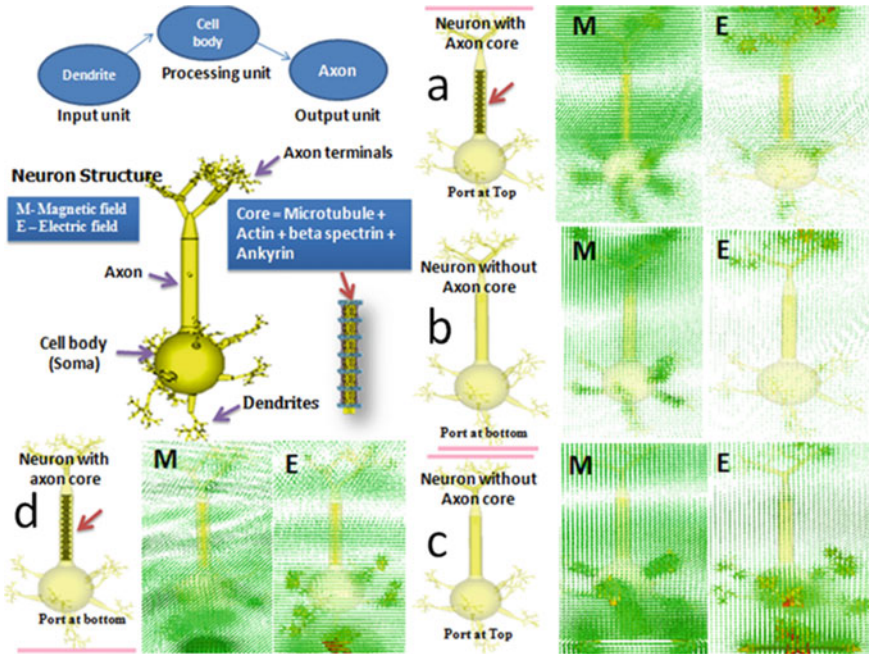
Fig. 1 Signal processing along the neuron [21], **a** and **c** Energy source applied on axon terminal of neuron with axon core and without axon core. **b** and **d** Energy source applied on dendrite branches of neuron without axon core and with axon core. **Panel c**: Port at axon terminal (port above, no axon); Waveguide port dimensional—300 × 300 nm; Simulation frequency range = 200–2000 THz; Boundaries condition = open space; Used solver = Time domain solver; Magnetic and electric field distribution scale at resonant frequency = 451.25 THz. **Panel b**: Port at dendritic terminal, port below, no axon; Port at bottom (dendrites); Waveguide port dimensional—300 × 300 nm; Simulation frequency range = 200–2000 THz; Boundaries condition = open space; **Used solver** = Time domain solver; Magnetic and electric field distribution scale at resonant frequency = 432.3 THz. **Panel a**: Port at axon terminal (top) with axon in the neuron core; Waveguide port dimensional—300 × 300 nm; Simulation frequency range = 200–2000 THz; Boundaries condition = open space

axon shown in Fig. 1. Figure 1d shows one interesting formation of ripples all along the axon, the reason is the conflict between the axon core and the lower dendritic port as both want to dominate and the result is a ripple. One should also note that irrespective of the port location or the existence of axon core or not, the magnetic energy distributes much more predominantly in the neuron surface. The topology of neuron membrane alone prefers a magnetic energy-dominated dielectric surface.

2.3 Connectome

A comprehensive structural map of brain's neural network is known as connectome [22, 23]. It is a database of all neuron structures or nerve bundles ~50 μm and below. The previous study has argued that connectome compresses the raw data to extract the maximal information [24]. Modern graph theory is a quantitative tool to define the characteristic of the network structure by using local contribution of individual nodes (nodes correspond to the neuronal elements) and edges (edges correspond to the function associated between the neuronal elements) through the network. A variety of the network's geometry provides the structural information and functional brain connectivity [25].

From MRI-derived structures, we simulated the connectome model of Rajat Jain: (25-years-old; acknowledgement UCSD) by editing the noisy structure and making it suitable to run in electromagnetic resonance software CST. Figure 2 shows electric and magnetic field simulations of the entire brain's neural wiring for four resonance frequencies, each row depicting a panel represents a particular frequency. We find that the electric field is delocalized at different functional domains in the brain while the magnetic field is concentrated in the midbrain region. This is a remarkable finding because researchers have always wondered how midbrain is regulating the rest of the brain, while it is processing the short-term response. Here, we find that while the electric field is neutral, it does not show any characteristic feature; the magnetic field and surface current concentrate in the midbrain around the hippocampus region and other functional modules of the midbrain.

For all four frequencies, the electric field confines itself into the fornix region and then branches out to the cortex region following specific energy transmission paths. We have tried to identify from the connectome data how different pathways are selected as a function of frequency. Here, due to the lack of space, we cannot detail all transmission routes but, in summary, by proper choice of resonance frequency, the connectome architecture by itself selects a particular neural pathway for signal transmission.

2.4 White Matter of Cerebral Hemisphere

The left hemisphere plays an important role in speech recognition, processing language, logical and analytical thinking, ordered sequential tasks, planning etc. while the right hemisphere primarily processes random thoughts, intuition and emotional thoughts, impulsive and creative behavior, imagination, creative writing and art etc. [26]. In Fig. 2, we have described the white matter of cerebral hemisphere of Rajat Jain (25) at four different resonance frequencies. It is obvious once again that the magnetic field dominates in the core brain structure with a faint localized electric field. The port is applied below the brain structure around brain stem mimicking the event of sensory input entering into the brain via spinal cord. Figure 3a shows
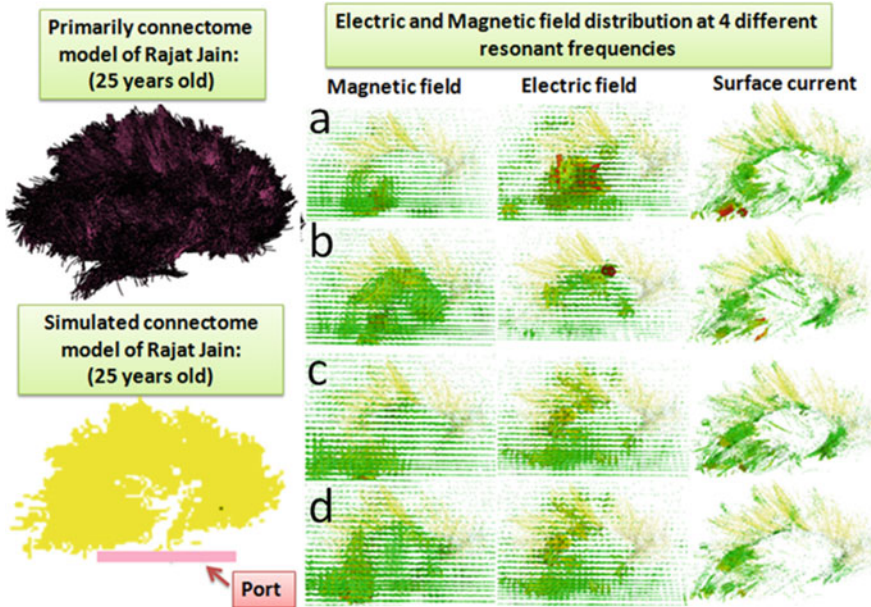
Fig. 2 Magnetic field, electric field, and surface current distribution on human connectome model which shows the characteristic pattern of fields distribution and surface current parameter at resonance frequencies. Simulation frequency range = 1–2500 THz; Waveguide port dimensional = 1.0e+004 × 1.0e+004; Boundaries condition = open space; **Used solver** = Time domain solver; Magnetic field, electric field, and surface current distribution scale at resonant frequency = 1325.2 THz (**panel a**, first row); 1719.2 THz (**panel b**, second row); 1972.2 THz (**panel c**, third row); 2250.3 THz (**panel d**, fourth row). At first resonance frequency, E-field is more dominating compared to M-field while in second resonance frequency, M-field is more dominating around the cavity of the connectome. At third resonance frequency, major contribution of E-field is around the boundary of that cavity, and M-field is randomly distributed around the model. The current distribution profile looks like a shape of hollow ellipsoid

that the magnetic fields select the surface around corpus callosum, specifically in the cingulate gyrus that covers the corpus callosum and cerebellum, the response coordination center of the brain but keeps the entire cortex untouched. No significant electric or magnetic energy flows or localizes in the cortex region for any resonance frequency. Even the surface current density locates around the cingulate gyrus, which is the base connection center, energetically, for the energy transmission all around the cortex as we see in the simulation of Fig. 3.
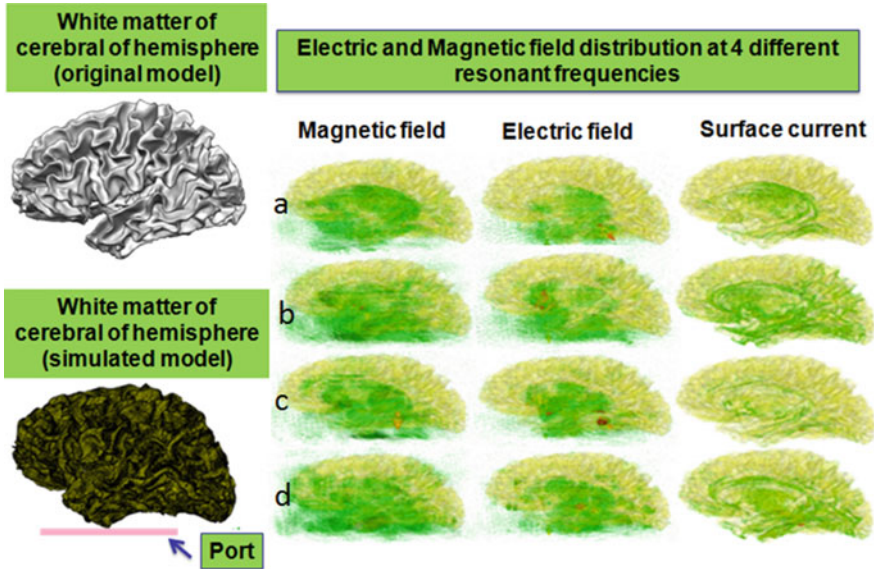
Fig. 3 White matter of cerebral of hemispheres: Magnetic field is more dominating compared to the electric field at the center of the hemisphere. Both fields exist at the cavity. Current distribution appears toward the inner boundary of hemisphere. Simulation frequency range = 1–5 THz, Waveguide port dimensional = 16e+004 × 2.6e+005 nm; Boundaries condition = open space; Used solver = Time domain solver. **Panel a**: 2.01 THz, **Panel b**: 2.25 THz, **Panel c**: 2.70 THz, and **Panel d**: 3.08 THz

2.5 Blood Vessel

Normally, in the brain models, the role of blood vessels is ignored just like glial cells are ignored, and the supremacy of neuron membrane continues. During the theoretical structural simulation of the human brain, we have made sure to integrate all electrical mechanical and magnetic field contributors in the decision making of the human brain [17]. To us, the blood vessel topology is the center of thermal noise source, which the brain elements use. We are concentrating on the THz band resonance of all brain components. In that context, the most important energy source topology is governed by the blood vessels. The walls of the blood vessel system are elastic that are also capable of increasing the diameter as the pressure rises [29]. The previous study suggests that a standing wave can be derived from average fluid flow through cross-sectional area of the blood vessel. The study related to the effect of magnetic field exposure on blood vessel in microvasoconstrictor is not clear. There are mechanical contractions and expansions of blood vessels due to heart beat that effects the magnetic field fluctuation in a geometric pattern in the brain. Half of the study cites the vasodilatory effect of magnetic field and the other half-cited magnetic field that could trigger either vasodilation or vasoconstriction [30]. Due to lack of investigation related to the effects of MFs on microcirculation and microvasculature,
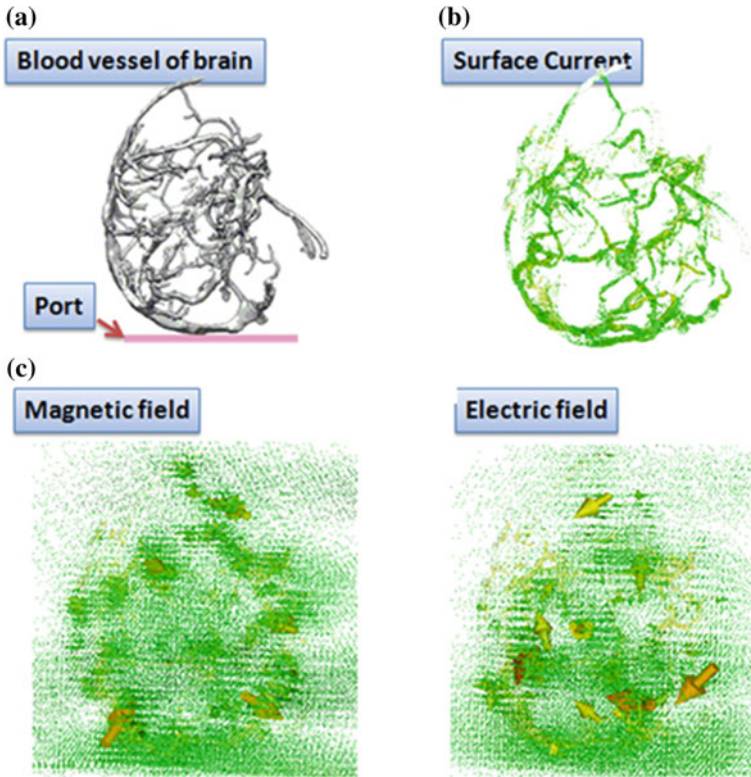
Fig. 4 a Original model of the blood vessel created by the structural data file. **b** Simulated model created by electromagnetic resonance software: Electric field and magnetic field profiles, magnetic field is more dominating compared to electric field, and both appear at a particular part of the veins. **c** Current profile of the blood vessel. Simulation frequency range $= 1$–2 THz; Boundary condition $=$ open space; **Used solver** $=$ time domain solver; Electric field, magnetic field, and surface current distribution scale at resonant frequency $= 1.42$ THz

here, we have analyzed how the topology of biomaterial (blood vessel) plays the role to perform their function. We theoretically studied the electric and magnetic field distributions on and around the blood vessel network of entire brain. At first, we collected structural data file and edited it to make it suitable to run in electromagnetic resonance software, CST, by keeping all conditions same. The distribution of both the fields is shown in Fig. 4.

By varying resonance frequencies, we have found that there is a virtual 3D cavity of energy distribution created by the blood vessels. There are, basically, two 3D cage-like structures formed, one by the electric field and another by the magnetic field. It is like two distinct 3D cages made up of wires that are somehow put together. The most interesting aspect of it is that both the cages generate energy barriers, in between them thin energy transmission pathways are built, it enables the system to manage energy locally at a very high resolution. Our finding suggests that our initial

speculation regarding the role of THz signals in governing information processing of biomaterials is apparently correct. We tried to go beyond neuron membrane as the sole contributor of the information processing in the human brain. Here, we are presenting only a few of the radical routes like blood vessel, axon core, microtubule, proteins, topology of midbrain, and ventricles. In future, we will include cortical columns neuroglia complex or supramolecular system, which directly communicates with the blood vessels to absorb energy.

2.6 Axon

Axon is an important wire like part of a neuron, which plays an important role in cellular communication [27, 28]. Normally, the concept of general readers is that everything inside a neuron is filled with cytoskeleton fluid, it is jelly like. This is not true for a neuron. In the 1990s, several studies were made by taking a cross section of a frozen neuron, which showed that the neuron core is primarily a solid structure made of microtubules and other neurofilaments, and they remain in a gap of ~50 nm. Still, textbooks on neurophysiology show internal structure of a neuron as empty. Recently, Xu et al. [21] have shown that not just the axon's initial segment but entire neuron's internal structure has ordered crystalline structure made of beta spectrin and actin, and ankyrin. These results prompted us to create artificial axon based on all these structural results in CST. The internal structure of axon is made of parallel microtubules in the core keeping a gap of ~50 nm as per the Cryo-TEM data, and then the lattice obtained by Xu et al. [21] is replicated around the core microtubule structure. The final complex architecture is studied by finding the electromagnetic resonance frequencies and then determining the distribution of electric and magnetic field on its surface.

Figure 5 shows the length variation data of an axon, we wanted to confirm whether that when a brain learns, neurons make new connections, break the old ones, and fundamental to all these processes is the change in the length of the core structure residing inside the neural branches. We find that as the length increases from 200 nm to a few micrometers, there is a wave-like interference of electric field and magnetic field, and a particular length shows prominent interference effects. The electromagnetic simulation of an axon show the clocking behavior of both electric and magnetic fields. It leads to the concept of node and antinodes. This is a brilliant exhibition of periodic interactive and silent modes. This is reflected in the directivity map in Fig. 5, where we can see that at certain lengths, the axon stops interacting and goes to a silent mode. Here, we are depicting a brief part of our extensive study with only one port that is pumping electromagnetic signal along its length.

Another interesting part of this study is that the electric field is dominant at a certain length and at that length the magnetic field is absent. There is a complementary interference between the electric and the magnetic. Now, we are investigating information about the role of axon in signal transmission and their topological role to perform own function.
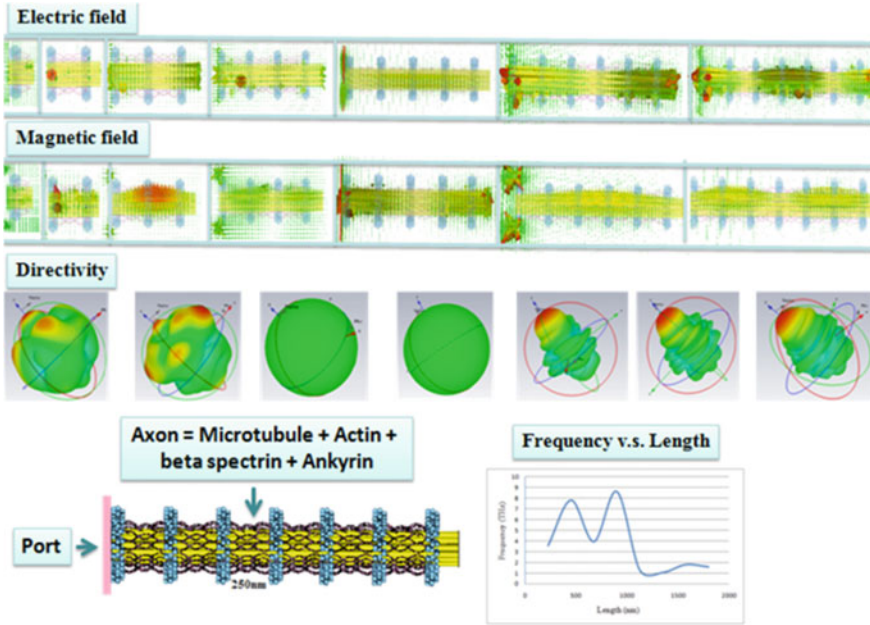
Fig. 5 At initial length of the axon, both electric field and magnetic field are dominant, and after increasing the length, magnetic field is dominated whereas electric fields are absent. Again increasing the length, electric field appears again and magnetic field is absent, similarly, it is repeated again and again: such type of phenomena lead to the concept of nodes and antinodes formation of electric field and magnetic field. Directivity is more at the initial length and after successive increase in length, it decreases, and after some particular length it comes again. If we plot a graph between frequency versus lengths, it almost shows the nature of the sine wave. Common condition for all axons (rings); Port dimensional $= 35 \times 34$ µm; Simulation frequency range $= 0$–10 THz; Boundaries condition $=$ open space; **Used solver** $=$ Time domain. In the row resonance frequencies are 3.5, 7.8 THz; 3.95, 8.6, 1.3 THz; and 1.07, 1.81 THz

2.7 Microtubule

Tubulin molecules in cells assemble and disassemble themselves, which generate a structure known as "Microtubule". Microtubule is a key part of axon described above. We have demonstrated that the microtubule stores and processes information just like a neuron [18, 19, 31]. Also, it has been demonstrated experimentally that even inside a living neuron cell, microtubule vibration survives [7, 32]. Using coaxial atom probes, it is also shown that the axon core communicates 250 µs faster than the onset of a neural firing. Therefore, there is another world of information processing via microtubule in the neurons. Microtubules are abundant in every single eukaryotic cell. Here, we create a microtubule structure in CST and run simulation to find the resonance frequencies. At the resonance frequencies, we image electric and magnetic field distributions at two different resonance frequencies (Fig. 6, panel a and panel b
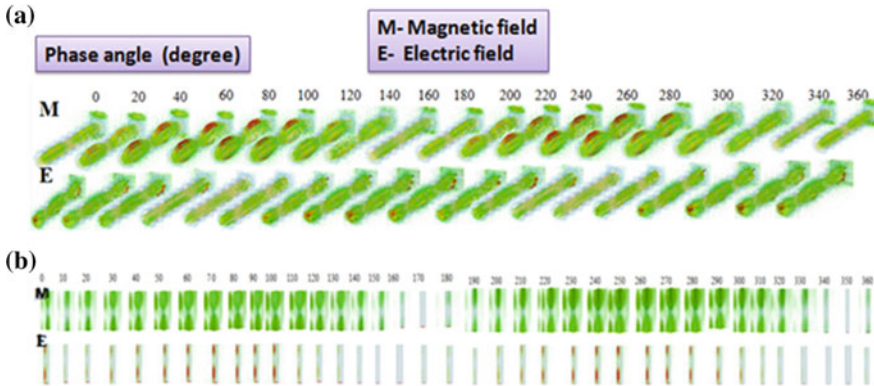
Fig. 6 **a** Clocking behavior of the electric and magnetic field of a microtubule: at the initial phase, magnetic field starts to rise from low value to high value; at 40°–100° magnetic field is maximum. After that, it goes down to the minimum lower peak, and it again increases to 220°–240° and goes down to decrease at the phase duration of 320°–360°. Simulation frequency range = 1–400 MHz; Waveguide port dimensional = 13 × 13 mm; Shape = helical form (diameter-27); Boundaries condition = open space; Used solver = Time domain solver; Magnetic and electric field distribution profile at resonant frequency = 199.25 THz. **b** Split of electric field and magnetic field: From figure (**b**), both fields are the function of phase, magnetic field is more dominating compared to electric field, which is contradictory of the fundamental concept. Simulation frequency range = 1–520 MHz; Waveguide port dimensional = 6 × 6 mm; Lattice arrangement = straight; Boundaries condition = open space; **Used solver** = Time domain solver

are generated at two different resonance frequencies). Axon-like clocking oscillations of the electric and magnetic fields are observed.

One important difference between the resonance behavior of microtubule and axon is that in axon, the topological changes are not that dynamic. Whereas the microtubule lattice changes its configuration so much that the mixture of lattice could modulate the resonance frequencies over a wide frequency range. We are not presenting the data here, but we have observed that mixture of lattice could tune the resonance frequencies.

We have obtained some more interesting results after doing parametric analysis. For example, length, diameter, pitch, lattice variation, and almost all the topological parameters deliver unique periodic effects. Clocking is fundamental to the resonance property. When phase makes a complete rotation, we find that the electric and the magnetic field distributes in a very particular topological way across the hardware. We have repeated similar study like axon and microtubule even in the single protein tubulin. We are not presenting it here because the plots appear similar. In future, we would report in a more detailed description, how port location, material property, and small changes in the symmetry of the structure changes the resonance frequency domain fundamentally. Note that we have already published that the microtubule primarily vibrates in the kHz, MHz, and in the GHz domain. Here, we are concentrating only in the THz domain. Later, we would show that a small change in symmetry could bring resonance from 100s of THz to a few MHz.
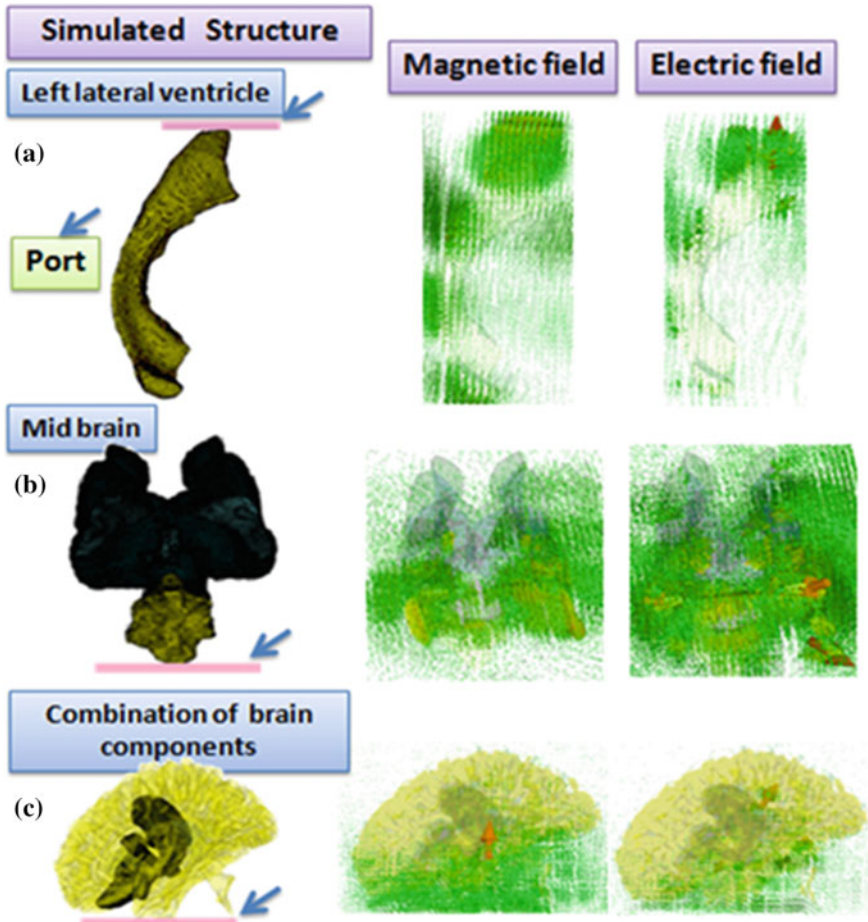
Fig. 7 **a** In case of ventricle, both the electric and magnetic fields are dominating. Both fields are propagating along the portion of ring. Left lateral ventricle; Port$=-6.0e+004 \times -6.00e+004$ nm; Frequency range$=1$–10 THz; Boundaries condition$=$open space; Used solver$=$Time domain solver; Magnetic and electric field distribution profile at resonant frequency$=2.04$ THz. **b** Electric field and magnetic field profile of midbrain by varying the location of the energy source around it. Simulation frequency range$=0$–3 THz; Boundaries condition$=$open space; **Used solver $=$** Time domain solver; Electric and magnetic field distribution scale at resonant frequency$=1.52$ THz. **c** Electric field and Magnetic field profile at one particular frequency on whole brain model. Simulation frequency range$=2$–7 THz; Boundaries condition$=$open space; Used solver$=$Time domain solver; Electric and magnetic field distribution scale at resonant frequency$=4.01$ THz

2.8 Ventricle, MidBrain, and Other Brain Components

The ventricular system of the human brain is made up of four interconnected cavities (left and right lateral ventricles, third ventricle, and fourth ventricle) connected to the

location of CSF production. Cerebrospinal fluid is the colorless fluid that is produced in choroid plexus of the brain that acts as a buffer for the cortex and provides the basic mechanical production of the brain inside the skull [33, 34]. The previous study suggests that the electric field is useful for the treatment of glioblastoma by constricting a realistic model of the head by using the MRI data and simulated by using the finite element method. Here, we have also created the realistic structure of the brain component's ventricles structure by using MRI data shown in Fig. 7a; midbrain (by using lateral ventricle, left, and right ventricle) and whole brain (by using midbrain and left, and right hemispherical) Fig. 7b, c.

Study of ventricles and cavities of the brain has been particularly challenging. Figure 7a shows clearly that there is a periodic ripple of energy density, this has been the central feature of all similar structures. We are presenting this simulation result because these are the only structure in the brain, which do not show any special energy localization. Figure 7b shows the midbrain and our study shows that its multiple components share energy in a characteristic manner. One could see, just like the connectome data presented in Fig. 2, we could see that at different frequencies, localization of electric and magnetic fields changes significantly at different constituent components of the midbrain. Another very interesting note is that electric field dominates in the midbrain over the magnetic field. Figure 7c is a very interesting study to end this article. Here, we have put the connectome, white matter and the midbrain together to find intricate changes in the resonance frequencies. We also wanted to learn how these components interact together in the whole brain, and how their individual responses vary. We found that when all components are put together, energy remains confined in the lower part of the brain structure. In future, we would report how change in the port location provides directed energy transfer toward functionally significant domains.

3   Conclusion and Future

We have carried out a pure materials science investigation to find out if we have considered the brain as a consciousness less material structure alone, then what would be its properties. Two major scale-free features have been observed regarding the brain. First is that the brain's electromagnetic resonance has a pronounced resonance in the THz (10^{12} Hz) domain, if one scans from μHz (10^{-6} Hz) to the pHz (10^{15} Hz). Second, the major finding is that at all scales, magnetic energy plays much more dominating role in the structure shadowing the electric field distribution. One of the prime conclusions of this study is that nowhere in any of the components we have studied, we found electromagnetic field is distributing homogeneously.

We outline major findings for each brain components: **Neuron**: Electromagnetic energy propagation is favored just in the opposite direction than that is established over a century. **Connectome**: By selecting resonance frequency, connectome could select particular neural pathway. **White matter**: Localize energy to the cingulate gyrus as a fundamental base to select the neural transmission pathways of cortex.

Blood vessel: A pair of virtual 3D cavity of energy distribution is created by the blood vessels, one electric another magnetic, and both are complementary to each other. **Axon**: Complementary constructive and destructive interference between electric and magnetic fields enable interactive and silent mode of an axon. **Microtubule**: Electromagnetic resonance has similar silent and interactive modes like the axon; however, it largely modifies its response by editing its surface lattice. **Ventricle, midbrain, and other brain components**: Ventricles provide cavity like effect. The existences of cavities therefore prove that not only the dielectric resonators but also the ventricles make brain a fusion of dielectric and cavity resonators. Midbrain is similar to connectome topology, and both work as if they need to provide a selective energy to certain neural branches or constituent components. When we have connected all brain components, we found that the whole system goes to a silent mode. We have not figured out the reason, but we are working on to find out the reason. Based on the studies done so far, we can comment that the origin of this silent mode is that we need multiple thermal ports around the blood vessels. Then, our artificial brain shown in Fig. 7c would turn live (clocking).

References

1. Abbott, L.F.: Lapique's introduction of the integrate-and-fire model neuron. Brain Res. Bull. **50**, 303–304 (1999)
2. Hodgkin, A.L., Huxley, A.F., Katz, B.: Measurements of current-voltage relations in the membrane of the giant axon of Loligo. J. Phys. **116**, 424–448 (1952)
3. Hamill, O.P., Marty, A., Neher, E., Sakmann, B., Sigworth, F.J.: Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches. Pflügers Archiv Eur. J. Phys. **391**, 85–100 (1981)
4. Sporns, O., Tononi, G., Kotter, R.: The human Connectome: a structural description of human brain. PLoS Comput. Biol. **1**(4), 245–251 (2005)
5. Sanes, S.R., Lichtman, J.W.: Can molecules explain long term potentiation? Nat. Neurosci. Rev. **2**, 597–605 (1999)
6. McCormick, D.A., Shu, Y., Yu, Y.: Hodgkin and Huxley model—still standing? Nature **445**, E1–E2, References on challenging the Hodgkin Huxley Action Potential Initiation in the Hodgkin-Huxley Model, Lucy J. Colwell mail, Michael P. Brenner Published (2009). https://doi.org/10.1371/journal.pcbi.1000265 (2007)
7. Ghosh, S., Sahu, S., Agrawal, L., Shiga, T., Bandyopadhyay, A.: Inventing a co-axial atomic resolution patch clamp to study a single resonating protein complex and ultra-low power communication deep inside a living neuron cell. J. Integr. Neurosci. **15**(4), 403–433 (2016)
8. Striegel, D.A., Hurdal, M.K.: Chemically based mathematical model for development of cerebral cortical folding patterns. PLoS Comput. Biol. (2009). https://doi.org/10.1371/journal.pcbi.1000524

9. Hurdal, M.K., Bowers, P.L., Stephenson, K., Sumners, D.W.L., Rehm, K., Schaper, K., Rottenberg, D.A.: Quasi-conformally flat mapping the human cerebellum. In: Taylor, C., Colchester, A. (eds.) Medical Image Computing and Computer-Assisted Intervention—MICCAI'99. Lecture Notes in Computer Science, pp. 279–286. Springer, Berlin (1999)

10. Van Essen, D.C.: Cause and effect in cortical folding. Nat. Rev. Neurosci. **8**, 12 (2007)

11. Çukur, T., Nishimoto, S., Huth, A.G., Gallant, J.L.: Attention during natural vision warps semantic representation across the human brain. Nat. Neurosci. **16**, 763–770 (2013)

12. Van Essen, D.C.: A tension-based theory of morphogenesis and compact wiring in the central nervous system. Nature **385**, 313–318 (1997)

13. Noctor, S., Martinez-Cerdeno, V., Ivic, L., Kriegstein, A.: Cortical neurons arise in symmetric and asymmetric division zones and migrate through specific phases. Nat. Rev. Neurosci. **7**, 136–144 (2004)

14. Basar, E.: Chaotic dynamics and resonance phenomena in brain function: progress, perspectives and thoughts. In: Basar, E. (ed.) Chaos in Brain Function, pp. 1–30. Springer-Verlag, Heidelberg (1990)

15. Hoke, M., Lehnertz, K., Pantev, C., Lütkenhöner, B.: Spatiotemporal aspects of synergetic processes in the auditory cortex as revealed by the magnetoencephalogram. In: Basar, E., Bullock, T.H. (eds.) Brain Dynamics, pp. 84–108. Springer-Verlag (1989)

16. Liebovitch, L.S., Fischbarg, J., Konairek, J.P., Todorova, I., Mei, W.: Fractal model of ion-channel kinetics. Biochim. Biophys. Acta **896**, 173–180 (1987)

17. Ghosh, S., Sahu, S., Fujita, D., Bandyopadhyay, A.: Design and operation of a brain like computer: a new class of frequency-fractal computing using wireless communication in a supramolecular organic, inorganic systems. Information **5**, 28–99 (2014)

18. Sahu, S., Ghosh, S., Hirata, K., Fujita, D., Bandyopadhyay, A.: Multi-level memory-switching properties of a single brain microtubule. Appl. Phys. Lett. **102**, 123701.1–123701.4 (2013)

19. Sahu, S., Ghosh, S., Fujita, D., Bandyopadhyay, A.: Live visualizations of single isolated tubulin protein self-assembly via tunneling current: effect of electromagnetic pumping during spontaneous growth of microtubule. Sci. Rep. **4**, 7303 (2014)

20. Stahl, S.M.: Structure and Function of Neurons, 3rd edn. Cambridge University Press. http://assets.cambridge.org/97805218/57024/excerpt/9780521857024_excerpt.pdf

21. Xu, K., Zhong, G., Zhuang, X.: Actin, spectrin and associated proteins form a periodic cytoskeleton structure in axons. Science **339**, 452–456 (2013)

22. Sporns, O., Tononi, G., Kotter, R.: The human connectome: a structural description of the human brain. PLoS Comput. Biol. **1**(4), 245–251 (2005)

23. Hagmann, P.: From diffusion MRI to brain connectomeics. Ph.D. Thesis, Ecole Polytechnique Federale de Lausanne (2005)

24. Sporns, O.: The human connectome: a complex network. http://dx.doi.org/10.1016/S0920-9964(12)70100-7

25. Bullmore, E., Sporns, O.: Complex brain networks: graph theoretical analysis of structural and functional systems. Nat. Rev. Neurosci. **10**(3), 186–198 (2009)

26. https://www.boundless.com/psychology/textbooks/boundless-psychologytextbook/biological-foundations-of-psychology-3/structure-and-function-of-the-brain-35/cerebral-hemispheres-and-lobes-of-the-brain-153-12688/

27. Yau, K.W.: Receptive fields, geometry and conduction block of sensory neurones in the central nervous system of the leech. J. Physiol. **263**(3), 513–538 (1976)

28. Debanne, D.: Information processing in the axon. Nat. Rev. Neurosci. **5**, 304–316 (2004)

29. Waves in blood vessels. http://hplgit.github.io/INF5620/doc/pub/sphinxwave/_main_wave009.html

30. McKay, J.C., Prato, F.S., Thomas, A.W.: A literature review: the effects of magnetic field exposure on blood flow and blood vessels in the microvasculature. Bioelectromagnetics **28**(2), 81–98 (2007)

31. Sahu, S., Ghosh, S., Ghosh, B., Aswani, K., Hirata, K., Fujita, D., Bandyopadhyay, A.: Atomic water channel controlling remarkable properties of a single brain microtubule: Correlating single protein to its supramolecular assembly. Biosens. Bioelectron. **47**, 141–148 (2013)

32. Agrawal, L., Sahu, S., Ghosh, S., Shiga, T., Fujita, D., Bandyopadhyay, A.: Inventing atomic resolution scanning dielectric microscopy to see a single protein complex operation live at resonance in a neuron without touching or adulterating the cell. J. Integrat. Neurosci. **15**(4), 435–462 (2016)
33. The ventricular system is a set of hollow cavities in the brain filled with cerebro spinal fluid. https://www.boundless.com/physiology/textbooks/boundless-anatomy-and-physiology-textbook/central-nervous-system-12/protection-of-the-brain-116/ventricles-636-9194/
34. Miranda, P.C., Mekonnen, A., Salvador, R., Basser, P.J.: Predicting the electric field distribution in the brain for the treatment of Glioblastoma. Phys. Med. Biol. **59**(15), 4137–4147 (2014)

Wireless Communication Through Microtubule Analogue Device: Noise-Driven Machines in the Bio-Systems

Komal Saxena, K. V. Karthik, Suryakant Kumar, D. Fujita and Anirban Bandyopadhyay

Abstract Looking beyond ionic communication in bio-systems that is limited to a narrow band of kHz frequency domain is our objective. Microtubule, a vital subcellular biomolecule found in almost all eukaryotic living systems, plays an important role in processing the cellular information. Therefore, here we introduce a microtubule analogue device, which wirelessly communicates with the neighboring microtubules and harvest energy from the noise present in the environment. The device is composed of spatially arranged lattice geometry with two different lattice parameters made of capacitors as tubulin protein analogue arranged on cylindrical shape structures. To demonstrate that the noise is harvested, both the devices are operated by noise, no ordered signal is applied in any measurement. Separation between both the devices is varied, while nearing the distance, the transmitted signal increases continuously and when they are taken further apart, the signal decreases gradually to null at ~140 cm. We image live, the generation and transmission of magnetic flux condensate between these two devices. This experiment is also repeated by inserting a magnetic shield 99.99% pure Ni sheet between two structures, which allows very less wireless transmission due to the shielding. The wireless communication frequency is in the range kHz–MHz.

Keywords Wireless power transfer · Magnetic condensate · Heat pipe effect
Noise harvesting · Microtubule

K. Saxena · K. V. Karthik · S. Kumar · D. Fujita · A. Bandyopadhyay (✉)
Advanced Key Technologies Division, National Institute for Materials Science,
1-2-1 Sengen, Tsukuba, Ibaraki 3050047, Japan
e-mail: anirban.bandyo@gmail.com

K. Saxena
Microwave Physics Laboratory, Dayalbagh Educational Institute, Dayalbagh,
Agra 282005, Uttar Pradesh, India

1   Introduction

Wireless communication is projected to bring the next industrial revolution [1]. There are plenty of challenges and opportunities [2]. Specially, in the field of biomedical engineering, wireless nanobots can do wonderful surgery [3] and [4]. Even it is demonstrated that organic nanobots could be used to destroy cancer cells and destroy beta plaques of Alzheimer's patient [5].

Biological systems are known to transfer signals by diffusing ions and process information using chemical reactions—a major part of its interaction is played by the protein's mechanical transformation of secondary structures. Biomaterials are made up of various components each with different dielectric constants, protein resonance was first measured around 1930s [6, 7]. Therefore, they are not like a tuning fork vibrating at a single frequency; shift in resonance frequencies reveal much more than static response [8]. One could confine water molecules via hydration [9] for the resonance measurement. Resonance peaks is a set of frequency that could act as a marker for a protein [10]. Consequently during resonance measurement a protein responds in a nanocavity in a complex manner [11, 12]. Measurements were made even in superconducting cavity [13]. Dispersion relation of a material is its dielectric response as a function of frequency; it reveals how different kinds of dielectric constituents in a material dominate in absorbing energy and resonantly vibrating at different frequencies [14]. Resonance of proteins survives in living cells [15]. After 26 years, we performed a similar experiment with advanced tools on tubulin and microtubule inside a live neuron cell [16]. Cells are not just chemical fluids [17, 18]. Dielectric and electronic properties have one-to-one correspondence [19]. Ionic interactions mostly occur in the kHz frequency domain as ionic waves take a long time to clock. Since existing biology mostly focuses on the ionic and molecular interactions, it looks only into the milliseconds domain. If there are other events happening at different time domains, then molecular biology does not take into account those factors. In fact, giant protein structures responsible for ionic transfers are all made of dipoles, which vibrate at several hundreds of nanoseconds to microsecond periods. The Fourier transform of such signal shows resonance frequency peaks in the MHz frequency domain [20, 21]. However, these dipoles are made of functional groups, which vibrate in picoseconds to nanoseconds time domain; we see resonance peaks in the Fourier transform plot. We speculated that a biological phenomenon would not be limited to a narrow kHz frequency domain or milliseconds time domain. We proposed that biological systems would operate in different time domains simultaneously and those time domains are geometrically connected. In this context, we offered a resonance chain where triplet of triplet grouping of resonance frequencies would be the most predominant topology, integrating the vibrations of a biological system [22]. If that were true, multiple communication modes would operate simultaneously in a single biological system. Apart from ionic diffusion and electrical pulse passing through a solid medium, there is a possibility of wireless transmission of signals.

Wireless transmission could happen in two ways. First, an antenna and receiver system radiates and absorbs electromagnetic energy in its structure. Then, the materials involved could be dielectric resonators and/or cavity resonators. Second, transferring the magnetic fields over a short distance directly, this route is popularly depicted as Tesla coil [23]. In this route, there are two ways. First, two conducting coils are kept side by side, and current passes through them, then due to inductive coupling, the magnetic energy could transfer. Alternately, there could be a pair of parallel plates; due to capacitive coupling, they could transfer magnetic fields. Both capacitive and inductive routes are studied in detail. Both the routes require a current flow. However, we have seen that if ionic motions are restricted, biomaterials do not allow flow of current. In presence of fluids, it is impossible to study conductivity of biomaterials because ionic currents through water find a direct less resistive transmission route via water, so they reject the more difficult biomaterial route. Therefore, the observed high current (above microampere) originates from water conductivity and not from biomaterials. When we developed the trick to study biomaterials under limited water layer 3–5 nm, we found that the conductivity of a material decreases several orders of magnitude (300 GΩ–10^{15} Ω). Therefore, biological systems are nonconducting intrinsically. This finding argues that they are not fit for wireless communication in the conventional routes as they cannot allow flow of current. Neither, biomaterials could be classical antenna or receiver, nor they could be Tesla coils. Moreover, water damps the electromagnetic signals, so an electromagnetic signal cannot pass through. If there exists at all any possibility of wireless communication, the only option would be magnetic wireless transmission. However, we need to identify how a critical insulator could communicate wirelessly when it cannot allow flow of current through itself.

Recently, attempts are being made to find options of wireless communication beyond strong coupling, exploring ideas to avoid exponential decay of wireless transmission as a function of distance [24], coupling magnetic resonance [25], switching multiple coils to enhance efficiency [26], and range modulation [27]. Here, we are exploring a new idea, if ultralow power driven by noise could transmit magnetic energy between two nonconducting materials. Biomaterials are rich in spiral structures, starting from DNA, alfa-helices, microtubule, actin-like microfilaments and neurofilaments, and even collagen. They look like Tesla coils, the only problem is that they do not allow the flow of currents as we expect. Current flow through the biomaterials is of the order of a few picoamperes even if 1 V is applied across the device. It is nearly impossible to investigate a biomaterial from all directions; they have a few nm-wide sizes, and hence accessing different parts simultaneously is nearly impossible. For that reason, we make 10^6 times larger replica of microtubule, predicted to process information in 1982 [28] and verified in 2013 [29], keeping its structural parameters identical to its biological counterparts. Then, we carry out rigorous investigations to find out if there is at all any possibility that wireless electromagnetic or magnetic communication could happen. One problem is that a microtubule is made of several tubulin proteins arranged all along the surface of an ordered water crystal core [21]. Therefore, we need to find an analogue of protein constituents for its artificial replica. We do it simply by taking capacitors. Even we keep the value of

the capacitor 106 times (microtesla) the capacitance of a tubulin protein, which is ~0.1 pT. In Sect. 2, the formulation for analyzing the different parameters is given. In Sect. 3, the experimental setup to demonstrate the noise is harvested through a microtubule analogue device and the design is discussed. In Sect. 4, various results based on the harvesting noise and generations of magnetic flux are discussed.

2 Formulation and Design Concept

2.1 The Concept of Microtubule Analogue

Preparation of an analogue microtubule has been patented (an inductor made of arrayed capacitors [30]). By wiring capacitors, which serve as an analogue of a protein, we solder them in a sequence keeping a distance around 1 cm and the long linear array of capacitors is rolled spirally along a cylinder surface keeping a pitch 1.03 cm accurately as shown in Fig. 1a. Thus, we keep pitch-to-diameter ratio as 1.03:1.00, which is abundantly used in the biomaterials as we have seen theoretically from the crystal structure data available for multiple spiral biomaterials. Now, we have reported earlier that microtubule switches its lattice parameters, following which we have replicated two lattice structures in two devices. It is known that a pair of identical coils would make maximum wireless power transfer. In this case, we are not interested in an absolute power transfer, but how exactly power transfer takes places, if at all any. From that context, we need microtubule analogues of different lattices.

There are two unique novelties in our work. First, our device is a junk material as far as electronics is concerned. If we try to pass low-frequency signals or near dc current, it faces an infinite resistance in a capacitor. However, if the frequency increases, the dielectrics switches polarize synchronously with the variation of the applied ac signal passing the wave. Therefore, up to a very high frequency, our device is simply a junk material. Moreover, as we connect a large number of capacitors in series, the lattice network also works as a capacitor and that blocks certain frequency range of ac signals further. Eventually, we end up in creating a system that blocks both ac and dc signals largely. This disadvantage could turn into an advantage. Blocking electrons critically may even help us in generating a field of a new kind. Second, we never use ac or dc signals from highly stable sources, but the noise generated from various noise generators is used for the measurement of signal transmission. Noise has profound effects on the capacitors.
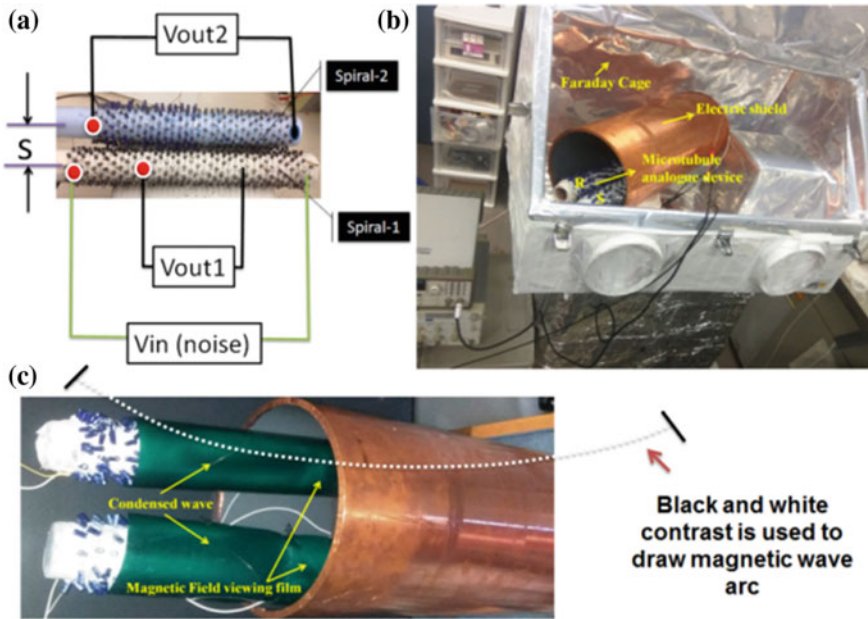
Fig. 1 **a** A pair of microtubule analogue device made of arrayed capacitors with diameter around 5.5 cm and length 30 cm. Vin is the circuit connection through which noise is applied as an input (20–80 mV). Vout is the wiring by which output voltage or signal or charge content measurement is carried out in the device. Another device is kept isolated and Vout is measured as ac signal through function generator, **b** entire device is kept inside a multilayered electromagnetic shielding chamber. Then, for critical measurement to confirm that the truly magnetic condensate is transmitting wirelessly from Spiral 1 to Spiral 2, we used another Cu chamber shown in the panel (**c**). On the green magnetic sensor sheet, if we keep long time, say hours (use new sheet for every experiment, because first experiment is very sensitive, and then there is a permanent hysteresis effect on the sheet), then the magnetic condensate wave is visible. The condensate looks like a waveform as we have shown using a dotted line how the arc looks like, and from the curvature of the condensed waveform we can calculate the wavelength of the condensed wave

2.2 The "Heat Pipe" Effect

Our target is to achieve "heat pipe effect" in the cylindrical surface. In a "heat pipe" effect, [31] a hollow pipe is filled with dielectric, where physical motion of quanta by radiation and transmission is associated with the phase ripples, where no physical objects move. Since the phase part transmits in parallel, the conductivity could reach 100 times that of silver [32]. Capacitors are well studied by exhibiting the heat pipe effect [33]. The phase transition of electrolyte works together to efficiently transfer heat between two neighboring capacitors and in its dielectric between two leads. Water channel increases the conductivity through microtubule by ~10^3 times, the effect of capillary water is significant in heat pipe effect [34]. Phase-assisted heat transfer is wireless and it does not get affected by the path resistance. Additionally,

a spiral path helps in thermal energy storage [35] and [36]. Therefore, we carefully select noise such that it is just sufficient to deplete the capacitors and engage the dielectric in rapid phase transition, so that three events happen. They are: (i) electronic motions are critically arrested, (ii) rapid phase transition increases the phase-assisted heat transfer, and (iii) electrons are critically arrested within the capacitor blocking leakage largely transforming the material from an extreme insulator (10^{12} Ω) to a super nonconductor (10^{15} Ω). The heat pipe effect-based transfer increases inversely with the depletion of electrons. These two fundamentals are background works for the finding presented here.

All along this study, we have been careful to use simple methods so that our striking finding that challenges the very idea of electronics is reproduced by school students everywhere. For that reason, we have bought market available magnetic sensor sheets and wrap up the devices, both the sources, where we apply a noise and the receiver, where we measure a wireless transmission. The magnetic sensor papers accumulate the magnetic moments from the surface of the artificial devices and provide direct evidence that a condensed magnetic wave has been created both in the source and in the destination device. We have used multilayered electromagnetic shields that block major parts of the external signal in a Faraday box. In addition, the study has been looking into the power transfer, estimating the magnetic flux, and finding the relationships between the magnetic condensed waves formed in the source and the destination.

3 Theory of Wireless Transmission

We built a preliminary theory of the phenomenon. Here, we briefly note a few aspects of it. Microtubule undergoes rapid phase changes on its surface. We can consider its surface as a composition of lattices of various lattice parameters and sizes. At a particular resonance frequency, there is an equilibrium, we get $N_r^{equ} = N_0 \exp\left(-\frac{\Delta G_{cluster}(r)}{k_B T}\right)$, this statistically accounts for the composition of lattice symmetries at resonance, where

$$\Delta G_{cluster}(r) = -\Delta G_{lc,v} \cdot \pi r^2 l_{cluster} + \sigma \pi r^2 l_{microtubule}. \qquad (1)$$

Here, $\Delta G_{lc,v}$ is due to phase change per unit cluster area on the microtubule surface and σ is the interfacial free energy propagating throughout the microtubule surface due to THz transmissions between tubulin proteins or equivalent capacitors. From Eq. (1), we get $\Delta G_{critical} = \frac{16\pi\sigma^3}{3\Delta G_{lc,v}^2}$, at radius of cluster of a particular lattice configuration would increase ($r > r_c$), i.e., greater than a critical radius, and Eq. (1) predicts a condensation above this limit. Condensation means that a particular lattice starts dominating its surface. The rate of condensation of a typical symmetry in a receiver microtubule's surface is given by

$$G = S_c k N_0 \frac{1}{N_c} \left(\frac{\Delta G_c}{3\pi k_B T} \right)^{1/2} \exp\left(-\frac{\Delta G_c}{k_B T} \right). \tag{2}$$

Here, N_c is the number of lattice points that first synchronizes its phase in the sender with the receiver's magnetic condensed wave's phase. Now, we need to find phase gain rate g_s from the lattice gain rate G. We have explained it in the following.

To understand the wireless communication between the microtubule analogue devices, we have inherently considered the coupled mode theory model of two resonant systems for the analysis. According to this coupled mode theory, these two resonant systems act as a source and receiver as shown in Fig. 1. The source system has a resonant frequency of ω_s. The overall system gain rate is expressed as $G = g_s - \gamma_s$, where g_s is the phase gain rate of the source and γ_s is the intrinsic loss rate of the source. The receiver lattice has a resonant frequency of ω_r. The overall loss rate of the system is expressed as $\gamma = \gamma_r + \gamma_{rl}$, where γ_r is the intrinsic loss rate of the receiver and γ_{rl} is the loss rate at receiver load. Power is transmitted from source to receiver with a transmission coupling coefficient rate α, which is a function of signal amplitude. Transmission coupling coefficient decreases exponentially as the separation distance between source and receiver increases. Here, we have supplied the noise signal to the source. If V_{noise} is the noise signal and V_r is output voltage of receiver, then the rate of change in signals is described by the following equation:

$$\frac{d}{dt} \begin{bmatrix} V_{noise} \\ V_r \end{bmatrix} = \begin{bmatrix} i\omega_s + g_s & -i\alpha \\ -i\alpha & i\omega_r - \gamma \end{bmatrix} \begin{bmatrix} V_{noise} \\ V_r \end{bmatrix}. \tag{3}$$

In Sect. 4, we report all the parameters' results.

4   Avoiding Artifacts

One prime source of artifacts is the environmental noise, as we deal with femto to autto watts. Wireless communications are happening between the source and the destination alone, to prove this, we need to take several precautions, human presence must be avoided, as human body acts as a giant capacitor and movement of a human body transmits power by discharging and charging capacitors. Any laboratory is full of machines, computers to characterization systems, and they radiate huge electromagnetic signals in the entire frequency range. Prior to the experiment, we made an electromagnetic-IR map of the whole room identifying the silent domain and carried out the work.

We used multiple layers of the following seven kinds of materials to create our own ultralow-noise zero Gauss chamber. The chips are kept inside zero Gauss chamber, inside a three-layered Faraday cage. The entire Faraday cage system shown in Fig. 1 is oriented toward the earth's magnetic field. A 2D inductive layer is added on the chamber to trap the ac noise.

(1) MCF7 layer (LF magnetic field, 30 dB (97%)); attenuation HF: 40 dB at 1 GHz, thickness 0.02 mm; permeability: $\mu4 = 25{,}000$; $\mu max = 100{,}000$; saturation polarization: 0.55 T; and composition: Co69, Fe4, Mo4, Nb1, Si16, and B7. The surface is grounded and it is conducting. (2) Mu layer: Nickel permalloy foil. (3) MCL61 layer: permeability $\mu2 = 10{,}000$; $\mu4 = 25{,}000$; $\mu max = 100{,}000$; saturation polarization: 0.55 T; Hc = 0.5 A/m; remnant Br/Bs = 0.7; and Curie temperature Tc = 225 °C. (4) EMF–RF and Ni–Cu ripstop shielding from Faraday defense, which is an electromagnetic shielding. (5) Pure Cu shield from Faraday defense. (6) Brass and aluminum metal sheet was used as core structure of the primary Faraday cage inside which a secondary cage was built. Ar gas 99.99% dry, atmosphere was created between primary and secondary cages. (8) ESD/EMP 7.0MIL material from Faraday defense was custom made at home from DIY kits, using multi-layered alternate aluminum and polyester/polyethylene coating as part of primary cage. Two aluminum metal layers in this five-layered bag provide maximum EMP bag protection,—33 lb puncture resistance (FTMS 101-C, 2065.1),—>40 db EMI attenuation (MIL-PFR-81705-REV.D),—surface resistivity (12 Ω/Sq). In (ASTM 1-257),—7 mil thickness (MIL-STD-3010C Method 1003),—moisture barrier (MIL-STD-3010C Method 3030), and heat sealing conditions:—temperature (400 °F, 204 °C),—time (0.6–4.5 s),—pressure (30–70 PSI, 206–482 kPa).

Finally, we studied wireless communication femtowatt (mV, pA) to auttowatt (mV, fA) following various types of measurements and measuring the same event following independent routes. We used amplifiers followed by power meters and sensors, apart from Smith charts of VNA to see if there is a real wireless power transmission. When a large number of capacitors are connected in series, it is expected that no current would flow, and therefore the devices are junk materials. Observation of wireless communication is remarkable and unprecedented.

5   Result and Discussion

Figure 1a shows the experimental setup and two spiral devices are made of 22 μF capacitors. We varied capacitance from 1–100 μF, changed diameters, length, pitch, but for clarity, in this first report, we confine into only one kind of device. One important aspect of this setup is that V_{in} and V_{out} are connected to different instruments (e.g., connect both to an oscilloscope, vector network analyzers, impedance analyzers, spectrum analyzers, etc.) to find the nature of signal produced by noise V_{in} and the transmitted signal. The Cu tube is rotated in Fig. 1b so that the devices kept inside are visible. A pair of analogue microtubule devices kept inside the Cu cylinder is enlarged in Fig. 1c and we can see that both the sender and the receiver devices are wrapped with the magnetic sensor sheet. One peculiarity of this sensor sheet is that it accumulates magnetic flux as the time pass by, it means if the magnetic condensate formed in the device is stable over time, the sheet would integrate and deliver the cumulative response.
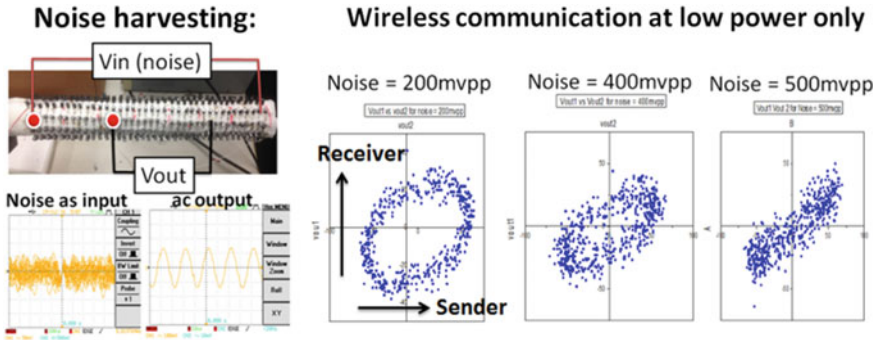
Fig. 2 Two panels are plotted side by side. In the left panel, there are three subpanels, one above shows the measuring circuit and the pair of panels below are the screenshots of the oscilloscope when the system was pumped with random noise (left) as Vin and output are taken as Vout, which is extremely low power (femtowatt) but ordered. This is an evidence that the arrayed capacitors in a lattice form are not a junk device. In the extreme right, three panels are indicating the Lissajous figures created during wireless communication between a pair of microtubule analogue devices. X-axis of an oscilloscope is fed with Vout1 and Y-axis of that oscilloscope is fed with Vout2. Together, they create Lissajous figures. An ellipse in the first panel suggests in phase transfer of wireless signal. Noise is increased 200–500 mV, and the transmitted signal nearly disappears. The unique phenomenon is observed only under noise

Figure 2 shows the noise harvesting basic experiment. Here, noise between 20 and 80 mV is applied in the 1 kHz–50 MHz range (Vin) and then the output is measured (Vout), which is a regularized ordered signal. The electronically junk device that does not have the ability to transmit electron converts noise of a particular type into an ordered signal. This is the most effective form of noise harvesting known to us. Three researchers were not told about the incredible phenomenon and were asked to repeat the experiments, between 2015 and 2017, and the three students reproduced the typical wireless communication parameters in a triple blind experiment. One of the most profound evidence that a unique wireless transmission is taking place in this electronically junk spiral topology is the three right panels of Fig. 2. Here, we have taken output signals from the source and the receiver and feed them in the X and Y axes of the oscilloscope to produce the Lissajous figure with 45° phase difference tells us about the phase locking transport of the ac signal. Now, conventional machines are not equipped with probes to determine the nature of the signal being transported. However, pure magnetic wave is really impossible to find. Only feasible easy to comprehend situation would be, if the magnetic signal is transmitted, which sends information of phase and then the signal rebuilds in the sender. Due to the magnetic nature of signal, phase locking is visible.

To confirm that the transmitted signal is magnetic in nature, we have inserted a pure magnetic metal shield of Ni sheet, and multiple other types of magnetic field absorbers and found that the transmitted signal is affected (Fig. 3a). The separation between the two coils is increased when the signal decreases to a threshold low value, but not zero. Currently, we are investigating distance independent transmission that
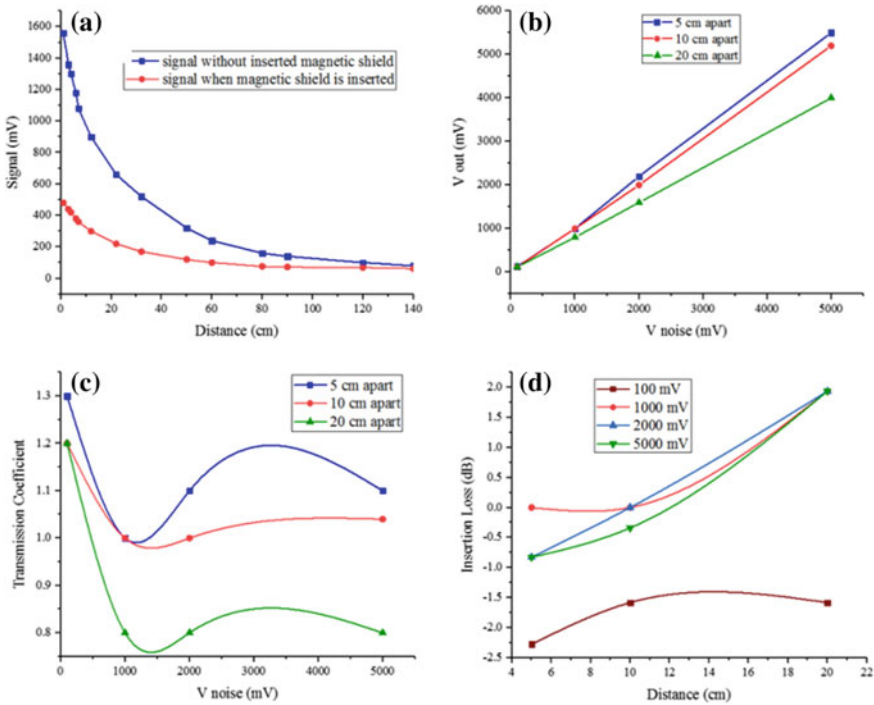
Fig. 3 **a** Two plots (blue, higher signal), when there is no magnetic shielding between the pair of devices that undergo wireless communication tests and when a shield is applied (red, lower signal). **b** Vout or transmitted power variation as a function of input noise for three different separations of sender and receiver coil. The amplitude applied to the device is much higher (around 5 V, because we tested the possibility of wireless communication with high power when device performance degrades due to hysteresis). **c** Transmission coefficient (ratio between output and input) is plotted as a function of noise input for three different separations. **d** Insertion loss (dB) as a function of the separation between the spiral coils for different noise inputs

is a possibility as observed by parity-based wireless communication researchers in other systems, or this is merely an artifact. Nevertheless, we confirm that power does get transmitted magnetically and wirelessly. Figure 3b is an interesting observation, here, a peculiar type of a nearly distance independent feature of wireless transmission is seen at low noise. Readers should note that this particular phenomenon of wireless transfer occurs best at less than 100 mV as shown in Fig. 2, all data in Fig. 3 are taken after tens of hours of continuous monitoring, to get very good image of condensate in our magnetic sensor.

Waiting for several hours provides a very good image of a standing wave of magnetic flux, and one could see transmission between coils even if noise is switched off, which is remarkable. However, the device characterization becomes difficult due to fatigue. Figure 3c is one example of a pair of saturated device undergoing operations for tens of hours. We can see a Gaussian-like response as a function of

Vin or Vnoise, it means even at a very high bias, there is a limiting domain where the phenomenon is observed. But, always, for the best response, a set of fresh devices work on 20–80 mV (shorting does not help in old devices). We have observed similar Gaussian optimization at several domains.

Exponential decay with distance suggests radiative transmission which varies in an inverse square manner with the distance. Figure 3c illustrates the noise voltage versus transmission coefficient graph. The transmission coefficient is the ratio of the transmitted output voltage to the applied noise voltage. Transmission coefficient more than 1 means that after reaching the destination the transmitted power increases. This is another remarkable feature of this new kind of wireless transmission. We activated both the device with noise and then mixed particular ac signal with the noise, even then, the receiver magnifies the signal. It means there is a clear harvesting of noise by the capacitor carpet on the cylindrical surface.

Figure 3d shows a new kind of optimization, which shows a peculiar feature at 100 mV. Here, it is also seen that at 100 mV the transmission coefficient is always greater than 1. Therefore, 100 mV may be considered as an operating voltage of noise, where at a particular separation, there is maximum wireless transmission and there is a domain where increasing the separation increases the transmission. This is possible if phase-assisted transport is happening during wireless transmission, and there is an interference of waves in the post-transmission scenario. One possibility would be phase-assisted transport of THz sensors, and using GaInN diode we have detected the transmission of THz signals across the capacitors. However, real thermal flow is trivial to be regulated in a nonlinear fashion, and the only possibility would be assisting in synchronizing all the capacitors on the cylindrical surface.

We have carried out detailed lattice variation study by changing the capacitor arrangement on the cylindrical surface. Joule heating of a capacitor is maximum during a dc current flow. However, under ripples of an ac signal or noise, a capacitor rapidly heats up and cools down. Rapid heating/cooling leads to a rapid phase transition inside the capacitor, like a heat pipe. Rate of phase transition could modulate conductivity by ~10^3 orders of magnitude [37]. The effects, phase regulation, and automated heat stabilization [38] enable "heat pipe" capacitors as noise harvesting system. In a microtubule, the arrayed proteins are rolled on a water crystal, it enhances the magnetic flux condensation, depicting the phase transfer-based transmission. Microtubule could act as a heat pipe under noise of particular (10–200 MHz; 5–70 GHz) frequency bandwidth, where water crystal activates.

Getting back to Fig. 3d, since it illustrates the distance between microtubule analogue devices versus insertion loss, which tells that how much signal is lost resulting from the insertion of a device, we see a gain instead of loss. Insertion loss of a two-port device can be calculated by the following equation: insertion loss (dB) $= -20\log|$Vout/Vnoise$|$. It is clear from Fig. 3d that there is a less signal loss at 100 mV input noise voltage. Insertion loss is associated with the deformation of the magnetic wave formed by the condensation of the magnetic wave. Figure 4a illustrates the input noise voltage versus wavelength of condensed wave that we have calculated from the arc, which we measure directly from the helical coil surface. Condensed wave is observed by the magnetic field viewing film, which is placed on
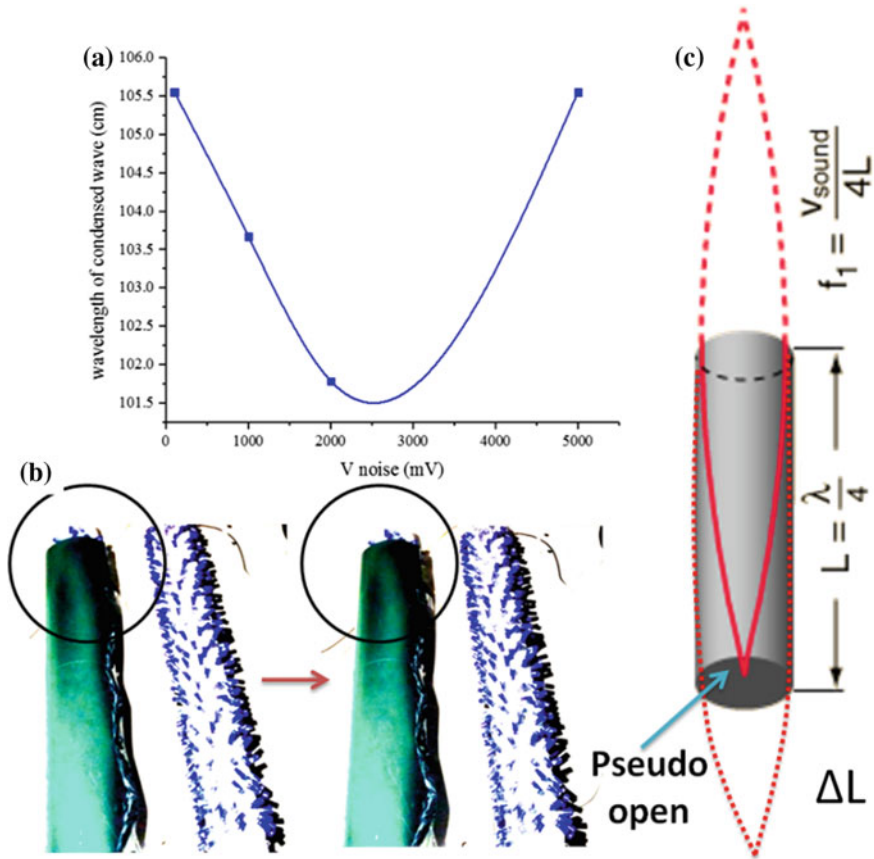
Fig. 4 **a** From the magnetic condensed wave visualized on the magnetic sensor sheet, arc is depicted as shown in the panel Fig. 1c, wavelength is calculated, which could be tuned using a noise amplitude as plotted. **b** Condensate is imaged live for hours, apparently, there is no change, but if there is an animated video, then we can see the change in the waveform shape. **c** Schematic presentation of the condensed wave formed in the microtubule analogue device, we see the condensed wave as if the device is very long (which is not real), and a part of the standing wave is cutoff. We do not know the reason, but it enables the device to generate waves much longer than the device's dimension

the microtubule analogue devices (shown in Fig. 1). More the noise bias is increased, the wavelength decreases and this means that the device squeezes the waveform to accommodate more magnetic energy on the cylindrical surface. This is a clever, intuitive, and a brilliant demonstration of energy management. We plan to explore this particular technology of energy management while developing the commercial version of noise harvesting devices. Topology of the lattices made of capacitors on the cylindrical surface has one unique advantage. There are plenty of different symmetries simultaneously superposed.

We repeat the context of coexistence of multiple symmetries on the cylindrical surface. In the nearest neighborhood scenario, we get one lattice symmetry made by the lattice points or capacitor location. However, there is a cable that is helically passing through noise from the beginning to the end of the array. A dc signal is blocked as soon as the first capacitor is charged, an ac signal would be blocked due to various capacitors formed by the network of wiring throughout the cylindrical surface, but the noise cannot be stopped. Therefore, noise plays a vital role in activating majority of the capacitors rapidly charging and discharging, which is also the foundation of the heat pipe effect. At the same time, noise-induced signals get regularized and several different reflected waves form and flow throughout the cylindrical surface. And we have compiled a random database of such waves only to conclude that the system generates various different waveforms. If we connect both the sender and receiver to the spectrum analyzer and observe 1 kHz–6 GHz simultaneously, we can identify a bunch of peaks not just one. Peaks are around 10–300 MHz. Here, in Fig. 4a when we are presenting the wavelength 100 cm it means we are arguing for around 300 MHz wave.

Another interesting observation that we made was that the magnetic wave that we image does not use the cylindrical surface boundary as a cavity, nor does it use it as a dielectric resonator. Large number of resonance peak and missing part of the waveform pointed out in Fig. 4b and is explained with a simple schematic in Fig. 4c , which suggests that the condensed magnetic wave originates from topology, and a hierarchical integration of phase is not limited to the dimension of the device.

6   Conclusion

We have shown here the magnetic condensate formation and its wireless transfer between a pair of analogue microtubule devices. There would be plenty of applications if a microscale version of this device is made, and one application would be the biomedical engineering and wireless power transfer technology for biomedical implants. In future, we will study modulation of the core of the cylinder with non-magnetic conducting elements mimicking the water crystal located at the core of a microtubular structure. Water has unique dielectric property [39]. We have already demonstrated that without the water channel microtubule loses all its remarkable electronic features [21]. Pentagon rings of water are fundamental to microtubule core, but are fundamental to other proteins too [40]. Water layer around protein molecule is essential and its dielectric property plays a vital role [41]. Protein hydration might have an impact on the wireless transmission and water molecules make an integral part of the molecule [42]. Buried water molecules hold protein's energy landscapes [43].

16-1-0003 (2016–2019) on the electromagnetic resonance-based communication and intelligence of biomaterials.

References

1. Simic, M., Bil, C., Vojisavljevic,V.: Investigation in wireless power transmission for UAV charging. Proc. Comput. Sci. **60**, 1846–1855 (2015)
2. Jawad, A.M., Nordin, R., Gharghan, S.K., Jawad, H.M., Ismail, M.: Opportunities and challenges for near-field wireless power transfer: a review. Energies **10**(1022), 1–28 (2017)
3. Ho, J.S., Yeh, A.J., Neofytou, E., Kim, S., Tanabe, Y., Patlolla, B., Beygui, R.E., Poon, A.S.Y.: Wireless power transfer to deep-tissue microimplants. PNAS **111**, 7974–7979 (2014)
4. Jiang, H., Zhang, J., Lan, D., Chao, K.K., Liou, S., Shahnasser, H., Fechter, R., Hirose, S., Harrison, M., Roy, S.: A low-frequency versatile wireless power transfer technology for biomedical implants. IEEE Trans. Biomed. Circuits Syst. **7**, 526–535 (2013)
5. Ghosh, S., Chatterjee, S., Roy, A., Ray, K., Swarnakar, S., Fujita, D., Bandyopahyay, A.: Resonant oscillation language of a futuristic nano-machine-module: eliminating cancer cells & Alzheimer Aβ plaques. Curr. Topic. Med. Chem. **15**, 534–541 (2015)
6. Zein, I., Wyman, J.: Studies on the dielectric constant of protein solutions. J. Biol. Chem. **76**, 443–476 (1931)
7. Elliott, M.A., Williams, J.W.: The dielectric behavior of solutions of the protein zein. J. Am. Chem. Soc. **61**, 718–725 (1939)
8. Vollmer, F., Braun, D., Libchaber, A.: Protein detection by optical shift of a resonant microcavity. Appl. Phys. Lett. **80**, 4057–4059 (2002)
9. Schirò, G., Cupane, A., Vitrano, E., Bruni, F.: Dielectric relaxations in confined hydrated myoglobin. J. Phys. Chem. B **113**, 9606–9613 (2009)
10. Kim, Y.-H., Yoon, S-II, Park S., Kim, H-S, Kim, Y-J, Jung, H.-I.: A simple and direct biomolecule detection scheme based on a microwave resonator. Sens. Actuat. B **130**, 823–828 (2008)
11. Lu, Q., Shu, F.-J., Zou, C.-L.: Dielectric Bow-tie nanocavity. Opt. Lett. **38**, 5311–5314 (2013)
12. Verma, R., Daya, K.S.: Rapid detection of pM concentration of insulin, using microwave whispering gallery mode. IEEE Sens. J. **17**, 2758–2765 (2017)
13. Zhai, Z., Kusko, C., Hakim, N., Sridhar, S., Revcolevschi, A., Vietkine, A.: Precision microwave dielectric and magnetic susceptibility measurements of correlated electronic materials using superconducting cavities. Rev. Sci. Instrum. **71**, 3151–3160 (2000)
14. Hanham, S.M., Watts, C., Otter, W.J., Lucyszyn, S., Klein, N.: Dielectric measurements of nanoliter liquids with a photonic crystal resonator at terahertz frequencies. Appl. Phys. Lett. **107**(032903), 1–5 (2015)
15. Burdette, E.C., Cain, F.L., Seals, J.: In vivo probe measurement technique at VHF through microwave frequencies. IEEE Trans. Microw. Theozy Tech. **28**, 414–427 (1980)
16. Ghosh, S., Sahu, S., Agrawal, L., Shiga, T., Bandyopadhyay, A.: Inventing a co-axial atomic resolution patch clamp to study a single resonating protein complex and ultra-low power communication deep inside a living neuron cell of integrative. Neuroscience **15**, 403–433 (2016)
17. Glanz, J.: Force-carrying web pervades living cell. Science **276**, 678–679 (1997)
18. Pokorný, J.: Endogenous electromagnetic forces in living cells: implications for transfer of reaction components. Electro-Magnetobiol. **20**, 59–73 (2001)
19. Pethig, R.: Dielectric and electronic properties of biological materials **139**, p. 376. Wiley, Chichester and New York (1979)
20. Sahu, S., Ghosh, S., Fujita, D., Bandyopadhyay, A.: Live visualizations of single isolated tubulin protein self-assembly via tunneling current: effect of electromagnetic pumping during spontaneous growth of microtubule. Sci. Rep. **4**(7303), 1–9 (2014)

21. Sahu, S., Ghosh, S., Ghosh, B., Aswani, K., Hirata, K., Fujita, D., Bandyopadhyay, A.: Atomic water channel controlling remarkable properties of a single brain microtubule: correlating single protein to its supramolecular assembly. Biosens. Bioelectron. **47**, 141–148 (2013)
22. Ghosh, S., Sahu, S., Fujita, D., Bandyopadhyay, A.: Design and operation of a brain like computer: a new class of frequency-fractal computing using wireless communication in a supramolecular organic, inorganic systems. Information **5**, 28–99 (2014)
23. Tesla, N.: Apparatus for transmitting electrical energy. US patent 1,119,732 (1914)
24. Assawaworrarit, S., Yu, X., Fan, S.: Robust wireless power transfer using a nonlinear parity-time-symmetric circuit. Nature **546**, 387–390 (2017)
25. Kurs, A., Karalis, A., Moffatt, R., Joannopoulos, J.D., Fisher, P., Soljačiic, M.: Wireless power transfer via strongly coupled magnetic resonance. Science **317**, 83–86 (2007)
26. Nair, V.V., Choi, J.R.: An efficiency enhancement technique for a wireless power transmission system based on a multiple coil switching technique. Energies **9**(156), 1–15 (2016)
27. Sample, A.P., Meyer, D.A., Smith, J.R.: Analysis, experimental results, and range adaptation of magnetically coupled resonators for wireless power transfer. IEEE Trans. Industr. Electron. **58**, 544–554 (2011)
28. Hameroff, S.R., Watt, R.C.: Information processing in microtubule. J. Theor. Biol. **98**, 549–561 (1982)
29. Sahu, S., Ghosh, S., Hirata, K., Fujita, D., Bandyopadhyay, A.: Multi-level memory-switching properties of a single brain microtubule. Appl. Phys. Lett. **102**(123701), 1–4 (2013)
30. Sahu, S., Fujita, D., Bandyopadhyay, A.: Inductor made of arrayed capacitors. 2010 Japanese patent has been issued on 20th August 2015 JP-511630 (world patent filed, this is the invention of fourth circuit element), US patent has been issued 9019685B2, 28 April (2015)
31. Heat pipe effect was invented long back by Gaugler, R.S., 1942; further improved by Grover, G., 1963
32. Catalog of Indek Corporation, 1998 holds this particular fact
33. Gasperi, M.L., Gollhardt, N.: Heat transfer model for capacitor banks. In: 33rd Annual Meeting of the IEEE IAS (1998)
34. Wang, Y., Gundevia, M.: Measurement of thermal conductivity and heat pipe effect in hydrophilic and hydrophobic carbon papers. Int. J. Heat Mass Trans. **60**, 134–142 (2013)
35. David, A.: Reay: thermal energy storage: the role of the heat pipe in performance enhancement. Int. J. Low-Carbon Technol. **10**, 99–109 (2015)
36. Zalba, B., Marin, J.M., Cabeza, L.F., Mehling, H.: Review on thermal energy storage with phase change materials, heat transfer analysis and applications. Appl. Therm. Eng. **23**, 251–283 (2003)
37. Udell, K.S.: Heat transfer in porous media considering phase change and capillarity—the heat pipe effect. Int. J. Heat Mass Trans. **28**, 485–495 (1985)
38. Parler, S.G.: Deriving life multipliers for Aluminum electrolytic capacitors. IEEE Power Electron. Soc. Newslett. **16**, 11–12 (2004)
39. Buchner, R., Barthel, J., Stauber, J.: The dielectric relaxation of water between 0 °C and 35 °C. Chem. Phys. Lett. **306**, 57–63 (1999)
40. Teeter, M.M.: Water structure of a hydrophobic protein at atomic resolution: pentagon rings of water molecules in crystals of crambin. Proc. Natl. Acad. Sci. U.S.A. **81**, 6014–6018 (1984)
41. Ebbinghaus, S., Kim, S.J., Heyden, M., Yu, X., Heugen, U., Gruebele, M., Leitner, D.M., Havenith, M.: An extended dynamical hydration shell around proteins. Proc. Natl. Acad. Sci. U.S.A. **104**, 20749–20752 (2007)
42. Otting, G., Liepinsh, E., Wuthrich, K.: Protein hydration in aqueous solution. Science **41**, 974–980 (1991)
43. Denisov, V., Peters, J., Hörlein, H.D., Halle, B.: Using buried water molecules to explore the energy landscape of proteins. Nature. Struct. Biol. **3**, 505–509 (1996)

Characterization of Stochastic Resonance by the Discrete General Beta Distribution

J. E. Lugo, R. Doti and Jocelyn Faubert

Abstract In this work, we use the Discrete General Beta Distribution (DGBD) to characterize stochastic resonance (SR). We present an SR model that uses the balance of energy (gain versus losses) between two signals, namely, a weak harmonic signal plus noise at different levels. The energy balance is done in two different states (representing below and above sensitivity threshold). The states are characterized by two wells divided by a barrier which are mathematically described by an asymmetrical Duffing potential. In order to have transitions between states, the presence of noise is necessary. With this model, the activity of neurons below and above the sensitivity threshold states is simulated and the time series is generated. Fourier analysis of these time series results in the well-known inverse-U-shaped behavior between signal and noise that characterizes SR. In the second part of the work, we ranked the values from the time series, simulated with the SR model, in descendent order from high to low values and then we calculate the fitting parameters for the DGBD for each noise condition and show that SR could also be characterized by DGBD parameters.

Keywords Stochastic resonance · Discrete general beta distribution
Transitions · Neurons

1   Introduction

The phenomenon known as stochastic resonance (SR) [1] is a nonlinear mechanism by which the inclusion of a stochastic signal achieves the identification of feeble stimuli. The maximum benefit is attained at some optimal noise level. Beyond that optimal level, noise intensity only deteriorate detection or the content of information is obtained. The SR signature is a weak harmonic stimulus that is a function of different noise levels with the shape of an inverted U. The SR phenomenon is ubiquitous in this universe including biology and technology. SR occurs in different

J. E. Lugo (✉) · R. Doti · J. Faubert
Faubert Lab, School of Optometry, Université de Montréal, Montréal H3C3J7, Canada
e-mail: eduardo.lugo@gmail.com

nano-physical systems [2], micro [3] and macro [4]. The brain is not the exception, and it also exists in several studies showing that the central part of the nervous system utilizes the noise to improve the detection of sensations [1, 5–10]. A clear example is the so-called crossmodal SR, which is omnipresent in persons involving distinguishable cortical regions and peripheral subsystems. For instance, using the similar auditory noise levels, the crossmodal SR has been found in proprioceptive, tactile, and visual sensorial systems [11, 12]. For instance, sensation transitions are defined by the fluctuation from subthreshold actions to a synchronized firing action in multisensory neurons (synchronized firing at a rate given by the harmonic signal). At first, the energy of their actions (conveyed by the feeble signals) is not sufficient to be registered. Once the stochastic signal permeates the central part of the nervous system, it creates a collective actuation between multisensory neurons, altering their initial operation. This collective behavior promotes perceptiveness transitions and the signals are therefore detected. In other words, the action generated by the mutual influence of the facilitating signal (for instance, auditory noise) and the excitatory signal (for example, tactile) at certain power level creates the conditions for the central part of the nervous system to discern of a contrariwise feeble signal.

Currently, a vast degree of investigation is committed to power law comportments, particularly concerning complex networks [13]. Nonetheless, in the most favorable cases, the power law tendency is reasonable but merely for a middle sequence of figures; the power law always breaks at the ends of data distributions [14]. From a scientific point of view, the study of both, the tail mathematical function and their breakdowns are of importance [15–19]. In the rank-ordered distributions, often connected to cumulative distribution functions, certain elements of a structure are assorted increasingly or downwardly consistent with its significance best known as rank. Recently, the Discrete General Beta Distribution (DGBD) was used to match a surprising number of diverse phenomena such as neuroscience, music, ecology, painting, genetics, urbanism, and social networks [13]. The DGBD uses three parameters and contains the product of two power laws competing again each other. The whole dataset defines both power laws. The work presented in [13] shows that the DGBD is a universal function that appears in apparently unrelated phenomena. Moreover, guided by their observations, they implemented a conflicting dynamic model that generates the same distribution. The model helped out identifying essential features beneath the distribution. In such example, the fitting parameter values of the distribution change accordingly with the state of the system dynamics. That is, the study suggests the possibility to distinguish between a disordered state and an ordered one. So, the DGBD can be used as a tool to track down systems' transitions from one state to another. This is the feature that will be used to characterize SR in this work.

This work is organized as follows: first, we describe an SR model; second, the DGBD is presented; third, we explain the mathematical methods used here to characterize SR; fourth, we present the results and discuss them; and finally, we give some conclusions and future work.

2 Neuronal Stochastic Resonance Mathematical Model

To model neuronal stochastic resonance, we will use a model that treats neurons dynamically consisting of random movements having low-amplitude levels (sub-threshold neuronal action), where evades occur at some time span [20]. The evades are best known as firings which are related with spikes. In this way, the effective multisensory neurons' responses are determined by a bistable potential as in [20]. The master equation in our model is given by

$$\ddot{x} = -V'(x) + \varepsilon\big[\gamma\,\mathrm{Cos}(\omega_0 t) + \sigma\,G(t) - \beta\dot{x}\big], \tag{1}$$

where the magnitude of the actions of neurons is characterized by x, $\dot{x}$ is the velocity magnitude of the actions of neurons, $V(x)$ is a double-well potential given by $V(x) = \alpha\big(-x^2/2 + x^4/4\big)$, ε is a readjustment parameter. $\mathrm{Cos}(\omega_0 t)$ is the weak excitatory signal representation, $G(t)$ represents the facilitation signal. It may be deterministic or stochastic signal, like white noise, and the parameters γ, σ, and β are adjustable. Inside brackets, we find the excitatory, facilitation, and energy losses, respectively. Deterministic signals have a repeating pattern or have limited bandwidth. Stochastic signals are represented by a random pattern and have a broad bandwidth like white noise [11].

We can transform the second-order Eq. (1) into a set of two first-order equations as

$$\begin{aligned} &x \leq 0 \\ &\dot{x} = x_1, \\ &\dot{x}_1 = -^-V'(x) + \varepsilon\big[\gamma\mathrm{Cos}(\omega_0 t) + \sigma\,G(t) - \beta x_1\big], \end{aligned} \tag{2a}$$

$$\begin{aligned} &x > 0 \\ &\dot{x} = x_1 \\ &\dot{x}_1 = -^+V'(x) \end{aligned}, \tag{2b}$$

where $^\pm V'(x) = {}^\pm\alpha\big(-x + x^3\big)$.

With the appropriate parameter values, Eqs. (2a) and (2b) simulate neuronal firings, and it gives solutions that have real aspects detected in the experimentation explained before. To accomplish useful neuronal firings simulations, the potential $V(x)$ is asymmetric, that is deeper for $x > 0$ than for $x \leq 0$.

3 The Discrete General Beta Distribution

Rank-ordered relations are used to observe how in a process some given properties decrease [15, 21]. A usual example is the Zipf law [13], which describes the frequency of using words in some particular language and it is represented by a power law

behavior. Nonetheless, this power behavior works only in the middle-rank range [22]. Cocho et al. showed that a vast number of diverse systems cannot be described by power laws, but they can be characterized by the following general rank-ordered distribution:

$$F(R) = A(n + 1 - R)^b / R^a, \tag{3}$$

where R is the rank number, n is the maximum number of values, A is a constant, and (a, b) two fitting exponents. This relation is a noncontinuous description of the continuous random variable generalized beta distribution, and it is called DGBD [23]. Moreover, in [13], a mathematical scheme is also presented that includes mutation-duplication characteristics and creates data fulfilling Eq. (3). They found that parameter a is associated with persistence. On the contrary, the parameter b is related to transformation. Nevertheless, the contest among transformation and persistence seems to be important, although not necessary feature.

4 Methods

4.1 Stochastic Resonance

Using the following parameters $^-\alpha = 1$, $^+\alpha = 49$, $\beta = 0.316$, $\omega_0 = 2\pi/10$, $\gamma = 0.095$, Gaussian-distributed noise with mean σ with standard deviation of 0.15σ, and $\sigma \in [0.28, 0.34, 0.4, 0.46, 0.52]$, we solved the set of Eqs. (2a), (2b) in MATLAB for each noise at (σ) level. Five simulations were run for each noise level. The simulation time was 25 s in steps of half a second. Initial conditions at $t = 0$, where $x(0) = 0$ and $x_1(0) = 0$. Once the time series $x(t)$ were generated, a fast Fourier transform algorithm was performed with 512 points for each 25 time series to locate the amplitude of the weak harmonic signal with frequency ω_0. These amplitudes were then averaged for each noise level. A plot was generated between each averaged amplitude and its corresponding noise level.

4.2 Discrete General Beta Distribution

For each noise level, the five time series were rank ordered in a descending fashion (from high amplitude to low amplitude). For the fitting, we have again used MATLAB, where we have performed the method of nonlinear least squares. We utilized a nonlinear regression model of Eq. (3), the dependent variable or the response is the amplitude of $x(t)$, the independent variable is the rank order r, and the nonlinear parameters were A, a, and b. To determine the nonlinear parameter estimates, we used the function $y_i = P(R_i, A, a, b) + \varepsilon_i$, where y_i, R_i, and ε_i represent the i-th

Fig. 1 Inverse-U-shaped behavior of the signal amplitude at 0.1 Hz with different noise levels
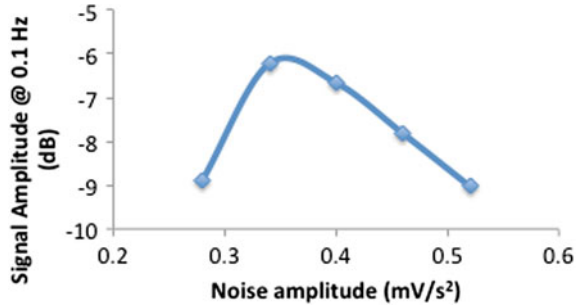
numerical xi amplitude value, rank, and residual or error. The function that is minimized is given by $\sum_{i=1}^{n} (y_i - P(R_i, A, a, b))^2$. Once parameters A, a, and b are known for each noise level, a plot was generated between each parameter value and its corresponding noise level.

5 Results

5.1 Stochastic Resonance

Figure 1 displays the established SR surjection; the greatest apex (resonance peak) is reached at $\sigma = 0.34$. Clearly, neuron activity amplitudes in the Fourier domain at a frequency of 0.1 Hz are modulated by the noise intensity.

5.2 Discrete General Beta Distribution

Figure 2 shows the behavior of parameters A, a, and b with different noise levels. Strikingly, parameters A are minimum and b are maximum at resonance (noise level $\sigma = 0.34$). Parameter a almost has the same value that parameter b at resonance.

6 Discussion

The mathematical model represented by the set of Eqs. (2a) and (2b) describe the well-known phenomenon of stochastic resonance. Clearly, it can be seen that at weak noise levels, the energy exchange between noise and signal is not sufficient to accomplish synchronicity. As an output, the subthreshold actions of neurons dominate and no firing occurs. Nonetheless, if the noise level is increased, then the firing is intensified

Fig. 2 (Top) DGBD parameter **a** behavior with different noise levels. (Bottom) DGBD parameter **a** and **b** behavior with different noise levels
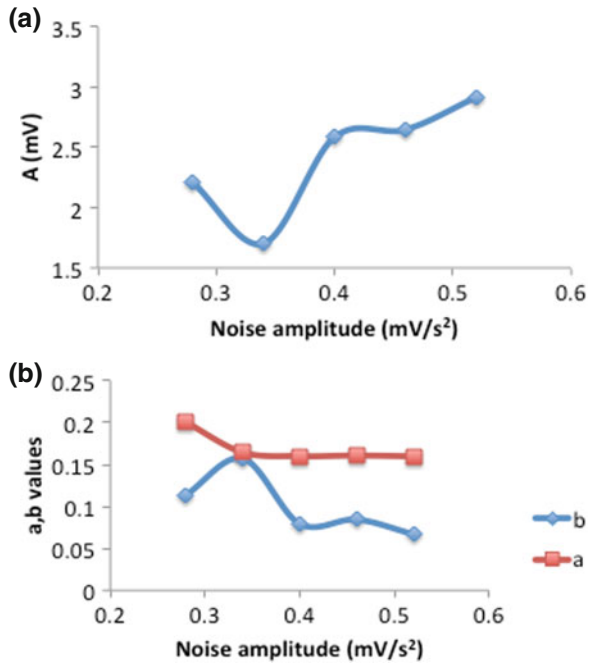
till it reaches a supreme. In that condition, the frequency of the signal pretty much matches the mean evasion rate. Above that condition, the stochastic firing happens at distinct frequencies. This means that there is a desynchronized power exchange between signal and noise, and the signal is buried below the action thresholds of the neurons. In summary, neuronal SR can be characterized by this model but the question now is the possibility to characterize it by other means. The answer is yes, as it is clearly demonstrated in Fig. 2. Parameters A from the DGBD can be used for this purpose; it presents a minimum exactly where the resonance peak appears. Parameter A represents a balance point between both functions forming the DGBG, namely, $g_1(R) = (n + 1 - R)^b$ and $g_2(R) = R^a$. For the range of values, we found for both parameters a and b, the function $g_1(R)$ decreases monotonically while $g_2(R)$ is a monotonically increasing function. The point value where both functions meet gives the value for A. On the contrary, parameter b values are smaller than a values except at the resonance peak noise value, where both values are almost the same. This condition might indicate a "permanence" transition according to [13]. That is, the probability that neuronal activity remains below the threshold is high for low-noise intensities but at the point where $a = b$, the probability to change from subthreshold activity to synchronize firing activity increases (synchronize firing at a rate given by the harmonic signal). As the noise is increased, the subthreshold activity is dominated again over unsynchronized firing activity. From this standpoint, the resonance condition is given when A is minimum or $a = b$, thus in that condition both functions $g_1(R)$ and $g_2(R)$ meets near to the midpoint of the DGBD at rank

number $R = (n+1)/2$, which is similar to the classical median concept in a frequency distribution of observed values or quantities.

7 Conclusions

We have shown in this work that the neuronal stochastic resonance can be characterized by the General Discrete Beta Distribution presented in [13]. The DGBD depends on three fitting parameters namely A, a, and b. We used a mathematical model for neuronal activity that presented the SR. That is, the SR model showed the traditional U-inverse shape between amplitude of the harmonic signal and noise. The model used the balance of energy (gain versus losses) between a weak harmonic signal and noise at different levels. The energy balance is done in two different states. One state represents the subthreshold activity and the other represents the suprathreshold activity (neuron firing). The states are characterized by two wells divided by a barrier which are mathematically described by an asymmetrical Duffing potential. The transitions between states are facilitated by the noise. SR can be characterized by using parameter A, which has a minimum value at the resonance peak (maximum peak in the U-inverse graph). This is equivalent to the condition where the other two parameters from the DGBD are equal. Mathematically, this is equivalent to find the median point of a frequency distribution of observed values. From this perspective, the resonance point is a transition point and actually the only point where neuronal activity transits from below threshold to synchronized suprathreshold activity (synchronize firing at rate given by the harmonic signal). The use of this SR characterization is straightforward. One can use the time series only and no Fourier or other mathematical analysis are required and just the DGBD fitting. From these results and the results from [13] we can envisage a method to characterize any other system that we know a priori and it transits from at least one state to a second one. For instance, think of human learning and how neuronal networks connect to a point where a person is capable of learning something. The emergence of these new learning states implies the existence of transition points that could be characterized by this distribution. In the future, some psychophysical experiments will be done to show this new way to characterize SR.

References

1. Moss, F., Ward, L., Sannita, W.: Stochastic resonance and sensory information processing: a tutorial and review of application. Clin. Neuro-physiol. **115**, 267–281 (2004)
2. Badzey, R.L., M, P.: Coherent signal amplification in bistable nanomechanical oscillators by stochastic resonance. Nature **437**, 995–998 (2005)

3. Simon, A.J., Libchaber, A.J.: Escape and synchronization of a brownian particle. Phys. Rev. Lett. **68**(23), 3375–3378 (1992)
4. Benzi, R., Stuera, A., Vulpiani, A.: J. Phys. A **14**, L543 (1981)
5. Harper, D.W.: Signal detection analysis of effect of white noise intensity on sensitivity to visual flicker. Percept. Mot. Skills **48**, 791–798 (1979)
6. Manjarrez, E., et al.: Effects of auditory noise on the psychophysical detection of visual signals: cross-modal stochastic resonance. Neurosci. Lett. **415**, 231–236 (2007)
7. Simonotto, E., et al.: Visual perception of stochastic resonance. Phys. Rev. Lett. **78**, 1186–1189 (1997)
8. Hidaka, I., Nozaki, D., Yamamoto, Y.: Functional stochastic resonance in the human brain: noise induced sensitization of baroreflex system. Phys. Rev. Lett. **85**, 3740–3743 (2000)
9. Collins, J.J., Imhoff, T.T., Peter, G.: Noise-mediated enhancements and decrements in human tactile sensation. Phys. Rev. E **56**, 923–926 (1997)
10. Kitajo, K., et al.: Behavioral stochastic resonance within the human brain. Phys. Rev. Lett. **90**, 218103 (2003)
11. Lugo, E., Doti, R., F, J.: Ubiquitous crossmodal stochastic resonance in humans: auditory noise facilitates tactile, visual and proprioceptive sensations. PLoS ONE **3**(8), e2860 (2008)
12. Lugo, J.E., et al.: Multisensory integration central processing modifies peripheral systems. Psychol. Sci. **19**, 989–997 (2008)
13. Martınez-Mekler, G., et al.: Universality of rank-ordering distributions in the arts and sciences. PLoS ONE **4**(3), e4791 (2009)
14. Newman, M.J.: Power laws, pareto distributions and Zipf's law. Contemp. Phys. **46**, 323–351 (2005)
15. Barabasi, A.-L., Albert, R.: Emergence of scaling in random networks. Science **286**, 509–512 (1999)
16. Sornette, D.: Critical markets crashes. Phys. Rep. **378**, 1–98 (2003)
17. Montroll, E.W., Shlesinger, M.F.: Maximum entropy formalism, fractals, scaling phenomena, and 1/f noise: a tale of tails. J. Stat. Phys. **32**, 209–230 (1983)
18. Hong, H., Ha, M., Park, H.: Finite-size scaling in complex networks. Phys. Rev. Lett. **98**(258701), 1–4 (2007)
19. Laherrere, J., Sornette, D.: Stretched exponential distribution in nature and economy: "fat tails" with characteristic scales. Eur. Phys. J. B **22**, 525–539 (1998)
20. Simiu, E.: Chaotic Transitions in Deterministic and Stochastic Dynamical Systems. Princeton University Press, New Jersey (2002)
21. Albert, R., Barabasi, A.-L.: Statistical mechanics of complex networks. Rev. Mod. Phys. **74**, 47–97 (2002)
22. Ferreri Cancho, F., Sole, R.: Two regimes in the frequency of words and the origins of complex lexicons: Zipf's law revisited. Quant. Linguist. **8**, 165–173 (2001)
23. McDonald, J.B.: Some generalized functions for the size distribution of income. Econometrica **52**, 647–664 (1984)

StaR: An EEGLAB Framework for the Measure Projection Toolbox (MPT) Statistical Analyses to be Performed in *R*

Yannick Roy, Jean-Claude Piponnier and Jocelyn Faubert

Abstract EEGLAB, a widely used toolbox in MATLAB (The Mathworks, Inc.), uses Independent Component Analysis (ICA) to decompose the EEG signal into sub-signals, and localizes brain sources of those sub-signals prior to independent component (IC) clustering for group study. In 2013, the Measure Projection Toolbox (MPT) was introduced as a new data-driven IC clustering toolbox for EEGLAB. Despite the numerous features and advantages offered by EEGLAB and the MPT, they both have limitations for statistical analyses with more than two independent variables. In order to work around those limitations, this paper introduces StaR, an EEGLAB framework for the MPT statistical analyses to be performed in *R*. StaR initially exports the data from different clusters generated by the MPT for different measures of interest (e.g., Event-Related Potentials (ERPs) and Event-Related Spectral Perturbations (ERSPs)) and formats the data such that further statistical analyses can be performed in *R*. Once in *R*, StaR uses linear mixed models as its default method to better handle missing values and intra-subject variability. Finally, StaR brings the results back into MATLAB to plot the results with the well-known and easy to interpret EEGLAB graphics. To make the whole process easy, StaR also offers an intuitive user interface that integrates into EEGLAB's menu.

Keywords EEG · EEGLAB · Measure projection toolbox · Mixed models · *R*

Y. Roy (✉) · J.-C. Piponnier · J. Faubert
Visual Psychophysics and Perception Laboratory, École d'Optométrie,
Université de Montréal, Montréal H3C3J7, QC, Canada
e-mail: yannick.roy@umontreal.ca

J.-C. Piponnier
e-mail: jean-claude.piponnier@umontreal.ca

J. Faubert
e-mail: jocelyn.faubert@umontreal.ca

1   Introduction

EEGLAB [1] is an open-source toolbox running in MATLAB (The Mathworks, Inc.) made for EEG signal analysis and conceived to go beyond traditional mean peak analysis. One of its major preprocessing steps is the Independent Component Analysis (ICA) [2]. A dipole fitting plugin, DIPFIT [3], is then used to locate the corresponding potential brain sources (modeled as dipoles) in a brain template.

EEGLAB uses the Independent Component (IC) clustering approach for multi-subject analysis in the brain source domain. The number of adjustable parameters available in its GUI to cluster the ICs might, however, introduce subjectivity into the process. To address this issue, Bigdely-Shamlo, Mullen, Kreutz-Delgado, and Makeig (2013) developed a plugin for EEGLAB, called the Measure Projection Toolbox (MPT), which proposes a simpler and objective clustering procedure.

Even though EEGLAB and MPT offer performing algorithms intended for processing individual datasets and for IC clustering, they both have some important limitations when it comes to creating and running multi-subject studies with complex statistical designs such as longitudinal designs.

These limitations are (1) EEGLAB does not allow one to create a "studyset" (dataset gathering all the individual datasets) including different ICA decompositions for a given subject; (2) the STUDY feature in EEGLAB, which handles the statistical modules for multi-subject comparisons (i.e., "study design"), which does not support more than two independent variables; (3) the MPT statistical feature only supports two independent variables (*Group* and *Dynamic*); (4) in the MPT, once the domains (i.e., clusters) are computed, only two types of univariate analysis are available (*t*-test and permutation); and (5) in the MPT, in the case of comparison of two different groups for one condition, missing values are replaced with zeros (0) in order to keep the number of datasets equal in both groups, and in the case of comparison of two different conditions for both groups gathered together, the datasets with even a single missing value are completely removed from the statistical analysis, which can have a significant impact on the results.

Since its release in 2013, the MPT has been used in only a few published studies [4–8]. We believe that the limitations mentioned above are one of the reasons why few researchers have used it.

As a solution to these limitations, we developed a framework for EEGLAB called StaR (Statistics in *R*). StaR provides a complete pipeline extracting the data from EEGLAB, the MPT, and MATLAB altogether, in order to perform statistical analyses in *R* framework, and bringing the results back in MATLAB to easily plot them in various ways using the StaR UI that leverages the EEGLAB graphics [9–12].

2 Methodology

This section describes the StaR framework that we have developed. The code can be downloaded from https://github.com/FaubertLab/StaR. Questions can also be sent to yannick.roy@umontreal.ca. If we were to receive positive feedback and requests to provide a complete toolbox, we would happily seek funding to do so.

2.1 StaR Framework Steps

Steps #1 and #2 consist of preprocessing the data and creating a studyset in EEGLAB, then performing the clustering in the MPT (see EEGLAB (sccn.ucsd.edu/wiki/EEGLAB) and MPT wiki pages (sccn.ucsd.edu/wiki/MPT).

Step #3 consists of exporting the data from the MPT to *R*. StaR tags missing values with a NaN (Not a Number) value when exporting them from the MPT.

Step #4 consists of creating the complex data frame in *R*.

Step #5 consists of carefully placing the MPT values in the long data structure previously created. The dimensionality of the data will automatically be switched around by StaR to facilitate its usability by the user and developer while also optimizing for parallel computing. In order to study the EEG signal in greater detail, the different measures of interest (e.g., ERP, ERSP, etc.) are analyzed in a point-by-point fashion (i.e., considering each data point constituting a given graphical representation of the measure for each subject). Thus, for each subject, the data frame may hold tens of thousands of values, and hence there is the necessity for parallel computing.

Step #6 and #7 consist of the actual statistical analysis and post hoc tests. We opted for the linear mixed model procedure (*lme4* package) to capture the within-subjects variability and to use as much of the data as possible, instead of removing subjects having missing values [13, 14]. Nevertheless, other tests can be used in *R* with small modifications in the code. Options of correction methods for multiple comparisons are also offered such as False Discovery Rate (FDR) [15] or Bonferroni.

Step #8 consists of bringing the results back into MATLAB to use EEGLAB graphics that are easy to recognize and interpret [10, 16]. One of the challenges here was the poor protocol of exchange between *R* and MATLAB. The data had to be linearized by carefully reshaping the matrices.

Step #9 consists of plotting the results. Handling more than two independent variables is not possible in EEGLAB and the MPT, therefore, even if the graphics are generated with EEGLAB functions, having the ability to plot this kind of graphics from a combination of three or four independent variables is new. The illustrations presented here come from a subset of datasets used in a longitudinal study initially comprising of four factors (*Group*, *Session*, and two condition-related crossed factors such as *Dynamic and Modulation*). In order to better handle the missing values, we have included a cluster-related factor (*Domain*).
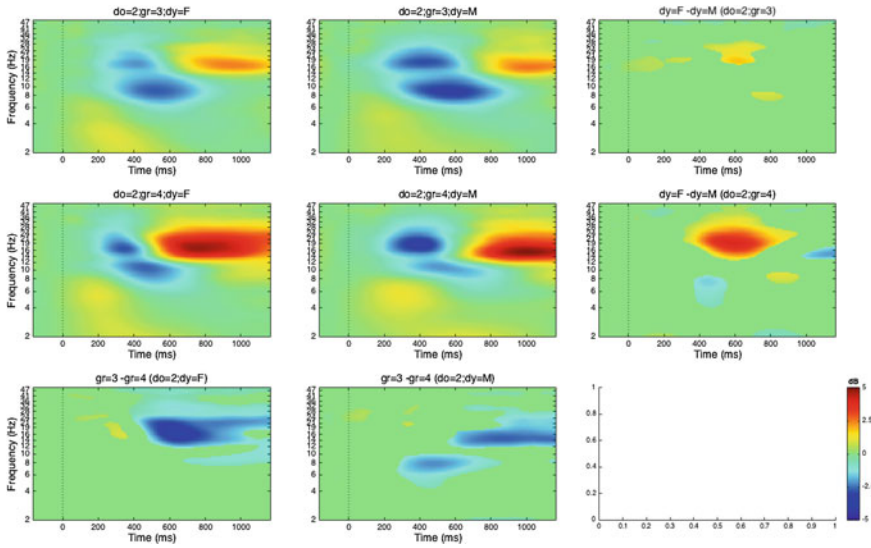
Fig. 1 ERSP 2 × 2. In this figure, the two factors are *Group* (2 levels: 3 and 4) and *Dynamic* (2 levels: M and F, where M stands for motion and F for flicker) for *Domain 2*. The effect of the factor *Dynamic within* a specific group is figured in the first two rows and the effect of the factor *Group* within a specific condition is figured in the first two columns. In the outer graphs, the green regions mean that it is not significant and the values are set to 0

Figure 1 shows an example of the data and statistical results for a given cluster (*domain 2*) from a combination of two factors (*Group* × *Dynamic*), both with two levels, while all other factors were kept combined. Any other pair of two variables could have been selected and would have produced a similar output. One could also fix the other variables to a specific value (e.g., *Session* = 1) instead of leaving them combined. It is also possible to select the option to only plot the significance mask showing only two colors (i.e., significant or not) instead of a color scale representing the difference between the two graphs.

Figure 2 shows an example of the data and statistical results for a given cluster (*domain 1*) from a combination of two factors (*Group* × *Modulation*), both with two levels, while all other factors were kept combined. Any other pair of two variables could have been selected and would have produced a similar output. One could also fix the other variables to a specific value (e.g., *Session* = 1) instead of leaving them combined.

2.2 StaR UI

Creating an intuitive user interface (UI) that makes it easy to plot different layouts of graphs based on the interaction of multiple independent variables, all with a dif-
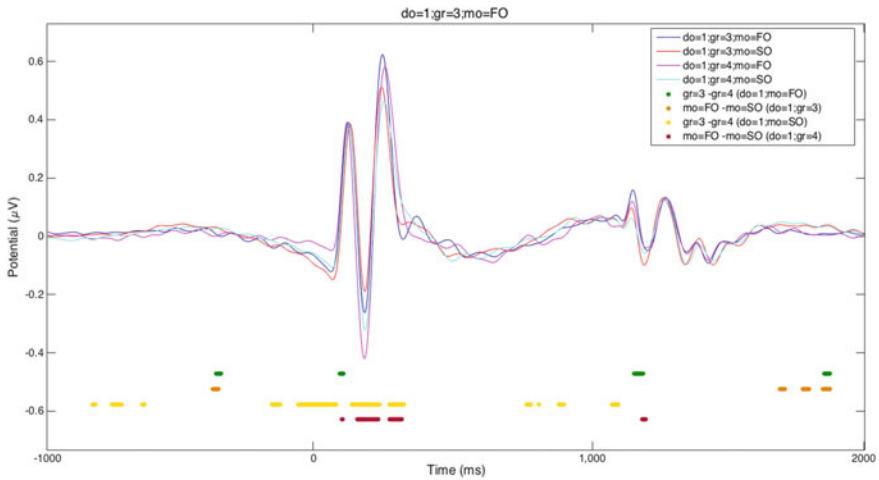
Fig. 2 ERP 2 × 2. In this figure, the two factors are *Group* (2 levels: 3 and 4) and *Modulation* (2 levels: FO, SO). Colored dots or lines below the curves indicate where the differences are significant. The legend indicates the independent variables being compared. For example, "gr = 3–gr = 4 (do = 1; mo = SO)" means that datasets from group 3 for domain 1 and second-order modulation were compared to datasets from group 4 for the same domain and modulation levels, all other factors combined

ferent number of values (e.g. 2 x 2, 2 x 3), is a challenge. Actually, when supporting multiple independent variables the combinations and possibilities quickly become manually unmanageable. The simple and intuitive StaR UI simplifies all these possible combinations and options.

2.3 Exploratory

We added a feature called *specific Dipoles* (Fig. 3) allowing the visualization of the domains generated by the MPT with the contributing dipoles from different groups and sessions represented with different colors. This feature does not compute anything but simply plots the dipoles already identified and located (i.e., source localization) with EEGLAB plug-in such as DIPFIT using a brain template such as the Montreal Neurological Institute, one prior to using the MPT and therefore before using StaR.

Such a tool gives the opportunity to quickly check if the data seem well-balanced in order to validate results. By selecting a specific domain with the groups and/or sessions of interest using different colors, one can see at a glance if (1) the groups seem to be equally represented and/or if (2) the sessions seem to be equally represented.
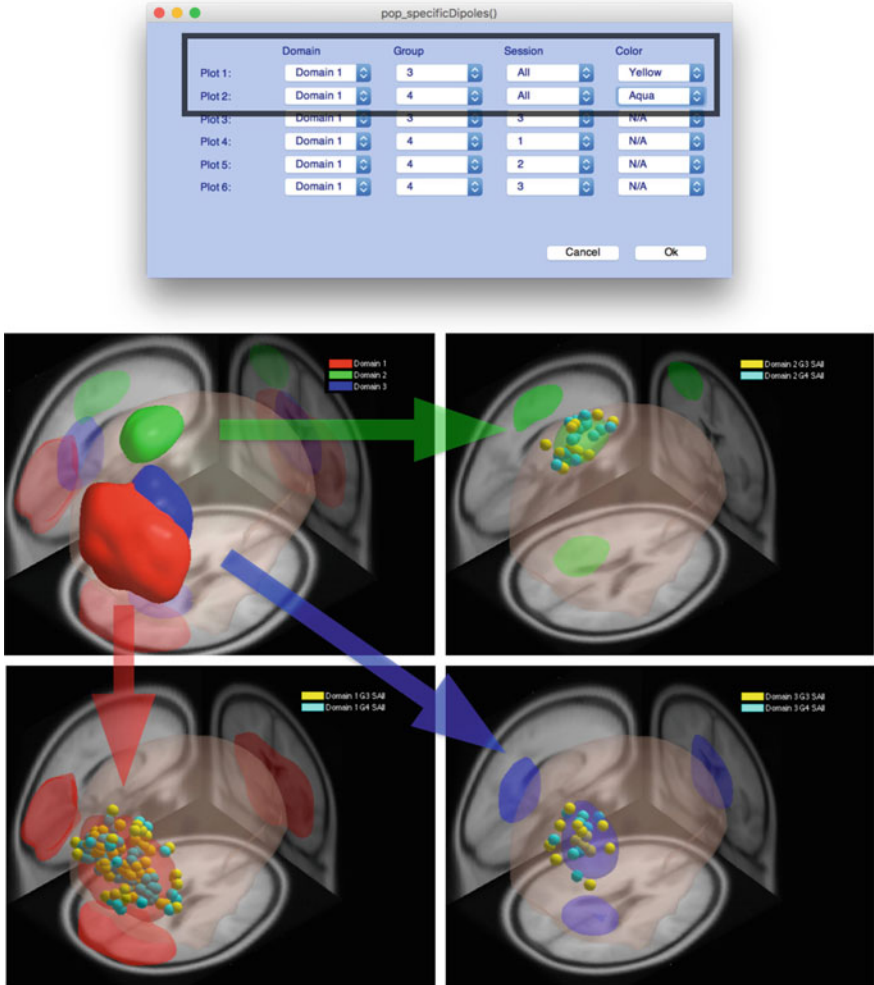
Fig. 3 *Specific Dipoles* UI and output. The UI (at the top) allows for the selection of a specific domain to show its contributing dipoles. The user can either select all groups for all sessions to see all the contributing dipoles of a specific domain, or can break down the analysis to a specific group and/or session

3 Discussion and Conclusion

StaR, the statistical framework introduced in this article is mainly targeting EEGLAB users, especially those interested in the use of the MPT as IC clustering tool. StaR creates a link between EEGLAB/MPT in MATLAB and *R*. It uses the "studyset" created in EEGLAB to perform statistical analyses of EEG signal characteristics of

interest (e.g., ERPs, ERSPs, power spectra, etc.) in *R*, and finally brings the results back into MATLAB in order to plot them using EEGLAB graphics.

The necessity for creating such a tool came from the limitations we encountered using EEGLAB and the MPT to create a longitudinal statistical study design with four factors (including two crossed condition-related factors).

The recent increase in computing power has made possible the use of more complex statistical designs like mixed-effect models, which are better at modeling within-subject variability and at dealing with missing values and unbalanced datasets than the classical ANOVA [17, 18]. Furthermore, contrary to the classical ANOVA, mixed-effect models do not violate underlying assumptions (e.g., linearity, sphericity, etc.) [19]. Else, mixed-effect models are well-suited for the analysis of longitudinal data containing missing values [20–22], and were thus particularly relevant in our case.

StaR was developed to be flexible and to easily support different statistical tests in *R* with minimum changes in the code while keeping all other steps untouched (e.g., exporting the data, plotting results with EEGLAB, etc.).

Finally, by making StaR freely accessible, we hope to encourage the research community to explore their EEG data beyond traditional ERP curves (peak amplitude and latency) obtained from the electrodes, especially by using the MPT as IC clustering tool.

References

1. Delorme, A., Makeig, S.: EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods **134**, 9–21 (2004). https://doi.org/10.1016/j.jneumeth.2003.10.009
2. Hyvärinen, A., Oja, E.: Independent component analysis: algorithms and applications. Neural Netw. **13**, 411–430 (2000). https://doi.org/10.1016/s0893-6080(00)00026-5
3. Oostenveld, R., Delorme, A., Makeig, S.: DIPFIT: equivalent dipole source localization of independent components (2003)
4. Chung, J.W., Ofori, E., Misra, G., Hess, C.W., Vaillancourt, D.E.: Beta-band activity and connectivity in sensorimotor and parietal cortex are important for accurate motor performance. Neuroimage **144**, 164–173 (2017). https://doi.org/10.1016/j.neuroimage.2016.10.008
5. Bogost, M.D., Burgos, P.I., Little, C.E., Woollacott, M.H., Dalton, B.H.: Electrocortical sources related to whole-body surface translations during a single- and dual-task paradigm. Front. Hum. Neurosci. **10**, 524 (2016). https://doi.org/10.3389/fnhum.2016.00524
6. Misra, G., Ofori, E., Chung, J.W., Coombes, S.A.: Pain-related suppression of beta oscillations facilitates voluntary movement. Cereb Cortex:bhw061. https://doi.org/10.1093/cercor/bhw061 (2016)
7. Balkan, O., Virji-Babul, N., Miyakoshi, M., Makeig, S., Garudadri, H.: Source-domain spectral EEG analysis of sports-related concussion via measure projection analysis. 2015 37th Annual International Conference IEEE Engineering Medicine Biological Society, vol. 2015, pp. 4053–4056. IEEE (2015). https://doi.org/10.1109/embc.2015.7319284
8. Misra, G., Wang, W., Archer, D.B., Roy, A., Coombes, S.A.: Automated classification of pain perception using high-density electroencephalography data. J. Neurophysiol. **117**, 786–795 (2017). https://doi.org/10.1152/jn.00650.2016
9. Delorme, A., Mullen, T., Kothe, C., Akalin Acar, Z., Bigdely-Shamlo, N., Vankov, A., et al.: EEGLAB, SIFT, NFT, BCILAB, and ERICA: new tools for advanced EEG processing. Comput. Intell. Neurosci. 130714 (2011). https://doi.org/10.1155/2011/130714

10. Onton, J., Makeig, S.: Information-based modeling of event-related brain dynamics. Prog. Brain Res. **159**, 99–120 (2006). https://doi.org/10.1016/s0079-6123(06)59007-7
11. Onton, J., Delorme, A., Makeig, S.: Frontal midline EEG dynamics during working memory. Neuroimage **27**, 341–356 (2005). https://doi.org/10.1016/j.neuroimage.2005.04.014
12. Makeig, S., Debener, S., Onton, J., Delorme, A.: Mining event-related brain dynamics. Trends Cogn. Sci. **8**, 204–210 (2004). https://doi.org/10.1016/j.tics.2004.03.008
13. Bates, D.: Computational Methods for Mixed Models, pp. 1–29 (2010)
14. Bates, D.M., Maechler, M., Bolker, B., Walker, S.: Fitting linear mixed-effects models using lme4. J. Stat. Softw. **67**, 1–48 (2015). https://doi.org/10.1177/009286150103500418
15. Benjamini, Y., Hochberg, Y.: Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. **57**, 289–300 (1995). https://doi.org/10.2307/2346101
16. Makeig, S.: Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. Electroencephalogr. Clin. Neurophysiol. **86**, 283–293 (1993). https://doi.org/10.1016/0013-4694(93)90110-h
17. Rissling, A.J., Miyakoshi, M., Sugar, C.A., Braff, D.L., Makeig, S., Light, G.A.: Cortical substrates and functional correlates of auditory deviance processing deficits in schizophrenia. Neuro Imag. Clin. **6**, 424–437 (2014). https://doi.org/10.1016/j.nicl.2014.09.006
18. Hasenstab, K., Sugar, C.A., Telesca, D., Mcevoy, K., Jeste, S., Şentürk, D.: Identifying longitudinal trends within EEG experiments. Biometrics **71**, 1090–1100 (2015). https://doi.org/10.1111/biom.12347
19. Picton, T.W., Bentin, S., Berg, P., Donchin, E., Hillyard, S.A., Johnson, R., et al.: Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. Psychophysiology **37**, 127–152 (2000). https://doi.org/10.1111/1469-8986.3720127
20. Gibbons, R.D., Hedeker, D., DuToit, S.: Advances in analysis of longitudinal data. Ann. Rev. Clin. **6**, 79–107 (2010). https://doi.org/10.1146/annurev.clinpsy.032408.153550.advances
21. Krueger, C., Tian, L.: A comparison of the general linear mixed model and repeated measures ANOVA using a dataset with multiple missing data points. Biol. Res. Nurs. **6**, 151–157 (2004). https://doi.org/10.1177/1099800404267682
22. Ibrahim, J.G., Molenberghs, G.: Missing data methods in longitudinal studies: a review. TEST **18**, 1–43 (2009). https://doi.org/10.1007/s11749-009-0138-x

Author Index